(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
01 November 2018 (01.11.2018)

WIPO | PCT

(10) International Publication Number
WO 2018/200888 A1

(51) International Patent Classification:
C12N 9/10 (2006.01) *C12P 7/22* (2006.01)
C12N 15/52 (2006.01) *C12P 17/06* (2006.01)

(21) International Application Number:
PCT/US2018/029668

(22) International Filing Date:
27 April 2018 (27.04.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/491,114 27 April 2017 (27.04.2017) US
62/569,532 07 October 2017 (07.10.2017) US

(71) Applicant: **REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, Oakland, California 94607 (US).

(72) Inventors: **KEASLING, Jay D.**; 2150 Shattuck Ave., Suite 510, Berkeley, California 94704 (US). **D'ESPAUX, Leo**; 2150 Shattuck Ave., Suite 510, Berkeley, California 94704 (US). **WANG, Jeff**; 2150 Shattuck Ave., Suite 510, Berkeley, California 94704 (US). **LUO, Xiaozhou**; 2150 Shattuck Ave., Suite 510, Berkeley, California 94704 (US). **REITER, Michael**; 2150 Shattuck Ave., Suite 510, Berkeley, California 94704 (US). **DENBY, Charles**; 2150 Shattuck Ave., Suite 510, Berkeley, California 94704 (US). **LECHNER, Anna**; 2150 Shattuck Ave., Suite 510, Berkeley, California 94704 (US).

(74) Agent: **OSMAN, Richard Aron**; 530 Lawrence Expy #332, Sunnyvale, California 94085 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

WO 2018/200888 A1

(54) Title: MICROORGANISMS AND METHODS FOR PRODUCING CANNABINOIDS AND CANNABINOID DERIVATIVES

(57) Abstract: The present disclosure provides genetically modified host cells that produce a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative. The present disclosure provides methods of synthesizing a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative.

MICROORGANISMS AND METHODS FOR PRODUCING CANNABINOIDS AND CANNABINOID DERIVATIVES

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 62/491,114, filed April 27, 2017, and U.S. Provisional Application No. 62/569,532, filed October 7, 2017, the contents of which are incorporated herein by reference in their entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

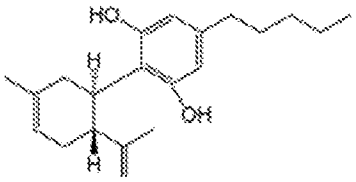
[0002] This invention was made with government support under Grant Numbers 1330914 and 1442724 awarded by the National Science Foundation. The government has certain rights in the invention.

REFERENCE TO A SEQUENCE LISTING

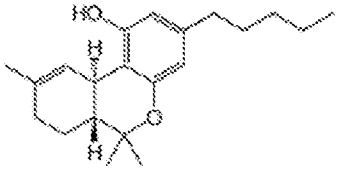
[0003] The Sequence Listing associated with this application is provided in text format in lieu of a paper copy, and is hereby incorporated by reference into the specification. The name of the text file containing the Sequence Listing is DEMT_001_00WOSeqList_ST25_txt. The text file is about 673 KB, was created on April 25, 2018, and is being submitted electronically via EFS-Web.

INTRODUCTION

[0004] Plants from the genus *Cannabis* have been used by humans for their medicinal properties for thousands of years. In modern times, the bioactive effects of *Cannabis* are attributed to a class of compounds termed "cannabinoids," of which there are hundreds of structural analogs including tetrahydrocannabinol (THC) and cannabidiol (CBD). These molecules and preparations of *Cannabis* material have recently found application as therapeutics for chronic pain, multiple sclerosis, cancer-associated nausea and vomiting, weight loss, appetite loss, spasticity, and other conditions.



Cannabidiol / CBD



Tetrahydrocannabinol / THC / Dronabinol / Marinol

[0005] The physiological effects of certain cannabinoids are thought to be mediated by their interaction with two cellular receptors found in humans and other animals. Cannabinoid receptor type 1 (CB1) is common in the brain, the reproductive system, and the eye. Cannabinoid receptor type 2 (CB2) is common in the immune system and mediates therapeutic effects related to inflammation in animal models. The discovery of cannabinoid receptors and their interactions with plant-derived cannabinoids predated the identification of endogenous ligands.

[0006] Besides THC and CBD, hundreds of other cannabinoids have been identified in *Cannabis*. However, many of these compounds exist at low levels and alongside more abundant cannabinoids, making it difficult to obtain pure samples from plants to study their therapeutic potential. Similarly, methods of chemically synthesizing these types of products has been cumbersome and costly, and tends to produce insufficient yield. Accordingly, additional methods of making pure cannabinoids, cannabinoid precursors, cannabinoid derivatives, or cannabinoid precursor derivatives are needed.

SUMMARY

[0007] The present disclosure provides methods, polypeptides, nucleic acids encoding said polypeptides, and genetically modified host cells for the production of cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives.

[0008] One aspect of the disclosure relates to a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase (GOT) polypeptide, wherein said GOT polypeptide catalyzes production of cannabigerolic acid from geranyl pyrophosphate (GPP) and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[0009] Another aspect of the disclosure relates to a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110.

[0010] One aspect of the disclosure relates to a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, the genetically modified host cell

comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100.

[0011] In certain embodiments of any of the foregoing or following, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a tetraketide synthase (TKS) polypeptide and one or more heterologous nucleic acids encoding an olivetolic acid (OAC) polypeptide, or one or more heterologous nucleic acids encoding a fusion TKS and OAC polypeptide. In some embodiments, the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76. In some embodiments, the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[0012] In certain embodiments of any of the foregoing or following, the genetically modified host cell further comprises one or more of the following: a) one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; b) one or more heterologous nucleic acids encoding a polypeptide that generates GPP; or c) one or more heterologous nucleic acids encoding a polypeptide that generates malonyl-CoA. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative, wherein the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is an acyl-activating enzyme (AAE) polypeptide. In some embodiments, the AAE polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90. In some embodiments, the AAE polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:92 or SEQ ID NO:149. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative, wherein the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is a fatty acyl-CoA ligase polypeptide. In some embodiments, the fatty acyl-CoA ligase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:145 or SEQ ID NO:147. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative, wherein the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is a fatty acyl-CoA synthetase (FAA) polypeptide. In some embodiments, the FAA polypeptide comprises an amino acid sequence having at least 50%

sequence identity to SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates GPP, wherein the polypeptide that generates GPP is a geranyl pyrophosphate synthetase (GPPS) polypeptide. In some embodiments, the GPPS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates malonyl-CoA, wherein the polypeptide that generates malonyl-CoA is an acetyl-CoA carboxylase-1 (ACC1) polypeptide. In some embodiments, the ACC1 polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207.

[0013] In certain embodiments of any of the foregoing or following, the genetically modified host cell further comprises one or more of the following: a) one or more heterologous nucleic acids encoding a HMG-CoA synthase (HMGS) polypeptide; b) one or more heterologous nucleic acids encoding a 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR) polypeptide; c) one or more heterologous nucleic acids encoding a mevalonate kinase (MK) polypeptide; d) one or more heterologous nucleic acids encoding a phosphomevalonate kinase (PMK) polypeptide; e) one or more heterologous nucleic acids encoding a mevalonate pyrophosphate decarboxylase (MVD) polypeptide; or f) one or more heterologous nucleic acids encoding a isopentenyl diphosphate isomerase (IDI) polypeptide. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an IDI polypeptide. In some embodiments, the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGR polypeptide. In some embodiments, the HMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:22. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGR polypeptide, wherein the HMGR polypeptide is a truncated HMGR (tHMGR) polypeptide. In some embodiments, the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGS polypeptide. In some embodiments, the HMGS

polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an MK polypeptide. In some embodiments, the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a PMK polypeptide. In some embodiments, the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a MVD polypeptide. In some embodiments, the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66.

[0014] In certain embodiments of any of the foregoing or following, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA. In some embodiments, the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide. In some embodiments, the acetoacetyl-CoA thiolase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25.

[0015] In certain embodiments of any of the foregoing or following, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a pyruvate dehydrogenase complex (PDC) polypeptide. In some embodiments, the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117.

[0016] In certain embodiments of any of the foregoing or following, the genetically modified host cell is a eukaryotic cell. In some embodiments, the eukaryotic cell is a yeast cell. In some embodiments, the yeast cell is *Saccharomyces cerevisiae*. In some embodiments, the *Saccharomyces cerevisiae* is a protease-deficient strain of *Saccharomyces cerevisiae*. In some embodiments, the genetically modified host cell is a plant cell.

[0017] In certain embodiments of any of the foregoing or following, the genetically modified host cell is a prokaryotic cell.

[0018] In certain embodiments of any of the foregoing or following, at least one of the one or more heterologous nucleic acids is integrated into the chromosome of the genetically modified host cell.

[0019] In certain embodiments of any of the foregoing or following, at least one of the one or more heterologous nucleic acids is maintained extrachromosomally.

[0020] In certain embodiments of any of the foregoing or following, two or more of the one or more heterologous nucleic acids are present in a single expression vector.

[0021] In certain embodiments of any of the foregoing or following, at least one of the heterologous nucleic acids is operably linked to an inducible promoter.

[0022] In certain embodiments of any of the foregoing or following, at least one of the heterologous nucleic acids is operably linked to a constitutive promoter.

[0023] In certain embodiments of any of the foregoing or following, culturing of the genetically modified host cell in a suitable medium provides for synthesis of the cannabinoid or the cannabinoid derivative in an increased amount compared to a non-genetically modified host cell cultured under similar conditions.

[0024] In certain embodiments of any of the foregoing or following, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a cannabinoid synthase polypeptide. In some embodiments, the cannabinoid synthase polypeptide is a tetrahydrocannabinolic acid (THCA) synthase polypeptide. In some embodiments, the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155. In some embodiments, the cannabinoid synthase polypeptide is a cannabidiolic acid (CBDA) synthase polypeptide. In some embodiments, the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[0025] In certain embodiments of any of the foregoing or following, the cannabinoid is cannabigerolic acid, cannabigerol, Δ^9-tetrahydrocannabinolic acid, Δ^9-tetrahydrocannabinol, Δ^8-tetrahydrocannabinolic acid, Δ^8-tetrahydrocannabinol, cannabidiolic acid, cannabidiol, cannabichromenic acid, cannabichromene, cannabinolic acid, cannabinol, cannabidivarinic acid, cannabidivarin, tetrahydrocannabivarinic acid, tetrahydrocannabivarin, cannabichromevarinic acid, cannabichromevarin, cannabigerovarinic acid, cannabigerovarin, cannabicyclolic acid, cannabicyclol, cannabielsoinic acid, cannabielsoin, cannabicitranic acid, or cannabicitran.

[0026] One aspect of the disclosure relates to a method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising: a) culturing the genetically modified host cell in a suitable medium; and b) recovering the produced cannabinoid or cannabinoid derivative.

[0027] Another aspect of the disclosure relates to a method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising: a) culturing the genetically modified host cell in a suitable medium comprising a carboxylic acid; b) recovering the produced cannabinoid or cannabinoid derivative.

[0028] One aspect of the disclosure relates to a method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising: a) culturing the genetically modified host cell in a suitable medium comprising olivetolic acid or an olivetolic acid derivative; b) recovering the produced cannabinoid or cannabinoid derivative.

[0029] Another aspect of the disclosure relates to a method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising: a) culturing a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide catalyzes production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, in a suitable medium; and b) recovering the produced cannabinoid or cannabinoid derivative.

[0030] One aspect of the disclosure relates to a method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising: a) culturing a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110 in a suitable medium; and b) recovering the produced cannabinoid or cannabinoid derivative.

[0031] Another aspect of the disclosure relates to a method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising: a) culturing a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100 in a suitable medium; and b) recovering the produced cannabinoid or cannabinoid derivative.

[0032] In certain embodiments of any of the foregoing or following, the suitable medium comprises a fermentable sugar. In some embodiments, the suitable medium comprises a pretreated cellulosic feedstock.

[0033] In certain embodiments of any of the foregoing or following, the suitable medium comprises a non-fermentable carbon source. In some embodiments, the non-fermentable carbon source comprises ethanol.

[0034] One aspect of the disclosure relates to an isolated or purified GOT polypeptide, wherein said GOT polypeptide catalyzes production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[0035] Another aspect of the disclosure relates to an isolated or purified polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110.

[0036] One aspect of the disclosure relates to an isolated or purified polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100.

[0037] Another aspect of the disclosure relates to an isolated or purified nucleic acid encoding a GOT polypeptide, wherein said GOT polypeptide catalyzes production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[0038] One aspect of the disclosure relates to an isolated or purified nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110.

[0039] Another aspect of the disclosure relates to an isolated or purified nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100.

[0040] One aspect of the disclosure relates to a vector comprising a nucleic acid encoding a GOT polypeptide, wherein said GOT polypeptide catalyzes production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[0041] Another aspect of the disclosure relates to a vector comprising a nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110.

[0042] One aspect of the disclosure relates to a vector comprising a nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100.

[0043] Another aspect of the disclosure relates to a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide catalyzes production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, into the genetically modified host cell.

[0044] One aspect of the disclosure relates to a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110 into the genetically modified host cell.

[0045] Another aspect of the disclosure relates to a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100 into the genetically modified host cell.

[0046] One aspect of the disclosure relates to a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing a vector comprising a nucleic acid encoding a GOT polypeptide, wherein said GOT polypeptide catalyzes production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82; a vector comprising a nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110; or a vector comprising a nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100, into the genetically modified host cell.

BRIEF DESCRIPTION OF THE DRAWINGS

[0047] FIG. 1 provides a schematic diagram of biosynthetic pathways for generating cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives.

[0048] FIG. 2 depicts intracellular olivetolic acid production using pathway 1a and a tetraketide synthase (TKS) polypeptide/olivetolic acid cyclase (OAC) polypeptide.

[0049] FIG. 3 depicts intracellular olivetolic acid production comparing pathway 1a and 1b.

[0050] FIG. 4 provides schematic depictions of 3 expression constructs for olivetolic acid production.

[0051] FIG. 5 provides schematic depictions of 2 expression constructs for olivetolic acid production.

[0052] FIG. 6 provides schematic depictions of 3 expression constructs for geranyl pyrophosphate (GPP) production.

[0053] FIG. 7 provides schematic depictions of 2 expression constructs for GPP production.

[0054] FIG. 8 provides schematic depictions of 3 expression constructs for cannabinoid production.

[0055] FIG. 9 depicts production of olivetolic acid using expression constructs 3 + 4 or expression constructs 3 + 5.

[0056] FIG. 10 depicts production of olivetolic acid using Construct 3 and culturing the cells in medium comprising hexanoate; or using Construct 1.

[0057] FIG. 11 is a schematic depiction of pathways for production of olivetolic acid derivatives by feeding various representative carboxylic acids, where the carboxylic acids are converted to their CoA forms by a promiscuous acyl-activating enzyme polypeptide (e.g., CsAAE1; CsAAE3), generating olivetolic acid derivatives.

[0058] FIG. 12 depicts various representative carboxylic acids with various functional groups that can be used as substrate for the biosynthesis of olivetolic acid or cannabinoid derivatives. FIG. 12 also depicts production of olivetolic acid or cannabinoid derivatives from these carboxylic acids.

[0059] FIG. 13 depicts various representative cannabinoid derivatives that can be generated by feeding different acids and the further derivatization of those derivatives with chemical reactions.

[0060] FIG. 14 depicts cannabinoid biosynthetic pathways utilizing neryl pyrophosphate (NPP) or GPP.

[0061] FIG. 15 depicts generation of cannabigerolic acid (CBGA) using a NphB polypeptide and the substrates olivetolic acid and GPP.

[0062] FIG. 16 depicts an expression construct to produce GPP.

[0063] FIG. 17 depicts an expression construct to produce hexanoyl-CoA and/or hexanoate.

[0064] FIG. 18 depicts an expression construct to produce hexanoyl-CoA and/or hexanoate.

[0065] FIG. 19 depicts an expression construct to produce olivetolic acid.

[0066] FIG. 20 depicts an expression construct to produce CBGA.

[0067] FIG. 21 depicts an expression construct to produce cannabidiolic acid (CBDA).

[0068] FIG. 22 depicts an expression construct to produce CBDA.

[0069] FIG. 23 depicts an expression construct to produce CBDA.

[0070] FIG. 24 depicts an expression construct to produce tetrahydrocannabinolic acid (THCA).

[0071] FIG. 25 depicts an expression construct to produce THCA.

[0072] FIGS. 26A, FIG. 26B, and FIG. 26C depict LC-MS traces illustrating the production of CBGA. These figures illustrate an LC-MS trace (m/z = 359.2) for ethyl acetate extraction of the yL444 strain (FIG. 26A), a 10 μM CBGA standard (FIG. 26B), and a mixture of ethyl acetate extraction of yL444 and 10 μM CBGA standard (FIG. 26C). Peaks observed at 9.2 minutes indicated the presence of CBGA in the ethyl acetate extraction of yL444.

[0073] FIG. 27 depicts the production of THCA with a THCA synthase polypeptide with an N-terminal truncation and a ProA signal sequence. The figure illustrates an LC-MS trace (m/z = 357.2144) for ethyl acetate extraction of yXL046 colony 1 (Duplicate 1, Top), yXL046 colony 2 (Duplicate 2, Middle), and a standard containing CBDA and THCA (Standard, Bottom). The peak at 7.9 mins indicated the presence of CBDA and the peak at 9.6 mins indicated the presence of THCA.

[0074] FIG. 28 depicts the production of CBDA with a CBDA synthase polypeptide with an N-terminal truncation and a ProA signal sequence. The figure illustrates an LC-MS trace (m/z = 357.2144) for ethyl acetate extraction of yXL047 colony 1 (Duplicate 1), a yXL047 colony 2 (Duplicate 2), a negative control (Negative) and a standard containing CBDA and THCA (Standard). The peak at 7.9 mins indicated the presence of CBDA and the peak at 9.6 mins indicated the presence of THCA.

[0075] FIGS. 29A and 29B depict expression constructs used in the production of the S21 strain. The expression constructs depicted in FIGS. 29A and 29B are also used in the production of following strains: S29, S31, S34, S35, S37, S38, S39, S41, S42, S43, S44, S45, S46, S47, S49, S50, S51, S78, S80, S81, S82, S83, S84, S85, S86, S87, S88, S89, S90,

S91, S94, S95, S97, S104, S108, S112, S114, S115, S116, S118, S123, S147, S164, S165, S166, S167, S168, S169, and S170.

[0076] FIGS. 30A, 30B, and 30C depict expression constructs used in the production of the S31 strain. The expression constructs depicted in FIGS. 30A, 30B, and 30C are also used in the production of following strains: S94, S95, and S97.

[0077] FIG. 31 depicts expression constructs used in the production of the S35 strain.

[0078] FIG. 32 depicts expression constructs used in the production of the S37 strain.

[0079] FIG. 33 depicts expression constructs used in the production of the S38 strain.

[0080] FIG. 34 depicts expression constructs used in the production of the S39 strain.

[0081] FIG. 35 depicts expression constructs used in the production of the S41 strain.

[0082] FIG. 36 depicts expression constructs used in the production of the S42 strain.

[0083] FIG. 37 depicts expression constructs used in the production of the S43 strain.

[0084] FIG. 38 depicts expression constructs used in the production of the S44 strain.

[0085] FIG. 39 depicts expression constructs used in the production of the S45 strain.

[0086] FIG. 40 depicts expression constructs used in the production of the S46 strain.

[0087] FIG. 41 depicts expression constructs used in the production of the S47 strain.

[0088] FIGS. 42A, 42B, and 42C depict expression constructs used in the production of the S49 strain.

[0089] FIGS. 43A, 43B, and 43C depict expression constructs used in the production of the S50 strain.

[0090] FIGS. 44A, 44B, and 44C depict expression constructs used in the production of the S51 strain. The expression constructs depicted in FIGS. 44A, 44B, and 44C are also used in the production of following strains: S78, S80, S81, S82, S83, S84, S85, S86, S87, S88, and S89.

[0091] FIG. 45 depicts expression constructs used in the production of the S78 strain.

[0092] FIG. 46 depicts expression constructs used in the production of the S80 strain.

[0093] FIG. 47 depicts expression constructs used in the production of the S81 strain.

[0094] FIG. 48 depicts expression constructs used in the production of the S82 strain.

[0095] FIG. 49 depicts expression constructs used in the production of the S83 strain.

[0096] FIG. 50 depicts expression constructs used in the production of the S84 strain.

[0097] FIG. 51 depicts expression constructs used in the production of the S85 strain.

[0098] FIG. 52 depicts expression constructs used in the production of the S86 strain.

[0099] FIG. 53 depicts expression constructs used in the production of the S87 strain.

[00100] FIG. 54 depicts expression constructs used in the production of the S88 strain.

[00101] FIG. 55 depicts expression constructs used in the production of the S89 strain.

[00102] FIGS. 56A, 56B, and 56C depict expression constructs used in the production of the S90 strain.

[00103] FIGS. 57A, 57B, and 57C depict expression constructs used in the production of the S91 strain.

[00104] FIG. 58 depicts expression constructs used in the production of the S94 strain.

[00105] FIG. 59 depicts expression constructs used in the production of the S95 strain.

[00106] FIG. 60 depicts expression constructs used in the production of the S97 strain.

[00107] FIG. 61 depicts expression constructs used in the production of the S104 strain.

[00108] FIG. 62 depicts expression constructs used in the production of the S108 strain.

[00109] FIG. 63 depicts expression constructs used in the production of the S112 strain.

[00110] FIG. 64 depicts expression constructs used in the production of the S114 strain.

[00111] FIG. 65 depicts expression constructs used in the production of the S115 strain.

[00112] FIG. 66 depicts expression constructs used in the production of the S116 strain.

[00113] FIG. 67 depicts expression constructs used in the production of the S118 strain.

[00114] FIG. 68 depicts expression constructs used in the production of the S123 strain.

[00115] FIG. 69 depicts expression constructs used in the production of the S147 strain.

[00116] FIG. 70 depicts expression constructs used in the production of the S164 strain.

[00117] FIG. 71 depicts expression constructs used in the production of the S165 strain.

[00118] FIG. 72 depicts expression constructs used in the production of the S166 strain.

[00119] FIG. 73 depicts expression constructs used in the production of the S167 strain.

[00120] FIG. 74 depicts expression constructs used in the production of the S168 strain.

[00121] FIG. 75 depicts expression constructs used in the production of the S169 strain.

[00122] FIG. 76 depicts expression constructs used in the production of the S170 strain.

[00123] FIG. 77 depicts the MS/MS spectrum of the CBGA peak produced from a CsPT4 polypeptide expressing strain (S29).

[00124] FIG. 78 depicts the MS/MS spectrum of an authentic CBGA standard.

[00125] FIG. 79 depicts CBGA produced by a CsGOT polypeptide at 1.06 min (top), CBGA produced by a CsPT4 polypeptide at 1.06 min (middle), and authentic CBGA standard at 1.06 min (bottom).

[00126] FIG. 80 depicts CBGA produced by a CsGOT polypeptide at 1.06 min (scale x 10^2 units).

[00127] FIG. 81 depicts CBGA produced by a CsPT4 polypeptide at 1.06 min (scale x 10^4 units)

[00128] FIG. 82 depicts an authentic CBGA standard at 1.06 min (scale x 10^4 units).

[00129] FIG. 83 depicts CBDA produced by S34 at 1.02 min (top) and an authentic CBDA standard at 1.02 min (bottom).

[00130] FIG. 84 depicts THCA produced from strain D123 at 1.29 min (top) and an authentic THCA standard at 1.29 min (bottom).

[00131] FIG. 85 depicts expression constructs used in the production of the S34 strain.

[00132] FIG. 86 depicts expression constructs used in the production of the S29 strain. The expression constructs depicted in FIG. 86 are also used in the production of following strains: S31, S34, S35, S37, S38, S39, S41, S42, S43, S44, S45, S46, S47, S49, S50, S51, S78, S80, S81, S82, S83, S84, S85, S86, S87, S88, S89, S90, S91, S94, S95, S97, and S123.

DETAILED DESCRIPTION

[00133] The present disclosure provides methods, polypeptides, nucleic acids encoding said polypeptides, and genetically modified host cells for producing cannabinoids, cannabinoid precursors, cannabinoid derivatives (e.g., non-naturally occurring cannabinoids), or cannabinoid precursor derivatives (e.g., non-naturally occurring cannabinoid precursors).

[00134] Geranyl pyrophosphate:olivetolic acid geranyltransferase (GOT, Enzyme Commission Number 2.5.1.102) polypeptides play an important role in the biosynthesis of cannabinoids, but reconstituting their activity in a genetically modified host cell has proven challenging, hampering progress in the production of cannabinoids or cannabinoid derivatives. Herein, novel genes encoding polypeptides of the disclosure that catalyze production of cannabigerolic acid (CBGA) from GPP and olivetolic acid have been identified, isolated, and characterized. Surprisingly, these polypeptides of the present disclosure can catalyze production of CBGA from GPP and olivetolic acid in an amount at least ten times higher than previously discovered *Cannabis* polypeptides that catalyze production of CBGA from GPP and olivetolic acid (see, for example, U.S. Patent Application Pub. No. US20120144523 and the GOT polypeptide, CsPT1, disclosed therein; SEQ ID NO:82 herein). The new polypeptides of the present disclosure that catalyze production of CBGA from GPP and olivetolic acid are GOT polypeptides (e.g., the CsPT4 polypeptide) and can generate cannabinoids and cannabinoid derivatives *in vivo* (e.g., within a genetically modified host cell) and *in vitro* (e.g., cell-free). These new GOT polypeptides, as well as nucleic acids encoding said GOT polypeptides, are useful in the methods and genetically modified host cells of the disclosure for producing cannabinoids or cannabinoid derivatives.

[00135] The methods of the disclosure may include using microorganisms genetically engineered (e.g., genetically modified host cells) to produce naturally-occurring and non-naturally occurring cannabinoids or cannabinoid precursors. Naturally-occurring cannabinoids and cannabinoid precursors and non-naturally occurring cannabinoids and cannabinoid precursors (e.g., cannabinoid derivatives and cannabinoid precursor derivatives) are challenging to synthesize using chemical synthesis due to their complex structures. The methods of the disclosure enable the construction of metabolic pathways inside living cells to produce bespoke cannabinoids, cannabinoid precursors, cannabinoid derivatives, or cannabinoid precursor derivatives from simple precursors such as sugars and carboxylic acids. One or more heterologous nucleic acids disclosed herein encoding one or more polypeptides disclosed herein can be introduced into host microorganisms allowing for the stepwise conversion of inexpensive feedstocks, e.g., sugar, into final products: cannabinoids, cannabinoid precursors, cannabinoid derivatives, or cannabinoid precursor derivatives. These products can be specified by the choice and construction of expression constructs or vectors comprising one or more heterologous nucleic acids disclosed herein, allowing for the efficient bioproduction of chosen cannabinoid precursors; cannabinoids, such as THC or

CBD and less common cannabinoid species found at low levels in *Cannabis*; or cannabinoid derivatives or cannabinoid precursor derivatives. Bioproduction also enables synthesis of cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives with defined stereochemistries, which is challenging to do using chemical synthesis.

[00136] The nucleic acids disclosed herein may include those encoding a polypeptide having at least one activity of a polypeptide present in the cannabinoid biosynthetic pathway, such as a GOT polypeptide (e.g., a CsPT4 polypeptide), responsible for the biosynthesis of the cannabinoid CBGA; a tetraketide synthase (TKS) polypeptide; an olivetolic acid cyclase (OAC) polypeptide; and a CBDA or THCA synthase polypeptide (see FIGS. 1 and 11). Nucleic acids disclosed herein may also include those encoding a polypeptide having at least one activity of a polypeptide involved in the synthesis of cannabinoid precursors. These polypeptides include, but are not limited to, polypeptides having at least one activity of a polypeptide present in the mevalonate pathway; polypeptides that generate acyl-CoA compounds or acyl-CoA compound derivatives (e.g., an acyl-activating enzyme polypeptide, a fatty acyl-CoA synthetase polypeptide, or a fatty acyl-CoA ligase polypeptide); polypeptides that generate GPP; polypeptides that generate malonyl-CoA; polypeptides that condense two molecules of acetyl-CoA to generate acetoacetyl-CoA, or pyruvate dehydrogenase complex polypeptides (see FIGS. 1 and 11).

[00137] The disclosure also provides for generation of cannabinoid precursor derivatives or cannabinoid derivatives, as well as cannabinoids or precursors thereof, with polypeptides that generate acyl-CoA compounds or acyl-CoA compound derivatives. In certain such embodiments, genetically modified host cells disclosed herein are modified with one or more heterologous nucleic acids encoding a polypeptide that generates acyl-CoA compounds or acyl-CoA compound derivatives. These polypeptides may permit production of hexanoyl-CoA, acyl-CoA compounds, derivatives of hexanoyl-CoA, or derivatives of acyl-CoA compounds. In some embodiments, hexanoic acid or carboxylic acids other than hexanoic acid are fed to genetically modified host cells expressing a polypeptide that generates acyl-CoA compounds or acyl-CoA compound derivatives (e.g., are present in the culture medium in which the cells are grown) to generate hexanoyl-CoA, acyl-CoA compounds, derivatives of hexanoyl-CoA, or derivatives of acyl-CoA compounds. These compounds are then converted to cannabinoid derivatives or cannabinoid precursor derivatives, as well as cannabinoids or precursors thereof, via one or more polypeptides

having at least one activity of a polypeptide present in the cannabinoid biosynthetic pathway or involved in the synthesis of cannabinoid precursors (see FIGS. 1 and 11).

[00138] Surprisingly, it was found that polypeptides that generate acyl-CoA compounds or acyl-CoA compound derivatives, as well as many polypeptides having at least one activity of a polypeptide present in the cannabinoid biosynthetic pathway, such as TKS polypeptides, OAC polypeptides, GOT polypeptides (e.g., a CsPT4 polypeptide), and CBDA or THCA synthase polypeptides, have broad substrate specificity. This broad substrate specificity permits generation of not only cannabinoids and cannabinoid precursors, but also cannabinoid derivatives and cannabinoid precursor derivatives that are not naturally occurring, both within a genetically modified host cell or in a cell-free reaction mixture comprising one or more of the polypeptides disclosed herein. Because of this broad substrate specificity, hexanoyl-CoA, acyl-CoA compounds, derivatives of hexanoyl-CoA, or derivatives of acyl-CoA compounds produced in genetically modified host cells by polypeptides that generate acyl-CoA compounds or acyl-CoA compound derivatives can be utilized by TKS and OAC polypeptides to make olivetolic acid or derivatives thereof. The olivetolic acid or derivatives thereof can then be utilized by a GOT polypeptide to afford cannabinoids or cannabinoid derivatives. Alternatively, olivetolic acid or derivatives thereof can be fed to genetically modified host cells comprising a GOT polypeptide to afford cannabinoids or cannabinoid derivatives. These cannabinoids or cannabinoid derivatives can then be converted to THCA or CDBA, or derivatives thereof, via a CBDA or THCA synthase polypeptide.

[00139] Besides allowing for the production of desired cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives, the present disclosure provides a more reliable and economical process than agriculture-based production. Microbial fermentations can be completed in days versus the months necessary for an agricultural crop, are not affected by climate variation or soil contamination (e.g., by heavy metals), and can produce pure products at high titer.

[00140] The present disclosure also provides a platform for the economical production of cannabinoid precursors, or derivatives thereof, and high-value cannabinoids including THC and CBD, as well as derivatives thereof. It also provides for the production of different cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives for which no viable method of production exists.

[00141] Additionally, the disclosure provides methods, genetically modified host cells, polypeptides, and nucleic acids encoding said polypeptides to produce cannabinoids,

cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives *in vivo* or *in vitro* from simple precursors. Nucleic acids disclosed herein encoding one or more polypeptides disclosed herein can be introduced into microorganisms (e.g., genetically modified host cells), resulting in expression or overexpression of the one or more polypeptides, which can then be utilized *in vitro* or *in vivo* for the production of cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives. In some embodiments, the *in vitro* methods are cell-free.

[00142] To produce cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives, and create biosynthetic pathways within genetically modified host cells, the genetically modified host cells may express or overexpress combinations of the heterologous nucleic acids disclosed herein encoding polypeptides disclosed herein.

Cannabinoid Biosynthesis

[00143] Nucleic acids encoding polypeptides having at least one activity of a polypeptide present in the cannabinoid biosynthesis pathway can be useful in the methods and genetically modified host cells disclosed herein for the synthesis of cannabinoids, cannabinoid precursors, cannabinoid derivatives, or cannabinoid precursor derivatives.

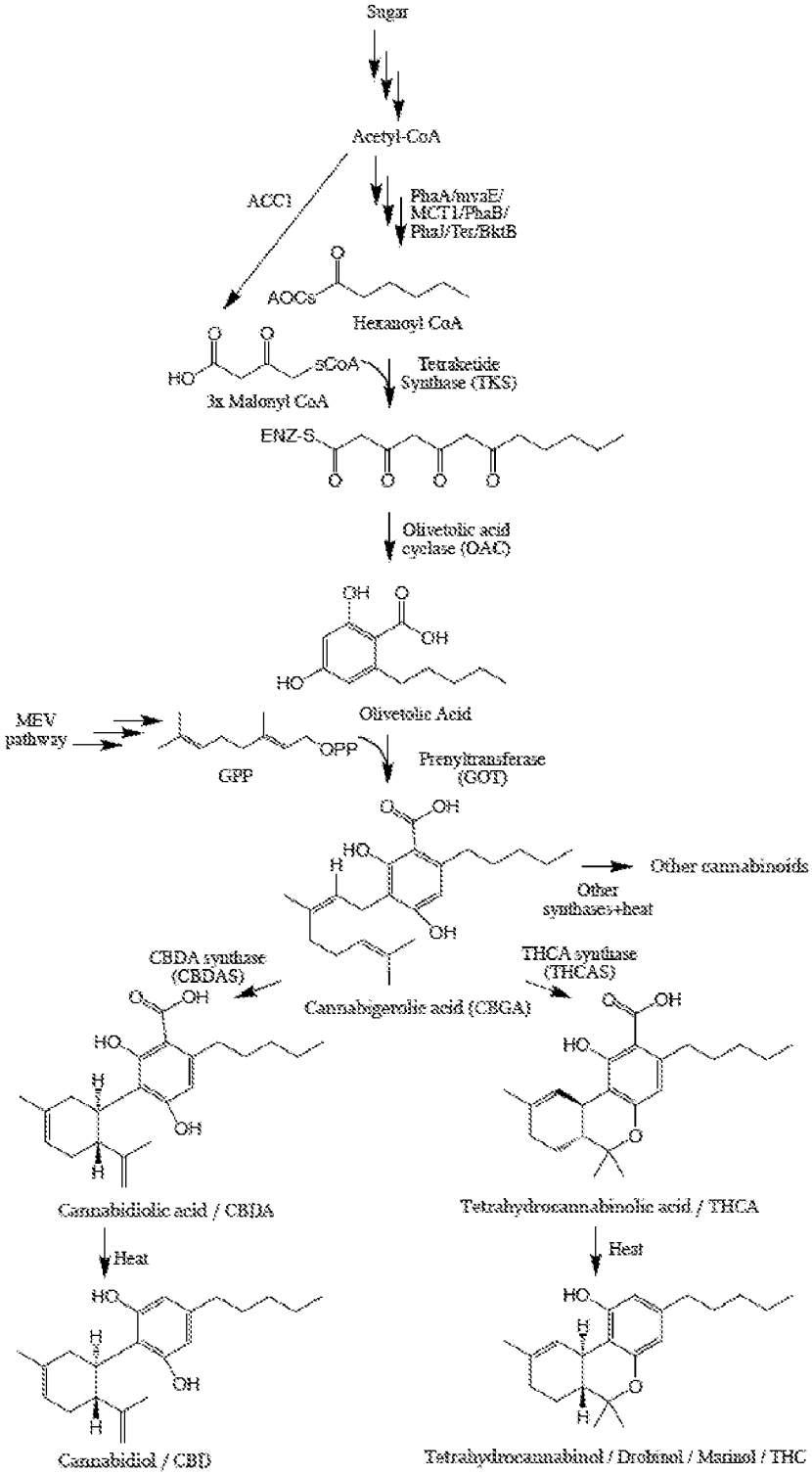
[00144] In *Cannabis*, cannabinoids are produced from the common metabolite precursors geranylpyrophosphate (GPP) and hexanoyl-CoA by the action of three polypeptides so far only identified in *Cannabis*. Hexanoyl-CoA and malonyl-CoA are combined to afford a 12-carbon tetraketide intermediate by a TKS polypeptide. This tetraketide intermediate is then cyclized by an OAC polypeptide to produce olivetolic acid. Olivetolic acid is then prenylated with the common isoprenoid precursor GPP by a GOT polypeptide (e.g., a CsPT4 polypeptide) to produce CBGA, the cannabinoid also known as the "mother cannabinoid." Different synthase polypeptides then convert CBGA into other cannabinoids, e.g., a THCA synthase polypeptide produces THCA, a CBDA synthase polypeptide produces CBDA, etc. In the presence of heat or light, the acidic cannabinoids can undergo decarboxylation, e.g., THCA producing THC or CBDA producing CBD.

[00145] GPP and hexanoyl-CoA can be generated through several pathways (see FIGS. 1 and 11). One or more nucleic acids encoding one or more polypeptides having at least one activity of a polypeptide present in these pathways can be useful in the methods and genetically modified host cells for the synthesis of cannabinoids, cannabinoid precursors, cannabinoid derivatives, or cannabinoid precursor derivatives.

[00146] Polypeptides that generate GPP or are part of a biosynthetic pathway that generates GPP may be one or more polypeptides having at least one activity of a polypeptide present in the mevalonate (MEV) pathway. The term "mevalonate pathway" or "MEV pathway," as used herein, may refer to the biosynthetic pathway that converts acetyl-CoA to isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). The mevalonate pathway comprises polypeptides that catalyze the following steps: (a) condensing two molecules of acetyl-CoA to generate acetoacetyl-CoA (e.g., by action of an acetoacetyl-CoA thiolase polypeptide); (b) condensing acetoacetyl-CoA with acetyl-CoA to form hydroxymethylglutaryl-CoA (HMG-CoA) (e.g., by action of a HMG-CoA synthase (HMGS) polypeptide); (c) converting HMG-CoA to mevalonate (e.g., by action of a HMG-CoA reductase (HMGR) polypeptide); (d) phosphorylating mevalonate to mevalonate 5-phosphate (e.g., by action of a mevalonate kinase (MK) polypeptide); (e) converting mevalonate 5-phosphate to mevalonate 5-pyrophosphate (e.g., by action of a phosphomevalonate kinase (PMK) polypeptide); (f) converting mevalonate 5-pyrophosphate to isopentenyl pyrophosphate (e.g., by action of a mevalonate pyrophosphate decarboxylase (MVD) polypeptide); and (g) converting isopentenyl pyrophosphate (IPP) to dimethylallyl pyrophosphate (DMAPP) (e.g., by action of an isopentenyl pyrophosphate isomerase (IDI) polypeptide) (FIGS. 1 and 11). A geranyl diphosphate synthase (GPPS) polypeptide then acts on IPP and/or DMAPP to generate GPP. Additionally, polypeptides that generate GPP or are part of a biosynthetic pathway that generates GPP may be one or more polypeptides having at least one activity of a polypeptide present in the deoxyxylulose-5-phosphate (DXP) pathway, instead of those of the MEV pathway (FIG. 1).

[00147] Polypeptides that generate hexanoyl-CoA may include polypeptides that generate acyl-CoA compounds or acyl-CoA compound derivatives (e.g., a hexanoyl-CoA synthase (HCS) polypeptide, an acyl-activating enzyme polypeptide, a fatty acyl-CoA synthetase polypeptide, or a fatty acyl-CoA ligase polypeptide). Hexanoyl-CoA may also be generated through pathways comprising one or more polypeptides that generate malonyl-CoA, such as an acetyl-CoA carboxylase (ACC) polypeptide. Additionally, hexanoyl-CoA may be generated with one or more polypeptides that are part of a biosynthetic pathway that produces hexanoyl-CoA, including, but not limited to: a malonyl CoA-acyl carrier protein transacylase (MCT1) polypeptide, a PaaH1 polypeptide, a Crt polypeptide, a Ter polypeptide, and a BktB polypeptide; a MCT1 polypeptide, a PhaB polypeptide, a PhaJ polypeptide, a Ter polypeptide, and a BktB polypeptide; a short chain fatty acyl-CoA thioesterase (SCFA-TE) polypeptide; or a fatty acid synthase (FAS) polypeptide (see FIGS.

1 and 11). Hexanoyl CoA derivatives, acyl-CoA compounds, or acyl-CoA compound derivatives may also be formed via such pathways and polypeptides.



Biosynthetic pathways to cannabinoids

[00148] GPP and hexanoyl-CoA may also be generated through pathways comprising polypeptides that condense two molecules of acetyl-CoA to generate acetoacetyl-CoA and pyruvate dehydrogenase complex polypeptides that generate acetyl-CoA from pyruvate (see FIGS. 1 and 11). Hexanoyl CoA derivatives, acyl-CoA compounds, or acyl-CoA compound derivatives may also be formed via such pathways.

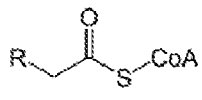
General Information

[00149] The practice of the present disclosure will employ, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, biochemistry, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature: "*Molecular Cloning: A Laboratory Manual*," second edition (Sambrook et al., 1989); "*Oligonucleotide Synthesis*" (M. J. Gait, ed., 1984); "*Animal Cell Culture*" (R. I. Freshney, ed., 1987); "*Methods in Enzymology*" (Academic Press, Inc.); "*Current Protocols in Molecular Biology*" (F. M. Ausubel et al., eds., 1987, and periodic updates); "*PCR: The Polymerase Chain Reaction*," (Mullis et al., eds., 1994). Singleton et al., *Dictionary of Microbiology and Molecular Biology* 2nd ed., J. Wiley & Sons (New York, N.Y. 1994), and March, *Advanced Organic Chemistry Reactions, Mechanisms and Structure* 4th ed., John Wiley & Sons (New York, N.Y. 1992), provide one skilled in the art with a general guide to many of the terms used in the present application.

[00150] "Cannabinoid" or "cannabinoid compound" as used herein may refer to a member of a class of unique meroterpenoids found until now only in *Cannabis sativa*. Cannabinoids may include, but are not limited to, cannabichromene (CBC) type (e.g. cannabichromenic acid), cannabigerol (CBG) type (e.g. cannabigerolic acid), cannabidiol (CBD) type (e.g. cannabidiolic acid), Δ^9-trans-tetrahydrocannabinol (Δ^9 -THC) type (e.g. Δ^9-tetrahydrocannabinolic acid), Δ^8-trans-tetrahydrocannabinol (Δ^8 -THC) type, cannabicyclol (CBL) type, cannabielsoin (CBE) type, cannabinol (CBN) type, cannabinodiol (CBND) type, cannabitriol (CBT) type, cannabigerolic acid (CBGA), cannabigerolic acid monomethylether (CBGAM), cannabigerol (CBG), cannabigerol monomethylether (CBGM), cannabigerovarinic acid (CBGVA), cannabigerovarin (CBGV), cannabichromenic acid (CBCA), cannabichromene (CBC), cannabichromevarinic acid (CBCVA), cannabichromevarin (CBCV), cannabidiolic acid (CBDA), cannabidiol (CBD), cannabidiol monomethylether (CBDM), cannabidiol-C_4 (CBD-C_4), cannabidivarinic acid (CBDVA), cannabidivarin (CBDV), cannabidiorcol (CBD-C_1), Δ^9 –tetrahydrocannabinolic acid A (THCA-A), Δ^9 –tetrahydrocannabinolic acid B (THCA-B), Δ^9 –tetrahydrocannabinol (THC),

Δ^9 –tetrahydrocannabinolic acid-C_4 (THCA-C_4), Δ^9 –tetrahydrocannabinol-C_4 (THC-C_4), Δ^9 –tetrahydrocannabivarinic acid (THCVA), Δ^9 –tetrahydrocannabivarin (THCV), Δ^9 –tetrahydrocannabiorcolic acid (THCA-C_1), Δ^9 –tetrahydrocannabiorcol (THC-C_1), Δ^7 –cis-iso-tetrahydrocannabivarin, Δ^8 –tetrahydrocannabinolic acid (Δ^8 –THCA), Δ^8 –tetrahydrocannabinol (Δ^8 –THC), cannabicyclolic acid (CBLA), cannabicyclol (CBL), cannabicyclovarin (CBLV), cannabielsoic acid A (CBEA-A), cannabielsoic acid B (CBEA-B), cannabielsoin (CBE), cannabielsoinic acid, cannabicitranic acid, cannabinolic acid (CBNA), cannabinol (CBN), cannabinol methylether (CBNM), cannabinol-C_4, (CBN-C_4), cannabivarin (CBV), cannabinol-C_2 (CNB-C_2), cannabiorcol (CBN-C_1), cannabinodiol (CBND), cannabinodivarin (CBVD), cannabitriol (CBT), 10-ethyoxy-9-hydroxy-delta-6a-tetrahydrocannabinol, 8,9-dihydroxyl-delta-6a-tetrahydrocannabinol, cannabitriolvarin (CBTVE), dehydrocannabifuran (DCBF), cannabifuran (CBF), cannabichromanon (CBCN), cannabicitran (CBT), 10-oxo-delta-6a-tetrahydrocannabinol (OTHC), delta-9-cis-tetrahydrocannabinol (cis-THC), 3,4,5,6-tetrahydro-7-hydroxy-alpha-alpha-2-trimethyl-9-n-propyl-2,6-methano-2H-1-benzoxocin-5-methanol (OH-iso-HHCV), cannabiripsol (CBR), and trihydroxy-delta-9-tetrahydrocannabinol (triOH-THC).

[00151] "Cannabinoid precursor" as used herein may refer to any intermediate present in the cannabinoid biosynthetic pathway before the production of the "mother cannabinoid," cannabigerolic acid (CBGA). Cannabinoid precursors may include, but are not limited to, GPP, olivetolic acid, hexanoyl-CoA, pyruvate, acetoacetyl-CoA, butyryl-CoA, acetyl-CoA, HMG-CoA, mevalonate, mevalonate-5-phosphate, mevalonate diphosphate, and malonyl-CoA.

[00152] An acyl-CoA compound as detailed herein may include compounds with the following structure: 

, wherein R is a fatty acid side chain optionally comprising one or more functional and/or reactive groups as disclosed herein (i.e., an acyl-CoA compound derivative).

[00153] As used herein, a hexanoyl CoA derivative, an acyl-CoA compound derivative, a cannabinoid derivative, or a cannabinoid precursor derivative (e.g., an olivetolic acid derivative) is produced by a genetically modified host cell disclosed herein or in a cell-free reaction mixture comprising one or more of the polypeptides disclosed herein and may refer to hexanoyl CoA, an acyl-CoA compound, a cannabinoid, or a cannabinoid precursor (e.g., olivetolic acid) comprising one or more functional and/or reactive groups. Functional groups may include, but are not limited to, azido, halo (e.g., chloride, bromide, iodide,

fluorine), methyl, alkyl (including branched and linear alkyl groups), alkynyl, alkenyl, methoxy, alkoxy, acetyl, amino, carboxyl, carbonyl, oxo, ester, hydroxyl, thio, cyano, aryl, heteroaryl, cycloalkyl, cycloalkenyl, cycloalkylalkenyl, cycloalkylalkynyl, cycloalkenylalkyl, cycloalkenylalkenyl, cycloalkenylalkynyl, heterocyclylalkenyl, heterocyclylalkynyl, heteroarylalkenyl, heteroarylalkynyl, arylalkenyl, arylalkynyl, heterocyclyl, spirocyclyl, heterospirocyclyl, thioalkyl, sulfone, sulfonyl, sulfoxide, amido, alkylamino, dialkylamino, arylamino, alkylarylamino, diarylamino, N-oxide, imide, enamine, imine, oxime, hydrazone, nitrile, aralkyl, cycloalkylalkyl, haloalkyl, heterocyclylalkyl, heteroarylalkyl, nitro, thioxo, and the like. See, e.g., FIGS. 12 and 13. Suitable reactive groups may include, but are not necessarily limited to, azide, carboxyl, carbonyl, amine, (e.g., alkyl amine (e.g., lower alkyl amine), aryl amine), halide, ester (e.g., alkyl ester (e.g., lower alkyl ester, benzyl ester), aryl ester, substituted aryl ester), cyano, thioester, thioether, sulfonyl halide, alcohol, thiol, succinimidyl ester, isothiocyanate, iodoacetamide, maleimide, hydrazine, alkynyl, alkenyl, and the like. A reactive group may facilitate covalent attachment of a molecule of interest. Suitable molecules of interest may include, but are not limited to, a detectable label; imaging agents; a toxin (including cytotoxins); a linker; a peptide; a drug (e.g., small molecule drugs); a member of a specific binding pair; an epitope tag; ligands for binding by a target receptor; tags to aid in purification; molecules that increase solubility; molecules that enhance bioavailability; molecules that increase *in vivo* half-life; molecules that target to a particular cell type; molecules that target to a particular tissue; molecules that provide for crossing the blood-brain barrier; molecules to facilitate selective attachment to a surface; and the like. Functional and reactive groups may be optionally substituted with one or more additional functional or reactive groups.

[00154] A cannabinoid derivative or cannabinoid precursor derivative produced by a genetically modified host cell disclosed herein or in a cell-free reaction mixture comprising one or more of the polypeptides disclosed herein may also refer a naturally-occurring cannabinoid or naturally-occurring cannabinoid precursor lacking one or more chemical moieties. Such chemical moieties may include, but are not limited to, methyl, alkyl, alkenyl, methoxy, alkoxy, acetyl, carboxyl, carbonyl, oxo, ester, hydroxyl, aryl, heteroaryl, cycloalkyl, cycloalkenyl, cycloalkylalkenyl, cycloalkenylalkyl, cycloalkenylalkenyl, heterocyclylalkenyl, heteroarylalkenyl, arylalkenyl, heterocyclyl, aralkyl, cycloalkylalkyl, heterocyclylalkyl, heteroarylalkyl, and the like. In some embodiments, a cannabinoid derivative or cannabinoid precursor derivative lacking one or more chemical moieties found

in a naturally-occurring cannabinoid or naturally-occurring cannabinoid precursor, and produced by a genetically modified host cell disclosed herein or in a cell-free reaction mixture comprising one or more of the polypeptides disclosed herein, may also comprise one or more of any of the functional and/or reactive groups described herein. Functional and reactive groups may be optionally substituted with one or more additional functional or reactive groups.

[00155] The term "nucleic acid" used herein, may refer to a polymeric form of nucleotides of any length, either ribonucleotides or deoxynucleotides. Thus, this term may include, but is not limited to, single-, double-, or multi-stranded DNA or RNA, genomic DNA, cDNA, genes, synthetic DNA or RNA, DNA-RNA hybrids, or a polymer comprising purine and pyrimidine bases or other naturally-occurring, chemically or biochemically modified, non- naturally-occurring, or derivatized nucleotide bases.

[00156] The terms "peptide," "polypeptide," and "protein" may be used interchangeably herein, and may refer to a polymeric form of amino acids of any length, which can include coded and non-coded amino acids, chemically or biochemically modified or derivatized amino acids, full-length polypeptides, fragments of polypeptides, or polypeptides having modified peptide backbones. The polypeptides disclosed herein may be presented as modified or engineered forms, including truncated or fusion forms, retaining the recited activities. The polypeptides disclosed herein may also be variants differing from a specifically recited "reference" polypeptide (e.g., a wild-type polypeptide) by amino acid insertions, deletions, mutations, and/or substitutions, but retains an activity that is substantially similar to the reference polypeptide.

[00157] As used herein, the term "heterologous" may refer to what is not normally found in nature. The term "heterologous nucleotide sequence" may refer to a nucleotide sequence not normally found in a given cell in nature. As such, a heterologous nucleotide sequence may be: (a) foreign to its host cell (i.e., is "exogenous" to the cell); (b) naturally found in the host cell (i.e., "endogenous") but present at an unnatural quantity in the cell (i.e., greater or lesser quantity than naturally found in the host cell); or (c) be naturally found in the host cell but positioned outside of its natural locus. The term "heterologous enzyme" or "heterologous polypeptide" may refer to an enzyme or polypeptide that is not normally found in a given cell in nature. The term encompasses an enzyme or polypeptide that is: (a) exogenous to a given cell (i.e., encoded by a nucleic acid that is not naturally present in the host cell or not naturally present in a given context in the host cell); and (b) naturally found in the host cell (e.g., the enzyme or polypeptide is encoded by a nucleic acid that is

endogenous to the cell) but that is produced in an unnatural amount (e.g., greater or lesser than that naturally found) in the host cell. As such, a heterologous nucleic acid may be: (a) foreign to its host cell (i.e., is "exogenous" to the cell); (b) naturally found in the host cell (i.e., "endogenous") but present at an unnatural quantity in the cell (i.e., greater or lesser quantity than naturally found in the host cell); or (c) be naturally found in the host cell but positioned outside of its natural locus.

[00158] "Operably linked" may refer to an arrangement of elements wherein the components so described are configured so as to perform their usual function. Thus, control sequences operably linked to a coding sequence are capable of effecting the expression of the coding sequence. The control sequences need not be contiguous with the coding sequence, so long as they function to direct the expression thereof. Thus, for example, intervening untranslated yet transcribed sequences can be present between a promoter sequence and the coding sequence and the promoter sequence can still be considered "operably linked" to the coding sequence.

[00159] "Isolated" may refer to polypeptides or nucleic acids that are substantially or essentially free from components that normally accompany them in their natural state. An isolated polypeptide or nucleic acid may be other than in the form or setting in which it is found in nature. Isolated polypeptides and nucleic acids therefore may be distinguished from the polypeptides and nucleic acids as they exist in natural cells. An isolated nucleic acid or polypeptide may further be purified from one or more other components in a mixture with the isolated nucleic acid or polypeptide, if such components are present.

[00160] A "genetically modified host cell" (also referred to as a "recombinant host cell") is a host cell into which has been introduced a heterologous nucleic acid, e.g., an expression vector or construct. For example, a prokaryotic host cell is a genetically modified prokaryotic host cell (e.g., a bacterium), by virtue of introduction into a suitable prokaryotic host cell of a heterologous nucleic acid, e.g., an exogenous nucleic acid that is foreign to (not normally found in nature in) the prokaryotic host cell, or a recombinant nucleic acid that is not normally found in the prokaryotic host cell; and a eukaryotic host cell is a genetically modified eukaryotic host cell, by virtue of introduction into a suitable eukaryotic host cell of a heterologous nucleic acid, e.g., an exogenous nucleic acid that is foreign to the eukaryotic host cell, or a recombinant nucleic acid that is not normally found in the eukaryotic host cell.

[00161] As used herein, a "cell-free system" may refer to a cell lysate, cell extract or other preparation in which substantially all of the cells in the preparation have been

disrupted or otherwise processed so that all or selected cellular components, e.g., organelles, proteins, nucleic acids, the cell membrane itself (or fragments or components thereof), or the like, are released from the cell or resuspended into an appropriate medium and/or purified from the cellular milieu. Cell-free systems can include reaction mixtures prepared from purified or isolated polypeptides and suitable reagents and buffers.

[00162] In some embodiments, conservative substitutions may be made in the amino acid sequence of a polypeptide without disrupting the three-dimensional structure or function of the polypeptide. Conservative substitutions may be accomplished by the skilled artisan by substituting amino acids with similar hydrophobicity, polarity, and R-chain length for one another. Additionally, by comparing aligned sequences of homologous proteins from different species, conservative substitutions may be identified by locating amino acid residues that have been mutated between species without altering the basic functions of the encoded proteins. The term "conservative amino acid substitution" may refer to the interchangeability in proteins of amino acid residues having similar side chains. For example, a group of amino acids having aliphatic side chains consists of glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains consists of serine and threonine; a group of amino acids having amide containing side chains consisting of asparagine and glutamine; a group of amino acids having aromatic side chains consists of phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains consists of lysine, arginine, and histidine; a group of amino acids having acidic side chains consists of glutamate and aspartate; and a group of amino acids having sulfur containing side chains consists of cysteine and methionine. Exemplary conservative amino acid substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, and asparagine-glutamine.

[00163] A polynucleotide or polypeptide has a certain percent "sequence identity" to another polynucleotide or polypeptide, meaning that, when aligned, that percentage of bases or amino acids are the same, and in the same relative position, when comparing the two sequences. Sequence identity can be determined in a number of different manners. To determine sequence identity, sequences can be aligned using various methods and computer programs (e.g., BLAST, T-COFFEE, MUSCLE, MAFFT, etc.), available over the world wide web at sites including ncbi.nlm.nili.gov/BLAST,ebi.ac.uk/Tools/msa/tcoffee/ebi.ac.uk/Tools/msa/muscle/mafft.cbrc.jp/alignment/software/. See, e.g., Altschul et al. (1990), J. Mol. Biol. 215:403-10.

[00164] Before the present disclosure is further described, it is to be understood that this disclosure is not limited to particular embodiments described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present disclosure will be limited only by the appended claims.

[00165] Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limit of that range and any other stated or intervening value in that stated range, is encompassed within the disclosure. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the disclosure, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the disclosure.

[00166] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. Although any methods and materials similar or equivalent to those described herein can also be used in the practice or testing of the present disclosure, the preferred methods and materials are now described. All publications mentioned herein are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited.

[00167] It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" may include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a cannabinoid compound" or "cannabinoid" may include a plurality of such compounds and reference to "the genetically modified host cell" may include reference to one or more genetically modified host cells and equivalents thereof known to those skilled in the art, and so forth. It is further noted that the claims may be drafted to exclude any optional element. As such, this statement is intended to serve as antecedent basis for use of such exclusive terminology as "solely," "only" and the like in connection with the recitation of claim elements, or use of a "negative" limitation.

[00168] It is appreciated that certain features of the disclosure, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the disclosure, which are, for brevity, described in the context of a single embodiment, may also be provided separately or in any suitable sub-combination. All combinations of the embodiments pertaining to the disclosure

are specifically embraced by the present disclosure and are disclosed herein just as if each and every combination was individually and explicitly disclosed. In addition, all sub-combinations of the various embodiments and elements thereof are also specifically embraced by the present disclosure and are disclosed herein just as if each and every such sub-combination was individually and explicitly disclosed herein.

Geranyl Pyrophosphate:Olivetolic Acid Geranyltransferase Polypeptides and Nucleic Acids Encoding Said Polypeptides

[00169] As described herein, novel polypeptides for catalyzing production of cannabigerolic acid from GPP and olivetolic acid have been identified and characterized. Surprisingly, these new polypeptides of the present disclosure can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than previously discovered *Cannabis* polypeptides that catalyze production of cannabigerolic acid from GPP and olivetolic acid (see, for example, U.S. Patent Application Pub. No. US20120144523 and the GOT polypeptide, CsPT1, disclosed therein; SEQ ID NO:82 herein). The new polypeptides of the present disclosure that catalyze production of cannabigerolic acid from GPP and olivetolic acid are new geranyl pyrophosphate:olivetolic acid geranyltransferase (GOT) polypeptides, the CsPT4 polypeptide and truncated versions thereof. These new polypeptides of the present disclosure can generate cannabinoids and cannabinoid derivatives *in vivo* (e.g., within a genetically modified host cell) and *in vitro* (e.g., cell-free).

[00170] These new GOT polypeptides, as well as nucleic acids encoding said GOT polypeptides, are useful in the methods and genetically modified host cells of the disclosure for producing cannabinoids or cannabinoid derivatives. In some embodiments, the GOT polypeptide of the disclosure cannot catalyze production of 5-geranyl olivetolic acid.

[00171] The CsPT4 polypeptide is remarkably different in sequence and activity than the previously identified CsPT1 polypeptide, also a GOT polypeptide. The CsPT1 polypeptide has only 57% homology to the CsPT4 polypeptide. Further, unlike the CsPT1 polypeptide, the activity of the CsPT4 polypeptide, or a truncated version thereof, can be readily reconstituted in a genetically modified host cell of the disclosure, permitting the production of cannabinoids or cannabinoid derivatives by the genetically modified host cells. A truncated version of the CsPT4 polypeptide, the CsPT4t polypeptide, lacking N-terminal amino acids 1-76 of the amino acid sequence set forth in SEQ ID NO:110 (the full-length CsPT4 polypeptide amino acid sequence) was found to readily catalyze the production of

cannabigerolic acid from GPP and olivetolic acid, with activity similar to that of the full-length CsPT4 polypeptide. However, other truncated versions of the CsPT4 polypeptide lacking N-terminal amino acids 1-112 (SEQ ID NO:211), 1-131 (SEQ ID NO:213), 1-142 (SEQ ID NO:215), 1-166 (SEQ ID NO:217), or 1-186 (SEQ ID NO:219) were unable to catalyze formation of cannabigerolic acid from GPP and olivetolic acid, suggesting that these truncation polypeptides lacked residues required for catalytic activity.

[00172] Surprisingly, it was found that the CsPT4 polypeptide, or a truncated version thereof, has broad substrate specificity, permitting generation of not only cannabinoids, but also cannabinoid derivatives. Because of this broad specificity, olivetolic acid or derivatives thereof produced in genetically modified host cells disclosed herein by TKS and OAC polypeptides can be utilized by a CsPT4 polypeptide, or a truncated version thereof, to afford cannabinoids and cannabinoid derivatives. Alternatively, olivetolic acid or derivatives thereof can be fed to genetically modified host cells disclosed herein comprising a CsPT4 polypeptide, or a truncated version thereof, to afford cannabinoids and cannabinoid derivatives. The cannabinoids and cannabinoid derivatives can then be converted to other cannabinoids or cannabinoid derivatives via a CBDA or THCA synthase polypeptide.

Isolated or Purified Nucleic Acids Encoding GOT Polypeptides

[00173] Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00174] Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a truncated CsPT4 polypeptide (CsPT4t polypeptide, lacking N-terminal amino acids 1-76 of the amino acid sequence set forth in SEQ ID NO:110), comprising the amino acid sequence set forth in SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 80%, at least 81%,

at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100.

[00175] Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 85% amino acid

sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof.

[00176] Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a full-length GOT polypeptide, a CsPT4 polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110.

[00177] Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a

CsPT4 polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof.

[00178] Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:111. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:111. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:111. Some embodiments of the disclosure relate to an isolated or purified CsPT4

nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:111.

[00179] Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:111. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:111. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof.

[00180] Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:225. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:225. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:225. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:225.

[00181] Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:225. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:225. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof.

[00182] Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:224. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:224, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:224.

[00183] Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ

ID NO:224, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:224, or a codon degenerate nucleotide sequence thereof.

[00184] Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:221. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:221, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:221. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:221. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:221.

[00185] Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:221. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:221. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:221, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:221, or a codon degenerate nucleotide sequence thereof.

[00186] Further included are nucleic acids that hybridize to the nucleic acids disclosed here. Hybridization conditions may be stringent in that hybridization will occur if there is at

least a 90%, 95%, or 97% sequence identity with the nucleotide sequence present in the nucleic acid encoding the polypeptides disclosed herein. The stringent conditions may include those used for known Southern hybridizations such as, for example, incubation overnight at 42 °C in a solution having 50% formamide, 5×SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5×Denhardt's solution, 10% dextran sulfate, and 20 micrograms/milliliter denatured, sheared salmon sperm DNA, following by washing the hybridization support in 0.1×SSC at about 65 °C. Other known hybridization conditions are well known and are described in Sambrook et al., Molecular Cloning: A Laboratory Manual, Third Edition, Cold Spring Harbor, N.Y. (2001).

[00187] The length of the nucleic acids disclosed herein may depend on the intended use. For example, if the intended use is as a primer or probe, for example for PCR amplification or for screening a library, the length of the nucleic acid will be less than the full length sequence, for example, 15-50 nucleotides. In certain such embodiments, the primers or probes may be substantially identical to a highly conserved region of the nucleotide sequence or may be substantially identical to either the 5' or 3' end of the nucleotide sequence. In some cases, these primers or probes may use universal bases in some positions so as to be "substantially identical" but still provide flexibility in sequence recognition. It is of note that suitable primer and probe hybridization conditions are well known in the art. Also included are cDNA molecules of the disclosed nucleic acids.

Isolated or Purified GOT Polypeptides

[00188] Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00189] Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid

sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100.

[00190] Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified GOT

polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof.

[00191] Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110.

[00192] Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an

isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof.

Vectors Comprising Nucleic Acids Encoding GOT Polypeptides

[00193] Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00194] Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:100. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid

sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100.

[00195] Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t

polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof.

[00196] Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:110. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110.

[00197] Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino

acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof.

[00198] Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:111. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:111. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at

least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:111. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:111.

[00199] Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:111. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:111. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof.

[00200] Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:225. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:225. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:225. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least

96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:225.

[00201] Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:225. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:225. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof.

[00202] Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:221. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:221, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:221. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:221. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:221.

[00203] Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:221. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ

ID NO:221. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:221, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:221, or a codon degenerate nucleotide sequence thereof.

[00204] Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:224. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:224, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:224.

[00205] Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:224, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:224, or a codon degenerate nucleotide sequence thereof.

Expression Constructs Comprising Nucleic Acids Encoding GOT Polypeptides

[00206] Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00207] Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:100. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100.

[00208] Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide,

comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof.

[00209] Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:110. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to

SEQ ID NO:110. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110.

[00210] Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising one or more

nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof.

[00211] Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:111. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:111. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:111. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:111.

[00212] Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:111. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:111. Some embodiments of the

disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof.

[00213] Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:225. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:225. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:225. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:225.

[00214] Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:225. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:225. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence

having at least 95% sequence identity to SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof.

[00215] Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:221. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:221, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:221. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:221. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:221.

[00216] Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:221. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:221. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:221, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:221, or a codon degenerate nucleotide sequence thereof.

[00217] Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID

NO: 224. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:224, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:224.

[00218] Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:224, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:224, or a codon degenerate nucleotide sequence thereof.

Polypeptides, Nucleic Acids, and Genetically Modified Host Cells for the Production of Cannabinoids, Cannabinoid Derivatives, Cannabinoid Precursors, or Cannabinoid Precursor Derivatives

[00219] The present disclosure provides genetically modified host cells for producing a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor

derivative. A genetically modified host cell of the present disclosure may be genetically modified with one or more heterologous nucleic acids disclosed herein encoding one or more polypeptides disclosed herein. Culturing of the genetically modified host cell in a suitable medium provides for synthesis of the cannabinoid, the cannabinoid derivative, the cannabinoid precursor, or the cannabinoid precursor derivative in a recoverable amount. In some embodiments, the genetically modified host cell of the disclosure produces a cannabinoid or a cannabinoid derivative.

[00220] The disclosure also provides nucleic acids, which can be introduced into microorganisms (e.g., genetically modified host cells), resulting in expression or overexpression of the one or more polypeptides, which can then be utilized *in vitro* (e.g., cell-free) or *in vivo* for the production of cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives. In certain such embodiments, cannabinoids or cannabinoid derivatives are produced.

[00221] One or more polypeptides which can be utilized for the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative are disclosed herein, and may include, but are not limited to: one or more polypeptides having at least one activity of a polypeptide present in the cannabinoid biosynthetic pathway, such as, a GOT polypeptide, a CBDA or THCA synthase polypeptide, a TKS polypeptide, and an OAC polypeptide; one or more polypeptides having at least one activity of a polypeptide present in the mevalonate (MEV) pathway; a polypeptide that generates acyl-CoA compounds or acyl-CoA compound derivatives (e.g., an acyl-activating enzyme polypeptide, a fatty acyl-CoA synthetase polypeptide, or a fatty acyl-CoA ligase polypeptide); a polypeptide that generates GPP; a polypeptide that generates malonyl-CoA; a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA, and a pyruvate dehydrogenase complex polypeptide. Additionally, polypeptides which can be utilized for the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative may be one or more polypeptides having at least one activity of a polypeptide present in the DXP pathway, instead of those of the MEV pathway.

[00222] Polypeptides which can be utilized for the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative may also include a hexanoyl-CoA synthase (HCS) polypeptide or one or more polypeptides that are part of a biosynthetic pathway that produces hexanoyl-CoA, including, but not limited to: a MCT1 polypeptide, a PaaH1 polypeptide, a Crt polypeptide, a Ter polypeptide, and a BktB

polypeptide; a MCT1 polypeptide, a PhaB polypeptide, a PhaJ polypeptide, a Ter polypeptide, and a BktB polypeptide; a short chain fatty acyl-CoA thioesterase (SCFA-TE) polypeptide; or a fatty acid synthase (FAS) polypeptide. Hexanoyl CoA derivatives, acyl-CoA compounds, or acyl-CoA compound derivatives may also be formed via such pathways and polypeptides.

[00223] Polypeptides which can be utilized for the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative may also include polypeptides that modulate NADH or NADPH redox balance, polypeptides that generate neryl pyrophosphate, and NphB polypeptides.

[00224] The disclosure also provides nucleic acids encoding said polypeptides which can be utilized for the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative. The disclosure also provides genetically modified host cells comprising one or more of said nucleic acids and polypeptides which can be utilized for the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative.

Geranyl Pyrophosphate:Olivetolic Acid Geranyltransferase (GOT) Polypeptides, Nucleic Acids, and Genetically Modified Host Cells Expressing Said Polypeptides

[00225] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase (GOT) polypeptide.

[00226] Exemplary GOT polypeptides disclosed herein may include a full-length GOT polypeptide, a fragment of a GOT polypeptide, a variant of a GOT polypeptide, a truncated GOT polypeptide, or a fusion polypeptide that has at least one activity of a GOT polypeptide. In some embodiments, the GOT polypeptide has aromatic prenyltransferase (PT) activity. In some embodiments, the GOT polypeptide modifies a cannabinoid precursor or a cannabinoid precursor derivative. In certain such embodiments, the GOT polypeptide modifies olivetolic acid or an olivetolic acid derivative. In some embodiments, the GOT polypeptide cannot catalyze the production of 5-geranyl olivetolic acid.

[00227] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, a

genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least 200-500 times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least 10, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 200, at least 300, at least 400, or at least 500 times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least 10-50, at least 50-100, at least 100-200, at least 100-300, at least 100-400, at least 200-400, at least 100-500, at least 200-500, or at least 300-500 times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00228] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:100 or SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:100 or SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100 or SEQ ID NO:110.

[00229] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:12, SEQ ID NO:82, SEQ ID NO:98, SEQ ID NO:99, or SEQ ID NO:223. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:12, SEQ ID NO:82, SEQ ID

NO:98, SEQ ID NO:99, or SEQ ID NO:223, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:12, SEQ ID NO:82, SEQ ID NO:98, SEQ ID NO:99, or SEQ ID NO:223.

[00230] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:13, SEQ ID NO:101, SEQ ID NO:102, or SEQ ID NO:103. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:13, SEQ ID NO:101, SEQ ID NO:102, or SEQ ID NO:103, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:13, SEQ ID NO:101, SEQ ID NO:102, or SEQ ID NO:103.

[00231] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:211, SEQ ID NO:213, SEQ ID NO:215, SEQ ID NO:217, or SEQ ID NO:219. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:211, SEQ ID NO:213, SEQ ID NO:215, SEQ ID NO:217, or SEQ ID NO:219, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least

99.9%, or 100% amino acid sequence identity to SEQ ID NO:211, SEQ ID NO:213, SEQ ID NO:215, SEQ ID NO:217, or SEQ ID NO:219.

[00232] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:12. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:12, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:12. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:12. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:12.

[00233] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:13. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:13, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:13. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:13. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:13.

[00234] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT1 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT1 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:82, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT1 polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT1 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT1 polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:82.

[00235] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a truncated CsPT1 (CsPT1_t75) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:223. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT1_t75 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:223, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT1_t75 polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:223. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT1_t75 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:223. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT1_t75 polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%,

at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:223.

[00236] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsGOTt75 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:98. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsGOTt75 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:98, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsGOTt75 polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:98. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsGOTt75 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:98. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsGOTt75 polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:98.

[00237] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsGOTt33 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:99. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsGOTt33 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:99, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsGOTt33 polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:99. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsGOTt33 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:99. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsGOTt33 polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least

88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:99.

[00238] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:100. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:100. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:100. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100.

[00239] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:100. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:100. In some embodiments, the GOT polypeptide

encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:100. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:100. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof.

[00240] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT7t polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:101. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT7t polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:101, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT7t polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:101. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT7t polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:101. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT7t polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at

least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:101.

[00241] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a H1PT1Lt polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:102. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a H1PT1Lt polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:102, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a H1PT1Lt polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:102. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a H1PT1Lt polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:102. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a H1PT1Lt polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:102.

[00242] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a H1PT2Lt polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:103. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a H1PT2Lt polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:103, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a H1PT2Lt polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:103. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a H1PT2Lt polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:103. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a H1PT2Lt polypeptide and comprises an

amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:103.

[00243] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110.

[00244] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having at least 75% amino

acid sequence identity to SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof.

[00245] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a truncated CsPT4 (CsPT4_t112) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:211. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t112 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:211, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t112 polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:211. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t112 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:211. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t112 polypeptide and comprises an amino acid sequence having at least

85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:211.

[00246] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a truncated CsPT4 (CsPT4_t131) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:213. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t131 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:213, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t131 polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:213. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t131 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:213. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t131 polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:213.

[00247] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a truncated CsPT4 (CsPT4_t142) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:215. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t142 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:215, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t142 polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:215. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t142 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:215. In

some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t142 polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:215.

[00248] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a truncated CsPT4 (CsPT4_t166) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:217. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t166 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:217, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t166 polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:217. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t166 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:217. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t166 polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:217.

[00249] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a truncated CsPT4 (CsPT4_t186) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:219. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t186 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:219, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t186 polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:219. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t186

polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:219. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t186 polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:219.

[00250] Exemplary GOT heterologous nucleic acids disclosed herein may include nucleic acids that encode a GOT polypeptide, such as, a full-length GOT polypeptide, a fragment of a GOT polypeptide, a variant of a GOT polypeptide, a truncated GOT polypeptide, or a fusion polypeptide that has at least one activity of a GOT polypeptide.

[00251] In some embodiments, the GOT polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the GOT polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the GOT polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of a GOT polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of a GOT polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of a GOT polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a GOT polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of a GOT polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a GOT polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has six copies of a GOT polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a GOT polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has seven copies of a GOT polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a GOT polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has eight copies of a GOT polypeptide-encoding heterologous nucleic acid.

[00252] In some embodiments, the one or more heterologous nucleic acids encoding a GOT polypeptide comprise a nucleotide sequence encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00253] In some embodiments, the one or more heterologous nucleic acids encoding a GOT polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:111, SEQ ID NO:221, SEQ ID NO:224, or SEQ ID NO:225. In some embodiments, the one or more heterologous nucleic acids encoding a GOT polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:111, SEQ ID NO:221, SEQ ID NO:224, or SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GOT polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:111, SEQ ID NO:221, SEQ ID NO:224, or SEQ ID NO:225.

[00254] In some embodiments, the one or more heterologous nucleic acids encoding a GOT polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:220 or SEQ ID NO:222. In some embodiments, the one or more heterologous nucleic acids encoding a GOT polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:220 or SEQ ID NO:222, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GOT polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:220 or SEQ ID NO:222.

[00255] In some embodiments, the one or more heterologous nucleic acids encoding a GOT polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:210, SEQ ID NO:212, SEQ ID NO:214, SEQ ID NO:216, or SEQ ID NO:218. In some embodiments, the one or more heterologous nucleic acids encoding a GOT polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:210, SEQ ID NO:212, SEQ ID NO:214, SEQ

ID NO:216, or SEQ ID NO:218, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GOT polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:210, SEQ ID NO:212, SEQ ID NO:214, SEQ ID NO:216, or SEQ ID NO:218.

[00256] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:111. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:111. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:111. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:111.

[00257] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:111. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:111. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof. In

some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof.

[00258] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:225. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:225. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:225. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:225.

[00259] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:225. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:225. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof.

[00260] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:221. In some

embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:221, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:221. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:221. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:221.

[00261] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:221. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:221. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:221, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:221, or a codon degenerate nucleotide sequence thereof.

[00262] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:224. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:224, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:224.

In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:224. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:224.

[00263] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:224. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:224. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:224, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:224, or a codon degenerate nucleotide sequence thereof.

[00264] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t112 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:210. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t112 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:210, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t112 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:210. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t112 polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least

99.9%, or 100% sequence identity to SEQ ID NO:210. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t112 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:210.

[00265] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t131 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:212. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t131 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:212, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t131 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:212. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t131 polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:212. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t131 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:212.

[00266] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t142 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:214. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t142 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:214, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t142 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:214. In some embodiments, the one or more heterologous

nucleic acids encoding a CsPT4_t142 polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:214. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t142 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:214.

[00267] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t166 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:216. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t166 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:216, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t166 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:216. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t166 polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:216. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t166 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:216.

[00268] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t186 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:218. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t186 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:218, or a codon

degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t186 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:218. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t186 polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:218. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t186 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:218.

[00269] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:220. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:220, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT1 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:220. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT1 polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:220. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT1 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:220.

[00270] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT1_t75 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:222. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT1_t75 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:222, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT1_t75 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:222. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT1_t75 polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:222. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT1_t75 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:222.

Cannabinoid Synthase Polypeptides, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00271] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a cannabinoid synthase polypeptide.

[00272] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than one cannabinoid synthase polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than two cannabinoid synthase polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than three cannabinoid synthase polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids

encoding two cannabinoid synthase polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding three cannabinoid synthase polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, or more cannabinoid synthase polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, or 3 cannabinoid synthase polypeptides.

[00273] In some embodiments, a cannabinoid synthase polypeptide is a tetrahydrocannabinolic acid synthase (THCAS) polypeptide. THCAS polypeptides can catalyze the conversion of cannabigerolic acid to THCA. Exemplary THCAS polypeptides disclosed herein may include a fragment of a THCAS polypeptide, a full-length THCAS polypeptide, a variant of a THCAS polypeptide, a truncated THCAS polypeptide, or a fusion polypeptide that has at least one activity of a THCAS polypeptide.

[00274] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a THCAS polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than one THCAS polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than two THCAS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than three THCAS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding two THCAS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding three THCAS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, or more THCAS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, or 3 THCAS polypeptides.

[00275] In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:14,

SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155, or a conservatively substituted amino acid sequence thereof. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155.

[00276] In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:14. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:14, or a conservatively substituted amino acid sequence thereof. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:14. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:14. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:14.

[00277] In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:86. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:86, or a conservatively substituted amino acid sequence thereof. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid

sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:86. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:86. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:86.

[00278] In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:155. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:155, or a conservatively substituted amino acid sequence thereof. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:155. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:155. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:155.

[00279] In some embodiments, the THCAS polypeptide may include a modified THCAS polypeptide with an N-terminal truncation to remove the secretion peptide and localize to cytoplasm. For example, in some embodiments, the THCAS polypeptide lacks N-terminal amino acids 1-28 of the amino acid sequence set forth in SEQ ID NO:14, or a corresponding signal peptide of another THCAS polypeptide.

[00280] In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:15. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous

nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:15, or a conservatively substituted amino acid sequence thereof. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:15. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:15. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:15.

[00281] In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO: 104. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO: 104, or a conservatively substituted amino acid sequence thereof. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:104. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:104. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:104.

[00282] In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO: 153. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO: 153, or a conservatively substituted amino acid sequence thereof. In some embodiments, the THCAS

polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:153. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:153. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:153.

[00283] Exemplary THCAS heterologous nucleic acids disclosed herein may include nucleic acids that encode a THCAS polypeptide, such as, a fragment of a THCAS polypeptide, a variant of a THCAS polypeptide, a full-length THCAS polypeptide, a truncated THCAS polypeptide, or a fusion polypeptide that has at least one activity of a THCAS polypeptide.

[00284] In some embodiments, the THCAS polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the THCAS polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the THCAS polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of a THCAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of a THCAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of a THCAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a THCAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of a THCAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has six copies of a THCAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has seven copies of a THCAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has eight copies of a THCAS polypeptide-encoding heterologous nucleic acid.

[00285] In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:85, SEQ ID NO:154, or SEQ ID NO:156. In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:85, SEQ ID NO:154, or SEQ ID NO:156, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:85, SEQ ID NO:154, or SEQ ID NO:156.

[00286] In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:85. In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:85, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:85. In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:85.

[00287] In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:154. In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:154, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:154. In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise a nucleotide sequence having at least 85%,

at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:154.

[00288] In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:156. In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:156, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:156. In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:156.

[00289] In some embodiments, a cannabinoid synthase polypeptide is cannabidiolic acid synthase (CBDAS) polypeptide. CBDAS polypeptides can catalyze the conversion of cannabigerolic acid to cannabidiolic acid (CBDA). Exemplary CBDAS polypeptides disclosed herein may include a full-length CBDAS polypeptide, a fragment of a CBDAS polypeptide, a variant of a CBDAS polypeptide, a truncated CBDAS polypeptide, or a fusion polypeptide that has at least one activity of a CBDAS polypeptide.

[00290] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a CBDAS polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than one CBDAS polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than two CBDAS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than three CBDAS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding two CBDAS polypeptides.

In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding three CBDAS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, or more CBDAS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, or 3 CDBAS polypeptides.

[00291] In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:88 or SEQ ID NO:151. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:88 or SEQ ID NO:151, or a conservatively substituted amino acid sequence thereof. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[00292] In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:88. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:88, or a conservatively substituted amino acid sequence thereof. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:88. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:88. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at

least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:88.

[00293] In some embodiments, the CBDAS polypeptide may include a modified CBDAS polypeptide with an N-terminal truncation to remove the secretion peptide and localize to cytoplasm. For example, in some embodiments, the CBDAS polypeptide lacks N-terminal amino acids 1-28 of the amino acid sequence set forth in SEQ ID NO:88, or a corresponding signal peptide of another CBDAS polypeptide.

[00294] In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:16. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:16, or a conservatively substituted amino acid sequence thereof. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:16. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:16. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:16.

[00295] In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:105. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:105, or a conservatively substituted amino acid sequence thereof. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:105. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:105. In some embodiments, the CBDAS

polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:105.

[00296] In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:151. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:151, or a conservatively substituted amino acid sequence thereof. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:151. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:151. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:151.

[00297] Exemplary CBDAS heterologous nucleic acids disclosed herein may include nucleic acids that encode a CBDAS polypeptide, such as, a full-length CBDAS polypeptide, a fragment of a CBDAS polypeptide, a variant of a CBDAS polypeptide, a truncated CBDAS polypeptide, or a fusion polypeptide that has at least one activity of a CBDAS polypeptide.

[00298] In some embodiments, the CBDAS polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the CBDAS polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the CBDAS polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of a CBDAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of a CBDAS polypeptide-encoding

heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of a CBDAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a CBDAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of a CBDAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has six copies of a CBDAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has seven copies of a CBDAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has eight copies of a CBDAS polypeptide-encoding heterologous nucleic acid.

[00299] In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:152 or SEQ ID NO:167. In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:152 or SEQ ID NO:167, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:152 or SEQ ID NO:167.

[00300] In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:87. In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:87, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:87. In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:87.

[00301] In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:152. In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:152, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:152. In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:152.

[00302] In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:167. In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:167, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:167. In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:167.

[00303] In some embodiments, at least one of the heterologous nucleic acids encoding a cannabinoid synthase polypeptide is operably linked to an inducible promoter. In some embodiments, at least one of the heterologous nucleic acids encoding a cannabinoid synthase polypeptide is operably linked to a constitutive promoter. In some embodiments, a signal peptide is linked to the N-terminus of a THCAS or CBDAS polypeptide or other cannabinoid synthase polypeptide.

Polypeptides that Generate Acyl-CoA Compounds or Acyl-CoA Compound Derivatives, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00304] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a polypeptide that generates acyl-CoA compounds or acyl-CoA compound derivatives. Such polypeptides may include, but are not limited to, acyl-activating enzyme (AAE) polypeptides, fatty acyl-CoA synthetases (FAA) polypeptides, or fatty acyl-CoA ligase polypeptides.

[00305] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding an AAE, FAA, or fatty acyl-CoA ligase polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than one AAE, FAA, or fatty acyl-CoA ligase polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than two AAE, FAA, or fatty acyl-CoA ligase polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than three AAE, FAA, or fatty acyl-CoA ligase polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding two AAE, FAA, or fatty acyl-CoA ligase polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding three AAE, FAA, or fatty acyl-CoA ligase polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, or more AAE, FAA, or fatty acyl-CoA ligase polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, or 3 AAE, FAA, or fatty acyl-CoA ligase polypeptides.

[00306] AAE polypeptides, FAA polypeptides, and fatty acyl-CoA ligase polypeptides can convert carboxylic acids to their CoA forms and generate acyl-CoA compounds or acyl-CoA compound derivatives. Promiscuous acyl-activating enzyme polypeptides, such as CsAAE1 and CsAAE3, FAA polypeptides, or fatty acyl-CoA ligase

polypeptides, may permit generation of cannabinoid derivatives (e.g., cannabigerolic acid derivatives) or cannabinoid precursor derivatives (e.g., olivetolic acid derivatives), as well as cannabinoids (e.g., cannabigerolic acid) or precursors thereof (e.g., olivetolic acid). In some embodiments, hexanoic acid or carboxylic acids other than hexanoic acid are fed to genetically modified host cells expressing an AAE polypeptide, FAA polypeptide, or fatty acyl-CoA ligase polypeptide (e.g., are present in the culture medium in which the cells are grown) to generate hexanoyl-CoA, acyl-CoA compounds, derivatives of hexanoyl-CoA, or derivatives of acyl-CoA compounds. In certain such embodiments, the cell culture medium comprising the genetically modified host cells comprises hexanoate. In some embodiments, the cell culture medium comprising the genetically modified host cells comprises a carboxylic acid other than hexanoate.

[00307] Exemplary AAE, FAA, or fatty acyl-CoA ligase polypeptides disclosed herein may include a full-length AAE, FAA, or fatty acyl-CoA ligase polypeptide; a fragment of a AAE, FAA, or fatty acyl-CoA ligase polypeptide; a variant of a AAE, FAA, or fatty acyl-CoA ligase polypeptide; a truncated AAE, FAA, or fatty acyl-CoA ligase polypeptide; or a fusion polypeptide that has at least one activity of an AAE, FAA, or fatty acyl-CoA ligase polypeptide.

[00308] In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE1 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:90. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE1 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:90, or a conservatively substituted amino acid sequence thereof. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE1 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:90. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE1 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:90. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE1 polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100%

amino acid sequence identity to SEQ ID NO:90. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE1 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:90.

[00309] In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:92 or SEQ ID NO:149. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:92 or SEQ ID NO:149, or a conservatively substituted amino acid sequence thereof. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:92 or SEQ ID NO:149.

[00310] In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:92. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:92, or a conservatively substituted amino acid sequence thereof. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:92. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:92. In some

embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:92.

[00311] In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:112. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:112, or a conservatively substituted amino acid sequence thereof. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:112. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:112. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:112. In these proceeding embodiments, the CsAAE3 polypeptide lacks the RELIQKVRSNM C-terminal amino acids.

[00312] In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:149. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:149, or a conservatively substituted amino acid sequence thereof. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:149. In some embodiments, the AAE

polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:149. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:149. In these proceeding embodiments, the CsAAE3 polypeptide lacks the RRELIQKVRSNM C-terminal amino acids.

[00313] In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:145 or SEQ ID NO:147. In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:145 or SEQ ID NO:147, or a conservatively substituted amino acid sequence thereof. In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:145 or SEQ ID NO:147.

[00314] In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:145. In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:145, or a conservatively substituted amino acid sequence thereof. In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:145. In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more

heterologous nucleic acids is a FADK polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:145. In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:145.

[00315] In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:147. In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:147, or a conservatively substituted amino acid sequence thereof. In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:147. In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:147. In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:147.

[00316] In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200, or a

conservatively substituted amino acid sequence thereof. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200.

[00317] In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:169. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:169, or a conservatively substituted amino acid sequence thereof. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:169. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:169. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2 polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:169. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:169.

[00318] In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a truncated FAA2 (tFAA2) polypeptide. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a tFAA2 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:194. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a tFAA2 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:194, or a conservatively substituted amino acid sequence thereof. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a tFAA2 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:194. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a tFAA2 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:194. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a tFAA2 polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:194. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a tFAA2 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:194.

[00319] In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a mutated FAA2 (FAA2mut) polypeptide. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2mut polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:196. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2mut polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:196, or a conservatively substituted amino acid sequence

thereof. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2mut polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:196. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2mut polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:196. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2mut polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:196. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2mut polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:196.

[00320] In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA1 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:192. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA1 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:192, or a conservatively substituted amino acid sequence thereof. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA1 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:192. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA1 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:192. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA1 polypeptide and comprises an amino acid sequence having at least 85%, at least

86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:192. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA1 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:192.

[00321] In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA3 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:198. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA3 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:198, or a conservatively substituted amino acid sequence thereof. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA3 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:198. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA3 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:198. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA3 polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:198. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA3 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least

97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:198.

[00322] In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA4 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:200. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA4 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:200, or a conservatively substituted amino acid sequence thereof. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA4 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:200. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA4 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:200. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA4 polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:200. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA4 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:200.

[00323] Exemplary AAE, FAA, or fatty acyl-CoA ligase heterologous nucleic acids disclosed herein may include nucleic acids that encode an AAE, FAA, or fatty acyl-CoA ligase polypeptide, such as, a full-length AAE, FAA, or fatty acyl-CoA ligase polypeptide; a fragment of a AAE, FAA, or fatty acyl-CoA ligase polypeptide; a variant of a AAE, FAA, or fatty acyl-CoA ligase polypeptide; a truncated AAE, FAA, or fatty acyl-CoA ligase polypeptide; or a fusion polypeptide that has at least one activity of an AAE, FAA, or fatty acyl-CoA ligase polypeptide.

[00324] In some embodiments, the AAE, FAA, or fatty acyl-CoA ligase polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the AAE, FAA, or fatty acyl-CoA ligase polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the AAE, FAA, or fatty acyl-CoA ligase polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of an AAE, FAA, or fatty acyl-CoA ligase polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of an AAE, FAA, or fatty acyl-CoA ligase polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of an AAE, FAA, or fatty acyl-CoA ligase polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of an AAE, FAA, or fatty acyl-CoA ligase polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of an AAE, FAA, or fatty acyl-CoA ligase polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has six copies of an AAE, FAA, or fatty acyl-CoA ligase polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has seven copies of an AAE, FAA, or fatty acyl-CoA ligase polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has eight copies of an AAE, FAA, or fatty acyl-CoA ligase polypeptide-encoding heterologous nucleic acid.

[00325] In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:164 or SEQ ID NO:165. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:164 or SEQ ID NO:165, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:164 or SEQ ID NO:165.

[00326] In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:89. In some

embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:89, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:89. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:89.

[00327] In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:164. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:164, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:164. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:164.

[00328] In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:165. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:165, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:165. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least

92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:165.

[00329] In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:150 or SEQ ID NO:166. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:150 or SEQ ID NO:166, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO: 150 or SEQ ID NO:166.

[00330] In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO: 91. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO: 91, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:91. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:91.

[00331] In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:150. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:150, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence

identity to SEQ ID NO:150. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:150.

[00332] In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:166. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:166, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:166. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:166.

[00333] In some embodiments, the one or more heterologous nucleic acids encoding a FADK polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:146 or SEQ ID NO:148. In some embodiments, the one or more heterologous nucleic acids encoding a FADK polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:146 or SEQ ID NO:148, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a FADK polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:146 or SEQ ID NO:148.

[00334] In some embodiments, the one or more heterologous nucleic acids encoding a FADK polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:146. In some embodiments, the one or more heterologous nucleic acids encoding a FADK polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:146, or a codon degenerate

nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a FADK polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:146. In some embodiments, the one or more heterologous nucleic acids encoding a FADK polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:146.

[00335] In some embodiments, the one or more heterologous nucleic acids encoding a FADK polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:148. In some embodiments, the one or more heterologous nucleic acids encoding a FADK polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:148, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a FADK polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:148. In some embodiments, the one or more heterologous nucleic acids encoding a FADK polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:148.

[00336] In some embodiments, the one or more heterologous nucleic acids encoding a FAA polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:168, SEQ ID NO:191, SEQ ID NO:193, SEQ ID NO:195, SEQ ID NO:197, or SEQ ID NO:199. In some embodiments, the one or more heterologous nucleic acids encoding a FAA polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:168, SEQ ID NO:191, SEQ ID NO:193, SEQ ID NO:195, SEQ ID NO:197, or SEQ ID NO:199, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a FAA polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to

SEQ ID NO:168, SEQ ID NO:191, SEQ ID NO:193, SEQ ID NO:195, SEQ ID NO:197, or SEQ ID NO:199.

[00337] In some embodiments, the one or more heterologous nucleic acids encoding a FAA2 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:168. In some embodiments, the one or more heterologous nucleic acids encoding a FAA2 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:168, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a FAA2 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:168. In some embodiments, the one or more heterologous nucleic acids encoding a FAA2 polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:168. In some embodiments, the one or more heterologous nucleic acids encoding a FAA2 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:168.

[00338] In some embodiments, the one or more heterologous nucleic acids encoding a tFAA2 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:193. In some embodiments, the one or more heterologous nucleic acids encoding a tFAA2 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:193, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a tFAA2 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:193. In some embodiments, the one or more heterologous nucleic acids encoding a tFAA2 polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:193. In some embodiments, the one or more heterologous nucleic acids

encoding a tFAA2 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:193.

[00339] In some embodiments, the one or more heterologous nucleic acids encoding a FAA2mut polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:195. In some embodiments, the one or more heterologous nucleic acids encoding a FAA2mut polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:195, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a FAA2mut polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:195. In some embodiments, the one or more heterologous nucleic acids encoding a FAA2mut polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:195. In some embodiments, the one or more heterologous nucleic acids encoding a FAA2mut polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:195.

[00340] In some embodiments, the one or more heterologous nucleic acids encoding a FAA1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:191. In some embodiments, the one or more heterologous nucleic acids encoding a FAA1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:191, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a FAA1 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:191. In some embodiments, the one or more heterologous nucleic acids encoding a FAA1 polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%,

at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:191. In some embodiments, the one or more heterologous nucleic acids encoding a FAA1 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:191.

[00341] In some embodiments, the one or more heterologous nucleic acids encoding a FAA3 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:197. In some embodiments, the one or more heterologous nucleic acids encoding a FAA3 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:197, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a FAA3 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:197. In some embodiments, the one or more heterologous nucleic acids encoding a FAA3 polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:197. In some embodiments, the one or more heterologous nucleic acids encoding a FAA3 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:197.

[00342] In some embodiments, the one or more heterologous nucleic acids encoding a FAA4 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:199. In some embodiments, the one or more heterologous nucleic acids encoding a FAA4 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:199, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic

acids encoding a FAA4 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:199. In some embodiments, the one or more heterologous nucleic acids encoding a FAA4 polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:199. In some embodiments, the one or more heterologous nucleic acids encoding a FAA4 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:199.

Polypeptides that Generate or Are Part of a Pathway that Generates Hexanoyl-CoA, Hexanoyl-CoA Derivatives, Acyl-CoA Compounds, or Acyl-CoA Compound Derivatives, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00343] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding one or more polypeptides that generate or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives.

[00344] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than one polypeptide that generates or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than two polypeptides that generate or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than

three polypeptides that generate or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than four polypeptides that generate or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than five polypeptides that generate or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding two polypeptides that generate or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding three polypeptides that generate or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding four polypeptides that generate or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding five polypeptides that generate or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, 4, 5 or more polypeptides that generate or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, 4, or 5 polypeptides that generate or are part of a biosynthetic pathway that generates hexanoyl-

CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives.

[00345] Exemplary polypeptides disclosed herein that generate or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives may include a full-length polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives; a fragment of a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives; a variant of a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives; a truncated polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives; or a fusion polypeptide that has at least one activity of a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives.

[00346] In some embodiments, the one or more polypeptides that generate hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives may include a hexanoyl-CoA synthase (HCS) polypeptide (e.g., as depicted in Box 1a of FIG. 1). In some embodiments, the one or more polypeptides that generate hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives is an HCS polypeptide and the cell culture medium comprising the genetically modified host cell comprises hexanoate. In some embodiments, the one or more polypeptides that generate hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives is an HCS polypeptide and the cell culture medium comprising the genetically modified host cell comprises a carboxylic acid other than hexanoate. In some embodiments, hexanoic acid or carboxylic acids other than hexanoic acid are fed to a genetically modified host cell expressing the HCS polypeptide (e.g., are present in the culture medium in which the cells are grown) to generate hexanoyl-CoA, acyl-CoA compounds, derivatives of hexanoyl-CoA, or derivatives of acyl-CoA compounds.

[00347] In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:1. In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:1, or a conservatively

substituted amino acid sequence thereof. In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:1.

[00348] In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids is a RevS polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:2. In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids is a RevS polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:2, or a conservatively substituted amino acid sequence thereof. In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids is a RevS polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:2.

[00349] In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids is an AflA polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:3. In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids is an AflA polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:3, or a conservatively substituted amino acid sequence thereof. In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids is an AflA polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:3.

[00350] In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids is an AflB polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:4. In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids is an AflB polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:4, or a conservatively substituted amino acid sequence thereof. In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids is an AflB polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:4.

[00351] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: i) one or more heterologous nucleic acids that encode an AflA polypeptide and ii) one or more heterologous nucleic acids that encode an AflB polypeptide.

[00352] In some embodiments, one or more polypeptides that generate hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives comprise a MCT1 polypeptide, a PaaH1 polypeptide, a Crt polypeptide, a Ter polypeptide, and a BktB polypeptide. See, e.g., Machado et al. (2012) *Metabolic Engineering* 14:504. In some embodiments, the PaaH1 (3-hydroxyacyl-CoA dehydrogenase) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:18 or SEQ ID NO:46. In some embodiments, the PaaH1 polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:18 or SEQ ID NO:46, or a conservatively substituted amino acid sequence thereof. In some embodiments, the PaaH1 polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%,

at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:18 or SEQ ID NO:46. In some embodiments, the Crt (crotonase) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:19 or SEQ ID NO:48. In some embodiments, the Crt polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:19 or SEQ ID NO:48, or a conservatively substituted amino acid sequence thereof. In some embodiments, the Crt polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:19 or SEQ ID NO:48. In some embodiments, the Ter (trans-2-enoyl-CoA reductase) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:20 or SEQ ID NO:50. In some embodiments, the Ter polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:20 or SEQ ID NO:50, or a conservatively substituted amino acid sequence thereof. In some embodiments, the Ter polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:20 or SEQ ID NO:50. In some embodiments, the BktB (β-ketothiolase) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:21 or SEQ ID NO:44. In some embodiments, the BktB polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:21 or SEQ ID NO:44, or a conservatively substituted amino acid sequence thereof. In some embodiments, the BktB polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%,

at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:21 or SEQ ID NO:44.

[00353] In some embodiments, the one or more polypeptides that generate hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives comprise a MCT1 polypeptide, a PhaB polypeptide, a PhaJ polypeptide, a Ter polypeptide, and a BktB polypeptide. In some embodiments, the PhaB (acetoacetyl-CoA reductase) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:94. In some embodiments, the PhaB polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:94, or a conservatively substituted amino acid sequence thereof. In some embodiments, the PhaB polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:94. In some embodiments, the PhaJ ((R)-specific enoyl-CoA hydratase) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:96. In some embodiments, the PhaJ polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:96, or a conservatively substituted amino acid sequence thereof. In some embodiments, the PhaJ polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:96. In some embodiments, the Ter (trans-2-

enoyl-CoA reductase) and the BktB (β-ketothiolase) polypeptides used are selected from the Ter and BktB polypeptides disclosed herein.

[00354] In some embodiments, the one or more polypeptides that generate hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives comprise a polypeptide that condenses an acetyl-CoA and a malonyl-CoA to generate acetoacetyl-CoA. Polypeptides that condense an acetyl-CoA and a malonyl-CoA to generate acetoacetyl-CoA may include a malonyl CoA-acyl carrier protein transacylase (MCT1) polypeptide. In some embodiments, the MCT1 polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:42. In some embodiments, the MCT1 polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:42, or a conservatively substituted amino acid sequence thereof. In some embodiments, the MCT1 polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:42. In some embodiments, the host cell is genetically modified with one or more heterologous nucleic acids encoding a polypeptide that condense an acetyl-CoA and a malonyl-CoA to generate acetoacetyl-CoA. In certain such embodiments, the polypeptide that condenses an acetyl-CoA and a malonyl-CoA to generate acetoacetyl-CoA is an MCT1 polypeptide.

[00355] The one or more polypeptides that generate hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives may also include a short chain fatty acyl-CoA thioesterase (SCFA-TE) polypeptide (e.g., as depicted in Box 1c of FIG. 1). In some embodiments, the SCFA-TE polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, or SEQ ID NO:31. In some embodiments, the SCFA-TE polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, or SEQ ID NO:31, or a

conservatively substituted amino acid sequence thereof. In some embodiments, the SCFA-TE polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, or SEQ ID NO:31.

[00356] In some embodiments, the one or more polypeptides that are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives comprise a fatty acid synthase polypeptide, such as a FAS1 or FAS2 polypeptide. In some embodiments, the FAS1 polypeptide encoded by the one or more heterologous nucleic acids is a FAS1 (I306A, R1834K) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:106. In some embodiments, the FAS1 polypeptide encoded by the one or more heterologous nucleic acids is a FAS1 (I306A, R1834K) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:106, or a conservatively substituted amino acid sequence thereof. In some embodiments, the FAS1 polypeptide encoded by the one or more heterologous nucleic acids is a FAS1 (I306A, R1834K) polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:106. In some embodiments, the FAS2 polypeptide encoded by the one or more heterologous nucleic acids is a FAS2 (G1250S) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:107. In some embodiments, the FAS2 polypeptide encoded by the one or more heterologous nucleic acids is a FAS2 (G1250S) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:107, or a conservatively substituted amino acid sequence thereof. In some embodiments, the FAS2 polypeptide encoded by the one or more heterologous nucleic acids is a FAS2 (G1250S) polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at

least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:107.

[00357] Exemplary heterologous nucleic acids disclosed herein may include nucleic acids that encode a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives, such as, a full-length polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives; a fragment of a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives; a variant of a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives; a truncated polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives; or a fusion polypeptide that has at least one activity of a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives.

[00358] In some embodiments, the polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the heterologous nucleic acid encoding a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of a heterologous nucleic acid encoding a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, the genetically modified host cell has two copies of a heterologous

nucleic acid encoding a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, the genetically modified host cell has three copies of a heterologous nucleic acid encoding a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, the genetically modified host cell has four copies of a heterologous nucleic acid encoding a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, the genetically modified host cell has five copies of a heterologous nucleic acid encoding a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives.

[00359] In some embodiments, the one or more heterologous nucleic acids encoding an MCT1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:41. In some embodiments, the one or more heterologous nucleic acids encoding an MCT1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:41, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an MCT1 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:41. In some embodiments, the one or more heterologous nucleic acids encoding a BktB polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:43. In some embodiments, the one or more heterologous nucleic acids encoding a BktB polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:43, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a BktB polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:43. In some embodiments, the one or more heterologous

nucleic acids encoding a PaaH1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:45. In some embodiments, the one or more heterologous nucleic acids encoding a PaaH1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:45, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a PaaH1 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:45. In some embodiments, the one or more heterologous nucleic acids encoding a Crt polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:47. In some embodiments, the one or more heterologous nucleic acids encoding a Crt polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:47, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a Crt polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:47. In some embodiments, the one or more heterologous nucleic acids encoding a Ter polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:49. In some embodiments, the one or more heterologous nucleic acids encoding a Ter polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:49, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a Ter polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:49. In some embodiments, the one or more heterologous nucleic acids encoding a PhaB polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:93. In some embodiments, the one or more heterologous nucleic acids encoding a PhaB polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:93, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more

heterologous nucleic acids encoding a PhaB polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:93. In some embodiments, the one or more heterologous nucleic acids encoding a PhaJ polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:95. In some embodiments, the one or more heterologous nucleic acids encoding a PhaJ polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:95, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a PhaJ polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:95.

Polypeptides that Generate Malonyl-CoA, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00360] In some embodiments, the host cell is genetically modified with one or more heterologous nucleic acids encoding a polypeptide that generates malonyl-CoA. In some embodiments, the polypeptide that generates malonyl-CoA is an acetyl-CoA carboxylate (ACC) polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding an ACC polypeptide.

[00361] Exemplary ACC polypeptides disclosed herein may include a full-length ACC polypeptide, a fragment of an ACC polypeptide, a variant of an ACC polypeptide, a truncated ACC polypeptide, or a fusion polypeptide that has at least one activity of an ACC polypeptide.

[00362] In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207, or a conservatively substituted amino

acid sequence thereof. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207.

[00363] In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:9. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:9, or a conservatively substituted amino acid sequence thereof. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:9. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:9. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:9.

[00364] In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 (S659A, S1157A) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:97. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 (S659A, S1157A) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:97, or a conservatively substituted amino acid sequence thereof. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 (S659A, S1157A) polypeptide and comprises an amino acid sequence having at least 50%, at least

55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:97. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 (S659A, S1157A) polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:97. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 (S659A, S1157A) polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:97.

[00365] In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 (S659A, S1157A) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:207. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 (S659A, S1157A) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:207, or a conservatively substituted amino acid sequence thereof. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 (S659A, S1157A) polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:207. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 (S659A, S1157A) polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:207. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 (S659A, S1157A) polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:207.

[00366] Exemplary ACC heterologous nucleic acids disclosed herein may include nucleic acids that encode an ACC polypeptide, such as, a full-length ACC polypeptide, a fragment of an ACC polypeptide, a variant of an ACC polypeptide, a truncated ACC polypeptide, or a fusion polypeptide that has at least one activity of an ACC polypeptide.

[00367] In some embodiments, the ACC polypeptide is overexpressed in the genetically modified host cell. See, e.g., Runguphan and Keasling (2014) *Metabolic Engineering* 21:103. Overexpression may be achieved by increasing the copy number of the ACC polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the ACC polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of an ACC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of an ACC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of an ACC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of an ACC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of an ACC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has six copies of an ACC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has seven copies of an ACC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has eight copies of an ACC polypeptide-encoding heterologous nucleic acid.

[00368] In some embodiments, the one or more heterologous nucleic acids encoding an ACC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:201. In some embodiments, the one or more heterologous nucleic acids encoding an ACC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:201, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an ACC polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:201. In some embodiments, the one or more heterologous nucleic acids encoding an ACC polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:201. In some embodiments, the one or more heterologous nucleic acids encoding an ACC polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at

least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:201.

Polypeptides that Condense an Acyl-CoA Compound or an Acyl-CoA Compound Derivative with Malonyl-CoA to Generate Olivetolic Acid or Derivatives of Olivetolic Acid, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00369] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding one or more polypeptides that condense an acyl-CoA compound, such as hexanoyl-CoA, or an acyl-CoA compound derivative, such as a hexanoyl-CoA derivative, with malonyl-CoA to generate olivetolic acid, or a derivative of olivetolic acid. Polypeptides that react an acyl-CoA compound or an acyl-CoA compound derivative with malonyl-CoA to generate olivetolic acid, or a derivative of olivetolic acid, may include TKS and OAC polypeptides. TKS and OAC polypeptides have been found to have broad substrate specificity, enabling production of cannabinoid derivatives or cannabinoid precursor derivatives, in addition to cannabinoids and cannabinoid precursors.

[00370] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a TKS polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than one TKS polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than two TKS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than three TKS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding two TKS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding three TKS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, or more TKS polypeptides. In some

embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, or 3 TKS polypeptides.

[00371] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding an OAC polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than one OAC polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than two OAC polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than three OAC polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding two OAC polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding three OAC polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, or more OAC polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, or 3 OAC polypeptides.

[00372] Exemplary TKS or OAC polypeptides disclosed herein may include a full-length TKS or OAC polypeptide, a fragment of a TKS or OAC polypeptide, a variant of a TKS or OAC polypeptide, a truncated TKS or OAC polypeptide, or a fusion polypeptide that has at least one activity of a TKS or OAC polypeptide.

[00373] In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:11 or SEQ ID NO:76. In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:11 or SEQ ID NO:76, or a conservatively substituted amino acid sequence thereof. In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least

99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:11 or SEQ ID NO:76.

[00374] In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:10 or SEQ ID NO:78. In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:10 or SEQ ID NO:78, or a conservatively substituted amino acid sequence thereof. In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00375] In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:11. In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:11, or a conservatively substituted amino acid sequence thereof. In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:11. In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:11. In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:11.

[00376] In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:76. In some embodiments, the TKS polypeptide encoded by the one or more heterologous

nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:76, or a conservatively substituted amino acid sequence thereof. In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:76. In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:76. In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:76.

[00377] In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:10. In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:10, or a conservatively substituted amino acid sequence thereof. In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:10. In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:10. In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:10.

[00378] In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:78. In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:78, or a conservatively substituted amino acid sequence thereof. In some embodiments, the OAC

polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:78. In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:78. In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:78.

[00379] In some embodiments, the TKS and OAC polypeptides are fused into a single polypeptide chain (a TKS/OAC fusion polypeptide). In some embodiments, the TKS/OAC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:80. In some embodiments, the TKS/OAC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:80, or a conservatively substituted amino acid sequence thereof. In some embodiments, the TKS/OAC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:80. In some embodiments, the TKS/OAC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:80. In some embodiments, the TKS/OAC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:80. In some embodiments, the TKS/OAC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least

99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:80.

[00380] Exemplary TKS or OAC heterologous nucleic acids disclosed herein may include nucleic acids that encode a TKS or OAC polypeptide, such as, a full-length TKS or OAC polypeptide, a fragment of a TKS or OAC polypeptide, a variant of a TKS or OAC polypeptide, a truncated TKS or OAC polypeptide, or a fusion polypeptide that has at least one activity of a TKS or OAC polypeptide.

[00381] In some embodiments, the TKS or OAC polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the TKS and/or OAC polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the TKS and/or OAC polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has six copies of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has seven copies of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has eight copies of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has nine copies of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has 10 copies of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has 11 copies of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has 12 copies of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid.

[00382] In some embodiments, the one or more heterologous nucleic acids encoding an OAC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:77 or SEQ

ID NO:163. In some embodiments, the one or more heterologous nucleic acids encoding an OAC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:77 or SEQ ID NO:163, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an OAC polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:77 or SEQ ID NO:163.

[00383] In some embodiments, the one or more heterologous nucleic acids encoding a TKS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:75. In some embodiments, the one or more heterologous nucleic acids encoding a TKS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:75, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a TKS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:75. In some embodiments, the one or more heterologous nucleic acids encoding a TKS polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:75.

[00384] In some embodiments, the one or more heterologous nucleic acids encoding a TKS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:162. In some embodiments, the one or more heterologous nucleic acids encoding a TKS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:162, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a TKS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:162. In some embodiments, the one or more heterologous nucleic acids encoding a TKS polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to

SEQ ID NO:162. In some embodiments, the one or more heterologous nucleic acids encoding a TKS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:162.

[00385] In some embodiments, the one or more heterologous nucleic acids encoding an OAC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:77. In some embodiments, the one or more heterologous nucleic acids encoding an OAC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:77, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an OAC polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:77. In some embodiments, the one or more heterologous nucleic acids encoding an OAC polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:77.

[00386] In some embodiments, the one or more heterologous nucleic acids encoding an OAC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:163. In some embodiments, the one or more heterologous nucleic acids encoding an OAC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:163, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an OAC polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:163. In some embodiments, the one or more heterologous nucleic acids encoding an OAC polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:163.

[00387] In some embodiments, the one or more heterologous nucleic acids encoding a TKS/OAC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:79. In some embodiments, the one or more heterologous nucleic acids encoding a TKS/OAC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:79, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a TKS/OAC polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:79. In some embodiments, the one or more heterologous nucleic acids encoding a TKS/OAC polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:79. In some embodiments, the one or more heterologous nucleic acids encoding a TKS/OAC polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:79.

Polypeptides that Generate Geranyl Pyrophosphate, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00388] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a polypeptide that generates GPP. In some embodiments, the polypeptide that generates GPP is a geranyl diphosphate synthase (GPPS) polypeptide. In some embodiments, the GPPS polypeptide also has a farnesyl diphosphate synthase (FPPS) polypeptide activity. In some embodiments, the GPPS polypeptide is modified such that it has reduced FPPS polypeptide activity (e.g., at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or more than 90%, less FPPS polypeptide activity) than the corresponding wild-type or parental GPPS polypeptide from which the modified GPPS polypeptide is derived. In some embodiments, the GPPS polypeptide is modified such that it has substantially no FPPS polypeptide activity. In some embodiments, a genetically modified host cell of the present disclosure is

genetically modified with one or more heterologous nucleic acids encoding a GPPS polypeptide.

[00389] Exemplary GPPS polypeptides disclosed herein may include a full-length GPPS polypeptide, a fragment of a GPPS polypeptide, a variant of a GPPS polypeptide, a truncated GPPS polypeptide, or a fusion polypeptide that has at least one activity of a GPPS polypeptide. In some embodiments, the one or more polypeptides that generate GPP or are part of a biosynthetic pathway that generates GPP are one or more polypeptides having at least one activity of a polypeptide present in the mevalonate (MEV) pathway. In some embodiments, the one or more polypeptides that generate GPP or are part of a biosynthetic pathway that generates GPP are one or more polypeptides having at least one activity of a polypeptide present in the DXP pathway.

[00390] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60.

[00391] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:121, SEQ ID NO:123, SEQ ID NO:125, SEQ ID NO:127, SEQ ID NO:129, SEQ ID NO:131, SEQ ID NO:133, SEQ ID NO:135, SEQ ID NO:137, SEQ ID NO:139, SEQ ID NO:141, SEQ ID NO:143, or SEQ ID NO:203. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:121, SEQ ID NO:123, SEQ ID NO:125, SEQ ID NO:127, SEQ ID NO:129, SEQ ID NO:131, SEQ ID NO:133, SEQ ID NO:135, SEQ ID NO:137, SEQ ID NO:139, SEQ ID NO:141, SEQ ID NO:143, or SEQ ID NO:203, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide

encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:121, SEQ ID NO:123, SEQ ID NO:125, SEQ ID NO:127, SEQ ID NO:129, SEQ ID NO:131, SEQ ID NO:133, SEQ ID NO:135, SEQ ID NO:137, SEQ ID NO:139, SEQ ID NO:141, SEQ ID NO:143, or SEQ ID NO:203.

[00392] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:5 or SEQ ID NO:6. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:5 or SEQ ID NO:6, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:5 or SEQ ID NO:6.

[00393] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: i) one or more heterologous nucleic acids that encode a GPPS polypeptide comprising an amino acid sequence as set forth in SEQ ID NO:5; and ii) one or more heterologous nucleic acids that encode a GPPS polypeptide comprising an amino acid sequence as set forth in SEQ ID NO:6.

[00394] In some embodiments, the GPPS (Erg20) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:7. In some embodiments, the GPPS (Erg20) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:7, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS (Erg20) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least

70%, or at least 75% amino acid sequence identity to SEQ ID NO:7. In some embodiments, the GPPS (Erg20) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:7. In some embodiments, the GPPS (Erg20) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:7.

[00395] In some embodiments, the GPPS (Erg20 (K197G)) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:8. In some embodiments, the GPPS (Erg20 (K197G)) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:8, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS (Erg20 (K197G)) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:8. In some embodiments, the GPPS (Erg20 (K197G)) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:8. In some embodiments, the GPPS (Erg20 (K197G)) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:8. The GPPS (Erg20 (K197G)) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:8 comprises a K197G amino acid substitution relative to the GPPS amino acid sequence set forth in SEQ ID NO:7. This mutation shifts the ratio of GPP to farnesyl diphosphate (FPP), increasing the production of the GPP required to produce CBDA.

[00396] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a

GPPS large subunit polypeptide and a GPPS small subunit polypeptide, where the GPPS large subunit polypeptide and the GPPS small subunit polypeptide together form a heterodimeric GPPS polypeptide. In some embodiments, the GPPS large subunit polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:72. In some embodiments, the GPPS large subunit polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:72, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS large subunit polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:72. In some embodiments, the GPPS small subunit polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:74. In some embodiments, the GPPS small subunit polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:74, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS small subunit polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:74.

[00397] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids is an ERG20mut (F96W, N127W) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:60. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids is an ERG20mut (F96W, N127W) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:60, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids is an ERG20mut (F96W, N127W) polypeptide and comprises an amino acid sequence having at least 50%, at least

55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:60. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids is an ERG20mut (F96W, N127W) polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:60. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids is an ERG20mut (F96W, N127W) polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:60. This mutation shifts the ratio of GPP to farnesyl diphosphate (FPP), increasing the production of the GPP required to produce CBDA.

[00398] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:121. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:121, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:121. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:121. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:121.

[00399] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:123. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:123, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS

polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:123. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:123. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:123.

[00400] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:125. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:125, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:125. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:125. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:125.

[00401] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:127. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:127, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at

least 75% amino acid sequence identity to SEQ ID NO:127. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:127. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:127.

[00402] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:129. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:129, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:129. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:129. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:129.

[00403] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:131. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:131, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:131. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an

amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:131. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:131.

[00404] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:133. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:133, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:133. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:133. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:133.

[00405] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:135. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:135, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:135. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:135. In some embodiments, the GPPS

polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:135.

[00406] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:137. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:137, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:137. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:137. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:137.

[00407] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:139. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:139, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:139. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:139. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least

90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:139.

[00408] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:141. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:141, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:141. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:141. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:141.

[00409] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:143. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:143, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:143. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:143. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at

least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:143.

[00410] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:203. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:203, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:203. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:203. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:203.

[00411] Exemplary GPPS heterologous nucleic acids disclosed herein may include nucleic acids that encode a GPPS polypeptide, such as, a full-length GPPS polypeptide, a fragment of a GPPS polypeptide, a variant of a GPPS polypeptide, a truncated GPPS polypeptide, or a fusion polypeptide that has at least one activity of a GPPS polypeptide.

[00412] In some embodiments, the GPPS polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the GPPS polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the GPPS polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of a GPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of a GPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of a GPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a GPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of a

GPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has six copies of a GPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has seven copies of a GPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has eight copies of a GPPS polypeptide-encoding heterologous nucleic acid.

[00413] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:122, SEQ ID NO:124, SEQ ID NO:126, SEQ ID NO:128, SEQ ID NO:130, SEQ ID NO:132, SEQ ID NO:134, SEQ ID NO:136, SEQ ID NO:138, SEQ ID NO:140, SEQ ID NO:142, SEQ ID NO:144, or SEQ ID NO:202. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:122, SEQ ID NO:124, SEQ ID NO:126, SEQ ID NO:128, SEQ ID NO:130, SEQ ID NO:132, SEQ ID NO:134, SEQ ID NO:136, SEQ ID NO:138, SEQ ID NO:140, SEQ ID NO:142, SEQ ID NO:144, or SEQ ID NO:202, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:122, SEQ ID NO:124, SEQ ID NO:126, SEQ ID NO:128, SEQ ID NO:130, SEQ ID NO:132, SEQ ID NO:134, SEQ ID NO:136, SEQ ID NO:138, SEQ ID NO:140, SEQ ID NO:142, SEQ ID NO:144, or SEQ ID NO:202.

[00414] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:71 and/or SEQ ID NO:73. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:71 and/or SEQ ID NO:73, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%,

at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:71 and/or SEQ ID NO:73.

[00415] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide (ERG20mut (F96W, N127W)) comprise the nucleotide sequence set forth in SEQ ID NO:59. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide (ERG20mut (F96W, N127W)) comprise the nucleotide sequence set forth in SEQ ID NO:59, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide (ERG20mut (F96W, N127W)) comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:59. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide (ERG20mut (F96W, N127W)) comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:59.

[00416] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide (ERG20mut (F96W, N127W)) comprise the nucleotide sequence set forth in SEQ ID NO:161. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide (ERG20mut (F96W, N127W)) comprise the nucleotide sequence set forth in SEQ ID NO:161, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide (ERG20mut (F96W, N127W)) comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:161. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide (ERG20mut (F96W, N127W)) comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:161. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide (ERG20mut (F96W, N127W)) comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least

96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:161.

[00417] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:122. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:122, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:122. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:122.

[00418] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:124. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:124, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:124. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:124.

[00419] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:126. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:126, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at

least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:126. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:126.

[00420] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:128. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:128, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:128. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:128.

[00421] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:130. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:130, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:130. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:130.

[00422] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:132. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:132, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:132. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:132.

[00423] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:134. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:134, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:134. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:134.

[00424] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:136. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:136, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:136. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS

polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:136.

[00425] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:138. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:138, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:138. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:138.

[00426] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:140. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:140, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:140. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:140.

[00427] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:142. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide

comprise the nucleotide sequence set forth in SEQ ID NO:142, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:142. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:142.

[00428] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:144. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:144, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:144. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:144.

[00429] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:202. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:202, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:202. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at

least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:202.

NphB Polypeptides, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00430] In some embodiments, a NphB polypeptide is used instead of a GOT polypeptide to generate cannabigerolic acid from GPP and olivetolic acid. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a NphB polypeptide.

[00431] Exemplary NphB polypeptides disclosed herein may include a full-length NphB polypeptide, a fragment of a NphB polypeptide, a variant of a NphB polypeptide, a truncated NphB polypeptide, or a fusion polypeptide that has at least one activity of a NphB polypeptide.

[00432] In some embodiments, the NphB polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:84. In some embodiments, the NphB polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:84, or a conservatively substituted amino acid sequence thereof. In some embodiments, the NphB polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:84.

[00433] Exemplary NphB heterologous nucleic acids disclosed herein may include nucleic acids that encode a NphB polypeptide, such as, a full-length NphB polypeptide, a fragment of a NphB polypeptide, a variant of a NphB polypeptide, a truncated NphB polypeptide, or a fusion polypeptide that has at least one activity of a NphB polypeptide.

[00434] In some embodiments, the NphB polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the NphB polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the NphB polypeptide-encoding heterologous nucleic acid to a

strong promoter. In some embodiments, the genetically modified host cell has one copy of a NphB polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of a NphB polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of a NphB polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a NphB polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of a NphB polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has six copies of a NphB polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has seven copies of a NphB polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has eight copies of a NphB polypeptide-encoding heterologous nucleic acid.

[00435] In some embodiments, the one or more heterologous nucleic acids encoding a NphB polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:83. In some embodiments, the one or more heterologous nucleic acids encoding a NphB polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:83, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a NphB polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:83.

Polypeptides that Generate Neryl Pyrophosphate or Cannabinerolic Acid, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00436] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a neryl pyrophosphate (NPP) synthase (NPPS) polypeptide (FIG. 11). NPP and olivetolic acid may be substrates to generate cannabinerolic acid (CBNRA). In some embodiments, a GOT polypeptide acts on NPP and an olivetolic acid derivative (as described elsewhere herein) to generate a CBNRA derivative. Cannabinerolic acid or derivatives thereof can serve as a

substrate for a CBDAS or THCAS polypeptide to generate CBDA or THCA, or derivatives thereof, respectively.

[00437] Exemplary NPPS polypeptides disclosed herein may include a fragment of a NPPS polypeptide, a variant of a NPPS polypeptide, a full-length NPPS polypeptide, a truncated NPPS polypeptide, or a fusion polypeptide that has at least one activity of a NPPS polypeptide.

[00438] In some embodiments, the NPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:70. In some embodiments, the NPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:70, or a conservatively substituted amino acid sequence thereof. In some embodiments, the NPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:70. In some embodiments, the NPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:70. In some embodiments, the NPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:70. In some embodiments, the NPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:70.

[00439] Exemplary NPPS heterologous nucleic acids disclosed herein may include nucleic acids that encode a NPPS polypeptide, such as, a full-length NPPS polypeptide, a fragment of a NPPS polypeptide, a variant of a NPPS polypeptide, a truncated NPPS polypeptide, or a fusion polypeptide that has at least one activity of a NPPS polypeptide.

[00440] In some embodiments, the NPPS polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the NPPS polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the NPPS polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of an NPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of an NPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of an NPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of an NPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of an NPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has six copies of an NPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has seven copies of an NPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has eight copies of an NPPS polypeptide-encoding heterologous nucleic acid.

[00441] In some embodiments, the one or more heterologous nucleic acids encoding a NPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:69. In some embodiments, the one or more heterologous nucleic acids encoding a NPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:69, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a NPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:69. In some embodiments, the one or more heterologous nucleic acids encoding a NPPS polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:69. In some embodiments, the one or more heterologous nucleic acids encoding a NPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least

95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:69.

Polypeptides that Generate Acetyl-CoA from Pyruvate, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00442] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a polypeptide that generates acetyl-CoA from pyruvate. Polypeptides that generate acetyl-CoA from pyruvate may include a pyruvate dehydrogenase complex (PDC) polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a PDC polypeptide.

[00443] Exemplary PDC polypeptides disclosed herein may include a full-length PDC polypeptide, a fragment of a PDC polypeptide, a variant of a PDC polypeptide, a truncated PDC polypeptide, or a fusion polypeptide that has at least one activity of a PDC polypeptide.

[00444] In some embodiments, the PDC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:117. In some embodiments, the PDC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:117, or a conservatively substituted amino acid sequence thereof. In some embodiments, the PDC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:117. In some embodiments, the PDC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:117. In some embodiments, the PDC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:117. In some embodiments, the PDC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%,

at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:117.

[00445] Exemplary PDC heterologous nucleic acids disclosed herein may include nucleic acids that encode a PDC polypeptide, such as, a full-length PDC polypeptide, a fragment of a PDC polypeptide, a variant of a PDC polypeptide, a truncated PDC polypeptide, or a fusion polypeptide that has at least one activity of a PDC polypeptide.

[00446] In some embodiments, the PDC polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the PDC polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the PDC polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of a PDC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of a PDC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of a PDC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a PDC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of a PDC polypeptide-encoding heterologous nucleic acid.

[00447] In some embodiments, the one or more heterologous nucleic acids encoding a PDC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:118. In some embodiments, the one or more heterologous nucleic acids encoding a PDC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:118, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a PDC polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:118. In some embodiments, the one or more heterologous nucleic acids encoding a PDC polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:118. In some embodiments, the one or more heterologous nucleic acids encoding a PDC polypeptide comprise a nucleotide sequence having at least 80%, at least

81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:118.

Polypeptides that Condense Two Molecules of Acetyl-CoA to Generate Acetoacetyl-CoA, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00448] In some embodiments, the host cell is genetically modified with one or more heterologous nucleic acids encoding a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA. In some embodiments, the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase (ERG10p) polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding an acetoacetyl-CoA thiolase polypeptide.

[00449] Exemplary acetoacetyl-CoA thiolase polypeptides disclosed herein may include a full-length acetoacetyl-CoA thiolase polypeptide, a fragment of an acetoacetyl-CoA thiolase polypeptide, a variant of an acetoacetyl-CoA thiolase polypeptide, a truncated acetoacetyl-CoA thiolase polypeptide, or a fusion polypeptide that has at least one activity of an acetoacetyl-CoA thiolase polypeptide.

[00450] In some embodiments, the acetoacetyl-CoA thiolase (ERG10p) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:25. In some embodiments, the acetoacetyl-CoA thiolase (ERG10p) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:25, or a conservatively substituted amino acid sequence thereof. In some embodiments, the acetoacetyl-CoA thiolase (ERG10p) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:25. In some embodiments, the acetoacetyl-CoA thiolase (ERG10p) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:25. In some embodiments, the acetoacetyl-CoA thiolase (ERG10p) polypeptide encoded by the one or

more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:25. In some embodiments, the acetoacetyl-CoA thiolase (ERG10p) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:25.

[00451] Exemplary acetoacetyl-CoA thiolase heterologous nucleic acids disclosed herein may include nucleic acids that encode an acetoacetyl-CoA thiolase polypeptide, such as, a full-length acetoacetyl-CoA thiolase polypeptide, a fragment of an acetoacetyl-CoA thiolase polypeptide, a variant of an acetoacetyl-CoA thiolase polypeptide, a truncated acetoacetyl-CoA thiolase polypeptide, or a fusion polypeptide that has at least one activity of an acetoacetyl-CoA thiolase polypeptide.

[00452] In some embodiments, the acetoacetyl-CoA thiolase polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the acetoacetyl-CoA thiolase polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the acetoacetyl-CoA thiolase polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of an acetoacetyl-CoA thiolase polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of an acetoacetyl-CoA thiolase polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of an acetoacetyl-CoA thiolase polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of an acetoacetyl-CoA thiolase polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of an acetoacetyl-CoA thiolase polypeptide-encoding heterologous nucleic acid.

[00453] In some embodiments, the one or more heterologous nucleic acids encoding an acetoacetyl-CoA thiolase (ERG10p) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:157. In some embodiments, the one or more heterologous nucleic acids encoding a ERG10p polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:157, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an acetoacetyl-CoA thiolase (ERG10p) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:157. In some embodiments, the one or more heterologous nucleic acids encoding an acetoacetyl-CoA thiolase (ERG10p) polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:157. In some embodiments, the one or more heterologous nucleic acids encoding an acetoacetyl-CoA thiolase (ERG10p) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:157.

[00454] In some embodiments, the one or more heterologous nucleic acids encoding an acetoacetyl-CoA thiolase (ERG10p) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:209. In some embodiments, the one or more heterologous nucleic acids encoding a ERG10p polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:209, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an acetoacetyl-CoA thiolase (ERG10p) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:209. In some embodiments, the one or more heterologous nucleic acids encoding an acetoacetyl-CoA thiolase (ERG10p) polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:209. In some embodiments, the one or more heterologous nucleic acids

encoding an acetoacetyl-CoA thiolase (ERG10p) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:209.

Mevalonate Pathway Polypeptides, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00455] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding one or more polypeptides having at least one activity of a polypeptide present in the mevalonate (MEV) pathway.

[00456] In some embodiments, the one or more polypeptides that generate GPP or are part of a biosynthetic pathway that generates GPP are one or more polypeptides having at least one activity of a polypeptide present in the mevalonate pathway. The mevalonate pathway may comprise polypeptides that catalyze the following steps: (a) condensing two molecules of acetyl-CoA to generate acetoacetyl-CoA (e.g., by action of an acetoacetyl-CoA thiolase polypeptide); (b) condensing acetoacetyl-CoA with acetyl-CoA to form hydroxymethylglutaryl-CoA (HMG-CoA) (e.g., by action of a HMGS polypeptide); (c) converting HMG-CoA to mevalonate (e.g., by action of an HMGR polypeptide); (d) phosphorylating mevalonate to mevalonate 5-phosphate (e.g., by action of a MK polypeptide); (e) converting mevalonate 5-phosphate to mevalonate 5-pyrophosphate (e.g., by action of a PMK polypeptide); (f) converting mevalonate 5-pyrophosphate to isopentenyl pyrophosphate (e.g., by action of a mevalonate pyrophosphate decarboxylase (MPD or MVD) polypeptide); and (g) converting isopentenyl pyrophosphate to dimethylallyl pyrophosphate (e.g., by action of an isopentenyl pyrophosphate isomerase (IDI) polypeptide).

[00457] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a MEV pathway polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than one MEV pathway polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more

heterologous nucleic acids encoding more than two MEV pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than three MEV pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than four MEV pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than five MEV pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than six MEV pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding all MEV pathway polypeptides.

[00458] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding two MEV pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding three MEV pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding four MEV pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding five MEV pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding six MEV pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, 4, 5, 6, or more MEV pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, 4, 5, or 6 MEV pathway polypeptides.

[00459] Exemplary MEV pathway polypeptides disclosed herein may include a full-length MEV pathway polypeptide, a fragment of a MEV pathway polypeptide, a variant of a MEV pathway polypeptide, a truncated MEV pathway polypeptide, or a fusion polypeptide that has at least one activity of a MEV pathway polypeptide. In some embodiments, the one or more MEV pathway polypeptides are selected from the group consisting of an

acetoacetyl-CoA thiolase polypeptide, a HMGS polypeptide, an HMGR polypeptide, an MK polypeptide, a PMK polypeptide, an MVD polypeptide, and an IDI polypeptide.

[00460] In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115, or a conservatively substituted amino acid sequence thereof. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115.

[00461] In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is a MvaS polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:23. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is a MvaS polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:23, or a conservatively substituted amino acid sequence thereof. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is a MvaS polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:23. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is a MvaS polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:23. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is a MvaS polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:23.

[00462] In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is a MvaS polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:56. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is a MvaS polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:56, or a conservatively substituted amino acid sequence thereof. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is a MvaS polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:56. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is a MvaS polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:56. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is a MvaS polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:56.

[00463] In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is an ERG13 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:24. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is an ERG13 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:24, or a conservatively substituted amino acid sequence thereof. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is an ERG13 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:24. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is an ERG13 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:24. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is an ERG13 polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at

least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:24.

[00464] In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is an ERG13 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:115. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is an ERG13 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:115, or a conservatively substituted amino acid sequence thereof. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is an ERG13 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:115. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is an ERG13 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:115. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is an ERG13 polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:115.

[00465] In some embodiments, the HMGR polypeptide encoded by the one or more heterologous nucleic acids is a MvaE polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:22. In some embodiments, the HMGR polypeptide encoded by the one or more heterologous nucleic acids is a MvaE polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:22, or a conservatively substituted amino acid sequence thereof. In some embodiments, the HMGR polypeptide encoded by the one or more heterologous nucleic acids is a MvaE polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:22. In some embodiments, the HMGR polypeptide encoded by the one or more heterologous nucleic acids is a MvaE polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:22. In some embodiments, the HMGR polypeptide encoded by the one or more heterologous nucleic acids is a MvaE polypeptide and comprises an amino acid sequence having at least 85%, at

least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:22. In some embodiments, the HMGR polypeptide encoded by the one or more heterologous nucleic acids is a MvaE polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:22.

[00466] In some embodiments, the HMGR polypeptide encoded by the one or more heterologous nucleic acids is a MvaE polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:54. In some embodiments, the HMGR polypeptide encoded by the one or more heterologous nucleic acids is a MvaE polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:54, or a conservatively substituted amino acid sequence thereof. In some embodiments, the HMGR polypeptide encoded by the one or more heterologous nucleic acids is a MvaE polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:54. In some embodiments, the HMGR polypeptide encoded by the one or more heterologous nucleic acids is a MvaE polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:54. In some embodiments, the HMGR polypeptide encoded by the one or more heterologous nucleic acids is a MvaE polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:54.

[00467] In some embodiments, the HMGR polypeptide is a truncated HMGR (tHMGR) polypeptide. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises the

amino acid sequence set forth in SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208, or a conservatively substituted amino acid sequence thereof. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208.

[00468] In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:17. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:17, or a conservatively substituted amino acid sequence thereof. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:17. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:17. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:17.

[00469] In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:52. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:52, or a conservatively substituted amino acid sequence thereof. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:52. In some embodiments, the

tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:52. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:52.

[00470] In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:113. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:113, or a conservatively substituted amino acid sequence thereof. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:113. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:113. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:113.

[00471] In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:208. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:208, or a conservatively substituted amino acid sequence thereof. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:208. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least

84% amino acid sequence identity to SEQ ID NO:208. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:208.

[00472] In some embodiments, the MK (ERG12) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:64. In some embodiments, the MK (ERG12) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:64, or a conservatively substituted amino acid sequence thereof. In some embodiments, the MK (ERG12) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:64. In some embodiments, the MK (ERG12) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:64. In some embodiments, the MK (ERG12) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:64. In some embodiments, the MK (ERG12) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:64.

[00473] In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:62 or SEQ ID NO:205. In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:62 or

SEQ ID NO:205, or a conservatively substituted amino acid sequence thereof. In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:62 or SEQ ID NO:205.

[00474] In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:62. In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:62, or a conservatively substituted amino acid sequence thereof. In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:62. In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:62. In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:62.

[00475] In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids is an ERG8 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:205. In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids is an ERG8 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:205, or a conservatively substituted amino acid sequence thereof. In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids is an ERG8 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:205. In some embodiments, the PMK

polypeptide encoded by the one or more heterologous nucleic acids is an ERG8 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:205. In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids is an ERG8 polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:205.

[00476] In some embodiments, a PMK polypeptide and MK polypeptide are fused into a single polypeptide chain (a PMK/MK fusion polypeptide). In some embodiments, the PMK/MK polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:68. In some embodiments, the PMK/MK polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:68, or a conservatively substituted amino acid sequence thereof. In some embodiments, the PMK/MK polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:68. In some embodiments, the PMK/MK polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:68. In some embodiments, the PMK/MK polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:68. In some embodiments, the PMK/MK polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:68.

[00477] In some embodiments, the MVD polypeptide encoded by the one or more heterologous nucleic acids is an ERG19 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:66. In some embodiments, the MVD polypeptide encoded by the one or more heterologous nucleic acids is an ERG19 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:66, or a conservatively substituted amino acid sequence thereof. In some embodiments, the MVD polypeptide encoded by the one or more heterologous nucleic acids is an ERG19 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:66. In some embodiments, the MVD polypeptide encoded by the one or more heterologous nucleic acids is an ERG19 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:66. In some embodiments, the MVD polypeptide encoded by the one or more heterologous nucleic acids is an ERG19 polypeptide and comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:66. In some embodiments, the MVD polypeptide encoded by the one or more heterologous nucleic acids is an ERG19 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:66.

[00478] In some embodiments, the IDI1 polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:58. In some embodiments, the IDI1 polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:58, or a conservatively substituted amino acid sequence thereof. In some embodiments, the IDI1 polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:58. In some embodiments, the IDI1 polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid

sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:58. In some embodiments, the IDI1 polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:58. In some embodiments, the IDI1 polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:58.

[00479] Exemplary MEV pathway heterologous nucleic acids disclosed herein may include nucleic acids that encode a MEV pathway polypeptide, such as, a full-length MEV pathway polypeptide, a fragment of a MEV pathway polypeptide, a variant of a MEV pathway polypeptide, a truncated MEV pathway polypeptide, or a fusion polypeptide that has at least one activity of a polypeptide that is part of the MEV pathway.

[00480] In some embodiments, the MEV pathway polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of a MEV pathway polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the MEV pathway polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of a MEV pathway polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of a MEV pathway polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of a MEV pathway polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a MEV pathway polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of a MEV pathway polypeptide-encoding heterologous nucleic acid.

[00481] In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (mvaS) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:55. In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (mvaS) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:55, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (mvaS) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:55. In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (mvaS) polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:55.

[00482] In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (ERG13) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:116 or SEQ ID NO:120. In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (ERG13) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:116 or SEQ ID NO:120, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (ERG13) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:116 or SEQ ID NO:120.

[00483] In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (ERG13) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:116. In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (ERG13) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:116, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (ERG13) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:116. In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (ERG13) polypeptide comprise a nucleotide sequence

having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:116.

[00484] In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (ERG13) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:120. In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (ERG13) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:120, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (ERG13) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:120. In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (ERG13) polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:120.

[00485] In some embodiments, the one or more heterologous nucleic acids encoding an HMGR (mvaE) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:53. In some embodiments, the one or more heterologous nucleic acids encoding an HMGR (mvaE) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:53, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an HMGR (mvaE) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:53. In some embodiments, the one or more heterologous nucleic acids encoding an HMGR (mvaE) polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:53.

[00486] In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:51, SEQ ID NO:114, or SEQ ID NO:119. In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise the nucleotide sequence set forth in SEQ ID

NO:51, SEQ ID NO:114, or SEQ ID NO:119, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:51, SEQ ID NO:114, or SEQ ID NO:119.

[00487] In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:51. In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:51, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:51. In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:51.

[00488] In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:114. In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:114, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:114. In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:114.

[00489] In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:119. In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:119, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:119. In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:119.

[00490] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with two or more heterologous nucleic acids that encode a tHMGR polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with two heterologous nucleic acids that encode a tHMGR polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with two or more heterologous nucleic acids that encode an HMGR polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with two heterologous nucleic acids that encode an HMGR polypeptide.

[00491] In some embodiments, the one or more heterologous nucleic acids encoding an MK (ERG12) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:63 or SEQ ID NO:206. In some embodiments, the one or more heterologous nucleic acids encoding an MK (ERG12) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:63 or SEQ ID NO:206, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an MK (ERG12) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:63 or SEQ ID NO:206.

[00492] In some embodiments, the one or more heterologous nucleic acids encoding an MK (ERG12) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:63. In some embodiments, the one or more heterologous nucleic acids encoding an MK (ERG12) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:63, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an MK (ERG12) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:63. In some embodiments, the one or more heterologous nucleic acids encoding an MK (ERG12) polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:63.

[00493] In some embodiments, the one or more heterologous nucleic acids encoding an MK (ERG12) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:206. In some embodiments, the one or more heterologous nucleic acids encoding an MK (ERG12) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:206, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an MK (ERG12) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:206. In some embodiments, the one or more heterologous nucleic acids encoding an MK (ERG12) polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:206.

[00494] In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:61, SEQ ID NO:160, or SEQ ID NO:204. In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:61, SEQ ID NO:160, or SEQ ID NO:204, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at

least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:61, SEQ ID NO:160, or SEQ ID NO:204.

[00495] In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:61. In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:61, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:61. In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:61.

[00496] In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:160. In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:160, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:160. In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:160.

[00497] In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:204. In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:204, or a codon

degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:204. In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:204.

[00498] In some embodiments, the one or more heterologous nucleic acids encoding a PMK/MK polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:67. In some embodiments, the one or more heterologous nucleic acids encoding a PMK/MK polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:67, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a PMK/MK polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:67. In some embodiments, the one or more heterologous nucleic acids encoding a PMK/MK polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:67. In some embodiments, the one or more heterologous nucleic acids encoding a PMK/MK polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:67.

[00499] In some embodiments, the one or more heterologous nucleic acids encoding an MVD (ERG19) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:65 or SEQ ID NO:158. In some embodiments, the one or more heterologous nucleic acids encoding an MVD (ERG19) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:65 or SEQ ID NO:158, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an MVD (ERG19)

polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:65 or SEQ ID NO:158.

[00500] In some embodiments, the one or more heterologous nucleic acids encoding an MVD (ERG19) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:65. In some embodiments, the one or more heterologous nucleic acids encoding an MVD (ERG19) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:65, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an MVD (ERG19) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:65. In some embodiments, the one or more heterologous nucleic acids encoding an MVD (ERG19) polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:65.

[00501] In some embodiments, the one or more heterologous nucleic acids encoding an MVD (ERG19) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:158. In some embodiments, the one or more heterologous nucleic acids encoding an MVD (ERG19) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:158, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an MVD (ERG19) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:158. In some embodiments, the one or more heterologous nucleic acids encoding an MVD (ERG19) polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:158.

[00502] In some embodiments, the one or more heterologous nucleic acids encoding an IDI1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:57 or SEQ ID

NO:159. In some embodiments, the one or more heterologous nucleic acids encoding an IDI1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:57 or SEQ ID NO:159, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an IDI1 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:57 or SEQ ID NO:159.

[00503] In some embodiments, the one or more heterologous nucleic acids encoding an IDI1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:57. In some embodiments, the one or more heterologous nucleic acids encoding an IDI1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:57, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an IDI1 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:57. In some embodiments, the one or more heterologous nucleic acids encoding an IDI1 polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:57.

[00504] In some embodiments, the one or more heterologous nucleic acids encoding an IDI1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:159. In some embodiments, the one or more heterologous nucleic acids encoding an IDI1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:159, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an IDI1 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:159. In some embodiments, the one or more heterologous nucleic acids encoding an IDI1 polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at

least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:159.

Polypeptides that Modulate NADH or NADPH Redox Balance, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00505] In some embodiments, the host cell is genetically modified with one or more heterologous nucleic acids encoding a polypeptide that modulates NADH or NADPH redox balance. GPP production has a redox imbalance in it that can be modulated by changing NADPH-using enzymes to NADH-using enzymes, bringing redox into better balance.

[00506] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a polypeptide that modulates NADH or NADPH redox balance. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than one polypeptide that modulates NADH or NADPH redox balance. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than two polypeptides that modulate NADH or NADPH redox balance. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than three polypeptides that modulate NADH or NADPH redox balance. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding two polypeptides that modulate NADH or NADPH redox balance. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding three polypeptides that modulate NADH or NADPH redox balance. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, or more polypeptides that modulate NADH or NADPH redox balance. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, or 3 polypeptides that modulate NADH or NADPH redox balance.

[00507] Exemplary polypeptides that modulate NADH or NADPH redox balance disclosed herein may include a full-length polypeptide that modulates NADH or NADPH redox balance, a fragment of a polypeptide that modulates NADH or NADPH redox balance,

a variant of a polypeptide that modulates NADH or NADPH redox balance, a truncated polypeptide that modulates NADH or NADPH redox balance, or a fusion polypeptide that has at least one activity of a polypeptide that modulates NADH or NADPH redox balance.

[00508] Exemplary heterologous nucleic acids disclosed herein may include nucleic acids that encode a polypeptide that modulates NADH or NADPH redox balance, such as, a full-length polypeptide that modulates NADH or NADPH redox balance, a fragment of a polypeptide that modulates NADH or NADPH redox balance, a variant of a polypeptide that modulates NADH or NADPH redox balance, a truncated polypeptide that modulates NADH or NADPH redox balance, or a fusion polypeptide that has at least one activity of a polypeptide that modulates NADH or NADPH redox balance.

[00509] In some embodiments, the polypeptide that modulates NADH or NADPH redox balance is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the heterologous nucleic acid encoding a polypeptide that modulates NADH or NADPH redox balance, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the polypeptide that modulates NADH or NADPH redox balance encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of a heterologous nucleic acid encoding a polypeptide that modulates NADH or NADPH redox balance. In some embodiments, the genetically modified host cell has two copies of a heterologous nucleic acid encoding a polypeptide that modulates NADH or NADPH redox balance. In some embodiments, the genetically modified host cell has three copies of a heterologous nucleic acid encoding a polypeptide that modulates NADH or NADPH redox balance. In some embodiments, the genetically modified host cell has four copies of a heterologous nucleic acid encoding a polypeptide that modulates NADH or NADPH redox balance. In some embodiments, the genetically modified host cell has five copies of a heterologous nucleic acid encoding a polypeptide that modulates NADH or NADPH redox balance.

DXP Pathway Polypeptides, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00510] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding one or more polypeptides having at least one activity of a polypeptide present in the deoxyxylulose-5-phosphate (DXP) pathway.

[00511] In some embodiments, the one or more polypeptides that generate GPP or are part of a biosynthetic pathway that generates GPP are polypeptides of the DXP pathway. The term "1-deoxy-D-xylulose 5-diphosphate pathway" or "DXP pathway" as used herein may refer to the pathway that converts glyceraldehyde-3-phosphate and pyruvate to IPP and DMAPP through a DXP pathway intermediate.

[00512] In the DXP pathway, pyruvate and D-glyceraldehyde-3-phosphate are converted via a series of reactions to IPP and DMAPP. The pathway involves action of the following polypeptides: a 1-deoxy-D-xylulose-5-phosphate synthase (Dxs) polypeptide, a 1-deoxy-D-xylulose-5-phosphate reductoisomerase (IspC) polypeptide, a 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase (IspD) polypeptide, a 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (IspE) polypeptide, a 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate synthase (IspF) polypeptide, a 1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate synthase (IspG) polypeptide, and an isopentenyl diphosphate isomerase (IspH) polypeptide.

[00513] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a DXP pathway polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than one DXP pathway polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than two DXP pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than three DXP pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than four DXP pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than five DXP pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than six DXP pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding all DXP pathway polypeptides.

[00514] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding two

DXP pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding three DXP pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding four DXP pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding five DXP pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding six DXP pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, 4, 5, 6, or more DXP pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, 4, 5, or 6 DXP pathway polypeptides.

[00515] Exemplary polypeptides disclosed herein that are part of the DXP pathway may include a full-length DXP pathway polypeptide, a fragment of a DXP pathway polypeptide, a variant of a DXP pathway polypeptide, a truncated DXP pathway polypeptide, or a fusion polypeptide that has at least one activity of a polypeptide that is part of the DXP pathway.

[00516] Examples of polypeptides of the DXP pathway are set forth in SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40.

[00517] In some embodiments, the 1-deoxy-D-xylulose-5-phosphate synthase (Dxs) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:32. In some embodiments, the 1-deoxy-D-xylulose-5-phosphate synthase (Dxs) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:32, or a conservatively substituted amino acid sequence thereof. In some embodiments, the 1-deoxy-D-xylulose-5-phosphate synthase (Dxs) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at

least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:32.

[00518] In some embodiments, the 1-deoxy-D-xylulose 5-phosphate reductoisomerase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:33. In some embodiments, the 1-deoxy-D-xylulose 5-phosphate reductoisomerase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:33, or a conservatively substituted amino acid sequence thereof. In some embodiments, the 1-deoxy-D-xylulose 5-phosphate reductoisomerase polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:33.

[00519] In some embodiments, the 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:34. In some embodiments, the 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:34, or a conservatively substituted amino acid sequence thereof. In some embodiments, the 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:34.

[00520] In some embodiments, the 4-diphosphocytidyl-2-C-methylerythritol kinase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:35. In some embodiments, the 4-diphosphocytidyl-2-C-methylerythritol kinase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:35, or a conservatively

substituted amino acid sequence thereof. In some embodiments, the 4-diphosphocytidyl-2-C-methylerythritol kinase polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:35.

[00521] In some embodiments, the 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:36. In some embodiments, the 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:36, or a conservatively substituted amino acid sequence thereof. In some embodiments, the 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:36.

[00522] In some embodiments, the 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:37. In some embodiments, the 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:37, or a conservatively substituted amino acid sequence thereof. In some embodiments, the 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least

99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:37.

[00523] In some embodiments, the 4-hydroxy-3-methylbut-2-enyl diphosphate reductase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:38. In some embodiments, the 4-hydroxy-3-methylbut-2-enyl diphosphate reductase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:38, or a conservatively substituted amino acid sequence thereof. In some embodiments, the 4-hydroxy-3-methylbut-2-enyl diphosphate reductase polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:38.

[00524] In some embodiments, the isopentenyl diphosphate (IPP) isomerase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:39. In some embodiments, the isopentenyl diphosphate (IPP) isomerase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:39, or a conservatively substituted amino acid sequence thereof. In some embodiments, the isopentenyl diphosphate (IPP) isomerase polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:39.

[00525] In some embodiments, the DXP pathway polypeptide is a mutated FPP synthase polypeptide. In some embodiments, the mutated FPP synthase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:40. In some embodiments, the mutated FPP synthase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth

in SEQ ID NO:40, or a conservatively substituted amino acid sequence thereof. In some embodiments, the mutated FPP synthase isomerase polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:40.

[00526] Exemplary DXP pathway heterologous nucleic acids disclosed herein may include nucleic acids that encode a DXP pathway polypeptide, such as, a full-length DXP pathway polypeptide, a fragment of a DXP pathway polypeptide, a variant of a DXP pathway polypeptide, a truncated DXP pathway polypeptide, or a fusion polypeptide that has at least one activity of a polypeptide that is part of the DXP pathway.

[00527] In some embodiments, the DXP pathway polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the DXP pathway polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the DXP pathway polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of a DXP pathway polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of a DXP pathway polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of a DXP pathway polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a DXP pathway polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of a DXP pathway polypeptide-encoding heterologous nucleic acid.

Genetically Modified Host Cells to Generate Cannabinoids, Cannabinoid Derivatives, Cannabinoid Precursors, or Cannabinoid Precursor Derivatives

[00528] The disclosure provides for genetically modified host cells for producing cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives. For producing cannabinoids, cannabinoid derivatives, cannabinoid precursors,

or cannabinoid precursor derivatives, genetically modified host cells disclosed herein may be genetically modified to express or overexpress one or more heterologous nucleic acids disclosed herein that encode one or more polypeptides disclosed herein. In some embodiments, the genetically modified host cell of the disclosure produces a cannabinoid or a cannabinoid derivative. The disclosure also provides genetically modified host cells genetically modified to express or overexpress one or more heterologous nucleic acids disclosed herein that encode one or more polypeptides disclosed herein.

[00529] In some embodiments, to produce cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives, expression or overexpression of one or more heterologous nucleic acids disclosed herein that encode one or more polypeptides disclosed herein in a genetically modified host cell may be done in combination with expression or overexpression by the genetically modified host cell of one or more other heterologous nucleic acids disclosed herein that encode one or more polypeptides disclosed herein. In certain such embodiments, the genetically modified host cell produces a cannabinoid or a cannabinoid derivative.

Exemplary Genetically Modified Host Cells Expressing a GOT Polypeptide, Wherein Said GOT Polypeptide Can Catalyze Production of Cannabigerolic Acid from Geranyl Pyrophosphate and Olivetolic Acid in an Amount at Least Ten Times Higher than a Polypeptide Comprising an Amino Acid Sequence Set Forth in SEQ ID NO:82

[00530] Some embodiments of the disclosure relate to a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00531] The disclosure also provides genetically modified host cells genetically modified to express or overexpress one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00532] In some embodiments, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in

an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, further comprises one or more heterologous nucleic acids encoding a TKS polypeptide and one or more heterologous nucleic acids encoding an OAC polypeptide. In certain such embodiments, the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76, and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00533] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, further comprises one or more of the following: a) one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; b) one or more heterologous nucleic acids encoding a polypeptide that generates GPP; or c) one or more heterologous nucleic acids encoding a polypeptide that generates malonyl-CoA. In certain such embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is an acyl-activating enzyme (AAE) polypeptide. In certain such embodiments, the AAE polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90. In some embodiments, the AAE polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:92 or SEQ ID NO:149. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is a fatty acyl-CoA ligase polypeptide. In certain such embodiments, the fatty acyl-CoA ligase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:145 or SEQ ID NO:147. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is a fatty acyl-CoA synthetase (FAA) polypeptide. In certain such embodiments, the FAA polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide. In certain such embodiments, the GPPS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60. In some embodiments, the polypeptide that generates malonyl-CoA is an ACC polypeptide. In

certain such embodiments, the ACC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207.

[00534] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, further comprises one or more of the following: a) one or more heterologous nucleic acids encoding a HMGS polypeptide; b) one or more heterologous nucleic acids encoding a 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR) polypeptide; c) one or more heterologous nucleic acids encoding a MK polypeptide; d) a PMK polypeptide; e) one or more heterologous nucleic acids encoding a MVD polypeptide; or f) one or more heterologous nucleic acids encoding an IDI polypeptide. In certain such embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an IDI polypeptide. In certain such embodiments, the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGR polypeptide. In certain such embodiments, the HMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:22. In some embodiments, the HMGR polypeptide is a truncated HMGR (tHMGR) polypeptide. In certain such embodiments, the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGS polypeptide. In certain such embodiments, the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an MK polypeptide. In certain such embodiments, the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a PMK polypeptide. In certain such embodiments, the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a MVD polypeptide. In certain such embodiments, the

MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66.

[00535] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, further comprises one or more heterologous nucleic acids encoding a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA. In certain such embodiments, the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide. In certain such embodiments, the acetoacetyl-CoA thiolase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25.

[00536] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, further comprises one or more heterologous nucleic acids encoding a PDC polypeptide. In certain such embodiments, the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117.

[00537] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, further comprises one or more heterologous nucleic acids encoding a cannabinoid synthase polypeptide. In certain such embodiments, the cannabinoid synthase polypeptide is a THCA synthase polypeptide. In certain such embodiments, the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155. In some embodiments, the cannabinoid synthase polypeptide is a CBDA synthase polypeptide. In certain such embodiments, the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

Exemplary Genetically Modified Host Cells Expressing a Polypeptide Comprising an Amino Acid Sequence Having Sequence Identity to SEQ ID NO:110

[00538] Some embodiments of the disclosure relate to a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof.

[00539] The disclosure also provides genetically modified host cells genetically modified to express or overexpress one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof.

[00540] In some embodiments, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid

sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110 further comprises one or more heterologous nucleic acids encoding a TKS polypeptide and one or more heterologous nucleic acids encoding an OAC polypeptide. In certain such embodiments, the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76, and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00541] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110 further comprises one or more of the following: a) one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; b) one or more heterologous nucleic acids encoding a polypeptide that generates GPP; or c) one or more heterologous nucleic acids encoding a polypeptide that generates malonyl-CoA. In certain such embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is an acyl-activating enzyme (AAE) polypeptide. In certain such embodiments, the AAE polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90. In some embodiments, the AAE polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:92 or SEQ ID NO:149. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is a fatty acyl-CoA ligase polypeptide. In certain such embodiments, the fatty acyl-CoA ligase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:145 or SEQ ID NO:147. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is a fatty acyl-CoA synthetase (FAA) polypeptide. In certain such embodiments, the FAA polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide. In certain such embodiments, the GPPS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID

NO:8, or SEQ ID NO:60. In some embodiments, the polypeptide that generates malonyl-CoA is an ACC polypeptide. In certain such embodiments, the ACC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207.

[00542] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110 further comprises one or more of the following: a) one or more heterologous nucleic acids encoding a HMGS polypeptide; b) one or more heterologous nucleic acids encoding a 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR) polypeptide; c) one or more heterologous nucleic acids encoding a MK polypeptide; d) one or more heterologous nucleic acids encoding a PMK polypeptide; e) one or more heterologous nucleic acids encoding a MVD polypeptide; or f) one or more heterologous nucleic acids encoding an IDI polypeptide. In certain such embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an IDI polypeptide. In certain such embodiments, the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGR polypeptide. In certain such embodiments, the HMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:22. In some embodiments, the HMGR polypeptide is a truncated HMGR (tHMGR) polypeptide. In certain such embodiments, the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGS polypeptide. In certain such embodiments, the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an MK polypeptide. In certain such embodiments, the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a PMK polypeptide. In certain such embodiments, the PMK polypeptide comprises an amino

acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a MVD polypeptide. In certain such embodiments, the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66.

[00543] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110 further comprises one or more heterologous nucleic acids encoding a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA. In certain such embodiments, the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide. In certain such embodiments, the acetoacetyl-CoA thiolase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25.

[00544] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110 further comprises one or more heterologous nucleic acids encoding a PDC polypeptide. In certain such embodiments, the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117.

[00545] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110 further comprises one or more heterologous nucleic acids encoding a cannabinoid synthase polypeptide. In certain such embodiments, the cannabinoid synthase polypeptide is a THCA synthase polypeptide. In certain such embodiments, the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155. In some

embodiments, the cannabinoid synthase polypeptide is a CBDA synthase polypeptide. In certain such embodiments, the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

Exemplary Genetically Modified Host Cells Expressing a Polypeptide Comprising an Amino Acid Sequence Having Sequence Identity to SEQ ID NO:100

[00546] The disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:100. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof.

[00547] The disclosure also provides genetically modified host cells genetically modified to express or overexpress one or more heterologous nucleic acids encoding GOT a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:100. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino

acid sequence set forth in SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof.

[00548] In some embodiments, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 further comprises one or more heterologous nucleic acids encoding a TKS polypeptide and one or more heterologous nucleic acids encoding an OAC polypeptide. In certain such embodiments, the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76, and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00549] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 further comprises one or more of the following: a) one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; b) one or more heterologous nucleic acids encoding a polypeptide that generates GPP; or c) one or more heterologous nucleic acids encoding a polypeptide that generates malonyl-CoA. In certain such embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is an acyl-activating enzyme (AAE) polypeptide. In certain such embodiments, the AAE polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90. In some embodiments, the AAE polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:92 or SEQ ID NO:149. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is a fatty acyl-CoA ligase polypeptide. In certain such embodiments, the fatty acyl-CoA ligase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:145 or SEQ ID NO:147. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is a fatty acyl-CoA synthetase (FAA) polypeptide. In certain such embodiments, the FAA polypeptide comprises an amino acid sequence having at least 50%

sequence identity to SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide. In certain such embodiments, the GPPS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60. In some embodiments, the polypeptide that generates malonyl-CoA is an ACC polypeptide. In certain such embodiments, the ACC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207.

[00550] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 further comprises one or more of the following: a) one or more heterologous nucleic acids encoding a HMGS polypeptide; b) one or more heterologous nucleic acids encoding a 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR) polypeptide; c) one or more heterologous nucleic acids encoding a MK polypeptide; d) one or more heterologous nucleic acids encoding a PMK polypeptide; e) one or more heterologous nucleic acids encoding a MVD polypeptide; or f) one or more heterologous nucleic acids encoding an IDI polypeptide. In certain such embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an IDI polypeptide. In certain such embodiments, the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGR polypeptide. In certain such embodiments, the HMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:22. In some embodiments, the HMGR polypeptide is a truncated HMGR (tHMGR) polypeptide. In certain such embodiments, the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGS polypeptide. In certain such embodiments, the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an MK polypeptide.

In certain such embodiments, the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a PMK polypeptide. In certain such embodiments, the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a MVD polypeptide. In certain such embodiments, the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66.

[00551] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 further comprises one or more heterologous nucleic acids encoding a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA. In certain such embodiments, the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide. In certain such embodiments, the acetoacetyl-CoA thiolase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25.

[00552] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 further comprises one or more heterologous nucleic acids encoding a PDC polypeptide. In certain such embodiments, the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117.

[00553] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 further comprises one or more heterologous nucleic acids encoding a cannabinoid synthase

polypeptide. In certain such embodiments, the cannabinoid synthase polypeptide is a THCA synthase polypeptide. In certain such embodiments, the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155. In some embodiments, the cannabinoid synthase polypeptide is a CBDA synthase polypeptide. In certain such embodiments, the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

Exemplary Genetically Modified Host Cells Expressing GOT Polypeptides

[00554] The present disclosure provides a genetically modified host cell that produces a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide and b) one or more heterologous nucleic acids that encode a THCA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155.

[00555] The present disclosure provides a genetically modified host cell that produces a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide and b) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some

embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[00556] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide and b) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00557] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide, b) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide, and c) one or more

heterologous nucleic acids that encode a THCA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; and the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155.

[00558] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide, b) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide, and c) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the TKS polypeptide comprises an

amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; and the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[00559] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide and b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60.

[00560] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide, b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP, and c) one or more heterologous nucleic acids that encode a THCA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%,

99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; and the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155.

[00561] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide, b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP, and c) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; and the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[00562] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that

generates an acyl-CoA compound or an acyl-CoA compound derivative; and c) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00563] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; c) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; and d) one or more heterologous nucleic acids that encode a THCA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%,

99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; and the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155.

[00564] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; c) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; and d) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA

compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; and the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[00565] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; c) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; and d) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; and the polypeptide that generates malonyl-CoA is an

ACC polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207.

[00566] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; c) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; and e) one or more heterologous nucleic acids that encode a THCA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; the polypeptide that generates malonyl-CoA is an ACC polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207; and the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155.

[00567] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; c) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; and e) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; the polypeptide that generates malonyl-CoA is an ACC polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207; and the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[00568] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT

polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; c) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; and d) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00569] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; c) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; and e) one or more heterologous nucleic acids that encode a THCA synthase polypeptide. In certain such embodiments, the GOT

polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; and the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155.

[00570] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; c) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; and e) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an

amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; and the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[00571] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; c) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; and e) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide

comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; and the polypeptide that generates malonyl-CoA is an ACC polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207.

[00572] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; c) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; e) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; and f) one or more heterologous nucleic acids that encode a THCA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%,

99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; the polypeptide that generates malonyl-CoA is an ACC polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207; the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; and the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155.

[00573] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; c) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; e) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; and f) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%,

99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; the polypeptide that generates malonyl-CoA is an ACC polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207; the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; and the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[00574] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; and e) one or more heterologous nucleic acids that encode a PDC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%,

99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; and the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117.

[00575] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a THCA synthase polypeptide; d) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; and f) one or more heterologous nucleic acids that encode a PDC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65%

(e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; and the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155.

[00576] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide; d) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; and f) one or more heterologous nucleic acids that encode a PDC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65%

(e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; and the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[00577] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; e) one or more heterologous nucleic acids that encode a PDC polypeptide; and f) one or more heterologous nucleic acids that encode a polypeptide that

generates an acyl-CoA compound or an acyl-CoA compound derivative. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; and the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117.

[00578] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT

polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a THCA synthase polypeptide; d) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; f) one or more heterologous nucleic acids that encode a PDC polypeptide; and g) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50%

sequence identity to SEQ ID NO:117; the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; and the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155.

[00579] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide; d) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; f) one or more heterologous nucleic acids that encode a PDC polypeptide; and g) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence

identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; and the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[00580] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; e) one or more heterologous nucleic acids that encode a PDC polypeptide; f) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; and g) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%,

85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00581] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a THCA synthase

polypeptide; d) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; f) one or more heterologous nucleic acids that encode a PDC polypeptide; g) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; and h) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID

NO:117; the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00582] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide; d) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; f) one or more heterologous nucleic acids that encode a PDC polypeptide; g) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; and h) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In

some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00583] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; e) one or more heterologous nucleic acids that encode a PDC polypeptide; and f) one or more heterologous nucleic acids that encode a TKS polypeptide

and one or more heterologous nucleic acids that encode an OAC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00584] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that

generates GPP; c) one or more heterologous nucleic acids that encode a THCA synthase polypeptide; d) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; f) a PDC polypeptide; and g) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; the THCA synthase polypeptide comprises an amino acid sequence having at least

50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00585] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide; d) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; f) one or more heterologous nucleic acids that encode a PDC polypeptide; and g) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity

to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00586] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; e) one or more heterologous nucleic acids that encode a PDC polypeptide; f) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; g) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; and h) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino

acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; the polypeptide that generates malonyl-CoA is an ACC polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00587] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a THCA synthase polypeptide; d) one or more heterologous nucleic acids that encode one or more of the

following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; f) one or more heterologous nucleic acids that encode a PDC polypeptide; g) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; h) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; and i) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID

NO:117; the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155; the polypeptide that generates malonyl-CoA is an ACC polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00588] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide; d) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; f) one or more heterologous nucleic acids that encode a PDC polypeptide; g) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; h) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; and i) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino

acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151; the polypeptide that generates malonyl-CoA is an ACC polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00589] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode one or more of the

following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; e) one or more heterologous nucleic acids that encode a PDC polypeptide; f) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; and g) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; the polypeptide that generates malonyl-CoA is an ACC polypeptide and comprises an amino acid sequence having at least 50% sequence

identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00590] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a THCA synthase polypeptide; d) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; f) one or more heterologous nucleic acids that encode a PDC polypeptide; g) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; and h) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity

to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155; the polypeptide that generates malonyl-CoA is an ACC polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00591] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide; d) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; f) one or more heterologous nucleic acids that encode a PDC polypeptide; g) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; and h) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%,

94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151; the polypeptide that generates malonyl-CoA is an ACC polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00592] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA;

d) one or more heterologous nucleic acids that encode one or more polypeptides that condense an acyl-CoA compound or an acyl-CoA compound derivative with malonyl-CoA to generate olivetolic acid or derivatives of olivetolic acid; e) one or more heterologous nucleic acids that encode a geranyl pyrophosphate:olivetolic acid transferase (GOT) polypeptide or an aromatic prenyltransferase polypeptide such as a NphB polypeptide; and f) one or more heterologous nucleic acids that encode a cannabinoid synthase polypeptide.

[00593] The present disclosure also provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; b) one or more heterologous nucleic acids that encode a polypeptide that generates neryl pyrophosphate (NPP); c) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; d) one or more heterologous nucleic acids that encode one or more polypeptides that condense an acyl-CoA compound or an acyl-CoA compound derivative and malonyl-CoA to generate olivetolic acid or derivatives of olivetolic acid; e) one or more heterologous nucleic acids that encode a GOT polypeptide or a NphB polypeptide; and f) one or more heterologous nucleic acids that encode a cannabinoid synthase polypeptide. In certain such embodiments, culturing of the genetically modified host cell in a suitable medium provides for synthesis of the cannabinoid or the cannabinoid derivative in a recoverable amount.

[00594] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; b) one or more heterologous nucleic acids that encode a TKS/OAC fusion polypeptide; and c) one or more heterologous nucleic acids that encode a GOT polypeptide or a NphB polypeptide. In certain such embodiments, culturing the genetically modified host cell in a medium comprising a carboxylic acid provides for synthesis of a cannabinoid derivative or cannabinoid in a recoverable amount. In some embodiments, the genetically modified host cell is further genetically modified with one or more heterologous nucleic acids that encode a THCAS or CBDAS polypeptide. In certain such embodiments, culturing the genetically modified host cell in a medium comprising a carboxylic acid provides for synthesis of a cannabinoid derivative or a cannabinoid in a recoverable amount.

Exemplary Genetically Modified Host Cells Expressing NPPS Polypeptides

[00595] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that condenses 2 molecules of acetyl-CoA to generate acetoacetyl-CoA; b) one or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; and c) one or more heterologous nucleic acids encoding a NPPS polypeptide.

[00596] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that condenses 2 molecules of acetyl-CoA to generate acetoacetyl-CoA; b) one or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; c) one or more heterologous nucleic acids encoding a NPPS polypeptide; and d) one or more heterologous nucleic acids encoding a TKS polypeptide and one or more heterologous nucleic acids encoding an OAC polypeptide.

[00597] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that condenses 2 molecules of acetyl-CoA to generate acetoacetyl-CoA; b) one or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; c) one or more heterologous nucleic acids encoding a NPPS polypeptide; d) one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; and e) one or more heterologous nucleic acids encoding a TKS polypeptide and one or more heterologous nucleic acids encoding an OAC polypeptide.

[00598] A GOT polypeptide, using NPP and olivetolic acid as substrates, can generate cannabinerolic acid (CBNRA). In some embodiments, a GOT polypeptide acts on NPP and an olivetolic acid derivative (as described elsewhere herein) to generate a CBNRA derivative. Cannabinerolic acid or derivatives thereof can serve as a substrate for a CBDAS or THCAS polypeptide to generate CBDA or THCA, or derivatives thereof, respectively.

[00599] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that condenses one molecule of acetyl-CoA and one molecule of malonyl-CoA to generate acetoacetyl-CoA; b) one or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; c) one or more heterologous nucleic acids encoding a NPPS polypeptide; and d) one or more heterologous nucleic acids encoding a GOT polypeptide. In certain such embodiments, the GOT polypeptide encoded by one or more

heterologous nucleic acids can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:110.

[00600] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that condenses 2 molecules of acetyl-CoA to generate acetoacetyl-CoA; b) one or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; c) one or more heterologous nucleic acids encoding a NPPS polypeptide; and d) one or more heterologous nucleic acids encoding a GOT polypeptide or a NphB polypeptide. In certain such embodiments, the GOT polypeptide encoded by one or more heterologous nucleic acids can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:110.

[00601] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that condenses 2 molecules of acetyl-CoA to generate acetoacetyl-CoA; b) one

or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; c) one or more heterologous nucleic acids encoding a NPPS polypeptide; d) one or more heterologous nucleic acids encoding a GOT polypeptide or a NphB polypeptide; and e) one or more heterologous nucleic acids encoding a CBDAS or THCAS polypeptide. In certain such embodiments, the GOT polypeptide encoded by one or more heterologous nucleic acids can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:110.

[00602] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that condenses one molecule of acetyl-CoA and one molecule of malonyl-CoA to generate acetoacetyl-CoA; b) one or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; c) one or more heterologous nucleic acids encoding a NPPS polypeptide; d) one or more heterologous nucleic acids encoding a TKS polypeptide and one or more heterologous nucleic acids encoding an OAC polypeptide; and e) one or more heterologous nucleic acids encoding a GOT polypeptide. In certain such embodiments, the GOT polypeptide encoded by one or more heterologous nucleic acids can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids

comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:110.

[00603] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that condenses 2 molecules of acetyl-CoA to generate acetoacetyl-CoA; b) one or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; c) one or more heterologous nucleic acids encoding a NPPS polypeptide; d) one or more heterologous nucleic acids encoding a TKS polypeptide and one or more heterologous nucleic acids encoding an OAC polypeptide; and e) one or more heterologous nucleic acids encoding a GOT polypeptide or a NphB polypeptide. In certain such embodiments, the GOT polypeptide encoded by one or more heterologous nucleic acids can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:110.

[00604] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that condenses 2 molecules of acetyl-CoA to generate acetoacetyl-CoA; b) one or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; c) one or more heterologous nucleic acids encoding a NPPS polypeptide; d) one or more heterologous nucleic acids encoding a TKS polypeptide and one or more heterologous nucleic acids encoding an OAC polypeptide; e) one or more heterologous nucleic acids encoding a GOT polypeptide or a NphB polypeptide; and f) one or more heterologous nucleic acids encoding a CBDAS or THCAS polypeptide. In certain such embodiments, the

GOT polypeptide encoded by one or more heterologous nucleic acids can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:110.

[00605] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that condenses one molecule of acetyl-CoA and one molecule of malonyl-CoA to generate acetoacetyl-CoA; b) one or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; c) one or more heterologous nucleic acids encoding a NPPS polypeptide; d) one or more heterologous nucleic acids encoding a TKS polypeptide and one or more heterologous nucleic acids encoding an OAC polypeptide; e) one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; and f) one or more heterologous nucleic acids encoding a GOT polypeptide. In certain such embodiments, the GOT polypeptide encoded by one or more heterologous nucleic acids can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%,

89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:110.

[00606] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified: a) with one or more heterologous nucleic acids encoding a polypeptide that condenses 2 molecules of acetyl-CoA to generate acetoacetyl-CoA; b) with one or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; c) with one or more heterologous nucleic acids encoding a NPPS polypeptide; d) with one or more heterologous nucleic acids encoding a TKS polypeptide and with one or more heterologous nucleic acids encoding an OAC polypeptide; e) with one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; and f) with one or more heterologous nucleic acids encoding a GOT polypeptide or a NphB polypeptide. In certain such embodiments, the GOT polypeptide encoded by one or more heterologous nucleic acids can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:110.

[00607] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that condenses 2 molecules of acetyl-CoA to generate acetoacetyl-CoA; b) one or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; c) one or more heterologous nucleic acids encoding a NPPS polypeptide; d) one or more heterologous nucleic acids encoding a TKS polypeptide and one or more heterologous nucleic acids encoding an OAC polypeptide; e) one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; f) one or more heterologous nucleic acids encoding a GOT polypeptide or a

NphB polypeptide; and g) one or more heterologous nucleic acids encoding a CBDAS or THCAS polypeptide. In certain such embodiments, the GOT polypeptide encoded by one or more heterologous nucleic acids can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:110.

Exemplary Genetically Modified Host Cells for Making Olivetolic Acid or Olivetolic Acid Derivatives

[00608] The present disclosure provides a genetically modified host cell for producing olivetolic acid or an olivetolic acid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative and b) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids encoding an OAC polypeptide. In certain such embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78. In certain such embodiments, culturing the genetically modified host cell disclosed herein in a medium comprising a carboxylic acid provides for synthesis of an olivetolic acid or olivetolic acid derivative in a recoverable amount.

[00609] The present disclosure provides a genetically modified host cell for producing olivetolic acid or an olivetolic acid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; and b) one or more heterologous nucleic acids that encode a TKS/OAC fusion polypeptide. In certain such embodiments, culturing the genetically modified host cell in a medium comprising a carboxylic acid provides for synthesis of the olivetolic acid or olivetolic acid derivative in a recoverable amount.

Suitable Host Cells

[00610] Parent host cells that are suitable for use in generating a genetically modified host cell of the present disclosure may include prokaryotic cells and eukaryotic cells. In some embodiments, the eukaryotic cells are yeast cells. In some embodiments, the eukaryotic cells are plant cells.

[00611] Host cells (including parent host cells and genetically modified host cells) are in some embodiments unicellular organisms, or are grown in culture as single cells. In some embodiments, the host cell is a eukaryotic cell. Suitable eukaryotic host cells may include, but are not limited to, yeast cells, insect cells, plant cells, fungal cells, and algal cells. Suitable eukaryotic host cells may include, but are not limited to, *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia* sp., *Saccharomyces cerevisiae*, *Saccharomyces* sp., *Hansenula polymorpha* (now known as *Pichia angusta*), *Kluyveromyces* sp., *Kluyveromyces lactis, Kluyveromyces marxianus*, *Schizosaccharomyces pombe*, *Dekkera bruxellensis*, *Arxula adeninivorans*, *Candida albicans*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium* sp., *Fusarium gramineum*, *Fusarium venenatum*, *Neurospora crassa*, *Chlamydomonas reinhardtii*, and the like. In some embodiments, the host cell is a protease-deficient strain of *Saccharomyces cerevisiae*. In some embodiments, the host cell is a eukaryotic cell other than a plant cell. In some embodiments, the eukaryotic cell is a plant cell. In some embodiments, the eukaryotic cell is a plant cell, where the plant cell is one that does not normally produce a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative. In some embodiments, the host cell is *Saccharomyces*

cerevisiae. In some embodiments, the genetically modified host cell disclosed herein is cultured *in vitro*.

[00612] In some embodiments, the host cell is a prokaryotic cell. Suitable prokaryotic cells may include, but are not limited to, any of a variety of laboratory strains of *Escherichia coli*, *Lactobacillus* sp., *Salmonella* sp., *Shigella* sp., and the like. See, e.g., Carrier et al. (1992) *J. Immunol.* 148:1176-1181; U.S. Patent No. 6,447,784; and Sizemore et al. (1995) *Science* 270:299-302. Examples of Salmonella strains which can be employed may include, but are not limited to, *Salmonella typhi* and *S. typhimurium*. Suitable Shigella strains may include, but are not limited to, *Shigella flexneri*, *Shigella sonnei*, and *Shigella disenteriae*. Typically, the laboratory strain is one that is non-pathogenic. Non-limiting examples of other suitable bacteria may include, but are not limited to, *Bacillus subtilis*, *Pseudomonas putida*, *Pseudomonas aeruginosa*, *Pseudomonas mevalonii*, *Rhodobacter sphaeroides*, *Rhodobacter capsulatus*, *Rhodospirillum rubrum*, *Rhodococcus* sp., and the like. In some embodiments, the host cell is *Escherichia coli*.

Genetic Modification of Host Cells

[00613] The present disclosure provides for a method of making a genetically modified host cell for producing a cannabinoid, a cannabinoid derivative, a cannabinoid precursor derivative, or a cannabinoid precursor, comprising introducing into the genetically modified host cell one or more heterologous nucleic acids disclosed herein. In some embodiments, the genetically modified host cell produces a cannabinoid or a cannabinoid derivative. The disclosure also provides a method for making a genetically modified host cell genetically modified to express or overexpress one or more heterologous nucleic acids disclosed herein that encode one or more polypeptides disclosed herein, comprising introducing into the genetically modified host cell one or more heterologous nucleic acids disclosed herein.

[00614] In some embodiments, the disclosure provides for a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, into the genetically modified host cell. In some embodiments, the present disclosure provides for a method of making a genetically modified host cell genetically modified to express or overexpress one or more

heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, comprising introducing one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, into the genetically modified host cell.

[00615] In some embodiments, the disclosure provides for a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110 into the genetically modified host cell. In some embodiments, the present disclosure provides for a method of making a genetically modified host cell genetically modified to express or overexpress one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110 comprising introducing one or more heterologous nucleic acids encoding the GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110 into the genetically modified host cell.

[00616] The disclosure provides for a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 into the genetically modified host cell. In some embodiments, the present disclosure provides for a method of making a genetically modified host cell genetically modified to express or

overexpress one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 comprising introducing one or more heterologous nucleic acids encoding the GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 into the genetically modified host cell.

[00617] To genetically modify a parent host cell to produce a genetically modified host cell of the present disclosure, one or more heterologous nucleic acids disclosed herein is introduced stably or transiently into a host cell, using established techniques. Such techniques may include, but are not limited to, electroporation, calcium phosphate precipitation, DEAE-dextran mediated transfection, liposome-mediated transfection, and the like. For stable transformation, a heterologous nucleic acid will generally further include a selectable marker, e.g., any of several well-known selectable markers such as neomycin resistance, ampicillin resistance, tetracycline resistance, chloramphenicol resistance, kanamycin resistance, and the like. In some embodiments, a parent host cell is genetically modified to produce a genetically modified host cell of the present disclosure using a CRISPR/Cas9 system to genetically modify a parent host cell with one or more heterologous nucleic acids disclosed herein.

[00618] One or more nucleic acids disclosed herein can be present in an expression vector or construct. Suitable expression vectors may include, but are not limited to, baculovirus vectors, bacteriophage vectors, plasmids, phagemids, cosmids, fosmids, bacterial artificial chromosomes, viral vectors (e.g. viral vectors based on vaccinia virus, poliovirus, adenovirus, adeno-associated virus, SV40, herpes simplex virus, and the like), P1-based artificial chromosomes, yeast plasmids, yeast artificial chromosomes, and any other vectors specific for specific hosts of interest (such as *E. coli* and yeast). Thus, for example, one or more nucleic acids encoding a mevalonate pathway gene product(s) is included in any one of a variety of expression vectors for expressing the mevalonate pathway gene product(s). Such vectors may include chromosomal, non-chromosomal, and synthetic DNA sequences.

[00619] Numerous additional suitable expression vectors are known to those of skill in the art, and many are commercially available. The following vectors are provided by way

of example; for bacterial host cells: pQE vectors (Qiagen), pBluescript plasmids, pNH vectors, lambda-ZAP vectors (Stratagene); pTrc99a, pKK223-3, pDR540, and pRIT2T (Pharmacia); for eukaryotic host cells: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG, and pSVLSV40 (Pharmacia). However, any other plasmid or other vector may be used so long as it is compatible with the host cell.

[00620] In some embodiments, one or more of the nucleic acids disclosed herein are present in a single expression vector. In some embodiments, two or more of the nucleic acids disclosed herein are present in a single expression vector. In some embodiments, three or more of the nucleic acids disclosed herein are present in a single expression vector. In some embodiments, four or more of the nucleic acids disclosed herein are present in a single expression vector. In some embodiments, five or more of the nucleic acids disclosed herein are present in a single expression vector. In some embodiments, six or more of the nucleic acids disclosed herein are present in a single expression vector. In some embodiments, seven or more of the nucleic acids disclosed herein are present in a single expression vector.

[00621] In some embodiments, two or more nucleic acids disclosed herein are in separate expression vectors. In some embodiments, three or more nucleic acids disclosed herein are in separate expression vectors. In some embodiments, four or more nucleic acids disclosed herein are in separate expression vectors. In some embodiments, five or more nucleic acids disclosed herein are in separate expression vectors. In some embodiments, six or more nucleic acids disclosed herein are in separate expression vectors. In some embodiments, seven or more nucleic acids disclosed herein are in separate expression vectors. In some embodiments, eight or more nucleic acids disclosed herein are in separate expression vectors. In some embodiments, nine or more nucleic acids disclosed herein are in separate expression vectors. In some embodiments, ten or more nucleic acids disclosed herein are in separate expression vectors.

[00622] In some embodiments, one or more of the nucleic acids disclosed herein are present in a single expression construct. In some embodiments, two or more of the nucleic acids disclosed herein are present in a single expression construct. In some embodiments, three or more of the nucleic acids disclosed herein are present in a single expression construct. In some embodiments, four or more of the nucleic acids disclosed herein are present in a single expression construct. In some embodiments, five or more of the nucleic acids disclosed herein are present in a single expression construct. In some embodiments, six or more of the nucleic acids disclosed herein are present in a single expression construct.

In some embodiments, seven or more of the nucleic acids disclosed herein are present in a single expression construct.

[00623] In some embodiments, two or more nucleic acids disclosed herein are in separate expression constructs. In some embodiments, three or more nucleic acids disclosed herein are in separate expression constructs. In some embodiments, four or more nucleic acids disclosed herein are in separate expression constructs. In some embodiments, five or more nucleic acids disclosed herein are in separate expression constructs. In some embodiments, six or more nucleic acids disclosed herein are in separate expression constructs. In some embodiments, seven or more nucleic acids disclosed herein are in separate expression constructs. In some embodiments, eight or more nucleic acids disclosed herein are in separate expression constructs. In some embodiments, nine or more nucleic acids disclosed herein are in separate expression constructs. In some embodiments, ten or more nucleic acids disclosed herein are in separate expression constructs.

[00624] The disclosure provides a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing a vector comprising one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, into the genetically modified host cell. The disclosure provides a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing a vector comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110 into the genetically modified host cell. The disclosure also provides a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing a vector comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 into the genetically modified host cell.

[00625] The disclosure also provides a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing a

vector comprising a CsPT4 heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:111 into the genetically modified host cell. The disclosure also provides a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing a vector comprising a CsPT4 heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:225 into the genetically modified host cell. The disclosure also provides a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing a vector comprising a CsPT4t heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:221 into the genetically modified host cell. The disclosure also provides a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing a vector comprising a CsPT4t heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:224 into the genetically modified host cell.

[00626] The disclosure provides a method of making a genetically modified host cell genetically modified to express or overexpress one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, comprising introducing a vector comprising one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, into the genetically modified host cell. The disclosure provides a method of making a genetically modified host cell genetically modified to express or overexpress one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110, comprising introducing a vector comprising one or more heterologous nucleic acids encoding the GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110 into the genetically modified host cell. The disclosure provides a method of making a genetically

modified host cell genetically modified to express or overexpress one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100, comprising introducing a vector comprising one or more heterologous nucleic acids encoding the GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 into the genetically modified host cell.

[00627] The disclosure also provides a method of making a genetically modified host cell genetically modified to express or overexpress a CsPT4 heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:111, comprising introducing a vector comprising a CsPT4 heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:111 into the genetically modified host cell. The disclosure also provides a method of making a genetically modified host cell genetically modified to express or overexpress a CsPT4 heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:225, comprising introducing a vector comprising a CsPT4 heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:225 into the genetically modified host cell. The disclosure also provides a method of making a genetically modified host cell genetically modified to express or overexpress a CsPT4t heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:221, comprising introducing a vector comprising a CsPT4t heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:221 into the genetically modified host cell. The disclosure also provides a method of making a genetically modified host cell genetically modified to express or overexpress a CsPT4t heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:224, comprising introducing a vector comprising a CsPT4t heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:224 into the genetically modified host cell.

[00628] In some embodiments, one or more of the heterologous nucleic acids disclosed herein is present in a high copy number plasmid, e.g., a plasmid that exists in about

10-50 copies per cell, or more than 50 copies per cell. In some embodiments, one or more of the heterologous nucleic acids disclosed herein is present in a low copy number plasmid. In some embodiments, one or more of the heterologous nucleic acids disclosed herein is present in a medium copy number plasmid.

[00629] Depending on the host/vector or host/construct system utilized, any of a number of suitable transcription and translation control elements, including constitutive and inducible promoters, transcription enhancer elements, transcription terminators, etc. may be used in the expression vector or construct (see e.g., Bitter et al. (1987) *Methods in Enzymology*, 153:516-544).

[00630] In some embodiments, heterologous nucleic acids disclosed herein are operably linked to a promoter. In some embodiments, the promoter is a constitutive promoter. In some embodiments, the promoter is an inducible promoter. In some embodiments, the promoter is functional in a prokaryotic cell. In some embodiments, the promoter is functional in a eukaryotic cell. In some embodiments, the promoter can be a strong driver of expression. In some embodiments, the promoter can be a weak driver of expression. In some embodiments, the promoter can be a medium driver of expression.

[00631] Suitable promoters for use in prokaryotic host cells may include, but are not limited to, a bacteriophage T7 RNA polymerase promoter; a trp promoter; a lac operon promoter; a hybrid promoter, e.g., a lac/tac hybrid promoter, a tac/trc hybrid promoter, a trp/lac promoter, a T7/lac promoter; a trc promoter; a tac promoter, and the like; an araBAD promoter; *in vivo* regulated promoters, such as an *ssaG* promoter or a related promoter (*see*, e.g., U.S. Patent Publication No. 20040131637), a *pagC* promoter (Pulkkinen and Miller, *J. Bacteriol.*, 1991: 173(1): 86-93; Alpuche-Aranda et al., PNAS, 1992; 89(21): 10079-83), a *nirB* promoter (Harborne et al. (1992) *Mol. Micro.* 6:2805-2813), and the like (*see*, e.g., Dunstan et al. (1999) *Infect. Immun.* 67:5133-5141; McKelvie et al. (2004) *Vaccine* 22:3243-3255; and Chatfield et al. (1992) *Biotechnol.* 10:888-892); a sigma70 promoter, e.g., a consensus sigma70 promoter (see, e.g., GenBank Accession Nos. AX798980, AX798961, and AX798183); a stationary phase promoter, e.g., a *dps* promoter, a *spv* promoter, and the like; a promoter derived from the pathogenicity island SPI-2 (*see*, e.g., WO96/17951); an *actA* promoter (*see*, e.g., Shetron-Rama et al. (2002) *Infect. Immun.* 70:1087-1096); an *rpsM* promoter (*see*, e.g., Valdivia and Falkow (1996). *Mol. Microbiol.* 22:367–378); a tet promoter (*see*, e.g., Hillen, W. and Wissmann, A. (1989) In Saenger, W. and Heinemann, U. (eds), *Topics in Molecular and Structural Biology, Protein–Nucleic Acid*

Interaction. Macmillan, London, UK, Vol. 10, pp. 143–162); an SP6 promoter (*see*, e.g., Melton et al. (1984) *Nucl. Acids Res.* 12:7035); and the like.

[00632] Suitable constitutive promoters for use in prokaryotic cells are known in the art and may also include, but are not limited to, a sigma70 promoter, e.g., a consensus sigma70 promoter.

[00633] Non-limiting examples of suitable eukaryotic promoters may include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector, construct, and promoter is well within the level of ordinary skill in the art. The expression vector or construct may also contain a ribosome binding site for translation initiation and a transcription terminator. The expression vector or construct may also include appropriate sequences for amplifying expression.

[00634] In yeast, a number of vectors or constructs containing constitutive or inducible promoters may be used. For a review see, Current Protocols in Molecular Biology, Vol. 2, 1988, Ed. Ausubel, et al., Greene Publish. Assoc. & Wiley Interscience, Ch. 13; Grant, et al., 1987, Expression and Secretion Vectors for Yeast, in Methods in Enzymology, Eds. Wu & Grossman, 31987, Acad. Press, N.Y., Vol. 153, pp.516-544; Glover, 1986, DNA Cloning, Vol. II, IRL Press, Wash., D.C., Ch. 3; and Bitter, 1987, Heterologous Gene Expression in Yeast, Methods in Enzymology, Eds. Berger & Kimmel, Acad. Press, N.Y., Vol. 152, pp. 673-684; and The Molecular Biology of the Yeast Saccharomyces, 1982, Eds. Strathern et al., Cold Spring Harbor Press, Vols. I and II. A constitutive yeast promoter such as ADH or LEU2 or an inducible promoter such as GAL may be used (Cloning in Yeast, Ch. 3, R. Rothstein In: DNA Cloning Vol. 11, A Practical Approach, Ed. DM Glover, 1986, IRL Press, Wash., D.C.). Alternatively, vectors may be used which promote integration of foreign DNA sequences into the yeast chromosome.

[00635] Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of *E. coli*, the *S. cerevisiae* TRP1 gene, etc.; and a promoter derived from a highly-expressed gene to direct transcription of the coding sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), α-factor, acid phosphatase, or heat shock proteins, among others.

[00636] Inducible promoters are well known in the art. Suitable inducible promoters may include, but are not limited to, the pL of bacteriophage λ; Plac; Ptrp; Ptac (Ptrp-lac hybrid promoter); an isopropyl-beta-D-thiogalactopyranoside (IPTG)-inducible promoter,

e.g., a *lacZ* promoter; a tetracycline-inducible promoter; an arabinose inducible promoter, e.g., P_{BAD} (see, e.g., Guzman et al. (1995) *J. Bacteriol.* 177:4121-4130); a xylose-inducible promoter, e.g., Pxyl (see, e.g., Kim et al. (1996) *Gene* 181:71-76); a GAL1 promoter; a tryptophan promoter; a lac promoter; an alcohol-inducible promoter, e.g., a methanol-inducible promoter, an ethanol-inducible promoter; a raffinose-inducible promoter; a heat-inducible promoter, e.g., heat inducible lambda P_L promoter, a promoter controlled by a heat-sensitive repressor (e.g., CI857-repressed lambda-based expression vectors; *see*, e.g., Hoffmann et al. (1999) *FEMS Microbiol Lett.* 177(2):327-34); and the like.

[00637] In addition, the expression vectors or constructs will in many embodiments contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in prokaryotic host cells such as *E. coli*.

[00638] In some embodiments, one or more heterologous nucleic acids disclosed herein is integrated into the genome of the genetically modified host cell disclosed herein. In some embodiments, one or more heterologous nucleic acids disclosed herein remains episomal (i.e., is not integrated into the genome of the genetically modified host cell). In some embodiments, at least one of the one or more heterologous nucleic acids disclosed herein is maintained extrachromosomally.

[00639] In some embodiments, a subject heterologous nucleic acid or a subject recombinant expression vector or construct comprises a promoter or other regulatory element(s) for expression in a plant cell. Non-limiting examples of suitable constitutive promoters that are functional in a plant cell is the cauliflower mosaic virus 35S promoter, a tandem 35S promoter (Kay et al., *Science* 236:1299 (1987)), a cauliflower mosaic virus 19S promoter, a nopaline synthase gene promoter (Singer et al., *Plant Mol. Biol.* 14:433 (1990); An, *Plant Physiol.* 81:86 (1986), an octopine synthase gene promoter, and a ubiquitin promoter. Suitable inducible promoters that are functional in a plant cell may include, but are not limited to, a phenylalanine ammonia-lyase gene promoter, a chalcone synthase gene promoter, a pathogenesis-related protein gene promoter, a copper-inducible regulatory element (Mett et al., *Proc. Natl. Acad. Sci. USA* 90:4567-4571 (1993); Furst et al., Cell 55:705-717 (1988)); tetracycline and chlor-tetracycline-inducible regulatory elements (Gatz et al., *Plant J.* 2:397-404 (1992); Röder et al., *Mol. Gen. Genet.* 243:32-38 (1994); Gatz, *Meth. Cell Biol.* 50:411-424 (1995)); ecdysone inducible regulatory elements (Christopherson et al., *Proc. Natl. Acad. Sci. USA* 89:6314-6318 (1992); Kreutzweiser et al.,

Ecotoxicol. Environ. Safety 28:14-24 (1994)); heat shock inducible regulatory elements (Takahashi et al., *Plant Physiol.* 99:383-390 (1992); Yabe et al., *Plant Cell Physiol.* 35:1207-1219 (1994); Ueda et al., *Mol. Gen. Genet.* 250:533-539 (1996)); and lac operon elements, which are used in combination with a constitutively expressed lac repressor to confer, for example, IPTG-inducible expression (Wilde et al., *EMBO J.* 11:1251-1259 (1992); a nitrate-inducible promoter derived from the spinach nitrite reductase gene (Back et al., *Plant Mol. Biol.* 17:9 (1991)); a light-inducible promoter, such as that associated with the small subunit of RuBP carboxylase or the LHCP gene families (Feinbaum et al., *Mol. Gen. Genet.* 226:449 (1991); Lam and Chua, *Science* 248:471 (1990)); a light-responsive regulatory element as described in U.S. Patent Publication No. 20040038400; a salicylic acid inducible regulatory elements (Uknes et al., *Plant Cell* 5:159-169 (1993); Bi et al., *Plant J.* 8:235-245 (1995)); plant hormone-inducible regulatory elements (Yamaguchi-Shinozaki et al., *Plant Mol. Biol.* 15:905 (1990); Kares et al., *Plant Mol. Biol.* 15:225 (1990)); and human hormone-inducible regulatory elements such as the human glucocorticoid response element (Schena et al., *Proc. Natl. Acad. Sci. USA* 88:10421 (1991).

[00640] Plant tissue-selective regulatory elements also can be included in a subject heterologous nucleic acid or a subject vector or construct. Suitable tissue-selective regulatory elements, which can be used to ectopically express a heterologous nucleic acid in a single tissue or in a limited number of tissues, may include, but are not limited to, a xylem-selective regulatory element, a tracheid-selective regulatory element, a fiber-selective regulatory element, a trichome-selective regulatory element (see, e.g., Wang et al. (2002) *J. Exp. Botany* 53:1891-1897), a glandular trichome-selective regulatory element, and the like.

[00641] Vectors that are suitable for use in plant cells are known in the art, and any such vector can be used to introduce a subject heterologous nucleic acid into a plant host cell. Suitable vectors may include, e.g., a Ti plasmid of *Agrobacterium tumefaciens* or an Ri_1 plasmid of *A. rhizogenes*. The Ti or Ri_1 plasmid is transmitted to plant cells on infection by *Agrobacterium* and is stably integrated into the plant genome. J. Schell, *Science,* 237:1176-83 (1987). Also suitable for use is a plant artificial chromosome, as described in, e.g., U.S. Patent No. 6,900,012.

[00642] As will be appreciated by the skilled artisan, slight changes in nucleotide sequence do not necessarily alter the amino acid sequence of the encoded polypeptide. It will be appreciated by persons skilled in the art that changes in the identities of nucleotides in a specific gene sequence that change the amino acid sequence of the encoded polypeptide may result in reduced or enhanced effectiveness of the genes and that, in some applications (e.g.,

anti-sense, co-suppression, or RNAi), partial sequences often work as effectively as full length versions. The ways in which the nucleotide sequence can be varied or shortened are well known to persons skilled in the art, as are ways of testing the effectiveness of the altered genes. In certain embodiments, effectiveness may easily be tested by, for example, conventional gas chromatography. All such variations of the genes are therefore included as part of the present disclosure.

Codon Usage

[00643] As is well known to those of skill in the art, it is possible to improve the expression of a heterologous nucleic acid in a host organism by replacing the nucleotide sequences coding for a particular amino acid (i.e., a codon) with another codon which is better expressed in the host organism (i.e., codon optimization). One reason that this effect arises due to the fact that different organisms show preferences for different codons. In some embodiments, a heterologous nucleic acid disclosed herein is modified or optimized such that the nucleotide sequence reflects the codon preference for the particular host cell. For example, the nucleotide sequence will in some embodiments be modified or optimized for yeast codon preference. See, e.g., Bennetzen and Hall (1982) *J. Biol. Chem.* 257(6): 3026-3031. As another non-limiting example, the nucleotide sequence will in some embodiments be modified or optimized for *E. coli* codon preference. See, e.g., Gouy and Gautier (1982) *Nucleic Acids Res.* 10(22):7055-7074; Eyre-Walker (1996) *Mol. Biol. Evol.* 13(6):864-872. See also Nakamura et al. (2000) *Nucleic Acids Res.* 28(1):292.

[00644] Statistical methods have been generated to analyze codon usage bias in various organisms and many computer algorithms have been developed to implement these statistical analyses in the design of codon optimized gene sequences (Lithwick G, Margalit H (2003) Hierarchy of sequence-dependent features associated with prokaryotic translation. Genome Research 13: 2665-73). Other modifications in codon usage to increase protein expression that are not dependent on codon bias have also been described (Welch et al. (2009). Design parameters to control synthetic gene expression in Escherichia coli. PLoS ONE 4: e7002).

[00645] In some embodiments, the codon usage of a coding sequence is modified such that the level of translation of the encoded mRNA is decreased. Reducing the level of translation of an mRNA by modifying codon usage is achieved by modifying the sequence to include codons that are rare or not commonly used by the host cell. Codon usage tables for many organisms are available that summarize the percentage of time a specific organism

uses a specific codon to encode for an amino acid. Certain codons are used more often than other, "rare" codons. The use of "rare" codons in a sequence generally decreases its rate of translation. Thus, e.g., the coding sequence is modified by introducing one or more rare codons, which affect the rate of translation, but not the amino acid sequence of the polypeptide translated. For example, there are 6 codons that encode for arginine: CGT, CGC, CGA, CGG, AGA, and AGG. In *E. coli* the codons CGT and CGC are used far more often (encoding approximately 40% of the arginines in *E. coli* each) than the codon AGG (encoding approximately 2% of the arginines in *E. coli*). Modifying a CGT codon within the sequence of a gene to an AGG codon would not change the sequence of the polypeptide, but would likely decrease the gene's rate of translation.

[00646] Further, it will be appreciated that this disclosure embraces the degeneracy of codon usage as would be understood by one of ordinary skill in the art and illustrated in the following table.

Codon Degeneracies

Amino Acid	Codons
Ala/A	GCT, GCC, GCA, GCG
Arg/R	CGT, CGC, CGA, CGG, AGA, AGG
Asn/N	AAT, AAC
Asp/D	GAT, GAC
Cys/C	TGT, TGC
Gln/Q	CAA, CAG
Glu/E	GAA, GAG
Gly/G	GGT, GGC, GGA, GGG
His/H	CAT, CAC
Ile/I	ATT, ATC, ATA
Leu/L	TTA, TTG, CTT, CTC, CTA, CTG
Lys/K	AAA, AAG
Met/M	ATG
Phe/F	TTT, TTC
Pro/P	CCT, CCC, CCA, CCG
Ser/S	TCT, TCC, TCA, TCG, AGT, AGC
Thr/T	ACT, ACC, ACA, ACG
Trp/W	TGG
Tyr/Y	TAT, TAC
Val/V	GTT, GTC, GTA, GTG

START	ATG
STOP	TAG, TGA, TAA

Genetically Modified Plants

[00647] The present disclosure provides genetically modified plants, where the genetically modified plants are genetically modified with one or more heterologous nucleic acids disclosed herein to generate a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative. In some embodiments, the genetically modified plant is a plant of a genus other than *Cannabis*.

[00648] The present disclosure provides a genetically modified plant, wherein the genetically modified plant is genetically modified with: a) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; d) one or more heterologous nucleic acids that encode one or more polypeptides that condense an acyl-CoA compound or an acyl-CoA compound derivative and malonyl-CoA to generate olivetolic acid or derivatives of olivetolic acid; e) one or more heterologous nucleic acids that encode a polypeptide that condenses GPP and olivetolic acid to generate cannabigerolic acid or derivatives thereof; or f) one or more heterologous nucleic acids that encode a cannabinoid synthase polypeptide, wherein the polypeptides are produced in the plant, and wherein production of the polypeptides results in production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by the genetically modified plant. In some embodiments, the plant is a monocot. In some embodiments, the plant is a dicot. In some embodiments, one or more of the polypeptide-encoding heterologous nucleic acids is operably linked to a constitutive promoter. In some embodiments, one or more of the polypeptide-encoding heterologous nucleic acids is operably linked to an inducible promoter. In some embodiments, one or more of the polypeptide-encoding heterologous nucleic acids is operably linked to a tissue-specific promoter. In some embodiments, the tissue-specific promoter is a trichome-specific promoter.

[00649] The present disclosure provides a method of producing a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative, the method comprising maintaining a transgenic plant under conditions that favor production of the encoded one or more polypeptides, wherein production of the encoded one or more

polypeptides results in production of the cannabinoid, the cannabinoid derivative, the cannabinoid precursor, or the cannabinoid precursor derivative.

[00650] In some embodiments, the genome of the transgenic plant comprises a subject heterologous nucleic acid. In some embodiments, the transgenic plant is homozygous for the genetic modification. In some embodiments, the transgenic plant is heterozygous for the genetic modification.

[00651] In some embodiments, a subject transgenic plant produces one or more transgene-encoded polypeptides disclosed herein that result in the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in an amount that is at least about 50%, at least about 2-fold, at least about 5-fold, at least about 10-fold, at least about 25-fold, at least about 50-fold, or at least about 100-fold, or higher, than the amount of the cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative produced by a control plant, e.g., a non-transgenic plant (a plant that does not include the transgene encoding the one or more polypeptides) of the same species.

[00652] In some embodiments, a subject transgenic plant is a transgenic version of a control, non-transgenic plant that normally produces a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative that is generated by, or is a downstream product of, transgene-encoded one or more polypeptides that produce a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative; where the transgenic plant produces the cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in an amount that is at least about 50%, at least about 2-fold, at least about 5-fold, at least about 10-fold, at least about 25-fold, at least about 50-fold, or at least about 100-fold, or higher, than the amount of the cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative produced by the control plant, e.g., a non-transgenic plant (a plant that does not include the transgene encoding the one or more polypeptides) of the same species.

[00653] Methods of introducing exogenous nucleic acids into plant cells are well known in the art. Such plant cells are considered "transformed." Suitable methods may include viral infection (such as double stranded DNA viruses), transfection, conjugation, protoplast fusion, electroporation, particle gun technology, calcium phosphate precipitation, direct microinjection, silicon carbide whiskers technology, *Agrobacterium*-mediated transformation, CRISPR/Cas9-mediated genome editing, and the like. The choice of method

is generally dependent on the type of cell being transformed and the circumstances under which the transformation is taking place (e.g., *in vitro*, *ex vivo*, or *in vivo*).

[00654] A CRISPR/Cas9 system can be used to generate a transgenic (genetically modified) plant of the present disclosure. CRISPR/Cas9 systems and methods are known in the art. See, e.g., Bortesi and Fischer (2015) *Biotechnol. Advances* 33:41; and Fan et al. (2015) *Sci. Reports* 5:12217.

[00655] Transformation methods based upon the soil bacterium *Agrobacterium tumefaciens* can be used for introducing an exogenous nucleic acid into a vascular plant. The wild type form of *Agrobacterium* contains a Ti (tumor-inducing) plasmid that directs production of tumorigenic crown gall growth on host plants. Transfer of the tumor-inducing T-DNA region of the Ti plasmid to a plant genome requires the Ti plasmid-encoded virulence genes as well as T-DNA borders, which are a set of direct DNA repeats that delineate the region to be transferred. An *Agrobacterium*-based vector is a modified form of a Ti plasmid, in which the tumor inducing functions are replaced by the nucleotide sequence of interest to be introduced into the plant host.

[00656] *Agrobacterium*-mediated transformation generally employs cointegrate vectors or, e.g., binary vector systems, in which the components of the Ti plasmid are divided between a helper vector, which resides permanently in the *Agrobacterium* host and carries the virulence genes, and a shuttle vector, which contains the gene of interest bounded by T-DNA sequences. A variety of binary vectors are well known in the art and are commercially available, for example, from Clontech (Palo Alto, Calif.). Methods of coculturing *Agrobacterium* with cultured plant cells or wounded tissue such as leaf tissue, root explants, hypocotyledons, stem pieces or tubers, for example, also are well known in the art. See., e.g., Glick and Thompson, (eds.), *Methods in Plant Molecular Biology and Biotechnology*, Boca Raton, Fla.: CRC Press (1993).

[00657] *Agrobacterium*-mediated transformation is useful for producing a variety of transgenic vascular plants (Wang et al., supra, 1995) including at least one species of *Eucalyptus* and forage legumes such as alfalfa (lucerne); birdsfoot trefoil, white clover, *Stylosanthes, Lotononis bainessii* and sainfoin.

[00658] Microprojectile-mediated transformation also can be used to produce a subject transgenic plant. This method, first described by Klein et al. (*Nature* 327:70-73 (1987)), relies on microprojectiles such as gold or tungsten that are coated with the desired heterologous nucleic acid by precipitation with calcium chloride, spermidine or polyethylene

glycol. The microprojectile particles are accelerated at high speed into an angiosperm tissue using a device such as the BIOLISTIC PD-1000 (Biorad; Hercules Calif.).

[00659] A subject heterologous nucleic acid may be introduced into a plant in a manner such that the heterologous nucleic acid is able to enter a plant cell(s), e.g., via an *in vivo* or *ex vivo* protocol. By "*in vivo*," it may mean that the heterologous nucleic acid is administered to a living body of a plant e.g. infiltration. By "*ex vivo*" it may mean that cells or explants are modified outside of the plant, and then such cells or organs are regenerated to a plant. A number of vectors or constructs suitable for stable transformation of plant cells or for the establishment of transgenic plants have been described, including those described in Weissbach and Weissbach, (1989) Methods for Plant Molecular Biology Academic Press, and Gelvin et al., (1990) Plant Molecular Biology Manual, Kluwer Academic Publishers. Specific examples may include those derived from a Ti plasmid of *Agrobacterium tumefaciens*, as well as those disclosed by Herrera-Estrella et al. (1983) Nature 303: 209, Bevan (1984) Nucl Acid Res. 12: 8711-8721, Klee (1985) Bio/Technolo 3: 637-642. Alternatively, non-Ti vectors can be used to transfer the DNA into plants and cells by using free DNA delivery techniques. By using these methods transgenic plants such as wheat, rice (Christou (1991) Bio/Technology 9:957-962) and corn (Gordon-Kamm (1990) Plant Cell 2: 603-618) can be produced. An immature embryo can also be a good target tissue for monocots for direct DNA delivery techniques by using the particle gun (Weeks et al. (1993) Plant Physiol 102: 1077-1084; Vasil (1993) Bio/Technolo 10: 667-674; Wan and Lemeaux (1994) Plant Physiol 104: 37-48 and for Agrobacterium-mediated DNA transfer (Ishida et al. (1996) Nature Biotech 14: 745-750). Exemplary methods for introduction of DNA into chloroplasts are biolistic bombardment, polyethylene glycol transformation of protoplasts, and microinjection (Danieli et al Nat. Biotechnol 16:345-348, 1998; Staub et al Nat. Biotechnol 18: 333-338, 2000; O'Neill et al Plant J. 3:729-738, 1993; Knoblauch et al Nat. Biotechnol 17: 906-909; U.S. Pat. Nos. 5,451,513, 5,545,817, 5,545,818, and 5,576,198; in Intl. Application No. WO 95/16783; and in Boynton et al., Methods in Enzymology 217: 510-536 (1993), Svab et al., Proc. Natl. Acad. Sci. USA 90: 913-917 (1993), and McBride et al., Proc. Natl. Acad. Sci. USA 91: 7301-7305 (1994)). Any vector suitable for the methods of biolistic bombardment, polyethylene glycol transformation of protoplasts and microinjection will be suitable as a targeting vector for chloroplast transformation. Any double stranded DNA vector may be used as a transformation vector, especially when the method of introduction does not utilize *Agrobacterium*.

[00660] Plants which can be genetically modified may include grains, forage crops, fruits, vegetables, oil seed crops, palms, forestry, and vines. Specific examples of plants which can be modified may include, but are not limited to, maize, banana, peanut, field peas, sunflower, tobacco, tomato, canola, tobacco, wheat, barley, oats, potato, soybeans, cotton, sorghum, lupin, and rice. Plants which can be genetically modified may include *Theobroma cacao*.

[00661] Also provided by the present disclosure are transformed plant cells, tissues, plants and products that contain the transformed plant cells. A feature of the subject transformed cells, and tissues and products that include the same is the presence of a subject heterologous nucleic acid integrated into the genome, and production by plant cells of one or more polypeptides that are utilized to generate a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative. Recombinant plant cells of the present disclosure are useful as populations of recombinant cells, or as a tissue, seed, whole plant, stem, fruit, leaf, root, flower, stem, tuber, grain, animal feed, a field of plants, and the like.

[00662] Also provided by the present disclosure is reproductive material of a subject transgenic plant, where reproductive material may include seeds, progeny plants and clonal material, where such material can give rise to a plant that produces a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative.

Methods of Producing a Cannabinoid, a Cannabinoid Precursor, a Cannabinoid Derivative, or a Cannabinoid Precursor Derivative

[00663] The present disclosure provides methods of producing a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative. The methods may involve culturing a genetically modified host cell of the present disclosure in a suitable medium and recovering the produced cannabinoid, the cannabinoid precursor, the cannabinoid precursor derivative, or the cannabinoid derivative. The methods may also involve cell-free production of cannabinoids, cannabinoid precursors, cannabinoid precursor derivatives, or cannabinoid derivatives using one or more polypeptides disclosed herein expressed or overexpressed by a genetically modified host cell of the disclosure.

[00664] The present disclosure provides methods of producing a cannabinoid or a cannabinoid derivative. The methods may involve culturing a genetically modified host cell of the present disclosure in a suitable medium and recovering the produced cannabinoid or cannabinoid derivative. The methods may also involve cell-free production of cannabinoids

or cannabinoid derivatives using one or more polypeptides disclosed herein expressed or overexpressed by a genetically modified host cell of the disclosure.

[00665] Cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives that can be produced with the methods or genetically modified host cells of the present disclosure may include, but are not limited to, cannabichromene (CBC) type (e.g. cannabichromenic acid), cannabigerol (CBG) type (e.g. cannabigerolic acid), cannabidiol (CBD) type (e.g. cannabidiolic acid), Δ^9-trans-tetrahydrocannabinol (Δ^9 -THC) type (e.g. Δ^9-tetrahydrocannabinolic acid), Δ^8-trans-tetrahydrocannabinol (Δ^8 -THC) type, cannabicyclol (CBL) type, cannabielsoin (CBE) type, cannabinol (CBN) type, cannabinodiol (CBND) type, cannabitriol (CBT) type, olivetolic acid, GPP, derivatives of any of the foregoing, and others as listed in Elsohly M.A. and Slade D., Life Sci. 2005 Dec 22;78(5):539-48. Epub 2005 Sep 30.

[00666] Cannabinoids or cannabinoid derivatives that can be produced with the methods or genetically modified host cells of the present disclosure may also include, but are not limited to, cannabigerolic acid (CBGA), cannabigerolic acid monomethylether (CBGAM), cannabigerol (CBG), cannabigerol monomethylether (CBGM), cannabigerovarinic acid (CBGVA), cannabigerovarin (CBGV), cannabichromenic acid (CBCA), cannabichromene (CBC), cannabichromevarinic acid (CBCVA), cannabichromevarin (CBCV), cannabidiolic acid (CBDA), cannabidiol (CBD), cannabidiol monomethylether (CBDM), cannabidiol-C_4 (CBD-C_4), cannabidivarinic acid (CBDVA), cannabidivarin (CBDV), cannabidiorcol (CBD-C_1), Δ^9 –tetrahydrocannabinolic acid A (THCA-A), Δ^9 –tetrahydrocannabinolic acid B (THCA-B), Δ^9 –tetrahydrocannabinol (THC), Δ^9 –tetrahydrocannabinolic acid-C_4 (THCA-C_4), Δ^9 –tetrahydrocannabinol-C_4 (THC-C_4), Δ^9 –tetrahydrocannabivarinic acid (THCVA), Δ^9 –tetrahydrocannabivarin (THCV), Δ^9 –tetrahydrocannabiorcolic acid (THCA-C_1), Δ^9 –tetrahydrocannabiorcol (THC-C_1), Δ^7 –cis-iso-tetrahydrocannabivarin, Δ^8 –tetrahydrocannabinolic acid (Δ^8 –THCA), Δ^8 –tetrahydrocannabinol (Δ^8 –THC), cannabicyclolic acid (CBLA), cannabicyclol (CBL), cannabicyclovarin (CBLV), cannabielsoic acid A (CBEA-A), cannabielsoic acid B (CBEA-B), cannabielsoin (CBE), cannabielsoinic acid, cannabicitranic acid, cannabinolic acid (CBNA), cannabinol (CBN), cannabinol methylether (CBNM), cannabinol-C_4, (CBN-C_4), cannabivarin (CBV), cannabinol-C_2 (CNB-C_2), cannabiorcol (CBN-C_1), cannabinodiol (CBND), cannabinodivarin (CBVD), cannabitriol (CBT), 10-ethyoxy-9-hydroxy-delta-6a-tetrahydrocannabinol, 8,9-dihydroxyl-delta-6a-tetrahydrocannabinol, cannabitriolvarin (CBTVE), dehydrocannabifuran (DCBF), cannabifuran (CBF), cannabichromanon (CBCN),

cannabicitran (CBT), 10-oxo-delta-6a-tetrahydrocannabinol (OTHC), delta-9-cis-tetrahydrocannabinol (cis-THC), 3,4,5,6-tetrahydro-7-hydroxy-alpha-alpha-2-trimethyl-9-n-propyl-2,6-methano-2H-1-benzoxocin-5-methanol (OH-iso-HHCV), cannabiripsol (CBR), trihydroxy-delta-9-tetrahydrocannabinol (triOH-THC), and derivatives of any of the foregoing.

[00667] Additional cannabinoid derivatives that can be produced with the methods or genetically modified host cells of the present disclosure may also include, but are not limited to, 2-geranyl-5-pentyl-resorcylic acid, 2-geranyl-5-(4-pentynyl)-resorcylic acid, 2-geranyl-5-(trans-2-pentenyl)-resorcylic acid, 2-geranyl-5-(4-methylhexyl)-resorcylic acid, 2-geranyl-5-(5-hexynyl) resorcylic acid, 2-geranyl-5-(trans-2-hexenyl)-resorcylic acid, 2-geranyl-5-(5-hexenyl)-resorcylic acid, 2-geranyl-5-heptyl-resorcylic acid, 2-geranyl-5-(6-heptynoic)-resorcylic acid, 2-geranyl-5-octyl-resorcylic acid, 2-geranyl-5-(trans-2-octenyl)-resorcylic acid, 2-geranyl-5-nonyl-resorcylic acid, 2-geranyl-5-(trans-2-nonenyl) resorcylic acid, 2-geranyl-5-decyl-resorcylic acid, 2-geranyl-5-(4-phenylbutyl)-resorcylic acid, 2-geranyl-5-(5-phenylpentyl)-resorcylic acid, 2-geranyl-5-(6-phenylhexyl)-resorcylic acid, 2-geranyl-5-(7-phenylheptyl)-resorcylic acid, (6aR,10aR)-1-hydroxy-6,6,9-trimethyl-3-propyl-6a,7,8,10a-tetrahydro-6H-dibenzo[b,d]pyran-2-carboxylic acid, (6aR,10aR)-1-hydroxy-6,6,9-trimethyl-3-(4-methylhexyl)-6a,7,8,10a-tetrahydro-6H-dibenzo[b,d]pyran-2-carboxylic acid, (6aR,10aR)-1-hydroxy-6,6,9-trimethyl-3-(5-hexenyl)-6a,7,8,10a-tetrahydro-6H-dibenzo[b,d]pyran-2-carboxylic acid, (6aR,10aR)-1-hydroxy-6,6,9-trimethyl-3-(5-hexenyl)-6a,7,8,10a-tetrahydro-6H-dibenzo[b,d]pyran-2-carboxylic acid, (6aR,10aR)-1-hydroxy-6,6,9-trimethyl-3-(6-heptynyl)-6a,7,8,10a-tetrahydro-6H-dibenzo[b,d]pyran-2-carboxylic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-6-(hexan-2-yl)-2,4-dihydroxybenzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-(2-methylpentyl)benzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-(3-methylpentyl)benzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-(4-methylpentyl)benzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-[(1E)-pent-1-en-1-yl]benzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-[(2E)-pent-2-en-1-yl]benzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-[(2E)-pent-3-en-1-yl]benzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-(pent-4-en-1-yl)benzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-propylbenzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-butylbenzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-hexylbenzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-heptylbenzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-

6-octylbenzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-nonanylbenzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-decanylbenzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-undecanylbenzoic acid, 6-(4-chlorobutyl)-3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxybenzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-[4-(methylsulfanyl)butyl]benzoic acid, and others as listed in Bow, E. W. and Rimoldi, J. M., "The Structure–Function Relationships of Classical Cannabinoids: CB1/CB2 Modulation," *Perspectives in Medicinal Chemistry* 2016:8 17–39 doi: 10.4137/PMC.S32171, incorporated by reference herein.

[00668] Cannabinoid precursor derivatives that can be produced with the methods or genetically modified host cells of the present disclosure may also include, but are not limited to, divarinolic acid, 5-pentyl-resorcylic acid, 5-(4-pentynyl)-resorcylic acid, 5-(trans-2-pentenyl)-resorcylic acid, 5-(4-methylhexyl)-resorcylic acid, 5-(5-hexynyl)-resorcylic acid, 5-(trans-2-hexenyl)-resorcylic acid, 5-(5-hexenyl)-resorcylic acid, 5-heptyl-resorcylic acid, 5-(6-heptynoic)-resorcylic acid, 5-octyl-resorcylic acid, 5-(trans-2-octenyl)-resorcylic acid, 5-nonyl-resorcylic acid, 5-(trans-2-nonenyl)-resorcylic acid, 5-decyl-resorcylic acid, 5-(4-phenylbutyl)-resorcylic acid, 5-(5-phenylpentyl)-resorcylic acid, 5-(6-phenylhexyl)-resorcylic acid, and 5-(7-phenylheptyl)-resorcylic acid.

[00669] Cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives that can be produced with the methods or genetically modified host cells of the present disclosure may also include, but are not limited to, polyketides or polyketide derivatives.

[00670] A cannabinoid derivative or cannabinoid precursor derivative may lack one or more chemical moieties found in a naturally-occurring cannabinoid or naturally-occurring cannabinoid precursor. Such chemical moieties may include, but are not limited to, methyl, alkyl, alkenyl, methoxy, alkoxy, acetyl, carboxyl, carbonyl, oxo, ester, hydroxyl, aryl, heteroaryl, cycloalkyl, cycloalkenyl, cycloalkylalkenyl, cycloalkenylalkyl, cycloalkenylalkenyl, heterocyclylalkenyl, heteroarylalkenyl, arylalkenyl, heterocyclyl, aralkyl, cycloalkylalkyl, heterocyclylalkyl, heteroarylalkyl, and the like. In some embodiments, a cannabinoid derivative or cannabinoid precursor derivative lacking one or more chemical moieties found in a naturally-occurring cannabinoid or naturally-occurring cannabinoid precursor, and produced by a genetically modified host cell disclosed herein or in a cell-free reaction mixture comprising one or more of the polypeptides disclosed herein, may also comprise one or more of any of the functional and/or reactive groups described

herein. Functional and reactive groups may be optionally substituted with one or more additional functional or reactive groups.

[00671] A cannabinoid derivative or cannabinoid precursor derivative may be a cannabinoid or cannabinoid precursor comprising one or more functional and/or reactive groups and is produced by a genetically modified host cell disclosed herein or in a cell-free reaction mixture comprising one or more of the polypeptides disclosed herein. Functional groups may include, but are not limited to, azido, halo (e.g., chloride, bromide, iodide, fluorine), methyl, alkyl, alkynyl, alkenyl, methoxy, alkoxy, acetyl, amino, carboxyl, carbonyl, oxo, ester, hydroxyl, thio, cyano, aryl, heteroaryl, cycloalkyl, cycloalkenyl, cycloalkylalkenyl, cycloalkylalkynyl, cycloalkenylalkyl, cycloalkenylalkenyl, cycloalkenylalkynyl, heterocyclylalkenyl, heterocyclylalkynyl, heteroarylalkenyl, heteroarylalkynyl, arylalkenyl, arylalkynyl, spirocyclyl, heterospirocyclyl, heterocyclyl, thioalkyl, sulfone, sulfonyl, sulfoxide, amido, alkylamino, dialkylamino, arylamino, alkylarylamino, diarylamino, N-oxide, imide, enamine, imine, oxime, hydrazone, nitrile, aralkyl, cycloalkylalkyl, haloalkyl, heterocyclylalkyl, heteroarylalkyl, nitro, thioxo, and the like. See, e.g., FIGS. 12 and 13. Suitable reactive groups may include, but are not necessarily limited to, azide, carboxyl, carbonyl, amine, (e.g., alkyl amine (e.g., lower alkyl amine), aryl amine), halide, ester (e.g., alkyl ester (e.g., lower alkyl ester, benzyl ester), aryl ester, substituted aryl ester), cyano, thioester, thioether, sulfonyl halide, alcohol, thiol, succinimidyl ester, isothiocyanate, iodoacetamide, maleimide, hydrazine, alkynyl, alkenyl, acetyl, and the like. In some embodiments, the reactive group is selected from a carboxyl, a carbonyl, an amine, an ester, a thioester, a thioether, a sulfonyl halide, an alcohol, a thiol, an alkyne, alkene, an azide, a succinimidyl ester, an isothiocyanate, an iodoacetamide, a maleimide, and a hydrazine. Functional and reactive groups may be optionally substituted with one or more additional functional or reactive groups.

[00672] A reactive group may facilitate covalent attachment of a molecule of interest. Suitable molecules of interest may include, but are not limited to, a detectable label; imaging agents; a toxin (including cytotoxins); a linker; a peptide; a drug (e.g., small molecule drugs); a member of a specific binding pair; an epitope tag; ligands for binding by a target receptor; tags to aid in purification; molecules that increase solubility; and the like. A linker may be a peptide linker or a non-peptide linker.

[00673] In some embodiments, a cannabinoid derivative or a cannabinoid precursor derivative comprising an azide may be reacted with a compound comprising an alkyne group via "click chemistry" to generate a product comprising a heterocycle, also known as an

azide-alkyne cycloaddition. In some embodiments, a cannabinoid derivative or a cannabinoid precursor derivative comprising an alkyne may be reacted with a compound comprising an azide group via click chemistry to generate a product comprising a heterocycle.

[00674] Additional molecules that may be desirable for attachment to a cannabinoid derivative or cannabinoid precursor derivative may include, but are not necessarily limited to, detectable labels (e.g., spin labels, fluorescence resonance energy transfer (FRET)-type dyes, e.g., for studying structure of biomolecules *in vivo*); small molecule drugs; cytotoxic molecules (e.g., drugs); imaging agents; ligands for binding by a target receptor; tags to aid in purification by, for example, affinity chromatography (e.g., attachment of a FLAG epitope); molecules that increase solubility (e.g., poly(ethylene glycol); molecules that enhance bioavailability; molecules that increase *in vivo* half-life; molecules that target to a particular cell type (e.g., an antibody specific for an epitope on a target cell); molecules that target to a particular tissue; molecules that provide for crossing the blood-brain barrier; and molecules to facilitate selective attachment to a surface, and the like.

[00675] In some embodiments, a molecule of interest comprises an imaging agent. Suitable imaging agents may include positive contrast agents and negative contrast agents. Suitable positive contrast agents may include, but are not limited to, gadolinium tetraazacyclododecanetetraacetic acid (Gd-DOTA); gadolinium-diethylenetriaminepentaacetic acid (Gd-DTPA); gadolinium-1,4,7-tris(carbonylmethyl)-10-(2'-hydroxypropyl)-1,4,7,10-tetraazacyclododecane (Gd-HP-DO3A); Manganese(II)-dipyridoxal diphosphate (Mn-DPDP); Gd-diethylenetriaminepentaacetate-bis(methylamide) (Gd-DTPA-BMA); and the like. Suitable negative contrast agents may include, but are not limited to, a superparamagnetic iron oxide (SPIO) imaging agent; and a perfluorocarbon, where suitable perfluorocarbons may include, but are not limited to, fluoroheptanes, fluorocycloheptanes, fluoromethylcycloheptanes, fluorohexanes, fluorocyclohexanes, fluoropentanes, fluorocyclopentanes, fluoromethylcyclopentanes, fluorodimethylcyclopentanes, fluoromethylcyclobutanes, fluorodimethylcyclobutanes, fluorotrimethylcyclobutanes, fluorobutanes, fluorocyclobutanse, fluoropropanes, fluoroethers, fluoropolyethers, fluorotriethylamines, perfluorohexanes, perfluoropentanes, perfluorobutanes, perfluoropropanes, sulfur hexafluoride, and the like.

[00676] Additional cannabinoid derivatives and cannabinoid precursor derivatives that can be produced with a method or genetically modified host cell of the present disclosure may include derivatives that have been modified via organic synthesis or an enzymatic route

to modify drug metabolism and pharmacokinetics (e.g. solubility, bioavailability, absorption, distribution, plasma half-life and metabolic clearance). Modification examples may include, but are not limited to, halogenation, acetylation and methylation.

[00677] The cannabinoids, cannabinoid derivatives, cannabinoid precursors, and cannabinoid precursor derivatives described herein further include all pharmaceutically acceptable isotopically labeled cannabinoids, cannabinoid derivatives, cannabinoid precursors, and cannabinoid precursor derivatives. An "isotopically-" or "radio-labeled" compound is a compound where one or more atoms are replaced or substituted by an atom having an atomic mass or mass number different from the atomic mass or mass number typically found in nature (i.e., naturally occurring). For example, in some embodiments, in the cannabinoids, cannabinoid derivatives, cannabinoid precursors, and cannabinoid precursor derivatives described herein hydrogen atoms are replaced or substituted by one or more deuterium or tritium. Certain isotopically labeled cannabinoids, cannabinoid derivatives, cannabinoid precursors, and cannabinoid precursor derivatives of this disclosure, for example, those incorporating a radioactive isotope, are useful in drug and/or substrate tissue distribution studies. The radioactive isotopes tritium, i.e., ^{3}H, and carbon 14, i.e., ^{14}C, are particularly useful for this purpose in view of their ease of incorporation and ready means of detection. Substitution with heavier isotopes such as deuterium, i.e., ^{2}H, may afford certain therapeutic advantages resulting from greater metabolic stability, for example, increased *in vivo* half-life or reduced dosage requirements, and hence may be preferred in some circumstances. Suitable isotopes that may be incorporated in cannabinoids, cannabinoid derivatives, cannabinoid precursors, and cannabinoid precursor derivatives described herein include but are not limited to ^{2}H (also written as D for deuterium), ^{3}H (also written as T for tritium), ^{11}C, ^{13}C, ^{14}C, ^{13}N, ^{15}N, ^{15}O, ^{17}O, ^{18}O, ^{18}F, ^{35}S, ^{36}Cl, ^{82}Br, ^{75}Br, ^{76}Br, ^{77}Br, ^{123}I, ^{124}I, ^{125}I, and ^{131}I. Substitution with positron emitting isotopes, such as ^{11}C, ^{18}F, ^{15}O, and ^{13}N, can be useful in Positron Emission Topography (PET) studies.

[00678] The methods of bioproduction disclosed herein enable synthesis of cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives with defined stereochemistries, which is challenging to do using chemical synthesis. Cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives disclosed herein may be enantiomers or disastereomers. The term "enantiomers" may refer to a pair of stereoisomers which are non-superimposable mirror images of one another. In some embodiments the cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives may be the (*S*)-enantiomer. In

some embodiments the cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives may be the (*R*)-enantiomer. In some embodiments, the cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives may be the (+) or (-) enantiomers. The term "diastereomers" may refer to the set of stereoisomers which cannot be made superimposable by rotation around single bonds. For example, cis- and trans- double bonds, endo- and exo- substitution on bicyclic ring systems, and compounds containing multiple stereogenic centers with different relative configurations may be considered to be diastereomers. The term "diastereomer" may refer to any member of this set of compounds. Cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives disclosed herein may include a double bond or a fused ring. In certain such embodiments, the double bond or fused ring may be cis or trans, unless the configuration is specifically defined. If the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative contains a double bond, the substituent may be in the E or Z configuration, unless the configuration is specifically defined.

[00679] In some embodiments when the cannabinoid, cannabinoid precursor, cannabinoid precursor derivative, or cannabinoid derivative is recovered from the cell lysate, from the culture medium, from both the cell lysate and the culture medium, or from a cell-free reaction mixture comprising one or more polypeptides disclosed herein, the recovered cannabinoid, cannabinoid precursor, cannabinoid precursor derivative, or cannabinoid derivative is in the form of a salt. In certain such embodiments, the salt is a pharmaceutically acceptable salt. In some embodiments, the salt of the recovered cannabinoid, cannabinoid precursor, cannabinoid precursor derivative, or cannabinoid derivative is then purified as disclosed herein.

[00680] The disclosure includes pharmaceutically acceptable salts of the cannabinoids, cannabinoid derivatives, cannabinoid precursors, and cannabinoid precursor derivatives described herein. "Pharmaceutically acceptable salts" refer to those salts which retain the biological effectiveness and properties of the free bases, which are not biologically or otherwise undesirable. Representative pharmaceutically acceptable salts include, but are not limited to, e.g., water-soluble and water-insoluble salts, such as the acetate, amsonate (4,4-diaminostilbene-2,2-disulfonate), benzenesulfonate, benzonate, bicarbonate, bisulfate, bitartrate, borate, bromide, butyrate, calcium, calcium edetate, camsylate, carbonate, chloride, citrate, clavulariate, dihydrochloride, edetate, edisylate, estolate, esylate, fiunarate, gluceptate, gluconate, glutamate, glycollylarsanilate, hexafluorophosphate, hexylresorcinate,

hydrabamine, hydrobromide, hydrochloride, hydroxynaphthoate, iodide, sethionate, lactate, lactobionate, laurate, magnesium, malate, maleate, mandelate, mesylate, methylbromide, methylnitrate, methylsulfate, mucate, napsylate, nitrate, N-methylglucamine ammonium salt, 3-hydroxy-2-naphthoate, oleate, oxalate, palmitate, pamoate (1,1-methene-bis-2-hydroxy-3-naphthoate, einbonate), pantothenate, phosphate/diphosphate, picrate, polygalacturonate, propionate, p-toluenesulfonate, salicylate, stearate, subacetate, succinate, sulfate, sulfosalicylate, suramate, tannate, tartrate, teoclate, tosylate, triethiodide, and valerate salts.

[00681] "Pharmaceutically acceptable salt" also includes both acid and base addition salts. "Pharmaceutically acceptable acid addition salt" refers to those salts which retain the biological effectiveness and properties of the free bases, which are not biologically or otherwise undesirable, and which are formed with inorganic acids such as, but are not limited to, hydrochloric acid, hydrobromic acid, sulfuric acid, nitric acid, phosphoric acid and the like, and organic acids such as, but not limited to, acetic acid, 2,2-dichloroacetic acid, adipic acid, alginic acid, ascorbic acid, aspartic acid, benzenesulfonic acid, benzoic acid, 4-acetamidobenzoic acid, camphoric acid, camphor-10-sulfonic acid, capric acid, caproic acid, caprylic acid, carbonic acid, cinnamic acid, citric acid, cyclamic acid, dodecylsulfuric acid, ethane-1,2-disulfonic acid, ethanesulfonic acid, 2-hydroxyethanesulfonic acid, formic acid, fumaric acid, galactaric acid, gentisic acid, glucoheptonic acid, gluconic acid, glucuronic acid, glutamic acid, glutaric acid, 2-oxo-glutaric acid, glycerophosphoric acid, glycolic acid, hippuric acid, isobutyric acid, lactic acid, lactobionic acid, lauric acid, maleic acid, malic acid, malonic acid, mandelic acid, methanesulfonic acid, mucic acid, naphthalene-l,5-disulfonic acid, naphthalene-2-sulfonic acid, l-hydroxy-2-naphthoic acid, nicotinic acid, oleic acid, orotic acid, oxalic acid, palmitic acid, pamoic acid, propionic acid, pyroglutamic acid, pyruvic acid, salicylic acid, 4-aminosalicylic acid, sebacic acid, stearic acid, succinic acid, tartaric acid, thiocyanic acid, p-toluenesulfonic acid, trifluoroacetic acid, undecylenic acid, and the like.

[00682] "Pharmaceutically acceptable base addition salt" refers to those salts which retain the biological effectiveness and properties of the free acids, which are not biologically or otherwise undesirable. These salts are prepared from addition of an inorganic base or an organic base to the free acid. Salts derived from inorganic bases include, but are not limited to, the sodium, potassium, lithium, ammonium, calcium, magnesium, iron, zinc, copper, manganese, aluminum salts and the like. For example, inorganic salts include, but are not limited to, ammonium, sodium, potassium, calcium, and magnesium salts. Salts derived from organic bases include, but are not limited to, salts of primary, secondary, and tertiary amines,

substituted amines including naturally occurring substituted amines, cyclic amines and basic ion exchange resins, such as ammonia, isopropylamine, trimethylamine, diethylamine, triethylamine, tripropylamine, diethanolamine, ethanolamine, deanol, 2-dimethylaminoethanol, 2-diethylaminoethanol, dicyclohexylamine, lysine, arginine, histidine, caffeine, procaine, hydrabamine, choline, betaine, benethamine, benzathine, ethylenediamine, glucosamine, methylglucamine, theobromine, triethanolamine, tromethamine, purines, piperazine, piperidine, N-ethylpiperidine, polyamine resins and the like.

Methods of Using Host Cells to Generate Cannabinoids, Cannabinoid Precursors, Cannabinoid Derivatives, or Cannabinoid Precursor Derivatives

[00683] The disclosure provides methods of producing a cannabinoid, a cannabinoid precursor, a cannabinoid precursor derivative, or a cannabinoid derivative in a genetically modified host cell, the method comprising: culturing a genetically modified host cell of the disclosure in a suitable medium and recovering the produced cannabinoid, cannabinoid precursor, cannabinoid precursor derivative, or cannabinoid derivative. In certain such embodiments, the produced cannabinoid, cannabinoid precursor, cannabinoid precursor derivative, or cannabinoid derivative is then purified as disclosed herein.

[00684] In some embodiments, culturing of the genetically modified host cells of the disclosure in a suitable medium provides for synthesis of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in an increased amount compared to a non-genetically modified host cell cultured under similar conditions.

[00685] The disclosure provides methods of producing a cannabinoid, a cannabinoid precursor, a cannabinoid precursor derivative, or a cannabinoid derivative, the method comprising: culturing a genetically modified host cell of the disclosure in a suitable medium comprising a carboxylic acid and recovering the produced cannabinoid, cannabinoid precursor, cannabinoid precursor derivative, or cannabinoid derivative. In certain such embodiments, the produced cannabinoid, cannabinoid precursor, cannabinoid precursor derivative, or cannabinoid derivative is then purified as disclosed herein.

[00686] In some embodiments, the cannabinoid, cannabinoid precursor, cannabinoid precursor derivative, or cannabinoid derivative is recovered from the cell lysate, from the culture medium, or from both the cell lysate and the culture medium. In certain such embodiments, the recovered cannabinoid, cannabinoid precursor, cannabinoid precursor derivative, or cannabinoid derivative is then purified as disclosed herein.

[00687] The disclosure provides methods of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising: culturing a genetically modified host cell of the disclosure in a suitable medium and recovering the produced cannabinoid or cannabinoid derivative. In certain such embodiments, the produced cannabinoid or cannabinoid derivative is then purified as disclosed herein.

[00688] In some embodiments, culturing of the genetically modified host cells of the disclosure in a suitable medium provides for synthesis of a cannabinoid or a cannabinoid derivative in an increased amount compared to a non-genetically modified host cell cultured under similar conditions.

[00689] The disclosure provides methods of producing a cannabinoid or a cannabinoid derivative, the method comprising: culturing a genetically modified host cell of the disclosure in a suitable medium comprising a carboxylic acid and recovering the produced cannabinoid or cannabinoid derivative. In certain such embodiments, the produced cannabinoid or cannabinoid derivative is then purified as disclosed herein.

[00690] In some embodiments, the cannabinoid or cannabinoid derivative is recovered from the cell lysate, from the culture medium, or from both the cell lysate and the culture medium. In certain such embodiments, the recovered cannabinoid or cannabinoid derivative is then purified as disclosed herein.

[00691] In some embodiments, the genetically modified host cell of the present disclosure is cultured in a suitable medium comprising a carboxylic acid to generate an acyl-CoA compound or an acyl-CoA compound derivative. In certain such embodiments, the genetically modified host cell is genetically modified with one or more heterologous nucleic acids encoding an AAE polypeptide, a FAA polypeptide, or a fatty acyl-CoA ligase polypeptide, as described herein. In some embodiments, the genetically modified host cells of the present disclosure may further convert an acyl-CoA compound or an acyl-CoA compound derivative to cannabinoids, cannabinoid precursors, cannabinoid precursor derivatives, or cannabinoid derivatives.

[00692] Carboxylic acids may include, but are not limited to, C_3-C_{18} fatty acids, butyric acid, isobutyric acid, valeric acid, hexanoic acid, heptanoic acid, octanoic acid, nonanoic acid, decanoic acid, undecanoic acid, lauric acid, myristic acid, C_{15}-C_{18} fatty acids, fumaric acid, itaconic acid, malic acid, succinic acid, maleic acid, malonic acid, glutaric acid, glucaric acid, oxalic acid, adipic acid, pimelic acid, suberic acid, azelaic acid, sebacic acid, dodecanedioic acid, glutaconic acid, ortho-phthalic acid, isophthalic acid, terephthalic acid, citric acid, isocitric acid, aconitic acid, tricarballylic acid, and

trimesic acid. Carboxylic acids may include C_4-C_{10} carboxylic acids. In some embodiments, the carboxylic acid is a C_4 carboxylic acid. In some embodiments, the carboxylic acid is a C_5 carboxylic acid. In some embodiments, the carboxylic acid is a C_6 carboxylic acid. In some embodiments, the carboxylic acid is a C_7 carboxylic acid. In some embodiments, the carboxylic acid is a C_8 carboxylic acid. In some embodiments, the carboxylic acid is a C_9 carboxylic acid. In some embodiments, the carboxylic acid is a C_{10} carboxylic acid. In some embodiments, the carboxylic acid is butyric acid. In some embodiments, the carboxylic acid is valeric acid. In some embodiments, the carboxylic acid is hexanoic acid. In some embodiments, the carboxylic acid is heptanoic acid. In some embodiments, the carboxylic acid is octanoic acid. In some embodiments, the carboxylic acid is nonanoic acid. In some embodiments, the carboxylic acid is decanoic acid. See, e.g., FIG. 12.

[00693] In some embodiments, the carboxylic acid comprises one or more functional and/or reactive groups to generate derivatives of hexanoyl-CoA or derivatives of acyl-CoA compounds. Functional groups may include, but are not limited to, azido, halo (e.g., chloride, bromide, iodide, fluorine), methyl, alkyl, alkynyl, alkenyl, methoxy, alkoxy, acetyl, amino, carboxyl, carbonyl, oxo, ester, hydroxyl, thio, cyano, aryl, heteroaryl, cycloalkyl, cycloalkenyl, cycloalkylalkenyl, cycloalkylalkynyl, cycloalkenylalkyl, cycloalkenylalkenyl, cycloalkenylalkynyl, heterocyclylalkenyl, heterocyclylalkynyl, heteroarylalkenyl, heteroarylalkynyl, arylalkenyl, arylalkynyl, spirocyclyl, heterospirocyclyl, heterocyclyl, thioalkyl, sulfone, sulfonyl, sulfoxide, amido, alkylamino, dialkylamino, arylamino, alkylarylamino, diarylamino, N-oxide, imide, enamine, imine, oxime, hydrazone, nitrile, aralkyl, cycloalkylalkyl, haloalkyl, heterocyclylalkyl, heteroarylalkyl, nitro, thioxo, and the like. See, e.g., FIGS. 12 and 13. Reactive groups may include, but are not necessarily limited to, azide, halogen, carboxyl, carbonyl, amine, (e.g., alkyl amine (e.g., lower alkyl amine), aryl amine), ester (e.g., alkyl ester (e.g., lower alkyl ester, benzyl ester), aryl ester, substituted aryl ester), cyano, thioester, thioether, sulfonyl halide, alcohol, thiol, succinimidyl ester, isothiocyanate, iodoacetamide, maleimide, hydrazine, alkynyl, alkenyl, and the like. In some embodiments, the reactive group is selected from a carboxyl, a carbonyl, an amine, an ester, thioester, thioether, a sulfonyl halide, an alcohol, a thiol, a succinimidyl ester, an isothiocyanate, an iodoacetamide, a maleimide, an azide, an alkyne, an alkene, and a hydrazine. Functional and reactive groups may be optionally substituted with one or more additional functional or reactive groups.

[00694] In some embodiments, the carboxylic acid is isotopically- or radio-labeled. In some embodiments, the carboxylic acid may be an enantiomer or disastereomer. In some

embodiments the carboxylic acid may be the (*S*)-enantiomer. In some embodiments the carboxylic acid may be the (*R*)-enantiomer. In some embodiments, the carboxylic acid may be the (+) or (-) enantiomer. In some embodiments, the carboxylic acid may include a double bond or a fused ring. In certain such embodiments, the double bond or fused ring may be cis or trans, unless the configuration is specifically defined. If the carboxylic acid contains a double bond, the substituent may be in the E or Z configuration, unless the configuration is specifically defined.

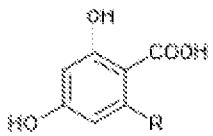
[00695] In some embodiments, the carboxylic acid comprises a C=C group. In some embodiments, the carboxylic acid comprises an alkyne group. In some embodiments, the carboxylic acid comprises an N_3 group. In some embodiments, the carboxylic acid comprises a halogen. In some embodiments, the carboxylic acid comprises a CN group. In some embodiments, the carboxylic acid comprises an iodide. In some embodiments, the carboxylic acid comprises a bromide. In some embodiments, the carboxylic acid comprises chloride. In some embodiments, the carboxylic acid comprises fluoride. In some embodiments, the carboxylic acid comprises a carbonyl. In some embodiments, the carboxylic acid comprises an acetyl. In some embodiments, the carboxylic acid comprises an alkyl group.

[00696] Carboxylic acids may include, but are not limited to, 2-methylhexanoic acid, 3-methylhexanoic acid, 4-methylhexanoic acid, 5-methylhexanoic acid, 2-hexenoic acid, 3-hexenoic acid, 4-hexenoic acid, 5-hexenoic acid, 5-chlorovaleric acid, 5-aminovaleric acid, 5-cyanovaleric acid, 5-(methylsulfanyl)valeric acid, 5-hydroxyvaleric acid, 5-phenylvaleric acid, 2,3-dimethylhexanoic acid, d_3-hexanoic acid, 5-chloropentanoic acid, 5-(methylsulfanyl)pentanoic acid, 4-pentynoic acid, trans-2-pentenoic acid, 5-hexynoic acid, trans-2-hexenoic acid, 6-heptynoic acid, trans-2-octenoic acid, trans-2-nonenoic acid, 4-phenylbutyric acid, 6-phenylhexanoic acid, 7-phenylyheptanoic acid, and the like. In some embodiments, the carboxylic acid is 2-methylhexanoic acid. In some embodiments, the carboxylic acid is 3-methylhexanoic acid. In some embodiments, the carboxylic acid is 4-methylhexanoic acid. In some embodiments, the carboxylic acid is 5-methylhexanoic acid. In some embodiments, the carboxylic acid is 2-hexenoic acid. In some embodiments, the carboxylic acid is 3-hexenoic acid. In some embodiments, the carboxylic acid is 4-hexenoic acid. In some embodiments, the carboxylic acid is 5-hexenoic acid. In some embodiments, the carboxylic acid is 5-chlorovaleric acid. In some embodiments, the carboxylic acid is 5-aminovaleric acid. In some embodiments, the carboxylic acid is 5-cyanovaleric acid. In some embodiments, the carboxylic acid is 5-(methylsulfanyl)valeric acid. In some embodiments, the carboxylic acid is 5-hydroxyvaleric acid. In some embodiments, the

carboxylic acid is 5-phenylvaleric acid. In some embodiments, the carboxylic acid is 2,3-dimethylhexanoic acid. In some embodiments, the carboxylic acid is d_3-hexanoic acid. In some embodiments, the carboxylic acid is 5-chloropentanoic acid. In some embodiments, the carboxylic acid is 5-(methylsulfanyl)pentanoic acid. In some embodiments, the carboxylic acid is 4-pentynoic acid. In some embodiments, the carboxylic acid is trans-2-pentenoic acid. In some embodiments, the carboxylic acid is 5-hexynoic acid. In some embodiments, the carboxylic acid is trans-2-hexenoic acid. In some embodiments, the carboxylic acid is 6-heptynoic acid. In some embodiments, the carboxylic acid is trans-2-octenoic acid. In some embodiments, the carboxylic acid is trans-2-nonenoic acid. In some embodiments, the carboxylic acid is 4-phenylbutyric acid. In some embodiments, the carboxylic acid is 6-phenylhexanoic acid. In some embodiments, the carboxylic acid is 7-phenylheptanoic acid.

[00697] The disclosure also provides methods of producing the following cannabinoid precursor or precursor derivatives: olivetolic acid or olivetolic acid derivatives. In certain such embodiments, the method comprises: culturing a genetically modified host cell of the disclosure in a suitable medium comprising a carboxylic acid, and recovering the produced olivetolic acid or olivetolic acid derivative. In certain such embodiments, the genetically modified host cell of the disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative, such as an AAE polypeptide, a FAA polypeptide, or a fatty acyl-CoA ligase polypeptide; b) one or more heterologous nucleic acids encoding a TKS polypeptide; and c) one or more heterologous nucleic acids encoding an OAC polypeptide. In some embodiments, the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative, such as an AAE polypeptide, a FAA polypeptide, or a fatty acyl-CoA ligase polypeptide; and b) one or more heterologous nucleic acids encoding a TKS/OAC fusion polypeptide.

[00698] In some embodiments, the olivetolic acid derivative produced by the methods or genetically modified host cells of the disclosure has the following formula:



where R is alkyl (e.g., C_1-C_{10} alkyl), substituted alkyl, alkyl ester, or alkyl-X, where X is a reactive or functional group, as disclosed herein. In some embodiments, the olivetolic acid or

olivetolic acid derivative is recovered from the cell lysate, from the culture medium, or from both the cell lysate and the culture medium. In certain such embodiments, the recovered olivetolic acid or olivetolic acid derivative is then purified as disclosed herein. In some embodiments, the olivetolic acid or olivetolic acid derivative is further converted by the genetically modified host cell to a cannabinoid derivative or a cannabinoid.

[00699] The disclosure also provides methods of producing a cannabinoid or a cannabinoid derivative, the method comprising: culturing a genetically modified host cell of the disclosure in a suitable medium comprising olivetolic acid or an olivetolic acid derivative and recovering the produced cannabinoid or cannabinoid derivative. In certain such embodiments, the produced cannabinoid or cannabinoid derivative is then purified as disclosed herein. The disclosure also provides methods of producing a cannabinoid derivative, the method comprising: culturing a genetically modified host cell of the disclosure in a suitable medium comprising olivetolic acid or an olivetolic acid derivative and recovering the produced cannabinoid derivative. In certain such embodiments, the produced cannabinoid derivative is then purified as disclosed herein.

[00700] In some embodiments, the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide and b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP (e.g., a GPPS polypeptide). In some embodiments, the olivetolic acid or olivetolic acid derivative is further converted to a cannabinoid derivative or a cannabinoid.

[00701] Olivetolic acid derivatives used herein may comprise one or more reactive and/or functional groups as disclosed herein. In some embodiments when the suitable medium comprises an olivetolic acid derivative, the olivetolic acid derivative is orsellinic acid. In some embodiments when the suitable medium comprises an olivetolic acid derivative, the olivetolic acid derivative is divarinic acid.

[00702] In some embodiments, a method of producing a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative may involve growing a transgenic (genetically modified) plant of the present disclosure under conditions that favor production of the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative. The cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative can be purified from the plant, or a part of the plant. The present disclosure provides food products made from a transgenic (genetically modified) plant of the present disclosure.

Exemplary Cell Culture Conditions

[00703] Suitable media may include standard culture media (e.g., Luria-Bertani broth, optionally supplemented with one or more additional agents, such as an inducer (e.g., where heterologous nucleic acids disclosed herein is under the control of an inducible promoter, etc.); standard yeast culture media; and the like). In some embodiments, the culture medium can be supplemented with a fermentable sugar (e.g., a hexose sugar, e.g., glucose, xylose, and the like). In some embodiments, the culture medium can be supplemented with hexanoate, carboxylic acids other than hexanoate, olivetolic acid, or olivetolic acid derivatives. In some embodiments, the culture medium can be supplemented with pretreated cellulosic feedstock (e.g., wheat grass, wheat straw, barley straw, sorghum, rice grass, sugarcane straw, bagasse, switchgrass, corn stover, corn fiber, grains, or any combination thereof). In some embodiments, the culture medium can be supplemented with oleic acid. In some embodiments, the suitable medium comprises a non-fermentable carbon source. In certain such embodiments, the non-fermentable carbon source comprises ethanol. In some embodiments, the suitable media comprises an inducer. In certain such embodiments, the inducer comprises galactose.

[00704] The carbon source in the suitable media can vary significantly, from simple sugars like glucose to more complex hydrolysates of other biomass, such as yeast extract. The addition of salts generally provide essential elements such as magnesium, nitrogen, phosphorus, and sulfur to allow the cells to synthesize polypeptides and nucleic acids. The suitable media can also be supplemented with selective agents, such as antibiotics, to select for the maintenance of certain plasmids and the like. For example, if a microorganism is resistant to a certain antibiotic, such as ampicillin or tetracycline, then that antibiotic can be added to the medium in order to prevent cells lacking the resistance from growing. The suitable media can be supplemented with other compounds as necessary to select for desired physiological or biochemical characteristics, such as particular amino acids and the like.

[00705] In some embodiments, genetically modified host cells disclosed herein are grown in minimal medium. As used herein, the terms "minimal medium" or "minimal media" may refer to growth medium containing the minimum nutrients possible for cell growth, generally, but not always, without the presence of one or more amino acids (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more amino acids). Minimal medium typically contains: (1) a carbon source for cellular (e.g. bacterial or yeast) growth; (2) various salts, which can vary among cellular (e.g. bacterial or yeast) species and growing conditions; and (3) water.

[00706] In some embodiments, genetically modified host cells disclosed herein are grown in rich medium or rich media. In certain such embodiments, the rich medium or rich media comprises yeast extract peptone dextrose (YPD) media comprising water, 10 g/L yeast extract, 20 g/L Bacto peptone, and 20 g/L dextrose (glucose). In some embodiments, the rich medium or rich media comprises YP + 20 g/L galactose and 1 g/L glucose. In some embodiments, the rich medium or rich media further comprises a carboxylic acid (e.g., 1 mM olivetolic acid, 1 mM olivetolic acid derivative, 2 mM hexanoic acid, or 2 mM of a carboxylic acid other than hexanoic acid). In some embodiments, rich medium or rich media affords more rapid cell growth compared to minimal media or minimal medium.

[00707] Materials and methods suitable for the maintenance and growth of the recombinant cells of the disclosure are described herein, e.g., in the Examples section. Other materials and methods suitable for the maintenance and growth of cell (e.g. bacterial or yeast) cultures are well known in the art. Exemplary techniques can be found in International Publication No. WO2009/076676, U.S. Patent Application No. 12/335,071 (U.S. Publ. No. 2009/0203102), WO 2010/003007, US Publ. No. 2010/0048964, WO 2009/132220, US Publ. No. 2010/0003716, Manual of Methods for General Bacteriology Gerhardt et al, eds), American Society for Microbiology, Washington, D.C. (1994) or Brock in Biotechnology: A Textbook of Industrial Microbiology, Second Edition (1989) Sinauer Associates, Inc., Sunderland, MA.

[00708] Standard cell culture conditions can be used to culture the genetically modified host cells disclosed herein (see, for example, WO 2004/033646 and references cited therein). In some embodiments, cells are grown and maintained at an appropriate temperature, gas mixture, and pH (such as at about 20 °C to about 37 °C, at about 0.04% to about 84% CO_2, at about 0% to about 100% dissolved oxygen, and at a pH between about 2 to about 9). In some embodiments, genetically modified host cells disclosed herein are grown at about 34 °C in a suitable cell culture medium. In some embodiments, genetically modified host cells disclosed herein are grown at about 20 °C to about 37 °C in a suitable cell culture medium. In some embodiments, genetically modified host cells disclosed herein are grown at about 20 °C, about 21 °C, about 22 °C, about 23 °C, about 24 °C, about 25 °C, about 26 °C, about 27 °C, about 28 °C, about 29 °C, about 30 °C, about 31 °C, about 32 °C, about 33 °C, about 34 °C, about 35 °C, about 36 °C, or about 37 °C in a suitable cell culture medium. In some embodiments, the pH ranges for fermentation are between about pH 3.0 to about pH 9.0 (such as about pH 3.0, about pH 3.5, about pH 4.0, about pH 4.5, about pH 5.0, about pH 5.5, about pH 6.0, about pH 6.5, about pH 7.0, about pH 7.5, about pH 8.0, about

pH 8.5, about pH 6.0 to about pH 8.0 or about 6.5 to about 7.0). In some embodiments, the pH ranges for fermentation are between about pH 4.5 to about pH 5.5. In some embodiments, the pH ranges for fermentation are between about pH 4.0 to about pH 6.0. In some embodiments, the pH ranges for fermentation are between about pH 3.0 to about pH 6.0. In some embodiments, the pH ranges for fermentation are between about pH 3.0 to about pH 5.5. In some embodiments, the pH ranges for fermentation are between about pH 3.0 to about pH 5.0. In some embodiments, the dissolved oxygen is between about 0% to about 10%, about 0% to about 20%, about 0% to about 30%, about 0% to about 40%, about 0% to about 50%, about 0% to about 60%, about 0% to about 70%, about 0% to about 80%, about 0% to about 90%, about 5% to about 10%, about 5% to about 20%, about 5% to about 30%, about 5% to about 40%, about 5% to about 50%, about 5% to about 60%, about 5% to about 70%, about 5% to about 80%, about 5% to about 90%, about 10% to about 20%, about 10% to about 30%, about 10% to about 40% or about 10% to about 50%. In some embodiments, the CO_2 level is between about 0.04% to about 0.1% CO_2, about 0.04% to about 1% CO_2, about 0.04% to about 5% CO_2, about 0.04% to about 10% CO_2, about 0.04% to about 20% CO_2, about 0.04% to about 30% CO_2, about 0.04% to about 40% CO_2, about 0.04% to about 50% CO_2, about 0.04% to about 60% CO_2, about 0.04% to about 70% CO_2, about 0.1% to about 5% CO_2, about 0.1% to about 10% CO_2, about 0.1% to about 20% CO_2, about 0.1% to about 30% CO_2, about 0.1% to about 40% CO_2, about 0.1% to about 50% CO_2, about 1% to about 5% CO_2, about 1% to about 10% CO_2, about 1% to about 20% CO_2, about 1% to about 30% CO_2, about 1% to about 40% CO_2, about 1% to about 50% CO_2, about 5% to about 10% CO_2, about 10% to about 20% CO_2, about 10% to about 30% CO_2, about 10% to about 40% CO_2, about 10% to about 50% CO_2, about 10% to about 60% CO_2, about 10% to about 70% CO_2, about 10% to about 80% CO_2, about 50% to about 60% CO_2, about 50% to about 70% CO_2, or about 50% to about 80% CO_2. genetically modified host cells disclosed herein disclosed herein can be grown under aerobic, anoxic, microaerobic, or anaerobic conditions based on the requirements of the cells.

[00709] Standard culture conditions and modes of fermentation, such as batch, fed-batch, or continuous fermentation that can be used are described in International Publication No. WO 2009/076676, U.S. Patent Application No. 12/335,071 (U.S. Publ. No. 2009/0203102), WO 2010/003007, US Publ. No. 2010/0048964, WO 2009/132220, US Publ. No. 2010/0003716, the contents of each of which are incorporated by reference herein in their entireties. Batch and Fed- Batch fermentations are common and well known in the

art and examples can be found in Brock, Biotechnology: A Textbook of Industrial Microbiology, Second Edition (1989) Sinauer Associates, Inc.

Production and Recovery of Produced Cannabinoids, Cannabinoid Precursors, Cannabinoid Derivatives or Cannabinoid Precursor Derivatives

[00710] In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by genetically modified host cells of the disclosure in a recoverable amount of from about 1 mg/L culture medium to about 1 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 mg/L culture medium to about 500 mg/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 mg/L culture medium to about 100 mg/L culture medium. For example, in some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 mg/L culture medium to about 5 mg/L culture medium, from about 5 mg/L culture medium to about 10 mg/L culture medium, from about 10 mg/L culture medium to about 25 mg/L culture medium, from about 25 mg/L culture medium to about 50 mg/L culture medium, from about 50 mg/L culture medium to about 75 mg/L culture medium, or from about 75 mg/L culture medium to about 100 mg/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 100 mg/L culture medium to about 150 mg/L culture medium, from about 150 mg/L culture medium to about 200 mg/L culture medium, from about 200 mg/L culture medium to about 250 mg/L culture medium, from about 250 mg/L culture medium to about 500 mg/L culture medium, from about 500 mg/L culture medium to about 750 mg/L culture medium, or from about 750 mg/L culture medium to about 1 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about from about 50 mg/L culture medium to about 100 mg/L culture medium, 50 mg/L culture medium to about 150

mg/L culture medium, from about 50 mg/L culture medium to about 200 mg/L culture medium, from about 50 mg/L culture medium to about 250 mg/L culture medium, from about 50 mg/L culture medium to about 500 mg/L culture medium, or from about 50 mg/L culture medium to about 750 mg/L culture medium.

[00711] In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 100 mg/L culture medium to about 500 mg/L culture medium, or more than 500 mg/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 500 mg/L culture medium to about 1 g/L culture medium, or more than 1 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 g/L culture medium to about 10 g/L culture medium, or more than 10 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 10 g/L culture medium to about 100 g/L culture medium, or more than 100 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 g/L culture medium to about 20 g/L culture medium, or more than 20 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 g/L culture medium to about 30 g/L culture medium, or more than 30 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 g/L culture medium to about 40 g/L culture medium, or more than 40 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 g/L culture medium to about 50 g/L culture medium, or more than 50 g/L culture medium. In some embodiments, a method of the present disclosure provides for

production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 g/L culture medium to about 60 g/L culture medium, or more than 60 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 g/L culture medium to about 70 g/L culture medium, or more than 70 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 g/L culture medium to about 80 g/L culture medium, or more than 80 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 g/L culture medium to about 90 g/L culture medium, or more than 90 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 10 g/L culture medium to about 20 g/L culture medium, or more than 20 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 10 g/L culture medium to about 30 g/L culture medium, or more than 30 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 10 g/L culture medium to about 40 g/L culture medium, or more than 40 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 10 g/L culture medium to about 50 g/L culture medium, or more than 50 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 10 g/L culture medium to about 60 g/L culture medium, or more than 60 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 10 g/L culture

medium to about 70 g/L culture medium, or more than 70 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 10 g/L culture medium to about 80 g/L culture medium, or more than 80 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 10 g/L culture medium to about 90 g/L culture medium, or more than 90 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 50 g/L culture medium to about 100 g/L culture medium, or more than 100 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 50 g/L culture medium to about 60 g/L culture medium, or more than 60 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 50 g/L culture medium to about 70 g/L culture medium, or more than 70 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 50 g/L culture medium to about 80 g/L culture medium, or more than 80 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 50 g/L culture medium to about 90 g/L culture medium, or more than 90 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 20 g/L culture medium to about 100 g/L culture medium, or more than 100 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 20 g/L culture medium to about 30 g/L culture medium, or more than 30 g/L culture medium. In some embodiments, a method of the present disclosure provides for

production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 20 g/L culture medium to about 40 g/L culture medium, or more than 40 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 20 g/L culture medium to about 50 g/L culture medium, or more than 50 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 20 g/L culture medium to about 60 g/L culture medium, or more than 60 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 20 g/L culture medium to about 70 g/L culture medium, or more than 70 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 20 g/L culture medium to about 80 g/L culture medium, or more than 80 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 20 g/L culture medium to about 90 g/L culture medium, or more than 90 g/L culture medium.

[00712] In some embodiments, the genetically modified host cell disclosed herein is cultured in a liquid medium comprising a precursor acid to generate acyl-CoA compounds or acyl-CoA compound derivatives. Suitable precursor acids may include, but are not limited to, carboxylic acids.

[00713] In some embodiments, a method of producing a cannabinoid, a cannabinoid derivative, cannabinoid precursor, cannabinoid precursor derivative, or cannabinoid precursor derivative may involve culturing a genetically modified yeast cell of the present disclosure under conditions that favor production of the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative; where the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative is produced by the genetically modified yeast cell and is present in the culture medium (e.g., a liquid culture medium) in which the genetically modified yeast cell is cultured. In some embodiments, the culture medium in which the genetically modified yeast cell is cultured

comprises a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in an amount of from 1 ng/L to 1 g/L (e.g., from 1 ng/L to 50 ng/L, from 50 ng/L to 100 ng/L, from 100 ng/L to 500 ng/L, from 500 ng/L to 1 μg/L, from 1 μg/L to 50 μg/L, from 50 μg/L to 100 μg/L, from 100 μg/L to 500 μg/L, from 500 μg/L to 1 mg/L, from 1 mg/L to 50 mg/L, from 50 mg/L to 100 mg/L, from 100 mg/L to 500 mg/L, or from 500 mg/L to 1 g/L). In some embodiments, the culture medium in which the genetically modified yeast cell is cultured comprises a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in an amount more than 1 g/L.

[00714] In some embodiments, a method of producing a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative may involve culturing a genetically modified yeast cell of the present disclosure under conditions that favor fermentation of a sugar, and under conditions that favor production of the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative; wherein the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative is produced by the genetically modified yeast cell and is present in alcohol produced by the genetically modified yeast cell. The present disclosure provides an alcoholic beverage produced by the genetically modified yeast cell, where the alcoholic beverage comprises the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative produced by the genetically modified yeast cell. Alcoholic beverages may include beer, wine, and distilled alcoholic beverages. In some embodiments, an alcoholic beverage of the present disclosure comprises a cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative in an amount of from 1 ng/L to 1 g/L (e.g., from 1 ng/L to 50 ng/L, from 50 ng/L to 100 ng/L, from 100 ng/L to 500 ng/L, from 500 ng/L to 1 μg/L, from 1 μg/L to 50 μg/L, from 50 μg/L to 100 μg/L, from 100 μg/L to 500 μg/L, from 500 μg/L to 1 mg/L, from 1 mg/L to 50 mg/L, from 50 mg/L to 100 mg/L, from 100 mg/L to 500 mg/L, or from 500 mg/L to 1 g/L). In some embodiments, an alcoholic beverage of the present disclosure comprises a cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative in an amount more than 1 g/L

[00715] In some embodiments, a method of the present disclosure provides for increased production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative. In certain such embodiments, culturing of the genetically modified host cell disclosed herein in a suitable medium provides for synthesis of

the cannabinoid, the cannabinoid derivative, the cannabinoid precursor, or the cannabinoid precursor derivative in an increased amount compared to a non-genetically modified host cell cultured under similar conditions. The production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by the genetically modified host cells disclosed herein may be increased by about 5% to about 1,000,000 folds compared to a non-genetically modified host cell cultured under similar conditions. The production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by the genetically modified host cells disclosed herein may be increased by about 10% to about 1,000,000 folds (e.g., about 50% to about 1,000,000 folds, about 1 to about 500,000 folds, about 1 to about 50,000 folds, about 1 to about 5,000 folds, about 1 to about 1,000 folds, about 1 to about 500 folds, about 1 to about 100 folds, about 1 to about 50 folds, about 5 to about 100,000 folds, about 5 to about 10,000 folds, about 5 to about 1,000 folds, about 5 to about 500 folds, about 5 to about 100 folds, about 10 to about 50,000 folds, about 50 to about 10,000 folds, about 100 to about 5,000 folds, about 200 to about 1,000 folds, about 50 to about 500 folds, or about 50 to about 200 folds) compared to the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by non-genetically modified host cells cultured under similar conditions. The production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by genetically modified host cells disclosed herein may also be increased by at least about any of 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 1 fold, 2 folds, 5 folds, 10 folds, 20 folds, 50 folds, 100 folds, 200 folds, 500 folds, 1000 folds, 2000 folds, 5000 folds, 10,000 folds, 20,000 folds, 50,000 folds, 100,000 folds, 200,000 folds, 500,000 folds, or 1,000,000 folds compared to the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by non-genetically modified host cells cultured under similar conditions.

[00716] In some embodiments, the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by genetically modified host cells of the disclosure may also be increased by at least about any of 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% compared to the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by non-genetically modified host cells cultured under similar conditions. In some embodiments, the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by genetically modified host cells disclosed

herein may also be increased by at least about any of 1-20%, 2-20%, 5-20%, 10-20%, 15-20%, 1-15%, 1-10%, 2-15%, 2-10%, 5-15%, 10-15%, 1-50%, 10-50%, 20-50%, 30-50%, 40-50%, 50-100%, 50-60%, 50-70%, 50-80%, or 50-90% compared to the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by non-genetically modified host cells cultured under similar conditions.

[00717] In some embodiments, production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by genetically modified host cells of the disclosure is determined by LC-MS analysis. In certain such embodiments, each cannabinoid, cannabinoid precursor, cannabinoid derivative, or cannabinoid precursor derivative is identified by retention time, determined from an authentic standard, and multiple reaction monitoring (MRM) transition.

[00718] In some embodiments, the genetically modified host cell of the disclosure is yeast cell. In certain such embodiments, the genetically modified host cell disclosed herein is cultured in a bioreactor. In some embodiments, the genetically modified host cell is cultured in a suitable medium supplemented with hexanoic acid, a carboxylic acid other than hexanoic acid, olivetolic acid, or an olivetolic acid derivative.

[00719] In some embodiments, the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative is recovered from a cell lysate, e.g., by lysing the genetically modified host cell disclosed herein and recovering the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative from the lysate. In other cases, the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative is recovered from the culture medium in which the genetically modified host cell disclosed herein is cultured. In other cases, the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative is recovered from both the cell lysate and the culture medium.

[00720] In some embodiments, the recovered cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative is then purified. In some embodiments, whole-cell broth from cultures comprising genetically modified host cells of the disclosure may be extracted with a suitable organic solvent to afford cannabinoids, cannabinoid precursors, cannabinoid derivatives, or cannabinoid precursor derivatives. Suitable organic solvents include, but are not limited to, hexane, heptane, ethyl acetate, petroleum ether, and di-ethyl ether, chloroform, and ethyl acetate. In some embodiments, the suitable organic solvent comprises hexane. In some embodiments, the suitable organic solvent may be added to the whole-cell broth from fermentations comprising genetically

modified host cells of the disclosure at a 10:1 ratio (10 parts whole-cell broth – 1 part organic solvent) and stirred for 30 minutes. In certain such embodiments, the organic fraction may be separated and extracted twice with an equal volume of acidic water (pH 2.5). The organic layer may then be separated and dried in a concentrator (rotary evaporator or thin film evaporator under reduced pressure) to obtain crude cannabinoid, cannabinoid precursor, cannabinoid derivative, or cannabinoid precursor derivative crystals. In certain such embodiments, the crude crystals may be heated to 105 °C for 15 minutes followed by 145 °C for 55 minutes to decarboxylate a crude cannabinoid or cannabinoid derivative. In certain such embodiments, the crude crystalline product may be re-dissolved and recrystallized in a suitable solvent (e.g., n-pentane) and filtered to remove any insoluble material. In certain such embodiments, the solvent may then be removed e.g. by rotary evaporation, to produce pure crystalline product.

[00721] In some embodiments, the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative is pure, e.g., at least about 40% pure, at least about 50% pure, at least about 60% pure, at least about 70% pure, at least about 80% pure, at least about 90% pure, at least about 95% pure, at least about 98%, or more than 98% pure, where "pure" in the context of a cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative may refer to a cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative that is free from other cannabinoids, cannabinoid derivatives, cannabinoid precursors, cannabinoid precursor derivatives macromolecules, contaminants, etc.

Cell-Free Methods of Generating Cannabinoids, Cannabinoid Precursors, Olivetolic Acid Derivatives, Olivetolic Acid, Cannabinoid Derivatives, or Cannabinoid Precursor Derivatives

[00722] The methods of the disclosure may involve cell-free production of cannabinoids, cannabinoid precursors, cannabinoid precursor derivatives, or cannabinoid derivatives using one or more polypeptides disclosed herein expressed or overexpressed by a genetically modified host cell of the disclosure. In some embodiments, one or more polypeptides disclosed herein expressed or overexpressed by a genetically modified host cell of the disclosure are used in a cell-free system for the production of cannabinoids, cannabinoid precursors, cannabinoid precursor derivatives, or cannabinoid derivatives. In certain such embodiments, appropriate starting materials for use in producing cannabinoids, cannabinoid precursors, cannabinoid precursor derivatives, or cannabinoid derivatives may

be mixed together with one or more polypeptides disclosed herein expressed or overexpressed by a genetically modified host cell of the disclosure in a suitable reaction vessel to effect the reaction. The one or more polypeptides disclosed herein expressed or overexpressed by a genetically modified host cell of the disclosure may be used in combination to effect a complete synthesis of a cannabinoid, cannabinoid precursor, cannabinoid precursor derivative, or cannabinoid derivative from the appropriate starting materials. In some embodiments, the cannabinoid, cannabinoid precursor, the cannabinoid precursor derivative, or cannabinoid derivative is recovered from a cell-free reaction mixture comprising one or more of the polypeptides disclosed herein.

[00723] In some embodiments, the recovered cannabinoids, cannabinoid precursors, cannabinoid precursor derivatives, or cannabinoid derivatives are then purified. In certain such embodiments, a cell-free reaction mixture comprising one or more of the polypeptides disclosed herein may be extracted with a suitable organic solvent to afford cannabinoids, cannabinoid precursors, cannabinoid derivatives, or cannabinoid precursor derivatives. Suitable organic solvents include, but are not limited to, hexane, heptane, ethyl acetate, petroleum ether, and di-ethyl ether, chloroform, and ethyl acetate. In some embodiments, the suitable organic solvent comprises hexane. In some embodiments, the suitable organic solvent may be added to the cell-free reaction mixture comprising one or more of the polypeptides disclosed herein at a 10:1 ratio (10 parts reaction mixture – 1 part organic solvent) and stirred for 30 minutes. In certain such embodiments, the organic fraction may be separated and extracted twice with an equal volume of acidic water (pH 2.5). The organic layer may then be separated and dried in a concentrator (rotary evaporator or thin film evaporator under reduced pressure) to obtain crude cannabinoid, cannabinoid precursor, cannabinoid derivative, or cannabinoid precursor derivative crystals. In certain such embodiments, the crude crystals may be heated to 105 °C for 15 minutes followed by 145 °C for 55 minutes to decarboxylate a crude cannabinoid or cannabinoid derivative. In certain such embodiments, the crude crystalline product may be re-dissolved and recrystallized in a suitable solvent (e.g., n-pentane) and filtered to remove any insoluble material. In certain such embodiments, the solvent may then be removed e.g. by rotary evaporation, to produce pure crystalline product.

[00724] In some embodiments, a prenyl group acceptor molecule, a prenyl group donor molecule and a GOT polypeptide may be mixed together in a suitable reaction vessel to effect the reaction. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten

times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In some embodiments, the prenyl group acceptor molecule is olivetolic acid or a derivative thereof. In some embodiments, the prenyl group donor molecule is GPP or a derivative thereof. In some embodiments, the reaction produces cannabigerolic acid or a derivative thereof.

[00725] In some embodiments, cell-free production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by one or more polypeptides disclosed herein expressed or overexpressed by a genetically modified host cell of the disclosure is determined by LC-MS analysis. In certain such embodiments, each cannabinoid, cannabinoid precursor, cannabinoid derivative, or cannabinoid precursor derivative is identified by retention time, determined from an authentic standard, and multiple reaction monitoring (MRM) transition.

EXAMPLES OF NON-LIMITING EMBODIMENTS OF THE DISCLOSURE

[00726] Embodiments, of the present subject matter disclosed herein may be beneficial alone or in combination, with one or more other embodiments. Without limiting the foregoing description, certain non-limiting embodiments of the disclosure, numbered I-1 to I-54, II-1 to II-55, and III-1 to III-81 are provided below. As will be apparent to those of skill in the art upon reading this disclosure, each of the individually numbered embodiments may be used or combined with any of the preceding or following individually numbered embodiments. This is intended to provide support for all such combinations of embodiments and is not limited to combinations of embodiments explicitly provided below. Some embodiments of the disclosure are of Embodiment I:

[00727] Embodiment I-1. A genetically modified host cell that produces a cannabinoid compound or a cannabinoid precursor, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode one or more polypeptides that generate hexanoyl-CoA or derivatives of hexanoyl-CoA; b) one or more heterologous nucleic acids that encode one or more polypeptides that generate geranyl pyrophosphate; c) one or more heterologous nucleic acids that encode one or more polypeptides that generate malonyl-CoA; d) one or more heterologous nucleic acids that

encode a fusion TKS/OAC polypeptide that condenses hexanoyl-CoA or its derivatives and malonyl-CoA to generate olivetolic acid or derivatives of olivetolic acid; e) one or more heterologous nucleic acids that encode a truncated geranyl pyrophosphate:olivetolic acid geranyltransferase (GOT) polypeptide or an NphB polypeptide; and f) one or more heterologous nucleic acids that encode a cannabinoid synthase polypeptide, wherein culturing of the genetically modified host cell in a suitable medium provides for synthesis of the cannabinoid compound or the cannabinoid precursor in a recoverable amount.

[00728] Embodiment I-2. The genetically modified host cell of Embodiment I-1, wherein the host cell is a eukaryotic cell.

[00729] Embodiment I-3. The genetically modified host cell of Embodiment I-2, wherein the host cell is a yeast cell.

[00730] Embodiment I-4. The genetically modified host cell of Embodiment I-3, wherein the host cell is *Saccharomyces cerevisiae*.

[00731] Embodiment I-5. The genetically modified host cell of Embodiment I-4, wherein the host cell is a protease-deficient strain of *Saccharomyces cerevisiae*.

[00732] Embodiment I-6. The genetically modified host cell of Embodiment I-2, wherein the host cell is a plant cell.

[00733] Embodiment I-7. The genetically modified host cell of Embodiment I-1, wherein the host cell is a prokaryotic cell.

[00734] Embodiment I-8. The genetically modified host cell of Embodiment I-1, wherein the one or more polypeptides that generate hexanoyl-CoA or a hexanoyl-CoA derivative is hexanoyl-CoA synthetase (HCS) polypeptide, and wherein the medium comprises hexanoate.

[00735] Embodiment I-9. The genetically modified host cell of Embodiment I-6, wherein the HCS polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, or SEQ ID NO:4.

[00736] Embodiment I-10. The genetically modified host cell of Embodiment I-1, wherein the one or more polypeptides that generate geranyl pyrophosphate comprise geranyl pyrophosphate synthetase (GPPS) polypeptide.

[00737] Embodiment I-11. The genetically modified host cell of Embodiment I-11, wherein the GPPS polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at

least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to one of the amino acid sequences set forth in SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8.

[00738] Embodiment I-12. The genetically modified host cell of Embodiment I-10, wherein the GPPS polypeptide is a dominant negative variant that reduces the ability of an endogenous GPPS polypeptide to function as a farnesyl pyrophosphate synthetase (FPPS) polypeptide.

[00739] Embodiment I-13. The genetically modified host cell of Embodiment I-10, wherein the GPPS polypeptide comprises a K197G amino acid substitution.

[00740] Embodiment I-14. The genetically modified host cell of Embodiment I-10, wherein the GPPS polypeptide is a heterodimeric protein comprising a GPPS large subunit polypeptide and a GPPS small subunit polypeptide, or a homodimeric or monomeric GPPS polypeptide.

[00741] Embodiment I-15. The genetically modified host cell of Embodiment I-1, wherein the one or more polypeptides that generate malonyl-CoA comprises acetyl-CoA carboxylase-1 (ACC1) polypeptide.

[00742] Embodiment I-16. The genetically modified host cell of Embodiment I-15, wherein the ACC1 polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:9.

[00743] Embodiment I-17. The genetically modified host cell of Embodiment I-1, wherein the one or more polypeptides that generate hexanoyl-CoA comprise an MCT1 polypeptide, a PaaH1 polypeptide, a Crt polypeptide, a Ter polypeptide, and a BktB polypeptide.

[00744] Embodiment I-18. The genetically modified host cell of Embodiment I-1, wherein the one or more polypeptides that generate hexanoyl-CoA comprise a MCT1 polypeptide, a PhaB polypeptide, a PhaJ polypeptide, a Ter polypeptide, and a BktB polypeptide.

[00745] Embodiment I-19. The genetically modified host cell of Embodiment I-17, wherein: i) the PaaH1 polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:18 or SEQ ID NO:46; ii) the Crt polypeptide

comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:19 or SEQ ID NO:48; iii) the Ter polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:20 or SEQ ID NO:50; and iv) the BktB polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:21 or SEQ ID NO:44.

[00746] Embodiment I-20. The genetically modified host cell of Embodiment I-18, wherein: i) the PhaB polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:94; ii) the PhaJ polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:96; iii) the Ter polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:20 or SEQ ID NO:50; and iv) the BktB polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:21 or SEQ ID NO:44.

[00747] Embodiment I-21. The genetically modified host cell of Embodiment I-1, wherein the host cell is genetically modified with a heterologous nucleic acid encoding one or more polypeptides that modulate NADH redox balance.

[00748] Embodiment I-22. The genetically modified host cell of Embodiment I-1, wherein the host cell is genetically modified with one or more of: i) one or more heterologous nucleic acids encoding a HMG-CoA synthase polypeptide; ii) one or more heterologous nucleic acids encoding a 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR) polypeptide; iii) one or more heterologous nucleic acids encoding an MK polypeptide; and

iv) one or more heterologous nucleic acids encoding an isopentenyl diphosphate isomerase (IDI) polypeptide.

[00749] Embodiment I-23. The genetically modified host cell of Embodiment I-22, wherein the host cell is genetically modified to overexpress a heterologous IDI polypeptide.

[00750] Embodiment I-24. The genetically modified host cell of Embodiment I-23, wherein the IDI polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:58.

[00751] Embodiment I-25. The genetically modified host cell of Embodiment I-22, wherein the host cell is genetically modified to overexpress a truncated HMGR (tHMGR) polypeptide.

[00752] Embodiment I-26. The genetically modified host cell of Embodiment I-25, wherein the tHMGR polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:17.

[00753] Embodiment I-27. The genetically modified host cell of Embodiment I-22, wherein the HMGR polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:22.

[00754] Embodiment I-28. The genetically modified host cell of Embodiment I-22, wherein the HMGS polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the MvaS polypeptide amino acid sequence set forth in SEQ ID NO:23, or wherein the HMGS polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the ERG13 polypeptide amino acid sequence set forth in SEQ ID NO:24.

[00755] Embodiment I-29. The genetically modified host cell of Embodiment I-22, wherein the MK polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least

95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the ERG12 polypeptide amino acid sequence set forth in SEQ ID NO:64.

[00756] Embodiment I-30. The genetically modified host cell of Embodiment I-1, wherein the host cell is genetically modified with a heterologous nucleic acid encoding one or more polypeptides that condense two molecules of acetyl-CoA to generate acetoacetyl-CoA.

[00757] Embodiment I-31. The genetically modified host cell of Embodiment I-30, wherein the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide.

[00758] Embodiment I-32. The genetically modified host cell of Embodiment I-31, wherein the acetoacetyl-CoA thiolase polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:25.

[00759] Embodiment I-33. The genetically modified host cell of Embodiment I-1, wherein the host cell is genetically modified with a heterologous nucleic acid encoding one or more polypeptides that condense one molecule of acetyl-CoA and one molecule of malonyl-CoA to generate acetoacetyl-CoA.

[00760] Embodiment I-34. The genetically modified host cell of any one of Embodiments I-1 to I-33, wherein at least one of the one or more heterologous nucleic acids is integrated into the chromosome of the host cell.

[00761] Embodiment I-35. The genetically modified host cell of any one of Embodiments I-1 to I-33, wherein at least one of the one or more heterologous nucleic acids is maintained extrachromosomally.

[00762] Embodiment I-36. The genetically modified host cell of any one of Embodiments I-1 to I-33, wherein two or more of the one or more heterologous nucleic acids are present in a single expression vector.

[00763] Embodiment I-37. The genetically modified host cell of any one of Embodiments I-1 to I-36, wherein the cannabinoid compound is cannabigerolic acid.

[00764] Embodiment I-38. The genetically modified host cell of Embodiment I-1, wherein the genetically modified host cell is genetically modified with a heterologous nucleic acid encoding a Δ^9-THCA synthase polypeptide.

[00765] Embodiment I-39. The genetically modified host cell of Embodiment I-38, wherein the THCA synthase polypeptide comprises an amino acid having at least 50%

(at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to SEQ ID NO:14 or SEQ ID NO:15.

[00766] Embodiment I-40. The genetically modified host cell of Embodiment I-1, wherein the genetically modified host cell is genetically modified with a heterologous nucleic acid encoding a CBDA synthase polypeptide.

[00767] Embodiment I-41. The genetically modified host cell of Embodiment I-40, wherein the CBDA synthase polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to SEQ ID NO:88 or SEQ ID NO:16.

[00768] Embodiment I-42. The genetically modified host cell of any one of Embodiments I-1 to I-41, wherein at least one of the heterologous nucleic acids is operably linked to an inducible promoter.

[00769] Embodiment I-43. The genetically modified host cell of any one of Embodiments I-1 to I-41, wherein at least one of the heterologous nucleic acids is operably linked to a constitutive promoter.

[00770] Embodiment I-44. The genetically modified host cell of Embodiment I-1, wherein the cannabinoid compound is cannabichromenic acid, cannabigerolic acid, Δ^9-tetrahydrocannabinolic acid, cannabidiolic acid, Δ^9-tetrahydrocannabinol, cannabidiol, or cannabichromene.

[00771] Embodiment I-45. A method of synthesizing a cannabinoid compound or cannabinoid precursor in a host cell, the method comprising: a) culturing a host cell of any one of Embodiments I-1 to I-40 in a suitable medium; and b) recovering the produced cannabinoid compound or cannabinoid precursor.

[00772] Embodiment I-46. The method of Embodiment I-45, wherein the medium comprises a fermentable sugar.

[00773] Embodiment I-47. The method of Embodiment I-45, wherein the medium comprises a pretreated cellulosic feedstock.

[00774] Embodiment I-48. A genetically modified host cell that produces an olivetolic acid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode an acyl-activating enzyme (AAE) polypeptide; and b) one or more heterologous nucleic acids that encode a TKS/OAC fusion polypeptide.

[00775] Embodiment I-49. The genetically modified host cell of Embodiment I-48, wherein the AAE polypeptide comprises an amino acid sequence having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to an AAE amino acid sequence set forth in SEQ ID NO:90 or SEQ ID NO:91.

[00776] Embodiment I-50. A genetically modified host cell that produces a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode an acyl-activating enzyme (AAE) polypeptide; b) one or more heterologous nucleic acids that encode a TKS/OAC fusion polypeptide; and c) one or more heterologous nucleic acids that encode a GOT polypeptide or an NphB polypeptide.

[00777] Embodiment I-51. The genetically modified host cell of Embodiment I-50, wherein the genetically modified host cell is genetically modified with a heterologous nucleic acid encoding a THCAS polypeptide.

[00778] Embodiment I-52. The genetically modified host cell of Embodiment I-50, wherein the genetically modified host cell is genetically modified with a heterologous nucleic acid encoding a CBDAS polypeptide.

[00779] Embodiment I-53. A method of producing an olivetolic acid derivative, the method comprising culturing the genetically modified host cell of Embodiment I-48 or Embodiment I-49 in a culture medium comprising a carboxylic acid.

[00780] Embodiment I-54. A method of producing a cannabinoid derivative, the method comprising culturing the genetically modified host cell of any one of Embodiments I-50 to I-52 in a culture medium comprising a carboxylic acid.

[00781] Some embodiments of the disclosure are of Embodiment II:

[00782] Embodiment II-1. A genetically modified host cell that produces a cannabinoid compound, a cannabinoid derivative, or a cannabinoid precursor, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode one or more polypeptides that generate hexanoyl-CoA or derivatives of hexanoyl-CoA; b) one or more heterologous nucleic acids that encode one or more polypeptides that generate geranyl pyrophosphate; c) one or more heterologous nucleic acids that encode one or more polypeptides that generate malonyl-CoA; d) one or more heterologous nucleic acids that encode a TKS polypeptide and an OAC polypeptide, or a fusion TKS and OAC polypeptide, that converts hexanoyl-CoA or its derivatives and malonyl-CoA to olivetolic acid or derivatives of olivetolic acid; e) one or more heterologous

nucleic acids that encode a geranyl pyrophosphate:olivetolic acid geranyltransferase (GOT) polypeptide or an NphB polypeptide; and f) one or more heterologous nucleic acids that encode a cannabinoid synthase polypeptide, wherein culturing of the genetically modified host cell in a suitable medium provides for synthesis of the cannabinoid compound, cannabinoid derivative, or the cannabinoid precursor in a recoverable amount.

[00783] Embodiment II-2. The genetically modified host cell of Embodiment II-1, wherein the host cell is a eukaryotic cell.

[00784] Embodiment II-3. The genetically modified host cell of Embodiment II-2, wherein the host cell is a yeast cell.

[00785] Embodiment II-4. The genetically modified host cell of Embodiment II-3, wherein the host cell is *Saccharomyces cerevisiae*.

[00786] Embodiment II-5. The genetically modified host cell of Embodiment II-4, wherein the host cell is a protease-deficient strain of *Saccharomyces cerevisiae*.

[00787] Embodiment II-6. The genetically modified host cell of Embodiment II-2, wherein the host cell is a plant cell.

[00788] Embodiment II-7. The genetically modified host cell of Embodiment II-1, wherein the host cell is a prokaryotic cell.

[00789] Embodiment II-8. The genetically modified host cell of Embodiment II-1, wherein the one or more polypeptides that generate hexanoyl-CoA or a hexanoyl-CoA derivative is a hexanoyl-CoA synthetase (HCS) polypeptide, and wherein the medium comprises hexanoate.

[00790] Embodiment II-9. The genetically modified host cell of Embodiment II-8, wherein the HCS polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, or SEQ ID NO:4.

[00791] Embodiment II-10. The genetically modified host cell of Embodiment II-1, wherein the one or more polypeptides that generate geranyl pyrophosphate comprise a geranyl pyrophosphate synthetase (GPPS) polypeptide.

[00792] Embodiment II-11. The genetically modified host cell of Embodiment II-11, wherein the GPPS polypeptide comprises an amino acid having at least 50% amino acid sequence identity to one of the amino acid sequences set forth in SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8.

[00793] Embodiment II-12. The genetically modified host cell of Embodiment II-10, wherein the GPPS polypeptide is a dominant negative variant that reduces the ability of

an endogenous GPPS polypeptide to function as a farnesyl pyrophosphate synthetase (FPPS) polypeptide.

[00794] Embodiment II-13. The genetically modified host cell of Embodiment II-10, wherein the GPPS polypeptide comprises a K197G amino acid substitution.

[00795] Embodiment II-14. The genetically modified host cell of Embodiment II-10, wherein the GPPS polypeptide is a heterodimeric protein comprising a GPPS large subunit polypeptide and a GPPS small subunit polypeptide, or a homodimeric or monomeric GPPS polypeptide.

[00796] Embodiment II-15. The genetically modified host cell of Embodiment II-1, wherein the one or more polypeptides that generate malonyl-CoA comprises an acetyl-CoA carboxylase-1 (ACC1) polypeptide.

[00797] Embodiment II-16. The genetically modified host cell of Embodiment II-15, wherein the ACC1 polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:9.

[00798] Embodiment II-17. The genetically modified host cell of Embodiment II-1, wherein the one or more polypeptides that generate hexanoyl-CoA comprise a MCT1 polypeptide, a PaaH1 polypeptide, a Crt polypeptide, a Ter polypeptide, and a BktB polypeptide.

[00799] Embodiment II-18. The genetically modified host cell of Embodiment II-1, wherein the one or more polypeptides that generate hexanoyl-CoA comprise a MCT1 polypeptide, a PhaB polypeptide, a PhaJ polypeptide, a Ter polypeptide, and a BktB polypeptide.

[00800] Embodiment II-19 The genetically modified host cell of Embodiment II-17, wherein: i) the PaaH1 polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:18 or SEQ ID NO:46; ii) the Crt polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:19 or SEQ ID NO:48; iii) the Ter polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:20 or SEQ ID NO:50; and iv) the BktB polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:21 or SEQ ID NO:44.

[00801] Embodiment II-20. The genetically modified host cell of Embodiment II-18, wherein: i) the PhaB polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:94; ii) the PhaJ

polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:96; iii) the Ter polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:20 or SEQ ID NO:50; and iv) the BktB polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:21 or SEQ ID NO:44.

[00802] Embodiment II-21. The genetically modified host cell of Embodiment II-1, wherein the host cell is genetically modified with a heterologous nucleic acid encoding one or more polypeptides that modulate NADH redox balance.

[00803] Embodiment II-22. The genetically modified host cell of Embodiment II-1, wherein the host cell is genetically modified with one or more of: i) one or more heterologous nucleic acids that encode an HMG-CoA synthase polypeptide; ii) one or more heterologous nucleic acids that encode a 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR) polypeptide; iii) one or more heterologous nucleic acids that encode an MK polypeptide; and iv) one or more heterologous nucleic acids that encode an isopentenyl diphosphate isomerase (IDI) polypeptide.

[00804] Embodiment II-23. The genetically modified host cell of Embodiment II-22, wherein the host cell is genetically modified to overexpress a heterologous IDI polypeptide.

[00805] Embodiment II-24. The genetically modified host cell of Embodiment II-23, wherein the IDI polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:58.

[00806] Embodiment II-25. The genetically modified host cell of Embodiment II-22, wherein the host cell is genetically modified to overexpress a truncated HMGR (tHMGR) polypeptide.

[00807] Embodiment II-26. The genetically modified host cell of Embodiment II-25, wherein the tHMGR polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:17.

[00808] Embodiment II-27. The genetically modified host cell of Embodiment II-22, wherein the HMGR polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:22.

[00809] Embodiment II-28. The genetically modified host cell of Embodiment II-22, wherein the HMGS polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the MvaS polypeptide amino acid sequence set forth in SEQ ID NO:23,

or wherein the HMGS polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the ERG13 polypeptide amino acid sequence set forth in SEQ ID NO:24.

[00810] Embodiment II-29. The genetically modified host cell of Embodiment II-22, wherein the MK polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the ERG12 polypeptide amino acid sequence set forth in SEQ ID NO:64.

[00811] Embodiment II-30. The genetically modified host cell of Embodiment II-1, wherein the host cell is genetically modified with a heterologous nucleic acid encoding one or more polypeptides that condense two molecules of acetyl-CoA to generate acetoacetyl-CoA.

[00812] Embodiment II-31. The genetically modified host cell of Embodiment II-30, wherein the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide.

[00813] Embodiment II-32. The genetically modified host cell of Embodiment II-31, wherein the acetoacetyl-CoA thiolase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25.

[00814] Embodiment II-33. The genetically modified host cell of Embodiment II-1, wherein the host cell is genetically modified with a heterologous nucleic acid encoding one or more polypeptides that condense one molecule of acetyl-CoA and one molecule of malonyl-CoA to generate acetoacetyl-CoA.

[00815] Embodiment II-34. The genetically modified host cell of any one of Embodiments II-1 to II-33, wherein at least one of the one or more heterologous nucleic acids is integrated into the chromosome of the host cell.

[00816] Embodiment II-35. The genetically modified host cell of any one of Embodiments II-1 to II-33, wherein at least one of the one or more heterologous nucleic acids is maintained extrachromosomally.

[00817] Embodiment II-36. The genetically modified host cell of any one of Embodiments II-1 to II-33, wherein two or more of the one or more heterologous nucleic acids are present in a single expression vector.

[00818] Embodiment II-37. The genetically modified host cell of any one of Embodiments II-1 to II-36, wherein the cannabinoid compound is cannabigerolic acid.

[00819] Embodiment II-38. The genetically modified host cell of Embodiment II-1, wherein the genetically modified host cell is genetically modified with a heterologous nucleic acid encoding a Δ^9-THCA synthase polypeptide.

[00820] Embodiment II-39. The genetically modified host cell of Embodiment II-38, wherein the THCA synthase polypeptide comprises an amino acid having at least 50% amino acid sequence identity to SEQ ID NO:14 or SEQ ID NO:15.

[00821] Embodiment II-40. The genetically modified host cell of Embodiment II-1, wherein the genetically modified host cell is genetically modified with a heterologous nucleic acid encoding a CBDA synthase polypeptide.

[00822] Embodiment II-41. The genetically modified host cell of Embodiment II-40, wherein the CBDA synthase polypeptide comprises an amino acid having at least 50% amino acid sequence identity to SEQ ID NO:88 or SEQ ID NO:16.

[00823] Embodiment II-42. The genetically modified host cell of any one of Embodiments II-1 to II-41, wherein at least one of the heterologous nucleic acids is operably linked to an inducible promoter.

[00824] Embodiment II-43. The genetically modified host cell of any one of Embodiments II-1 to II-41, wherein at least one of the heterologous nucleic acids is operably linked to a constitutive promoter.

[00825] Embodiment II-44. The genetically modified host cell of Embodiment II-1, wherein the cannabinoid compound is cannabichromenic acid, cannabigerolic acid, Δ^9-tetrahydrocannabinolic acid, cannabidiolic acid, Δ^9-tetrahydrocannabinol, cannabidiol, or cannabichromene.

[00826] Embodiment II-45. A method of synthesizing a cannabinoid compound or cannabinoid precursor in a host cell, the method comprising: a) culturing a host cell of any one of Embodiments II-1 to II-40 in a suitable medium; and b) recovering the produced cannabinoid compound or cannabinoid precursor.

[00827] Embodiment II-46. The method of Embodiment II-45, wherein the medium comprises a fermentable sugar.

[00828] Embodiment II-47. The method of Embodiment II-45, wherein the medium comprises a pretreated cellulosic feedstock.

[00829] Embodiment II-48. A genetically modified host cell that produces an olivetolic acid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode an acyl-activating enzyme

(AAE) polypeptide; and b) one or more heterologous nucleic acids that encode a TKS/OAC fusion polypeptide.

[00830] Embodiment II-49. The genetically modified host cell of Embodiment II-48, wherein the AAE polypeptide comprises an amino acid sequence having at least 50% amino acid sequence identity to an AAE amino acid sequence set forth in SEQ ID NO:90 or SEQ ID NO:91.

[00831] Embodiment II-50. A genetically modified host cell that produces a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode an acyl-activating enzyme (AAE) polypeptide; b) one or more heterologous nucleic acids that encode a TKS/OAC fusion polypeptide; and c) one or more heterologous nucleic acids that encode a GOT polypeptide or an NphB polypeptide.

[00832] Embodiment II-51. The genetically modified host cell of Embodiment II-50, wherein the genetically modified host cell is genetically modified with a heterologous nucleic acid encoding a THCAS polypeptide.

[00833] Embodiment II-52. The genetically modified host cell of Embodiment II-50, wherein the genetically modified host cell is genetically modified with a heterologous nucleic acid encoding a CBDAS polypeptide.

[00834] Embodiment II-53. A method of producing an olivetolic acid derivative, the method comprising culturing the genetically modified host cell of Embodiment II-48 or II-49 in a culture medium comprising a carboxylic acid.

[00835] Embodiment II-54. A method of producing a cannabinoid derivative, the method comprising culturing the genetically modified host cell of any one of Embodiments II-50 to II-52 in a culture medium comprising a carboxylic acid.

[00836] Embodiment II-55. The genetically modified host cell of Embodiment II-1, wherein the geranyl pyrophosphate olivetolic acid geranyltransferase (GOT) polypeptide or the NphB polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence of a GOT polypeptide, including CsPT4t polypeptide, or NphB polypeptide sequence, respectively, disclosed herein.

[00837] Some embodiments of the disclosure are of Embodiment III:

[00838] Embodiment III-1. A genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide, wherein said geranyl pyrophosphate:olivetolic acid

geranyltransferase polypeptide can catalyze production of cannabigerolic acid from geranyl pyrophosphate and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00839] Embodiment III-2. A genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110.

[00840] Embodiment III-3. A genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100.

[00841] Embodiment III-4. The genetically modified host cell of any one of Embodiments III-1 to III-3, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a tetraketide synthase (TKS) polypeptide and one or more heterologous nucleic acids encoding an olivetolic acid (OAC) polypeptide, or one or more heterologous nucleic acids encoding a fusion TKS and OAC polypeptide.

[00842] Embodiment III-5. The genetically modified host cell of Embodiment III-4, wherein the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76.

[00843] Embodiment III-6. The genetically modified host cell of Embodiment III-4 or III-5, wherein the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00844] Embodiment III-7. The genetically modified host cell of any one of Embodiments III-1 to III-6, wherein the genetically modified host cell further comprises one or more of the following: a) one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; b) one or more heterologous nucleic acids encoding a polypeptide that generates geranyl pyrophosphate; or c) one or more heterologous nucleic acids encoding a polypeptide that generates malonyl-CoA.

[00845] Embodiment III-8. The genetically modified host cell of Embodiment III-7, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA

compound derivative, wherein the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is an acyl-activating enzyme (AAE) polypeptide.

[00846] Embodiment III-9. The genetically modified host cell of Embodiment III-8, wherein the AAE polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90.

[00847] Embodiment III-10. The genetically modified host cell of Embodiment III-8, wherein the AAE polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:92 or SEQ ID NO:149.

[00848] Embodiment III-11. The genetically modified host cell of Embodiment III-7, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative, wherein the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is a fatty acyl-CoA ligase polypeptide.

[00849] Embodiment III-12. The genetically modified host cell of Embodiment III-11, wherein the fatty acyl-CoA ligase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:145 or SEQ ID NO:147.

[00850] Embodiment III-13. The genetically modified host cell of Embodiment III-7, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative, wherein the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is a fatty acyl-CoA synthetase (FAA) polypeptide.

[00851] Embodiment III-14. The genetically modified host cell of Embodiment III-13, wherein the FAA polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200.

[00852] Embodiment III-15. The genetically modified host cell of any one of Embodiments III-7 to III-14, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates geranyl pyrophosphate, wherein the polypeptide that generates geranyl pyrophosphate is a geranyl pyrophosphate synthetase (GPPS) polypeptide.

[00853] Embodiment III-16. The genetically modified host cell of Embodiment III-15, wherein the GPPS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60.

[00854] Embodiment III-17. The genetically modified host cell of any one of Embodiments III-7 to III-16, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates malonyl-CoA, wherein the polypeptide that generates malonyl-CoA is an acetyl-CoA carboxylase-1 (ACC1) polypeptide.

[00855] Embodiment III-18. The genetically modified host cell of Embodiment III-17, wherein the ACC1 polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207.

[00856] Embodiment III-19. The genetically modified host cell of any one of Embodiments III-1 to III-18, wherein the genetically modified host cell further comprises one or more of the following: a) one or more heterologous nucleic acids encoding a HMG-CoA synthase (HMGS) polypeptide; b) one or more heterologous nucleic acids encoding a 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR) polypeptide; c) one or more heterologous nucleic acids encoding a mevalonate kinase (MK) polypeptide; d) one or more heterologous nucleic acids encoding a phosphomevalonate kinase (PMK) polypeptide; e) one or more heterologous nucleic acids encoding a mevalonate pyrophosphate decarboxylase (MVD) polypeptide; or f) one or more heterologous nucleic acids encoding a isopentenyl diphosphate isomerase (IDI) polypeptide.

[00857] Embodiment III-20. The genetically modified host cell of Embodiment III-19, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an IDI polypeptide.

[00858] Embodiment III-21. The genetically modified host cell of Embodiment III-20, wherein the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58.

[00859] Embodiment III-22. The genetically modified host cell of any one of Embodiments III-19 to III-21, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGR polypeptide.

[00860] Embodiment III-23. The genetically modified host cell of Embodiment III-22, wherein the HMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:22.

[00861] Embodiment III-24. The genetically modified host cell of any one of Embodiments III-19 to III-21, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGR polypeptide, wherein the HMGR polypeptide is a truncated HMGR (tHMGR) polypeptide.

[00862] Embodiment III-25. The genetically modified host cell of Embodiment III-24, wherein the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208.

[00863] Embodiment III-26. The genetically modified host cell of any one of Embodiments III-19 to III-25, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGS polypeptide.

[00864] Embodiment III-27. The genetically modified host cell of Embodiment III-26, wherein the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115.

[00865] Embodiment III-28. The genetically modified host cell of any one of Embodiments III-19 to III-27, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an MK polypeptide.

[00866] Embodiment III-29. The genetically modified host cell of Embodiment III-28, wherein the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64.

[00867] Embodiment III-30. The genetically modified host cell of any one of Embodiments III-19 to III-29, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a PMK polypeptide.

[00868] Embodiment III-31. The genetically modified host cell of Embodiment III-30, wherein the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205.

[00869] Embodiment III-32. The genetically modified host cell of any one of Embodiments III-19 to III-31, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a MVD polypeptide.

[00870] Embodiment III-33. The genetically modified host cell of Embodiment III-32, wherein the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66.

[00871] Embodiment III-34. The genetically modified host cell of any one of Embodiments III-1 to III-33, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA.

[00872] Embodiment III-35. The genetically modified host cell of Embodiment III-34, wherein the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide.

[00873] Embodiment III-36. The genetically modified host cell of Embodiment III-35, wherein the acetoacetyl-CoA thiolase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25.

[00874] Embodiment III-37. The genetically modified host cell of any one of Embodiments III-1 to III-36, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a pyruvate dehydrogenase complex (PDC) polypeptide.

[00875] Embodiment III-38. The genetically modified host cell of Embodiment III-37, wherein the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117.

[00876] Embodiment III-39. The genetically modified host cell of any one of Embodiments III-1 to III-38, wherein the genetically modified host cell is a eukaryotic cell.

[00877] Embodiment III-40. The genetically modified host cell of Embodiment III-39, wherein the eukaryotic cell is a yeast cell.

[00878] Embodiment III-41. The genetically modified host cell of Embodiment III-40, wherein the yeast cell is *Saccharomyces cerevisiae*.

[00879] Embodiment III-42. The genetically modified host cell of Embodiment III-41, wherein the *Saccharomyces cerevisiae* is a protease-deficient strain of *Saccharomyces cerevisiae*.

[00880] Embodiment III-43. The genetically modified host cell of any one of Embodiments III-1 to III-39, wherein the genetically modified host cell is a plant cell.

[00881] Embodiment III-44. The genetically modified host cell of any one of Embodiments III-1 to III-38, wherein the genetically modified host cell is a prokaryotic cell.

[00882] Embodiment III-45. The genetically modified host cell of any one of Embodiments III-1 to III-44, wherein at least one of the one or more heterologous nucleic acids is integrated into the chromosome of the genetically modified host cell.

[00883] Embodiment III-46. The genetically modified host cell of any one of Embodiments III-1 to III-44, wherein at least one of the one or more heterologous nucleic acids is maintained extrachromosomally.

[00884] Embodiment III-47. The genetically modified host cell of any one of Embodiments III-1 to III-44, wherein two or more of the one or more heterologous nucleic acids are present in a single expression vector.

[00885] Embodiment III-48. The genetically modified host cell of any one of Embodiments III-1 to III-44, wherein at least one of the heterologous nucleic acids is operably linked to an inducible promoter.

[00886] Embodiment III-49. The genetically modified host cell of any one of Embodiments III-1 to III-44, wherein at least one of the heterologous nucleic acids is operably linked to a constitutive promoter.

[00887] Embodiment III-50. The genetically modified host cell of any one of Embodiments III-1 to III-49, wherein culturing of the genetically modified host cell in a suitable medium provides for synthesis of the cannabinoid or the cannabinoid derivative in an increased amount compared to a non-genetically modified host cell cultured under similar conditions.

[00888] Embodiment III-51. The genetically modified host cell of any one of Embodiments III-1 to III-50, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a cannabinoid synthase polypeptide.

[00889] Embodiment III-52. The genetically modified host cell of Embodiment III-51, wherein the cannabinoid synthase polypeptide is a tetrahydrocannabinolic acid (THCA) synthase polypeptide.

[00890] Embodiment III-53. The genetically modified host cell of Embodiment III-52, wherein the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155.

[00891] Embodiment III-54. The genetically modified host cell of Embodiment III-51, wherein the cannabinoid synthase polypeptide is a cannabidiolic acid (CBDA) synthase polypeptide.

[00892] Embodiment III-55. The genetically modified host cell of Embodiment III-54, wherein the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[00893] Embodiment III-56. The genetically modified host cell of any one of Embodiments III-1 to III-55, wherein the cannabinoid is cannabigerolic acid, cannabigerol, Δ^9-tetrahydrocannabinolic acid, Δ^9-tetrahydrocannabinol, Δ^8-tetrahydrocannabinolic acid, Δ^8-tetrahydrocannabinol, cannabidiolic acid, cannabidiol, cannabichromenic acid, cannabichromene, cannabinolic acid, cannabinol, cannabidivarinic acid, cannabidivarin, tetrahydrocannabivarinic acid, tetrahydrocannabivarin, cannabichromevarinic acid,

cannabichromevarin, cannabigerovarinic acid, cannabigerovarin, cannabicyclolic acid, cannabicyclol, cannabielsoinic acid, cannabielsoin, cannabicitranic acid, or cannabicitran.

[00894] Embodiment III-57. The genetically modified host cell of Embodiment III-56, wherein the cannabinoid is cannabigerolic acid.

[00895] Embodiment III-58. A method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising: a) culturing the genetically modified host cell of any one of Embodiments III-1 to III-57 in a suitable medium; and b) recovering the produced cannabinoid or cannabinoid derivative.

[00896] Embodiment III-59. A method of producing a cannabinoid or a cannabinoid derivative, the method comprising: a) culturing the genetically modified host cell of any one of Embodiments III-1 to III-57 in a suitable medium comprising a carboxylic acid; b) recovering the produced cannabinoid or cannabinoid derivative.

[00897] Embodiment III-60. A method of producing a cannabinoid or a cannabinoid derivative, the method comprising: a) culturing the genetically modified host cell of any one of Embodiments III-1 to III-57 in a suitable medium comprising olivetolic acid or an olivetolic acid derivative; b) recovering the produced cannabinoid or cannabinoid derivative.

[00898] Embodiment III-61. A method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising: a) culturing a genetically modified host cell comprising one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide, wherein said geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide can catalyze production of cannabigerolic acid from geranyl pyrophosphate and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, in a suitable medium; and b) recovering the produced cannabinoid or cannabinoid derivative.

[00899] Embodiment III-62. A method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising: a) culturing a genetically modified host cell comprising one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110 in a suitable medium; and b) recovering the produced cannabinoid or cannabinoid derivative.

[00900] Embodiment III-63. A method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising: a) culturing a genetically modified host cell comprising one or more heterologous nucleic acids encoding a

geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100 in a suitable medium; and b) recovering the produced cannabinoid or cannabinoid derivative.

[00901] Embodiment III-64. The method of any one of Embodiments III-58 to III-63, wherein the suitable medium comprises a fermentable sugar.

[00902] Embodiment III-65. The method of any one of Embodiments III-58 to III-63, wherein the suitable medium comprises a pretreated cellulosic feedstock.

[00903] Embodiment III-66. The method of any one of Embodiments III-58 to III-63, wherein the suitable medium comprises a non-fermentable carbon source.

[00904] Embodiment III-67. The method of Embodiment III-66, wherein the non-fermentable carbon source comprises ethanol.

[00905] Embodiment III-68. The method of any one of Embodiments III-58 to III-67, wherein the cannabinoid is cannabigerolic acid, cannabigerol, Δ^9-tetrahydrocannabinolic acid, Δ^9-tetrahydrocannabinol, Δ^8-tetrahydrocannabinolic acid, Δ^8-tetrahydrocannabinol, cannabidiolic acid, cannabidiol, cannabichromenic acid, cannabichromene, cannabinolic acid, cannabinol, cannabidivarinic acid, cannabidivarin, tetrahydrocannabivarinic acid, tetrahydrocannabivarin, cannabichromevarinic acid, cannabichromevarin, cannabigerovarinic acid, cannabigerovarin, cannabicyclolic acid, cannabicyclol, cannabielsoinic acid, cannabielsoin, cannabicitranic acid, or cannabicitran.

[00906] Embodiment III-69. An isolated or purified geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide, wherein said geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide can catalyze production of cannabigerolic acid from geranyl pyrophosphate and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00907] Embodiment III-70. An isolated or purified polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110.

[00908] Embodiment III-71. An isolated or purified polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100.

[00909] Embodiment III-72. An isolated or purified nucleic acid encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide, wherein said geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide can catalyze production of cannabigerolic acid from geranyl pyrophosphate and olivetolic acid in an amount at least ten

times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00910] Embodiment III-73. An isolated or purified nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110.

[00911] Embodiment III-74. An isolated or purified nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100.

[00912] Embodiment III-75. A vector comprising a nucleic acid encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide, wherein said geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide can catalyze production of cannabigerolic acid from geranyl pyrophosphate and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00913] Embodiment III-76. A vector comprising a nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110.

[00914] Embodiment III-77. A vector comprising a nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100.

[00915] Embodiment III-78. A method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide, wherein said geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide can catalyze production of cannabigerolic acid from geranyl pyrophosphate and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, into the genetically modified host cell.

[00916] Embodiment III-79. A method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110 into the genetically modified host cell.

[00917] Embodiment III-80. A method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100 into the genetically modified host cell.

[00918] Embodiment III-81. A method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing the vector of any one of Embodiments III-75 to III-77 into the genetically modified host cell.

[00919] Provided in Table 1 are amino acid and nucleotide sequences disclosed herein. Where a genus and/or species is noted, the sequence should not be construed to be limited only to the specified genus and/or species, but also includes other genera and/or species expressing said sequence.

Table 1: Amino acid and nucleotide sequences of the disclosure

SEQ ID	SEQUENCE
SEQ ID NO:1 hexanoyl-CoA synthetase (HCS) GenBank AFD33359 *Cannabis sativa*	MALELPHLLPYKLVKGQTLVAQAARAELASSSSSSVILKSNFINNNYIN YCNNNNNNERRLVVRRDWETMASSPSHSRNNNDIRTINHLRHVDSMA TMPSGAGKIPRLNAVILGEALATEENDLVFPTDEFSQQAHVPSPQKYLE MYKRSIEDPAGFWSEIASQFYWKQKWDDSVYSENLDVSKGRVNIEWF KGGITNICYNCLDKNVEAGLGDKIALYWEGNDTGFDDSLTYSQLLHK VCQLANYLKDMGVQKGDAVVIYLPMLLELPITMLACARIGAVHSVVF AGFSAESLSQRIIDCKPKVVITCNAVKRGPKIIHLKDIVDAALVESAKTG VPIDTCLVYENQLAMKRDITKWQDGRDIWWQDVIPKYPTECAVEWV DAEDPLFLLYTSGSTGKPKGVLHTTGGYMVYTATTFKYAFDYKPSDV YWCTADCGWITGHSYVTYGPLLNGASCIVFEGAPNYPDSGRCWDIVD KYKVTIFYTAPTLVRSLMRDGDEYVTRYSRKSLRILGSVGEPINPSAWR WFYNVVGDSRCPISDTWWQTETGGFMITPLPGAWPQKPGSATFPFFGV KPVIVDEKGVEIEGECSGYLCVKGSWPGAFRTLYGDYERYETTYFKPF TGYYFTGDGCSRDKDGYHWLTGRVDDVINVSGHRIGTAEVESALVSH PKCAEAAVVGIEHEVKGQAIYAFVTLVEGEPYSEELRKSLILSVRKQIG AFAAPERIHWAPGLPKTRSGKIMRRILRKIASGQLDELGDTSTLADPNV VEQLISLSNC
SEQ ID NO:2 RevS BAK64635.1| putative CoA ligase [Streptomyces sp. SN-593]	MELALPAELAPTLPEALRLRSEQQPDTVAYVFLRDGETPEETLTYGRL DRAARARAAALEAAGLAGGTAVLLYPSGLEFVAALLGCMYAGTAGA PVQVPTRRRGMERARRIADDAGAKTILTTTAVKREVEEHFADLLTGLT VIDTESLPDVPDDAPAVRLPGPDDVALLQYTSGSTGDPKGVEVTHANF RANVAETVELWPVRSDGTVVNWLPLFHDMGLMFGVVMPLFTGVPAY LMAPQSFIRRPARWLEAISRFRGTHAAAPSFAYELCVRSVADTGLPAG LDLSSWRVAVNGAEPVRWTAVADFTEAYAPAGFRPQAMCPGYGLAE NTLKLSGSPEDRPPTLLRADAAALQDGRVVPLTGPGTDGVRLVGSGVT VPSSRVAVVDPGTGTEQPAGRVGEIWINGPCVARGYHGRPAESAESFG ARIAGQEARGTWLRTGDLGFLHDGEVFVAGRLKDVVIHQGRNFYPQD IELSAEVSDRALHPNCAAAFALDDGRTERLVLLVEADGRALRNGGAD ALRARVHDAVWDRQRLRIDEIVLLRRGALPKTSSGKVQRRLARSRYL

SEQ ID	SEQUENCE
	DGEFGPAPAREA
SEQ ID NO:3 hexanoyl-CoA synthase (AflA); GenBank AAL99898 *Aspergillus* sp.	MVIQGKRLAASSIQLLASSLDAKKLCYEYDERQAPGVTQITEEAPTEQP PLSTPPSLPQTPNISPISASKIVIDDVALSRVQIVQALVARKLKTAIAQLP TSKSIKELSGGRSSLQNELVGDIHNEFSSIPDAPEQILLRDFGDANPTVQ LGKTSSAAVAKLISSKMPSDFNANAIRAHLANKWGLGPLRQTAVLLY AIASEPPSRLASSSAAEEYWDNVSSMYAESCGITLRPRQDTMNEDAMA SSAIDPAVVAEFSKGHRRLGVQQFQALAEYLQIDLSGSQASQSDALVA ELQQKVDLWTAEMTPEFLAGISPMLDVKKSRRYGSWWNMARQDVLA FYRRPSYSEFVDDALAFKVFLNRLCNRADEALLNMVRSLSCDAYFKQ GSLPGYHAASRLLEQAITSTVADCPKARLILPAVGPHTTITKDGTIEYAE APRQGVSGPTAYIQSLRQGASFIGLKSADVDTQSNLTDALLDAMCLAL HNGISFVGKTFLVTGAGQGSIGAGVVRLLLEGGARVLVTTSREPATTS RYFQQMYDNHGAKFSELRVVPCNLASAQDCEGLIRHVYDPRGLNWD LDAILPFAAASDYSTEMHDIRGQSELGHRLMLVNVFRVLGHIVHCKRD AGVDCHPTQVLLPLSPNHGIFGGDGMYPESKLALESLFHRIRSESWSDQ LSICGVRIGWTRSTGLMTAHDIIAETVEEHGIRTFSVAEMALNIAMLLT PDFVAHCEDGPLDADFTGSLGTLGSIPGFLAQLHQKVQLAAEVIRAVQ AEDEHERFLSPGTKPTLQAPVAPMHPRSSLRVGYPRLPDYEQEIRPLSP RLERLQDPANAVVVVGYSELGPWGSARLRWEIESQGQWTSAGYVEL AWLMNLIRHVNDESYVGWVDTQTGKPVRDGEIQALYGDHIDNHTGIR PIQSTSYNPERMEVLQEVAVEEDLPEFEVSQLTADAMRLRHGANVSIR PSGNPDACHVKLKRGAVILVPKTVPFVWGSCAGELPKGWTPAKYGIPE NLIHQVDPVTLYTICCVAEAFYSAGITHPLEVFRHIHLSELGNFIGSSMG GPTKTRQLYRDVYFDHEIPSDVLQDTYLNTPAAWVNMLLLGCTGPIKT PVGACATGVESIDSGYESIMAGKTKMCLVGGYDDLQEEASYGFAQLK ATVNVEEEIACGRQPSEMSRPMAESRAGFVEAHGCGVQLLCRGDIALQ MGLPIYAVIASSAMAADKIGSSVPAPGQGILSFSRERARSSMISVTSRPS SRSSTSSEVSDKSSLTSITSISNPAPRAQRARSTTDMAPLRAALATWGLT IDDLDVASLHGTSTRGNDLNEPEVIETQMRHLGRTPGRPLWAICQKSV TGHPKAPAAAWMLNGCLQVLDSGLVPGNRNLDTLDEALRSASHLCFP TRTVQLREVKAFLLTSFGFGQKGGQVVGVAPKYFFATLPRPEVEGYYR KVRVRTEAGDRAYAAAVMSQAVVKIQTQNPYDEPDAPRIFLDPLARIS QDPSTGQYRFRSDATPALDDDALPPPGEPTELVKGISSAWIEEKVRPHM SPGGTVGVDLVPLASFDAYKNAIFVERNYTVRERDWAEKSADVRAAY ASRWCAKEAVFKCLQTHSQGAGAAMKEIEIEHGGNGAPKVKLRGAA QTAARQRGLEGVQLSISYGDDAVIAVALGLMSGAS
SEQ ID NO:4 hexanoyl-CoA synthase (AflB) AAS66003.1| fatty acid synthase beta subunit [Aspergillus sp.]	MGSVSREHESIPIQAAQRGAARICAAFGGQGSNNLDVLKGLLELYKRY GPDLDELLDVASNTLSQLASSPAAIDVHEPWGFDLRQWLTTPEVAPSK EILALPPRSFPLNTLLSLALYCATCRELELDPGQFRSLLHSSTGHSQGIL AAVAITQAESWPTFYDACRTVLQISFWIGLEAYLFTPSSAASDAMIQDC IEHGEGLLSSMLSVSGLSRSQVERVIEHVNKGLGECNRWVHLALVNSH EKFVLAGPPQSLWAVCLHVRRIRADNDLDQSRILFRNRKPIVDILFLPIS APFHTPYLDGVQDRVIEALSSASLALHSIKIPLYHTGTGSNLQELQPHQ LIPTLIRAITVDQLDWPLVCRGLNATHVLDFGPGQTCSLIQELTQGTGV SVIQLTTQSGPKPVGGHLAAVNWEAEFGLRLHANVHGAAKLHNRMT TLLGKPPVMVAGMTPTTVRWDFVAAVAQAGYHVELAGGGYHAERQ FEAEIRRLATAIPADHGITCNLLYAKPTTFSWQISVIKDLVRQGVPVEGI TIGAGIPSPEVVQECVQSIGLKHISFKPGSFEAIHQVIQIARTHPNFLIGLQ WTAGRGGGHHSWEDFHGPILATYAQIRSCPNILLVVGSGFGGGPDTFP YLTGQWAQAFGYPCMPFDGVLLGSRMMVAREAHTSAQAKRLIIDAQ GVGDADWHKSFDEPTGGVVTVNSEFGQPIHVLATRGVMLWKELDNR VFSIKDTSKRLEYLRNHRQEIVSRLNADFARPWFAVDGHGQNVELED MTYLEVLRRLCDLTYVSHQKRWVDPSYRILLLDFVHLLRERFQCAIDN

SEQ ID	SEQUENCE
	PGEYPLDIIVRVEESLKDKAYRTLYPEDVSLLMHLFSRRDIKPVPFIPRL DERFETWFKKDSLWQSEDVEAVIGQDVQRIFIIQGPMAVQYSISDDESV KDILHNICNHYVEALQADSRETSIGDVHSITQKPLSAFPGLKVTTNRVQ GLYKFEKVGAVPEMDVLFEHIVGLSKSWARTCLMSKSVFRDGSRLHN PIRAALQLQRGDTIEVLLTADSEIRKIRLISPTGDGGSTSKVVLEIVSNDG QRVFATLAPNIPLSPEPSVVFCFKVDQKPNEWTLEEDASGRAERIKALY MSLWNLGFPNKASVLGLNSQFTGEELMITTDKIRDFERVLRQTSPLQL QSWNPQGCVPIDYCVVIAWSALTKPLMVSSLKCDLLDLLHSAISFHYA PSVKPLRVGDIVKTSSRILAVSVRPRGTMLTVSADIQRQGQHVVTVKS DFFLGGPVLACETPFELTEEPEMVVHVDSEVRRAILHSRKWLMREDRA LDLLGRQLLFRLKSEKLFRPDGQLALLQVTGSVFSYSPDGSTTAFGRV YFESESCTGNVVMDFLHRYGAPRAQLLELQHPGWTGTSTVAVRGPRR SQSYARVSLDHNPIHVCPAFARYAGLSGPIVHGMETSAMMRRIAEWAI GDADRSRFRSWHITLQAPVHPNDPLRVELQHKAMEDGEMVLKVQAF NERTEERVAEADAHVEQETTAYVFCGQGSQRQGMGMDLYVNCPEAK ALWARADKHLWEKYGFSILHIVQNNPPALTVHFGSQRGRRIRANYLR MMGQPPIDGRHPPILKGLTRNSTSYTFSYSQGLLMSTQFAQPALALME MAQFEWLKAQGVVQKGARFAGHSLGEYAALGACASFLSFEDLISLIFY RGLKMQNALPRDANGHTDYGMLAADPSRIGKGFEEASLKCLVHIIQQ ETGWFVEVVNYNINSQQYVCAGHFRALWMLGKICDDLSCHPQPETVE GQELRAMVWKHVPTVEQVPREDRMERGRATIPLPGIDIPYHSTMLRGE IEPYREYLSERIKVGDVKPCELVGRWIPNVVGQPFSVDKSYVQLVHGIT GSPRLHSLLQQMA
SEQ ID NO:5 AAF08793.1|AF182828_1 geranyl diphosphate synthase large subunit [Mentha x piperita]	MSALVNPVAKWPQTIGVKDVHGGRRRRSRSTLFQSHPLRTEMPFSLYF SSPLKAPATFSVSAVYTKEGSEIRDKDPAPSTSPAFDFDGYMLRKAKSV NKALEAAVQMKEPLKIHESMRYSLLAGGKRVRPMLCIAACELVGGDE STAMPAACAVEMIHTMSLMHDDLPCMDNDDLRRGKPTNHMAFGESV AVLAGDALLSFAFEHVAAATKGAPPERIVRVLGELAVSIGSEGLVAGQ VVDVCSEGMAEVGLDHLEFIHHHKTAALLQGSVVLGAILGGGKEEEV AKLRKFANCIGLLFQVVDDILDVTKSSKELGKTAGKDLVADKTTYPKL IGVEKSKEFADRLNREAQEQLLHFHPHRAAPLIALANYIAYRDN
SEQ ID NO:6 AAF08792.1|AF182827_1 geranyl diphosphate synthase small subunit [Mentha x piperita]	MAINLSHINSKTCFPLKTRSDLSRSSSARCMPTAAAAAFPTIATAAQSQ PYWAAIEADIERYLKKSITIRPPETVFGPMHHLTFAAPATAASTLCLAA CELVGGDRSQAMAAAAAIHLVHAAAYVHEHLPLTDGSRPVSKPAIQH KYGPNVELLTGDGIVPFGFELLAGSVDPARTDDPDRILRVIIEISRAGGP EGMISGLHREEEIVDGNTSLDFIEYVCKKKYGEMHACGAACGAILGGA AEEEIQKLRNFGLYQGTLRGMMEMKNSHQLIDENIIGKLKELALEELG GFHGKNAELMSSLVAEPSLYAA
SEQ ID NO:7 Erg20: farnesylpyro phosphate synthase (Saccharomyces sp.)	MASEKEIRRERFLNVFPKLVEELNASLLAYGMPKEACDWYAHSLNYN TPGGKLNRGLSVVDTYAILSNKTVEQLGQEEYEKVAILGWCIELLQAY FLVADDMMDKSITRRGQPCWYKVPEVGEIAINDAFMLEAAIYKLLKS HFRNEKYYIDITELFHEVTFQTELGQLMDLITAPEDKVDLSKFSLKKHS FIVTF**K**TAYYSFYLPVALAMYVAGITDEKDLKQARDVLIPLGEYFQIQD DYLDCFGTPEQIGKIGTDIQDNKCSWVINKALELASAEQRKTLDENYG KKDSVAEAKCKKIFNDLKIEQLYHEYEESIAKDLKAKISQVDESRGFKA DVLTAFLNKVYKRSK
SEQ ID NO:8 Mutated Erg20: farnesylpyro phosphate synthase	MASEKEIRRERFLNVFPKLVEELNASLLAYGMPKEACDWYAHSLNYN TPGGKLNRGLSVVDTYAILSNKTVEQLGQEEYEKVAILGWCIELLQAY FLVADDMMDKSITRRGQPCWYKVPEVGEIAINDAFMLEAAIYKLLKS HFRNEKYYIDITELFHEVTFQTELGQLMDLITAPEDKVDLSKFSLKKHS FIVTF**G**TAYYSFYLPVALAMYVAGITDEKDLKQARDVLIPLGEYFQIQD

SEQ ID	SEQUENCE
(K197G)	DYLDCFGTPEQIGKIGTDIQDNKCSWVINKALELASAEQRKTLDENYG KKDSVAEAKCKKIFNDLKIEQLYHEYEESIAKDLKAKISQVDESRGFKA DVLTAFLNKVYKRSK
SEQ ID NO:9 GenBank EXX73400 acetyl-CoA carboxylase (ACC1) Rhizophagus irregularis DAOM 197198w	MRAQAHLGGGLKRIETQHQKGKLTARERAELLLDPGSFNEYDTFVEH QCTDFGMDKNKIIGDGVVTGHGTINGRRVFTFSQDFTAFGGSLSKMHA QKICKIMDKAMLVGAPVIGLNDSGGARIQEGVDSLAGYADIFQRNVLS SGVVPQLSLIMGPCAGGAVYSPALTDFTFMVRDTSYLFVTGPEVVKAV CNEDVTQEELGGANTHTVISGVAHAAFENDIEAIQRIRDFMDFLPLSNR EQAPTRYSDDPIDREDPSLNHIIPVDSTKAYDMREIITRLIDDGHFFEIMP DYAKNIVVGFARMGGKTVSIVGNQPLVSSGVLDINSSVKAARFVRFCD AFNIPIITLVDVPGFLPGTAQEHNGIIRHGAKLLYAYAEATVPKITIITRK AYGGAYDVMSSKHLRGDMNYSWPTGEIAVMGAKGAVEIIFRHVEDR TQSEHEYIDKFANPIPAAQRGYIDDIILPAATRKRIIEDLFVLSHKQLPLI YKKHDNCPL
SEQ ID NO:10 Olivetolic acid cyclase (OAC) GenBank AFN42527 *Cannabis sativa*	MAVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVT QKNKEEGYTHIVEVTFESVETIQDYIIHPAHVGFGDVYRSFWEKLLIFD YTPRK
SEQ ID NO:11 Tetraketide synthase (TKS) GenBank B1Q2B6 Cannabis sativa	MNHLRAEGPASVLAIGTANPENILLQDEFPDYYFRVTKSEHMTQLKEK FRKICDKSMIRKRNCFLNEEHLKQNPRLVEHEMQTLDARQDMLVVEV PKLGKDACAKAIKEWGQPKSKITHLIFTSASTTDMPGADYHCAKLLGL SPSVKRVMMYQLGCYGGGTVLRIAKDIAENNKGARVLAVCCDIMACL FRGPSESDLELLVGQAIFGDGAAAVIVGAEPDESVGERPIFELVSTGQTI LPNSEGTIGGHIREAGLIFDLHKDVPMLISNNIEKCLIEAFTPIGISDWNSI FWITHPGGKAILDKVEEKLHLKSDKFVDSRHVLSEHGNMSSSTVLFVM DELRKRSLEEGKSTTGDGFEWGVLFGFGPGLTVERVVVRSVPIKY
SEQ ID NO:12 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase (GOT)	MNSIRAATTNQTEPPESDNHSVATKILNFGKACWKLQRPYTIIAFTSCA CGLFGKELLHNTNLISWSLMFKAFFFLVAILCIASFTTTINQIYDLHIDRI NKPDLPLASGEISVNTAWIMSIIVALFGLIITIKMKGGPLYIFGYCFGIFG GIVYSVPPFRWKQNPSTAFLLNFLAHIITNFTFYYASRAALGLPFELRPS FTFLLAFMKSMGSALALIKDASDVEGDTKFGISTLASKYGSRNLTLFCS GIVLLSYVAAILAGIIWPQAFNSNVMLLSHAILAFWLILQTRDFALTNY DPEAGRRFYEFMWKLYYAEYLVYVFIGS
SEQ ID NO:13 Engineered geranyl pyrophosphate olivetolic acid geranyltransferase (GOT)	MKDQRGNSIRASAQIEDRPPESGNLSALTNVKDFVSVCWEYVRPYTAK GVIICSSCLFGRELLENPNLFSWPLIFKAFFFLVAILCIASFTTTINQIYDL HIDRINKPDLPLASGEISVNTAWIMSIIVALFGLIITIKMKGGPLYIFGYCF GIFGGIVYSVPPFRWKQNPSTAFLLNFLAHIITNFTFYYASRAALGLPFE LRPSFTFLLAFMKSMGSALALIKDASDVEGDTKFGISTLASKYGSRNLT LFCSGIVLLSYVAAILAGIIWPQAFNSNVMLLSHAILAFWLILQTRDFAL TNYDPEAGRRFYEFMWKLYYAEYLVYVFIGS
SEQ ID NO:14 Mutant tetrahydro cannabinolic acid synthase (THCAS)	MNCSAFSFWFVCKIIFFFLSFHIQISIANPRENFLKCFSKHIPNNVANPKL VYTQHDQLYMSILNSTIQNLRFISDTTPKPLVIVTPSNNSHIQATILCSKK VGLQIRTRSGGHDAEGMSYISQVPFVVVDLRNMHSIKIDVHSQTAWVE AGATLGEVYYWINEKNENLSFPGGYCPTVGVGGHFSGGGYGALMRN YGLAADNIIDAHLVNVDGKVLDRKSMGEDLFWAIRGGGGENFGIIAA WKIKLVAVPSKSTIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLVLM THFITKNITDNHGKNKTTVHGYFSSIFHGGVDSLVDLMNKSFPELGIKK

SEQ ID	SEQUENCE
	TDCKEFSWIDTTIFYSGVVNFNTANFKKEILLDRSAGKKTAFSIKLDYV KKPIPETAMVKILEKLYEEDVGAGMYVLYPYGGIMEEISESAIPFPHRA GIMYELWYTASWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLN YRDLDLGKTNHASPNNYTQARIWGEKYFGKNFNRLVKVKTKVDPNN FFRNEQSIPPLPPHHHGS
SEQ ID NO:15 Truncated tetrahydro cannabinolic acid synthase (THCAS)	MNPRENFLKCFSKHIPNNVANPKLVYTQHDQLYMSILNSTIQNLRFISD TTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTRSGGHDAEGMSYISQVP FVVVDLRNMHSIKIDVHSQTAWVEAGATLGEVYYWINEKNENLSFPG GYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDR KSMGEDLFWAIRGGGGENFGIIAAWKIKLVAVPSKSTIFSVKKNMEIH GLVKLFNKWQNIAYKYDKDLVLMTHFITKNITDNHGKNKTTVHGYFS SIFHGGVDSLVDLMNKSFPELGIKKTDCKEFSWIDTTIFYSGVVNFNTA NFKKEILLDRSAGKKTAFSIKLDYVKKPIPETAMVKILEKLYEEDVGAG MYVLYPYGGIMEEISESAIPFPHRAGIMYELWYTASWEKQEDNEKHIN WVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNHASPNNYTQARIWG EKYFGKNFNRLVKVKTKVDPNNFFRNEQSIPPLPPHHHGS
SEQ ID NO:16 Truncated cannabidiolic acid synthase (CBDAS)	MNPRENFLKCFSQYIPNNATNLKLVYTQNNPLYMSVLNSTIHNLRFTS DTTPKPLVIVTPSHVSHIQGTILCSKKVGLQIRTRSGGHDSEGMSYISQV PFVIVDLRNMRSIKIDVHSQTAWVEAGATLGEVYYWVNEKNENLSLA AGYCPTVCAGGHFGGGGYGPLMRNYGLAADNIIDAHLVNVHGKVLD RKSMGEDLFWALRGGGAESFGIIVAWKIRLVAVPKSTMFSVKKIMEIH ELVKLVNKWQNIAYKYDKDLLLMTHFITRNITDNQGKNKTAIHTYFSS VFLGGVDSLVDLMNKSFPELGIKKTDCRQLSWIDTIIFYSGVVNYDTD NFNKEILLDRSAGQNGAFKIKLDYVKKPIPESVFVQILEKLYEEDIGAG MYALYPYGGIMDEISESAIPFPHRAGILYELWYICSWEKQEDNEKHLN WIRNIYNFMTPYVSKNPRLAYLNYRDLDIGINDPKNPNNYTQARIWGE KYFGKNFDRLVKVKTLVDPNNFFRNEQSIPPLPRHRHGS
SEQ ID NO:17 Truncated 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR)	MVLTNKTVISGSKVKSLSSAQSSSSGPSSSSEEDDSRDIESLDKKIRPLEE LEALLSSGNTKQLKNKEVAALVIHGKLPLYALEKKLGDTTRAVAVRR KALSILAEAPVLASDRLPYKNYDYDRVFGACCENVIGYMPLPVGVIGP LVIDGTSYHIPMATTEGCLVASAMRGCKAINAGGGATTVLTKDGMTR GPVVRFPTLKRSGACKIWLDSEEGQNAIKKAFNSTSRFARLQHIQTCLA GDLLFMRFRTTTGDAMGMNMISKGVEYSLKQMVEEYGWEDMEVVS VSGNYCTDKKPAAINWIEGRGKSVVAEATIPGDVVRKVLKSDVSALV ELNIAKNLVGSAMAGSVGGFNAHAANLVTAVFLALGQDPAQNVESSN CITLMKEVDGDLRISVSMPSIEVGTIGGGTVLEPQGAMLDLLGVRGPH ATAPGTNARQLARIVACAVLAGELSLCAALAAGHLVQSHMTHNRKPA EPTK PNNLDATDINRLKDGSVTCIKS
SEQ ID NO:18 PaaH1: 3-hydroxyacyl-CoA dehydrogenase (Ralstonia sp.)	MSIRTVGIVGAGTMGNGIAQACAVVGLNVVMVDISDAAVQKGVATV ASSLDRLIKKEKLTEADKASALARIKGSTSYDDLKATDIVIEAATENYD LKVKILKQIDGIVGENVIIASNTSSISITKLAAVTSRADRFIGMHFFNPVP VMALVELIRGLQTSDTTHAAVEALSKQLGKYPITVKNSPGFVVNRILCP MINEAFCVLGEGLASPEEIDEGMKLGCNHPIGPLALADMIGLDTMLAV MEVLYTEFADPKYRPAMLMREMVAAGYLGRKTGRGVYVYSK
SEQ ID NO:19 Crt: crotonase (Clostridium sp.)	MELNNVILEKEGKVAVVTINRPKALNALNSDTLKEMDYVIGEIENDSE VLAVILTGAGEKSFVAGADISEMKEMNTIEGRKFGILGNKVFRRLELLE KPVIAAVNGFALGGGCEIAMSCDIRIASSNARFGQPEVGLGITPGFGGT QRLSRLVGMGMAKQLIFTAQNIKADEALRIGLVNKVVEPSELMNTAK EIANKIVSNAPVAVKLSKQAINRGMQCDIDTALAFESEAFGECFSTEDQ KDAMTAFIEKRKIEGFKNR

SEQ ID	SEQUENCE
SEQ ID NO:20 Ter: trans-2-enoyl-CoA reductase (Treponema sp.)	MIVKPMVRNNICLNAHPQGCKKGVEDQIEYTKKRITAEVKAGAKAPK NVLVLGCSNGYGLASRITAAFGYGAATIGVSFEKAGSETKYGTPGWY NNLAFDEAAKREGLYSVTIDGDAFSDEIKAQVIEEAKKKGIKFDLIVYS LASPVRTDPDTGIMHKSVLKPFGKTFTGKTVDPFTGELKEISAEPANDE EAAATVKVMGGEDWERWIKQLSKEGLLEEGCITLAYSYIGPEATQAL YRKGTIGKAKEHLEATAHRLNKENPSIRAFVSVNKGLVTRASAVIPVIP LYLASLFKVMKEKGNHEGCIEQITRLYAERLYRKDGTIPVDEENRIRID DWELEEDVQKAVSALMEKVTGENAESLTDLAGYRHDFLASNGFDVE GINYEAEVERFDRI
SEQ ID NO:21 BktB: beta-ketothiolase (Ralstonia sp.)	MTREVVVVSGVRTAIGTFGGSLKDVAPAELGALVVREALARAQVSGD DVGHVVFGNVIQTEPRDMYLGRVAAVNGGVTINAPALTVNRLCGSGL QAIVSAAQTILLGDTDVAIGGGAESMSRAPYLAPAARWGARMGDAGL VDMMLGALHDPFHRIHMGVTAENVAKEYDISRAQQDEAALESHRRAS AAIKAGYFKDQIVPVVSKGRKGDVTFDTDEHVRHDATIDDMTKLRPV FVKENGTVTAGNASGLNDAAAAVVMMERAEAERRGLKPLARLVSYG HAGVDPKAMGIGPVPATKIALERAGLQVSDLDVIEANEAFAAQACAV TKALGLDPAKVNPNGSGISLGHPIGATGALITVKALHELNRVQGRYAL VTMCIGGGQGIAAIFERI
SEQ ID NO:22 MvaE: acetyl-CoA acetyltransferase/HMG-CoA reductase (Enterococcus sp.)	MKEVVMIDAARTPIGKYRGSLSPFTAVELGTLVTKGLLDKTKLKKDKI DQVIFGNVLQAGNGQNVARQIALNSGLPVDVPAMTINEVCGSGMKAV ILARQLIQLGEAELVIAGGTESMSQAPMLKPYQSETNEYGEPISSMVND GLTDAFSNAHMGLTAEKVATQFSVSREEQDRYALSSQLKAAHAVEAG VFSEEIIPVKISDEDVLSEDEAVRGNSTLEKLGTLRTVFSEEGTVTAGNA SPLNDGASVVILASKEYAENNNLPYLATIKEVAEVGIDPSIMGIAPIKAI QKLTDRSGMNLSTIDLFEINEAFAASSIVVSQELQLDEEKVNIYGGAIAL GHPIGASGARILTTLAYGLLREQKRYGIASLCIGGGLGLAVLLEANMEQ THKDVQKKKFYQLTPSERRSQLIEKNVLTQETALIFQEQTLSEELSDHM IENQVSEVEIPMGIAQNFQINGKKKWIPMATEEPSVIAAASNGAKICGNI CAETPQRLMRGQIVLSGKSEYQAVINAVNHRKEELILCANESYPSIVKR GGGVQDISTREFMGSFHAYLSIDFLVDVKDAMGANMINSILESVANKL REWFPEEEILFSILSNFATESLASACCEIPFERLGRNKEIGEQIAKKIQQA GEYAKLDPYRAATHNKGIMNGIEAVVAATGNDTRAVSASIHAYAARN GLYQGLTDWQIKGDKLVGKLTVPLAVATVGGASNILPKAKASLAMLD IDSAKELAQVIAAVGLAQNLAALRALVTEGIQKGHMGLQARSLAISIG AIGEEIEQVAKKLREAEKMNQQTAIQILEKIREK
SEQ ID NO:23 MvaS: HMG-CoA synthase (Lactobacillus plantarum)	MKIGIDKLHFATSHLYVDMAELATARQAEPDKYLIGIGQSKMAVIPPS QDVVTLAANAAAPMLTATDIAAIDLLVVGTESGIDNSKASAIYVAKLL GLSQRVRTIEMKEACYAATAGVQLAQDHVRVHPDKKALVIGSDVAR YGLNTPGEPTQGGGAVAMLISADPKVLVLGTESSLLSEDVMDFWRPL YHTEALVDGKYSSNIYIDYFQDVFKNYLQTTQTSPDTLTALVFHLPYT KMGLKALRSVLPLVDAEKQAQWLAHFEHARQLNRQVGNLYTGSLYL SLLSQLLTDPQLQPGNRLGLFSYGSGAEGEFYTGVIQPDYQTGLDHGLP QRLARRRRVSVAEYEALFSHQLQWRADDQSVSYADDPHRFVLTGQK NEQRQYLDQQV
SEQ ID NO:24 Erg13: HMG-CoA synthase (Saccharomyces cerevisiae)	MKLSTKLCWCGIKGRLRPQKQQQLHNTNLQMTELKKQKTAEQKTRP QNVGIKGIQIYIPTQCVNQSELEKFDGVSQGKYTIGLGQTNMSFVNDRE DIYSMSLTVLSKLIKSYNIDTNKIGRLEVGTETLIDKSKSVKSVLMQLFG ENTDVEGIDTLNACYGGTNALFNSLNWIESNAWDGRDAIVVCGDIAIY DKGAARPTGGAGTVAMWIGPDAPIVFDSVRASYMEHAYDFYKPDFTS EYPYVDGHFSLTCYVKALDQVYKSYSKKAISKGLVSDPAGSDALNVL KYFDYNVFHVPTCKLVTKSYGRLLYNDFRANPQLFPEVDAELATRDY DESLTDKNIEKTFVNVAKPFHKERVAQSLIVPTNTGNMYTASVYAAFA SLLNYVGSDDLQGKRVGLFSYGSGLAASLYSCKIVGDVQHIIKELDITN

SEQ ID	SEQUENCE
	KLAKRITETPKDYEAAIELRENAHLKKNFKPQGSIEHLQSGVYYLTNID DKFRRSYDVKK
SEQ ID NO:25 Erg10p: acetoacetyl CoA thiolase [Saccharomyces cerevisiae]	MSQNVYIVSTARTPIGSFQGSLSSKTAVELGAVALKGALAKVPELDAS KDFDEIIFGNVLSANLGQAPARQVALAAGLSNHIVASTVNKVCASAMK AIILGAQSIKCGNADVVVAGGCESMTNAPYYMPAARAGAKFGQTVLV DGVERDGLNDAYDGLAMGVHAEKCARDWDITREQQDNFAIESYQKS QKSQKEGKFDNEIVPVTIKGFRGKPDTQVTKDEEPARLHVEKLRSART VFQKENGTVTAANASPINDGAAAVILVSEKVLKEKNLKPLAIIKGWGE AAHQPADFTWAPSLAVPKALKHAGIEDINSVDYFEFNEAFSVVGLVNT KILKLDPSKVNVYGGAVALGHPLGCSGARVVVTLLSILQQEGGKIGVA AICNGGGGASSIVIEKI
SEQ ID NO:26 SCFA-TE: Short chain fatty acyl-CoA thioesterase From BMC Biochem. 2011 Aug 10;12:44. doi: 10.1186/1471-2091-12-44., e.g.: Bacteroides sp. (GenBank:CAH09236, Subfamily F)	MSDDKKIGSYKFIAEPFHVDFNGRLTMGVLGNHLLNCAGFHASERGF GIATLNEDNYTWVLSRLAIDLEEMPYQYEEFTVQTWVENVYRLFTDR NFAIIDKDGKKIGYARSVWAMINLNTRKPADLLTLHGGSIVDYVCDEP CPIEKPSRIKVATDQPCAKLTAKYSDIDINGHVNSIRYIEHILDLFPIDLY KSKRIQRFEMAYVAESYYGDELSFFEEEVSENEYHVEIKKNGSEVVCR AKVKFV
SEQ ID NO:27 SCFA-TE: Short chain fatty acyl-CoA thioesterase B. thetaiotaomicron (GenBank:AAO77182, Subfamily F)	MSEENKIGTYQFVAEPFHVDFNGRLTMGVLGNHLLNCAGFHASDRGF GIATLNEDNYTWVLSRLAIELDEMPYQYEKFSVQTWVENVYRLFTDR NFAVIDKDGKKIGYARSVWAMINLNTRKPADLLALHGGSIVDYICDEP CPIEKPSRIKVTSNQPVATLTAKYSDIDINGHVNSIRYIEHILDLFPIELYQ TKRIRRFEMAYVAESYFGDELSFFCDEVSENEFHVEVKKNGSEVVCRS KVIFE
SEQ ID NO:28 SCFA-TE: Short chain fatty acyl-CoA thioesterase Bryantella formatexigen (GenBank: EET61113, Subfamily H)	MIYMAYQYRSRIRYSEIGEDKKLTLPGLVNYFQDCSTFQSEALGIGLDT LGARQRAWLLASWKIVIDRLPRLGEEVVTETWPYGFKGFQGNRNFRM LDQEGHTLAAAASVWIYLNVESGHPCRIDGDVLEAYELEEELPLGPFS RKIPVPEESTERDSFLVMRSHLDTNHHVNNGQYILMAEEYLPEGFKVK QIRVEYRKAAVLHDTIVPFVCTEPQRCTVSLCGSDEKPFAVVEFSE
SEQ ID NO:29 SCFA-TE: Short chain fatty acyl-CoA thioesterase L. brevis (GenBank: ABJ63754, Subfamily J)	MAANEFSETHRVVYYEADDTGQLTLAMLINLFVLVSEDQNDALGLST AFVQSHGVGWVVTQYHLHIDELPRTGAQVTIKTRATAYNRYFAYREY WLLDDAGQVLAYGEGIWVTMSYATRKITTIPAEVMAPYHSEEQTRLP RLPRPDHFDEAVNQTLKPYTVRYFDIDGNGHVNNAHYFDWMLDVLP ATFLRAHHPTDVKIRFENEVQYGHQVTSELSQAAALTTQHMIKVGDLT AVKATIQWDNR
SEQ ID NO:30 SCFA-TE: Short chain fatty acyl-CoA	MATLGANASLYSEQHRITYYECDRTGRATLTTLIDIAVLASEDQSDAL GLTTEMVQSHGVGWVVTQYAIDITRMPRQDEVVTIAVRGSAYNPYFA YREFWIRDADGQQLAYITSIWVMMSQTTRRIVKILPELVAPYQSEVVK

SEQ ID	SEQUENCE
thioesterase L. plantarum (GenBank:CAD63310, Subfamily J)	RIPRLPRPISFEATDTTITKPYHVRFFDIDPNRHVNNAHYFDWLVDTLPATFLLQHDLVHVDVRYENEVKYGQTVTAHANILPSEVADQVTTSHLIEVDDEKCCEVTIQWRTLPEPIQ
SEQ ID NO:31 SCFA-TE: Short chain fatty acyl-CoA thioesterase Streptococcus dysgalactiae (GenBank:BAH81730, Subfamily J)	MGLSYREDIKLPFELCDVKSDIKFPLLLDYCLTVSGRQSAQLGRSNDYLLEQYGLIWIVTDYEATIHRLPHFQETITIETKALSYNKFFCYRQFYIYDQEGGLLVDILAYFALLNPDTRKVATIPEDLVAPFETDFVKKLHRVPKMPLLEQSIDRDYYVRYFDIDMNGHVNNSKYLDWMYDVLGCEFLKTHQPLKMTLKYVKEVSPGGQITSSYHLDQLTSYHQITSDGQLNAQAMIEWRAIKQTESEID
SEQ ID NO:32 DXS 1-deoxy-D-xylulose-5-phosphate synthase gene (dxs-AC# 16128405) Escherichia coli	MSFDIAKYPTLALVDSTQELRLLPKESLPKLCDELRRYLLDSVSRSSGHFASGLGTVELTVALHYVYNTPFDQLIWDVGHQAYPHKILTGRRDKIGTIRQKGGLHPFPWRGESEYDVLSVGHSSTSISAGIGIAVAAEKEGKNRRTVCVIGDGAITAGMAFEAMNHAGDIRPDMLVILNDNEMSISENVGALNNHLAQLLSGKLYSSLREGGKKVFSGVPPIKELLKRTEEHIKGMVVPGTLFEELGFNYIGPVDGHDVLGLITTLKNMRDLKGPQFLHIMTKKGRGYEPAEKDPITFHAVPKFDPSSGCLPKSSGGLPSYSKIFGDWLCETAAKDNKLMAITPAMREGSGMVEFSRKFPDRYFDVAIAEQHAVTFAAGLAIGGYKPIVAIYSTFLQRAYDQVLHDVAIQKLPVLFAIDRAGIVGADGQTHQGAFDLSYLRCIPEMVIMTPSDENECRQMLYTGYHYNDGPSAVRYPRGNAVGVELTPLEKLPIGKGIVKRRGEKLAILNFGTLMPEAAKVAESLNATLVDMRFVKPLDEALILEMAASHEALVTVEENAIMGGAGSGVNEVLMAHRKPVPVLNIGLPDFFIPQGTQEEMRAELGLDAAGMEAKIKAWLA
SEQ ID NO:33 DXR/IspC 1-deoxy-D-xylulose 5-phosphate reductoisomerase [Escherichia coli] AC# 16128166	MKQLTILGSTGSIGCSTLDVVRHNPEHFRVVALVAGKNVTRMVEQCLEFSPRYAVMDDEASAKLLKTMLQQQGSRTEVLSGQQAACDMAALEDVDQVMAAIVGAAGLLPTLAAIRAGKTILLANKESLVTCGRLFMDAVKQSKAQLLPVDSEHNAIFQSLPQPIQHNLGYADLEQNGVVSILLTGSGGPFRETPLRDLATMTPDQACRHPNWSMGRKISVDSATMMNKGLEYIEARWLFNASASQMEVLIHPQSVIHSMVRYQDGSVLAQLGEPDMRTPIAHTMAWPNRVNSGVKPLDFCKLSALTFAAPDYDRYPCLKLAMEAFEQGQAATTALNAANEITVAAFLAQQIRFTDIAALNLSVLEKMDMREPQCVDDVLSVDANAREVARKEVMRLAS
SEQ ID NO:34 IspD 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase [Escherichia coli] AC# 190908496	MATTHLDVCAVVPAAGFGRRMQTECPKQYLSIGNQTILEHSVHALLAHPRVKRVVIAISPGDSRFAQLPLANHPRITVVDGGEERADSVLAGLKAAGDAQWVLVHDAARPCLHQDDLARLLALSETSRTGGILAAPVRDTMKRAEPGKNAIAHTVDRNGLWHALTPQFFPRELLHDCLTRALNEGATITDEASALEYCGFHPQLVEGRADNIKVTRPEDLALAEFYLTRTIHQENT
SEQ ID NO:35 IspE 4-diphosphocytidyl-2-C-methylerythritol kinase [Escherichia coli] AC# 4062791	MRTQWPSPAKLNLFLYITGQRADGYHTLQTLFQFLDYGDTISIELRDDGDIRLLTPVEGVEHEDNLIVRAARLLMKTAADSGRLPTGSGANISIDKRLPMGGGLGGGSSNAATVLVALNHLWQCGLSMDELAEMGLTLGADVPVFVRGHAAFAEGVGEILTPVDPPEKWYLVAHPGVSIPTPVIFKDPELPRNTPKRSIETLLKCEFSNDCEVIARKRFREVDAVLSWLLEYAPSRLTGTGACVFAEFDTESEARQVLEQAPEWLNGFVAKGANLSPLHRAML
SEQ ID NO:36	MRIGHGFDVHAFGGEGPIIIGGVRIPYEKGLLAHSDGDVVLHALTDALL

SEQ ID	SEQUENCE
IspF 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase [Escherichia coli F11] AC# 190908583	GAAALGDIGKLFPDTDPAFKGADSRELLREAWRRIQAKGYALGNVDV TIIAQAPRMLPHIPQMRVFIAEDLGCHMDDVNVKATTTEKLGFTGRGE GIACEAVALLIKATK
SEQ ID NO:37 IspG 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase [Escherichia coli F11] CDU37657	MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQI KALERVGADIVRVSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALK VAEYGVDCLRINPGNIGNEERIRMVVDCARDKNIPIRIGVNAGSLEKDL QEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVFLAVESY RLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLA ADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLE DIITPMDVSIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLD NNDMIDQLEARIRAKASQLDEARRIDVQQVEK
SEQ ID NO:38 IspH 4-hydroxy-3-methylbut-2-enyl diphosphate reductase [Escherichia coli F11] AC# 190905591	MQILLANPRGFCAGVDRAISIVENALAIYGAPIYVRHEVVHNRYVVDS LRERGAIFIEQISEVPDGAILIFSAHGVSQAVRNEAKSRDLTVFDATCPL VTKVHMEVARASRRGEESILIGHAGHPEVEGTMGQYSNPEGGMYLVE SPDDVWKLTVKNEEKLSFMTQTTLSVDDTSDVIDALRKRFPKIVGPRK DDICYATTNRQEAVRALAEQAEVVLVVGSKNSSNSNRLAELAQRMGK RAFLIDDATDIQEEWVKEAKCVGVTAGASAPDILVQNVVARLQQLGG GEAIPLEGREENIVFEVPKELRVDIREVD
SEQ ID NO:39 IDI: isopentenyl diphosphate (IPP) isomerase (GenBank AKF73239) Escherichia coli	MQTEHVILLNAQGVPTGTLEKYAAHTADTRLHLAFSSWLFNAKGQLL VTRRALSKKAWPGVWTNSVCGHPQLGESNEDAVIRRCRYELGVEITPP ESIYPDFRYRATDPSGIVENEVCPVFAARTTSALQINDDEVMDYQWCD LADVLHGIDATPWAFSPWMVMQATNREARKRLSAFTQLK
SEQ ID NO:40 Mutated IspA* FPP synthase (S81F) for GPP production (ispA-AC# NP_414955)	MDFPQQLEACVKQANQALSRFIAPLPFQNTPVVETMQYGALLGGKRL RPFLVYATGHMFGVSTNTLDAPAAAVECIHAYSLIHDDLPAMDDDDL RRGLPTCHVKFGEANAILAGDALQTLAFSILSDADMPEVSDRDRISMIS ELASASGIAGMCGGQALDLDAEGKHVPLDALERIHRHKTGALIRAAV RLGALSAGDKGRRALPVLDKYAESIGLAFQVQDDILDVVGDTATLGK RQGADQQLGKSTYPALLGLEQARKKARDLIDDARQSLKQLAEQSLDT SALEALADYIIQRNK
SEQ ID NO:41 Malonyl CoA-acyl carrier protein transacylase (MCT1) Saccharomyces cerevisiae	atgaagctactaaccttcccaggtcaagggacctccatctccatttcgatattaaaagcgataataagaaacaaat caagagaattccaaacaatactgagtcagaacggcaaggaatcaaatgatctattgcagtacatcttccagaacc cttccagccccggaagcattgcagtctgctccaacctttctatcaattgtaccagatactctcgaatccttctgatcc tcaagatcaagcaccaaaaaatatgactaagatcgattcccccgacaagaaagacaatgaacaatgttacctttg ggtcactcgctaggcgagttaacatgtctgagtgttaattcactgttttcgttaaaggatctttttgatattgctaatttta gaaataagttaatggtaacatctactgaaaagtacttagtagcccacaatatcaacagatccaacaaatttgaaatg tgggcactctcttctccgagggccacagatttaccgcaagaagtgcaaaaactactaaattcccctaatttattatc atcttcacaaaataccatttctgtagcaaatgcaaattcagtaaagcaatgtgtagtcaccggtctggttgatgattta gagtccttaagaacagaattgaacttaaggttcccgcgtttaagaattacagaattaactaacccatacaacatccc cttccataatagcactgtgttgaggcccgttcaggaaccactctatgactacatttgggatatattaaagaaaaacg gaactcacacgttgatggagttgaaccatccaataatagctaacttagatggtaatatatcttactatattcatcatgc cctagatagattcgttaagtgttcaagcaggactgtgcaattcaccatgtgttatgataccataaactctggaaccc cagtggaaattgataagagtatttgctttggcccgggcaatgtgatttataaccttattcggagaaattgtccccaag tggacactatagaatacacctctttagcaactatagacgcttatcacaaggcggcagaggagaacaaagattga
SEQ ID NO:42 malonyl CoA-acyl carrier protein	MKLLTFPGQGTSISISILKAIIRNKSREFQTILSQNGKESNDLLQYIFQNPS SPGSIAVCSNLFYQLYQILSNPSDPQDQAPKNMTKIDSPDKKDNEQCYL LGHSLGELTCLSVNSLFSLKDLFDIANFRNKLMVTSTEKYLVAHNINRS

SEQ ID	SEQUENCE
transacylase (MCT1) Saccharomyces cerevisiae	NKFEMWALSSPRATDLPQEVQKLLNSPNLLSSSQNTISVANANSVKQC VVTGLVDDLESLRTELNLRFPRLRITELTNPYNIPFHNSTVLRPVQEPLY DYIWDILKKNGTHTLMELNHPIIANLDGNISYYIHHALDRFVKCSSRTV QFTMCYDTINSGTPVEIDKSICFGPGNVIYNLIRRNCPQVDTIEYTSLATI DAYHKAAEENKD*
SEQ ID NO:43 Artificial beta-ketothiolase (BktB) nucleotide sequence	atgactagagaagttgtcgtcgtttccggtgtccgtaccgctatcggtactttcggtggttccttaaaggatgttgct cctgctgaattgggtgctttagttgttagagaagctttggccagagcccaagtctccggtgacgacgttggtcacg tcgttttcggtaacgtcatccaaactgaaccacgtgacatgtacttgggtagagtcgccgctgttaacggtggtgtc accatcaacgctcctgccttaactgttaacagattatgtggttccggtttacaagctattgtctctgccgcccaaact atcttgttgggtgatactgacgttgctattggtggtggtgctgaatctatgtctagagctccatacttggctccagctg cccgttggggtgctagaatgggtgacgccggtttggtcgatatgatgttgggtgccttgcatgatcctttccacag aatccacatgggtgttaccgctgaaaacgttgctaaggaatacgatatctctagagctcaacaagatgaagccgc tttagaatctcacagacgtgcctccgccgctattaaggctggttacttcaaggaccaaattgttccagttgtctctaa gggtcgtaaaggtgatgttacctttgatactgacgaacacgttagacacgacgccactattgacgatatgactaaa ttaagaccagtctttgttaaggagaatggtaccgttactgctggtaacgcttctggtttgaacgatgccgccgctgc cgttgttatgatggaaagagctgaagccgaaagacgtggtttaaagccattggccagattagtctcctacggtca cgctggtgtcgacccaaaggctatgggtatcggtccagttcctgctactaagattgctttagaaagagctggtttg caagtttctgacttggacgtcatcgaagccaacgaagccttcgctgctcaagcttgtgctgtcaccaaggctttgg gtttggatccagctaaagttaaccctaatggttctggtatttccttgggtcacccaatcggtgctaccggtgctttaat cactgttaaagccttacacgaattgaacagagttcaaggtagatacgctttggtcactatgtgcatcggtggtggtc aaggtatcgctgctatcttcgaaagaatcggatcctaa
SEQ ID NO:44 Engineered beta-ketothiolase (BktB)	MTREVVVVSGVRTAIGTFGGSLKDVAPAELGALVVREALARAQVSGD DVGHVVFGNVIQTEPRDMYLGRVAAVNGGVTINAPALTVNRLCGSGL QAIVSAAQTILLGDTDVAIGGGAESMSRAPYLAPAARWGARMGDAGL VDMMLGALHDPFHRIHMGVTAENVAKEYDISRAQQDEAALESHRRAS AAIKAGYFKDQIVPVVSKGRKGDVTFDTDEHVRHDATIDDMTKLRPV FVKENGTVTAGNASGLNDAAAAVVMMERAEAERRGLKPLARLVSYG HAGVDPKAMGIGPVPATKIALERAGLQVSDLDVIEANEAFAAQACAV TKALGLDPAKVNPNGSGISLGHPIGATGALITVKALHELNRVQGRYAL VTMCIGGGQGIAAIFERIGS*
SEQ ID NO:45 Artificial PaaH1: 3-hydroxyacyl-CoA dehydrogenase nucleotide sequence	atgtccatcagaactgtcggtattgttggtgctggtactatgggtaacggtattgctcaagcctgtgctgtcgtcggt ttgaacgtcgtcatggtcgacatttctgacgctgctgttcaaaagggtgttgctactgtcgcttcctctttggacagat taattaagaaggaaaagttgaccgaagccgacaaggcctctgccttggccagaattaagggttccacttcttatg acgacttgaaagctaccgacattgttatcgaagctgctactgaaaactacgatttgaaagttaagatcttgaagcaa attgatggtatcgtcggtgagaacgtcattattgcttctaacacttcctccatttctatcactaaattagccgccgtca cctctagagccgacagatttatcggtatgcacttctttaatccagttccagtcatggctttggtcgaattaattagagg tttgcaaacctccgacaccacccacgccgccgttgaagctttgtctaagcaattgggtaagtacccaatcaccgtt aaaaattccccaggtttcgttgtcaaccgtattttgtgcccaatgatcaatgaagctttctgtgtcttgggtgagggtt tggcctccccagaagaaatcgatgaaggtatgaagttaggttgtaaccaccctattggtcctttagccttggccga catgatcggtttagacactatgttggccgttatggaagtcttgtacactgaattcgctgacccaaagtacagaccag ctatgttaatgagagaaatggttgctgccggttatttgggtagaaagactggtcgtggtgtttatgtctactctaaag gatc
SEQ ID NO:46 Engineered PaaH1: 3-hydroxyacyl-CoA dehydrogenase	MSIRTVGIVGAGTMGNGIAQACAVVGLNVVMVDISDAAVQKGVATV ASSLDRLIKKEKLTEADKASALARIKGSTSYDDLKATDIVIEAATENYD LKVKILKQIDGIVGENVIIASNTSSISITKLAAVTSRADRFIGMHFFNPVP VMALVELIRGLQTSDTTHAAVEALSKQLGKYPITVKNSPGFVVNRILCP MINEAFCVLGEGLASPEEIDEGMKLGCNHPIGPLALADMIGLDTMLAV MEVLYTEFADPKYRPAMLMREMVAAGYLGRKTGRGVYVYSKGI
SEQ ID NO:47 Artificial crotonase (Crt) nucleotide	atggaattgaacaacgttatttttggaaaaggaaggtaaggtcgctgtcgttactatcaacagaccaaaggctttaa acgctttgaactctgacaccttgaaagaaatggattatgttatcggtgaaatcgaaaatgactctgaagttttggcc gttatcttgactggtgctggtgaaaaatctttcgttgctggtgctgacatttctgaaatgaaggagatgaataccatt gaaggtagaaagttcggtatcttgggtaacaaggtttttagaagattggaattgttggaaaaaccagtcatcgctg

SEQ ID	SEQUENCE
sequence	ctgttaacggtttcgctttaggtggtggttgtgaaatcgctatgtcctgtgacattcgtatcgcctcctccaatgctag attcggtcaaccagaagttggtttaggtattactccaggtttcggtggtacccaaagattgtctagattggtcggtat gggtatggctaagcaattaattttcactgctcaaaacattaaggctgatgaagccttacgtattggtttggtcaacaa ggtcgttgaaccatctgaattgatgaataccgctaaggaaattgctaacaaaattgtttctaatgccccagttgctgt caagttgtccaagcaagctattaacagaggtatgcaatgtgatattgacactgctttggctttcgaatccgaagcttt tggtgaatgtttttctaccgaagatcaaaaggatgctatgaccgctttcatcgagaagagaaagatcgaaggtttc aaaaacagaggatcctaa
SEQ ID NO:48 Engineered crotonase (Crt)	MELNNVILEKEGKVAVVTINRPKALNALNSDTLKEMDYVIGEIENDSE VLAVILTGAGEKSFVAGADISEMKEMNTIEGRKFGILGNKVFRRLELLE KPVIAAVNGFALGGGCEIAMSCDIRIASSNARFGQPEVGLGITPGFGGT QRLSRLVGMGMAKQLIFTAQNIKADEALRIGLVNKVVEPSELMNTAK EIANKIVSNAPVAVKLSKQAINRGMQCDIDTALAFESEAFGECFSTEDQ KDAMTAFIEKRKIEGFKNRGS*
SEQ ID NO:49 Artificial Ter: trans-2-enoyl-CoA reductase nucleotide sequence	atgattgtcaaaccaatggttcgtaacaacatttgtttaaatgcccacccacaaggttgtaagaagggtgttgaaga tcaaatcgaatacactaaaaagagaattaccgctgaagttaaagctggtgctaaggccccaaagaacgttttggtt ttgggttgttccaacggttacggtttggcctccagaattactgctgcttttggttacggtgccgctaccatcggtgtct ctttcgaaaaggccggttccgaaactaagtacggtactccaggttggtacaataacttggctttcgatgaagctgc taagagagaagaggtttgtattccgttactattgacggtgatgccttttctgacgaaatcaaagctcaagtcatcgaag aagccaaaaagaaaggtatcaagttcgatttgattgtctactctttagcctctcctgttagaactgatccagatactg gtattatgcacaaatccgttttgaagccattcggtaagaccttcactggtaaaactgtcgatcctttcactggtgaatt aaaggaaatctctgctgaacctgccaacgacgaagaagctgctgccactgttaaggttatgggtggtgaagact gggaaagatggatcaagcaattatctaaggaaggtttgttggaagaaggttgtatcaccttggcttactcttacatc ggtccagaagctacccaagctttgtacagaaagggtaccattggtaaggctaaagaacacttggaggctactgc tcatagattgaacaaggaaaatccatccatcagagcctttgtttccgtcaataaaggtttggtcactagagcctctg ccgtcattccagttatccctttatacttggcttctttgtttaaagtcatgaaggaaaagggtaaccatgaaggttgtat cgaacaaatcactcgtttgtacgctgaacgtttatacagaaaggacggtaccatccctgtcgatgaagaaaacag aatcagaatcgacgattgggaattggaagaagatgttcaaaaagccgtttccgccttgatggaaaaggtcaccg gtgaaaatgccgaatccttgactgacttagctggttacagacatgacttttagcttctaatggtttcgatgttgaagg tattaactatgaggctgaagtcgaaagatttgacagaatcggatcctaa
SEQ ID NO:50 Engineered Ter: trans-2-enoyl-CoA reductase	MIVKPMVRNNICLNAHPQGCKKGVEDQIEYTKKRITAEVKAGAKAPK NVLVLGCSNGYGLASRITAAFGYGAATIGVSFEKAGSETKYGTPGWY NNLAFDEAAKREGLYSVTIDGDAFSDEIKAQVIEEAKKKGIKFDLIVYS LASPVRTDPDTGIMHKSVLKPFGKTFTGKTVDPFTGELKEISAEPANDE EAAATVKVMGGEDWERWIKQLSKEGLLEEGCITLAYSYIGPEATQAL YRKGTIGKAKEHLEATAHRLNKENPSIRAFVSVNKGLVTRASAVIPVIP LYLASLFKVMKEKGNHEGCIEQITRLYAERLYRKDGTIPVDEENRIRID DWELEEDVQKAVSALMEKVTGENAESLTDLAGYRHDFLASNGFDVE GINYEAEVERFDRIGS*
SEQ ID NO:51 Truncated 3-hydroxy-3-methyl-glutaryl-CoA reductase (tHMG1)	atggcgcgtgaccaattggtgaaaactgaagtcaccaagaagtcttttactgctcctgtacaaaaggcttctacac cagttttaaccaataaaacagtcatttctggatcgaaagtcaaaagtttatcatctgcgcaatcgagctcatcagga ccttcatcatctagtgaggaagatgattcccgcgatattgaaagcttggataagaaaatacgtcctttagaagaatt agaagcattattaagtagtggaaatacaaaacaattgaagaacaaagaggtcgctgccttggttattcacggtaa gttacctttgtacgctttggagaaaaaattaggtgatactacgagagcggttgcggtacgtaggaaggctctttca attttggcagaagctcctgtattagcatctgatcgtttaccatataaaaattatgactacgaccgcgtatttggcgctt gttgtgaaaatgttataggttacatgcctttgcccgttggtgttataggccccttggttatcgatggtacatcttatcat ataccaatggcaactacagagggttgtttggtagcttctgccatgcgtggctgtaaggcaatcaatgctggcggt ggtgcaacaactgttttaactaaggatggtatgacaagaggcccagtagtccgtttcccaactttgaaaagatctg gtgcctgtaagatatggttagactcagaagagggacaaaacgcaattaaaaaagcttttaactctacatcaagattt gcacgtctgcaacatattcaaacttgtctagcaggagatttgtt
SEQ ID NO:52 Truncated 3-hydroxy-3-methyl-glutaryl-	MARDQLVKTEVTKKSFTAPVQKASTPVLTNKTVISGSKVKSLSSAQSS SSGPSSSSEEDDSRDIESLDKKIRPLEELEALLSSGNTKQLKNKEVAALV IHGKLPLYALEKKLGDTTRAVAVRRKALSILAEAPVLASDRLPYKNYD

SEQ ID	SEQUENCE
CoA reductase (tHMG1)	YDRVFGACCENVIGYMPLPVGVIGPLVIDGTSYHIPMATTEGCLVASA MRGCKAINAGGGATTVLTKDGMTRGPVVRFPTLKRSGACKIWLDSEE GQNAIKKAFNSTSRFARLQHIQTCLAGDL
SEQ ID NO:53 Artificial acetyl-CoA acetyltransferase/ HMG-CoA reductase (mvaE) nucleotide sequence	atgaagactgtcgttatcatagatgccttgagaacaccaatcggtaaatacaaaggttcattatcccaagtttccgc cgttgacttaggtactcatgttactacacaattgttgaagagacactccacaatcagtgaagaaatcgatcaagtca tattcggtaacgtattgcaagctggtaatggtcaaaacccagccagacaaatagctatcaattctggtttatcacat gaaattcctgctatgacagtaaacgaagtttgtggttcaggcatgaaagcagtcattttggccaagcaattgataca attaggtgaagcagaagttttaatcgccggtggtatagaaaacatgagtcaagctccaaaattgcaaagattcaat tacgaaactgaatcttacgatgcacctttctcttcgatgatgtatgatggtttgactgacgctttttctggtcaagcaat gggtttaacagctgaaaatgtcgcagaaaagtaccatgtaaccagagaagaacaagatcaattttccgttcacag tcaattaaaagctgcacaagcacaagccgaaggtattttcgccgacgaaatagctccattggaagtttctggtaca ttagtcgaaaaggatgaaggtattagacctaactccagtgttgaaaaattgggtactttgaagacagtattcaagg aagacggtacagttaccgctggtaatgcctctaccattaacgatggtgctagtgcattgattatagcttctcaagaa tatgccgaagctcatggtttgccatacttagctatcattagagatagtgtagaagttggtattgacccagcatacatg ggtatctctcctataaaagcaatccaaaagttgttagccagaaaccaattgaccactgaagaaattgatttgtacga aattaacgaagcatttgccgctacatcaatcgttgtccaaagagaattggcattgccagaagaaaaggttaacatc tatggtggtggtatctccttgggtcacgctataggtgcaaccggtgccagattgttgacttccttaagttaccaattg aaccaaaaggaaaagaaatacggtgttgcttctttatgcattggtggtggtttgggtttagcaatgttgttagaaaga ccacaacaaaagaaaaattctagattctaccaaatgtcccctgaagaaagattggcctcattgttaaatgaaggtc aaatttccgcagatactaagaaagaatttgaaaacaccgctttatcttcacaaatcgcaaaccatatgatcgaaaac caaatctctgaaacagaagttccaatgggtgtcggtttgcacttaactgtcgatgaaacagactatttggtaccaat ggctaccgaagaacctagtgttatcgcagccttatctaatggtgctaagatagcacaaggttttaagactgttaacc aacaaagattgatgagaggtcaaatcgtattctacgatgttgctgacccagaatcattaatcgataagttgcaagta agagaagccgaagtttttcaacaagctgaattgtcttacccttcaatagttaagagaggtggtggtttgagagattt gcaatacagaacttttgacgaatccttcgtcagtgtagatttcttagttgatgtcaaggacgccatgggtgctaatat tgttaacgcaatgttggaaggtgtcgccgaattgtttagagaatggttcgctgaacaaaagattttgttttctatcttgt caaactacgctacagaatctgtagttaccatgaaaactgcaattccagtttccagattgagtaagggttctaacggt agagaaatcgctgaaaagattgttttggcatcaagatatgcctccttagacccttacagagctgttactcataataa gggtataatgaacggtatcgaagctgtcgtattagcaaccggtaatgatactagagcagtatctgcctcatgtcac gcattcgccgttaaggaaggtagataccaaggtttgacatcatggaccttggatggtgaacaattaattggtgaaa tatccgttccattggctttagcaactgttggtggtgctacaaaagtcttgcctaagagtcaagctgcagccgatttgt tagccgtcactgacgctaaggaattgtctagagttgtcgctgcagtaggtttagctcaaaatttggccgctttaaga gcattggtttcagaaggtattcaaaaaggtcatatggctttgcaagcaagatccttagccatgacagttggtgctac cggtaaagaagtcgaagccgtagctcaacaattaaaaagacaaaagacaatgaaccaagacagagcaatggc tatattaaacgatttgagaaagcaataa
SEQ ID NO:54 MvaE: acetyl-CoA acetyltransferase/ HMG-CoA reductase (Enterococcus sp.)	MKTVVIIDALRTPIGKYKGSLSQVSAVDLGTHVTTQLLKRHSTISEEID QVIFGNVLQAGNGQNPARQIAINSGLSHEIPAMTVNEVCGSGMKAVIL AKQLIQLGEAEVLIAGGIENMSQAPKLQRFNYETESYDAPFSSMMYDG LTDAFSGQAMGLTAENVAEKYHVTREEQDQFSVHSQLKAAQAQAEGI FADEIAPLEVSGTLVEKDEGIRPNSSVEKLGTLKTVFKEDGTVTAGNAS TINDGASALIIASQEYAEAHGLPYLAIIRDSVEVGIDPAYMGISPIKAIQK LLARNQLTTEEIDLYEINEAFAATSIVVQRELALPEEKVNIYGGGISLGH AIGATGARLLTSLSYQLNQKEKKYGVASLCIGGGLGLAMLLERPQQK KNSRFYQMSPEERLASLLNEGQISADTKKEFENTALSSQIANHMIENQI SETEVPMGVGLHLTVDETDYLVPMATEEPSVIAALSNGAKIAQGFKTV NQQRLMRGQIVFYDVADPESLIDKLQVREAEVFQQAELSYPSIVKRGG GLRDLQYRTFDESFVSVDFLVDVKDAMGANIVNAMLEGVAELFREWF AEQKILFSILSNYATESVVTMKTAIPVSRLSKGSNGREIAEKIVLASRYA SLDPYRAVTHNKGIMNGIEAVVLATGNDTRAVSASCHAFAVKEGRYQ GLTSWTLDGEQLIGEISVPLALATVGGATKVLPKSQAAADLLAVTDAK ELSRVVAAVGLAQNLAALRALVSEGIQKGHMALQARSLAMTVGATG KEVEAVAQQLKRQKTMNQDRAMAILNDLRKQ*
SEQ ID NO:55	atgacaattgggattgataaaattagtttttttgtgccccttattatattgatatgacggcactggctgaagccagaa

SEQ ID	SEQUENCE
MvaS: HMG-CoA synthase (Enterococcus sp.)	atgtagaccctggaaaatttcatattggtattgggcaagaccaaatggcggtgaacccaatcagccaagatattgt gacatttgcagccaatgccgcagaagcgatcttgaccaaagaagataaagaggccattgatatggtgattgtcg ggactgagtccagtatcgatgagtcaaaagcggccgcagttgtcttacatcgtttaatggggattcaacctttcgct cgctctttcgaaatcaaggaagcttgttacggagcaacagcaggcttacagttagctaagaatcacgtagccttac atccagataaaaaagtcttggtcgtagcggcagatattgcaaaatatggcttaaattctggcggtgagcctacaca aggagctggggcggttgcaatgttagttgctagtgaaccgcgcattttggctttaaaagaggataatgtgatgctg acgcaagatatctatgacttttggcgtccaacaggccacccgtatcctatggtcgatggtcctttgtcaaacgaaa cctacatccaatcttttgcccaagtctgggatgaacataaaaaacgaaccggtcttgattttgcagattatgatgcttt agcgttccatattccttacacaaaaatgggcaaaaaagccttattagcaaaaatctccgaccaaactgaagcaga acaggaacgaattttagcccgttatgaagaaagtatcgtctatagtcgtcgcgtaggaaacttgtatacgggttca ctttatctgggactcatttccctttagaaaatgcaacgactttaaccgcaggcaatcaaattggtttattcagttatgg ttctggtgctgtcgctgaatttttcactggtgaattagtagctggttatcaaaatcatttacaaaaagaaactcatttag cactgctggataatcggacagaactttctatcgctgaatatgaagccatgtttgcagaaactttagacacagacatt gatcaaacgttagaagatgaattaaaatatagtatttctgctattaataataccgttcgttcttatcgaaactaa
SEQ ID NO:56 MvaS: HMG-CoA synthase (Enterococcus sp.)	MTIGIDKISFFVPPYYIDMTALAEARNVDPGKFHIGIGQDQMAVNPISQ DIVTFAANAAEAILTKEDKEAIDMVIVGTESSIDESKAAAVVLHRLMGI QPFARSFEIKEACYGATAGLQLAKNHVALHPDKKVLVVAADIAKYGL NSGGEPTQGAGAVAMLVASEPRILALKEDNVMLTQDIYDFWRPTGHP YPMVDGPLSNETYIQSFAQVWDEHKKRTGLDFADYDALAFHIPYTKM GKKALLAKISDQTEAEQERILARYEESIVYSRRVGNLYTGSLYLGLISLL ENATTLTAGNQIGLFSYGSGAVAEFFTGELVAGYQNHLQKETHLALLD NRTELSIAEYEAMFAETLDTDIDQTLEDELKYSISAINNTVRSYRN*
SEQ ID NO:57 Isopentenyl pyrophosphate isomerase (Sc_IDI1) Saccharomyces sp.	atgactgccgacaacaatagtatgccccatggtgcagtatctagttacgccaaattagtgcaaaaccaaacacct gaagacattttggaagagtttcctgaaattattccattacaacaaagacctaatacccgatctagtgagacgtcaaa tgacgaaagcggagaaacatgtttttctggtcatgatgaggagcaaattaagttaatgaatgaaaattgtattgtttt ggattgggacgataatgctattggtgccggtaccaagaaagtttgtcatttaatggaaaatattgaaaagggtttac tacatcgtgcattctccgtctttattttcaatgaacaaggtgaattacttttacaacaaagagccactgaaaaaataac tttccctgatctttggactaacacatgctgctctcatccactatgtattgatgacgaattaggtttgaagggtaagcta gacgataagattaagggcgctattactgcggcggtgagaaaactagatcatgaattaggtattccagaagatgaa actaagacaaggggtaagtttcacttttaaacagaatccattacatggcaccaagcaatgaaccatggggtgaa catgaaattgattacatcctattttataagatcaacgctaaagaaaacttgactgtcaacccaaacgtcaatgaagtt agagacttcaaatgggtttcaccaaatgatttgaaaactatgtttgctgacccaagttacaagtttacgccttggttta agattatttgcgagaattacttattcaactggtgggagcaattagatgacctttctgaagtggaaaatgacaggcaa attcatagaatgctataa
SEQ ID NO:58 Isopentenyl pyrophosphate isomerase (Sc_IDI1) Saccharomyces sp.	MTADNNSMPHGAVSSYAKLVQNQTPEDILEEFPEIIPLQQRPNTRSSET SNDESGETCFSGHDEEQIKLMNENCIVLDWDDNAIGAGTKKVCHLME NIEKGLLHRAFSVFIFNEQGELLLQQRATEKITFPDLWTNTCCSHPLCID DELGLKGKLDDKIKGAITAAVRKLDHELGIPEDETKTRGKFHFLNRIHY MAPSNEPWGEHEIDYILFYKINAKENLTVNPNVNEVRDFKWVSPNDLK TMFADPSYKFTPWFKIICENYLFNWWEQLDDLSEVENDRQIHRML
SEQ ID NO:59 Mutant farnesyl pyrophosphate synthase (Erg20mut, F96W, N127W)	atggcttcagaaaaagaaattaggagagagagattcttgaacgttttccctaaattagtagaggaattgaacgcat cgcttttggcttacggtatgcctaaggaagcatgtgactggtatgcccactcattgaactacaacactccaggcgg taagctaaatagaggtttgtccgttgtggacacgtatgctattctctccaacaagaccgttgaacaattggggcaa gaagaatacgaaaaggttgccattctaggttggtgcattgagttgttgcaggcttactggttggtcgccgatgatat gatggacaagtccattaccagaagaggccaaccatgttggtacaaggttcctgaagttggggaaattgccatctg ggacgcattcatgttagaggctgctatctacaagcttttgaaatctcacttcagaaacgaaaaatactacatagatat caccgaattgttccatgaggtcaccttccaaaccgaattgggccaattgatggacttaatcactgcacctgaagac aaagtcgacttgagtaagttctccctaaagaagcactccttcatagttactttcaagactgcttactattctttctactt gcctgtcgcattggccatgtacgttgccggtatcacggatgaaaaggatttgaaacaagccagagatgtcttgatt ccattgggtgaatacttccaaattcaagatgactacttagactgcttcggtaccccagaacagatcggtaagatcg gtacagatatccaagataacaaatgttcttgggtaatcaacaaggcattggaacttgcttccgcagaacaaagaa agactttagacgaaaattacggtaagaaggactcagtcgcagaagccaaatgcaaaaagattttcaatgacttga aaattgaacagctataccacgaatatgaagagtctattgccaaggatttgaaggccaaaatttctcaggtcgatga

SEQ ID	SEQUENCE
	gtctcgtggcttcaaagctgatgtcttaactgcgttcttgaacaaagtttacaagagaagcaaatag
SEQ ID NO:60 Mutant farnesyl pyrophosphate synthase (Erg20mut, F96W, N127W)	MASEKEIRRERFLNVFPKLVEELNASLLAYGMPKEACDWYAHSLNYN TPGGKLNRGLSVVDTYAILSNKTVEQLGQEEYEKVAILGWCIELLQAY WLVADDMMDKSITRRGQPCWYKVPEVGEIAIWDAFMLEAAIYKLLKS HFRNEKYYIDITELFHEVTFQTELGQLMDLITAPEDKVDLSKFSLKKHS FIVTFKTAYYSFYLPVALAMYVAGITDEKDLKQARDVLIPLGEYFQIQD DYLDCFGTPEQIGKIGTDIQDNKCSWVINKALELASAEQRKTLDENYG KKDSVAEAKCKKIFNDLKIEQLYHEYEESIAKDLKAKISQVDESRGFKA DVLTAFLNKVYKRSK*
SEQ ID NO:61 Phosphomevalonate kinase (Sc_ERG8) Saccharomyces sp.	atgtcagagttgagagccttcagtgccccagggaaagcgttactagctggtggatatttagttttagatacaaaata tgaagcatttgtagtcggattatcggcaagaatgcatgctgtagcccatccttacggttcattgcaagggtctgata agtttgaagtgcgtgtgaaaagtaaacaatttaaagatggggagtggctgtaccatataagtcctaaaagtggctt cattcctgtttcgataggcggatctaagaacccttttcattgaaaaagttatcgctaacgtatttagctactttaaacct acatggacgactactgcaatagaaacttgttcgttattgatattttctctgatgatgcctaccattctcaggaggatag cgttaccgaacatcgtggcaacagaagattgagttttcattcgcacagaattgaagaagttcccaaaacagggct gggctcctcggcaggtttagtcacagttttaactacagctttggcctccttttttgtatcggacctggaaaataatgta gacaaatatagagaagttattcataatttagcacaagttgctcattgtcaagctcagggtaaaattggaagcgggtt tgatgtagcggcggcagcatatggatctatcagatatagaagattcccacccgcattaatctctaatttgccagata ttggaagtgctacttacggcagtaaactggcgcatttggttgatgaagaagactggaatattacgattaaaagtaa ccatttaccttcgggattaactttatggatgggcgatattaagaatggttcagaaacagtaaaactggtccagaag gtaaaaaattggtatgattcgcatatgccagaaagcttgaaaatatatacagaactcgatcatgcaaattctagattt atggatggactatctaaactagatcgcttacacgagactcatgacgattacagcgatcagatatttgagtctcttga gaggaatgactgtacctgtcaaaagtatcctgaaatcacagaagttagagatgcagttgccacaattagacgttcc tttagaaaaataactaaagaatctggtgccgatatcgaacctcccgtacaaactagcttattggatgattgccagac cttaaaaggagttcttacttgcttaatacctggtgctggtggttatgacgccattgcagtgattactaagcaagatgtt gatcttagggctcaaaccgctaatgacaaaagattttctaaggttcaatggctggatgtaactcaggctgactggg gtgttaggaaagaaaaagatccggaaacttatcttgataaataa
SEQ ID NO:62 Phosphomevalonate kinase (Sc_ERG8) Saccharomyces sp.	MSELRAFSAPGKALLAGGYLVLDTKYEAFVVGLSARMHAVAHPYGSL QGSDKFEVRVKSKQFKDGEWLYHISPKSGFIPVSIGGSKNPFIEKVIANV FSYFKPNMDDYCNRNLFVIDIFSDDAYHSQEDSVTEHRGNRRLSFHSH RIEEVPKTGLGSSAGLVTVLTTALASFFVSDLENNVDKYREVIHNLAQ VAHCQAQGKIGSGFDVAAAAYGSIRYRRFPPALISNLPDIGSATYGSKL AHLVDEEDWNITIKSNHLPSGLTLWMGDIKNGSETVKLVQKVKNWYD SHMPESLKIYTELDHANSRFMDGLSKLDRLHETHDDYSDQIFESLERN DCTCQKYPEITEVRDAVATIRRSFRKITKESGADIEPPVQTSLLDDCQTL KGVLTCLIPGAGGYDAIAVITKQDVDLRAQTANDKRFSKVQWLDVTQ ADWGVRKEKDPETYLDK*
SEQ ID NO:63 ERG12 – mevalonate kinase (Saccharomyces sp.)	atgtcattaccgttcttaacttctgcaccgggaaaggttattatttttggtgaacactctgctgtgtacaacaagcctg ccgtcgctgctagtgtgtctgcgttgagaacctacctgctaataagcgagtcatctgcaccagatactattgaattg gacttcccggacattagctttaatcataagtggtccatcaatgatttcaatgccatcaccgaggatcaagtaaactc ccaaaaattggccaaggctcaacaagccaccgatggcttgtctcaggaactcgttagtcttttggatccgttgttag ctcaactatccgaatccttccactaccatgcagcgttttgtttcctgtatatgtttgtttgcctatgcccccatgccaag aatattaagttttctttaaagtctactttacccatcggtgctgggttgggctcaagcgcctctatttctgtatcactggc cttagctatggcctacttgggggggttaataggatctaatgacttggaaaagctgtcagaaaacgataagcatata gtgaatcaatgggccttcataggtgaaaagtgtattcacggtacccccttcaggaatagataacgctgtggccactt atggtaatgccctgctatttgaaaaagactcacataatggaacaataaacacaaacaatttttaagttcttagatgattt cccagccattccaatgatcctaacctatactagaattccaaggtctacaaaagatcttgttgctcgcgttcgtgtgtt ggtcaccgagaaatttcctgaagttatgaagccaattctagatgccatgggtgaatgtgccctacaaggcttagag atcatgactaagttaagtaaatgtaaaggcaccgatgacgaggctgtagaaactaataatgaactgtatgaacaa ctattggaattgataagaataaatcatggactgcttgtctcaatcggtgtttctcatcctggattagaacttattaaaaa tctgagcgatgatttgagaattggctccacaaaacttaccggtgctggtggcggcggttgctctttgactttgttac gaagagacattactcaagagcaaattgacagcttcaaaaagaaattgcaagatgattttagttacgagacatttga aacagacttgggtgggactggctgctgtttgttaagcgcaaaaaatttgaataaagatcttaaaatcaaatccctag

SEQ ID	SEQUENCE
	tattccaattatttgaaaataaaactaccacaaagcaacaaattgacgatctattattgccaggaaacacgaatttac catggacttcataa
SEQ ID NO:64 ERG12 – mevalonate kinase (Saccharomyces sp.)	MSLPFLTSAPGKVIIFGEHSAVYNKPAVAASVSALRTYLLISESSAPDTI ELDFPDISFNHKWSINDFNAITEDQVNSQKLAKAQQATDGLSQELVSLL DPLLAQLSESFHYHAAFCFLYMFVCLCPHAKNIKFSLKSTLPIGAGLGS SASISVSLALAMAYLGGLIGSNDLEKLSENDKHIVNQWAFIGEKCIHGT PSGIDNAVATYGNALLFEKDSHNGTINTNNFKFLDDFPAIPMILTYTRIP RSTKDLVARVRVLVTEKFPEVMKPILDAMGECALQGLEIMTKLSKCK GTDDEAVETNNELYEQLLELIRINHGLLVSIGVSHPGLELIKNLSDDLRI GSTKLTGAGGGGCSLTLLRRDITQEQIDSFKKKLQDDFSYETFETDLGG TGCCLLSAKNLNKDLKIKSLVFQLFENKTTTKQQIDDLLLPGNTNLPW TS*
SEQ ID NO:65 Mevalonate pyrophosphate decarboxylase (Sc_ERG19) Saccharomyces sp.	atgaccgtttacacagcatccgttaccgcacccgtcaacatcgcaacccttaagtattgggggaaaagggacac gaagttgaatctgcccaccaattcgtccatatcagtgactttatcgcaagatgacctcagaacgttgacctctgcg gctactgcacctgagtttgaacgcgacactttgtggttaaatggagaaccacacagcatcgacaatgaaagaact caaaattgtctgcgcgacctacgccaattaagaaaggaaatggaatcgaaggacgcctcattgcccacattatct caatggaaactccacattgtctccgaaaataactttcctacagcagctggtttagcttcctccgctgctggctttgct gcattggtctctgcaattgctaagttataccaattaccacagtcaacttcagaaatatctagaatagcaagaaaggg gtctggttcagcttgtagatcgttgtttggcggatacgtggcctgggaaatgggaaaagctgaagatggtcatgat tccatggcagtacaaatcgcagacagctctgactggcctcagatgaaagcttgtgtcctagttgtcagcgatatta aaaaggatgtgagttccactcagggtatgcaattgaccgtggcaacctccgaactatttaaagaaagaattgaac atgtcgtaccaaagagatttgaagtcatgcgtaaagccattgttgaaaaagatttcgccacctttgcaaaggaaac aatgatggattccaactctttccatgccacatgtttggactctttccctccaatattctacatgaatgacacttccaag cgtatcatcagttggtgccacaccattaatcagttttacggagaaacaatcgttgcatacacgtttgatgcaggtcc aaatgctgtgttgtactacttagctgaaaatgagtcgaaactctttgcatttatctataaattgtttggctctgttcctgg atgggacaagaaatttactactgagcagcttgaggctttcaaccatcaatttgaatcatctaactttactgcacgtga attggatcttgagttgcaaaaggatgttgccagagtgattttaactcaagtcggttcaggcccacaagaaacaaac gaatctttgattgacgcaaagactggtctaccaaaggaataa
SEQ ID NO:66 Mevalonate pyrophosphate decarboxylase (Sc_ERG19) Saccharomyces sp.	MTVYTASVTAPVNIATLKYWGKRDTKLNLPTNSSISVTLSQDDLRTLT SAATAPEFERDTLWLNGEPHSIDNERTQNCLRDLRQLRKEMESKDASL PTLSQWKLHIVSENNFPTAAGLASSAAGFAALVSAIAKLYQLPQSTSEI SRIARKGSGSACRSLFGGYVAWEMGKAEDGHDSMAVQIADSSDWPQ MKACVLVVSDIKKDVSSTQGMQLTVATSELFKERIEHVVPKRFEVMR KAIVEKDFATFAKETMMDSNSFHATCLDSFPPIFYMNDTSKRIISWCHT INQFYGETIVAYTFDAGPNAVLYYLAENESKLFAFIYKLFGSVPGWDK KFTTEQLEAFNHQFESSNFTARELDLELQKDVARVILTQVGSGPQETNE SLIDAKTGLPKE*
SEQ ID NO:67 Engineered phosphomevalonate kinase/mevalonate kinase (Erg8-T2A-Erg12)	atgtcagagttgagagccttcagtgccccagggaaagcgttactagctggtggatatttagttttagatacaaaata tgaagcatttgtagtcggattatcggcaagaatgcatgctgtagcccatccttacggttcattgcaagggtctgata agtttgaagtgcgtgtgaaaagtaaacaatttaaagatggggagtggctgtaccatataagtcctaaaagtggctt cattcctgtttcgataggcggatctaagaaccctttcattgaaaaagttatcgctaacgtatttagctactttaaaccta acatggacgactactgcaatagaaacttgttcgttattgatattttctctgatgatgcctaccattctcaggaggatag cgttaccgaacatcgtggcaacagaagattgagttttcattcgcacagaattgaagaagttcccaaaacagggct gggctcctcggcaggtttagtcacagttttaactacagctttggcctccttttttgtatcggacctggaaaataatgta gacaaatatagagaagttattcataatttagcacaagttgctcattgtcaagctcagggtaaaattggaagcgggtt tgatgtagcggcggcagcatatggatctatcagatatagaagattcccacccgcattaatctctaatttgccagata ttggaagtgctacttacggcagtaaactggcgcatttggttgatgaagaagactggaatattacgattaaaagtaa ccatttaccttcgggattaactttatggatgggcgatattaagaatggttcagaaacagtaaaactggtccagaag gtaaaaaattggtatgattcgcatatgccagaaagcttgaaaatatatacagaactcgatcatgcaaattctagattt atggatggactatctaaactagatcgcttacacgagactcatgacgattacagcgatcagatatttgagtctcttga gaggaatgactgtacctgtcaaaagtatcctgaaatcacagaagttagagatgcagttgccacaattagacgttcc tttagaaaaataactaaagaatctggtgccgatatcgaacctcccgtacaaactagcttattggatgattgccagac cttaaaaggagttcttacttgcttaatacctggtgctggtggttatgacgccattgcagtgattactaagcaagatgtt

SEQ ID	SEQUENCE
	gatcttagggctcaaaccgctaatgacaaaagattttctaaggttcaatggctggatgtaactcaggctgactggg gtgttaggaaagaaaaagatccggaaacttatcttgataaaaagcttgagggcagaggaagtcttctaacatgcg gtgacgtggaggagaatcccggccctgctagcatgtcattaccgttcttaacttctgcaccgggaaaggttattatt tttggtgaacactctgctgtgtacaacaagcctgccgtcgctgctagtgtgtctgcgttgagaacctacctgctaat aagcgagtcatctgcaccagatactattgaattggacttcccggacattagctttaatcataagtggtccatcaatg atttcaatgccatcaccgaggatcaagtaaactcccaaaaattggccaaggctcaacaagccaccgatggcttgt ctcaggaactcgttagtcttttggatccgttgttagctcaactatccgaatccttccactaccatgcagcgttttgtttc ctgtatatgtttgtttgcctatgcccccatgccaagaatattaagttttctttaaagtctactttacccatcggtgctggg ttgggctcaagcgcctctatttctgtatcactggccttagctatggcctacttgggggggttaataggatctaatgac ttggaaaagctgtcagaaaacgataagcatatagtgaatcaatgggccttcataggtgaaaagtgtattcacggta ccccttcaggaatagataacgctgtggccacttatggtaatgccctgctatttgaaaaagactcacataatggaac aataaacacaaacaattttaagttcttagatgatttcccagccattccaatgatcctaacctatactagaattccaagg tctacaaaagatcttgttgctcgcgttcgtgtgttggtcaccgagaaatttcctgaagttatgaagccaattctagat gccatgggtgaatgtgccctacaaggcttagagatcatgactaagttaagtaaatgtaaaggcaccgatgacga ggctgtagaaactaataatgaactgtatgaacaactattggaattgataagaataaatcatggactgcttgtctcaat cggtgtttctcatcctggattagaacttattaaaaatctgagcgatgatttgagaattggctccacaaaacttaccgg tgctggtggcggcggttgctctttgactttgttacgaagagacattactcaagagcaaattgacagcttcaaaaag aaattgcaagatgattttagttacgagacatttgaaacagacttgggtgggactggctgctgtttgttaagcgcaaa aaatttgaataaagatcttaaaatcaaatccctagtattccaattatttgaaaataaaactaccacaaagcaacaaatt gacgatctattattgccaggaaacacgaatttaccatggacttcataa
SEQ ID NO:68 Engineered phosphomevalonate kinase/mevalonate kinase (Erg8-T2A-Erg12)	MSELRAFSAPGKALLAGGYLVLDTKYEAFVVGLSARMHAVAHPYGSL QGSDKFEVRVKSKQFKDGEWLYHISPKSGFIPVSIGGSKNPFIEKVIANV FSYFKPNMDDYCNRNLFVIDIFSDDAYHSQEDSVTEHRGNRRLSFHSH RIEEVPKTGLGSSAGLVTVLTTALASFFVSDLENNVDKYREVIHNLAQ VAHCQAQGKIGSGFDVAAAAYGSIRYRRFPPALISNLPDIGSATYGSKL AHLVDEEDWNITIKSNHLPSGLTLWMGDIKNGSETVKLVQKVKNWYD SHMPESLKIYTELDHANSRFMDGLSKLDRLHETHDDYSDQIFESLERN DCTCQKYPEITEVRDAVATIRRSFRKITKESGADIEPPVQTSLLDDCQTL KGVLTCLIPGAGGYDAIAVITKQDVDLRAQTANDKRFSKVQWLDVTQ ADWGVRKEKDPETYLDKKLEGRGSLLTCGDVEENPGPASMSLPFLTS APGKVIIFGEHSAVYNKPAVAASVSALRTYLLISESSAPDTIELDFPDISF NHKWSINDFNAITEDQVNSQKLAKAQQATDGLSQELVSLLDPLLAQLS ESFHYHAAFCFLYMFVCLCPHAKNIKFSLKSTLPIGAGLGSSASISVSLA LAMAYLGGLIGSNDLEKLSENDKHIVNQWAFIGEKCIHGTPSGIDNAV ATYGNALLFEKDSHNGTINTNNFKFLDDFPAIPMILTYTRIPRSTKDLV ARVRVLVTEKFPEVMKPILDAMGECALQGLEIMTKLSKCKGTDDEAV ETNNELYEQLLELIRINHGLLVSIGVSHPGLELIKNLSDDLRIGSTKLTG AGGGGCSLTLLRRDITQEQIDSFKKKLQDDFSYETFETDLGGTGCCLLS AKNLNKDLKIKSLVFQLFENKTTTKQQIDDLLLPGNTNLPWTS*
SEQ ID NO:69 Artificial neryl pyrophosphate (NPP) synthase (NPPS) nucleotide sequence	atgtgctcacttaatttgcaaacggaaaagctatgctatgaagacaatgacaatgacttggacgaggaactgatg ccgaagcacatagcgctaatcatggatggtaatagacgttgggcaaaagacaagggcttagaagtgtacgaag ggcacaaacatataatcccgaaactaaaagaaatatgtgacatatcctccaagttggggattcagatcatcacag cgttcgcgttctccacagagaactggaagagatccaaggaggaagtcgatttcctattgcagatgtttgaagaaat ctatgacgaatttagccgttctggggtgagagtgagtatcatcggatgcaaaagcgatttgccgatgacccttcaa aaatgtatcgcattgacagaggaaacgacgaaaggcaataagggattacacctggtcatagcacttaactacgg tgggtattacgatatcctacaagcaacgaagtccattgtaaacaaggctatgaatggtttattggacgttgaagaca tcaataaaaatctgttcgaccaagaattagaaagcaaatgccctaaccctgacttgctgatcagaactgggggag aacagagggtctctaattttcttctatggcaattggcttatactgagttctatttttaccaatactttattccctgactttgg tgaagaggacctgaaagaagccatcatgaattttcaacagagacaccgtagattcggaggacatacttattga
SEQ ID NO:70 Neryl pyrophosphate (NPP) synthase	MCSLNLQTEKLCYEDNDNDLDEELMPKHIALIMDGNRRWAKDKGLE VYEGHKHIIPKLKEICDISSKLGIQIITAFAFSTENWKRSKEEVDFLLQMF EEIYDEFSRSGVRVSIIGCKSDLPMTLQKCIALTEETTKGNKGLHLVIAL NYGGYYDILQATKSIVNKAMNGLLDVEDINKNLFDQELESKCPNPDLL

SEQ ID	SEQUENCE
(NPPS) Solanum sp.	IRTGGEQRVSNFLLWQLAYTEFYFTNTLFPDFGEEDLKEAIMNFQQRH RRFGGHTY*
SEQ ID NO:71 Artificial geranylgeranyl pyrophosphate synthase large subunit (GPPSlsu) nucleotide sequence	atgagcaccgtgaatctgacctgggtgcagacgtgctctatgttcaaccagggcgggcgttcccgttcattgtca accttcaacttaaatctgtaccatccattgaagaaaacgcctttctctatccagacacctaagcagaaaaggccaa cttccccttctcatctatcagtgccgtattaacggagcaggaagcagtaaaggagggtgacgaggaaaaaagc atatttaacttcaaatcttatatggttcagaaagctaatagcgtgaatcaggcactagattctgcggtgttattgagag accccattatgatacatgaatctatgcgttactctttgcttgcgggcggcaagcgtgtcagaccgatgttatgcttaa gtgcgtgcgagttagtaggaggtaaagagtctgtagcaatgcccgcagcatgtgctgtagaaatgatacacaca atgtcactgattcacgatgatcttccttgcatggataacgacgatcttcgtagaggtaagccaaccaaccacaagg tattcggggaagacgtggcagttttagcaggagacgcgctactagcgttcgcgtttgaacacatggcagttagca cagtaggagttccagcagcaaaaatagttagggctataggagagttagcaaagtccatcggtagcgagggcctt gttgccggacaggtagttgatatcgatagtgaagggttggctaacgtgggactagaacaactggagttcatccac ctacacaagacaggggcactgcttgaagcgagtgttgtacttggggctattctgggggggaggaacagatgagg aggtagaaaaactacgtagttttgccaggtgtataggactactatttcaagttgtagatgatatccttgacgtcacga agagtagtcaagagttaggaaaaacagcagggaaagatctagttgccgataaagtaacctaccccaggctaat gggtatcgataaatctcgtgagttcgccgaacaattaaatactgaggctaagcaacatttaagcgggtttgatccta ttaaggctgcgccgctgattgctctagcaaactatattgcatatagacagaactga
SEQ ID NO:72 Geranylgeranyl pyrophosphate synthase large subunit (GPPSlsu) Cannabis sativa	MSTVNLTWVQTCSMFNQGGRSRSLSTFNLNLYHPLKKTPFSIQTPKQK RPTSPFSSISAVLTEQEAVKEGDEEKSIFNFKSYMVQKANSVNQALDSA VLLRDPIMIHESMRYSLLAGGKRVRPMLCLSACELVGGKESVAMPAA CAVEMIHTMSLIHDDLPCMDNDDLRRGKPTNHKVFGEDVAVLAGDA LLAFAFEHMAVSTVGVPAAKIVRAIGELAKSIGSEGLVAGQVVDIDSE GLANVGLEQLEFIHLHKTGALLEASVVLGAILGGGTDEEVEKLRSFAR CIGLLFQVVDDILDVTKSSQELGKTAGKDLVADKVTYPRLMGIDKSRE FAEQLNTEAKQHLSGFDPIKAAPLIALANYIAYRQN*
SEQ ID NO:73 Artificial geranylgeranyl pyrophosphate synthase small subunit (GPPSssu) nucleotide sequence	atggctgtttacaacctttcaatcaactgttctcccagattcgtccatcatgtatacgtgccccatttacatgtaaatc aaataagagcctgagccatgtccccatgagaatcacgatgtcaaagcagcatcatcactcatactttgcctctaca acggcagatgtcgatgcccatctaaaacaatcaatcacaattaaaccccccgttgtctgtccacgaagccatgtata actttatcttcagtacgccaccgaatttggcgccatcattatgtgtcgcagcatgtgaattggttgggggtcaccag ggacaggcgatggcagcggccagcgcattaagggtaatacatgctagcatcgttacccacgatcaccttccgtt aacgggaaggccaaaccccacctcaccggaggccgctacgcacaattcctataatccaaacatacagttgttatt acctgacgccattacacccttcgggtttgagctattagcgtccagtgatgatcttacacacaacaagagtgagaga gttcttagggtgatcgttgaatttacgaggactttcggttccagaggcactatagacgcccaataccacgaaaagt tggctagtaggtttgatgtggatagccatgaggcaaagaccgtaggatgggggcattacccatcattgaaaaag gagggagccatgcacgcatgtgctgctgcctgcggagcaatattgggtgaggctcatgaagaagaagtggaa aaattgcgtacattcgggctgtatgtcggcatgatccaaggttatgcgaacagattcatcatgagcagtacagag gagaaaaaagaggctgacaggataattgaggagcttaccaatttagcgcgtcaggagctgaaatacttcgatgg aaggaacctagaaccgttttcaacattcttgttccgtttgtag
SEQ ID NO:74 Geranylgeranyl pyrophosphate synthase small subunit (GPPSssu) Cannabis sativa	MAVYNLSINCSPRFVHHVYVPHFTCKSNKSLSHVPMRITMSKQHHHSY FASTTADVDAHLKQSITIKPPLSVHEAMYNFIFSTPPNLAPSLCVAACEL VGGHQGQAMAAASALRVIHASIVTHDHLPLTGRPNPTSPEAATHNSYN PNIQLLLPDAITPFGFELLASSDDLTHNKSERVLRVIVEFTRTFGSRGTID AQYHEKLASRFDVDSHEAKTVGWGHYPSLKKEGAMHACAAACGAIL GEAHEEEVEKLRTFGLYVGMIQGYANRFIMSSTEEKKEADRIIEELTNL ARQELKYFDGRNLEPFSTFLFRL*
SEQ ID NO:75 Artificial tetraketide synthase (TKS) nucleotide sequence	atgaatcatttaagagctgaaggtccagcctccgttttggccatcggtaccgctaaccctgaaaacattttgttgca agacgaattcccagactactacttcagagtcactaagtccgaacacatgacccaattgaaggagaagttcagaa agatttgtgacaagtccatgattagaaagagaaactgtttcttgaacgaagaacacttgaagcaaaacccaagatt ggttgaacatgaaatgcaaactttggacgctagacaagacatgttggttgttgaagtccctaagttgggtaaggat gcctgtgctaaggccattaaagaatggggtcaacctaagtccaagattacccacttgattttcacctctgcctccac cactgacatgcctggtgctgattaccactgcgctaagttattgggtttgtctccatccgttaagagagttatgatgta ccaattgggttgctacggtggtggtactgttttaagaattgctaaggatattgctgaaaacaacaagggtgccaga

SEQ ID	SEQUENCE
	gtcttagctgtctgctgtgacattatggcttgtttattcagaggtccatctgaatccgacttggaattgttggttggtca agctatcttcggtgacggtgctgctgccgttattgttggtgctgaaccagacgaatccgttggtgaaagaccaattt ttgaattggtttccaccggtcaaactattttgccaaattccgaaggtaccatcggtggtcatatcagagaagccggt ttgatcttcgacttacataaggatgtcccaatgttgatctctaacaacattgaaaagtgtttgatcgaagcttttaccc caattggtatttctgactggaactctatcttctggattacccatcctggtggtaaggctattttggataaggtcgagga aaaattgcacttgaagtctgacaagttcgttgactctagacacgtcttgtccgaacatggtaatatgtcctcttccac cgttttattcgttatggatgagttgagaaagagatccttagaagaaggtaagtccaccaccggtgatggttttgagt ggggtgttttgttcggtttcggtccaggtttgaccgtcgaaagagttgttgttagatctgtcccaattaagtacggat cc
SEQ ID NO:76 Artificial tetraketide synthase (TKS)	MNHLRAEGPASVLAIGTANPENILLQDEFPDYYFRVTKSEHMTQLKEK FRKICDKSMIRKRNCFLNEEHLKQNPRLVEHEMQTLDARQDMLVVEV PKLGKDACAKAIKEWGQPKSKITHLIFTSASTTDMPGADYHCAKLLGL SPSVKRVMMYQLGCYGGGTVLRIAKDIAENNKGARVLAVCCDIMACL FRGPSESDLELLVGQAIFGDGAAAVIVGAEPDESVGERPIFELVSTGQTI LPNSEGTIGGHIREAGLIFDLHKDVPMLISNNIEKCLIEAFTPIGISDWNSI FWITHPGGKAILDKVEEKLHLKSDKFVDSRHVLSEHGNMSSSTVLFVM DELRKRSLEEGKSTTGDGFEWGVLFGFGPGLTVERVVVRSVPIKYGS
SEQ ID NO:77 Artificial olivetolic acid cyclase (OAC) nucleotide sequence	atggccgtcaagcacttgatcgttttgaagttcaaggatgaaatcactgaagctcaaaaggaagaattcttcaaaa cctacgtcaacttagtcaatattattccagccatgaaggacgtctattggggtaaggacgttactcaaaagaataa ggaggaaggttatactcatatcgttgaggtcactttcgaatctgttgagactattcaagactacatcatccacccag cccacgttggtttcggtgatgtttatcgttccttctgggaaaaattgttgatcttcgactacacccctagaaagggat cc
SEQ ID NO:78 Artificial olivetolic acid cyclase (OAC)	MAVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVT QKNKEEGYTHIVEVTFESVETIQDYIIHPAHVGFGDVYRSFWEKLLIFD YTPRKGS
SEQ ID NO:79 Fusion tetraketide synthase-olivetolic acid cyclase (TKS-OAC)	atgaatcatttaagagctgaaggtccagcctccgttttggccatcggtaccgctaaccctgaaaacattttgttgca agacgaattcccagactactacttcagagtcactaagtccgaacacatgacccaattgaaggagaagttcagaa agatttgtgacaagtccatgattagaaagagaaactgtttcttgaacgaagaacacttgaagcaaaacccaagatt ggttgaacatgaaatgcaaactttggacgctagacaagacatgttggttgttgaagtccctaagttgggtaaggat gcctgtgctaaggccattaaagaatggggtcaacctaagtccaagattacccacttgattttcacctctgcctccac cactgacatgcctggtgctgattaccactgcgctaagttattgggtttgtctccatccgttaagagagttatgatgta ccaattgggttgctacggtggtggtactgttttaagaattgctaaggatattgctgaaaacaacaagggtgccaga gtcttagctgtctgctgtgacattatggcttgtttattcagaggtccatctgaatccgacttggaattgttggttggtca agctatcttcggtgacggtgctgctgccgttattgttggtgctgaaccagacgaatccgttggtgaaagaccaattt ttgaattggtttccaccggtcaaactattttgccaaattccgaaggtaccatcggtggtcatatcagagaagccggt ttgatcttcgacttacataaggatgtcccaatgttgatctctaacaacattgaaaagtgtttgatcgaagcttttaccc caattggtatttctgactggaactctatcttctggattacccatcctggtggtaaggctattttggataaggtcgagga aaaattgcacttgaagtctgacaagttcgttgactctagacacgtcttgtccgaacatggtaatatgtcctcttccac cgttttattcgttatggatgagttgagaaagagatccttagaagaaggtaagtccaccaccggtgatggttttgagt ggggtgttttgttcggtttcggtccaggtttgaccgtcgaaagagttgttgttagatctgtcccaattaagtacgcag ccacaagcggttctacgggctccacgggctctaccggcagtgggaggagcactgggtcaacgggatcaacag gtagtggaagatcacacatggttgccgtcaagcacttgatcgttttgaagttcaaggatgaaatcactgaagctca aaaggaagaattcttcaaaacctacgtcaacttagtcaatattattccagccatgaaggacgtctattggggtaag gacgttactcaaaagaataaggaggaaggttatactcatatcgttgaggtcactttcgaatctgttgagactattca agactacatcatccacccagcccacgttggtttcggtgatgtttatcgttccttctgggaaaaattgttgatcttcgac tacacccctagaaagggtaactcgagagcttttgattaa
SEQ ID NO:80 Fusion tetraketide synthase-olivetolic acid cyclase (TKS-OAC)	MNHLRAEGPASVLAIGTANPENILLQDEFPDYYFRVTKSEHMTQLKEK FRKICDKSMIRKRNCFLNEEHLKQNPRLVEHEMQTLDARQDMLVVEV PKLGKDACAKAIKEWGQPKSKITHLIFTSASTTDMPGADYHCAKLLGL SPSVKRVMMYQLGCYGGGTVLRIAKDIAENNKGARVLAVCCDIMACL FRGPSESDLELLVGQAIFGDGAAAVIVGAEPDESVGERPIFELVSTGQTI LPNSEGTIGGHIREAGLIFDLHKDVPMLISNNIEKCLIEAFTPIGISDWNSI

SEQ ID	SEQUENCE
	FWITHPGGKAILDKVEEKLHLKSDKFVDSRHVLSEHGNMSSSTVLFVM DELRKRSLEEGKSTTGDGFEWGVLFGFGPGLTVERVVVRSVPIKYAAT SGSTGSTGSTGSGRSTGSTGSTGSGRSHMVAVKHLIVLKFKDEITEAQK EEFFKTYVNLVNIIPAMKDVYWGKDVTQKNKEEGYTHIVEVTFESVET IQDYIIHPAHVGFGDVYRSFWEKLLIFDYTPRKGNSRAFD*
SEQ ID NO:81 Artificial geranyl pyrophosphate olivetolic acid geranyltransferase (GOT) nucleotide sequence	atgggattgtccagcgtgtgcaccttctcattccaaaccaactaccatacacttctcaatccgcacaataataaccc gaaaaccagcttattatgttatagacacccgaagacgcctattaagtacagttataacaactttcctagcaagcatt gctctactaaaagttttcatctgcaaaacaagtgctctgagtccttgagtatagcaaagaatagcattagagctgca acgacaaatcaaaccgagccgccggagtctgataaccatagtgtggcgaccaagatactaaattttggcaaagc gtgttggaagctacaacgaccttatactattatcgcgtttacgagttgtgcatgtgggctgttcgggaaagagctctt gcacaatacaaacttaatcagttggagtttgatgttcaaagcatttttttttctcgtcgctatcttatgtatcgcgtcattt accacgaccataaatcaaatatacgatctgcatatcgatcgtatcaataagcccgacctcccactggcctcaggt gaaatttccgttaacacggcgtggattatgagtataatcgtagcactatttggacttattataaccatcaaaatgaag ggcggtcctctatacatttttggatattgttttgggatttttggaggtatagtctattccgtccccccattcagatggaa acaaaaacccgtccaccgctttccttttaaatttcttggcacatatcatcacaaacttcacgttttactatgccagccga gccgcactgggactcccgttcgagttgcgtccgtcattcaccttccttttagcttttatgaaatctatgggaagcgct ttagctttaattaaggacgcgagcgacgtggaaggggacacgaaattcggtataagcacgctggcttcaaaatat ggaagtcgtaatctcactctattttgttctgggattgtactcctaagttacgtagctgcgatactcgcaggcattatat ggccacaagctttcaactccaacgtaatgttgctatcacatgcaatcttggccttctggctcatccttcaaactagag attttgcactaacgaactacgatccagaagcgggtcgtcgattttacgaatttatgtggaaactgtactatgctgagt acctcgtctatgtgttcata
SEQ ID NO:82 geranyl pyrophosphate olivetolic acid geranyltransferase (GOT) (CsPT1) *Cannabis sativa* 395 aa WO 2011/017798	MGLSSVCTFSFQTNYHTLLNPHNNNPKTSLLCYRHPKTPIKYSYNNFPS KHCSTKSFHLQNKCSESLSIAKNSIRAATTNQTEPPESDNHSVATKILNF GKACWKLQRPYTIIAFTSCACGLFGKELLHNTNLISWSLMFKAFFFLVA ILCIASFTTTINQIYDLHIDRINKPDLPLASGEISVNTAWIMSIIVALFGLII TIKMKGGPLYIFGYCFGIFGGIVYSVPPFRWKQNPSTAFLLNFLAHIITN FTFYYASRAALGLPFELRPSFTFLLAFMKSMGSALALIKDASDVEGDTK FGISTLASKYGSRNLTLFCSGIVLLSYVAAILAGIIWPQAFNSNVMLLSH AILAFWLILQTRDFALTNYDPEAGRRFYEFMWKLYYAEYLVYVFI
SEQ ID NO:83 Artificial aromatic prenyltransferase (NphB-ScCO) nucleotide sequence	atgtctgaggcggcagacgtagagagagtatacgctgctatggaggaagcggctggattattgggggtggctt gtgccagagacaagatatatccgttactgtctactttccaggacactcttgtagaaggagggagtgtggtggtgttt agtatggcatcaggccgtcattcaacagagctagatttcagtatatctgtgccaacaagtcacggtgatccatacg caaccgtagtcgagaagggtcttttcccggcaacagggcatcctgtagatgatttgcttgccgacacacagaag cacctgcccgtctccatgttcgcaatcgatggtgaggtgaccggaggatttaaaaagacttacgctttcttcccga ctgacaatatgccaggagttgccgagttgagtgcaataccatccatgccgccagcagtcgcggagaacgccga attgttcgcccgttacggcttggacaaagtccaaatgactagtatggactataaaaagaggcaggtgaatctatatt tcagcgaactttctgcccaaaccttggaggcggagagcgttttagcccttgttagggagttagggctacacgtcc cgaatgagttgggtttgaaattttgtaagcgtagcttttcagtatatccgacgctgaactgggaaactggaaagatt gacaggctatgctttgcagtgatttctaatgaccctacgcttgtaccttcctcagacgagggcgacatcgagaaatt ccacaactatgccacaaaagctccgtatgcctacgtcggcgaaaaacgtactctagtatacggtttgactctgagt cccaaggaagagtattacaagctaggagcgtactatcatatcactgatgtgcaacgtggcttgctgaaagccttc gactccttagaggac
SEQ ID NO:84 Aromatic prenyltransferase NphB-ScCO (Streptomyces sp.)	MSEAADVERVYAAMEEAAGLLGVACARDKIYPLLSTFQDTLVEGGSV VVFSMASGRHSTELDFSISVPTSHGDPYATVVEKGLFPATGHPVDDLL ADTQKHLPVSMFAIDGEVTGGFKKTYAFFPTDNMPGVAELSAIPSMPP AVAENAELFARYGLDKVQMTSMDYKKRQVNLYFSELSAQTLEAESVL ALVRELGLHVPNELGLKFCKRSFSVYPTLNWETGKIDRLCFAVISNDPT LVPSSDEGDIEKFHNYATKAPYAYVGEKRTLVYGLTLSPKEEYYKLGA YYHITDVQRGLLKAFDSLED

SEQ ID	SEQUENCE
SEQ ID NO:85 Artificial Tetrahydro cannabinolic acid synthase (THCAS) nucleotide sequence	atgaactgttccgcgtttagtttctggttcgtgtgcaagatcatcttctttttctaagcttcaacattcaaatcagcatc gcgaatcctcaggagaacttcctgaagtgtttctcagaatacataccaaataatcccgccaatcctaaatttatatat acccaacatgatcagctatacatgagtgtattgaactctacgattcagaatctaagattcacatctgatacaacgcc gaaacctctagtaatcgtgacaccgtctaatgtctcccatattcaagcttctatcttgtgctcaaagaaagtcggtctt caaataaggacacgttctggcgggcatgacgccgagggcatgtcatatatcagccaagtaccatttgtagtcgtg gatttaagaaacatgcattctataaaaatcgacgttcactcccaaacggcatgggtggaagctggagcgacactg ggggaggtgtactactggatcaatgaaaagaacgaaaatttttccttccccggaggatattgtccgacagttggg gtgggggccacttctctggcggcgggtacggcgctctgatgcgtaattatggactggccgcagataacataat cgacgcgcatttggtgaacgttgacgggaaggttttggataggaagtctatgggagaggacctattctgggcaat tagaggcggaggaggagagaattttggtattattgctgcatggaagattaaattggttgcggtgccgagtaaaag taccatcttttccgtcaagaaaaacatggagattcacggactagttaagctgtttaataaatggcaaaacatcgcct ataagtacgacaaagatttggttctgatgacgcatttcataactaagaatataactgataatcacggcaagaataag accactgtgcacggttattttagttcaatattccatggcggcgttgactcccttgtcgatttgatgaataagagcttcc ctgaattgggtatcaagaagacagactgcaaagaattctcctggattgatacgactatcttctattcaggggtcgtg aatttcaacactgcgaatttcaaaaaggagatattgttagaccgttccgcgggaaaaaaaactgcgttttctattaa actagattatgtgaaaaaaccgattcctgagacagccatggttaagattcttgaaaaattgtatgaagaggatgtcg gggtcggtatgtacgtcctttaccttacggaggaatcatggaagaaatatccgaatctgcaattcctttcccgcat cgtgccggtattatgtatgagctatggtacaccgctagctgggagaagcaggaagataacgagaagcatatcaa ttgggtgaggtctgtgtataattttacaacaccatacgtcagtcaaaaccctagattggcctatcttaactatcgtgat ctggacttgggaaaaacaaatccagaatccccaaataactacactcaagcccgtatatggggcgagaagtactt cggcaaaaatttcaatagactggtcaaagttaagacgaaagcagaccctaataatttcttccgtaacgaacaatca attcccccgcttccgccacaccatcac
SEQ ID NO:86 Tetrahydro cannabinolic acid synthase (THCAS) *Cannabis sativa*	MNCSAFSFWFVCKIIFFFLSFNIQISIANPQENFLKCFSEYIPNNPANPKFI YTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQASILCSKK VGLQIRTRSGGHDAEGMSYISQVPFVVVDLRNMHSIKIDVHSQTAWVE AGATLGEVYYWINEKNENFSFPGGYCPTVGVGGHFSGGGYGALMRN YGLAADNIIDAHLVNVDGKVLDRKSMGEDLFWAIRGGGGENFGIIAA WKIKLVAVPSKSTIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLVLM THFITKNITDNHGKNKTTVHGYFSSIFHGGVDSLVDLMNKSFPELGIKK TDCKEFSWIDTTIFYSGVVNFNTANFKKEILLDRSAGKKTAFSIKLDYV KKPIPETAMVKILEKLYEEDVGVGMYVLYPYGGIMEEISESAIPFPHRA GIMYELWYTASWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLN YRDLDLGKTNPESPNNYTQARIWGEKYFGKNFNRLVKVKTKADPNNF FRNEQSIPPLPPHHH
SEQ ID NO:87 Artificial Cannabidiolic acid synthase (CBDAS) nucleotide sequence	atgaaatgttctactttcagtttttggttcgtgtgtaagatcatcttttttcttttttcagcttcaatatacagacaagtatcgc caatccaagagaaaatttcttaaaatgtttttcacagtacatccctaataacgccactaacctgaaattagtgtacac ccaaaataatcctctttatatgtctgttttaaactccacgatccataatttaaggtttacatcagatacgacaccaaag cccttggtaatcgtgactcccagccacgtgagccacatacaggggaccatcctgtgctctaagaaagtaggcttg cagatcaggacaagatccggtggacacgacagtgagggaatgtcctatatttcacaagtccccttcgttatagta gatctgaggaacatgaggtccattaagattgatgtgcactcacaaacggcttgggttgaagctggagccacattg ggagaggtttattactgggtgaatgagaagaacgagaacctttcattagcagcgggatattgtcccacggtgtgc gcaggtgggcatttcgggggaggagggtacggccctttgatgagaaattacgggctagcggcagacaacatc atcgacgcccatctggtgaacgtgcatggaaaagtactggacagaaagtcaatgggcgaggacctgttttgggc tttgagaggggcggtgcagagtcatttggcatcatagttgcatggaaaatcagacttgttgccgtcccaaagtcc acaatgttctctgttaagaaaatcatggagatacacgaattggtgaaattagtgaataaatggcaaaacatagcgt acaagtacgacaaagacttactgctgatgacacactttatcacccgtaatattacagataatcagggtaagaacaa aaccgcgatccatacatatttttcatccgtttttctaggcggtgtcgattcattagtagatctgatgaacaaatctttcc ccgaacttggtatcaaaaagactgattgcagacagttatcatggattgatacaataattttctattctggtgtcgtaaa ttacgataccgataatttttaataaggaaatactattagatcgttccgctgggcagaatggtgcattcaagataaaact tgattatgtcaaaaagcccattccagagagtgtctttgtgcagatccttgagaagttgtatgaagaagacattggtg cagggatgtacgcgctatatccgtacgggggtattatggacgagatttctgagagcgccataccattcccacaca gagcaggaattttatacgagttatggtatatctgctcatgggaaaaacaggaagacaacgagaagcacttaaact ggatacgtaatatctataatttttatgaccccatacgtatcaaaaaatccgcgtcttgcgtaccttaactacagggacc

SEQ ID	SEQUENCE
	tggacataggtataaacgacccaaaaaatcccaataattacacccaagctagaatctgggggagaagtatttcg gtaagaactttgaccgtttggtaaaagtcaaaactctggtcgatccgaacaatttcttccgtaacgagcaatccata cctccgctaccgagacatagacat
SEQ ID NO:88 GenBank A6P6V9 Cannabidiolic acid synthase (CBDAS) Cannabis sativa	MKCSTFSFWFVCKIIFFFFSFNIQTSIANPRENFLKCFSQYIPNNATNLKL VYTQNNPLYMSVLNSTIHNLRFTSDTTPKPLVIVTPSHVSHIQGTILCSK KVGLQIRTRSGGHDSEGMSYISQVPFVIVDLRNMRSIKIDVHSQTAWV EAGATLGEVYYWVNEKNENLSLAAGYCPTVCAGGHFGGGGYGPLMR NYGLAADNIIDAHLVNVHGKVLDRKSMGEDLFWALRGGGAESFGIIV AWKIRLVAVPKSTMFSVKKIMEIHELVKLVNKWQNIAYKYDKDLLLM THFITRNITDNQGKNKTAIHTYFSSVFLGGVDSLVDLMNKSFPELGIKK TDCRQLSWIDTIIFYSGVVNYDTDNFNKEILLDRSAGQNGAFKIKLDYV KKPIPESVFVQILEKLYEEDIGAGMYALYPYGGIMDEISESAIPFPHRAGI LYELWYICSWEKQEDNEKHLNWIRNIYNFMTPYVSKNPRLAYLNYRD LDIGINDPKNPNNYTQARIWGEKYFGKNFDRLVKVKTLVDPNNFFRNE QSIPPLPRHRH
SEQ ID NO:89 Artificial acyl-activating enzyme (CsAAE1) nucleotide sequence	atgggaaaaaactacaaaagtctggactccgtcgtcgcgtcagacttcattgccctaggcataacatcagaggta gcggaaaccttacacggcagactagccgagattgtttgtaactacggggcggctactccccagacttggatcaa tatagccaatcacatattaagccccgatttgccgtttttcccttcaccaaatgttgttctacggctgctataaggacttt ggaccagcgccccccgcgtggattcctgatccggagaaagttaaatccacgaatcttggggcattactagaaaa acgtggcaaagaattcctaggagttaaatataaggaccccatatcttccttttcacactttcaagaattttcagttaga aacccagaggtttactggaggacagtattaatggatgagatgaagataagctttagtaaggatccggagtgtattc tgcgtagagatgacattaacaatcctggcggaagtgaatggctgcctggtgggtacctgaatagtgctaagaact gtttaaacgtcaactctaataaaaaattgaatgatacaatgattgtatggagagacgaagggaacgatgacctacc attgaacaagctgactctagatcagctacgtaaacgtgtatggttggtcgggtacgcgctggaggagatgggatt agaaaaaggatgcgcaattgctatcgacatgcctatgcatgtggacgcggtagtcatttacttggccattgtccta gcgggttacgtcgtcgtttcaattgcagacagcttttctgcacccgaaatcagtacccgtctgcgtttgtctaaagct aaggcaatatttacccaagaccatataattagaggcaagaagcgtataccgttgtacagtagggttgtagaggca aagtcacccatggctattgtgataccatgctctggctctaatataggagcggagcttagagatggtgacatctcct gggattactttcttgaacgtgctaaggagtttaaaaactgtgaatttactgcaagagagcagcccgtggatgcata cacaaacatattgttctccagcggtactacgggagaacctaaagcaatacctttggacacaagctacacccctttaa agcggccgctgacggatggtcccacctggatatcaggaagggtgacgtcatagtttggccgactaacctggga tggatgatgggcccttggctggtttacgctagccttctgaatggggccagcattgcattgtacaatggctcaccgc ttgtatcaggcttcgcgaagttcgtacaggacgccaaggtaacaatgctaggcgtagttccgtccatagttaggtc ttggaagagcacgaactgcgttagtggctacgattggagcactattcgttgtttcagctcttctggcgaggccagc aacgttgatgaatatttgtggttgatggggagagcgaactacaaacctgttattgagatgtgcggcggaactgag attgggggagcattctccgccggttcttttctacaagcccaaagtttatcctcttttagcagccagtgcatgggctgt acactatacattctggacaaaaatggttatccgatgccgaaaaacaagcccggcatcggagaactggccctagg acccgtgatgttcggcgctagtaagacgttgttgaatgggaatcaccacgacgtttatttttaagggaatgccaactt tgaatggcgaagtacttcgtagacacggagacatctttgagttgacttcaaacggttactaccacgctcatggacg tgccgatgatacgatgaacattgggggaattaaaatttcatccatagaaatagaacgtgtgtgtaacgaagtcgat gatcgtgtattcgagactacagcgatcggtgtcccaccgttgggtgggggaccagaacaattggtaatctttttgt tctgaaagactccaacgatacgaccatcgacctaaatcagctgaggctatcctttaatctgggcttgcagaaaaa gctaaatcctttattcaaagtcactagagttgttcctttatcttcattaccaagaactgcaacaaataaaataatgcgta gagttctaaggcagcagtttagtcatttcgaa
SEQ ID NO:90 Acyl-activating enzyme (CsAAE1) Cannabis sativa	MGKNYKSLDSVVASDFIALGITSEVAETLHGRLAEIVCNYGAATPQTW INIANHILSPDLPFSLHQMLFYGCYKDFGPAPPAWIPDPEKVKSTNLGA LLEKRGKEFLGVKYKDPISSFSHFQEFSVRNPEVYWRTVLMDEMKISFS KDPECILRRDDINNPGGSEWLPGGYLNSAKNCLNVNSNKKLNDTMIV WRDEGNDDLPLNKLTLDQLRKRVWLVGYALEEMGLEKGCAIAIDMP MHVDAVVIYLAIVLAGYVVVSIADSFSAPEISTRLRLSKAKAIFTQDHII RGKKRIPLYSRVVEAKSPMAIVIPCSGSNIGAELRDGDISWDYFLERAK EFKNCEFTAREQPVDAYTNILFSSGTTGEPKAIPWTQATPLKAAADGW SHLDIRKGDVIVWPTNLGWMMGPWLVYASLLNGASIALYNGSPLVSG

SEQ ID	SEQUENCE
	FAKFVQDAKVTMLGVVPSIVRSWKSTNCVSGYDWSTIRCFSSSGEASN VDEYLWLMGRANYKPVIEMCGGTEIGGAFSAGSFLQAQSLSSFSSQCM GCTLYILDKNGYPMPKNKPGIGELALGPVMFGASKTLLNGNHHDVYF KGMPTLNGEVLRRHGDIFELTSNGYYHAHGRADDTMNIGGIKISSIEIE RVCNEVDDRVFETTAIGVPPLGGGPEQLVIFFVLKDSNDTTIDLNQLRL SFNLGLQKKLNPLFKVTRVVPLSSLPRTATNKIMRRVLRQQFSHFE
SEQ ID NO:91 Artificial acyl-activating enzyme (CsAAE3) nucleotide sequence	atggaaaagtctggttatggtagagatggtatctacaggtctttaagaccaccattgcatttgccaaacaacaaca acttgtccatggtcagtttcttgttcagaaactcttcttcctacccacaaaaaccagccttgattgactctgaaactaa tcaaatcttgtccttctcccacttcaaatccaccgttattaaggtttctcacggtttcttgaacttgggtatcaagaaga atgactggttgatctacgctccaaactctattcatttcccagtttgctttttgggtattattgcttctggtgctattgctact acttccaacccattatacaccgtcagtgaattgtctaagcaagtcaaggattctaacccaaagttgattatcaccgtt ccacaattattggaaaaggtcaagggtttcaacttgccaaccattttgattggtccagactcagaacaagaatcctc ttcagataaggttatgaccttcaacgatttggttaacttgggtggttcttctggttctgaatttccaatcgttgatgactt caagcaatctgatactgctgctttgttgtactcttctggtactactggtatgtctaaaggttggttgactcacaagaac tttatcgcctcttctttgatggttaccatggaacaagacttggttggtgaaatggataacgttttcttgtgcttcttgcca atgttccatgttttcggtttggccattattacctacgctcaattgcaaagaggtaacactgttatttccgccagattcga tttggaaaagatgttgaaggacgtcgaaaagtatgttactcatttgtggtggcctccagttattttggctttgtctaaa aactccatggttaagttcaacttgtcatccatcaagtacattggttcaggtgctgctccattgggtaaggatttgatg gaagaatgttctaaatggccatacggtatagttgctcaaggttacggtatgactgaaacttgtggtatcgtttctatg gaagatatcagaggtggtaagagaaattctggttcagctggtatgttggcttcaggtgttgaagctcaaatagtttc tgttgataccttgaaaccattgccaccaaatcaattgggtgaaatttgggttaagggtccaaatatgatgcaaggtt acttcaacaatccacaagctaccaagttgaccattgataagaaaggttgggttcatactggtgacttgggttacttt gatgaagatggtcacttgtactgggacagaatcaaagaattgattaagtacaagggttttcaagtcgctccagctg aattggaaggtttgttggtttctcatccagaaatattggatgcctggattccatttccagatgctgaagctggtgaag ttccagttgcttattggagatcaccaaactcttcattgactgaaaacgacgtcaagaagttcattgctggtcaagttg cttctttcaagagattgagaaaggtcaccttcatcaactctgttccaaaatctgcttccggtaagatcttgagaagag aattgatccaaaaggtcagatccaatatg
SEQ ID NO:92 Acyl-activating enzyme (CsAAE3) Cannabis sativa	MEKSGYGRDGIYRSLRPPLHLPNNNNLSMVSFLFRNSSSYPQKPALIDS ETNQILSFSHFKSTVIKVSHGFLNLGIKKNDWLIYAPNSIHFPVCFLGIIA SGAIATTSNPLYTVSELSKQVKDSNPKLIITVPQLLEKVKGFNLPTILIGP DSEQESSSDKVMTFNDLVNLGGSSGSEFPIVDDFKQSDTAALLYSSGTT GMSKGWLTHKNFIASSLMVTMEQDLVGEMDNVFLCFLPMFHVFGLAI ITYAQLQRGNTVISARFDLEKMLKDVEKYVTHLWVPPVILALSKNSM VKFNLSSIKYIGSGAAPLGKDLMEECSKWPYGIVAQGYGMTETCGIVS MEDIRGGKRNSGSAGMLASGVEAQIVSVDTLKPLPPNQLGEIWVKGPN MMQGYFNNPQATKLTIDKKGWVHTGDLGYFDEDGHLYWDRIKELIK YKGFQVAPAELEGLLVSHPEILDAWIPFPDAEAGEVPVAYWRSPNSSL TENDVKKFIAGQVASFKRLRKVTFINSVPKSASGKILRRELIQKVRSNM
SEQ ID NO:93 Artificial acetoacetyl-CoA reductase (PhaB) nucleotide sequence	atgacgcagagaatcgcctatgtaacgggtgggatgggtgggataggaaccgccatatgtcagagactagcaa aggacggattcagggttgtagccggttgcggtcctaatagtccaagaagagagaaatggttggaacagcaaaa agctctaggatttgattttatagcatcagaagggaatgttgctgactgggattctacaaagacggcatttgacaaag tgaaatctgaagtcggcgaggtcgatgtcctaattaacaacgccggcatcaccagagatgtggttttcaggaaga tgactagggctgactgggacgccgtgatagacacaaatttgacgagcttgttcaacgtcacaaagcaagtaattg acggcatggcagatcgtgggtggggaaggatagtcaatatctccagcgtcaacggtcagaaaggccagttcgg acagactaactactccacagcgaaggctggcttacacggattcacgatggccttggcccaagaggtggctacta aaggggtgactgtgaacacagtgtcaccaggatacatcgcgacggatatggtcaaagctattagacaagatgtc ctggacaagattgttgccactattcccgtaaagaggcttgggttaccagaagagatagcttcaatttgcgcttggct atctagtgaggaatcagggttcagcactggggcggacttttcattaaacggtggattacacatgggaggatcc
SEQ ID NO:94 Mutant acetoacetyl-CoA reductase (PhaB)	MTQRIAYVTGGMGGIGTAICQRLAKDGFRVVAGCGPNSPRREKWLEQ QKALGFDFIASEGNVADWDSTKTAFDKVKSEVGEVDVLINNAGITRD VVFRKMTRADWDAVIDTNLTSLFNVTKQVIDGMADRGWGRIVNISSV NGQKGQFGQTNYSTAKAGLHGFTMALAQEVATKGVTVNTVSPGYIA TDMVKAIRQDVLDKIVATIPVKRLGLPEEIASICAWLSSEESGFSTGADF

SEQ ID	SEQUENCE
	SLNGGLHMGGS
SEQ ID NO:95 Artificial (R)-specific enoyl-CoA hydratase (PhaJ)	atgtctgcccagagtctggaagtcggtcaaaaagcaagactgtcaaaaagatttggggcggcagaggtagcgg cgttcgcggcgctgtctgaggattttaatccactgcacttagatcctgcgttcgccgcgacaacagcattcgagag gcccatcgtgcacggcatgctacttgcctctttgttctcaggtctactgggtcaacagttacctgggaaaggaagc atctatctgggacagtcattgtcttttaagctgcccgtcttcgtcggcgatgaggtgacagcagaagtagaagtca cagcattgagggaagacaagcctattgcgacccttactactcgtatttttactcagggcggagccttagcagtgac aggagaagctgtagtaaaactaccaggatcc
SEQ ID NO:96 Mutant (R)-specific enoyl-CoA hydratase (PhaJ)	MSAQSLEVGQKARLSKRFGAAEVAAFAALSEDFNPLHLDPAFAATTA FERPIVHGMLLASLFSGLLGQQLPGKGSIYLGQSLSFKLPVFVGDEVTA EVEVTALREDKPIATLTTRIFTQGGALAVTGEAVVKLPGS
SEQ ID NO:97 Mutated acetyl-CoA carboxylase (ACC1) (S659A, S1157A)	MSEESLFESSPQKMEYEITNYSERHTELPGHFIGLNTVDKLEESPLRDFV KSHGGHTVISKILIANNGIAAVKEIRSVRKWAYETFGDDRTVQFVAMA TPEDLEANAEYIRMADQYIEVPGGTNNNNYANVDLIVDIAERADVDA VWAGWGHASENPLLPEKLSQSKRKVIFIGPPGNAMRSLGDKISSTIVAQ SAKVPCIPWSGTGVDTVHVDEKTGLVSVDDDIYQKGCCTSPEDGLQK AKRIGFPVMIKASEGGGGKGIRQVEREEDFIALYHQAANEIPGSPIFIMK LAGRARHLEVQLLADQYGTNISLFGRDCSVQRRHQKIIEEAPVTIAKAE TFHEMEKAAVRLGKLVGYVSAGTVEYLYSHDDGKFYFLELNPRLQVE HPTTEMVSGVNLPAAQLQIAMGIPMHRISDIRTLYGMNPHSASEIDFEF KTQDATKKQRRPIPKGHCTACRITSEDPNDGFKPSGGTLHELNFRSSSN VWGYFSVGNNGNIHSFSDSQFGHIFAFGENRQASRKHMVVALKELSIR GDFRTTVEYLIKLLETEDFEDNTITTGWLDDLITHKMTAEKPDPTLAVI CGAATKAFLASEEARHKYIESLQKGQVLSKDLLQTMFPVDFIHEGKRY KFTVAKSGNDRYTLFINGSKCDIILRQLADGGLLIAIGGKSHTIYWKEE VAATRLSVDSMTTLLEVENDPTQLRTPSPGKLVKFLVENGEHIIKGQPY AEIEVMKMQMPLVSQENGIVQLLKQPGSTIVAGDIMAIMTLDDPSKVK HALPFEGMLPDFGSPVIEGTKPAYKFKSLVSTLENILKGYDNQVIMNAS LQQLIEVLRNPKLPYSEWKLHISALHSRLPAKLDEQMEELVARSLRRG AVFPARQLSKLIDMAVKNPEYNPDKLLGAVVEPLADIAHKYSNGLEA HEHSIFVHFLEEYYEVEKLFNGPNVREENIILKLRDENPKDLDKVALTV LSHSKVSAKNNLILAILKHYQPLCKLSSKVSAIFSTPLQHIVELESKATA KVALQAREILIQGALPSVKERTEQIEHILKSSVVKVAYGSSNPKRSEPDL NILKDLIDSNYVVFDVLLQFLTHQDPVVTAAAAQVYIRRAYRAYTIGDI RVHEGVTVPIVEWKFQLPSAAFSTFPTVKSKMGMNRAVSVADLSYVA NSQSSPLREGILMAVDHLDDVDEILSQSLEVIPRHQSSSNGPAPDRSGSS ASLSNVANVCVASTEGFESEEEILVRLREILDLNKQELINASIRRITFMF GFKDGSYPKYYTFNGPNYNENETIRHIEPALAFQLELGRLSNFNIKPIFT DNRNIHVYEAVSKTSPLDKRFFTRGIIRTGHIRDDISIQEYLTSEANRLM SDILDNLEVTDTSNSDLNHIFINFIAVFDISPEDVEAAFGGFLERFGKRLL RLRVSSAEIRIIIKDPQTGAPVPLRALINNVSGYVIKTEMYTEVKNAKGE WVFKSLGKPGSMHLRPIATPYPVKEWLQPKRYKAHLMGTTYVYDFPE LFRQASSSQWKNFSADVKLTDDFFISNELIEDENGELTEVEREPGANAI GMVAFKITVKTPEYPRGRQFVVVANDITFKIGSFGPQEDEFFNKVTEYA RKRGIPRIYLAANSGARIGMAEEIVPLFQVAWNDAANPDKGFQYLYLT SEGMETLKKFDKENSVLTERTVINGEERFVIKTIIGSEDGLGVECLRGSG LIAGATSRAYHDIFTITLVTCRSVGIGAYLVRLGQRAIQVEGQPIILTGA PAINKMLGREVYTSNLQLGGTQIMYNNGVSHLTAVDDLAGVEKIVEW MSYVPAKRNMPVPILETKDTWDRPVDFTPTNDETYDVRWMIEGRETE SGFEYGLFDKGSFFETLSGWAKGVVVGRARLGGIPLGVIGVETRTVEN LIPADPANPNSAETLIQEPGQVWHPNSAFKTAQAINDFNNGEQLPMMIL ANWRGFSGGQRDMFNEVLKYGSFIVDALVDYKQPIIIYIPPTGELRGGS WVVVDPTINADQMEMYADVNARAGVLEPQGMVGIKFRREKLLDTM

SEQ ID	SEQUENCE
	NRLDDKYRELRSQLSNKSLAPEVHQQISKQLADRERELLPIYGQISLQF ADLHDRSSRMVAKGVISKELEWTEARRFFFWRLRRRLNEEYLIKRLSH QVGEASRLEKIARIRSWYPASVDHEDDRQVATWIEENYKTLDDKLKG LKLESFAQDLAKKIRSDHDNAIDGLSEVIKMLSTDDKEKLLKTLK*
SEQ ID NO:98 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsGOTt75	MAATTNQTEPPESDNHSVATKILNFGKACWKLQRPYTIIAFTSCACGLF GKELLHNTNLISWSLMFKAFFFLVAILCIASFTTTINQIYDLHIDRINKPD LPLASGEISVNTAWIMSIIVALFGLIITIKMKGGPLYIFGYCFGIFGGIVYS VPPFRWKQNPSTAFLLNFLAHIITNFTFYYASRAALGLPFELRPSFTFLL AFMKSMGSALALIKDASDVEGDTKFGISTLASKYGSRNLTLFCSGIVLL SYVAAILAGIIWPQAFNSNVMLLSHAILAFWLILQTRDFALTNYDPEAG RRFYEFMWKLYYAEYLVYVFI*
SEQ ID NO:99 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsGOTt33	MSHPKTPIKYSYNNFPSKHCSTKSFHLQNKCSESLSIAKNSIRAATTNQT EPPESDNHSVATKILNFGKACWKLQRPYTIIAFTSCACGLFGKELLHNT NLISWSLMFKAFFFLVAILCIASFTTTINQIYDLHIDRINKPDLPLASGEIS VNTAWIMSIIVALFGLIITIKMKGGPLYIFGYCFGIFGGIVYSVPPFRWK QNPSTAFLLNFLAHIITNFTFYYASRAALGLPFELRPSFTFLLAFMKSMG SALALIKDASDVEGDTKFGISTLASKYGSRNLTLFCSGIVLLSYVAAILA GIIWPQAFNSNVMLLSHAILAFWLILQTRDFALTNYDPEAGRRFYEFM WKLYYAEYLVYVFI*
SEQ ID NO:100 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4t	MSAGSDQIEGSPHHESDNSIATKILNFGHTCWKLQRPYVVKGMISIACG LFGRELFNNRHLFSWGLMWKAFFALVPILSFNFFAAIMNQIYDVDIDRI NKPDLPLVSGEMSIETAWILSIIVALTGLIVTIKLKSAPLFVFIYIFGIFAG FAYSVPPIRWKQYPFTNFLITISSHVGLAFTSYSATTSALGLPFVWRPAF SFIIAFMTVMGMTIAFAKDISDIEGDAKYGVSTVATKLGARNMTFVVS GVLLLNYLVSISIGIIWPQVFKSNIMILSHAILAFCLIFQTRELALANYAS APSRQFFEFIWLLYYAEYFVYVFI*
SEQ ID NO:101 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT7t	MSTDTANQTEPPESNTKYSVVTKILSFGHTCWKLQRPYTFIGVISCACG LFGRELFHNTNLLSWSLMLKAFSSLMVILSVNLCTNIINQITDLDIDRIN KPDLPLASGEMSIETAWIMSIIVALTGLILTIKLNCGPLFISLYCVSILVG ALYSVPPFRWKQNPNTAFSSYFMGLVIVNFTCYYASRAAFGLPFEMSP PFTFILAFVKSMGSALFLCKDVSDIEGDSKHGISTLATRYGAKNITFLCS GIVLLTYVSAILAAIIWPQAFKSNVMLLSHATLAFWLIFQTREFALTNY NPEAGRKFYEFMWKLHYAEYLVYVFI*
SEQ ID NO:102 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase HlPT1Lt	MDRPPESGNLSALTNVKDFVSVCWEYVRPYTAKGVIICSSCLFGRELL ENPNLFSWPLIFRALLGMLAILGSCFYTAGINQIFDMDIDRINKPDLPLV SGRISVESAWLLTLSPAIIGFILILKLNSGPLLTSLYCLAILSGTIYSVPPFR WKKNPITAFLCILMIHAGLNFSVYYASRAALGLAFVWSPSFSFITAFITF MTLTLASSKDLSDINGDRKFGVETFATKLGAKNITLLGTGLLLLNYVA AISTAIIWPKAFKSNIMLLSHAILAFSLFFQARELDRTNYTPEACKSFYEF IWILFSAEYVVYLFI*
SEQ ID NO:103 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase HlPT2t	MGHLPRPNSLTAWSHQSEFPSTIVTKGSNFGHASWKFVRPIPFVAVSIIC TSLFGAELLKNPNLFSWQLMFDAFQGLVVILLYHIYINGLNQIYDLESD RINKPDLPLAAEEMSVKSAWFLTIFSAVASLLLMIKLKCGLFLTCMYCC YLVIGAMYSVPPFRWKMNTFTSTLWNFSEIGIGINFLINYASRATLGLPF QWRPPFTFIIGFVSTLSIILSILKDVPDVEGDKKVGMSTLPVIFGARTIVL VGSGFFLLNYVAAIGVAIMWPQAFKGYIMIPAHAIFASALIFKTWLLDK ANYAKEASDSYYHFLWFLMIAEYILYPFIST*
SEQ ID NO:104 Truncated tetrahydrocannabinolic acid synthase	MNFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPK PLVIVTPSNVSHIQASILCSKKVGLQIRTRSGGHDAEGMSYISQVPFVVV DLRNMHSIKIDVHSQTAWVEAGATLGEVYYWINEKNENFSFPGGYCP TVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMG

SEQ ID	SEQUENCE
THCASt28	EDLFWAIRGGGGENFGIIAAWKIKLVAVPSKSTIFSVKKNMEIHGLVKL FNKWQNIAYKYDKDLVLMTHFITKNITDNHGKNKTTVHGYFSSIFHGG VDSLVDLMNKSFPELGIKKTDCKEFSWIDTTIFYSGVVNFNTANFKKEI LLDRSAGKKTAFSIKLDYVKKPIPETAMVKILEKLYEEDVGVGMYVLY PYGGIMEEISESAIPFPHRAGIMYELWYTASWEKQEDNEKHINWVRSV YNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARIWGEKYFG KNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*
SEQ ID NO:105 Truncated cannabidiolic acid synthase CBDASt28*	MNPRENFLKCFSQYIPNNATNLKLVYTQNNPLYMSVLNSTIHNLRFTS DTTPKPLVIVTPSHVSHIQGTILCSKKVGLQIRTRSGGHDSEGMSYISQV PFVIVDLRNMRSIKIDVHSQTAWVEAGATLGEVYYWVNEKNENLSLA AGYCPTVCAGGHFGGGGYGPLMRNYGLAADNIIDAHLVNVHGKVLD RKSMGEDLFWALRGGGAESFGIIVAWKIRLVAVPKSTMFSVKKIMEIH ELVKLVNKWQNIAYKYDKDLLLMTHFITRNITDNQGKNKTAIHTYFSS VFLGGVDSLVDLMNKSFPELGIKKTDCRQLSWIDTIIFYSGVVNYDTD NFNKEILLDRSAGQNGAFKIKLDYVKKPIPESVFVQILEKLYEEDIGAG MYALYPYGGIMDEISESAIPFPHRAGILYELWYICSWEKQEDNEKHLN WIRNIYNFMTPYVSKNPRLAYLNYRDLDIGINDPKNPNNYTQARIWGE KYFGKNFDRLVKVKTLVDPNNFFRNEQSIPPLPRHRH*
SEQ ID NO:106 Mutated fatty acid synthase (FAS1, I306A, R1834K)	MDAYSTRPLTLSHGSLEHVLLVPTASFFIASQLQEQFNKILPEPTEGFAA DDEPTTPAELVGKFLGYVSSLVEPSKVGQFDQVLNLCLTEFENCYLEG NDIHALAAKLLQENDTTLVKTKELIKNYITARIMAKRPFDKKSNSALFR AVGEGNAQLVAIFGGQGNTDDYFEELRDLYQTYHVLVGDLIKFSAETL SELIRTTLDAEKVFTQGLNILEWLENPSNTPDKDYLLSIPISCPLIGVIQL AHYVVTAKLLGFTPGELRSYLKGATGHSQGLVTAVAIAETDSWESFFV SVRKAITVLFFGGVRCYEAYPNTSLPPSILEDSLENNEGVPSPMLSISNL TQEQVQDYVNKTNSHLPAGKQVEISLVNGAKNLVVSGPPQSLYGLNL TLRKAKAPSGLDQSRIPFSERKLKFSNRFLPVASPFHSHLLVPASDLINK DLVKNNVSFNAKDIQIPVYDTFDGSDLRVLSGSISERIVDCIIRLPVKWE TTTQFKATHILDFGPGGASGLGVLTHRNKDGTGVRVIVAGTLDINPDD DYGFKQEIFDVTSNGLKKNPNWLEEYHPKLIKNKSGKIFVETKFSKLIG RPPLLVPGMTPCTVSPDFVAATTNAGYTIELAGGGYFSAAGMTAAIDS VVSQIEKGSTFGINLIYVNPFMLQWGIPLIKELRSKGYPIQFLTIGAGVPS LEVASEYIETLGLKYLGLKPGSIDAISQVINIAKAHPNFPIALQWTGGRG GGHHSFEDAHTPMLQMYSKIRRHPNIMLIFGSGFGSADDTYPYLTGEW STKFDYPPMPFDGFLFGSRVMIAKEVKTSPDAKKCIAACTGVPDDKWE QTYKKPTGGIVTVRSEMGEPIHKIATRGVMLWKEFDETIFNLPKNKLV PTLEAKRDYIISRLNADFQKPWFATVNGQARDLATMTYEEVAKRLVE LMFIRSTNSWFDVTWRTFTGDFLRRVEERFTKSKTLSLIQSYSLLDKPD EAIEKVFNAYPAAREQFLNAQDIDHFLSMCQNPMQKPVPFVPVLDRRF EIFFKKDSLWQSEHLEAVVDQDVQRTCILHGPVAAQFTKVIDEPIKSIM DGIHDGHIKKLLHQYYGDDESKIPAVEYFGGESPVDVQSQVDSSSVSE DSAVFKATSSTDEESWFKALAGSEINWRHASFLCSFITQDKMFVSNPIR KVFKPSQGMVVEISNGNTSSKTVVTLSEPVQGELKPTVILKLLKENIIQ MEMIENRTMDGKPVSLPLLYNFNPDNGFAPISEVMEDRNQRIKEMYW KLWIDEPFNLDFDPRDVIKGKDFEITAKEVYDFTHAVGNNCEDFVSRP DRTMLAPMDFAIVVGWRAIIKAIFPNTVDGDLLKLVHLSNGYKMIPGA KPLQVGDVVSTTAVIESVVNQPTGKIVDVVGTLSRNGKPVMEVTSSFF YRGNYTDFENTFQKTVEPVYQMHIKTSKDIAVLRSKEWFQLDDEDFD LLNKTLTFETETEVTFKNANIFSSVKCFGPIKVELPTKETVEIGIVDYEA GASHGNPVVDFLKRNGSTLEQKVNLENPIPIAVLDSYTPSTNEPYARVS GDLNPIHVSRHFASYANLPGTITHGMFSSASVRALIENWAADSVSSRVR GYTCQFVDMVLPNTALKTSIQHVGMINGRKLIKFETRNEDDVVVLTGE AEIEQPVTTFVFTGQGSQEQGMGMDLYKTSKAAQDVWNRADNHFKD

SEQ ID	SEQUENCE
	TYGFSILDIVINNPVNLTIHFGGEKGKRIRENYSAMIFETIVDGKLKTEKI FKEINEHSTSYTFRSEKGLLSATQFTQPALTLMEKAAFEDLKSKGLIPA DATFAGHSLGEYAALASLADVMSIESLVEVVFYFGMTMQVAVPRDEL GRSNYGMIAINPGRVAASFSQEALQYVVERVGKRTGWLVEIVNYNVE NQQYVAAGDLRALDTVTNVLNFIKLQKIDIIELQKSLSLEEVEGHLFEII DEASKKSAVKPRPLKLERGFACIPLVGISVPFHSTYLMNGVKPFKSFLK KNIIKENVKVARLAGKYIPNLTAKPFQVTKEYFQDVYDLTGSEPIKEIID NWEKYEQS*
SEQ ID NO:107 Mutated fatty acid synthase (FAS2, G1250S)	MKPEVEQELAHILLTELLAYQFASPVRWIETQDVFLKDFNTERVVEIGP SPTLAGMAQRTLKNKYESYDAALSLHREILCYSKDAKEIYYTPDPSEL AAKEEPAKEEAPAPTPAASAPAPAAAAPAPVAAAAPAAAAAEIADEPV KASLLLHVLVAHKLKKSLDSIPMSKTIKDLVGGKSTVQNEILGDLGKE FGTTPEKPEETPLEELAETFQDTFSGALGKQSSSLLSRLISSKMPGGFTIT VARKYLQTRWGLPSGRQDGVLLVALSNEPAARLGSEADAKAFLDSM AQKYASIVGVDLSSAASASGAAGAGAAAGAAMIDAGALEEITKDHKV LARQQLQVLARYLKMDLDNGERKFLKEKDTVAELQAQLDYLNAELG EFFVNGVATSFSRKKARTFDSSWNWAKQSLLSLYFEIIHGVLKNVDRE VVSEAINIMNRSNDALIKFMEYHISNTDETKGENYQLVKTLGEQLIENC KQVLDVDPVYKDVAKPTGPKTAIDKNGNITYSEEPREKVRKLSQYVQ EMALGGPITKESQPTIEEDLTRVYKAISAQADKQDISSSTRVEFEKLYSD LMKFLESSKEIDPSQTTQLAGMDVEDALDKDSTKEVASLPNKSTISKTV SSTIPRETIPFLHLRKKTPAGDWKYDRQLSSLFLDGLEKAAFNGVTFKD KYVLITGAGKGSIGAEVLQGLLQGGAKVVVTTSRFSKQVTDYYQSIYA KYGAKGSTLIVVPFNQGSKQDVEALIEFIYDTEKNGGLGWDLDAIIPFA AIPEQGIELEHIDSKSEFAHRIMLTNILRMMGCVKKQKSARGIETRPAQ VILPMSPNHGTFGGDGMYSESKLSLETLFNRWHSESWANQLTVCGAII GWTRGTGLMSANNIIAEGIEKMGVRTFSQKEMAFNLLGLLTPEVVELC QKSPVMADLNGGLQFVPELKEFTAKLRKELVETSEVRKAVSIETALEH KVVNGNSADAAYAQVEIQPRANIQLDFPELKPYKQVKQIAPAELEGLL DLERVIVVTGFAEVGPWGSARTRWEMEAFGEFSLEGCVEMAWIMGFI SYHNGNLKGRPYTGWVDSKTKEPVDDKDVKAKYETSILEHSGIRLIEP ELFNGYNPEKKEMIQEVIVEEDLEPFEASKETAEQFKHQHGDKVDIFEI PETGEYSVKLLKGATLYIPKALRFDRLVAGQIPTGWNAKTYGISDDIIS QVDPITLFVLVSVVEAFIASGITDPYEMYKYVHVSEVGNCSGSSMGGV SALRGMFKDRFKDEPVQNDILQESFINTMSAWVNMLLISSSGPIKTPVG ACATSVESVDIGVETILSGKARICIVGGYDDFQEEGSFEFGNMKATSNT LEEFEHGRTPAEMSRPATTTRNGFMEAQGAGIQIIMQADLALKMGVPI YGIVAMAATATDKIGRSVPAPGKGILTTAREHHSSVKYASPNLNMKYR KRQLVTREAQIKDWVENELEALKLEAEEIPSEDQNEFLLERTREIHNEA ESQLRAAQQQWGNDFYKRDPRIAPLRGALATYGLTIDDLGVASFHGTS TKANDKNESATINEMMKHLGRSEGNPVIGVFQKFLTGHPKGAAGAW MMNGALQILNSGIIPGNRNADNVDKILEQFEYVLYPSKTLKTDGVRAV SITSFGFGQKGGQAIVVHPDYLYGAITEDRYNEYVAKVSAREKSAYKF FHNGMIYNKLFVSKEHAPYTDELEEDVYLDPLARVSKDKKSGSLTFNS KNIQSKDSYINANTIETAKMIENMTKEKVSNGGVGVDVELITSINVEND TFIERNFTPQEIEYCSAQPSVQSSFAGTWSAKEAVFKSLGVKSLGGGAA LKDIEIVRVNKNAPAVELHGNAKKAAEEAGVTDVKVSISHDDLQAVA VAVSTKK*
SEQ ID NO:108 MBPtag	MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKF PQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWD AVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGK SALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKA GLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDT

SEQ ID	SEQUENCE
	SKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLL TDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNI PQMSAFWYAVRTAVINAASGRQTVDEALKDAQTRITK
SEQ ID NO:109 ProA tag	MIFDGTTMSIAIGLLSTLGIGAEA
SEQ ID NO:110 Geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4 Cannibis sativa	MGLSLVCTFSFQTNYHTLLNPHNKNPKNSLLSYQHPKTPIIKSSYDNFP SKYCLTKNFHLLGLNSHNRISSQSRSIRAGSDQIEGSPHHESDNSIATKIL NFGHTCWKLQRPYVVKGMISIACGLFGRELFNNRHLFSWGLMWKAFF ALVPILSFNFFAAIMNQIYDVDIDRINKPDLPLVSGEMSIETAWILSIIVA LTGLIVTIKLKSAPLFVFIYIFGIFAGFAYSVPPIRWKQYPFTNFLITISSH VGLAFTSYSATTSALGLPFVWRPAFSFIIAFMTVMGMTIAFAKDISDIEG DAKYGVSTVATKLGARNMTFVVSGVLLLNYLVSISIGIIWPQVFKSNIM ILSHAILAFCLIFQTRELALANYASAPSRQFFEFIWLLYYAEYFVYVFI*
SEQ ID NO:111 Artificial geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4 nucleotide sequence	atgggtttatctttggtctgcaccttctcctttcaaactaactaccacactttattgaatccacataataagaatcctaa gaactctttattgtcctaccaacacccaaagactcctattatcaagtcctcttacgataacttcccatctaagtactgtt tgactaagaatttccatttgttgggtttgaattctcacaacagaatttcctcccaatcccgttctattagagccggttct gatcaaatcgaaggttcccctcatcatgagtccgataactccattgctactaaaattttaaatttcggtcatacttgttg gaagttgcaacgtccttacgttgtcaagggtatgatctctattgcttgtggtttgttcggtagagaagttgtttaacaac agacacttgttctcttggggtttgatgtggaaagctttcttcgctttggtcccaattttgtctttcaatttcttcgccgcca tcatgaaccaaatctacgatgttgatatcgaccgtatcaacaagccagacttacctttagtttccggtgaaatgtcca ttgaaactgcttggatcttgtctatcattgttgccttgactggtttaattgttactattaagttgaagtccgctccattgttt gtcttcatctacatcttcggtatcttcgctggtttcgcttactccgtcccacctattagatggaaacaatatccttttacc aatttcttgatcactatttcctctcatgttggtttggctttcacttcttactctgccaccacttctgctttaggtttgcctttc gtttggcgtcctgccttctctttcattattgctttcatgactgtcatgggtatgactattgcctttgctaaagacatttctg atatcgaaggtgatgctaagtacggtgtctctaccgttgctaccaagttaggtgctagaaatatgacttttgttgtttc tggtgtcttattgttgaactacttggtttctatctctattggtatcatttggccacaagttttcaagtctaacattatgatct tgtctcatgctattttggctttctgtttgatctttcaaactcgtgaattagccttagccaattatgcctctgccccatccc gtcaatttttcgaattcatctggttgttatactatgccgaatacttcgtttacgtcttcatttaa
SEQ ID NO:112 Truncated acyl activating enzyme AAE (CsAAE3 truncation)	MEKSGYGRDGIYRSLRPPLHLPNNNNLSMVSFLFRNSSSYPQKPALIDS ETNQILSFSHFKSTVIKVSHGFLNLGIKKNDWLIYAPNSIHFPVCFLGIIA SGAIATTSNPLYTVSELSKQVKDSNPKLIITVPQLLEKVKGFNLPTILIGP DSEQESSSDKVMTFNDLVNLGGSSGSEFPIVDDFKQSDTAALLYSSGTT GMSKGWLTHKNFIASSLMVTMEQDLVGEMDNVFLCFLPMFHVFGLAI ITYAQLQRGNTVISARFDLEKMLKDVEKYVTHLWWPPVILALSKNSM VKFNLSSIKYIGSGAAPLGKDLMEECSKWPYGIVAQGYGMTETCGIVS MEDIRGGKRNSGSAGMLASGVEAQIVSVDTLKPLPPNQLGEIWVKGPN MMQGYFNNPQATKLTIDKKGWVHTGDLGYFDEDGHLYWDRIKELIK YKGFQVAPAELEGLLVSHPEILDAWIPFPDAEAGEVPVAYWRSPNSSL TENDVKKFIAGQVASFKRLRKVTFINSVPKSASGKILR
SEQ ID NO:113 Truncated 3-hydroxy-3-methyl-glutaryl-CoA reductase (tHMG1)	MSDQLVKTEVTKKSFTAPVQKASTPVLTNKTVISGSKVKSLSSAQSSSS GPSSSSEEDDSRDIESLDKKIRPLEELEALLSSGNTKQLKNKEVAALVIH GKLPLYALEKKLGDTTRAVAVRRKALSILAEAPVLASDRLPYKNYDY DRVFGACCENVIGYMPLPVGVIGPLVIDGTSYHIPMATTEGCLVASAM RGCKAINAGGGATTVLTKDGMTRGPVVRFPTLKRSGACKIWLDSEEG QNAIKKAFNSTSRFARLQHIQTCLAGDLLFMRFRTTTGDAMGMNMISK GVEYSLKQMVEEYGWEDMEVVSVSGNYCTDKKPAAINWIEGRGKSV VAEATIPGDVVRKVLKSDVSALVELNIAKNLVGSAMAGSVGGFNAHA ANLVTAVFLALGQDPAQNVESSNCITLMKEVDGDLRISVSMPSIEVGTI GGGTVLEPQGAMLDLLGVRGPHATAPGTNARQLARIVACAVLAGELS LCAALAAGHLVQSHMTHNRKPAEPTKPNNLDATDINRLKDGSVTCIKS *

SEQ ID	SEQUENCE
SEQ ID NO:114 Truncated 3-hydroxy-3-methyl-glutaryl-CoA reductase (tHMG1)	atgtccgatcaattagtcaagaccgaagtcaccaagaagtccttcaccgccccagttcaaaaagcctccactcca gtcttaactaacaagaccgttatttctggttccaaggttaagtctttatcctccgctcaatcctcctcttccggtccatc ttcttcttctgaagaagatgattcccgtgacattgagtccttggataagaaaatcagacctttggaagaattagaag ctttgttatcctctggtaatactaagcaattgaaaaacaaggaagttgctgctttggttattcacggtaaattacctttg tacgctttagagaagaagttaggtgacactacccgtgctgtcgccgttagaagaaaagctttgtctattttagctga ggcccctgttttggcttctgacagattaccatacaagaattacgattacgatagagtttcggtgcctgttgtgagaa cgttatcggttatatgccattaccagtcggtgttatcggtccattggttattgacggtacctcttaccacatcccaatg gctactactgaaggttgtttagtcgcctccgccatgagaggttgtaaggctatcaatgctggtggtggtgctactac cgtcttgactaaggatggtatgactagaggtccagttgtccgttttccaactttgaaaagatctggtgcttgtaagatt tggttggattctgaagaaggtcaaaatgccattaagaaggctttcaattccacctctagatttgccagattacaacat attcaaacctgtttagccggtgatttgttgttcatgagattcagaactactactggtgatgctatgggtatgaacatga tctctaagggtgtcgaatattctttaaaacaaatggttgaagagtatggttgggaagacatggaggtcgtctctgtct ctggtaactactgtactgataagaaaccagctgctatcaactggatcgaaggtcgtggtaagtctgttgttgccga agctactattccaggtgatgttgttagaaaggttttaaaatccgatgtctctgccttggttgagttgaacattgctaaa aacttggttggttctgctatggctggttctgtcggtggttttaatgcccatgccgccaacttagtcaccgccgttttctt agctttgggtcaagatccagctcaaaatgtcgaatcctccaactgtatcactttgatgaaagaggtcgacggtgac ttgcgtatctctgtttccatgccatctatcgaagttggtactatcggtggtggtactgttttggagccacaaggtgcta tgttggacttattgggtgttagaggtccacacgccactgctcctggtaccaacgccagacaattagctagaatcgt tgcctgtgccgtcttagctggtgagttgtctttatgtgctgccttagctgctggtcacttggtccaatcccacatgact cataacagaaagccagctgaacctaccaagcctaacaacttggatgccaccgatattaatcgtttaaaagatggtt ctgtcacttgcattaagtcctaa
SEQ ID NO:115 HMG-CoA synthase (Sc_ERG13) Saccharomyces sp.	MTELKKQKTAEQKTRPQNVGIKGIQIYIPTQCVNQSELEKFDGVSQGK YTIGLGQTNMSFVNDREDIYSMSLTVLSKLIKSYNIDTNKIGRLEVGTE TLIDKSKSVKSVLMQLFGENTDVEGIDTLNACYGGTNALFNSLNWIES NAWDGRDAIVVCGDIAIYDKGAARPTGGAGTVAMWIGPDAPIVFDSV RASYMEHAYDFYKPDFTSEYPYVDGHFSLTCYVKALDQVYKSYSKKA ISKGLVSDPAGSDALNVLKYFDYNVFHVPTCKLVTKSYGRLLYNDFRA NPQLFPEVDAELATRDYDESLTDKNIEKTFVNVAKPFHKERVAQSLIVP TNTGNMYTASVYAAFASLLNYVGSDDLQGKRVGLFSYGSGLAASLYS CKIVGDVQHIIKELDITNKLAKRITETPKDYEAAIELRENAHLKKNFKP QGSIEHLQSGVYYLTNIDDKFRRSYDVKK*
SEQ ID NO:116 Artificial HMG-CoA synthase (Sc_ERG13) nucleotide sequence	atgaccgaattgaagaagcaaaagactgctgaacaaaaaacccgtccacaaaatgttggtatcaagggtattca aatctacattccaactcaatgcgtcaaccaatctgaattggaaaaatttgatggtgtttctcaaggtaaatacactatt ggtttgggtcaaactaatatgtccttcgttaacgacagagaagatatttactccatgtccttgaccgtcttgtccaaat tgattaagtcttataatattgacaccaacaagatcggtagattggaagttggtactgaaactttgattgataagtctaa gtctgttaagtctgttttaatgcaattgttcggtgaaaatactgacgttgaaggtattgacactttgaacgcttgttacg gtggtactaatgctttatttaactctttgaactggattgaatccaacgcttgggacggtagagatgccattgtcgtttg tggtgacattgctatctatgacaagggtgccgctcgtccaactggtggtgctggtaccgttgctatgtggatcggt cctgacgccccaatcgtttttgattctgttcgtgcttcttacatggagcatgcttacgacttttataagccagactttac ctccgaatatccatacgtcgatggtcatttctctttgacctgctacgttaaagccttagatcaagtctacaagtcttact ccaagaaggccatttccaagggtttagtctccgatccagctggttccgatgctttaaacgttttaaagtacttcgatt acaacgtttttcatgtccctacttgtaaattggttaccaaatcttacggtagattattgtacaacgatttcagagctaat ccacaattatttccagaagtcgatgctgagttggctactagagattacgacgaatccttgaccgacaaaaatattga aaagacttttgttaacgttgctaagccatttcacaaagagagagttgcccaatctttgattgtcccaactaatactggt aatatgtatactgcttctgtttacgctgcctttgcttctttgttgaactatgtcggttctgacgacttacaaggtaagcgt gtcggtttgttctcctacggttccggtttggctgcctctttgtattcttgtaagattgtcggtgatgttcaacacatcatc aaggaattggatatcaccaataaattggccaagagaatcactgaaactcctaaagactatgaagctgctatcgaat tgagagaaaatgctcatttaaagaaaaactttaaaccacaaggttctattgaacacttgcaatccggtgtttactact taactaacatcgatgacaagttccgtagatcctacgacgtcaagaagtaa
SEQ ID NO:117 Pyruvate dehydrogenase	MSYTVGTYLAERLVQIGLKHHFAVAGDYNLVLLDNLLLNKNMEQVY CCNELNCGFSAEGYARAKGAAAAVVTYSVGALSAFDAIGGAYAENLP VILISGAPNNNDHAAGHVLHHALGKTDYHYQLEMAKNITAAAEAIYT PEEAPAKIDHVIKTALREKKPVYLEIACNIASMPCAAPGPASALFNDEA

SEQ ID	SEQUENCE
complex (Zm_PDC) *Zymomonas mobilis*	SDEASLNAAVEETLKFIANRDKVAVLVGSKLRAAGAEEAAVKFADAL GGAVATMAAAKSFFPEENPHYIGTSWGEVSYPGVEKTMKEADAVIAL APVFNDYSTTGWTDIPDPKKLVLAEPRSVVVNGIRFPSVHLKDYLTRL AQKVSKKTGALDFFKSLNAGELKKAAPADPSAPLVNAEIARQVEALLT PNTTVIAETGDSWFNAQRMKLPNGARVEYEMQWGHIGWSVPAAFGY AVGAPERRNILMVGDGSFQLTAQEVAQMVRLKLPVIIFLINNYGYTIEV MIHDGPYNNIKNWDYAGLMEVFNGNGGYDSGAGKGLKAKTGGELAE AIKVALANTDGPTLIECFIGREDCTEELVKWGKRVAAANSRKPVNKLL *
SEQ ID NO:118 Artificial pyruvate dehydrogenase complex (Zm_PDC) nucleotide sequence	atgtcctacaccgttggtacctacttagctgagcgtttggtccaaatcggtttgaagcaccatttcgccgttgctggt gattacaacttggtcttgttagataatttattattgaacaagaacatggaacaagtctactgctgtaatgaattgaact gtggtttctctgctgaaggttatgctagagctaaaggtgccgctgccgctgttgtcacttactctgttggtgctttgtc tgccttcgacgctattggtggtgcttacgccgagaatttacctgttatttttaatttctggtgcccctaacaataacgat catgctgctggtcatgttttacaccacgctttgggtaaaactgactaccattatcaattagagatggccaaaaacat caccgccgctgccgaggccatttacactccagaagaagccccagccaaaattgatcacgtcatcaaaaccgcc ttgagagagaaaaaacctgtttacttggaaatcgcctgtaatatcgcctctatgccttgcgccgctcctggtcctgc ttccgccttattcaacgatgaggcttctgatgaagcttccttaaacgctgctgttgaggagactttaaagttcatcgct aatagagataaggtcgctgttttagtcggttctaagttgcgtgctgccggtgccgaggaagctgctgttaaattcgc cgatgctttaggtggtgctgtcgccaccatggccgccgccaaatccttttccctgaagaaaacccacactacatc ggtacttcttggggtgaagtctcttacccaggtgtcgaaaagactatgaaggaagccgatgccgtcatcgccttg gccccagtttttaatgattattccaccactggttggactgatatcccagatcctaaaaagttagttttagccgagccta gatccgttgttgttaacggtattagattcccttccgttcacttgaaggattacttaactagattggctcaaaaggtttcc aagaagaccggtgctttggacttttcaaatctttgaacgccggtgagttaaagaaggccgcccctgctgacccat ctgctccattggttaacgctgagattgctagacaagtcgaagctttattgaccccaaacactaccgttatcgccgaa actggtgactcttggtttaatgctcaaagaatgaagttaccaaatggtgccagagttgagtacgaaatgcaatggg gtcatatcggttggtctgtcccagctgcttttggttatgctgttggtgcccctgagagaagaaacatcttgatggttg gtgacggttccttccaattgactgctcaagaagtcgctcaaatggttagattaaaattaccagtcatcatcttcttgat caataactacggttacactatcgaagtcatgattcacgatggtccttacaataatattaagaactgggactatgctg gtttgatggaagtctttaatggtaacggtggttacgattccggtgctggtaagggtttaaaggctaagactggtggt gaattagctgaagccattaaggttgccttggctaacaccgacggtcctactttaatcgaatgtttcattggtagaga ggattgtaccgaagagttagttaagtggggtaagagagttgccgctgctaattcccgtaagcctgtcaataaattg ttataa
SEQ ID NO:119 Truncated 3-hydroxy-3-methyl-glutaryl-CoA reductase (tHMG1)	atgcaattggtgaagactgaagtcaccaagaagtcttttactgctcctgtacaaaaggcttctacaccagttttaac caataaaacagtcatttctggatcgaaagtcaaaagtttatcatctgcgcaatcgagctcatcaggaccttcatcat ctagtgaggaagatgattcccgcgatattgaaagcttggataagaaaatacgtcctttagaagaattagaagcatt attaagtagtggaaatacaaaacaattgaagaacaaagaggtcgctgccttggttattcacggtaagttacctttgt acgctttggagaaaaaattaggtgatactacgagagcggttgcggtacgtaggaaggctctttcaattttggcaga agctcctgtattagcatctgatcgtttaccatataaaaattatgactacgaccgcgtatttggcgcttgttgtgaaaat gttataggttacatgcctttgcccgttggtgttataggccccttggttatcgatggtacatcttatcatataccaatggc aactacagagggttgtttggtagcttctgccatgcgtggctgtaaggcaatcaatgctggcggtggtgcaacaac tgttttaactaaggatggtatgacaagaggcccagtagtccgtttcccaactttgaaaagatctggtgcctgtaaga tatggttagactcagaagagggacaaaacgcaattaaaaaagcttttaactctacatcaagatttgcacgtctgca acatattcaaacttgtctagcaggagatttactcttcatgagatttagaacaactactggtgacgcaatgggtatgaa tatgatttctaagggtgtcgaatactcattaaagcaaatggtagaagagtatggctgggaagatatggaggttgtct ccgtttctggtaactactgtaccgacaaaaaaccagctgccatcaactggatcgaaggtcgtggtaagagtgtcg tcgcagaagctactattcctggtgatgttgtcagaaaagtgttaaaaagtgatgtttccgcattggttgagttgaaca ttgctaagaatttggttggatctgcaatggctgggtctgttggtggatttaacgcacatgcagctaatttagtgacag ctgttttcttggcattaggacaagatcctgcacaaaatgtcgaaagttccaactgtataacattgatgaaagaagtg gacggtgatttgagaatttccgtatccatgccatccatcgaagtaggtaccatcggtggtggtactgttctagaacc acaaggtgccatgttggacttattaggtgtaagaggcccacatgctaccgctcctggtaccaacgcacgtcaatt agcaagaatagttgcctgtgccgtcttggcaggtgaattatccttatgtgctgccctagcagccggccatttggttc aaagtcatatgacccacaacaggaaacctgctgaaccaacaaaacctaacaatttggacgccactgatataaatc gtttgaaagatgggtccgtcacctgcattaaatcctaa

SEQ ID	SEQUENCE
SEQ ID NO:120 HMG-CoA synthase (Sc_ERG13) Saccharomyces sp.	atgactgaactaaaaaaacaaaagaccgctgaacaaaaaaccagacctcaaaatgtcggtattaaaggtatcca aatttacatcccaactcaatgtgtcaaccaatctgagctagagaaatttgatggcgtttctcaaggtaaatacacaat tggtctgggccaaaccaacatgtcttttgtcaatgacagagaagatatctactcgatgtccctaactgttttgtctaa gttgatcaagagttacaacatcgacaccaacaaaattggtagattagaagtcggtactgaaactctgattgacaag tccaagtctgtcaagtctgtcttgatgcaattgtttggtgaaaacactgacgtcgaaggtattgacacgcttaatgcc tgttacggtggtaccaacgcgttgttcaactctttgaactggattgaatctaacgcatgggatggtagagacgccat tgtagtttgcggtgatattgccatctacgataagggtgccgcaagaccaaccggtggtgccggtactgttgctatg tggatcggtcctgatgctccaattgtatttgactctgtaagagcttcttacatggaacacgcctacgatttttacaagc cagatttcaccagcgaatatccttacgtcgatggtcatttttcattaacttgttacgtcaaggctcttgatcaagtttac aagagttattccaagaaggctatttctaaagggttggttagcgatcccgctggttcggatgctttgaacgttttgaaa tatttcgactacaacgttttccatgttccaacctgtaaattggtcacaaaatcatacggtagattactatataacgattt cagagccaatcctcaattgttcccagaagttgacgccgaattagctactcgcgattatgacgaatctttaaccgata agaacattgaaaaaacttttgttaatgttgctaagccattccacaaagagagagttgcccaatctttgattgttccaa caaacacaggtaacatgtacaccgcatctgtttatgccgcctttgcatctctattaaactatgttggatctgacgactt acaaggcaagcgtgttggtttatttttcttacggttccggtttagctgcatctctatattcttgcaaaattgttggtgacgt ccaacatattatcaaggaattagatattactaacaaattagccaagagaatcaccgaaactccaaaggattacgaa gctgccatcgaattgagagaaaatgcccatttgaagaagaacttcaaacctcaaggttccattgagcatttgcaaa gtggtgtttactacttgaccaacatcgatgacaaatttagaagatcttacgatgttaaaaaataat
SEQ ID NO:121 Geranylgeranyl pyrophosphate synthase (Cr_GPPS) Catharanthus sp.	MLFSRGLYRIARTSLNRSRLLYPLQSQSPELLQSFQFRSPIGSSQKVSGF RVIYSCVSSALANVGQQVQRQSNSVAEEPLDPFSLVADELSILANRLRS MVVAEVPKLASAAEYFFKLGVEGKRFRPTVLLLMATAIDAPISRTPPD TSLDTLSTELRLRQQSIAEITEMIHVASLLHDDVLDDAETRRGIGSLNFV MGNKLAVLAGDFLLSRACVALASLKNTEVVSLLATVVEHLVTGETMQ MTTTSDQRCSMEYYMQKTYYKTASLISNSCKAIALLAGQTSEVAMLA YEYGKNLGLAFQLIDDVLDFTGTSASLGKGSLSDIRHGIVTAPILFAIEE FPELRAVVDEGFENPYNVDLALHYLGKSRGIQRTRELAIKHANLASDA IDSLPVTDDEHVLRSRRALVELTQRVITRRK*
SEQ ID NO:122 Artificial geranylgeranyl pyrophosphate synthase (Cr_GPPS) nucleotide sequence	atgttattctctcgtggtttatacagaatcgccagaacttctttgaacagatcccgtttgttgtaccctttacaatctcaa tctcctgaattgttacaatccttccaattcagatctccaatcggttcctctcaaaaggtttccggtttcagagttatcta ctcctgcgtttcctctgctttagctaacgttggtcaacaagtccaaagacaatctaattccgttgctgaagaacctttg gacccattctccttggttgccgatgaattatccattttagctaacagattgcgttctatggtcgtcgctgaagttccaa agttagcctccgccgccgaatatttcttcaagttgggtgtcgagggtaaaagattcagaccaactgttttgttgttaa tggccaccgccattgatgccccaatctctagaaccccacctgacacctccttagatactttatccaccgaattgcgt ttgagacaacaatctatcgccgaaattactgaaatgattcatgtcgcttccttgttgcacgatgatgttttggatgatg ctgaaactagaagaggtattggttctttaaattttgtcatgggtaacaaattggctgttttggccggtgacttcttattat ctagagcttgtgttgccttagcttctttgaaaaacactgaagtcgtctccttgttagccactgtcgttgaacacttagtt actggtgagactatgcaaatgactaccacctccgatcaaagatgttctatggaatactacatgcaaaagacctatt acaagactgcctctttgatttctaactcctgtaaagccattgccttgttagctggtcaaacttctgaagttgccatgttg gcttacgaatacggtaaaaacttgggtttggctttccaattgattgatgatgttttggatttcactggtacttctgcttcc ttaggtaaaggttctttgtctgatattcgtcacggtatcgttaccgccccaatcttgttcgctattgaagaattcccag agttaagagctgttgttgacgaaggtttcgaaaacccttacaatgttgacttagccttgcactacttgggtaaatcta gaggtattcaacgtaccagagaattagccattaaacatgctaacttagcctctgacgccattgactctttaccagtc actgatgatgagcacgtcttacgttccagacgtgccttagttgaattgactcaaagagttattactagaagaaagta a
SEQ ID NO:123 Geranylgeranyl pyrophosphate synthase (Mi_GPPS1) Mangifera indica	MLFSYGLSRISINPRASLLTCRWLLSHLTGSLSPSTSSHTISDSVHKVWG CREAYTWSVPALHGFRHQIHHQSSSLIEDQLDPFSLVADELSLVANRLR SMVVTEVPKLASAAEYFFKMGVEGKRFRPAVLLLMATALNVHVLEPL PEGAGDALMTELRTRQQCIAEITEMIHVASLLHDDVLDDADTRRGIGS LNLVMGNKLAVLAGDFLLSRACVALASLKNTEVVSLLATVVEHLVTG ETMQMTTSSDQRCSMEYYMQKTYYKTASLISNSCKAIALLAGQSAEV AMLAFEFGKNLGLAYQLIDDVLDFTGTSASLGKGSLSDIRHGIVTAPIL FAMEEFPQLRAVIDQGFENPSNVDVALEYLGKSRGIQRTRELATNHAN LAAAAIDALPKTDNEEVRKSRRALLDLTQRVITRNK*

SEQ ID	SEQUENCE
SEQ ID NO:124 Artificial geranylgeranyl pyrophosphate synthase (Mi_GPPS1) nucleotide sequence	atgttattctcttatggtttatctcgtatttctattaaccctcgtgcctctttattgacttgtagatggttattatcccatttga ctggttctttatctccttccacttcttcccatactatttctgactccgtccataaagtctggggttgcagagaagcctat acttggtctgtcccagctttacatggttttagacatcaaatccaccatcaatcctcttccttgattgaagatcaattaga cccattctccttggtcgccgatgagttgtccttggttgctaaccgtttaagatctatggttgtcactgaagtccctaaa ttagcctctgccgccgaatacttttttcaagatgggtgtcgaaggtaagcgtttcagaccagctgtcttgttgttaatg gccactgccttaaacgttcatgttttggaacctttgcctgaaggtgctggtgacgctttaatgaccgagttgagaac ccgtcaacaatgcattgctgaaatcactgagatgatccacgtcgcctctttattgcatgacgatgttttagacgacg ctgatactagaagaggtattggttctttgaacttggttatgggtaacaaattggccgttttggccggtgatttcttgtta tcccgtgcttgcgttgctttagcttctttgaagaacactgaagttgtttctttgttggccaccgtcgttgaacacttagtt actggtgagactatgcaaatgaccacctcttctgaccaaagatgttccatggaatattacatgcaaaaaacttatta caaaaccgcctccttgatttctaactcctgtaaagccatcgccttattagctggtcaatctgctgaagttgccatgtta gccttcgagtttggtaagaacttgggtttagcttaccaattgatcgatgatgtcttggatttaccggtacctctgcttc tttgggtaagggttccttgtccgacattagacacggtattgttaccgccccaatcttattcgctatggaagagtttcc acaattgagagctgttatcgaccaaggtttcgagaacccatctaacgttgacgtcgccttagagtatttaggtaaat ctagaggtatccaacgtacccgtgaattagctactaaccatgctaacttagccgccgccgccatcgatgccttgc ctaaaaccgataatgaagaagtccgtaagtccagacgtgctttattagatttgactcaaagagtcatcaccagaaa caaatag
SEQ ID NO:125 Geranylgeranyl pyrophosphate synthase (Mi_GPPS2) Mangifera indica	MPFVVPRRNRSLSVSAVLTKEETLREEEEDPKPVFDFKSYMLQKGNSV NQALDAVVSIREPKKIHEAMRYSLLAGGKRVRPVLCIAACELVGGNES MAMPAACAVEMIHTMSLIHDDLPCMDNDDLRRGKPTNHKVFGEDVA VLAGDALLAFSFENMAVSTVGVLPSRVVKAVGELAKSIGIEGLVAGQV VDINSEGLKEVGLDHLEFIHQHKTAALLEGSVVLGAILGGGSDDEVEK LRTFARCIGLLFQVVDDILDVTKSSRELGKTAGKDLVADKVTYPKLLGI EKSRELADKLNKDAQQQLSGFDQEKAAPLIALSNYIAYRQN*
SEQ ID NO:126 Artificial geranylgeranyl pyrophosphate synthase (Mi_GPPS2) nucleotide sequence	atgccattcgttgttcctagaagaaaccgttctttgtccgtttccgccgttttgaccaaggaagaaactttaagagag gaagaagaagatccaaagccagttttcgacttcaaatcttacatgttacaaaagggtaattctgttaatcaagctttg gatgctgtcgtttccattagagaacctaagaaaatccatgaggctatgcgttactctttgttggctggtggtaagag agttcgtcctgttttgtgtattgccgcctgtgaattggtcggtggtaacgaatctatggctatgccagccgcctgtgc tgtcgaaatgatccacactatgtccttgattcacgatgatttgccatgtatggataatgacgatttgcgtcgtggtaa acctaccaaccataaagttttcggtgaagacgtcgccgttttggctggtgacgctttattagctttttccttcgagaac atggccgtttccactgttggtgtcttaccatccagagttgtcaaggctgttggtgaattggccaagtctatcggtatt gaaggtttggttgccggtcaagtcgtcgatattaattctgagggtttaaaagaggtcggtttagatcacttagaattt atccatcaacacaaaaccgctgctttgttggagggttctgttgttttgggtgctattttaggtggtggttctgatgatg aagtcgaaaagttgcgtaccttgctagatgtatcggtttgttgtttcaagttgttgacgatattttggatgtcactaag tcttccagagaattgggtaagactgccggtaaagatttggttgctgataaagttacttatccaaagttgttaggtattg aaaagtctcgtgaattggccgataagttaaacaaggatgctcaacaacaattatccggttttgatcaagagaaggc tgccccttaatcgctttgtccaattacatcgcctacagacaaaactag
SEQ ID NO:127 Truncated geranylgeranyl pyrophosphate synthase Cs2_GPPS_NTrunc	MVIAEVPKLASAAEYFFKMGVEGKRFRPTVLLLMATALNVRVPEPLH DGVEDASATELRTRQQCIAEITEMIHVASLLHDDVLDDADTRRGIGSL NFVMGNKLAVLAGDFLLSRACVALASLKNTEVVTLLATVVEHLVTGE TMQMTTSSDQRCSMDYYMQKTYYKTASLISNSCKAIALLAGQTAEVA ILAFDYGKNLGLAYQLIDDVLDFTGTSASLGKGSLSDIRHGIITAPILFA MEEFPQLRTVVEQGFEDSSNVDIALEYLGKSRGIQKTRELAVKHANLA AAAIDSLPENNDEDVTKSRRALLDLTHRVITRNK*
SEQ ID NO:128 Truncated geranylgeranyl pyrophosphate synthase Cs2_GPPS_NTrunc	atggtcattgctgaagttcctaaattagcctctgccgccgaatacttcttcaagatgggtgtcgagggtaagagatt tcgtcctaccgttttgttgttaatggccaccgccttaaacgtcagagtccctgaaccattacatgatggtgttgaaga tgcctctgccaccgagttgagaactagacaacaatgtattgctgaaatcaccgagatgattcacgttgcctctttgtt gcacgatgatgttttggatgatgctgatacccgtcgtggtatcggttctttgaactttgtcatgggtaacaagttggct gtcttggctggtgatttcttattgtctcgtgcctgcgttgccttagcctctttaaaaaatacgaagttgttacttttattg gccactgttgttgagcacttggttactggtgaaactatgcaaatgaccacctcttccgaccaacgttgttccatgga ctattacatgcaaaagacctactacaagaccgcttctttgatttccaattcttgtaaagccattgccttattagctggtc aaactgctgaagttgccatcttggccttcgactacggtaaaaacttgggtttagcttaccaattgattgatgacgtttt agattttactggtacttctgcttctttgggtaaaggttctttatccgatattcgtcatggtatcattaccgctccaatctta

SEQ ID	SEQUENCE
	ttcgctatggaagaatttcctcaattgcgtactgtcgttgaacaaggtttcgaagactcctccaacgttgacattgcc ttagaatacttgggtaagtctcgtggtattcaaaagacccgtgaattagccgttaaacatgccaacttagccgccg ccgccatcgattccttgcctgaaaacaacgatgaggatgtcaccaagtcccgtcgtgctttgttagatttaactcac agagttattacccgtaacaagtaa
SEQ ID NO:129 Geranylgeranyl pyrophosphate synthase (Qr_GPPS) Quercus sp.	MLFSRISRIRRPGSNGFRWFLSHKTHLQFLNPPAYSYSSTHKVLGCREIF SWGLPALHGFRHNIHHQSSSIVEEQNDPFSLVADELSMVANRLRSMVV TEVPKLASAAEYFFKMGVEGKRFRPTVLLLMATAMNISILEPSLRGPG DALTTELRARQQRIAEITEMIHVASLLHDDVLDDADTRRGIGSLNFVM GNKLAVLAGDFLLSRACVALASLKNTEVVSLLAKVVEHLVTGETMQ MTTTCEQRCSMEYYMQKTYYKTASLISNSCKAIALLGGQTSEVAMLA YEYGKNLGLAYQLIDDVLDFTGTSASLGKGSLSDIRHGIITAPILFAMEE FPQLREVVDRGFDDPANVDVALDYLGKSRGIQRARELAKKHANIAAE AIDSLPESNDEDVRKSRRALLDLTERVITRTK*
SEQ ID NO:130 Artificial geranylgeranyl pyrophosphate synthase (Qr_GPPS) nucleotide sequence	atgttgttctctcgtatttctcgtatccgtagaccaggttctaatggtttcagatggttcttgtcccataagactcattta caattcttgaaccctccagcttattcctactcttccactcataaggtcttgggttgtagagaaatttttttcctggggttta cctgccttacatggtttcagacacaacattcaccaccaatcttcctctattgttgaagaacaaaatgacccttttctctttt ggtcgctgatgagttgtccatggttgctaacagattgcgttctatggttgttactgaagttcctaaattagcctccgcc gctgaatacttttttaaaatgggtgttgaaggtaagagattcagaccaactgttttattgttgatggctaccgccatga acatttccatcttagaaccatctttgagaggtccaggtgacgctttgaccactgaattgagagccagacaacaaag aattgctgaaattaccgagatgatccacgttgcttccttgttgcacgatgacgttttggatgacgctgatactagaag aggtattggttccttaaacttttgtcatgggtaataaattagctgttttggctggtgatttttttgttatctcgtgcctgtgttg ctttagcttctttgaagaacaccgaagttgtctccttgttagccaaggtcgtcgaacacttggttactggtgaaactat gcaaatgaccactacttgtgaacaaagatgttccatggaatactacatgcaaaagacttactataagaccgcttctt taatttccaactcctgtaaagccattgctttattaggtggtcaaacttctgaggtcgctatgttagcctacgaatatgg taaaaacttgggtttagcttaccaattgattgatgatgtcttggatttcactggtacttctgcttccttgggtaagggttc cttgtctgatattagacatggtatcattactgctccaattttgtttgctatggaagaattcccacaattacgtgaagttgt cgatagaggtttcgacgatcctgccaacgtcgatgttgccttggactacttgggtaagtctagaggtatccaaaga gccagagagttagctaaaaaacacgctaacattgctgccgaagccatcgactctttgccagaatccaacgacga ggacgtcagaaagtcccgtcgtgctttgttggacttgaccgaaagagtcattactcgtactaagtaa
SEQ ID NO:131 Truncated geranylgeranyl pyrophosphate synthase Pa_GPPS_Ntrunc	MYTRCILRDKYSRFNLRRKFFTSAKSINALNGLPDSGNPRGESNGISQF EIQQVFRCKEYIWIDRHKFHDVGFQAHHKGSITDEEQVDPFSLVADELS ILANRLRSMILTEIPKLGTAAEYFFKLGVEGKRFRPMVLLLMASSLTIGI PEVAADCLRKGLDEEQRLRQQRIAEITEMIHVASLLHDDVLDDADTRR GVGSLNFVMGNKLAVLAGDFLLSRASVALASLKNTEVVELLSKVLEH LVTGEIMQMTNTNEQRCSMEYYMQKTFYKTASLMANSCKAIALIAGQ PAEVCMLAYDYGRNLGLAYQLLDDVLDFTGTTASLGKGSLSDIRQGIV TAPILFALEEFPQLHDVINRKFKKPGDIDLALEFLGKSDGIRKAKQLAA QHAGLAAFSVESFPPSESEYVKLCRKALIDLSEKVITRTR*
SEQ ID NO:132 Truncated geranylgeranyl pyrophosphate synthase Pa_GPPS_N_trunc	atgtatacccgttgcattttaagagacaagtattctcgtttcaacttgagacgtaaattcttcacttccgctaaatccat caatgccttgaatggtttacctgactctggtaaccctagaggtgaatctaacggtatctcccaattcgaaattcaac aagttttccgttgtaaagaatacatttggatcgatcgtcacaagttccacgatgttggttttcaagctcatcacaagg gttccatcactgacgaggaacaagttgacccttttttctttagtcgctgatgaattgtccatcttagctaatcgtttaaga tccatgatcttaaccgagattccaaagttaggtaccgctgccgaatacttttttcaagttgggtgtcgaaggtaagag atttagaccaatggtttttgttgttgatggcctcctctttaactattggtatccctgaagttgccgctgattgtttgcgtaa gggtttggacgaagaacaaagattacgtcaacaacgtatcgctgaaattactgaaatgattcatgtcgcctctttgt tgcacgatgatgttttggatgacgccgatactagacgtggtgttggttccttgaactttgttatgggtaacaagttgg ctgttttagccggtgatttcttgttatctagagcttctgttgccttagcttctttaaagaacactgaggttgttgagttatt gtctaaggttttggagcacttagtcactggtgagatcatgcaaatgactaacactaatgaacaaagatgttctatgg aatattacatgcaaaagactttctacaagaccgcctctttgatggctaattcttgtaaagccattgccttgatcgctgg tcaacctgccgaagtctgcatgttggcctacgactacggtagaaacttgggtttagcttatcaattattggatgacgt ttggatttcactggtaccactgcttctttaggtaagggttccttatccgacatcagacaaggtattgttactgccccct attttattcgctttggaagaattccctcaattacacgacgtcatcaaccgtaagttcaaaaaaccaggtgacatcgat ttggccttggaatttttgggtaagtctgatggtatccgtaaagccaaacaattggctgctcaacatgctggtttagct

SEQ ID	SEQUENCE
	gcctttctgtcgaatcctttccaccatctgaatccgaatacgttaagttatgtagaaaggccttgatcgatttgtctga aaaggtcattactcgtaccagataa
SEQ ID NO:133 Geranylgeranyl pyrophosphate synthase Ag_GPPS Abies grandis	MAYSAMATMGYNGMAASCHTLHPTSPLKPFHGASTSLEAFNGEHMG LLRGYSKRKLSSYKNPASRSSNATVAQLLNPPQKGKKAVEFDFNKYM DSKAMTVNEALNKAIPLRYPQKIYESMRYSLLAGGKRVRPVLCIAACE LVGGTEELAIPTACAIEMIHTMSLMHDDLPCIDNDDLRRGKPTNHKIFG EDTAVTAGNALHSYAFEHIAVSTSKTVGADRILRMVSELGRATGSEGV MGGQMVDIASEGDPSIDLQTLEWIHIHKTAMLLECSVVCGAIIGGASEI VIERARRYARCVGLLFQVVDDILDVTKSSDELGKTAGKDLISDKATYP KLMGLEKAKEFSDELLNRAKGELSCFDPVKAAPLLGLADYVAFRQN*
SEQ ID NO:134 Artificial geranylgeranyl pyrophosphate synthase (Ag_GPPS) nucleotide sequence	atggcttattctgctatggctactatgggttacaacggtatggctgcttcttgtcacactttacacccaacttctccatt gaaacctttcacggtgcttctacttccttggaagccttcaatggtgaacacatgggtttgttaagaggttattctaag cgtaagttgtcctcttacaaaaatccagcttctcgttcctccaatgctaccgtcgctcaattattgaacccaccacaa aagggtaagaaggctgttgaatttgacttcaataagtatatggattctaaggctatgaccgtcaacgaggctttgaa taaagccatcccattgcgttacccacaaaagatctacgaatctatgagatattctttgttagctggtggtaagagagt ccgtccagttttgtgtatcgccgcttgtgaattagtcggtggtactgaggagttagctattccaaccgcctgtgccat cgaaatgatccacaccatgtctttgatgcacgatgatttgccatgtatcgacaacgatgacttgagacgtggtaaa cctaccaatcataagattttcggtgaagatactgctgttactgccggtaacgctttacactcttacgccttcgaacat attgctgtttctacttccaagactgttggtgctgatagaattttgagaatggtttctgaattaggtcgtgctactggttc cgaaggtgttatgggtggtcaaatggtcgatattgcttctgaaggtgacccttccattgatttgcaaactttagaatg gatccacatccacaagactgctatgttattagaatgttctgttgtctgtggtgccatcatcggtggtgcttctgaaatt gttattgagagagccagacgttatgctcgttgtgtcggtttattgtttcaagttgttgacgacattttagatgttaccaa atcttctgacgaattgggtaaaactgctggtaaagatttaatctccgataaagccacctaccctaagttgatgggttt ggagaaggccaaagagttttccgatgaattattaaacagagctaaaggtgaattgtcttgcttcgatccagttaag gctgccccattgttaggtttggctgactacgttgccttcagacaaaactaa
SEQ ID NO:135 Truncated geranylgeranyl pyrophosphate synthase Pb_GPPS_NTrunc	MAAIFPSIPSNFKPPQISQTLTRRRRPNRTLCTATSDQSYLSASSADIYSH LLRSLPATIHPSVKAPIHSLLSSPIPPTIAPPLCLAATELVGGNPNSAINAA CAIHLIHAVTHTRTAPPLAEFSPGVLLMTGDGLLVLAYEMLARSPAVD ADTSVRVLKEVARTAAAVAAAYEGGREGELAAGAAACGVILGGGNE EEVERGRRVGMFAGKMELVEAEVELRLGFEDAKAGAVRRLLEEMRF TQSFVNVRNPFYGK*
SEQ ID NO:136 Truncated geranylgeranyl pyrophosphate synthase Pb_GPPS_NTrunc	atggctgctatctttccatccattccatccaacttcaaaccacctcaaatctctcaaactttgaccagacgtagaaga ccaaaccgtactttatgtactgccacctctgaccaatcttacttgtccgcttcttctgccgacatttattctcatttgtta agatctttaccagctactattcatccatctgttaaagccccaatccattctttattgtcctctccaattcctccaaccatc gctccacctttgtgtttagctgctaccgaattggttggtggtaaccctaactctgccattaacgccgcctgtgccatt catttgattcatgctgttactcatactagaaccgctccaccattagctgaattttctcctggtgttttgttgatgactggt gatggtttattagttttggcttacgagatgttggccagatccccagctgttgatgccgatacttctgtccgtgttttgaa ggaagtcgctagaaccgccgccgccgtcgccgctgcttatgaaggtggtagagaagggtgaattagctgccggt gccgctgcttgtggtgtcattttgggtggtggtaacgaagaagaggtcgaaagaggtcgtagagtcggtatgttc gctggtaaaatggaattagttgaagctgaagtcgaattgagattgggtttcgaagatgctaaagccggtgccgtta gaagattgttggaagaaatgcgtttcacccaatcttttgtcaacgttagaaaccctttttatggtaagtaa
SEQ ID NO:137 Geranylgeranyl pyrophosphate synthase (Ai_GPPS) Azadirachta indica	MLFSRGLSRISRIPRNSLIGCRWLVSYRPDTILSGSSHSVGDSTQKVLGC REAYLWSLPALHGIRHQIHQQSSSLIEEELDPFSLVADELSLVANRLRS MVVAEVPKLASAAEYFFKMGVEGKRFRPTVLLLMASALNVQVPQPLS DGVGDALTTELRTRQQCIAEITEMIHVASLLHDDVLDDADTRRGIGSL NFVMGNKLAVLAGDFLLSRACVALASLKNTEVVSLLATVVEHLVTGE TMQMTTTAEQRRSMDYYMQKTYYKTASLISNSCKAIALLAGQTTEVA MLAFDYGKNLGLAFQLIDDVLDFTGTSASLGKGSLSDIRHGIVTAPILF AMEEFPELRKVVDKGFDDPSNVDIALEYLGKSRGIQRTRELAQKHANL ATVALDSLPESNDDDVKKSRRALLDLAQRVITRNK*
SEQ ID NO:138	atgttgttttccagaggtttatctcgtatttccagaatcccacgtaactctttgatcggttgtagatggttagtttcttacc

SEQ ID	SEQUENCE
Artificial geranylgeranyl pyrophosphate synthase (Ai_GPPS) nucleotide sequence	gtcctgataccattttatctggttcctctcactccgttggtgactctactcaaaaggttttaggttgtcgtgaagcttact tgtggtctttaccagccttgcacggtattagacaccaaattcatcaacaatcctcttctttgattgaagaagaattgga tccattctctttagttgctgatgaattgtctttagtcgctaaccgtttgagatccatggtcgtcgctgaagtcccaaaat tagcctccgccgccgagtacttcttcaagatgggtgttgagggtaagagattccgtccaactgtcttattgttgatg gcctccgccttaaacgttcaagtcccacaacctttgtctgacggtgttggtgatgctttgactaccgagttgagaac tagacaacaatgcattgctgagattactgaaatgatccatgttgcttctttgttgcatgacgacgttttggatgatgct gacactagacgtggtatcggttctttgaacttcgttatgggtaacaagttggctgtcttggctggtgatttcttgttgtc cagagcctgtgttgctttagcttccttgaagaatactgaggttgtctctttgttggccaccgttgttgaacacttggtc accggtgaaactatgcaaatgactactactgctgaacaaagacgttccatggattattacatgcaaaagacttact ataagaccgcctctttgatttccaactcttgtaaagccattgccttgttagctggtcaaactaccgaagttgctatgtt ggctttcgattacggtaagaatttgggtttagcttttcaattgatcgatgacgtcttggatttactggtacctctgcttc tttaggtaaaggttccttgtctgatattagacacggtatcgttaccgctccaattttattcgctatggaagaattccca gaattaagaaaggttgttgataagggttttgacgacccttccaacgttgacattgctttggagtatttgggtaagtct agaggtattcaaagaaccagagaattggctcaaaaacatgccaatttggccaccgtcgccttggattctttaccag aatccaacgacgacgatgttaagaagtctcgtagagctttattggacttggctcaaagagttattactagaaacaa gtaa
SEQ ID NO:139 Truncated geranylgeranyl pyrophosphate synthase Es_GPPS_NTrunc	MRRSGSATAAAAATLARHANACCRARSPALGLLPGAAASSSTHRAAL SSNSGHGGDGSGHYDAAMRRRESCASRSRHRWSGQEAAAASATTTT ARRAPGGVAGASGQGSAAGSVRALSSSFLADAVRETATNHCIDRVVN GGLDGSVPVDKDTPTVEVQDFVYDIDFAQRPSGASQSLADGPDPFELV SAELAGLSDGIKSLIGTEHAVLNAAAKYFFELDGGKKIRPTMVILMSQA CNSNSQQVRPDVQPGTELVNPLQLRLAEITEMIHAASLFHDDVIDEADT RRGVPSVNKVFGNKLAILAGDFLLARSSMSLARLRSLESVELMSAAIE HLVKGEVLQMRPTEDGGGAFEYYVRKNYYKTGSLMANSCKASAVLG QHDLEVQEVAFEYGKRVGLAFQLVDDILDFEGNTFTLGKPALNDLRQ GLATAPVLLAAEQQPGLAKLISRKFRGPGDVDEALELVHRSDGIARAK EVAVVQAEKAMSAILTLHDSPAQNALVQLAHKIVNRNH*
SEQ ID NO:140 Truncated geranylgeranyl pyrophosphate synthase Es_GPPS_NTrunc	atgcgtagatccggttccgctaccgccgctgccgctgccaccttagccagacacgccaacgcctgttgtagagc ccgttccccagctttaggtttgttgcctggtgccgccgcttcttcctctactcacagagccgccttgtcttctaattct ggtcatggtggtgatggttccggtcattacgacgctgctatgagaagaagagaatcttgcgcttccagatctcgtc acagatggtccggtcaagaagctgccgccgcctccgccactaccaccaccgctcgtcgtgctccaggtggtgt cgccggtgcttctggtcaaggttctgctgccggttctgttagagccttatcctcttctttttttagccgatgccgttcgtg aaaccgctactaaccactgtatcgaccgtgttgtcaacggtggtttggacggttctgtcccagtcgataaagatac cccaactgtcgaagttcaagactttgtttatgatattgactttgctcaacgtccatccggtgcctctcaatctttagctg acggtccagatccattcgagttagtttccgctgagttggccggtttgtctgatggtattaagtctttgattggtaccga acatgctgtcttgaacgccgccgccaaatatttcttcgaattagatggtggtaaaaagatcagacctactatggttat cttaatgtcccaagcttgtaactctaattcccaacaagttcgtcctgacgttcaaccaggtactgaattagtcaatcct ttgcaattaagattggctgaaatcaccgagatgattcatgctgcttctttattccacgacgatgttattgatgaggctg atactagacgtggtgtcccttctgttaataaagttttcggtaacaaattagccatcttggccggtgacttcttattggct agatcctctatgtccttggcccgtttaagatccttggagtccgtcgaattgatgtccgccgctatcgaacacttggtc aaaggtgaagttttacaaatgcgtccaactgaggacggtggtggtgctttcgagtactacgtcagaaaaaattact acaagactggttctttgatggctaactcctgtaaggcctccgccgttttaggtcaacacgacttagaagtccaaga ggtcgcttttgaatacggtaagagagtcggtttggctttccaattggttgacgatattttagattttgaaggtaatactt tcactttgggtaagccagctttaaacgacttgagacaaggtttagccactgcccctgtcttgttagctgctgaacaa caacctggtttagctaaattgatctccagaaagtttagaggtcctggtgatgtcgatgaagctttggaattggtcca cagatccgacggtattgctagagctaaggaggttgctgttgtccaagccgaaaaagctatgtctgccattttgacc ttgcatgactccccagctcaaaatgctttggttcaattggctcacaaaatcgtcaatcgtaaccattag
SEQ ID NO:141 Geranylgeranyl pyrophosphate synthase (Si_GPPS) Solanum sp.	MIFSKGLAQISRNRFSRCRWLFSLRPIPQLHQSNHIHDPPKVLGCRVIHS WVSNALSGIGQQIHQQSTAVAEEQVDPFSLVADELSLLTNRLRSMVVA EVPKLASAAEYFFKLGVEGKRFRPTVLLLMATALNVQIPRSAPQVDVD SFSGDLRTRQQCIAEITEMIHVASLLHDDVLDDADTRRGIGSLNFVMG NKLAVLAGDFLLSRACVALASLKNTEVVCLLATVVEHLVTGETMQMT TSSDERCSMEYYMQKTYYKTASLISNSCKAIALLAGHSAEVSVLAFDY

SEQ ID	SEQUENCE
	GKNLGLAFQLIDDVLDFTGTSATLGKGSLSDIRHGIVTAPILYAMEEFP QLRTLVDRGFDDPVNVEIALDYLGKSRGIQRTRELARKHASLASAAID SLPESDDEEVQRSRRALVELTHRVITRTK*
SEQ ID NO:142 Artificial geranylgeranyl pyrophosphate synthase (Si_GPPS) nucleotide sequence	atgatcttttccaagggtttagctcaaatctctcgtaatagattctctcgttgcagatggttattctctttgcgtccaattc ctcaattacaccaatccaatcacatccacgacccaccaaaagttttgggttgtcgtgtcattcactcttgggtttctaa tgccttgtctggtatcggtcaacaaatccatcaacaatctactgccgttgccgaggaacaagtcgacccttttttctttt ggttgctgatgagttatccttgttaaccaacagattgagatccatggttgtcgctgaagtccctaagttagcctccgc cgctgagtatttctttaagttaggtgtcgaaggtaaacgtttccgtccaactgtcttgttgttgatggccactgcctta aacgtccaaattcctcgttctgctccacaagttgacgttgactctttttctggtgacttgagaactagacaacaatgta tcgctgaaattactgaaatgattcacgtcgcctctttgttgcatgatgacgtcttagatgatgctgatactagaagag gtattggttccttaaattttgttatgggtaataagttggctgttttggctggtgatttcttgttatccagagcctgcgtcg ccttagcctccttgaagaacaccgaagttgtctgtttattggccaccgttgtcgaacatttggttaccggtgaaacta tgcaaatgactacctcctccgatgaaagatgttccatggaatactacatgcaaaagacctactataagactgcctct ttgatttctaactcttgtaaagccattgccttgttagccggtcactctgctgaagtttctgtcttggccttcgattacggt aagaacttaggtttggcttttcaattgatcgacgatgttttggacttcaccggtacctctgctactttgggtaaaggtt ccttgtccgatatcagacatggtatcgttactgctcctatttgtatgctatggaagaattccctcaattacgtactttg gttgacagaggtttcgatgatccagttaatgttgagatcgctttggattacttgggtaaatcccgtggtattcaaaga actagagaattagccagaaagcatgcctctttagcctctgccgccatcgattccttgcctgaatccgacgatgagg aagttcaaagatctcgtagagctttggtcgaattgacccatagagtcattactcgtactaagtaa
SEQ ID NO:143 Geranylgeranyl pyrophosphate synthase Hb_GPPS Hevea brasiliensis	MQFLRGLSPISRSGLRLFLSRQLYPFPVANSSQLLGDSTQKVFNRRETY SWSLVDSHGFKQQIHHQSSFLSEEPLDPFSLVADELSLVANRLRAMLVS EVPKLASAAEYFFKMGVEGKRLRPTVLLLMATALNVHIHEPMPNGVG DTLGAELRTRQQCIAEITEMIHVASLLHDDVLDDADTRRGIGSLNFVM GNKVAVLAGDFLLSRACVALASLKNTEVVSLLATVVEHLVTGETMQ MTSTSEQRCSMDHYMQKTYYKTASLISDSCKAIALLAGQTTEVAMLA FEYGKSLGLAFQLIDDVLDFTGTSASLGKGSLSDIRHVIRLSLI*
SEQ ID NO:144 Artificial geranylgeranyl pyrophosphate synthase (Hb_GPPS) nucleotide sequence	atgcaatttttgagaggtttgtcccctatttccagatccggtttgcgtttattcttatctcgtcaattatatccattcccag tcgccaactcctcccaattattaggtgactctactcaaaaggtttttaacagacgtgagacttactcttggtctttggt cgactctcacggttttaagcaacaaattcatcaccaatcctctttttttgtctgaagaaccattggatccattctctttgg ttgctgatgaattatccttggtcgctaacagattgcgtgctatgttggtttctgaagtcccaaaattagcctccgccg ctgaatatttttttcaagatgggtgttgaaggtaagagattgcgtccaaccgtcttgttattaatggccactgctttaaa cgttcatatccatgaacctatgcctaacggtgttggtgacactttgggtgccgaattgagaactagacaacaatgc atcgctgaaatcaccgaaatgatccatgttgcttctttattacatgacgacgttttagacgatgccgataccagaag aggtattggttctttgaacttcgttatgggtaacaaggttgctgttttggccggtgacttttttgttgtccagagcttgtgt tgccttagcttctttgaagaatacccgaagtcgtttcttttattggccaccgtcgtcgaacacttggttactggtgagact atgcaaatgacctccacttctgagcaacgttgttccatggatcattatatgcaaaagacttactataagaccgcttcc ttaatctctgattcctgtaaagccatcgccttgttagctggtcaaactaccgaggtcgccatgttggccttcgaatat ggtaagtctttgggtttagcttttcaattaatcgacgatgttttagacttcaccggtacttctgcttccttgggtaaggg ttctttgtccgacattagacacgttattagattatccttaatttaa
SEQ ID NO:145 Mutant medium-chain fatty acid CoA ligase Ec_FADK_v1	MHPTGPHLGPDVLFRESNMKVTLTFNEQRRAAYRQQGLWGDASLAD YWQQTARAMPDKIAVVDNHGASYTYSALDHAASCLANWMLAKGIES GDRIAFQLPGWCEFTVIYLACLKIGAVSVPLLPSWREAELVWVLNKCQ AKMFFAPTLFKQTRPVDLILPLQNQLPQLQQIVGVDKLAPATSSLSLSQI IADNTSLTTAITTHGDELAAVLFTSGTEGLPKGVMLTHNNILASERAYC ARLNLTWQDVFMMPAPLGHATGFLHGVTAPFLIGARSVLLDIFTPDAC LALLEQQRCTCMLGATPFVYDLLNVLEKQPADLSALRFFLCGGTTIPK KVARECQQRGIKLLSVYGSTESSPHAVVNLDDPLSRFMHTDGYAAAG VEIKVVDDARKTLPPGCEGEEASRGPNVFMGYFDEPELTARALDEEG WYYSGDLCRMDEAGYIKITGRKKDIIVRGGENISSREVEDILLQHPKIH DACVVAMSDERLGERSCAYVVLKAPHHSLSLEEVVAFFSRKRVAKYK YPEHIVVIEKLPRTTSGKIQKFLLRKDIMRRLTQDVCEEIE*
SEQ ID NO:146	atgcatccaactggtccacacttaggtcctgatgtcttatttagagaatctaatatgaaagtcactttgacctttaatg

SEQ ID	SEQUENCE
Artificial medium-chain fatty acid CoA ligase Ec_FADK_v1 nucleotide sequence	aacaaagacgtgccgcttacagacaacaaggtttgtggggtgacgcttctttggctgactactggcaacaaactg ctagagctatgccagacaagatcgccgttgtcgataaccacggtgcttcttatacctactctgctttggatcatgcc gcttcttgtttggctaattggatgttggctaagggtatcgaatctggtgatcgtattgcttttcaattgccaggttggtg tgaatttaccgttatctacttggcttgtttgaagattggtgctgtttctgtcccattgttgccatcttggagagaagccg aattggtttgggttttgaacaaatgccaagctaagatgttctttgctccaaccttgttcaagcaaactagaccagttg acttgattttacctttacaaaatcaattaccacaattgcaacaaatcgttggtgttgacaagttagctccagccacctc ctctttgtccttgtcccaaattatcgctgataatacttctttaaccaccgctatcactactcacggtgatgagttggctg ctgttttgttcacttccggtactgagggtttgccaaagggtgttatgttgacccacaataacattttggcttccgaaag agcttattgtgctcgtttgaacttgacctggcaagatgttttcatgatgccagctccattgggtcatgctactggtttct tgcacggtgttactgccccattcttgattggtgctagatctgtcttgttggatatctttaccccagacgcttgcttagct ttattggaacaacaaagatgtacctgtatgttaggtgctactccatttgtttacgatttattgaacgttttggaaaaaca accagctgatttgtctgccttgagattctttttgtgtggtggtactactattccaaagaaagttgctagagaatgccaa caaagaggtatcaagttgttgtccgtctatggttccactgaatcttctcctcatgctgttgtcaatttagatgacccatt gtctagattcatgcacaccgatggttacgccgctgctggtgttgagattaaggttgtcgacgatgctagaaagacc ttacctccaggttgtgaaggtgaagaagcctctagaggtccaaatgtctttatgggttacttcgacgagccagaatt gactgctagagctttagatgaggaaggttggtattactctggtgatttgtgtagaatggatgaagctggttacattaa aatcactggtagaaagaaggacattattgttagaggtggtgaaaatatctcctccagagaagttgaagatattttatt gcaacacccaaagattcatgatgcttgtgttgttgctatgtccgatgagagattaggtgaaagatcttgtgcttacgt tgttttgaaggctccacatcactctttgtctttagaagaagtcgttgctttcttctctagaaagagagtcgccaagtac aagtacccagaacacattgttgttatcgaaaaattgcctagaactacttctggtaaaattcaaaaattcttgttgaga aaggatatcatgagacgtttgacccaagatgtctgtgaagaaattgaataa
SEQ ID NO:147 Medium-chain fatty acid CoA ligase Ec_FADK_v2 Escherichia coli	MHPTGPHLGPDVLFRESNMKVTLTFNEQRRAAYRQQGLWGDASLAD YWQQTARAMPDKIAVVDNHGASYTYSALDHAASCLANWMLAKGIES GDRIAFQLPGWCEFTVIYLACLKIGAVSVPLLPSWREAELVWVLNKCQ AKMFFAPTLFKQTRPVDLILPLQNQLPQLQQIVGVDKLAPATSSLSLSQI IADNTSLTTAITTHGDELAAVLFTSGTEGLPKGVMLTHNNILASERAYC ARLNLTWQDVFMMPAPLGHATGFLHGVTAPFLIGARSVLLDIFTPDAC LALLEQQRCTCMLGATPFVYDLLNVLEKQPADLSALRFFLCGGTTIPK KVARECQQRGIKLLSVYGSTESSPHAVVNLDDPLSRFMHTDGYAAAG VEIKVVDDARKTLPPGCEGEEASRGPNVFMGYFDEPELTARALDEEG WYYSGDLCRMDEAGYIKITGRKKDIIVRGGENISSREVEDILLQHPKIH DACVVAMSDERLGERSCAYVVLKAPHHSLSLEEVVAFFSRKRVAKYK YPEHIVVIEKLPRTTSGKIQKFLLRKDIMRRLTQDVCEEIE*
SEQ ID NO:148 Artificial medium-chain fatty acid CoA ligase Ec_FADK_v2 nucleotide sequence	atgcatccaactggtcctcacttaggtccagatgtcttattcagagaatctaacatgaaagtcactttaacttttaacg aacaacgtagagctgcttatagacaacaaggtttgtggggtgatgcttccttggctgactactggcaacaaactgc tagagccatgccagataaaattgccgttgttgacaatcacggtgcttcttacacttattctgccttagatcacgctgc ttcctgtttagctaactggatgttagctaagggtattgaatccggtgatagaattgctttccaattgccaggttggtgc gaatttactgtcatttatttagcttgtttaaagattggtgccgtctccgtccctttgttgccatcctggagagaggccg agttggtttgggttttaaacaagtgtcaagctaaaatgttctttgctcctaccttgttcaagcaaaccagaccagttga cttaattttgccattacaaaaccaattaccacaattgcaacaaatcgtcggtgttgataaattagctccagccacttct tctttgtccttatcccaaattattgctgataacacttctttaactactgctattactactcacggtgatgaattggccgct gttttgttcacttccggtactgaaggtttgcctaaaggtgtcatgttgactcacaacaacattttggcctctgaaaga gcttactgtgcccgtttaaatttgacctggcaagatgtcttcatgatgcctgctccattgggtcacgctaccggtttct tacacggtgtcactgccccattcttgatcggtgctcgttctgttttattggatatctttactccagatgcttgcttggcttt attggaacaacaaagatgtacctgcatgttaggtgctactcctttcgtctatgatttattgaacgtcttagaaaaacaa ccagctgatttatccgctttaagattctttttgtgtggtggtactactatcccaaaaaaggtcgccagagaatgtcaa caaagaggtattaaattattgtccgtttatggttccactgaatcttcccctcatgctgttgtcaatttagacgacccttt gtccagattcatgcacactgatggttacgccgctgctggtgtcgaaatcaaggttgttgatgacgctagaaaaact ttaccacctggttgcgaaggtgaagaggcttccagaggtccaaacgtctttatgggttactttgatgaaccagaatt gactgccagagctttggatgaggaaggttggtattattctggtgatttgtgtagaatggatgaagccggttacatca agatcaccggtagaaagaaagacatcatcgttagaggtggtgaaaacatttcttctagagaagttgaagacatttt gttgcaacacccaaagatccacgacgcttgtgtcgtcgccatgtctgacgaaagattgggtgaacgttcttgtgct tacgtcgtcttgaaagccccacaccactctttgtctttggaagaagtcgttgctttttctctcgtaagcgtgttgcca

SEQ ID	SEQUENCE
	agtacaagtacccagagcacatcgttgttattgaaaaattgcctcgtactacttccggtaagattcaaaagttcttatt acgtaaggacatcatgagaagattgactcaagacgtctgcgaagaaattgaataa
SEQ ID NO:149 Truncated acyl activating enzyme (Cs_AAE3_Ctrunc)	MEKSGYGRDGIYRSLRPPLHLPNNNNLSMVSFLFRNSSSYPQKPALIDS ETNQILSFSHFKSTVIKVSHGFLNLGIKKNDWLIYAPNSIHFPVCFLGIIA SGAIATTSNPLYTVSELSKQVKDSNPKLIITVPQLLEKVKGFNLPTILIGP DSEQESSSDKVMTFNDLVNLGGSSGSEFPIVDDFKQSDTAALLYSSGTT GMSKGWLTHKNFIASSLMVTMEQDLVGEMDNVFLCFLPMFHVFGLAI ITYAQLQRGNTVISARFDLEKMLKDVEKYVTHLWWPPVILALSKNSM VKFNLSSIKYIGSGAAPLGKDLMEECSKWPYGIVAQGYGMTETCGIVS MEDIRGGKRNSGSAGMLASGVEAQIVSVDTLKPLPPNQLGEIWVKGPN MMQGYFNNPQATKLTIDKKGWVHTGDLGYFDEDGHLYWDRIKELIK YKGFQVAPAELEGLLVSHPEILDAWIPFPDAEAGEVPVAYWRSPNSSL TENDVKKFIAGQVASFKRLRKVTFINSVPKSASGKIL
SEQ ID NO:150 Truncated acyl activating enzyme (Cs_AAE3_Ctrunc)	atggaaaaatctggttatggtagagacggtatctacagatccttgcgtcctccattacacttgccaaacaataataa cttatctatggtttccttttttgttccgtaactcttcctcttacccacaaaaacctgctttgattgactccgaaaccaatca aatcttgtccttttcccacttcaaatctactgtcattaaagtctctcacggtttcttgaacttaggtattaagaagaacg actggttgatctacgctcctaattccatccactttccagtttgtttcttgggtatcattgcttctggtgccattgctacca cttctaaccctttatacactgtttctgagttatctaagcaagttaaagattctaacccaaaattgattatcactgtccca caattattagaaaaggtcaagggtttcaatttaccaaccattttaatcggtccagactccgaacaagagtcttcttcc gataaagttatgacttttaacgacttagttaacttgggtggttcttctggttctgagttcccaatcgtcgatgatttcaa gcaatctgacaccgccgctttattgtattcctctggtactactggtatgtctaagggttggttgactcacaaaaacttt atcgcttcctctttgatggttaccatggaacaagacttggttggtgaaatggataacgtcttcttgtgtttttaccaat gttccatgttttcggtttagctatcattacttacgctcaattacaaagaggtaacactgtcatctctgctcgttttgactt agaaaagatgttgaaagacgttgaaaagtacgttactcacttgtggtggcctcctgttatttttagctttgtctaagaat tctatggttaaattcaacttgtcctctatcaagtacattggttctggtgccgctccattaggtaaggacttgatggaag aatgttctaaatggccttacggtatcgtcgctcaaggttacggtatgactgaaacttgtggtatcgtttctatggaag acatcagaggtggtaagcgtaactccggttctgctggtatgttggcttccggtgttgaagcccaaattgtttctgtc gatactttgaaaccttgccacctaaccaattaggtgaaatttgggttaaaggtcctaacatgatgcaaggttacttc aataaccctcaagctactaagttaactattgataagaagggttgggttcatactggtgatttgggttacttcgatgaa gatggtcatttgtactgggatagaatcaaagaattaattaagtataaaggtttccaagttgccccagctgaattgga aggtttgttggtttctcatcctgaaattttagatgcttggattccttttcccagacgctgaagccggtgaagttccagtt gcttactggagatcccctaactcttccttgactgaaaacgacgtcaagaagttcatcgctggtcaagttgcttccttt aagagattaagaaaagtcaccttcatcaactccgttccaaagtctgcttccggtaagattttg
SEQ ID NO:151 Truncated cannabidiolic acid synthase Cs_CBDASt28	MSNPRENFLKCFSQYIPNNATNLKLVYTQNNPLYMSVLNSTIHNLRFTS DTTPKPLVIVTPSHVSHIQGTILCSKKVGLQIRTRSGGHDSEGMSYISQV PFVIVDLRNMRSIKIDVHSQTAWVEAGATLGEVYYWVNEKNENLSLA AGYCPTVCAGGHFGGGGYGPLMRNYGLAADNIIDAHLVNVHGKVLD RKSMGEDLFWALRGGGAESFGIIVAWKIRLVAVPKSTMFSVKKIMEIH ELVKLVNKWQNIAYKYDKDLLLMTHFITRNITDNQGKNKTAIHTYFSS VFLGGVDSLVDLMNKSFPELGIKKTDCRQLSWIDTIIFYSGVVNYDTD NFNKEILLDRSAGQNGAFKIKLDYVKKPIPESVFVQILEKLYEEDIGAG MYALYPYGGIMDEISESAIPFPHRAGILYELWYICSWEKQEDNEKHLN WIRNIYNFMTPYVSKNPRLAYLNYRDLDIGINDPKNPNNYTQARIWGE KYFGKNFDRLVKVKTLVDPNNFFRNEQSIPPLPRHRH*
SEQ ID NO:152 Truncated cannabidiolic acid synthase Cs_CBDASt28	atgtctaatccaagagagaatttcttaaagtgtttttctcaatacatcccaaacaatgctactaacttaaagttggttta cactcaaaataacccattgtacatgtctgtcttgaactctaccattcacaatttgcgttttacttctgacaccacccccta agccattagttattgttaccccatcccacgtctctcacatccaaggtactattttgtgttctaaaaaggttggtttgcaa attagaactagatctggtggtcacgactccgagggtatgtcttacatctctcaagttccattcgttattgtcgacttgc gtaacatgcgttccatcaaaatcgatgttcactcccaaactgcttgggtcgaagccggtgccactttaggtgaggt ttattactgggtcaatgagaagaatgagaatttgtccttggctgctggttattgtccaaccgtctgtgctggtggtcat tttggtggtggtggttacggtccattaatgagaaactatggtttggctgccgataacattatcgacgctcacttggtt aatgtccacggtaaggtcttagatagaaaatccatgggtgaggacttgttctgggctttgagaggtggtggtgctg agtcctttggtatcatcgttgcttggaaaattcgtttagttgctgtcccaaaatctactatgttttctgttaagaagatca

SEQ ID	SEQUENCE
	tggaaattcacgagttggttaagttggttaataagtggcaaaatattgcctacaagtatgacaaagacttgttattga tgactcacttcatcactagaaacatcaccgataaccaaggtaaaaataaaactgctatccatacctacttctcctcc gttttcttgggtggtgtcgactccttagttgatttgatgaacaaatcttttcctgaattaggtatcaagaagactgattgt cgtcaattgtcctggattgataccattatcttttactctggtgtcgtcaattacgacaccgataatttcaataaggaaat tttattggacagatctgccggtcaaaacggtgctttcaagatcaagttggactacgttaaaaaaccaatcccagaat ccgtctttgtccaaattttggagaagttatacgaggaagacatcggtgctggtatgtatgccttatatccatacggtg gtattatggatgaaatttccgaatctgctatcccatttccacatcgtgctggtattttgtatgaattatggtacatttgttc ctgggaaaagcaagaagataacgagaagcacttgaattggatcagaaatatctacaatttcatgactccttacgttt ctaagaatcctcgtttggcttacttgaactacagagatttggacatcggtattaatgacccaaagaacccaaataac tatactcaagctagaatttggggtgaaaagtacttcggtaaaaactttgacagattggttaaggttaagactttagtt gatccaaataacttcttcagaaatgaacaatccatcccaccattgcctagacacagacactaa
SEQ ID NO:153 Truncated tetrahydro cannabinolic acid synthase Cs_THCASt28	MSNPRENFLKCFSKHIPNNVANPKLVYTQHDQLYMSILNSTIQNLRFIS DTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTRSGGHDAEGMSYISQV PFVVVDLRNMHSIKIDVHSQTAWVEAGATLGEVYYWINEKNENLSFP GGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLD RKSMGEDLFWAIRGGGGENFGIIAAWKIKLVAVPSKSTIFSVKKNMEIH GLVKLFNKWQNIAYKYDKDLVLMTHFITKNITDNHGKNKTTVHGYFS SIFHGGVDSLVDLMNKSFPELGIKKTDCKEFSWIDTTIFYSGVVNFNTA NFKKEILLDRSAGKKTAFSIKLDYVKKPIPETAMVKILEKLYEEDVGAG MYVLYPYGGIMEEISESAIPFPHRAGIMYELWYTASWEKQEDNEKHIN WVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNHASPNNYTQARIWG EKYFGKNFNRLVKVKTKVDPNNFFRNEQSIPPLPPHHH*
SEQ ID NO:154 Truncated tetrahydrocannabinolic acid synthase Cs_THCASt28	atgtctaaccctcgtgagaacttcttgaaatgtttctccaaacatatcccaaacaatgtcgctaaccctaagttagttt acactcaacatgatcaattatatatgtctatcttgaactctaccatccaaaacttgagattcatctccgataccacccc aaaaccattggttattgttaccccatccaacaattctcatattcaagctaccattttgtgctccaaaaaggtcggtttg caaatccgtactagatctggtggtcacgatgctgaaggtatgtcttacatttcccaagtcccattcgttgttgtcgatt taagaaatatgcactctatcaaaatcgacgttcactctcaaactgcttgggttgaagccggtgccactttaggtgag gtttactactggattaacgaaaagaatgaaaacttatcctttccaggtggttactgtccaactgttggtgttggtggtc acttctctggtggtggttatggtgccttgatgagaaactacggtttagctgctgataatattatcgacgctcacttggt taatgtcgacggtaaggttttggacagaaaatccatgggtgaagatttattctgggccattagaggtggtggtggt gaaaacttcggtatcattgctgcttggaaaattaaattggtcgctgtcccatccaagtctactattttctccgtcaaga aaaacatggaaattcatggtttggttaaattattcaacaagtggcaaaacattgcttacaaatacgacaaagactta gttttgatgacccacttcattactaaaaacattaccgacaaccatggtaaaaataaaactactgttcacggttacttct cttccatttttcatggtggtgtcgactccttggtcgatttaatgaacaaatctttccctgagttgggtatcaagaagac cgactgtaaagaattctcttggatcgacactactattttctactctggtgtcgttaacttcaacaccgctaatttcaaga aggaaatttttattagatagatccgctggtaaaaagaccgctttctctatcaaattagactacgttaaaaaaccaatcc cagaaaccgctatggtcaaaatcttggaaaaattatatgaagaagacgttggtgccggtatgtacgtcttatatcca tatggtggtattatggaagagatctctgaatccgctatccctttccacacagagccggtattatgtacgaattatgg tacactgcttcctgggagaaacaagaagataatgaaaagcacattaactgggttagatctgtttacaacttcactac tccatacgtctctcaaaacccaagattagcctacttaaactaccgtgatttggatttaggtaaaactaatcacgcttc cccaaacaactacacccaagctagaatttggggtgagaagtactttggtaagaacttcaaccgtttagtcaaggtc aagactaaagttgatccaaacaatttttttcagaaacgaacaatctatcccaccttttaccaccacaccaccattag
SEQ ID NO:155 GenBank AB057805.1 Tetrahydro cannabinolic acid synthase (THCAS, Cs_THCAS_full) *Cannabis sativa*	MNCSAFSFWFVCKIIFFFLSFHIQISIANPRENFLKCFSKHIPNNVANPKL VYTQHDQLYMSILNSTIQNLRFISDTTPKPLVIVTPSNNSHIQATILCSKK VGLQIRTRSGGHDAEGMSYISQVPFVVVDLRNMHSIKIDVHSQTAWVE AGATLGEVYYWINEKNENLSFPGGYCPTVGVGGHFSGGGYGALMRN YGLAADNIIDAHLVNVDGKVLDRKSMGEDLFWAIRGGGGENFGIIAA WKIKLVAVPSKSTIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLVLM THFITKNITDNHGKNKTTVHGYFSSIFHGGVDSLVDLMNKSFPELGIKK TDCKEFSWIDTTIFYSGVVNFNTANFKKEILLDRSAGKKTAFSIKLDYV KKPIPETAMVKILEKLYEEDVGAGMYVLYPYGGIMEEISESAIPFPHRA GIMYELWYTASWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLN YRDLDLGKTNHASPNNYTQARIWGEKYFGKNFNRLVKVKTKVDPNN

SEQ ID	SEQUENCE
	FFRNEQSIPPLPPHHH*
SEQ ID NO:156 Artificial Tetrahydrocannabinolic acid synthase Cs_THCAS_full nucleotide sequence	atgaattgttctgctttctctttctggttcgtttgtaagatcatctttttcttcttatctttccatattcaaatctctatcgctaa ccctcgtgagaacttcttgaaatgtttctccaaacatatcccaaacaatgtcgctaaccctaagttagtttacactca acatgatcaattatatatgtctatcttgaactctaccatccaaaacttgagattcatctccgataccaccccaaaacc attggttattgttaccccatccaacaattctcatattcaagctaccattttgtgctccaaaaaggtcggtttgcaaatcc gtactagatctggtggtcacgatgctgaaggtatgtcttacatttcccaagtcccattcgttgttgtcgatttaagaaa tatgcactctatcaaaatcgacgttcactctcaaactgcttgggttgaagccggtgccactttaggtgaggtttacta ctggattaacgaaaagaatgaaaacttatcctttccaggtggttactgtccaactgttggtgttggtggtcacttctct ggtggtggttatggtgccttgatgagaaactacggtttagctgctgataatattatcgacgctcacttggttaatgtc gacggtaaggttttggacagaaaatccatgggtgaagatttattctgggccattagaggtggtggtggtgaaaact tcggtatcattgctgcttggaaaattaaattggtcgctgtcccatccaagtctactattttctccgtcaagaaaaacat ggaaattcatggtttggttaaattattcaacaagtggcaaaacattgcttacaaatacgacaaagacttagttttgat gacccacttcattactaaaaacattaccgacaaccatggtaaaaataaaactactgttcacggttacttctcttccatt tttcatggtggtgtcgactccttggtcgatttaatgaacaaatctttccctgagttgggtatcaagaagaccgactgt aaagaattctcttggatcgacactactattttctactctggtgtcgttaacttcaacaccgctaatttcaagaaggaaa ttttattagatagatccgctggtaaaaagaccgctttctctatcaaattagactacgttaaaaaaccaatcccagaaa ccgctatggtcaaaatcttggaaaaattatatgaagaagacgttggtgccggtatgtacgtcttatatccatatggtg gtattatggaagagatctctgaatccgctatccctttccacacagagccggtattatgtacgaattatggtacactg cttcctgggagaaacaagaagataatgaaaagcacattaactgggttagatctgtttacaacttcactactccatac gtctctcaaaacccaagattagcctacttaaactaccgtgatttggatttaggtaaaactaatcacgcttccccaaac aactacacccaagctagaatttggggtgagaagtactttggtaagaacttcaaccgtttagtcaaggtcaagacta aagttgatccaaacaatttttttcagaaacgaacaatctatcccaccttttaccaccacaccaccattag
SEQ ID NO:157 Artificial Erg10p: acetoacetyl CoA thiolase nucleotide sequence	atgtcccaaaatgtttacattgtttctactgctagaactcctatcggttccttccaaggttccttatcttccaaaactgcc gtcgaattgggtgccgttgccttgaaaggtgctttagctaaagttccagagttagacgcttccaaagatttcgatga aattatcttcggtaacgttttatccgctaacttgggtcaagctccagccagacaagttgccttggctgccggtttgtc taatcacatcgttgcttctactgtcaacaaagtttgtgcctctgctatgaaagctatcattttaggtgcccaatctatta aatgtggtaatgctgacgttgttgtcgctggtggttgtgagtccatgaccaacgccccttactacatgccagccgc cagagccggtgccaaattcggtcaaactgttttggttgacggtgttgaaagagatggtttgaacgatgcctatgac ggtttggctatgggtgttcacgctgaaaagtgtgctagagactgggacattaccagagaacaacaagataatttc gctattgaatcttaccaaaagtcccaaaaatctcaaaaggaaggtaagtttgacaatgaaatcgttccagttactat caagggttttcgtggtaagcctgatactcaagtcaccaaggatgaagaaccagcccgtttacacgtcgaaaagtt gagatctgccagaaccgttttccaaaaagaaaacggtaccgttactgctgccaatgcttctccaatcaacgatggt gccgctgctgttatttttagtctctgagaaggttttgaaggagaaaaatttgaagcctttagccatcattaagggttgg ggtgaagctgctcaccaaccagctgatttcacttgggccccttctttagctgtcccaaaggctttaaaacacgctg gtattgaagatatcaactctgttgactacttcgaattcaatgaagctttctctgtcgtcggtttggtcaatacaaaatc ttgaagttggatccttctaaggttaacgtttacggtggtgctgtcgccttaggtcacccttttaggttgttctggtgcta gagttgttgtcaccttgttgtccattttacaacaagaaggtggtaagatcggtgttgctgctatctgtaacggtggtg gtggtgcttcttccattgtcatcgaaaagatctag
SEQ ID NO:158 Artificial mevalonate pyrophosphate decarboxylase (Sc_ERG19) nucleotide sequence	atgactgtctacactgcctccgttactgcccctgtcaacattgccaccttgaagtattggggtaaaagagatactaa attgaacttaccaactaactcctccatttctgtcactttgtctcaagatgatttgagaaccttgacttccgctgccacc gcccctgaatttgagagagatactttgtggttaaatggtgaacctcattctattgacaacgaaagaacccaaaact gtttacgtgacttgagacaattgcgtaaggaaatggaatctaaagacgcttctttacctaccttgtctcaatggaaat tgcatatcgtttctgaaaataacttccctactgctgccggtttggcttcctccgctgctggttttgctgctttagtttctg ccatcgccaaattatatcaattgccacaatccacttccgaaatctctagaatcgctagaaaaggttccggttctgctt gtagatccttgttcggtggttacgttgcttgggaaatgggtaaagctgaagacggtcatgattctatggccgttcaa attgccgactcctccgattggcctcaaatgaaagcttgtgtcttggttgtctccgatatcaaaaaggatgtctcttcta ctcaaggtatgcaattaactgttgccacttccgaattgttcaaagagcgtatcgaacacgttgttccaaagagatttg aagttatgagaaaagctatcgtcgaaaaggacttcgctacctttgccaaggagactatgatggattctaactccttc cacgctacttgtttggattcctttccacctattttctacatgaatgacacctccaaacgtattatctcttggtgtcacacc attaaccaattttatggtgaaactatcgtcgcttacactttcgatgccggtccaaacgctgtcttgtactatttggctga aaacgaatccaagttatttgcttttatctataagttgttcggttccgtccctggttgggacaagaaattcaccactgaa caattggaagctttcaaccaccaattcgaatcttccaatttcactgctagagaattagatttggaattacaaaaggat gtcgctagagtcatcttaactcaagttggttccggtccacaagaaactaacgaatctttgattgatgctaaaactggt

SEQ ID	SEQUENCE
	ttgcctaaagaataa
SEQ ID NO:159 Artificial isopentenyl pyrophosphate isomerase Sc_IDI1 nucleotide sequence	atgaccgctgacaacaactccatgccacatggtgctgtctcctcctacgctaaattagtccaaaaccaaaccccctg aagacattttagaagagttccctgaaatcattccattgcaacaaagaccaaacactagatcctccgagacttctaa cgatgaatctggtgaaacttgtttttctggtcatgatgaagaacaaatcaagttgatgaacgagaattgtattgttttg gactgggatgacaacgctatcggtgctggtaccaaaaaggtctgtcacttgatggaaaacatcgaaaagggtttg ttgcatagagccttttccgtcttcatcttcaacgaacaaggtgagttattattgcaacaaagagccactgaaaaaatc acctttccagatttatggaccaacacctgttgctcccatccattgtgtattgatgatgaattgggtttgaaaggtaagt tggacgacaagattaaaggtgccatcaccgccgctgttcgtaagttagaccatgaattgggtatccctgaagacg aaactaagactagaggtaaattccatttcttgaatcgtattcactacatggctccttccaatgaaccatggggtgaa cacgaaatcgactacattttgttttacaaaattaatgctaaagaaaatttaaccgttaacccaaacgtcaacgaggtt agagatttcaagtgggtctctccaaacgatttgaagactatgttcgctgacccatcctacaagttcactccatggttt aagatcatctgtgaaaactatttgtttaactggtgggagcaattggacgacttatctgaagttgaaaatgatcgtcaa attcaccgtatgttgtaa
SEQ ID NO:160 Artificial phosphomevalonate kinase Sc_ERG8 nucleotide sequence	atgtccgagttaagagccttctccgctcctggtaaagccttattagctggtggttacttagtcttggatactaaatatg aagccttcgtcgtcggtttatctgccagaatgcatgccgtcgcccatccatacggttccttgcaaggttctgacaag tttgaggtccgtgtcaagtctaaacaattcaaagatggtgaatggttgtatcatatttctccaaaatccggtttcattcc agtttctatcggtggttctaagaacccattcatcgaaaaagtcatcgctaacgttttctcttacttcaagcctaatatg gatgattattgcaatagaaatttattcgttattgatatcttctccgatgacgcctatcattcccaagaagactctgttac cgagcatagaggtaacagaagattatctttccactctcacagaattgaagaagttccaaaaactggtttaggttctt ctgctggtttagtcaccgttttaaccactgccttggcttctttctttgtttccgacttagaaaataacgtcgacaagtat cgtgaagtcatccacaacttggcccaagttgctcattgtcaagctcaaggtaagattggttccggtttcgatgttgct gccgccgcctacggttccatcagatatagaagattccctccagctttgatttctaacttaccagatattggttctgcta cttatggttccaagttggctcacttggttgacgaagaagattggaacattaccatcaagtccaatcacttgccatctg gtttaactttgtggatgggtgatatcaagaacggttctgaaactgtcaaattggtccaaaaggtcaaaaattggtac gattcccatatgccagagtctttgaagatctatactgaattggaccacgctaactctcgtttcatggatggtttgtcta agttggacagattgcatgaaactcacgacgactactctgaccaaattttcgagtccttggaaagaaacgactgca cttgtcaaaagtatccagaaatcaccgaggttagagatgccgttgctactattagaagatccttcagaaagattacc aaggaatctggtgctgatattgagcctccagttcaaacttctttgttggatgattgccaaactttaaaaggtgttttaa cttgtttaattcctggtgctggtggttacgacgccatcgccgttatcaccaaacaagacgtcgacttaagagccca aactgccaacgacaaaagattctccaaggttcaatggttggacgtcactcaagctgattggggtgttagaaaaga aaaggacccagagacttacttggataaatag
SEQ ID NO:161 Mutant farnesyl pyrophosphate synthase (Erg20mut, F96W, N127W)	atggcttctgagaaggagattcgtcgtgagagattcttgaatgtttttcctaaattagtcgaggaattgaacgcttctt tgttggcttatggtatgcctaaggaagcttgtgattggtatgctcactccttgaattataatactccaggtggtaaatt gaaccgtggtttgtctgttgttgacacttacgctattttatctaacaagaccgtcgagcaattgggtcaagaagagta tgaaaaggtcgctattttaggttggtgtattgaattgttgcaagcttactggttggttgccgatgacatgatggacaa gtctattactcgtcgtggtcaaccttgctggtataaggtcccagaggttggtgaaattgctatctgggacgctttcat gttggaagctgctatctataaattgttgaaatcccacttcagaaacgagaaatactacattgacatcaccgagttgtt ccacgaagtcactttccaaactgagttaggtcaattaatggacttgatcaccgctccagaagacaaagttgacttgt ccaagttttccttgaaaaagcactctttcatcgttactttcaagactgcttattactctttctacttaccagttgccttggc tatgtacgtcgccggtatcactgacgaaaaggacttgaagcaagctcgtgacgttttgattccattaggtgaatattt ccaaatccaagatgactacttagactgttttggtaccctgaacaaatcggtaagatcggtactgatattcaagata acaagtgctcttgggttatcaacaaggctttagagttagcctccgccgaacaacgtaaaactttagatgaaaacta cggtaaaaaagactctgttgctgaggccaagtgtaagaagatttttaacgatttaaaaatcgaacaattgtatcacg aatatgaagagtccattgctaaggatttgaaggctaaaatttctcaagttgacgaatcccgtggtttcaaagctgac gttttgactgctttttttaaacaaggtttacaagcgttccaaataa
SEQ ID NO:162 Artificial tetraketide synthase (TKS) nucleotide sequence	atgaaccatttaagagctgagggtccagcttccgtcttggctatcggtactgctaatccagagaacattttattacaa gatgagtttccagattactatttccgtgttactaagtccgagcatatgacccaattgaaagaaaagttccgtaaaatc tgtgataaatctatgattagaaaaagaaactgctttttaaacgaagaacacttgaagcaaaacccaagattagttga acacgagatgcaaaccttggacgctagacaagatatgttggttgtcgaggttcctaaattgggtaaagacgcctg tgctaaagctatcaaagagtggggtcaacctaagtccaagatcactcacttaatcttcacttccgcttccaccactg acatgcctggtgctgattaccactgtgccaagttgttgggtttgtctccttctgtcaagagagttatgatgtaccaatt aggttgttacggtggtggtactgtcttaagaattgctaaggacatcgctgaaaacaacaaaggtgctagagtttta gccgtttgttgtgacatcatggcttgtttatttcgtggtccatctgaatctgacttggagttgttggttggtcaagctatt

SEQ ID	SEQUENCE
	tttggtgatggtgccgctgccgtcatcgttggtgctgagccagatgaatccgttggtgaaagaccaattttcgaatt agtctctactggtcaaactattttgccaaactccgagggtactatcggtggtcatattcgtgaagccggtttaatcttt gatttgcacaaagacgttccaatgttgatctctaacaacatcgaaaagtgtttaattgaggcttttactccaattggta tctctgactggaactctatcttctggatcactcatccaggtggtaaggctatcttggacaaggttgaagaaaaatta catttaaagtccgataaattcgtcgattctcgtcatgttttgtctgaacacggtaacatgtcttcctccactgtcttgttt gttatggatgaattacgtaagagatctttggaggagggtaagtctactactggtgatggtttcgaatggggtgttttg ttcggtttcggtcctggtttgactgttgaacgtgttgttgttagatctgttccaattaagtactag
SEQ ID NO:163 Artificial olivetolic acid cyclase (OAC) nucleotide sequence	atggccgtcaaacacttgatcgtcttaaaattcaaggatgaaattactgaagctcaaaaagaagagttcttcaaaa cctatgtcaatttagtcaacattattcctgctatgaaggacgtttactggggtaaggatgtcacccaaaagaacaag gaagaaggttacactcacattgttgaagtcactttcgaatctgttgaaactatccaagattatattatccacccagct catgtcggttttggtgatgtttacagatctttttgggaaaaattgttgatctttgactatactccaagaaaataa
SEQ ID NO:164 Artificial acyl-activating enzyme Cs_AAE1_v1 nucleotide sequence	atgggtaagaattacaagtccttagactctgttgttgcttctgactttattgctttaggtattacttccgaagttgctgaa accttacacggtagattggctgaaattgtttgcaactacggtgctgctacccctcaaacttggattaacattgctaat catattttgtctccagatttgccattttctttacaccaaatgttgttctacggttgttacaaggatttcggtcctgctcctc cagcttggattcctgatccagaaaaagtcaaatctactaacttgggtgctttgttggaaaagagaggtaaggagttt ttgggtgttaagtacaaggacccaatttcttctttctctcacttccaagaattctctgttagaaaccctgaagtttactg gagaactgttttgatggatgagatgaagatttctttttctaaggacccagagtgtatcttaagaagagacgacattaa caatccaggtggttctgagtggttaccaggtggttacttgaactctgccaaaaattgcttgaacgttaactctaaca agaaattgaatgacactatgattgtctggagagatgagggtaacgatgatttgcctttgaataaattgactttggatc aattgagaaaaagagtctggttggttggttacgctttggaagaaatgggtttagaaaaaggttgtgctatcgccatc gatatgcctatgcacgttgatgctgttgttatttatttggctattgttttagctggttatgttgttgtttccatcgccgactc cttctctgctccagaaatctccaccagattgagattgtctaaagccaaagccattttcacccaagaccacatcatta gaggtaagaagcgtattccattgtattctcgtgttgttgaagctaaatctcctatggctatcgtcatcccatgctctgg ttctaacatcggtgctgaattaagagacggtgatatttcttgggactacttttagaaagagctaaagaattcaaaaa ctgcgagtttactgctagagaacaacctgtcgacgcttatactaatattttattctcttctggtactactggtgaaccta aggctattccatggacccaagctactcctttgaaagccgctgctgatggttggtcccatttagacatcagaaaagg tgatgtcatcgtctggccaactaacttaggttggatgatgggtccatggttagtctacgcttctttgttgaatggtgcc tctatcgccttatataatggttccccttagtctctggttttgctaaattcgttcaagatgctaaggttaccatgttaggt gttgtcccttctatcgttagatcttggaaatctactaactgtgtttctggttacgactggtccactattcgttgtttctcttc ttctggtgaagcttccaatgtcgatgagtacttatggttaatgggtcgtgctaactacaagccagtcatcgaaatgt gcggtggtactgaaattggtggtgctttttccgctggttctttttacaagcccaatccttgtcttccttctcctctcaat gtatgggttgtactttatatatcttagataagaatggttaccctatgcctaaaaacaagccaggtattggtgaattagc tttgggtcctgttatgtttggtgcttctaaaaccttgttaaatggtaatcatcacgacgtttacttcaaaggtatgcctac tttgaacggtgaggttttgagacgtcatggtgatattttcgaattaacttccaacggttattatcacgctcacggtaga gctgatgatactatgaacattggtggtattaagatctcttccatcgaaattgagagagtttgtaacgaggttgacgat cgtgttttcgaaactactgctattggtgtccctccttaggtggtggtccagaacaattggttatctttttcgtcttgaag gactccaacgacaccactatcgacttaaaccaattaagattgtctttcaacttgggtttgcaaaagaagttgaatcc attatttaaggttactcgtgtcgttccattgtcctccttgccaagaactgctaccaacaagattatgcgtagagtcttg agacaacaattctctcactttgagtaa
SEQ ID NO:165 Artificial acyl-activating enzyme Cs_AAE1_v2 nucleotide sequence	atgggtaagaactacaaatccttagattccgtcgtcgcttctgatttcatcgctttgggtattacttctgaagttgctga aaccttgcatggtagattggctgaaattgtctgtaactacggtgctgctaccccacaaacttggatcaacattgcta accacatcttatcccctgacttgccattctccttacaccaaatgttgttctacggttgttataaagatttcggtccagct cctcctgcttggattcctgacccagagaaggttaagtctactaatttaggtgctttgttagagaagagaggtaagga atttttaggtgttaagtataaagatccaatttcttccttctctcacttccaagaattttctgttagaaacccagaagtttac tggagaactgttttgatggatgaaatgaagatctcttttccaaggacccagagtgtattttgagacgtgatgacatc aacaatccaggtggttctgagtggttaccaggtggttacttgaactctgccaagaattgtttgaacgttaactctaac aaaaagttgaacgataccatgattgtttggagagacgaaggtaacgatgatttgccattgaataagttaaccttgg atcaattgagaaaaagagtctggttagtcggttacgctttggaagagatgggtttggaaaagggttgtgctatcgc catcgatatgccaatgcatgttgatgctgttgttatctatttggccattgttttggctggttacgttgttgtttccatcgct gactccttctctgctccagaaatttctactagattaagattgtctaaagccaaagccattttcactcaagaccatattat tagaggtaagaaaagaattccattgtattccagagttgttgaagctaaatccccaatggccatcgtcatcccatgct ctggttctaatattggtgccgaattgagagacggtgatatctcttgggactacttttggagcgtgctaaagaattta aaaactgcgaattcaccgccagagaacaaccagttgacgcctacactaacattttgttttcttctggtactactggt

SEQ ID	SEQUENCE
	gaacctaaggctattccatggactcaagctactccattgaaagccgccgccgatggttggtcccacttagatatta gaaagggtgatgtcatcgtctggcctactaacttgggttggatgatgggtccttggttggtttacgcttccttattga acggtgcctctatcgctttatataatggttccccttagtttctggttttgctaaattcgttcaagatgctaaggttactat gttgggtgtcgtcccatccattgtccgttcctggaagtctaccaattgtgtttctggttatgattggtctactattcgttg ttttttcttcctctggtgaagcttctaatgtcgatgaatatttgtggttaatgggtagagctaactacaagccagttattg aaatgtggtggtactgaaattggtggtgctttctctgctggttccttttgcaagctcaatccttgtcttctttctcctc ccaatgtatgggttgcactttatacatcttggacaagaatggttaccctatgccaaagaataaaccaggtattggtg aattggctttgggtccagtcatgttcggtgcttctaagactttgttgaacggtaaccatcatgacgtctacttcaagg gtatgcctaccttgaacggtgaagttttaagacgtcacggtgacattttcgaattgacttccaacggttattatcatgc tcacggtagagctgacgacactatgaacatcggtggtattaagatctcttctatcgaaattgaaagagtttgcaacg aggttgatgatcgtgtcttcgaaaccactgctattggtgtccctcctttaggtggtggtcctgagcaattggttatttttc tttgtcttaaaggattctaacgacaccactattgacttaaatcaattgagattgtccttcaatttgggtttgcaaaagaa gttgaacccattattcaaggttactcgtgtcgttcctttgtcctctttgccaagaaccgctaccaataaaattatgaga cgtgttttgcgtcaacaattctctcactttgaataa
SEQ ID NO:166 Artificial acyl-activating enzyme Cs_AAE3 nucleotide sequence	atggaaaaatctggttatggtagagacggtatctacagatccttgcgtcctccattacacttgccaaacaataataa cttatctatggtttccttttttgttccgtaactcttcctcttacccacaaaaacctgctttgattgactccgaaaccaatca aatcttgtccttttcccacttcaaatctactgtcattaaagtctctcacggtttcttgaacttaggtattaagaagaacg actggttgatctacgctcctaattccatccacttccagtttgtttcttgggtatcattgcttctggtgccattgctacca cttctaaccctttatacactgtttctgagttatctaagcaagttaaagattctaacccaaaattgattatcactgtccca caattattagaaaaggtcaagggtttcaatttaccaaccattttaatcggtccagactccgaacaagagtcttcttcc gataaagttatgacttttaacgacttagttaacttgggtggttcttctggttctgagttcccaatcgtcgatgatttcaa gcaatctgacaccgccgctttattgtattcctctggtactactggtatgtctaagggttggttgactcacaaaaacttt atcgcttcctctttgatggttaccatggaacaagacttggttggtgaaatggataacgtcttcttgtgtttttttaccaat gttccatgttttcggtttagctatcattacttacgctcaattacaaagaggtaacactgtcatctctgctcgttttgactt agaaaagatgttgaaagacgttgaaaagtacgttactcacttgtggtggcctcctgttatttttagctttgtctaagaat tctatggttaaattcaacttgtcctctatcaagtacattggttctggtgccgctccattaggtaaggacttgatggaag aatgttctaaatggccttacggtatcgtcgctcaaggttacggtatgactgaaacttgtggtatcgtttctatggaag acatcagaggtggtaagcgtaactccggttctgctggtatgttggcttccggtgttgaagcccaaattgtttctgtc gatactttgaaaccttttgccacctaaccaattaggtgaaatttgggttaaaggtcctaacatgatgcaaggttacttc aataaccctcaagctactaagttaactattgataagaagggttgggttcatactggtgatttgggttacttcgatgaa gatggtcatttgtactgggatagaatcaaagaattaattaagtataaaggtttccaagttgccccagctgaattgga aggtttgttggtttctcatcctgaaattttagatgcttggattcctttcccagacgctgaagccggtgaagttccagtt gcttactggagatcccctaactcttccttgactgaaaacgacgtcaagaagttcatcgctggtcaagttgcttccttt aagagattaagaaaagtcaccttcatcaactccgttccaaagtctgcttccggtaagattttgagaagagaattaat ccaaaaggttcgttccaacatgtag
SEQ ID NO:167 Artificial cannabidiolic acid synthase (CBDAS) nucleotide sequence	atgaaatgttccacctttttctttctggtttgtttgtaagatcatcttcttcttcttctccttcaacatccaaacttccatcgct aatccaagagagaatttcttaaagtgtttttctcaatacatcccaaacaatgctactaacttaaagttggtttacactca aaataacccattgtacatgtctgtcttgaactctaccattcacaatttgcgttttacttctgacaccacccctaagccat tagttattgttacccccatcccacgtctctcacatccaaggtactattttgtgttctaaaaaggttggtttgcaaattaga actagatctggtggtcacgactccgagggtatgtcttacatctctcaagttccattcgttattgtcgacttgcgtaaca tgcgttccatcaaaatcgatgttcactcccaaactgcttgggtcgaagccggtgccactttaggtgaggtttattact gggtcaatgagaagaatgagaatttgtccttggctgctggttattgtccaaccgtctgtgctggtggtcattttggtg gtggtggttacggtccattaatgagaaactatggtttggctgccgataacattatcgacgctcacttggttaatgtcc acggtaaggtcttagatagaaaatccatgggtgaggacttgttctgggctttgagaggtggtggtgctgagtccttt ggtatcatcgttgcttggaaaattcgtttagttgctgtcccaaaatctactatgttttctgttaagaagatcatggaaatt cacgagttggttaagttggttaataagtggcaaaatattgcctacaagtatgacaaagacttgttattgatgactcac ttcatcactagaaacatcaccgataaccaaggtaaaaataaaactgctatccatacctacttctcctccgttttcttgg gtggtgtcgactccttagttgatttgatgaacaaatcttttcctgaattaggtatcaagaagactgattgtcgtcaattg tcctggattgataccattatcttttactctggtgtcgtcaattacgacaccgataatttcaataaggaaattttattggac agatctgccggtcaaaacggtgctttcaagatcaagttggactacgttaaaaaaccaatcccagaatccgtctttgt ccaaattttggagaagttatacgaggaagacatcggtgctggtatgtatgccttatatccatacggtggtattatgg atgaaatttccgaatctgctatcccatttccacatcgtgctggtattttgtatgaattatggtacatttgttcctgggaaa agcaagaagataacgagaagcacttgaattggatcagaaatatctacaatttcatgactccttacgtttctaagaat

SEQ ID	SEQUENCE
	cctcgtttggcttacttgaactacagagatttggacatcggtattaatgacccaaagaacccaaataactatactca agctagaatttggggtgaaaagtacttcggtaaaaactttgacagattggttaaggttaagactttagttgatccaaa taacttcttcagaaatgaacaatccatcccaccattgcctagacacagacactaa
SEQ ID NO:168 Medium chain fatty acyl-CoA synthetase Sc_FAA2 nucleotide sequence Saccharomyces sp.	atggccgctccagattatgcacttaccgatttaattgaatcggatcctcgtttcgaaagtttgaagacaagattagcc ggttacaccaaaggctctgatgaatatattgaagagctatactctcaattaccactgaccagctaccccaggtaca aaacatttttaaagaaacaggcggttgccatttcgaatccggataatgaagctggttttagctcgatttataggagtt ctctttcttctgaaaatctagtgagctgtgtggataaaaacttaagaactgcatacgatcacttcatgttttctgcaag gagatggcctcaacgtgactgtttaggttcaaggccaattgataaagccacaggcacctgggaggaaacattcc gtttcgagtcgtactccacggtatctaaaagatgtcataatatcggaagtggtatattgtctttggtaaacacgaaaa ggaaacgtcctttggaagccaatgattttgttgttgctatcttatcacacaacaaccctgaatggatcctaacagattt ggcctgtcaggcctattctctaactaacacggctttgtacgaaacattaggtccaaacacctccgagtacatattga atttaaccgaggcccccattctgatttttgcaaaatcaaatatgtatcatgtattgaagatggtgcctgatatgaaattt gttaatactttggtttgtatggatgaattaactcatgacgagctccgtatgctaaatgaatcgttgctacccgttaagt gcaactctctcaatgaaaaaatcacatttttttcattggagcaggtagaacaagttggttgctttaacaaaattcctgc aattccacctaccccagattccttgtatactatttcgtttacttctggtactacaggtttacctaaaggtgtggaaatgt ctcacagaaacattgcgtctgggatagcatttgcttttctaccttcagaataccgccagataaaagaaaccaaca gttatatgatatgtgtttttgccattggctcatatttttgaaagaatggttattgcgtatgatctagccatcgggtttgga ataggcttcttacataaaccagacccaactgtattggtagaggatttgaagattttgaaaccttacgcggttgccct ggttcctagaatattaacacggtttgaagccggtataaaaaatgctttggataaatcgactgtccagaggaacgta gcaaatactatattggattctaaatcggccagatttaccgcaagaggtggtccagataaatcgattatgaattttcta gtttatcatcgcgtattgattgataaaatcagagactctttaggtttgtccaataactcgtttataattaccggatcagc tcccatatctaaagataccttactatttttaagaagcgccttggatattggtataagacagggctacggcttaactga aacttttgctggtgtctgtttaagcgaaccgtttgaaaaagatgtcggatcttgtggtgccataggtatttctgcaga atgtagattgaagtctgttccagaaatgggttaccatgccgacaaggatttaaaaggtgaactgcaaattcgtggc ccacaggtttttgaaagatattttaaaaatccgaatgaaacttcaaaagccgttgaccaagatggttggttttccacg ggagatgttgcatttatcgatgcaaaaggtcgcatcagcgtcattgatcgagtcaagaacttttttcaagctagcaca tggtgaatatattgctccagagaaaatcgaaaatatttatttatcatcatgcccctatatcacgcaaatatttgtctttg gagatcctttgaagacatttttagttggcatcgttggtgttgatgttgatgcagcgcaaccgattttagctgcaaagc acccagaggtgaaaacgtggactaaggaagtgctagtagaaaacttaaatcgtaataaaaagctaaggaagga atttttaaacaaaattaataaatgcatcgatgggctacaaggatttgaaaaattgcacaacatcaaagtcggacttg agcctttgactctcgaggatgatgttgtgacgccaacttttaaaataaagcgtgccaaagcatcaaaattcttcaaa gatacattagaccaactatacgccgaaggttcactagtcaagacagaaaagctttag
SEQ ID NO:169 Medium chain fatty acyl-CoA synthetase Sc_FAA2 Saccharomyces sp.	MAAPDYALTDLIESDPRFESLKTRLAGYTKGSDEYIEELYSQLPLTSYP RYKTFLKKQAVAISNPDNEAGFSSIYRSSLSSENLVSCVDKNLRTAYDH FMFSARRWPQRDCLGSRPIDKATGTWEETFRFESYSTVSKRCHNIGSGI LSLVNTKRKRPLEANDFVVAILSHNNPEWILTDLACQAYSLTNTALYE TLGPNTSEYILNLTEAPILIFAKSNMYHVLKMVPDMKFVNTLVCMDEL THDELRMLNESLLPVKCNSLNEKITFFSLEQVEQVGCFNKIPAIPPTPDS LYTISFTSGTTGLPKGVEMSHRNIASGIAFAFSTFRIPPDKRNQQLYDMC FLPLAHIFERMVIAYDLAIGFGIGFLHKPDPTVLVEDLKILKPYAVALVP RILTRFEAGIKNALDKSTVQRNVANTILDSKSARFTARGGPDKSIMNFL VYHRVLIDKIRDSLGLSNNSFIITGSAPISKDTLLFLRSALDIGIRQGYGL TETFAGVCLSEPFEKDVGSCGAIGISAECRLKSVPEMGYHADKDLKGE LQIRGPQVFERYFKNPNETSKAVDQDGWFSTGDVAFIDAKGRISVIDR VKNFFKLAHGEYIAPEKIENIYLSSCPYITQIFVFGDPLKTFLVGIVGVD VDAAQPILAAKHPEVKTWTKEVLVENLNRNKKLRKEFLNKINKCIDGL QGFEKLHNIKVGLEPLTLEDDVVTPTFKIKRAKASKFFKDTLDQLYAE GSLVKTEKL*
SEQ ID NO:170 MBPtag	atgaaaatcgaagagggtaaattggtcatctggatcaatggtgacaaaggttacaacggtttggctgaagtcggt aaaaaattcgagaaagacactggtattaaggttaccgtcgaacacccagataagttggaagaaaaatttccacaa gttgccgctactggtgatggtccagacatcattttctgggcccacgacagatttggtggttatgctcaatctggtttg ttagccgagatcaccccagacaaagcctttcaagataaattatacccatttacctgggatgctgtccgttacaacg gtaagttgatcgcttacccaatcgccgttgaagctttgtctttaatctacaataaagacttattgccaaaccctccaaa

SEQ ID	SEQUENCE
	gacctgggaagaaattcctgccttggataaggaattaaaggctaaaggtaaatctgccttaatgttcaacttacaa gagccttactttacttggccattgattgctgctgatggtggttatgcttttaagtacgaaaatggtaaatacgacatta aagatgttggtgttgacaatgccggtgctaaagccggtttaactttcttagtcgacttgatcaagaacaagcacatg aatgctgacactgattattctatcgctgaagccgccttcaacaagggtgaaactgctatgactatcaatggtccttg ggcctggtctaatattgacacctccaaagtcaactacggtgttactgtcttaccaactttcaaaggtcaaccttccaa gccatttgtcggtgttttgtctgctggtattaacgctgcctctccaaacaaagaattggccaaggaatttttggaaaa ctacttgttgactgacgaaggtttagaggctgttaacaaagacaaaccattgggtgctgtcgccttgaaatcctac gaagaagaattagccaaggatccaagaatcgccgctaccatggaaaatgctcaaaaaggtgaaattatgccaa acattccacaaatgtccgcttttttggtacgctgttagaactgctgttattaatgctgcttctggtagacaaactgtcga tgaagctttgaaggacgctcaaaccagaatcactaag
SEQ ID NO:171 GS12 Linker	ggaggtggaggaggtggttccggaggaggtggttct
SEQ ID NO:172 GS12 Linker	GGGGGGSGGGGS
SEQ ID NO:173 GB1 tag	atgtctgacacttacaagttgatcttgaacggtaagactttgaaaggtgaaactactaccgaagctgttgatgctgc cactgctgaaaaggtttttaagcaatacgccaatgataacggtgtcgacggtgaatggacttacgatgatgccact aagacttttaccgttactgaa
SEQ ID NO:174 GB1 tag	MSDTYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWT YDDATKTFTVTE
SEQ ID NO:175 MFalpha1_1-19	atgagatttccttcaatttttactgcagttttattcgcagcatcctccgcattagct
SEQ ID NO:176 MFalpha1_1-19	MRFPSIFTAVLFAASSALA
SEQ ID NO:177 MFalpha1_1-89	atgagatttccttcaatttttactgcagttttattcgcagcatcctccgcattagctgctccagtcaacactacaacag aagatgaaacggcacaaattccggctgaagctgtcatcggttacttagatttagaaggggatttcgatgttgctgtt ttgccattttccaacagcacaaataacgggttattgtttataaatactactattgccagcattgctgctaaagaagaa ggggtatctttggataaaagagaggctgaagct
SEQ ID NO:178 MFalpha1_1-89	MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYLDLEGDFD VAVLPFSNSTNNGLLFINTTIASIAAKEEGVSLDKREAEA
SEQ ID NO:179 DasherGFP	atgaccgcactaacagaaggagctaaactattcgaaaaggagattccttacattacagaattagagggtgatgtc gaaggaatgaaattcattatcaagggcgagggtactggtgacgctactaccggtacgattaaagcaaagtacat ctgtacaacaggtgaccttcctgttccgtgggctactctggtgagcactttgtcttatggagttcaatgttttgctaaa taccttcgcacattaaagacttttttcaaaagtgcaatgcctgagggctatactcaggagagaacaatatctttcga aggagatggtgtgtataagactagggctatggtcacgtatgaaagaggatccatctacaatagagtaactttaact ggtgaaaacttcaaaaaggacggtcacatccttagaaagaatgttgcctttcaatgcccaccatccatcttgtacat tttgccagacacagttaacaatggtatcagagttgagtttaaccaagcttatgacatagagggtgtcaccgaaaag ttggttacaaaatgttcacagatgaatcgtcccctggcaggatcagctgccgtccatatcccacgttaccatcatat cacttatcataccaagctgtccaaagatcgtgatgagagaagggatcacatgtgtttggttgaagtggtaaaggc cgtggatttggatacttaccaaggttga
SEQ ID NO:180 DasherGFP	MTALTEGAKLFEKEIPYITELEGDVEGMKFIIKGEGTGDATTGTIKAKYI CTTGDLPVPWATLVSTLSYGVQCFAKYPSHIKDFFKSAMPEGYTQERT ISFEGDGVYKTRAMVTYERGSIYNRVTLTGENFKKDGHILRKNVAFQC PPSILYILPDTVNNGIRVEFNQAYDIEGVTEKLVTKCSQMNRPLAGSAA VHIPRYHHITYHTKLSKDRDERRDHMCLVEVVKAVDLDTYQG*

SEQ ID	SEQUENCE
SEQ ID NO:181 ER1 tag	atattagagcaacctctgaaatttgtgcttactgcggccgtcgtgctcttgacgacgtcggttctttgttgtgtagtatt tacataa
SEQ ID NO:182 ER1 tag	ILEQPLKFVLTAAVVLLTTSVLCCVVFT*
SEQ ID NO:183 ER2 tag	tctacctctgaaaaccaaagtaaaggtagtggtacattggttgtcatattggccattttaatgctaggtgttgcttatta tttgttgaacgaataa
SEQ ID NO:184 ER2 tag	STSENQSKGSGTLVVILAILMLGVAYYLLNE*
SEQ ID NO:185 PM1 tag	tggtacaaggatctaaaaatgaagatgtgtctggctttagtaatcatcatattgcttgttgtaatcatcgtccccattg ctgttcactttagtcgataa
SEQ ID NO:186 PM1 tag	WYKDLKMKMCLALVIIILLVVIIVPIAVHFSR*
SEQ ID NO:187 VC1 tag	aatataaaagaaataatgtggtggcagaaggtcaaaaatattacgttattaactttcactattatactatttgtaagtg ctgctttcatgtttttctatctgtggtaa
SEQ ID NO:188 VC1 tag	NIKEIMWWQKVKNITLLTFTIILFVSAAFMFFYLW*
SEQ ID NO:189 PEX8 tag	tctaaattataa
SEQ ID NO:190 PEX8 tag	SKL*
SEQ ID NO:191 Long chain fatty acyl-CoA synthetase Sc_FAA1 Saccharomyces cerevisiae	atggttgctcaatataccgttccagttgggaaagccgccaatgagcatgaaactgctccaagaagaaattatcaat gccgcgagaagccgctcgtcagaccgcctaacacaaagtgttccactgtttatgagtttgttctagagtgctttca gaagaacaaaaattcaaatgctatgggttggagggatgttaaggaaattcatgaagaatccaaatcggttatgaa aaaagttgatggcaaggagacttcagtggaaaagaaatggatgtattatgaactatcgcattatcattataattcatt tgaccaattgaccgatatcatgcatgaaattggtcgtgggttggtgaaaataggattaaagcctaatgatgatgac aaattacatctttacgcagccacttctcacaagtggatgaagatgttcttaggagcgcagtctcaaggtattcctgtc gtcactgcctacgatactttgggagagaaagggctaattcattctttggtgcaaacggggtctaaggccatttttac cgataactctttattaccatccttgatcaaaccagtgcaagccgctcaagacgtaaaatacataattcatttcgattc catcagttctgaggacaggaggcaaagtggtaagatctatcaatctgctcatgatgccatcaacagaattaaaga agttagacctgatatcaagacctttagctttgacgacatcttgaagctaggtaaagaatcctgtaacgaaatcgatg ttcatccacctggcaaggatgatctttgttgcatcatgtatacgtctggttctacaggtgagccaaagggtgttgtctt gaaacattcaaatgttgtcgcaggtgttggtggtgcaagtttgaatgttttgaagtttgtgggcaatacgaccgtgt tatctgtttttttgccactagctcatattttttgaattggttttcgaactattgtccttttattgggggggcctgcattggttatg ccaccgtaaaaactttaactagcagctctgtgagaaattgtcaaggtgatttgcaagaattcaagcccacaatcat ggttggtgtcgccgctgtttgggaaacagtgagaaaagggatcttaaaccaaattgataatttgcccttcctcacc aagaaaatcttctggaccgcgtataataccaagttgaacatgcaacgtctccacatccctggtggcggcgccttа ggaaacttggttttcaaaaaaatcagaactgccacaggtggccaattaagatatttgttaaacggtggttctccaat cagtcgggatgctcaggaattcatcacaaatttaatctgccctatgcttattggttacggtttaaccgagacatgcg ctagtaccaccatcttggatcctgctaattttgaactcggcgtcgctggtgacctaacaggttgtgttaccgtcaaa ctagttgatgttgaagaattaggttattttgctaaaaacaaccaaggtgaagtttggatcacaggtgccaatgtcac gcctgaatattataagaatgaggaagaaacttctcaagctttaacaagcgatggttggttcaagaccggtgacatc ggtgaatgggaagcaaatggccatttgaaaataattgacaggaagaaaaacttggtcaaaacaatgaacggtga

SEQ ID	SEQUENCE
	atatatcgcactcgagaaattagagtccgtttacagatctaacgaatatgttgctaacatttgtgtttatgccgaccaa tctaagactaagccagttggtattattgtaccaaatcatgctccattaacgaagcttgctaaaaagttgggaattatg gaacaaaaagacagttcaattaatatcgaaaattatttggaggatgcaaaattgattaaagctgtttattctgatctttt gaagacaggtaaagaccaaggtttggttggcattgaattactagcaggcatagtgttctttgacggcgaatggact ccacaaaacggttttgttacgtccgctcagaaattgaaaagaaaagacattttgaatgctgtcaaagataaagttga cgccgtttatagttcgtcttaa
SEQ ID NO:192 Long chain fatty acyl-CoA synthetase Sc_FAA1 Saccharomyces cerevisiae	MVAQYTVPVGKAANEHETAPRRNYQCREKPLVRPPNTKCSTVYEFVL ECFQKNKNSNAMGWRDVKEIHEESKSVMKKVDGKETSVEKKWMYY ELSHYHYNSFDQLTDIMHEIGRGLVKIGLKPNDDDKLHLYAATSHKW MKMFLGAQSQGIPVVTAYDTLGEKGLIHSLVQTGSKAIFTDNSLLPSLI KPVQAAQDVKYIIHFDSISSEDRRQSGKIYQSAHDAINRIKEVRPDIKTF SFDDILKLGKESCNEIDVHPPGKDDLCCIMYTSGSTGEPKGVVLKHSNV VAGVGGASLNVLKFVGNTDRVICFLPLAHIFELVFELLSFYWGACIGY ATVKTLTSSSVRNCQGDLQEFKPTIMVGVAAVWETVRKGILNQIDNLP FLTKKIFWTAYNTKLNMQRLHIPGGGALGNLVFKKIRTATGGQLRYLL NGGSPISRDAQEFITNLICPMLIGYGLTETCASTTILDPANFELGVAGDL TGCVTVKLVDVEELGYFAKNNQGEVWITGANVTPEYYKNEEETSQAL TSDGWFKTGDIGEWEANGHLKIIDRKKNLVKTMNGEYIALEKLESVY RSNEYVANICVYADQSKTKPVGIIVPNHAPLTKLAKKLGIMEQKDSSIN IENYLEDAKLIKAVYSDLLKTGKDQGLVGIELLAGIVFFDGEWTPQNG FVTSAQKLKRKDILNAVKDKVDAVYSSS*
SEQ ID NO:193 Truncated medium chain fatty acyl-CoA synthetase Sc_FAA2_Ctrunc	atggccgctccagattatgcacttaccgatttaattgaatcggatcctcgtttcgaaagtttgaagacaagattagcc ggttacaccaaaggctctgatgaatatattgaagagctatactctcaattaccactgaccagctaccccaggtaca aaacatttttaaagaaacaggcggttgccatttcgaatccggataatgaagctggttttagctcgatttataggagtt ctctttcttctgaaaatctagtgagctgtgtggataaaaacttaagaactgcatacgatcacttcatgttttctgcaag gagatggcctcaacgtgactgtttaggttcaaggccaattgataaagccacaggcacctgggaggaaacattcc gtttcgagtcgtactccacggtatctaaaagatgtcataatatcggaagtggtatattgtctttggtaaacacgaaaa ggaaacgtcctttggaagccaatgattttgttgttgctatcttatcacacaacaaccctgaatggatcctaacagattt ggcctgtcaggcctattctctaactaacacggctttgtacgaaacattaggtccaaacacctccgagtacatattga atttaaccgaggcccccattctgatttttgcaaaatcaaatatgtatcatgtattgaagatggtgcctgatatgaaattt gttaatactttggtttgtatggatgaattaactcatgacgagctccgtatgctaaatgaatcgttgctacccgttaagt gcaactctctcaatgaaaaaatcacatttttttcattggagcaggtagaacaagttggttgctttaacaaaattcctgc aattccacctaccccagattccttgtatactatttcgtttacttctggtactacaggtttacctaaaggtgtggaaatgt ctcacagaaacattgcgtctgggatagcatttgctttttctaccttcagaataccgccagataaaagaaaccaaca gttatatgatatgtgtttttttgccattggctcatatttttgaaagaatggttattgcgtatgatctagccatcgggtttgga ataggcttcttacataaaccagacccaactgtattggtagaggatttgaagattttgaaaccttacgcggttgccct ggttcctagaatattaacacggtttgaagccggtataaaaaatgctttggataaatcgactgtccagaggaacgta gcaaatactatattggattctaaatcggccagatttaccgcaagaggtggtccagataaatcgattatgaatttctta gtttatcatcgcgtattgattgataaaatcagagactctttaggtttgtccaataactcgtttataattaccggatcagc tcccatatctaaagataccttactatttttaagaagcgccttggatattggtataagacagggctacggcttaactga aacttttgctggtgtctgtttaagcgaaccgtttgaaaaagatgtcggatcttgtggtgccataggtatttctgcaga atgtagattgaagtctgttccagaaatgggttaccatgccgacaaggatttaaaaggtgaactgcaaattcgtggc ccacaggtttttgaaagatattttaaaaatccgaatgaaacttcaaaagccgttgaccaagatggttggttttccacg ggagatgttgcatttatcgatgcaaaaggtcgcatcagcgtcattgatcgagtcaagaacttttcaagctagcaca tggtgaatatattgctccagagaaaatcgaaaatatttatttatcatcatgcccctatatcacgcaaatatttgtctttg gagatccttgaagacatttttagttggcatcgttggtgttgatgttgatgcagcgcaaccgattttagctgcaaagc acccagaggtgaaaacgtggactaaggaagtgctagtagaaaacttaaatcgtaataaaaagctaaggaagga atttttaaacaaaattaataaatgcatcgatgggctacaaggatttgaaaaattgcacaacatcaaagtcggacttg agcctttgactctcgaggatgatgttgtgacgccaacttttaaaataaagcgtgccaaagcatcaaaattcttcaaa gatacattagaccaactatacgccgaaggttcactagtcaagacatag
SEQ ID NO:194 Truncated medium chain fatty acyl-CoA	MAAPDYALTDLIESDPRFESLKTRLAGYTKGSDEYIEELYSQLPLTSYP RYKTFLKKQAVAISNPDNEAGFSSIYRSSLSSENLVSCVDKNLRTAYDH FMFSARRWPQRDCLGSRPIDKATGTWEETFRFESYSTVSKRCHNIGSGI

SEQ ID	SEQUENCE
synthetase Sc_FAA2_Ctrunc	LSLVNTKRKRPLEANDFVVAILSHNNPEWILTDLACQAYSLTNTALYE TLGPNTSEYILNLTEAPILIFAKSNMYHVLKMVPDMKFVNTLVCMDEL THDELRMLNESLLPVKCNSLNEKITFFSLEQVEQVGCFNKIPAIPPTPDS LYTISFTSGTTGLPKGVEMSHRNIASGIAFAFSTFRIPPDKRNQQLYDMC FLPLAHIFERMVIAYDLAIGFGIGFLHKPDPTVLVEDLKILKPYAVALVP RILTRFEAGIKNALDKSTVQRNVANTILDSKSARFTARGGPDKSIMNFL VYHRVLIDKIRDSLGLSNNSFIITGSAPISKDTLLFLRSALDIGIRQGYGL TETFAGVCLSEPFEKDVGSCGAIGISAECRLKSVPEMGYHADKDLKGE LQIRGPQVFERYFKNPNETSKAVDQDGWFSTGDVAFIDAKGRISVIDR VKNFFKLAHGEYIAPEKIENIYLSSCPYITQIFVFGDPLKTFLVGIVGVD VDAAQPILAAKHPEVKTWTKEVLVENLNRNKKLRKEFLNKINKCIDGL QGFEKLHNIKVGLEPLTLEDDVVTPTFKIKRAKASKFFKDTLDQLYAE GSLVKT*
SEQ ID NO:195 Mutated medium chain fatty acyl-CoA synthetase Sc_FAA2_Cmut	atggccgctccagattatgcacttaccgatttaattgaatcggatcctcgtttcgaaagtttgaagacaagattagcc ggttacaccaaaggctctgatgaatatattgaagagctatactctcaattaccactgaccagctaccccaggtaca aaacatttttaaagaaacaggcggttgccatttcgaatccggataatgaagctggttttagctcgatttataggagtt ctctttcttctgaaaatctagtgagctgtgtggataaaaacttaagaactgcatacgatcacttcatgttttctgcaag gagatggcctcaacgtgactgtttaggttcaaggccaattgataaagccacaggcacctgggaggaaacattcc gtttcgagtcgtactccacggtatctaaaagatgtcataatatcggaagtggtatattgtctttggtaaacacgaaaa ggaaacgtcctttggaagccaatgattttgttgttgctatcttatcacacaacaaccctgaatggatcctaacagattt ggcctgtcaggcctattctctaactaacacggctttgtacgaaacattaggtccaaacacctccgagtacatattga atttaaccgaggcccccattctgatttttgcaaaatcaaatatgtatcatgtattgaagatggtgcctgatatgaaattt gttaatactttggtttgtatggatgaattaactcatgacgagctccgtatgctaaatgaatcgttgctacccgttaagt gcaactctctcaatgaaaaaatcacatttttttcattggagcaggtagaacaagttggttgctttaacaaaattcctgc aattccacctaccccagattccttgtatactatttcgtttacttctggtactacaggtttacctaaaggtgtggaaatgt ctcacagaaacattgcgtctgggatagcatttgctttttctaccttcagaataccgccagataaaagaaaccaaca gttatatgatatgtgttttttgccattggctcatatttttgaaagaatggttattgcgtatgatctagccatcgggtttgga ataggcttcttacataaaccagacccaactgtattggtagaggatttgaagattttgaaaccttacgcggttgccct ggttcctagaatattaacacggtttgaagccggtataaaaaatgctttggataaatcgactgtccagaggaacgta gcaaatactatattggattctaaatcggccagatttaccgcaagaggtggtccagataaatcgattatgaattttcta gtttatcatcgcgtattgattgataaaatcagagactctttaggtttgtccaataactcgtttataattaccggatcagc tcccatatctaaagatacctactatttttaagaagcgccttggatattggtataagacagggctacggcttaactga aactttgctggtgtctgtttaagcgaaccgtttgaaaaagatgtcggatcttgtggtgccataggtatttctgcaga atgtagattgaagtctgttccagaaatgggttaccatgccgacaaggatttaaaaggtgaactgcaaattcgtggc ccacaggtttttgaaagatattttaaaaatccgaatgaaacttcaaaagccgttgaccaagatggttggttttccacg ggagatgttgcatttatcgatgcaaaaggtcgcatcagcgtcattgatcgagtcaagaacttttttcaagctagcaca tggtgaatatattgctccagagaaaatcgaaaatatttatttatcatcatgcccctatatcacgcaaatatttgtctttg gagatcctttgaagacatttttagttggcatcgttggtgttgatgttgatgcagcgcaaccgattttagctgcaaagc acccagaggtgaaaacgtggactaaggaagtgctagtagaaaacttaaatcgtaataaaaagctaaggaagga atttttaaacaaaattaataaatgcatcgatgggctacaaggatttgaaaaattgcacaacatcaaagtcggacttg agcctttgactctcgaggatgatgttgtgacgccaacttttaaaataaagcgtgccaaagcatcaaaattcttcaaa gatacattagaccaactatacgccgaaggttcactagtcaagacagaaaagcttaaatag
SEQ ID NO:196 Mutated medium chain fatty acyl-CoA synthetase Sc_FAA2_Cmut	MAAPDYALTDLIESDPRFESLKTRLAGYTKGSDEYIEELYSQLPLTSYP RYKTFLKKQAVAISNPDNEAGFSSIYRSSLSSENLVSCVDKNLRTAYDH FMFSARRWPQRDCLGSRPIDKATGTWEETFRFESYSTVSKRCHNIGSGI LSLVNTKRKRPLEANDFVVAILSHNNPEWILTDLACQAYSLTNTALYE TLGPNTSEYILNLTEAPILIFAKSNMYHVLKMVPDMKFVNTLVCMDEL THDELRMLNESLLPVKCNSLNEKITFFSLEQVEQVGCFNKIPAIPPTPDS LYTISFTSGTTGLPKGVEMSHRNIASGIAFAFSTFRIPPDKRNQQLYDMC FLPLAHIFERMVIAYDLAIGFGIGFLHKPDPTVLVEDLKILKPYAVALVP RILTRFEAGIKNALDKSTVQRNVANTILDSKSARFTARGGPDKSIMNFL VYHRVLIDKIRDSLGLSNNSFIITGSAPISKDTLLFLRSALDIGIRQGYGL TETFAGVCLSEPFEKDVGSCGAIGISAECRLKSVPEMGYHADKDLKGE

SEQ ID	SEQUENCE
	LQIRGPQVFERYFKNPNETSKAVDQDGWFSTGDVAFIDAKGRISVIDR VKNFFKLAHGEYIAPEKIENIYLSSCPYITQIFVFGDPLKTFLVGIVGVD VDAAQPILAAKHPEVKTWTKEVLVENLNRNKKLRKEFLNKINKCIDGL QGFEKLHNIKVGLEPLTLEDDVVTPTFKIKRAKASKFFKDTLDQLYAE GSLVKTEKLK*
SEQ ID NO:197 Long-chain fatty acyl-CoA synthetase Sc_FAA3 Saccharomyces cerevisiae	atgtccgaacaacactctgtcgcagtcggtaaagctgctaatgagcacgagactgcccctaggagaaatgttag agtcaagaagcggcccttaattagaccattgaactcgtcagcatctacgctgtatgaatttgccctagagtgtttca acaagggtggaaaacgagatggtatggcttggagagatgtcatcgagattcatgagacaaagaaaaccattgtg agaaaggtagacggcaaggataaatctatagaaaagacatggctgtattatgaaatgtcaccatataaaatgatg acctaccaggaactgatctgggtgatgcacgatatgggccgtgggctggcaaaaataggcatcaagcccaatg gagaacacaaattccacatcttcgcatctacttcccataaatggatgaagattttccttggttgcatatcccagggta tccccgtagtaaccgcgtatgatactttgggtgagagcggtttgattcactccatggttgaaaccgagtctgctgct attttcactgataatcaattattggctaaaatgatagtgcctttgcaatctgctaaagatatcaaatttcttatccataac gaacctatcgaccccaatgacagaagacaaaacggcaaactttacaaggctgctaaggatgccattaataagat cagagaagttaggccagacataaaaatttatagttttgaagaagttgtcaagataggtaaaaaaagtaaagatga ggtcaaacttcatccacctgagccaaaagatttggcttgtatcatgtacacctcgggctcgatcagtgcaccaaaa ggtgtagtattgactcattataatattgtttcgggtatcgctggtgtaggtcacaacgtctttggatggatcggctcta cagaccgtgttttgtcgttcttgccattggctcatatttttgaactggtctttgaattcgaagccttttactggaacggta ttcttgggtacggtagtgttaagactttgactaatacttcgactcgtaattgtaagggtgacctggttgagtttaagcc tactattatgatcggtgtggctgccgtttgggaaactgtgagaaaagctattttggaaaagatcagcgatttaactc ccgtactccaaaagatttttttggtctgcctatagtatgaaagaaaagagtgtaccatgcaccgggtttttaagtcgta tggtcttcaagaaagtcagacaagccaccggtggtcatcttaagtatattatgaacggtgggtctgcgatcagtatt gatgctcagaaattcttttctatcgtcctgtgtcctatgattatcggttacggccttactgaaacagttgcgaatgcttg tgttttggagcctgatcatttcgaatatggtatagttggtgatcttgttggatcggtcactgccaaattggtggatgtta aggacctaggttattatgcaaaaaacaatcaaggtgaattgcttctaaagggtgcgccggtctgttctgaatattat aagaatccaatagaaacggcggtctctttcacttacgatggatggtttcgtactggtgatattgttgaatggactccc aagggacaacttaaaattattgatagaagaaagaatttggttaaaaccctaaatggtgaatatattgcattagaaaa gttagaatctgtttacaggtcaaactcctatgtgaaaaatatctgtgtttatgccgatgaaagtagggttaaaccggt gggtattgtggtacccaacccaggacccctatctaaatttgctgtcaaattgcgtattatgaaaaagggtgaagac atcgaaaactatatccatgacaaagcattacgaaatgctgttttcaaagagatgatcgcaacagccaaatctcaag gtttggttggtattgaactattatgtggtattgttttctttgatgaagaatggacacctgaaaatggctttgtcacatctg ctcaaaaattaaagagaagagaaatcttagccgctgttaaatcagaagtcgaaagggtttacaaagaaaattctta g
SEQ ID NO:198 Long-chain fatty acyl-CoA synthetase Sc_FAA3 Saccharomyces cerevisiae	MSEQHSVAVGKAANEHETAPRRNVRVKKRPLIRPLNSSASTLYEFALE CFNKGGKRDGMAWRDVIEIHETKKTIVRKVDGKDKSIEKTWLYYEMS PYKMMTYQELIWVMHDMGRGLAKIGIKPNGEHKFHIFASTSHKWMKI FLGCISQGIPVVTAYDTLGESGLIHSMVETESAAIFTDNQLLAKMIVPLQ SAKDIKFLIHNEPIDPNDRRQNGKLYKAAKDAINKIREVRPDIKIYSFEE VVKIGKKSKDEVKLHPPEPKDLACIMYTSGSISAPKGVVLTHYNIVSGI AGVGHNVFGWIGSTDRVLSFLPLAHIFELVFEFEAFYWNGILGYGSVK TLTNTSTRNCKGDLVEFKPTIMIGVAAVWETVRKAILEKISDLTPVLQK IFWSAYSMKEKSVPCTGFLSRMVFKKVRQATGGHLKYIMNGGSAISID AQKFFSIVLCPMIIGYGLTETVANACVLEPDHFEYGIVGDLVGSVTAKL VDVKDLGYYAKNNQGELLLKGAPVCSEYYKNPIETAVSFTYDGWFRT GDIVEWTPKGQLKIIDRRKNLVKTLNGEYIALEKLESVYRSNSYVKNIC VYADESRVKPVGIVVPNPGPLSKFAVKLRIMKKGEDIENYIHDKALRN AVFKEMIATAKSQGLVGIELLCGIVFFDEEWTPENGFVTSAQKLKRREI LAAVKSEVERVYKENS*
SEQ ID NO:199 Long-chain fatty acyl-CoA synthetase Sc_FAA4	atgaccgaacaatattccgttgcagttggcgaagccgacaatgagcatgaaaccgctccaagaagaaatatcag ggttaaagacaagcctttgattagacccataaactcctcagcatctacactgtacgaattcgccctggaatgttttac caaaggtggtaagagagacggtatggcatggagagatattatagatatacatgagacgaaaaaaaccatagtca agagggtggatggtaaggataagcccatcgaaaaaacatggttgtactacgaactgactccctacataaccatg acatacgaggagatgatctgcgtaatgcacgacattggacgtgggctgataaagattggtgttaaacctaacggt

SEQ ID	SEQUENCE
Saccharomyces cerevisiae	gagaacaagttccacatctttgcctctacatctcacaagtggatgaaaactttcttggttgcatgtcacaaggtattc ctgtggtcaccgcgtacgacactttgggtgagagcggtttgattcactccatggtggaaacggattccgtcgccat tttcacggacaaccagctgttgtccaaattagcagttcctttgaaaaccgccaagaacgtaaaattcgtcattcaca acgaacccatcgatccaagtgacaaaagacaaaatggtaagctttacaaggctgccaaggatgctgttgacaaa atcaaggaagttagaccggacataaaaatctacagtttcgatgaaattattgagataggtaaaaaggccaaggac gaggttgaattgcatttccccaagcctgaagatccagcttgtatcatgtacacttctggttccactggtacaccaaa gggtgtggtattgacacattacaacattgtagctggtattggtggtgtgggccataacgttatcggatggattggcc caacagaccgtattatcgcattcttgccattggctcatatttttgaattaatctttgaattcgaagcgttctactggaat ggtatcctagggtacgccactgtcaagactttaaccccaacttctacacgtaattgccaaggtgacctgatggagt ttaaacctaccgtaatggtaggtgttgccgcagtttgggaaacagtgagaaaaggtatcctggccaagatcaacg aattgcccggttggtctcaaacgcttttctggactgtctatgctttgaaagagagaaatataccatgcagcggcttg ctgagtgggttgatcttcaagagaatcagagaagcaaccggtggaaacttaaggtttattctgaacggtgggtct gcaatcagcatagacgcccaaaaattcctctccaaccttctatgtcctatgctcattggatatgggctaactgaggg tgtggctaatgcctgtgtcctggagcctgaacattttgattacggtattgctggtgaccttgtcggaactattacagc taaattggtggatgtcgaagatttgggctattttgccaagaataaccaaggtgaattgctgttaaagggtgcaccca tctgttctgaatactataagaatcctgaagaaactgctgcggcctttaccgatgatggctggttccgtaccggtgat atcgctgaatggacccccaagggacaaattaagatcattgatagaaagaaaaatttggtcaagaccttaaacggt gagtacattgcattggaaaaattagaatccatttacagatcaaatccttacgtccaaaacatctgtgtctacgctgat gaaaacaaagttaagcctgtcggtattgtggtccctaacttaggacacttgtctaagctggctatcgaattaggtat aatggtaccaggtgaagatgtcgaaagctatatccatgaaaagaagctacaggatgccgtttgcaaagatatgct gtcaactgccaaatctcaaggcttgaatggtattgaattattatgtggcattgttttctttgaagaagaatggactcca gaaaacggtcttgttacatccgcccaaaaattaaagagaagagatattctagcggctgtcaagccagatgtggaa agagtttataaagaaaacacttaa
SEQ ID NO:200 Sc_FAA4 Saccharomyces cerevisiae	MTEQYSVAVGEADNEHETAPRRNIRVKDKPLIRPINSSASTLYEFALEC FTKGGKRDGMAWRDIIDIHETKKTIVKRVDGKDKPIEKTWLYYELTPY ITMTYEEMICVMHDIGRGLIKIGVKPNGENKFHIFASTSHKWMKTFLG CMSQGIPVVTAYDTLGESGLIHSMVETDSVAIFTDNQLLSKLAVPLKTA KNVKFVIHNEPIDPSDKRQNGKLYKAAKDAVDKIKEVRPDIKIYSFDEII EIGKKAKDEVELHFPKPEDPACIMYTSGSTGTPKGVVLTHYNIVAGIGG VGHNVIGWIGPTDRIIAFLPLAHIFELIFEFEAFYWNGILGYATVKTLTPT STRNCQGDLMEFKPTVMVGVAAVWETVRKGILAKINELPGWSQTLF WTVYALKERNIPCSGLLSGLIFKRIREATGGNLRFILNGGSAISIDAQKF LSNLLCPMLIGYGLTEGVANACVLEPEHFDYGIAGDLVGTITAKLVDV EDLGYFAKNNQGELLLKGAPICSEYYKNPEETAAAFTDDGWFRTGDIA EWTPKGQIKIIDRKKNLVKTLNGEYIALEKLESIYRSNPYVQNICVYAD ENKVKPVGIVVPNLGHLSKLAIELGIMVPGEDVESYIHEKKLQDAVCK DMLSTAKSQGLNGIELLCGIVFFEEEWTPENGLVTSAQKLKRRDILAA VKPDVERVYKENT*
SEQ ID NO:201 Mutated acetyl-CoA carboxylase (ACC1) (S659A, S1157A)	atgagcgaagaaagcttattcgagtcttctccacagaagatggagtacgaaattacaaactactcagaaagacat acagaacttccaggtcatttcattggcctcaatacagtagataaactagaggagtccccgttaagggactttgttaa gagtcacggtggtcacacggtcatatccaagatcctgatagcaaataatggtattgccgccgtgaaagaaattag atccgtcagaaaatgggcatacgagacgttcggcgatgacagaaccgtccaattcgtcgccatggccacccca gaagatctggaggccaacgcagaatatatccgtatggccgatcaatacattgaagtgccaggtggtactaataat aacaactacgctaacgtagacttgatcgtagacatcgccgaaagagcagacgtagacgccgtatgggctggct ggggtcacgcctccgagaatccactattgcctgaaaaattgtcccagtctaagaggaaagtcatctttattgggcc tccaggtaacgccatgaggtctttaggtgataaaatctcctctaccattgtcgctcaaagtgctaaagtcccatgtat tccatggtctggtaccggtgttgacaccgttcacgtggacgagaaaaccggtctggtctctgtcgacgatgacat ctatcaaaagggttgttgtacctctcctgaagatggtttacaaaaggccaagcgtattggttttcctgtcatgattaa ggcatccgaaggtggtggtggtaaaggtatcagacaagttgaacgtgaagaagatttcatcgctttataccacca ggcagccaacgaaattccaggctcccccattttcatcatgaagttggccggtagagcgcgtcacttggaagttca actgctagcagatcagtacggtacaaatatttccttgttcggtagagactgttccgttcagagacgtcatcaaaaaa ttatcgaagaagcaccagttacaattgccaaggctgaaacatttcacgagatggaaaaggctgccgtcagactg gggaaactagtcggttatgtctctgccggtaccgtggagtatctatattctcatgatgatggaaaattctacttttag

SEQ ID	SEQUENCE
	aattgaacccaagattacaagtcgagcatccaacaacggaaatggtctccggtgttaacttacctgcagctcaatt acaaatcgctatgggtatccctatgcatagaataagtgacattagaactttatatggtatgaatcctcattctgcctca gaaatcgatttcgaattcaaaactcaagatgccaccaagaaacaaagaagacctattccaaagggtcattgtacc gcttgtcgtatcacatcagaagatccaaacgatggattcaagccatcgggtggtactttgcatgaactaaacttcc gttcttcctctaatgtttggggttacttctccgtgggtaacaatggtaatattcactccttttcggactctcagttcggc catatttttgcttttggtgaaaatagacaagcttccaggaaacacatggttgttgccctgaaggaattgtccattagg ggtgatttcagaactactgtggaatacttgatcaaacttttggaaactgaagatttcgaggataacactattaccacc ggttggttggacgatttgattactcataaaatgaccgctgaaaagcctgatccaactcttgccgtcatttgcggtgc cgctacaaaggctttcttagcatctgaagaagcccgccacaagtatatcgaatccttacaaaagggacaagttcta tctaaagacctactgcaaactatgttccctgtagattttatccatgagggtaaaagatacaagttcaccgtagctaaa tccggtaatgaccgttacacattatttatcaatggttctaaatgtgatatcatactgcgtcaactatctgatggtggtct tttgattgccataggcggtaaatcgcataccatctattggaaagaagaagttgctgctacaagattatccgttgact ctatgactactttgttggaagttgaaaacgatccaacccagttgcgtactccatcccctggtaaattggttaaattctt ggtggaaaatggtgaacacattatcaagggccaaccatatgcagaaattgaagttatgaaaatgcaaatgccttt ggtttctcaagaaaatggtatcgtccagttattaaagcaacctggttctaccattgttgcaggtgatatcatggctatt atgactcttgacgatccatccaaggtcaagcacgctctaccatttgaaggtatgctgccagattttggttctccagtt atcgaaggaaccaaacctgcctataaattcaagtcattagtgtctactttggaaaacattttgaagggttatgacaa ccaagttattatgaacgcttccttgcaacaattgatagaggttttgagaaatccaaaactgccttactcagaatgga aactacacatctctgctttacattcaagattgcctgctaagctagatgaacaaatggaagagttagttgcacgttcttt gagacgtggtgctgttttcccagctagacaattaagtaaattgattgatatggccgtgaagaatcctgaatacaacc ccgacaaattgctgggcgccgtcgtggaaccattggcggatattgctcataagtactctaacgggttagaagccc atgaacattctatatttgtccatttcttggaagaatattacgaagttgaaaagttattcaatggtccaaatgttcgtgag gaaaatatcattctgaaattgcgtgatgaaaaccctaaagatctagataaagttgcgctaactgttttgtctcattcga aagtttcagcgaagaataacctgatcctagctatcttgaaacattatcaaccattgtgcaagttatcttctaaagtttct gccattttctctactcctctacaacatattgttgaactagaatctaaggctaccgctaaggtcgctctacaagcaaga gaaattttgattcaaggcgctttaccttcggtcaaggaaagaactgaacaaattgaacatatcttaaaatcctctgtt gtgaaggttgcctatggctcatccaatccaaagcgctctgaaccagatttgaatatcttgaaggacttgatcgattc taattacgttgtgttcgatgttttacttcaattcctaacccatcaagacccagttgtgactgctgcagctgctcaagtct atattcgtcgtgcttatcgtgcttacaccataggagatattagagttcacgaaggtgtcacagttccaattgttgaat ggaaattccaactaccttcagctgcgttctccacctttccaactgttaaatctaaaatgggtatgaacagggctgttt ctgtttcagatttgtcatatgttgcaaacagtcagtcatctccgttaagagaaggtattttgatggctgtggatcattta gatgatgttgatgaaattttgtcacaaagtttggaagttattcctcgtcaccaatcttcttctaacggacctgctcctga tcgttctggtagctccgcatcgttgagtaatgttgctaatgtttgtgttgcttctacagaaggtttcgaatctgaagag gaaattttggtaaggttgagagaaattttggatttgaataagcaggaattaatcaatgcttctatccgtcgtatcacat ttatgttcggttttaaagatgggtcttatccaaagtattatacttttaacggtccaaattataacgaaaatgaaacaatt cgtcacattgagccggctttggccttccaactggaattaggaagattgtccaacttcaacattaaaccaattttcact gataatagaaacatccatgtctacgaagctgttagtaagacttctccattggataagagattctttacaagaggtatt attagaacgggtcatatccgtgatgacatttctattcaagaatatctgacttctgaagctaacagattgatgagtgat atattggataatttagaagtcaccgacacttcaaattctgatttgaatcatatcttcatcaacttcattgcggtgtttgat atctctccagaagatgtcgaagccgccttcggtggtttcttagaaagatttggtaagagattgttgagattgcgtgtt tcttctgccgaaattagaatcatcatcaaagatcctcaaacaggtgccccagtaccattgcgtgccttgatcaataa cgtttctggttatgttatcaaaacagaaatgtacaccgaagtcaagaacgcaaaaggtgaatgggtatttaagtctt tgggtaaacctggatccatgcatttaagacctattgctactccttaccctgttaaggaatggttgcaaccaaaacgtt ataaggcacacttgatgggtaccacatatgtctatgacttcccagaattattccgccaagcatcgtcatcccaatgg aaaaatttctctgcagatgttaagttaacagatgatttctttatttccaacgagttgattgaagatgaaaacggcgaat taactgaggtggaaagagaacctggtgccaacgctattggtatggttgcctttaagattactgtaaagactcctga atatccaagaggccgtcaatttgttgttgttgctaacgatatcacattcaagatcggttcctttggtccacaagaaga cgaattcttcaataaggttactgaatatgctagaaagcgtggtatcccaagaatttacttggctgcaaactcaggtg ccagaattggtatggctgaagagattgttccactatttcaagttgcatggaatgatgctgccaatccggacaaggg cttccaatacttatacttaacaagtgaaggtatggaaactttaaagaaatttgacaaagaaaattctgttctcactgaa cgtactgttataaacggtgaagaaagatttgtcatcaagacaattattggttctgaagatgggttaggtgtcgaatgt ctacgtggatctggtttaattgctggtgcaacgtcaagggcttaccacgatatcttcactatcaccttagtcacttgta gatccgtcggtatcggtgcttatttggttcgtttgggtcaaagagctattcaggtcgaaggccagccaattatttttaa ctggtgctcctgcaatcaacaaaatgctgggtagagaagtttatacttctaacttacaattgggtggtactcaaatca

SEQ ID	SEQUENCE
	tgtataacaacggtgtttcacatttgactgctgttgacgatttagctggtgtagagaagattgttgaatggatgtcttat gttccagccaagcgtaatatgccagttcctatcttggaaactaaagacacatgggatagaccagttgatttcactcc aactaatgatgaaacttacgatgtaagatggatgattgaaggtcgtgagactgaaagtggatttgaatatggtttgtt tgataaagggtctttctttgaaactttgtcaggatgggccaaaggtgttgtcgttggtagagcccgtcttggtggtat tccactgggtgttattggtgttgaaacaagaactgtcgagaacttgattcctgctgatccagctaatccaaatagtg ctgaaacattaattcaagaacctggtcaagtttggcatccaaactccgccttcaagactgctcaagctatcaatgac tttaacaacggtgaacaattgccaatgatgattttggccaactggagaggtttctctggtggtcaacgtgatatgttc aacgaagtcttgaagtatggttcgtttattgttgacgcattggtggattacaaacaaccaattattatctatatcccac ctaccggtgaactaagaggtggttcatgggttgttgtcgatccaactatcaacgctgaccaaatggaaatgtatgc cgacgtcaacgctagagctggtgttttggaaccacaaggtatggttggtatcaagttccgtagagaaaaattgctg gacaccatgaacagattggatgacaagtacagagaattgagatctcaattatccaacaagagtttggctccagaa gtacatcagcaaatatccaagcaattagctgatcgtgagagagaactattgccaatttacggacaaatcagtcttc aatttgctgatttgcacgataggtcttcacgtatggtggccaagggtgttatttctaaggaactggaatggaccgag gcacgtcgtttcttcttctggagattgagaagaagattgaacgaagaatatttgattaaaaggttgagccatcaggt aggcgaagcatcaagattagaaaagatcgcaagaattagatcgtggtaccctgcttcagtggaccatgaagatg ataggcaagtcgcaacatggattgaagaaaactacaaaacttttggacgataaactaaagggtttgaaattagagt cattcgctcaagacttagctaaaaagatcagaagcgaccatgacaatgctattgatggattatctgaagttatcaag atgttatctaccgatgataaagaaaaattgttgaagactttgaaataa
SEQ ID NO:202 Truncated geranylgeranyl pyrophosphate synthase Ag_GPPS_Ntrunc	atgtcttttgacttcaataagtatatggattctaaggctatgaccgtcaacgaggctttgaataaagccatcccattg cgttacccacaaaagatctacgaatctatgagatattctttgttagctggtggtaagagagtccgtccagttttgtgt atcgccgcttgtgaattagtcggtggtactgaggagttagctattccaaccgcctgtgccatcgaaatgatccaca ccatgtctttgatgcacgatgatttgccatgtatcgacaacgatgacttgagacgtggtaaacctaccaatcataag attttcggtgaagatactgctgttactgccggtaacgctttacactcttacgccttcgaacatattgctgtttctacttc caagactgttggtgctgatagaattttgagaatggtttctgaattaggtcgtgctactggttccgaaggtgttatggg tggtcaaatggtcgatattgcttctgaaggtgacccttccattgatttgcaaactttagaatggatccacatccacaa gactgctatgttattagaatgttctgttgtctgtggtgccatcatcggtggtgcttctgaaattgttattgagagagcc agacgttatgctcgttgtgtcggtttattgtttcaagttgttgacgacattttagatgttaccaaatcttctgacgaattg ggtaaaactgctggtaaagatttaatctccgataaagccacctaccctaagttgatgggtttggagaaggccaaa gagttttccgatgaattattaaacagagctaaaggtgaattgtcttgcttcgatccagttaaggctgccccattgtta ggtttggctgactacgttgccttcagacaaaactaa
SEQ ID NO:203 Truncated geranylgeranyl pyrophosphate synthase Ag_GPPS_Ntrunc	MSFDFNKYMDSKAMTVNEALNKAIPLRYPQKIYESMRYSLLAGGKRV RPVLCIAACELVGGTEELAIPTACAIEMIHTMSLMHDDLPCIDNDDLRR GKPTNHKIFGEDTAVTAGNALHSYAFEHIAVSTSKTVGADRILRMVSE LGRATGSEGVMGGQMVDIASEGDPSIDLQTLEWIHIHKTAMLLECSVV CGAIIGGASEIVIERARRYARCVGLLFQVVDDILDVTKSSDELGKTAGK DLISDKATYPKLMGLEKAKEFSDELLNRAKGELSCFDPVKAAPLLGLA DYVAFRQN*
SEQ ID NO:204 Phosphomevalonate kinase Sc_ERG8 Saccharomyces cerevisiae	ttatttatcaagataagtttccggatctttttctttcctaacaccccagtcagcctgagttacatccagccattgaaccctt agaaaatcttttgtcatcagcggtttgagccctaagatcaacatcttgcttagcaatcactgcaatggcgtcataac caccagcaccaggtattaagcaagtaagaactccttttaaggtctggcaatcatccaataagctagtttgtacggg aggttcgatatcggcaccagattctttagttatttttctaaaggaacgtctaattgtggcaactgcatctctaacttctg tgatctcaggatacttttgacaggtacagtcattcctctcaagagactcaaatatctgatcgctgtaatcgtcatgag tctcgtgtaagcgatctagtttagatagtccatccataaatctagaatttgcatgatcgagttctgtatatattttcaag ctttccggcatatgcgaatcataccaatttttttaccttctggaccagttttactgtttctgaaccattcttaatatcgccc atccataaagttaatcccgaaggtaaatggttacttttaatcgttatattccagtcttcttcattaaccaaatgcgccag tttactgccgtaagtagcacttccaatatctggcaaattagagattaatgcgggtgggaatcttctatatctgatagat ccatatgctgccgccgctacatcaaacccgcttccaattttaccctgagcttgacaatgagcaacttgtgataaatt atgaataacttctctatatttgtctacattattttccaggtccgatacaaaaaaggaggccaaagctgtagttaaaact gtgactaaacctgccgaggagcccagccctgttttgggaacttcttcaattctgtgcgaatgaaaactcaatcttct gttgccacgatgttcggtaacgctgtcctcctgagaatggtaggcatcatcagagaaaatatcaataacgaacaa gtttctattgcagtagtcgtccatgttaggcttaaagtagctaaatacgttagcgataacttttttcaatgaaagggttct tagatccgcctatcgaaacaggaatgaagccagttttaggacttatatggtacagccactccccatctttaaattgtt

SEQ ID	SEQUENCE
	tacttttcacacgcacttcaaacttatcagactcttgcaatgaaccgtaaggatgggctacagcatgcattcttgcc gataatccgactacaaatgcttcatatttcggatctaaaactaaatatccaccagctagtaacgctttccctggggc actgaaggctctcaactctgacat
SEQ ID NO:205 Phosphomevalonate kinase Sc_ERG8 Saccharomyces cerevisiae	MSELRAFSAPGKALLAGGYLVLDPKYEAFVVGLSARMHAVAHPYGSL QESDKFEVRVKSKQFKDGEWLYHISPKTGFIPVSIGGSKNPFIEKVIANV FSYFKPNMDDYCNRNLFVIDIFSDDAYHSQEDSVTEHRGNRRLSFHSH RIEEVPKTGLGSSAGLVTVLTTALASFFVSDLENNVDKYREVIHNLSQV AHCQAQGKIGSGFDVAAAAYGSIRYRRFPPALISNLPDIGSATYGSKLA HLVNEEDWNITIKSNHLPSGLTLWMGDIKNGSETVKLVQKVKNWYDS HMPESLKIYTELDHANSRFMDGLSKLDRLHETHDDYSDQIFESLERND CTCQKYPEITEVRDAVATIRRSFRKITKESGADIEPPVQTSLLDDCQTLK GVLTCLIPGAGGYDAIAVIAKQDVDLRAQTADDKRFSKVQWLDVTQA DWGVRKEKDPETYLDK*
SEQ ID NO:206 Mevalonate kinase Erg12 Saccharomyces cerevisiae	atgtcattaccgttcttaacttctgcaccgggaaaggttattatttttggtgaacactctgctgtgtacaacaagcctg ccgtcgctgctagtgtgtctgcgttgagaacctacctgctaataagcgagtcatctgcaccagatactattgaattg gacttcccggacattagctttaatcataagtggtccatcaatgatttcaatgccatcaccgaggatcaagtaaactc ccaaaaattggccaaggctcaacaagccaccgatggcttgtctcaggaactcgttagtcttttggatccgttgttag ctcaactatccgaatccttccactaccatgcagcgttttgtttcctgtatatgtttgtttgcctatgcccccatgccaag aatattaagttttctttaaagtctactttacccatcggtgctgggttgggctcaagcgcctctatttctgtatcactggc cttagctatggcctacttgggggggttaataggatctaatgacttggaaaagctgtcagaaaacgataagcatata gtgaatcaatgggccttcataggtgaaaagtgtattcacggtacccccttcaggaatagataacgctgtggccactt atggtaatgccctgctatttgaaaaagactcacataatggaacaataaacacaaacaatttaagttcttagatgattt cccagccattccaatgatcctaacctatactagaattccaaggtctacaaaagatcttgttgctcgcgttcgtgtgtt ggtcaccgagaaatttcctgaagttatgaagccaattctagatgccatgggtgaatgtgccctacaaggcttagag atcatgactaagttaagtaaatgtaaaggcaccgatgacgaggctgtagaaactaataatgaactgtatgaacaa ctattggaattgataagaataaatcatggactgcttgtctcaatcggtgtttctcatcctggattagaacttattaaaaa tctgagcgatgatttgagaattggctccacaaaacttaccggtgctggtggcggcggttgctctttgactttgttac gaagagacattactcaagagcaaattgacagtttcaaaaagaaattgcaagatgattttagttacgagacatttgaa acagacttgggtgggactggctgctgtttgttaagcgcaaaaaatttgaataaagatcttaaaatcaaatccctagt attccaattatttgaaaataaaactaccacaaagcaacaaattgacgatctattattgccaggaaacacgaatttac catggacttcataa
SEQ ID NO:207 Mutated acetyl-CoA carboxylase (ACC1) (S659A, S1157A)	MSEESLFESSPQKMEYEITNYSERHTELPGHFIGLNTVDKLEESPLRDFV KSHGGHTVISKILIANNGIAAVKEIRSVRKWAYETFGDDRTVQFVAMA TPEDLEANAEYIRMADQYIEVPGGTNNNNYANVDLIVDIAERADVDA VWAGWGHASENPLLPEKLSQSKRKVIFIGPPGNAMRSLGDKISSTIVAQ SAKVPCIPWSGTGVDTVHVDEKTGLVSVDDDIYQKGCCTSPEDGLQK AKRIGFPVMIKASEGGGGKGIRQVEREEDFIALYHQAANEIPGSPIFIMK LAGRARHLEVQLLADQYGTNISLFGRDCSVQRRHQKIIEEAPVTIAKAE TFHEMEKAAVRLGKLVGYVSAGTVEYLYSHDDGKFYFLELNPRLQVE HPTTEMVSGVNLPAAQLQIAMGIPMHRISDIRTLYGMNPHSASEIDFEF KTQDATKKQRRPIPKGHCTACRITSEDPNDGFKPSGGTLHELNFRSSSN VWGYFSVGNNGNIHSFSDSQFGHIFAFGENRQASRKHMVVALKELSIR GDFRTTVEYLIKLLETEDFEDNTITTGWLDDLITHKMTAEKPDPTLAVI CGAATKAFLASEEARHKYIESLQKGQVLSKDLLQTMFPVDFIHEGKRY KFTVAKSGNDRYTLFINGSKCDIILRQLSDGGLLIAIGGKSHTIYWKEEV AATRLSVDSMTTLLEVENDPTQLRTPSPGKLVKFLVENGEHIIKGQPYA EIEVMKMQMPLVSQENGIVQLLKQPGSTIVAGDIMAIMTLDDPSKVKH ALPFEGMLPDFGSPVIEGTKPAYKFKSLVSTLENILKGYDNQVIMNASL QQLIEVLRNPKLPYSEWKLHISALHSRLPAKLDEQMEELVARSLRRGA VFPARQLSKLIDMAVKNPEYNPDKLLGAVVEPLADIAHKYSNGLEAHE HSIFVHFLEEYYEVEKLFNGPNVREENIILKLRDENPKDLDKVALTVLS HSKVSAKNNLILAILKHYQPLCKLSSKVSAIFSTPLQHIVELESKATAKV ALQAREILIQGALPSVKERTEQIEHILKSSVVKVAYGSSNPKRSEPDLNI

SEQ ID	SEQUENCE
	LKDLIDSNYVVFDVLLQFLTHQDPVVTAAAAQVYIRRAYRAYTIGDIR VHEGVTVPIVEWKFQLPSAAFSTFPTVKSKMGMNRAVSVSDLSYVAN SQSSPLREGILMAVDHLDDVDEILSQSLEVIPRHQSSSNGPAPDRSGSSA SLSNVANVCVASTEGFESEEEILVRLREILDLNKQELINASIRRITFMFGF KDGSYPKYYTFNGPNYNENETIRHIEPALAFQLELGRLSNFNIKPIFTDN RNIHVYEAVSKTSPLDKRFFTRGIIRTGHIRDDISIQEYLTSEANRLMSDI LDNLEVTDTSNSDLNHIFINFIAVFDISPEDVEAAFGGFLERFGKRLLRL RVSSAEIRIIIKDPQTGAPVPLRALINNVSGYVIKTEMYTEVKNAKGEW VFKSLGKPGSMHLRPIATPYPVKEWLQPKRYKAHLMGTTYVYDFPEL FRQASSSQWKNFSADVKLTDDFFISNELIEDENGELTEVEREPGANAIG MVAFKITVKTPEYPRGRQFVVVANDITFKIGSFGPQEDEFFNKVTEYAR KRGIPRIYLAANSGARIGMAEEIVPLFQVAWNDAANPDKGFQYLYLTS EGMETLKKFDKENSVLTERTVINGEERFVIKTIIGSEDGLGVECLRGSG LIAGATSRAYHDIFTITLVTCRSVGIGAYLVRLGQRAIQVEGQPIILTGA PAINKMLGREVYTSNLQLGGTQIMYNNGVSHLTAVDDLAGVEKIVEW MSYVPAKRNMPVPILETKDTWDRPVDFTPTNDETYDVRWMIEGRETE SGFEYGLFDKGSFFETLSGWAKGVVVGRARLGGIPLGVIGVETRTVEN LIPADPANPNSAETLIQEPGQVWHPNSAFKTAQAINDFNNGEQLPMMIL ANWRGFSGGQRDMFNEVLKYGSFIVDALVDYKQPIIIYIPPTGELRGGS WVVVDPTINADQMEMYADVNARAGVLEPQGMVGIKFRREKLLDTM NRLDDKYRELRSQLSNKSLAPEVHQQISKQLADRERELLPIYGQISLQF ADLHDRSSRMVAKGVISKELEWTEARRFFFWRLRRRLNEEYLIKRLSH QVGEASRLEKIARIRSWYPASVDHEDDRQVATWIEENYKTLDDKLKG LKLESFAQDLAKKIRSDHDNAIDGLSEVIKMLSTDDKEKLLKTLK*
SEQ ID NO:208 Truncated 3-hydroxy-3-methyl-glutaryl-CoA reductase (Sc_tHMG1)	MQLVKTEVTKKSFTAPVQKASTPVLTNKTVISGSKVKSLSSAQSSSSGP SSSSEEDDSRDIESLDKKIRPLEELEALLSSGNTKQLKNKEVAALVIHGK LPLYALEKKLGDTTRAVAVRRKALSILAEAPVLASDRLPYKNYDYDR VFGACCENVIGYMPLPVGVIGPLVIDGTSYHIPMATTEGCLVASAMRG CKAINAGGGATTVLTKDGMTRGPVVRFPTLKRSGACKIWLDSEEGQN AIKKAFNSTSRFARLQHIQTCLAGDLLFMRFRTTTGDAMGMNMISKGV EYSLKQMVEEYGWEDMEVVSVSGNYCTDKKPAAINWIEGRGKSVVA EATIPGDVVRKVLKSDVSALVELNIAKNLVGSAMAGSVGGFNAHAAN LVTAVFLALGQDPAQNVESSNCITLMKEVDGDLRISVSMPSIEVGTIGG GTVLEPQGAMLDLLGVRGPHATAPGTNARQLARIVACAVLAGELSLC AALAAGHLVQSHMTHNRKPAEPTKPNNLDATDINRLKDGSVTCIKS*
SEQ ID NO:209 Erg10p: acetoacetyl CoA thiolase [Saccharomyces cerevisiae].	atgtctcagaacgtttacattgtatcgactgccagaaccccaattggttcattccagggttctctatcctccaagaca gcagtggaattgggtgctgttgctttaaaaggcgccttggctaaggttccagaattggatgcatccaaggattttga cgaaattatttttggtaacgttctttctgccaatttgggccaagctccggccagacaagttgctttggctgccggtttg agtaatcatatcgttgcaagcacagttaacaaggtctgtgcatccgctatgaaggcaatcattttgggtgctcaatc catcaaatgtggtaatgctgatgttgtcgtagctggtggttgtgaatctatgactaacgcaccatactacatgccag cagcccgtgcgggtgccaaatttggccaaactgttcttgttgatggtgtcgaaagagatgggttgaacgatgcgt acgatggtctagccatgggtgtacacgcagaaaagtgtgcccgtgattgggatattactagagaacaacaagac aattttgccatcgaatcctaccaaaaatctcaaaaatctcaaaaggaaggtaaattcgacaatgaaattgtacctgtt accattaagggatttagaggtaagcctgatactcaagtcacgaaggacgaggaacctgctagattacacgttgaa aaattgagatctgcaaggactgttttccaaaaagaaaacggtactgttactgccgctaacgcttctccaatcaacg atggtgctgcagccgtcatcttggtttccgaaaaagttttgaaggaaaagaatttgaagcctttggctattatcaaag gttggggtgaggccgctcatcaaccagctgattttacatgggctccatctcttgcagttccaaaggctttgaaacat gctggcatcgaagacatcaattctgttgattactttgaattcaatgaagccttttcggttgtcggtttggtgaacacta agattttgaagctagacccatctaaggttaatgtatatggtggtgctgttgctctaggtcacccattgggttgttctgg tgctagagtggttgttacactgctatccatcttacagcaagaaggaggtaagatcggtgttgccgccatttgtaatg gtggtggtggtgcttcctctattgtcattgaaaagatatga
SEQ ID NO:210 Artificial truncated	atgtcttacgttgtcaagggtatgatctctattgcttgtggtttgttcggtagagaattgtttaacaacagacacttgttc tcttggggtttgatgtggaaagctttcttcgctttggtcccaattttgtctttcaatttcttcgccgccatcatgaaccaa

SEQ ID	SEQUENCE
geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t112	atctacgatgttgatatcgaccgtatcaacaagccagacttacctttagtttccggtgaaatgtccattgaaactgctt ggatcttgtctatcattgttgccttgactggtttaattgttactattaagttgaagtccgctccattgtttgtcttcatctac atcttcggtatcttcgctggtttcgcttactccgtcccacctattagatggaaacaatatccttttaccaatttcttgatc actatttcctctcatgttggtttggctttcacttcttactctgccaccacttctgctttaggtttgcctttcgtttggcgtcc tgccttctctttcattattgctttcatgactgtcatgggtatgactattgcctttgctaaagacatttctgatatcgaaggt gatgctaagtacggtgtctctaccgttgctaccaagttaggtgctagaaatatgacttttgttgtttctggtgtcttatt gttgaactacttggtttctatctctattggtatcatttggccacaagttttcaagtctaacattatgatcttgtctcatgct attttggctttctgtttgatctttcaaactcgtgaattagccttagccaattatgcctctgccccatcccgtcaatttttcg aattcatctggttgttatactatgccgaatacttcgtttacgtcttcatttaa
SEQ ID NO:211 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t112	MSYVVKGMISIACGLFGRELFNNRHLFSWGLMWKAFFALVPILSFNFF AAIMNQIYDVDIDRINKPDLPLVSGEMSIETAWILSIIVALTGLIVTIKLK SAPLFVFIYIFGIFAGFAYSVPPIRWKQYPFTNFLITISSHVGLAFTSYSAT TSALGLPFVWRPAFSFIIAFMTVMGMTIAFAKDISDIEGDAKYGVSTVA TKLGARNMTFVVSGVLLLNYLVSISIGIIWPQVFKSNIMILSHAILAFCLI FQTRELALANYASAPSRQFFEFIWLLYYAEYFVYVFI
SEQ ID NO:212 Artificial truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t131 nucleotide sequence	atgtctaacaacagacacttgttctcttggggtttgatgtggaaagctttcttcgctttggtcccaattttgtctttcaatt tcttcgccgccatcatgaaccaaatctacgatgttgatatcgaccgtatcaacaagccagacttacctttagtttccg gtgaaatgtccattgaaactgcttggatcttgtctatcattgttgccttgactggtttaattgttactattaagttgaagtc cgctccattgtttgtcttcatctacatcttcggtatcttcgctggtttcgcttactccgtcccacctattagatggaaaca atatccttttaccaatttcttgatcactatttcctctcatgttggtttggctttcacttcttactctgccaccacttctgcttta ggtttgcctttcgtttggcgtcctgccttctctttcattattgctttcatgactgtcatgggtatgactattgcctttgctaa agacatttctgatatcgaaggtgatgctaagtacggtgtctctaccgttgctaccaagttaggtgctagaaatatga cttttgttgtttctggtgtcttattgttgaactacttggtttctatctctattggtatcatttggccacaagttttcaagtcta acattatgatcttgtctcatgctattttggctttctgtttgatctttcaaactcgtgaattagccttagccaattatgcctct gccccatcccgtcaatttttcgaattcatctggttgttatactatgccgaatacttcgtttacgtcttcatttaa
SEQ ID NO:213 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t131	MSNNRHLFSWGLMWKAFFALVPILSFNFFAAIMNQIYDVDIDRINKPD LPLVSGEMSIETAWILSIIVALTGLIVTIKLKSAPLFVFIYIFGIFAGFAYS VPPIRWKQYPFTNFLITISSHVGLAFTSYSATTSALGLPFVWRPAFSFIIA FMTVMGMTIAFAKDISDIEGDAKYGVSTVATKLGARNMTFVVSGVLL LNYLVSISIGIIWPQVFKSNIMILSHAILAFCLIFQTRELALANYASAPSR QFFEFIWLLYYAEYFVYVFI
SEQ ID NO:214 Artificial truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t142 nucleotide sequence	atgtcttggaaagctttcttcgctttggtcccaattttgtctttcaatttcttcgccgccatcatgaaccaaatctacgat gttgatatcgaccgtatcaacaagccagacttacctttagtttccggtgaaatgtccattgaaactgcttggatcttgt ctatcattgttgccttgactggtttaattgttactattaagttgaagtccgctccattgtttgtcttcatctacatcttcggt atcttcgctggtttcgcttactccgtcccacctattagatggaaacaatatccttttaccaatttcttgatcactatttcct ctcatgttggtttggctttcacttcttactctgccaccacttctgctttaggtttgcctttcgtttggcgtcctgccttctct ttcattattgctttcatgactgtcatgggtatgactattgcctttgctaaagacatttctgatatcgaaggtgatgctaa gtacggtgtctctaccgttgctaccaagttaggtgctagaaatatgacttttgttgtttctggtgtcttattgttgaacta cttggtttctatctctattggtatcatttggccacaagttttcaagtctaacattatgatcttgtctcatgctatttggcttt ctgtttgatctttcaaactcgtgaattagccttagccaattatgcctctgccccatcccgtcaatttttcgaattcatctg gttgttatactatgccgaatacttcgtttacgtcttcatttaa
SEQ ID NO:215 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t142	MSWKAFFALVPILSFNFFAAIMNQIYDVDIDRINKPDLPLVSGEMSIETA WILSIIVALTGLIVTIKLKSAPLFVFIYIFGIFAGFAYSVPPIRWKQYPFTN FLITISSHVGLAFTSYSATTSALGLPFVWRPAFSFIIAFMTVMGMTIAFA KDISDIEGDAKYGVSTVATKLGARNMTFVVSGVLLLNYLVSISIGIIWP QVFKSNIMILSHAILAFCLIFQTRELALANYASAPSRQFFEFIWLLYYAE YFVYVFI

SEQ ID	SEQUENCE
SEQ ID NO:216 Artificial truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t166 nucleotide sequence	Atgtctgatgttgatatcgaccgtatcaacaagccagacttacctttagtttccggtgaaatgtccattgaaactgct tggatcttgtctatcattgttgccttgactggtttaattgttactattaagttgaagtccgctccattgtttgtcttcatcta catcttcggtatcttcgctggtttcgcttactccgtcccacctattagatggaaacaatatccttttaccaatttcttgat cactatttcctctcatgttggtttggctttcacttcttactctgccaccacttctgctttaggtttgcctttcgtttggcgtc ctgccttctctttcattattgctttcatgactgtcatgggtatgactattgcctttgctaaagacatttctgatatcgaag gtgatgctaagtacggtgtctctaccgttgctaccaagttaggtgctagaaatatgacttttgttgtttctggtgtctta ttgttgaactacttggtttctatctctattggtatcatttggccacaagttttcaagtctaacattatgatcttgtctcatgc tattttggctttctgtttgatctttcaaactcgtgaattagccttagccaattatgcctctgccccatcccgtcaatttttc gaattcatctggttgttatactatgccgaatacttcgtttacgtcttcatttaa
SEQ ID NO:217 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t166	MSDVDIDRINKPDLPLVSGEMSIETAWILSIIVALTGLIVTIKLKSAPLFV FIYIFGIFAGFAYSVPPIRWKQYPFTNFLITISSHVGLAFTSYSATTSALGL PFVWRPAFSFIIAFMTVMGMTIAFAKDISDIEGDAKYGVSTVATKLGAR NMTFVVSGVLLLNYLVSISIGIIWPQVFKSNIMILSHAILAFCLIFQTREL ALANYASAPSRQFFEFIWLLYYAEYFVYVFI
SEQ ID NO:218 Artificial truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t186 nucleotide sequence	atgtctattgaaactgcttggatcttgtctatcattgttgccttgactggtttaattgttactattaagttgaagtccgctc cattgtttgtcttcatctacatcttcggtatcttcgctggtttcgcttactccgtcccacctattagatggaaacaatatc cttttaccaatttcttgatcactatttcctctcatgttggtttggctttcacttcttactctgccaccacttctgctttaggttt gcctttcgtttggcgtcctgccttctctttcattattgctttcatgactgtcatgggtatgactattgcctttgctaaagac atttctgatatcgaaggtgatgctaagtacggtgtctctaccgttgctaccaagttaggtgctagaaatatgacttttg ttgtttctggtgtcttattgttgaactacttggtttctatctctattggtatcatttggccacaagttttcaagtctaacatta tgatcttgtctcatgctatttttggctttctgtttgatctttcaaactcgtgaattagccttagccaattatgcctctgcccc atcccgtcaatttttcgaattcatctggttgttatactatgccgaatacttcgtttacgtcttcatttaa
SEQ ID NO:219 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t186	MSIETAWILSIIVALTGLIVTIKLKSAPLFVFIYIFGIFAGFAYSVPPIRWK QYPFTNFLITISSHVGLAFTSYSATTSALGLPFVWRPAFSFIIAFMTVMG MTIAFAKDISDIEGDAKYGVSTVATKLGARNMTFVVSGVLLLNYLVSI SIGIIWPQVFKSNIMILSHAILAFCLIFQTRELALANYASAPSRQFFEFIWL LYYAEYFVYVFI
SEQ ID NO:220 Artificial geranyl pyrophosphate olivetolic acid geranyltransferase CsGOT (CsPT1) nucleotide sequence	atgggtttatcttccgtttgtacttttttctttccaaactaactaccacactttgttaaatccacacaacaacaaccctaaa acctccttgttatgttacagacacccaaagacccctattaaatactcctacaacaacttcccatccaaacactgctc cactaagtcctttcacttgcaaaacaagtgttctgaatccttgtccattgccaagaactctattcgtgccgctactact aaccaaactgagccacctgaatccgataaccactccgtcgccaccaagatcttgaattttggtaaagcttgctgga aattgcaaagaccatacactattattgctttcacttcctgtgcttgtggtttattcggtaaggaattattgcataacacc aacttgatttcttggtccttaatgttcaaagccttcttctttttagttgccattttatgtattgcttctttcactactactattaa tcaaatttacgatttgcacattgacagaatcaataagcctgacttgccattagcttccggtgaaatttctgttaacact gcttggatcatgtccatcattgtcgctttgttcggtttaattatcaccatcaaaatgaagggtggtcctttgtacatcttc ggttattgcttcggtattttcggtggtattgtctactctgtcccaccattcagatggaagcaaaacccatccactgcct ttttgttgaatttcttggctcacatcattaccaattttacttttctactatgcctcccgtgctgctttaggtttgcctttgagt tacgtccatccttcacttttttattggcttttatgaagtccatgggttctgctttagccttaattaaggacgcctctgacg ttgaaggtgatactaagttcggtatctctactttagcctctaagtacggttctcgtaacttgaccttgttctgttctggta ttgtcttgttgtcttacgtcgccgctattttggccggtatcatctggccacaagctttcaactctaacgttatgttgttgt ctcatgctatcttagctttctggttgatcttacaaaccagagacttcgctttgactaactacgacccagaagccggtc gtagattctacgaattcatgtggaaattgtactacgccgagtacttggtctacgttttcatttag
SEQ ID NO:221	atgtctgctggctctgaccaaattgaaggttccccgcatcacgaatcagataatagtattgccacaaagatcttaaa

SEQ ID	SEQUENCE
Artificial truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4t nucleotide sequence	ctttgggcatacatgttggaaattacaaaggccctacgtcgtcaaaggaatgataagcatcgcttgcggtctgttc ggaagggaattatttaacaataggcatctattcagctgggggttaatgtggaaagctttcttcgcgttagtgccaat cctaagctttaacttttcgccgccatcatgaaccagatttatgatgttgatatcgacaggataaataagccagatctt ccattggtatccggtgaaatgtcaatagaaactgcatggatattatctattatcgttgcgctgaccggactgatagta acaatcaaattgaaatctgcacccctgtttgtttttatatatatatttggtattttcgctggattcgcttactcagtgccac ctatcaggtggaagcagtacccattcacgaattttctgatcacgatctctagccacgtcgggttagcgttcacatctt actctgcaaccacgagtgccttggggcttcctttcgtctggcgtccagcttttagttttatcattgcctttatgaccgta atgggaatgacgatcgcattcgcaaaggacatttctgacatagaggggatgcaaaatacggtgtctccactgtg gcgacaaaattaggagctaggaatatgactttcgtggtgtccggtgtattattactaaattatctggtatctataagta tcggcatcatatggccgcaagtgtttaaatccaacattatgatactgagtcatgctatttttggctttttgtctgatttttc agacgcgtgagttggcgcttgcaaactatgcctctgcgcccagcaggcagttttttgaattcatatggttattgtact atgccgagtatttcgtctacgtatttatttaa
SEQ ID NO:222 Artificial truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsGOT_t75 (CsPT1_t75) nucleotide sequence	atgtctgccgctactactaaccaaactgagccacctgaatccgataaccactccgtcgccaccaagatcttgaatt ttggtaaagcttgctggaaattgcaaagaccatacactattattgctttcacttcctgtgcttgtggtttattcggtaag gaattattgcataacaccaacttgatttcttggtccttaatgttcaaagccttcttctttttagttgccatttttatgtattgct tctttcactactactattaatcaaatttacgatttgcacattgacagaatcaataagcctgacttgccattagcttccgg tgaaatttctgttaacactgcttggatcatgtccatcattgtcgctttgttcggtttaattatcaccatcaaaatgaagg gtggtcctttgtacatcttcggttattgcttcggtattttcggtggtattgtctactctgtcccaccattcagatggaagc aaaacccatccactgccttttgttgaatttcttggctcacatcattaccaattttactttctactatgcctcccgtgctg ctttaggtttgccttttgagttacgtccatccttcacttttttattggcttttatgaagtccatgggttctgctttagccttaa ttaaggacgcctctgacgttgaaggtgatactaagttcggtatctctactttagcctctaagtacggttctcgtaactt gaccttgttctgttctggtattgtcttgttgtcttacgtcgccgctattttggccggtatcatctggccacaagctttcaa ctctaacgttatgttgttgtctcatgctatcttagctttctggttgatcttacaaaccagagacttcgctttgactaacta cgacccagaagccggtcgtagattctacgaattcatgtggaaattgtactacgccgagtacttggtctacgttttca tttag
SEQ ID NO:223 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsGOT_t75 (CsPT1_t75)	MSAATTNQTEPPESDNHSVATKILNFGKACWKLQRPYTIIAFTSCACGL FGKELLHNTNLISWSLMFKAFFFLVAILCIASFTTTINQIYDLHIDRINKP DLPLASGEISVNTAWIMSIIVALFGLIITIKMKGGPLYIFGYCFGIFGGIV YSVPPFRWKQNPSTAFLLNFLAHIITNFTFYYASRAALGLPFELRPSFTF LLAFMKSMGSALALIKDASDVEGDTKFGISTLASKYGSRNLTLFCSGIV LLSYVAAILAGIIWPQAFNSNVMLLSHAILAFWLILQTRDFALTNYDPE AGRRFYEFMWKLYYAEYLVYVFI
SEQ ID NO:224 Artificial truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t76 (CsPT4t) nucleotide sequence	Atgtccgccggttctgatcaaatcgaaggttcccctcatcatgagtccgataactccattgctactaaaattttaaat ttcggtcatacttgttggaagttgcaacgtccttacgttgtcaagggtatgatctctattgcttgtggtttgttcggtag agaattgtttaacaacagacacttgttctcttggggtttgatgtggaaagctttcttcgctttggtcccaattttgtctttc aatttcttcgccgccatcatgaaccaaatctacgatgttgatatcgaccgtatcaacaagccagacttacctttagttt ccggtgaaatgtccattgaaactgcttggatcttgtctatcattgttgccttgactggtttaattgttactattaagttga agtccgctccattgtttgtcttcatctacatcttcggtatcttcgctggtttcgcttactccgtcccacctattagatgga aacaatatccttttaccaatttcttgatcactatttcctctcatgttggtttggctttcacttcttactctgccaccacttct gctttaggtttgcctttcgtttggcgtcctgccttctctttcattattgctttcatgactgtcatgggtatgactattgcctt tgctaaagacatttctgatatcgaaggtgatgctaagtacggtgtctctaccgttgctaccaagttaggtgctagaa atatgacttttgttgtttctggtgtcttattgttgaactacttggtttctatctctattggtatcatttggccacaagttttca agtctaacattatgatcttgtctcatgctattttggctttctgtttgatctttcaaactcgtgaattagccttagccaattat gcctctgccccatcccgtcaatttttcgaattcatctggttgttatactatgccgaatacttcgtttacgtcttcatttaa
SEQ ID NO:225 Geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4	atgggactctcattagtttgtacctttcatttcaaactaattatcatactttattaaaccctcataataagaatcccaaa aactcattattatcttatcaacaccccaaaacaccaataattaaatcctcttatgataatttttccctctaaatattgcttaa ccaagaactttcatttacttggactcaattcacacaacagaataagctcacaatcaaggtccattagggcaggtag cgatcaaattgaaggttctcctcatcatgaatctgataattcaatagcaactaaaattttaaatttttggacatacttgtt ggaaacttcaaagaccatatgtagtaaaagggatgatttcaatcgcttgtggtttgtttgggagagagttgttcaata acagacatttattcagttggggtttgatgtggaaggcattctttgctttggtgcctatattgtccttcaatttctttgcag caatcatgaatcaaatttacgatgtggacatcgacaggataaacaagcctgatctaccactagtttcaggggaaat

SEQ ID	SEQUENCE
Cannabis sativa	gtcaattgaaacagcttggattttgagcataattgtggcactaactgggttgatagtaactataaaattgaaatctgc accacttttgttttcatttacatttttggtatatttgctgggtttgcctattctgttccaccaattagatggaagcaatatc cttttaccaatttttctaattaccatatcgagtcatgtgggcttagctttcacatcatattctgcaaccacatcagctcttg gtttaccatttgtgtggaggcctgcttttagtttcatcatagcattcatgacagttatgggtatgactattgcttttgcca aagatatttcagatattgaaggcgacgccaaatatggggtatcaactgttgcaaccaaattaggtgctaggaacat gacatttgttgtttctggagttcttcttctaaactacttggtttctatatctattgggataatttggcctcaggttttcaaga gtaacataatgatactttctcatgcaatcttagcattttgcttaatcttccagactcgtgagcttgctctagcaaattac gcctcggcgccaagcagacaattcttcgagtttatctggttgctatattatgctgaatactttgtatatgtatttatataa

EXAMPLES

[00920] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the present disclosure, and are not intended to limit the scope of what the inventors regard as their disclosure nor are they intended to represent that the experiments below are all or the only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (e.g. amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Celsius, and pressure is at or near atmospheric. Standard abbreviations may be used, e.g., bp, base pair(s); kb, kilobase(s); s or sec, second(s); min, minute(s); h or hr, hour(s); aa, amino acid(s); kb, kilobase(s); bp, base pair(s); nt, nucleotide(s); and the like.

General Methods of the Examples

Yeast transformation methods

[00921] Each DNA construct comprising one or more heterologous nucleic acids disclosed herein (e.g., constructs detailed in Table 11) was integrated into Saccharomyces cerevisiae (CEN.PK2) with standard molecular biology techniques in an optimized lithium acetate (LiAc) transformation. Briefly, cells were grown overnight in yeast extract peptone dextrose (YPD) media at 30 °C with shaking (200 rpm), diluted to an OD600 of 0.1 in 100 mL YPD, and grown to an OD600 of 0.6 - 0.8. For each transformation, 5 mL of culture was harvested by centrifugation, washed in 5 mL of sterile water, spun down again, resuspended in 1 mL of 100 mM LiAc, and transferred to a microcentrifuge tube. Cells were spun down (13,000 x g) for 30 seconds, the supernatant was removed, and the cells were resuspended in a transformation mix consisting of 240 µL 50% PEG, 36 µL 1M LiAc, 10 µL boiled salmon sperm DNA, and 74 µL of donor DNA. Following a heat shock at 42 °C for 40 minutes, cells were recovered overnight in YPD media before plating on selective media. DNA integration was confirmed by colony PCR with primers specific to the integrations.

Yeast culturing conditions

[00922] Yeast colonies verified to contain the expected DNA assembly comprising one or more heterologous nucleic acids disclosed herein, genetically modified host cells, were picked into 96-well microtiter plates containing 360 µL of YPD (10 g/L yeast extract, 20 g/L Bacto peptone, 20 g/L dextrose (glucose)) and sealed with a breathable film seal. Cells were cultured at 30 °C in a high capacity microtiter plate incubator shaking at 1000 rpm and 80% humidity for 3 days until the cultures reached carbon exhaustion. The growth-saturated cultures were subcultured into fresh plates containing YPGAL and either olivetolic acid or hexanoic acid, or an olivetolic acid derivative or a carboxylic acid other than hexanoic acid (10 g/L yeast extract, 20 g/L Bacto peptone, 20 g/L galactose, 1 g/L glucose and either 1 mM olivetolic acid or 2 mM hexanoic acid, or 1 mM of an olivetolic acid derivative or 2 mM of a carboxylic acid other than hexanoic acid), by taking 14.4 µL from the saturated cultures and diluting into 360 µL of fresh media and sealed with a breathable film seal. Genetically modified host cells in the production media were cultured at 30 °C in a high capacity microtiter plate shaker at 1000 rpm and 80% humidity for an additional 3 days prior to extraction and analysis. Upon completion, 100 µL of whole cell broth was diluted into 900 µL of methanol, sealed with a foil seal, and shaken at 1500 rpm for 60 seconds to extract the cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives. After shaking, the plate was centrifuged at 1000 x g for 60 seconds to remove any solids. After centrifugation, 12 µL of supernatant was transferred to a fresh assay plate containing 228 µL of methanol, sealed with a foil seal, shaken for 60 seconds at 900 rpm, and analyzed by LC-MS.

Analytical methods

[00923] Samples were analyzed by LC-MS mass spectrometer (Agilent 6470) using an Agilent Poroshell 120 Phenyl Hexyl 2.1 x 50 mm, 1.9 µm analytical column with the following gradient (Mobile Phase A: LC-MS grade water with 0.1% formic acid; Mobile Phase B: LC-MS grade acetonitrile with 0.1% formic acid):

Time (minutes)	**%B**
0	40
0.1	40
0.6	60
1	65
1.01	95
2.01	95
2.02	40

Time (minutes)	%B
2.5	40

[00924] The mass spectrometer was operated in negative ion multiple reaction monitoring mode. Each cannabinoid, cannabinoid precursor, cannabinoid derivative, or cannabinoid precursor derivative was identified by retention time, determined from an authentic standard, and multiple reaction monitoring (MRM) transition:

Compound Name	Q1 Mass (Da)	Q3 Mass (Da)
CBGA	359.2	341.1
CBGA	359.2	315.2
CBDA	357.2	339.1
CBDA	357.2	245.1
THCA	357.0	313.0

Recovery and purifications

[00925] Whole-cell broth from cultures comprising genetically modified host cells of the disclosure are extracted with a suitable organic solvent to afford cannabinoids, cannabinoid precursors, cannabinoid derivatives, or cannabinoid precursor derivatives. Suitable organic solvents include, but are not limited to, hexane, heptane, ethyl acetate, petroleum ether, and di-ethyl ether, chloroform, and ethyl acetate. The suitable organic solvent, such as hexane, is added to the whole-cell broth from fermentations comprising genetically modified host cells of the disclosure at a 10:1 ratio (10 parts whole-cell broth – 1 part organic solvent) and stirred for 30 minutes. The organic fraction is separated and extracted twice with an equal volume of acidic water (pH 2.5). The organic layer is then separated and dried in a concentrator (rotary evaporator or thin film evaporator under reduced pressure) to obtain crude cannabinoid, cannabinoid precursor, cannabinoid derivative, or cannabinoid precursor derivative crystals. The crude crystals may then be heated to 105 °C for 15 minutes followed by 145 °C for 55 minutes to decarboxylate a crude cannabinoid or cannabinoid derivative. The crude crystalline product is re-dissolved and recrystallized in a suitable solvent (e.g., n-pentane) and filtered through a 1 µm filter to remove any insoluble material. The solvent is then removed e.g. by rotary evaporation, to produce pure crystalline product.

In vitro enzyme assay and cell-free production of cannabinoids or cannabinoid derivatives

[00926] In some embodiments, genetically modified host cells, e.g., genetically modified yeast cells, verified to comprise one or more heterologous nucleic acids encoding a

GOT polypeptide that catalyzes production of cannabigerolic acid from geranyl pyrophosphate and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82 or a polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110, are cultured in 96-well microtiter plates containing 360 µL of YPD (10 g/L yeast extract, 20 g/L Bacto peptone, 20 g/L dextrose (glucose)) and sealed with a breathable film seal. Cells are then cultured at 30 °C in a high capacity microtiter plate incubator shaking at 1000 rpm and 80% humidity for 3 days until the cultures reach carbon exhaustion. The growth-saturated cultures are then subcultured into 200 mL of YPGAL media to an OD600 of 0.2 and incubated with shaking for 20 hours at 30 °C. Cells are then harvested by centrifugation at 3000 x g for 5 minutes at 4 °C. Harvested cells are then resuspended in 50 mL buffer (50 mM Tris-HCl, 1 mM EDTA, 0.1 M KCl, pH 7.4, 125 units Benzonase) and then lysed (Emulsiflex C3, Avestin, INC., 60 bar, 10 min). Cells debris is removed by centrifugation (10,000 × g, 10 min, 4 °C). Subsequently, the supernatant is then subjected to ultracentrifugation (150,000 × g, 1 h, 4 °C, Beckman Coulter L-90K, TI-70). The resulting membrane fractions of the GOT polypeptide that catalyzes production of cannabigerolic acid from geranyl pyrophosphate and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82 or the polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110 are then resuspended in 3.3 mL buffer (10 mM Tris-HCl, 10 mM $MgCl_2$, pH 8.0, 10% glycerol) and solubilized with a tissue grinder. Then, 0.02% (v/v) of the respective membrane preparations are then dissolved in reaction buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, pH 8.5) and substrate (500 µM olivetolic acid, 500 µM GPP) to a total volume of 50 µL and incubated for 1 hour at 30 °C. Assays are then extracted by adding two reaction volumes of ethyl acetate followed by vortexing and centrifugation. The organic layer is evaporated for 30 minutes, resuspended in acetonitrile/H_2O/formic acid (80:20:0.05%) and filtered with Ultrafree® -MC columns (0.22 µm pore size, PVDF membrane material). Cannabinoids or cannabinoid derivatives are then detected via LC-MS and/or recovered and purified.

Yeast cultivation in a bioreactor

[00927] Single yeast colonies comprising genetically modified host cells disclosed herein are grown in 15 mL of Verduyn medium (originally described by Verduyn et al, Yeast 8(7): 501-17) with 50 mM succinate (pH 5.0) and 2% glucose in a 125 mL flask at 30 °C, with shaking at 200 rpm to an OD600 between 4 to 9. Glycerol is then added to the culture to a concentration of 20% and 1 mL vials of the genetically modified host cell suspension are stored at −80 °C. One to two vials of genetically modified host cells are thawed and grown in Verduyn medium with 50 mM succinate (pH 5.0) and 4% sucrose for 24 hours, then sub-cultured to an OD600 reading of 0.1 in the same media. After 24 hours of growth at 30 °C with shaking, 65 mL of culture is used to inoculate a 1.3-liter fermenter (Eppendorf DASGIP Bioreactor) with 585 mL of Verduyn fermentation media containing 20 g/L galactose supplemented with hexanoic acid (2 mM), a carboxylic acid other than hexanoic acid (2 mM), olivetolic acid (1 mM), or an olivetolic acid derivative (1 mM). A poly-alpha-olefin may be added to the fermenter as an extractive agent. The fermenter is maintained at 30 °C and pH 5.0 with addition of NH_4OH. In an initial batch phase, the fermenter is aerated at 0.5 volume per volume per minute air (VVM) and agitation ramped to maintain 30% dissolved oxygen. After the initial sugar is consumed, the rise in dissolved oxygen triggers feeding of galactose + hexanoic acid (800 g galactose per liter + 9.28 g hexanoic acid per liter) at 10 g galactose per liter per hour in pulses of 10 g galactose per liter doses (alternatively, rather than feeding the genetically modified host cells disclosed herein hexanoic acid, olivetolic acid, an olivetolic acid derivative, or a carboxylic acid other than hexanoic acid is fed to the genetically modified host cells).

[00928] Between pulses, the feed rate is lowered to 5 g galactose per liter per hour. Upon a 10% rise in dissolved oxygen, the feed rate is resumed at 10 g L^{-1} $hour^{-1}$. As genetically modified host cell density increases, dissolved oxygen is allowed to reach 0%, and the pulse dose is increased to 50 g galacose per liter. Oxygen transfer rate is maintained at rates representative of full-scale conditions of 100 mM per liter per hour by adjusting agitation as volume increased. Feed rate is adjusted dynamically to meet demand using an algorithm that alternates between a high feed rate and low feed rate. During the low feed rate, genetically modified host cells should consume galactose and hexanoic acid, or, alternatively, olivetolic acid, an olivetolic acid derivative, or a carboxylic acid other than hexanoic acid, and any overflow metabolites accumulated during the high feed rate. A rise in dissolved oxygen triggers the high feed rate to resume. The length of time spent in the low feed rate reflects the extent to which genetically modified host cells are over- or under-fed in

the prior high feed rate pulse; this information is then monitored and used to tune the high feed rate up or down, keeping the low feed rate within a defined range.

[00929] Over time, the feed rate matches sugar and hexanoic acid, or, alternatively, olivetolic acid, an olivetolic acid derivative, or a carboxylic acid other than hexanoic acid, demand from genetically modified host cells. This algorithm ensures minimal net accumulation of fermentation products other than cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives; biomass; and CO_2. In some embodiments, the process continues for 5 to 14 days. In certain such embodiments, accumulated broth is removed daily and assayed for biomass and cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative concentration. A concentrated solution of $NH_4H_2PO_4$, trace metals and vitamins are added periodically to maintain steady state concentrations.

Example 1 – Synthesis of Olivetolic Acid or Derivatives Thereof

[00930] The cannabinoid pathway is composed of four biosynthetic steps using the precursors hexanoyl-CoA, malonyl-CoA, and geranyl pyrophosphate (FIG. 1, Box 4). *Saccharomyces cerevisiae* has previously been engineered to produce high levels of malonyl-CoA. To increase geranyl pyrophosphate supply, the engineering strategy as outlined in FIG. 1, Box 3, was carried out. In addition, a heterologous nucleic acid encoding an ACC1 polypeptide was overexpressed to increase flux to malonyl-CoA.

[00931] FIG. 1: Diagram illustrating biosynthetic pathways for converting sugar or hexanoate to the cannabinoids Δ^9-THC and CBD

[00932] To date, engineering biosynthesis of the precursor hexanoyl-CoA in *Saccharomyces cerevisiae* has not been described. Strategies for hexanoyl-CoA biosynthesis in *Saccharomyces cerevisiae*, as outlined in FIG. 1, were conceived: Pathway 1a) hexanoyl-CoA synthetase polypeptide from *C. sativa* converts hexanoate to hexanoyl-CoA (FIG. 1, box 1a). A heterologous nucleic acid encoding a hexanoyl-CoA synthetase polypeptide from *C. sativa* was integrated into *S. cerevisiae*. The resulting cells were fed hexanoate to increase hexanoyl-CoA supply (FIG. 2).

[00933] FIG. 2: Intracellular olivetolic acid production using pathway 1a and TKS/OAC polypeptides. Yeast strains yWL004 expressing a heterologous nucleic acid encoding a TKS polypeptide and a heterologous nucleic acid encoding an OAC polypeptide and yWL009 expressing a heterologous nucleic acid encoding a TKS polypeptide, a heterologous nucleic acid encoding an OAC polypeptide, and a heterologous nucleic acid

encoding a HCS polypeptide were grown in YPG and either in the absence or presence of 1mM hexanoate for olivetolic acid production. Addition of hexanoate leads to six-fold increase in olivetolic acid production in strain yWL004, indicating an endogenous acyl-CoA ligase activity. Chromosomal integration of the heterologous nucleic acid encoding a hexanoyl-CoA synthetase polypeptide leading to strain yWL009 shows an additional two-fold increase in olivetolic acid production due to increased hexanoyl-CoA supply.

[00934] Biosynthesis of hexanoyl-CoA from fermentable sugars was increased by integrating pathway 1b, which comprises four enzymes encoding a reverse β-oxidation pathway that has been optimized in *E. coli* for production of hexanol (FIG. 1, box 1b) (Machado et al. 2012). LC-MS analysis confirmed that this pathway was also functional in *S. cerevisiae* and its activity is comparable to pathway 1a, since olivetolic acid yields were similar between the engineered strains yWL009 and yWL0013 (FIG. 3).

[00935] FIG. 3: Intracellular olivetolic acid production comparing pathway 1a and 1b. Yeast strain yWL009 expressing a heterologous nucleic acid encoding a TKS polypeptide, a heterologous nucleic acid encoding an OAC polypeptide, and a heterologous nucleic acid encoding a HCS polypeptide was grown in producing conditions (YPG with 1mM hexanoate) and compared to yWL013 expressing a heterologous nucleic acid encoding a TKS polypeptide, a heterologous nucleic acid encoding a OAC polypeptide, and the hexanoyl-CoA supply pathway 1b grown in non-producing (YPD) as well as producing (YPG) conditions. Chromosomal integration of the hexanoyl-CoA pathway yields similar levels in olivetolic acid production as strain yWL009 when grown in YPG with 1 mM hexanoate.

Table 2: List of strains used in this study

yWL004	Cen.PK2, ACC1::TKS-OAC, tHMGR::MvaE/S,
yWL009	Cen.PK2, ACC1::TKS-OAC, tHMGR::MvaE/S, URA3::HCS
yWL0013	CenPK2, ACC1::TKS-OAC, URA3::HexCoA

[00936] Codon optimized genes were synthesized and used in this study for the polypeptides listed in Table 3.

Table 3: List of polypeptides used in this study

Polypeptide	**Function**	**Original host**
BktB	β-ketothiolase	*Ralstonia eutropha*

Polypeptide	Function	Original host
PaaH1	3-Hydroxyacyl-CoA dehydrogenase	*R. eutropha*
Crt	Crotonase	*Clostridium acetobutylicum*
Ter	Trans-2-enoyl-CoA reductase	*Treponema denticola*
HCS	Hexanoyl-CoA synthetase	*Cannabis sativa*
ERG10	Acetyl-CoA acetyltransferase	*Saccharomyces cerevisiae*
ERG13	HMG-CoA synthase	*S. cerevisiae*
tHMG1	HMG-CoA reductase	*S. cerevisiae*
ERG12	Mevalonate kinase	*S. cerevisiae*
IDI1	Isopentenyl diphosphate:dimethylallyl diphosphate isomerase	*S. cerevisiae*
ERG20	Farnesylpyrophosphate synthetase	*S. cerevisiae*
MvaE	acetyl-CoA acetyltransferase/HMG-CoA reductase	*Escherichia coli*
MvaS	HMG-CoA synthase	*E. coli*
TKS	Tetraketide Synthase (Type III PKS)	*C. sativa*
OAC	Olivetolic acid cyclase	*C. sativa*
GOT	geranyl pyrophosphate:olivetolate geranyltransferase	*C. sativa*
Δ^9-THCAS	Δ^9-tetrahyrdocannabinoidic acid synthase	*C. sativa*
CBDAS	cannabidiolic acid synthase	*C. sativa*
DXS	1-deoxy-D-xylulose-5-phosphate synthase gene	*E. coli*
IspC	1-deoxy-D-xylulose 5-phosphate reductoisomerase	*E. coli*
IspD	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	*E. coli*
IspE	4-diphosphocytidyl-2-C-methylerythritol kinase	*E. coli*
IspF	2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	*E. coli*
IspG	4-hydroxy-3-methylbut-2-en-1-yl diphosphate	*E. coli*

Polypeptide	Function	Original host
	synthase	
IspH	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	*E. coli*
IDI	Isopentenyl diphosphate (IPP) isomerase	*E. coli*
IspA*	mutated FPP synthase (S81F) for GPP production	*E. coli*
AflA	Hexanoyl-CoA synthase, subunit A	*Aspergillus parasiticus*
AflB	Hexanoyl-CoA synthase, subunit B	*A. parasiticus*
SCFA-TE	Short chain fatty acyl-CoA Thioesterase	Various microbes

Example 2 – Synthesis of Olivetolic Acid or Derivatives Thereof or Cannabinoids or Derivatives Thereof

[00937] Multiple polypeptides in pathway 1b require NADH as a co-factor. In order to maximize flux through pathway 1b, other biosynthetic pathways that compete for NADH supply are modified (FIG. 1, Box 2). One target can be the ethanol pathway, mediated by various alcohol dehydrogenase polypeptides, but may also include other pathways that consume NADH, such as the glycerol biosynthesis pathway.

[00938] Another route conceived towards hexanoyl-CoA is described in pathway 1c: The alfatoxin biosynthetic gene cluster (iterative type I PKS) encodes a fatty acid synthase-based mechanism (FasA and FasB) for production of hexanoyl-CoA. In some embodiments, a heterologous nucleic acid encoding a thioesterase polypeptide and a heterologous nucleic acid encoding a CoA ligase polypeptide similar to a C6-tolerant thioesterase polypeptide (see BMC Biochem. 2011 Aug 10;12:44. doi: 10.1186/1471-2091-12-44) and a heterologous nucleic acid encoding a HCS polypeptide are expressed to facilitate release of hexanoyl-ACP and activate free hexanoate to its acyl-CoA compound. Additionally, various type II PKS biosynthetic pathways (e.g. benastatin, R1128) contain a FabH-like KSIII (e.g. BenQ, ZhuH), AT and ACP component, which are crucial for providing and selecting the rare hexanoate PKS starter unit. Lastly, the type I PKS pathway for reveromycin biosynthesis encodes the fatty acyl-CoA ligase RevS polypeptide and the FabH-like KASIII component RevR polypeptide, which are suggested to provide hexanoyl-CoA via fatty acid degradation as well as de novo fatty acid biosynthesis.

[00939] To avoid competitive consumption of hexanoyl-CoA via β-oxidation, the fatty acid degradation pathway is engineered to have lowered activity. Alternatively, yeast are grown in presence of oleic acid to avoid competition for fatty acids as energy source.

[00940] The pathway of four genes encoding the NADH pathway for hexanoyl-CoA production, including polypeptides PaaH1, Crt, Ter, and BktB, was constructed under the control of Gal1, Gal10, Gal7, and TEF2 promoters, respectively. FIG. 4. The whole cassette was inserted between the upstream and downstream homology region of ADE2 and was integrated into the genome of *S. cerevisiae* using CRISPR/Cas9 to generate yXL001 (using Construct 1/pXL044 as shown in FIG. 4). The pathway of four genes encoding the NADPH pathway (including PhaB, PhaJ, Ter, and BktB polypeptides) was introduced into to *S. cerevisiae* in the same way to generate yXL002 (using Construct 1/pXL072 as shown in FIG. 4). The MCT1 gene under the control of Gal1 promoter flanked by the 1622b homology region (Construct 2; FIG. 4) was introduced into the genome of yXL001 and yXL002 using CRISPR/Cas9 to generate yXL003 and yXL004 (FIG. 4).

[00941] A cassette encoding TKS and OAC genes under the control of Gal1 and Gal10 promoters flanked by ACC1 homology region (Construct 4; FIG. 5) was introduced into the genome of yXL003 and yXL004 using CRISPR/Cas9 to generate yXL005 and yXL006. A heterologous nucleic acid encoding a TKS-OAC fusion polypeptide under the control of a Gal1 promoter (Construct 5; FIG. 5) was introduced into yXL003 and yXL004 to generate yXL007 and yXL008. The resulting strains were inoculated into 10 mL YP medium supplemented with 2% dextrose. After an overnight culture at 30 °C and centrifugation at 3,000 × g for 5 mins, the pellet was resuspended into YP medium supplemented with 2% galactose. After two days expression, the culture supernatant was extracted with equal volume of ethyl acetate, and, after evaporation and filtration, the samples were analyzed by LC-MS, which showed the production of a significant amount of olivetolic acid (FIG. 9 and FIG. 10).

[00942] CsAAE (Construct 3; FIG. 4), TKS, and OAC genes (Construct 4; FIG. 5) were introduced into the genome of *S. cerevisiae* using CRISPR/Cas9 to generate yXL009, which can produce higher level of olivetolic acid in the presence of exogenously supplied hexanoate (FIG. 11).

[00943] In addition, by supplementing the growth medium with various aliphatic acids, from C_4-C_{10}, various olivetolic acid derivatives can be produced from yXL009 (FIG. 11 and FIG. 12). Some of the olivetolic acid derivatives can be further modified by biological or chemical means to covalently attach to other compounds. For example, click

chemistry can be performed on the olivetolic derivative containing alkyne functional group. The olivetolic derivative is dissolved in biology grade dimethyl sulfoxide (DMSO) and treated with a DMSO solution of crosslinker containing an azide group (1.0 equiv.), TBTA (DMSO: tBuOH 1:1), $CuSO_4$ $5H_2O$, sodium ascorbate and HEPES-KOH pH: 7.0 (final HEPES-KOH ≈250 mM). The reaction is placed on a water bath at 37 °C for 12 to 16 hours. Liquid chromatograph-mass spectrometry (LC-MS) analysis of the reaction mixture shows reaction completion after 16 hours to obtain the further modified olivetolic acid.

[00944] The GPPS large subunit (GPPSlsu) and small subunit (GPPSssu) genes from *Cannabis sativa* under the control of Gal1 and Gal10 promoters flanked by ADE1 homology region (Construct 10; FIG. 7) were introduced into yXL008 and yXL009 to generate yXL010 and yXL011. A cassette encoding a NphB polypeptide and a THCAS polypeptide under the control of Gal1 and Gal10 promoters flanked by 1014a homology region (Construct 12; FIG. 8) was introduced into the genome of yXL010 and yXL011 to generate yXL012 and yXL013 using CRISPR/Cas9. The resulting strains were inoculated into 10 mL YP medium supplemented with 2% dextrose. After an overnight culture at 30 °C and centrifugation at 3,000 × g for 5 mins, the pellet was resuspended into YP medium supplemented with 2% galactose. After two days expression, the culture supernatant was extracted with equal volume of ethyl acetate, and, after evaporation and filtration, the samples were analyzed by LC-MS, which showed that the overexpression of NphB in yXL010 resulted in the production of cannabigerolic acid (FIGS. 14 and 15). In the presence of a THCAS polypeptide, the cannabigerolic acid was transformed into THCA or into THC. With yXL013, C_4-C_{10} acids were added to the expression medium, resulting in the production of cannabigerolic acid derivatives, which were then modified by a THCAS polypeptide to produce THCA or THC derivatives. Those derivatives can then be further modified by chemical reactions (FIG. 13).

Example 3 – Synthesis of Cannabinoid Precursors, Cannabinoids, or Derivatives of the Foregoing

[00945] To recreate cannabinoid production in microorganisms, chassis *S. cerevisiae* strains were developed containing metabolic pathways for the production of (1) GPP through the mevalonate (Mva) pathway, (2), olivetolic acid or derivatives, (3) CBGA or derivatives, and (4) different cannabinoids or cannabinoid derivatives produced by cannabinoid synthase polypeptides.

Production of GPP

[00946] A GPP-overproducing strain, GTY23, was produced by overexpressing Mva pathway genes and introducing a repressible promoter on ERG9. A previously described ERG20 F96W-N127W mutant, ERG20mut, was added to provide a source of GPP precursor in the cell (FIG. 16). This strain was used to screen GOT polypeptide candidates.

Production of Olivetolic Acid or Derivatives Thereof

[00947] Olivetolic acid was produced from sugar by introducing genes CsTKS and CsOAC, and pathways to produce hexanoyl-CoA. Pathways for the production of hexanoate and hexanoyl-CoA are known in the art (e.g., Gajewski et al, "Engineering fungal de novo fatty acid synthesis for short chain fatty acid production," Nature Communications 2017). To produce olivetolic acid or its derivatives, rather than using hexanoyl-CoA pathways, a previously reported acyl-CoA ligase polypeptide, such as a CsAAE1 or CsAAE3 polypeptide, was introduced and exogenously fed cells hexanoate or a carboxylic acid other than hexanoate (FIGS. 17-19). These pathways allow for the production of non-naturally occurring cannabinoids.

Production of CBGA

[00948] The mother cannabinoid CBGA, or derivatives thereof, was produced by a GOT polypeptide. A *C. sativa* GOT polypeptide was identified in the 1990s, yet no report was identified describing reconstituting GOT polypeptide activity *in vivo*. Twenty-five polypeptide variants were screened for *in vivo* production of CBGA in strains containing GPP pathways and exogenously fed olivetolic acid. These genes were all chromosomally integrated driven by GAL1 promoters and screened for activity in yeast extract peptone galactose (YPG) media. GC-MS and LC-MS analysis demonstrated *in vivo* production of CBGA from a CsPT4t polypeptide (FIGS 26A-C). The gene sequence of the CsPT4t polypeptide is referred to as a GOT polypeptide (FIG. 20). yL444 was the strain used in the production of CBGA and expresses the following genotype: CEN.PK2-1D {1114a::GAL1p-CsPT4t-TDH1t; 308a::GAL1p-ERG20(F96W-N127W)-TDH1t; erg9::KanMX_CTR3p-ERG9; leu2-3,112::His3MX6_GAL1p-ERG19/GAL10p-ERG8; ura3-52::ura3/GAL1p-MvaS(A110G)/GAL10p-MvaE; his3_1::hphMX4_GAL1p-ERG12/GAL10p-IDI1; MATa} (FIGS. 6 and 20). LC-MS was carried out as follows (FIGS 26A-C):

Column info: 2015 Kinetex XB-C18 2.1x100 mm RES6 method 10.6 min

Method info:

0-5.6 mins, 45%-73%B, 0.2 mL/min

5.6-6.2 mins, 73%-97%B, 0.2 mL/min

6.2-11.3 mins, 97%B, 0.3 mL/min

11.3-12.7, 97-45%B, 0.3 mL/min

12.7-15.5, 45%B, 0.3 mL/min

A: H2O+0.05% TFA

Production of THCA and CBDA

[00949] Cannabinoid synthase genes have been identified from the *Cannabis* genome (including but not limited to THCA synthase (THCAS), CBDA synthase (CBDAS), JP450547, JP454863, JP471546, JP452622). To produce THCA and CBDA, the corresponding THCA synthase and CBDA synthase, respectively, were introduced into a strain producing CBGA containing a heterologous nucleic acid encoding a CsPT4t polypeptide. The synthases were introduced as N-terminal truncated polypeptides with polypeptide tags, e.g., ProA signal sequence (MIFDGTTMSIAIGLLSTLGIGAEA, from proteinase A with UniProt accession number F2QUG8) attached and the transcription of both synthases were under the control of GAL10 promoter. The final plasmid constructs were named as pESC-ProA-THCAS and pESC-ProA-CBDAS. Both plasmids were transformed individually into the above-mentioned strain, which has high CBGA production in the presence of olivetolic acid, to give strains yXL046 and yXL047 (FIGS. 21-25).

[00950] After confirming the transformation by PCR of THCAS or CBDAS, two colonies from each culture were inoculated into a defined medium (SC-Leu + 2% Dextrose) and were incubated at 30 °C with shaking at 800 RPM. After two-day growth, the cultures were back-diluted 1:50 into inducing medium (SC-Leu + 2% galactose + 1 mM olivetolic acid + $CuSO_4$) and incubated at 30 °C with shaking at 800 RPM for 4 days. After 4-day incubation, equal volume of ethyl acetate was added to the expression cultures and the mixtures were subjected to three rounds of bead beating. Then the mixtures were then spun down at 5000 RPM and the organic layers were sent for LC-MS analysis, which showed the production of THCA and CBGA from the corresponding cultures (FIGS. 27 and 28).

Example 4 – Generation of a Base Yeast Strain Capable of High Flux to CBGA with Olivetolic Acid Feeding

[00951] CBGA production strains were created from wild-type *Saccharomyces cerevisiae* strain (CEN.PK2) by expressing genes of the mevalonate pathway polypeptides and a GOT polypeptide under control of the GAL1 or GAL10 promoter. The S21 strain comprised the following chromosomally integrated mevalonate pathway genes from *S. cerevisiae*: ERG10, ERG13, truncated HMG1 (tHMGR), ERG12, ERG8, ERG19, and IDI1.

The S21 strain additionally comprised the chromosomally integrated pyruvate decarboxylase (PDC) from *Zymomonas mobilis* to increase flux from pyruvate towards acetyl-CoA.

[00952] To generate additionally strains, a mutant form of ERG20, ERG20mut, which preferentially generates GPP was added to the S21 strain with the following chromosomally integrated GOTs from *C. sativa*: CsPT1 (S164), a truncated CsPT1 (CsPT1_t75, S165), or CsPT4 (S29). Constructs used in S29, S164, and S165 are shown in Table 11.

[00953] Yeast colonies verified to contain the expected DNA assembly comprising one or more heterologous nucleic acids disclosed herein were picked into 96-well microtiter plates containing 360 µL of YPD (10 g/L yeast extract, 20 g/L Bacto peptone, 20 g/L dextrose (glucose)) and sealed with a breathable film seal. Cells were cultured at 30 °C in a high capacity microtiter plate incubator shaking at 1000 rpm and 80% humidity for 3 days until the cultures reached carbon exhaustion. The growth-saturated cultures were subcultured into fresh plates containing YPGAL and olivetolic acid (10 g/L yeast extract, 20 g/L Bacto peptone, 20 g/L galactose, 1 g/L glucose and 1 mM olivetolic acid) by taking 14.4 µL from the saturated cultures and diluting into 360 µL of fresh media and sealed with a breathable film seal. Genetically modified host cells in the production media were cultured at 30 °C in a high capacity microtiter plate shaker at 1000 rpm and 80% humidity for an additional 3 days prior to extraction and analysis. Upon completion, 100 µL of whole cell broth was diluted into 900 µL of methanol, sealed with a foil seal, and shaken at 1500 rpm for 60 seconds to extract the cannabinoids. After shaking, the plate was centrifuged at 1000 x g for 60 seconds to remove any solids. After centrifugation, 12 µL of supernatant was transferred to a fresh assay plate containing 228 µL of methanol, sealed with a foil seal, shaken for 60 seconds at 900 rpm, and analyzed by LC-MS.

[00954] Samples were analyzed by LC-MS mass spectrometer (Agilent 6470) using an Agilent Poroshell 120 Phenyl Hexyl 2.1 x 50 mm, 1.9 µm analytical column with the following gradient (Mobile Phase A: LC-MS grade water with 0.1% formic acid; Mobile Phase B: LC-MS grade acetonitrile with 0.1% formic acid):

Time (minutes)	**%B**
0	40
0.1	40
0.6	60
1	65
1.01	95
2.01	95
2.02	40
2.5	40

[00955] The mass spectrometer was operated in negative ion multiple reaction monitoring mode. Each cannabinoid was identified by retention time, determined from an authentic standard, and MRM transition (see FIGS. 77 and 78):

Compound Name	Q1 Mass (Da)	Q3 Mass (Da)
CBGA	359.2	341.1
CBGA	359.2	315.2

[00956] CsPT1 polypeptide and CsPT1_t75 polypeptide produced equivalent amounts of CBGA *in vivo* (1.3 mg/L CBGA). However, CsPT4 polypeptide produced 216 mg/L CBGA *in vivo* (see FIGS 79-82).

Example 5 – Determining the Minimal Catalytic Domain of CsPT4

[00957] To determine the minimal catalytic domain of CsPT4 polypeptide required for the conversion of GPP and olivetolic acid to CBGA, multiple N-terminal truncations of the CsPT4 polypeptide were generated (see Table 11) and expressed *in vivo* in the S21 strain with feeding of 1 mM olivetolic acid. Only full length CsPT4 polypeptide and CsPT4_t76 polypeptide (CsPT4t) displayed activity *in vivo* (Table 4).

Table 4: Screening of CsPT4 truncated polypeptides

CsPT4 construct	Strain	Peak intensity
CsPT4	S29	8901
CsPT4_t76	S147	6859
CsPT4_t112	S166	19
CsPT4_t131	S167	24
CsPT4_t142	S168	20
CsPT4_t166	S169	21
CsPT4_t186	S170	29

Example 6 – Generation of a Base Yeast Strain Capable of High Flux to CBGA with Hexanoic Acid (Caproic Acid) Feeding

[00958] To convert the high flux strain for the production of CBGA with olivetolic acid (S29) to a high flux stain for the production of CBGA with hexanoic acid, genes responsible for the production to olivetolic acid from fatty acids were expressed using the GAL1 or GAL10 promoter in S29. The strain comprised the following chromosomally integrated olivetolic acid pathway genes from *C. sativa*: three copies of TKS and three copies of OAC. Three different strains were generated with two copies of *C. sativa* AAE1

(S78), two copies of *C. sativa* AAE3 (S81), or two copies of *S. cerevisiae* FAA2 (S83) (see Table 11 for information on the strains). The strains were grown and tested as in Examples 4 and 5 but with 2 mM hexanoic acid added to the media instead of 1 mM olivetolic acid. Production of CBGA by the strains was observed (Table 5).

Table 5: Generation of CBGA

		Titer (mg/L)		
Feed compound	**Product**	**AAE1v1 (S78)**	**AAE3-Ctrunc (S81)**	**FAA2 (S83)**
Hexanoic acid	CBGA	38.5	32.1	35.1

Example 7 – Generation of a Base Yeast Strain Capable of High Flux to CBDA and THCA

[00959] To convert the high flux strain for the production of CBGA to a high flux strain for the production of CBDA or THCA, a heterologous nucleic acid encoding CBDA synthase polypeptide (S34) or THCA synthase polypeptide (S123) was added to Strain S29 (see Table 11 for information on the strains). The strains were tested as in Examples 4 and 5 with 1 mM olivetolic acid in the media. CBDA and THCA were produced by the strains, as shown in FIGS. 83 and 84.

Example 8 – Feeding of Cannabinoid Precursor Derivatives to Yeast to Produce Rare and Non-Naturally Occurring CBGA Derivatives

[00960] Strains from Example 6 (S78, S81, and S83) were grown as in Examples 4 and 5 but with 2 mM of a carboxylic acid (detailed in Table 6) added to the media and analyzed as in Example 4. Table 6 details the products produced by the strains (product peak intensity).

Table 6: CBGA Derivatives Produced

Feed compound	**Product (IUPAC name)**	**Transition 1**	**Transition 2**	**AAE1v1 (S78)**	**AAE3-Ctrunc (S81)**	**FAA2 (S83)**
2-methyl hexanoic acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-6-(hexan-2-yl)-2,4-dihydroxybenzoic acid	373 --> 355	373 --> 329	1136	1255	1301

4-methyl hexanoic acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-(3-methylpentyl)benzoic acid	373 --> 355	373 --> 329	82453	91493	82517
5-methyl hexanoic acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-(4-methylpentyl)benzoic acid	373 --> 355	373 --> 329	76145	77270	77145
2-hexenoic acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-[(1E)-pent-1-en-1-yl]benzoic acid	357 --> 339	357 --> 313	311	536	588
3-hexenoic acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-[(2E)-pent-2-en-1-yl]benzoic acid	357 --> 339	357 --> 313	90422	104366	112440
5-hexenoic acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-(pent-4-en-1-yl)benzoic acid	357 --> 339	357 --> 313	302499	325854	365798
butanoic acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-propylbenzoic acid	331 --> 313	331 --> 287	92181	106229	103368
pentanoic acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-butylbenzoic acid	345 --> 327	345 --> 301	224003	232206	236366
heptanoic acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-hexylbenzoic acid	373 --> 355	373 --> 329	66544	67766	66570
octanoic acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-heptylbenzoic acid	387 --> 369	387 --> 343	4225	3212	3603
5-chloro pentanoic acid	6-(4-chlorobutyl)-3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxybenzoic acid	379 -->361	379 --> 335	1023	947	902

5-(methyl sulfanyl) pentanoic acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-[4-(methylsulfanyl) butyl]benzoic acid	391 --> 373	391 --> 347	18396	18704	19412

Example 9 – Feeding of Cannabinoid Precursor Derivatives to Yeast to Produce Rare and Non-Naturally Occurring CBDA Derivatives

[00961] Strains with (S34) or without a CBDA synthase polypeptide (S29) were tested as in Examples 4 and 5 with 1 mM of an olivetolic acid derivative (detailed in Table 7). Table 7 details the products produced by the strains.

Table 7: CBDA Derivatives Produced

Feed compound	**Product (IUPAC name)**	**Transition 1**	**Transition 2**	**CBGA derivative titer (mg/L) (S29)**	**CBDA derivative titer (mg/L) (S34)**
Orsellinic Acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-methylbenzoic acid	303 --> 285	303 --> 259	1.86	1.05
Divarinic Acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-propylbenzoic acid	331 --> 313	331 --> 287	29.54	3.06

Example 10 – Feeding of Cannabinoid Precursors to Yeast to Produce CBDA or CBGA

[00962] As there are numerous ways to produce cannabinoid precursors (e.g., GPP), a number of different genes were tested *in vivo* to optimize cannabinoid production. Different GPP synthase polypeptides, CBDA synthase polypeptides, TKS polypeptides, OAC polypeptides, medium and long chain fatty acyl-CoA synthetase polypeptides were tested in various combinations as described below (see Table 11 for information on the strains).

[00963] Strains were constructed with different GPP synthase polypeptides to identify the best producer of GPP for production of CBGA when fed 1 mM olivetolic acid. Strain S21 was transformed with heterologous nucleic acids encoding a CsPT4 polypeptide and a GPP synthase polypeptide. CBGA titer was measured as described in Example 4. CBGA titers, titer standard deviations (SD) and number of replicates tested are indicated in Table 8.

Table 8: Production of CBGA

Product	Strain	Titer (mg/L)	SD	n
CBGA	S29	215.6	12.2	8
CBGA	S114	6.8	0.7	3
CBGA	S116	15.5	2.0	4
CBGA	S108	8.5	1.7	4
CBGA	S112	9.9	1.6	4
CBGA	S104	10.2	1.6	3
CBGA	S115	9.2	1.9	4
CBGA	S118	5.1	NA	1

[00964] To optimize production of CBDA, strain S29 was transformed with a series of constructs with two copies of a CBDA synthase polypeptide encoding heterologous nucleic acid and grown as in Examples 4 and 5 with 1 mM olivetolic acid. CBDA was measured as described in Example 4. CBDA peak intensity, peak intensity standard deviations (SD) and number of replicates tested are indicated in Table 9.

Table 9: Production of CBDA

Product	Strain	Peak Area	SD	n
CBDA	S34	1651	329	4
CBDA	S35	831	72	4
CBDA	S37	505	26	4
CBDA	S38	658	31	4
CBDA	S39	1274	85	4
CBDA	S41	2129	462	4
CBDA	S42	72	4	4
CBDA	S43	419	481	4
CBDA	S44	758	68	4
CBDA	S45	1253	177	4
CBDA	S46	670	112	4
CBDA	S47	300	15	4

[00965] To optimize production of CBGA from hexanoic acid, different combinations of TKS polypeptide, OAC polypeptide, and medium and long chain fatty acyl-CoA synthetase polypeptide were tested *in vivo*. All strains were daughters or granddaughters of strain S29. All strains were tested as described in Example 4 with 2 mM hexanoic acid added to the production media. CBGA titers, titer standard deviations (SD) and number of replicates tested are indicated in Table 10.

Table 10: Production of CBGA

Product	Strain	Titer (mg/L)	SD	n
CBGA	S31	53.6	12.2	8

Product	Strain	Titer (mg/L)	SD	n
CBGA	S49	55.7	9.3	8
CBGA	S50	22.9	7.0	8
CBGA	S90	67.5	2.8	4
CBGA	S91	63.5	4.2	4
CBGA	S78	38.5	2.5	4
CBGA	S80	37.5	1.8	4
CBGA	S81	32.1	5.8	4
CBGA	S82	35.1	7.0	4
CBGA	S83	35.1	2.6	4
CBGA	S84	36.4	3.5	4
CBGA	S85	34.4	4.3	4
CBGA	S86	36.6	1.8	4
CBGA	S87	32.2	4.9	4
CBGA	S88	40.9	1.4	4
CBGA	S89	39.3	2.7	4
CBGA	S94	59.6	7.9	8
CBGA	S95	58.5	9.2	8
CBGA	S97	72.9	5.5	8

Table 11: Constructs and strains used in the Examples

Strain (Contructs)	Parent Strain*	Polypeptide SEQ ID NOs (Nucleotide SEQ ID NOs)
S21 (FIGS. 29A and 29B)		Sc_tHMG1: SEQ ID NO:208 (SEQ ID NO:119) Sc_ERG13: SEQ ID NO:115 (SEQ ID NO:120) Sc_ERG10: SEQ ID NO:25 (SEQ ID NO:209) Sc_MVD1 (Sc_ERG19): SEQ ID NO:66 (SEQ ID NO:65) Sc_IDI1: SEQ ID NO:58 (SEQ ID NO:57) Zm_PDC: SEQ ID NO:117 (SEQ ID NO:118) Sc_ERG8: SEQ ID NO:205 (SEQ ID NO:204) Sc_ERG12: SEQ ID NO:64 (SEQ ID NO:206)
S29 (FIG. 86)	S21	Cs_PT4: SEQ ID NO:110 (SEQ ID NO:111) Sc_ERG20_mut: SEQ ID NO:60 (SEQ ID NO:161)
S31 (FIGS. 30A, 30B, and 30C)	S29	Cs_OAC: SEQ ID NO:10 (SEQ ID NO:163) Cs_TKS: SEQ ID NO:11 (SEQ ID NO:162) Cs_AAE1_v1: SEQ ID NO:90 (SEQ ID NO:164) Sc_FAA2: SEQ ID NO:169 (SEQ ID NO:168)
S34 (FIG. 85)	S29	Cs_CBDAS_co1: SEQ ID NO:88 (SEQ ID NO:167)
S35 (FIG. 31)	S29	Cs_CBDAS_t28: SEQ ID NO:151 (SEQ ID NO:152)

Strain (Contructs)	Parent Strain*	Polypeptide SEQ ID NOs (Nucleotide SEQ ID NOs)
S37 (FIG. 32)	S29	MBP_co1: SEQ ID NO:108 (SEQ ID NO:170) Cs_CBDAS_t28: SEQ ID NO:151 (SEQ ID NO:152) GS12: SEQ ID NO:172 (SEQ ID NO:171)
S38 (FIG. 33)	S29	Cs_CBDAS_t28: SEQ ID NO:151 (SEQ ID NO:152) GB1: SEQ ID NO:174 (SEQ ID NO:173) GS12: SEQ ID NO:172 (SEQ ID NO:171)
S39 (FIG. 34)	S29	Cs_CBDAS_t28: SEQ ID NO:151 (SEQ ID NO:152) Sc_MFalpha1_1-19: SEQ ID NO:176 (SEQ ID NO:175)
S41 (FIG. 35)	S29	Cs_CBDAS_t28: SEQ ID NO:151 (SEQ ID NO:152) Sc_MFalpha1_1-89: SEQ ID NO:178 (SEQ ID NO:177)
S42 (FIG. 36)	S29	Cs_CBDAS_t28: SEQ ID NO:151 (SEQ ID NO:152) DasherGFP: SEQ ID NO:180 (SEQ ID NO:179) GS12: SEQ ID NO:172 (SEQ ID NO:171)
S43 (FIG. 37)	S29	Cs_CBDAS_t28: SEQ ID NO:151 (SEQ ID NO:152) GS12: SEQ ID NO:172 (SEQ ID NO:171) ER1_tag: SEQ ID NO:182 (SEQ ID NO:181)
S44 (FIG. 38)	S29	Cs_CBDAS_t28: SEQ ID NO:151 (SEQ ID NO:152) GS12: SEQ ID NO:172 (SEQ ID NO:171) ER2_tag: SEQ ID NO:184 (SEQ ID NO:183)
S45 (FIG. 39)	S29	Cs_CBDAS_t28: SEQ ID NO:151 (SEQ ID NO:152) GS12: SEQ ID NO:172 (SEQ ID NO:171) PM1_tag: SEQ ID NO:186 (SEQ ID NO:185)
S46 (FIG. 40)	S29	Cs_CBDAS_t28: SEQ ID NO:151 (SEQ ID NO:152) GS12: SEQ ID NO:172 (SEQ ID NO:171) VC1_tag: SEQ ID NO:188 (SEQ ID NO:187)
S47 (FIG. 41)	S29	Cs_CBDAS_t28: SEQ ID NO:151 (SEQ ID NO:152)

Strain (Contructs)	Parent Strain*	Polypeptide SEQ ID NOs (Nucleotide SEQ ID NOs)
		PEX8_tag: SEQ ID NO:190 (SEQ ID NO:189)
S49 (FIGS. 42A, 42B, and 42C)	S29	Cs_OAC: SEQ ID NO:10 (SEQ ID NO:163) Cs_TKS: SEQ ID NO:11 (SEQ ID NO:162) Cs_AAE1_v1: SEQ ID NO:90 (SEQ ID NO:164) Cs_AAE_v1: SEQ ID NO:90 (SEQ ID NO:164)
S50 (FIGS. 43A, 43B, and 43C)	S29	Cs_OAC: SEQ ID NO:10 (SEQ ID NO:163) Cs_TKS: SEQ ID NO:11 (SEQ ID NO:162) Sc_FAA2: SEQ ID NO:169 (SEQ ID NO:168)
S51 (FIGS. 44A, 44B, and 44C)	S29	Cs_OAC: SEQ ID NO:10 (SEQ ID NO:163) Cs_TKS: SEQ ID NO:11 (SEQ ID NO:162)
S78 (FIG. 45)	S51	Cs_AAE1_v1: SEQ ID NO:90 (SEQ ID NO:164) Cs_AAE_v1: SEQ ID NO:90 (SEQ ID NO:164) GB1: SEQ ID NO:174 (SEQ ID NO:173)
S80 (FIG. 46)	S51	Cs_AAE3: SEQ ID NO:92 (SEQ ID NO:166)
S81 (FIG. 47)	S51	Cs_AAE3_Ctrunc: SEQ ID NO:149 (SEQ ID NO:150)
S82 (FIG. 48)	S51	Sc_FAA1: SEQ ID NO:192 (SEQ ID NO:191) FAA1: SEQ ID NO:192 (SEQ ID NO:191)
S83 (FIG. 49)	S51	Sc_FAA2: SEQ ID NO:169 (SEQ ID NO:168)
S84 (FIG. 50)	S51	Sc_FAA2_Ctrunc: SEQ ID NO:194 (SEQ ID NO:193)
S85 (FIG. 51)	S51	Sc_FAA2_Cmut: SEQ ID NO:196 (SEQ ID NO:195) Sc_FAA2: SEQ ID NO:169 (SEQ ID NO:168)
S86 (FIG. 52)	S51	Sc_FAA3: SEQ ID NO:198 (SEQ ID NO:197)
S87 (FIG. 53)	S51	Sc_FAA4: SEQ ID NO:200 (SEQ ID NO:199)
S88 (FIG. 54)	S51	Cs_AAE1_v1: SEQ ID NO:90 (SEQ ID NO:164) Sc_ACC1_act: SEQ ID NO:207 (SEQ ID NO:201)
S89 (FIG. 55)	S51	Sc_FAA2: SEQ ID NO:169 (SEQ ID NO:168) Sc_ACC1_act: SEQ ID NO:207 (SEQ ID NO:201)
S90 (FIGS. 56A, 56B, and 56C)	S29	Cs_OAC: SEQ ID NO:10 (SEQ ID NO:163) Cs_TKS: SEQ ID NO:11 (SEQ ID NO:162) Cs_AAE1_v1: SEQ ID NO:90 (SEQ ID NO:164)

Strain (Contructs)	Parent Strain*	Polypeptide SEQ ID NOs (Nucleotide SEQ ID NOs)
		Cs_AAE_v1: SEQ ID NO:90 (SEQ ID NO:164)
S91 (FIGS. 57A, 57B, and 57C)	S29	Cs_OAC: SEQ ID NO:10 (SEQ ID NO:163) Cs_TKS: SEQ ID NO:11 (SEQ ID NO:162) Sc_FAA2: SEQ ID NO:169 (SEQ ID NO:168)
S94 (FIG. 58)	S31	Cs_PT4_full: SEQ ID NO:110 (SEQ ID NO:111)
S95 (FIG. 59)	S31	GB1: SEQ ID NO:174 (SEQ ID NO:173) Cs_OAC: SEQ ID NO:10 (SEQ ID NO:163)
S97 (FIG. 60)	S31	Cs_OAC: SEQ ID NO:10 (SEQ ID NO:163) Cs_TKS: SEQ ID NO:11 (SEQ ID NO:162) GS12: SEQ ID NO:172 (SEQ ID NO:171)
S104 (FIG. 61)	S21	Cs_PT4: SEQ ID NO:110 (SEQ ID NO:111) Ag_GPPS: SEQ ID NO:133 (SEQ ID NO:134) GB1: SEQ ID NO:174 (SEQ ID NO:173)
S108 (FIG. 62)	S21	Cs_PT4: SEQ ID NO:110 (SEQ ID NO:111) Hb_GPPS: SEQ ID NO:143 (SEQ ID NO:144) GB1: SEQ ID NO:174 (SEQ ID NO:173)
S112 (FIG. 63)	S21	Cs_PT4: SEQ ID NO:110 (SEQ ID NO:111) Cs_GPPS_NTrunc: SEQ ID NO:127 (SEQ ID NO:128)
S114 (FIG. 64)	S21	Cs_PT4: SEQ ID NO:110 (SEQ ID NO:111) Pa_GPPS_NTrunc: SEQ ID NO:131 (SEQ ID NO:132)
S115 (FIG. 65)	S21	Cs_PT4: SEQ ID NO:110 (SEQ ID NO:111) Ag_GPPS_NTrunc: SEQ ID NO:203 (SEQ ID NO:202)
S116 (FIG. 66)	S21	Cs_PT4: SEQ ID NO:110 (SEQ ID NO:111) Pb_GPPS_NTrunc: SEQ ID NO:135 (SEQ ID NO:136)
S118 (FIG. 67)	S21	Cs_PT4: SEQ ID NO:110 (SEQ ID NO:111) Es_GPPS_NTrunc: SEQ ID NO:139 (SEQ ID NO:140)
S123 (FIG. 68)	S29	Cs_THCAS_full: SEQ ID NO:155 (SEQ ID NO:156)
S147 (FIG. 69)	S21	Cs_PT4: SEQ ID NO:100 (SEQ ID NO:224) Sc_ERG20_mut: SEQ ID NO:60 (SEQ ID NO:161)
S164 (FIG. 70)	S21	Cs_PT4: SEQ ID NO:82 (SEQ ID NO:220)

Strain (Contructs)	Parent Strain*	Polypeptide SEQ ID NOs (Nucleotide SEQ ID NOs)
		Sc_ERG20_mut: SEQ ID NO:60 (SEQ ID NO:161)
S165 (FIG. 71)	S21	Cs_PT4: SEQ ID NO:223 (SEQ ID NO:222) Sc_ERG20_mut: SEQ ID NO:60 (SEQ ID NO:161)
S166 (FIG. 72)	S21	Cs_PT4: SEQ ID NO:211 (SEQ ID NO:210) Sc_ERG20_mut: SEQ ID NO:60 (SEQ ID NO:161)
S167 (FIG. 73)	S21	Cs_PT4: SEQ ID NO:213 (SEQ ID NO:212) Sc_ERG20_mut: SEQ ID NO:60 (SEQ ID NO:161)
S168 (FIG. 74)	S21	Cs_PT4: SEQ ID NO:215 (SEQ ID NO:214) Sc_ERG20_mut: SEQ ID NO:60 (SEQ ID NO:161)
S169 (FIG. 75)	S21	Cs_PT4: SEQ ID NO:217 (SEQ ID NO:216) Sc_ERG20_mut: SEQ ID NO:60 (SEQ ID NO:161)
S170 (FIG. 76)	S21	Cs_PT4: SEQ ID NO:219 (SEQ ID NO:218) Sc_ERG20_mut: SEQ ID NO:60 (SEQ ID NO:161)

* If a strain has a parent strain, it is a child strain. All of the constructs present in the parent strain are also all present in the child strain.

[00966] While the present disclosure has been described with reference to the specific embodiments thereof, it should be understood by those skilled in the art that various changes may be made and equivalents may be substituted without departing from the true spirit and scope of the disclosure. In addition, many modifications may be made to adapt a particular situation, material, composition of matter, process, process step or steps, to the objective, spirit and scope of the present disclosure. All such modifications are intended to be within the scope of the claims appended hereto.

CLAIMS

What is claimed is:

1. A genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide, wherein said geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide can catalyze production of cannabigerolic acid from geranyl pyrophosphate and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

2. A genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110.

3. A genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100.

4. The genetically modified host cell of any one of claims 1-3, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a tetraketide synthase (TKS) polypeptide and one or more heterologous nucleic acids encoding an olivetolic acid (OAC) polypeptide, or one or more heterologous nucleic acids encoding a fusion TKS and OAC polypeptide.

5. The genetically modified host cell of claim 4, wherein the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76.

6. The genetically modified host cell of claim 4 or 5, wherein the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

7. The genetically modified host cell of any one of claims 1-6, wherein the genetically modified host cell further comprises one or more of the following:

a) one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative;

b) one or more heterologous nucleic acids encoding a polypeptide that generates geranyl pyrophosphate; or

c) one or more heterologous nucleic acids encoding a polypeptide that generates malonyl-CoA.

8. The genetically modified host cell of claim 7, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative, wherein the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is an acyl-activating enzyme (AAE) polypeptide.

9. The genetically modified host cell of claim 8, wherein the AAE polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90.

10. The genetically modified host cell of claim 8, wherein the AAE polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:92 or SEQ ID NO:149.

11. The genetically modified host cell of claim 7, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative, wherein the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is a fatty acyl-CoA ligase polypeptide.

12. The genetically modified host cell of claim 11, wherein the fatty acyl-CoA ligase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:145 or SEQ ID NO:147.

13. The genetically modified host cell of claim 7, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative, wherein the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is a fatty acyl-CoA synthetase (FAA) polypeptide.

14. The genetically modified host cell of claim 13, wherein the FAA polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200.

15. The genetically modified host cell of any one of claims 7-14, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates geranyl pyrophosphate, wherein the polypeptide that generates geranyl pyrophosphate is a geranyl pyrophosphate synthetase (GPPS) polypeptide.

16. The genetically modified host cell of claim 15, wherein the GPPS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60.

17. The genetically modified host cell of any one of claims 7-16, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates malonyl-CoA, wherein the polypeptide that generates malonyl-CoA is an acetyl-CoA carboxylase-1 (ACC1) polypeptide.

18. The genetically modified host cell of claim 17, wherein the ACC1 polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207.

19. The genetically modified host cell of any one of claims 1-18, wherein the genetically modified host cell further comprises one or more of the following: a) one or more heterologous nucleic acids encoding a HMG-CoA synthase (HMGS) polypeptide; b) one or more heterologous nucleic acids encoding a 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR) polypeptide; c) one or more heterologous nucleic acids encoding a mevalonate kinase (MK) polypeptide; d) one or more heterologous nucleic acids encoding a phosphomevalonate kinase (PMK) polypeptide; e) one or more heterologous nucleic acids encoding a mevalonate pyrophosphate decarboxylase (MVD) polypeptide; or f) one or more heterologous nucleic acids encoding a isopentenyl diphosphate isomerase (IDI) polypeptide.

20. The genetically modified host cell of claim 19, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an IDI polypeptide.

21. The genetically modified host cell of claim 20, wherein the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58.

22. The genetically modified host cell of any one of claims 19-21, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGR polypeptide.

23. The genetically modified host cell of claim 22, wherein the HMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:22.

24. The genetically modified host cell of any one of claims 19-21, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGR polypeptide, wherein the HMGR polypeptide is a truncated HMGR (tHMGR) polypeptide.

25. The genetically modified host cell of claim 24, wherein the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208.

26. The genetically modified host cell of any one of claims 19-25, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGS polypeptide.

27. The genetically modified host cell of claim 26, wherein the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115.

28. The genetically modified host cell of any one of claims 19-27, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an MK polypeptide.

29. The genetically modified host cell of claim 28, wherein the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64.

30. The genetically modified host cell of any one of claims 19-29, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a PMK polypeptide.

31. The genetically modified host cell of claim 30, wherein the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205.

32. The genetically modified host cell of any one of claims 19-31, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a MVD polypeptide.

33. The genetically modified host cell of claim 32, wherein the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66.

34. The genetically modified host cell of any one of claims 1-33, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA.

35. The genetically modified host cell of claim 34, wherein the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide.

36. The genetically modified host cell of claim 35, wherein the acetoacetyl-CoA thiolase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25.

37. The genetically modified host cell of any one of claims 1-36, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a pyruvate dehydrogenase complex (PDC) polypeptide.

38. The genetically modified host cell of claim 37, wherein the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117.

39. The genetically modified host cell of any one of claims 1-38, wherein the genetically modified host cell is a eukaryotic cell.

40. The genetically modified host cell of claim 39, wherein the eukaryotic cell is a yeast cell.

41. The genetically modified host cell of claim 40, wherein the yeast cell is *Saccharomyces cerevisiae*.

42. The genetically modified host cell of claim 41, wherein the *Saccharomyces cerevisiae* is a protease-deficient strain of *Saccharomyces cerevisiae*.

43. The genetically modified host cell of any one of claims 1-39, wherein the genetically modified host cell is a plant cell.

44. The genetically modified host cell of any one of claims 1-38, wherein the genetically modified host cell is a prokaryotic cell.

45. The genetically modified host cell of any one of claims 1-44, wherein at least one of the one or more heterologous nucleic acids is integrated into the chromosome of the genetically modified host cell.

46. The genetically modified host cell of any one of claims 1-44, wherein at least one of the one or more heterologous nucleic acids is maintained extrachromosomally.

47. The genetically modified host cell of any one of claims 1-44, wherein two or more of the one or more heterologous nucleic acids are present in a single expression vector.

48. The genetically modified host cell of any one of claims 1-44, wherein at least one of the heterologous nucleic acids is operably linked to an inducible promoter.

49. The genetically modified host cell of any one of claims 1-44, wherein at least one of the heterologous nucleic acids is operably linked to a constitutive promoter.

50. The genetically modified host cell of any one of claims 1-49, wherein culturing of the genetically modified host cell in a suitable medium provides for synthesis of the cannabinoid or the cannabinoid derivative in an increased amount compared to a non-genetically modified host cell cultured under similar conditions.

51. The genetically modified host cell of any one of claims 1-50, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a cannabinoid synthase polypeptide.

52. The genetically modified host cell of claim 51, wherein the cannabinoid synthase polypeptide is a tetrahydrocannabinolic acid (THCA) synthase polypeptide.

53. The genetically modified host cell of claim 52, wherein the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155.

54. The genetically modified host cell of claim 51, wherein the cannabinoid synthase polypeptide is a cannabidiolic acid (CBDA) synthase polypeptide.

55. The genetically modified host cell of claim 54, wherein the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

56. The genetically modified host cell of any one of claims 1-55, wherein the cannabinoid is cannabigerolic acid, cannabigerol, Δ^9-tetrahydrocannabinolic acid, Δ^9-tetrahydrocannabinol, Δ^8-tetrahydrocannabinolic acid, Δ^8-tetrahydrocannabinol, cannabidiolic acid, cannabidiol, cannabichromenic acid, cannabichromene, cannabinolic acid, cannabinol, cannabidivarinic acid, cannabidivarin, tetrahydrocannabivarinic acid, tetrahydrocannabivarin, cannabichromevarinic acid, cannabichromevarin, cannabigerovarinic acid, cannabigerovarin, cannabicyclolic acid, cannabicyclol, cannabielsoinic acid, cannabielsoin, cannabicitranic acid, or cannabicitran.

57. The genetically modified host cell of claim 56, wherein the cannabinoid is cannabigerolic acid.

58. A method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising:

a) culturing the genetically modified host cell of any one of claims 1-57 in a suitable medium; and

b) recovering the produced cannabinoid or cannabinoid derivative.

59. A method of producing a cannabinoid or a cannabinoid derivative, the method comprising:

a) culturing the genetically modified host cell of any one of claims 1-57 in a suitable medium comprising a carboxylic acid;

b) recovering the produced cannabinoid or cannabinoid derivative.

60. A method of producing a cannabinoid or a cannabinoid derivative, the method comprising:

a) culturing the genetically modified host cell of any one of claims 1-57 in a suitable medium comprising olivetolic acid or an olivetolic acid derivative;

b) recovering the produced cannabinoid or cannabinoid derivative.

61. A method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising:

a) culturing a genetically modified host cell comprising one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide, wherein said geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide can catalyze production of cannabigerolic acid from geranyl pyrophosphate and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, in a suitable medium; and

b) recovering the produced cannabinoid or cannabinoid derivative.

62. A method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising:

a) culturing a genetically modified host cell comprising one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110 in a suitable medium; and

b) recovering the produced cannabinoid or cannabinoid derivative.

63. A method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising:

a) culturing a genetically modified host cell comprising one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100 in a suitable medium; and

b) recovering the produced cannabinoid or cannabinoid derivative.

64. The method of any one of claims 58-63, wherein the suitable medium comprises a fermentable sugar.

65. The method of any one of claims 58-63, wherein the suitable medium comprises a pretreated cellulosic feedstock.

66. The method of any one of claims 58-63, wherein the suitable medium comprises a non-fermentable carbon source.

67. The method of claim 66, wherein the non-fermentable carbon source comprises ethanol.

68. The method of any one of claims 58-67, wherein the cannabinoid is cannabigerolic acid, cannabigerol, Δ^9-tetrahydrocannabinolic acid, Δ^9-tetrahydrocannabinol, Δ^8-tetrahydrocannabinolic acid, Δ^8-tetrahydrocannabinol, cannabidiolic acid, cannabidiol, cannabichromenic acid, cannabichromene, cannabinolic acid, cannabinol, cannabidivarinic acid, cannabidivarin, tetrahydrocannabivarinic acid, tetrahydrocannabivarin, cannabichromevarinic acid, cannabichromevarin, cannabigerovarinic acid, cannabigerovarin, cannabicyclolic acid, cannabicyclol, cannabielsoinic acid, cannabielsoin, cannabicitranic acid, or cannabicitran.

69. An isolated or purified geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide, wherein said geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide can catalyze production of cannabigerolic acid from geranyl pyrophosphate and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

70. An isolated or purified polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110.

71. An isolated or purified polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100.

72. An isolated or purified nucleic acid encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide, wherein said geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide can catalyze production of cannabigerolic acid from geranyl pyrophosphate and olivetolic acid in an amount at least ten

times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

73. An isolated or purified nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110.

74. An isolated or purified nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100.

75. A vector comprising a nucleic acid encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide, wherein said geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide can catalyze production of cannabigerolic acid from geranyl pyrophosphate and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

76. A vector comprising a nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110.

77. A vector comprising a nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100.

78. A method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide, wherein said geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide can catalyze production of cannabigerolic acid from geranyl pyrophosphate and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, into the genetically modified host cell.

79. A method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase

polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110 into the genetically modified host cell.

80. A method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100 into the genetically modified host cell.

81. A method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing the vector of any one of claims 75-77 into the genetically modified host cell.

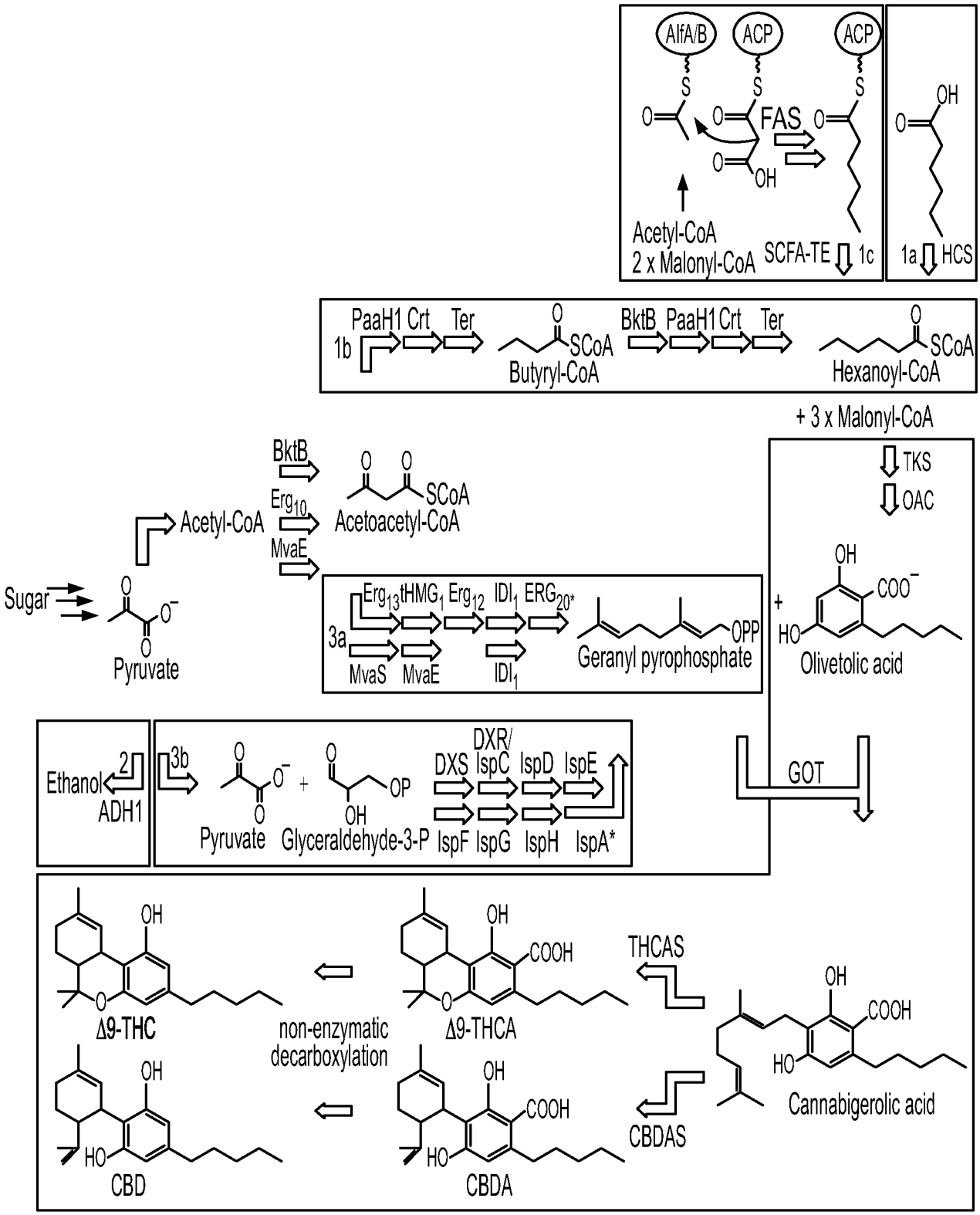

AlfA/B
ACP
ACP
FAS
Acetyl-CoA
2 x Malonyl-CoA
SCFA-TE
1c
1a
HCS
PaaH1 Crt Ter
1b
Butyryl-CoA
BktB PaaH1 Crt Ter
Hexanoyl-CoA
+ 3 x Malonyl-CoA
TKS
OAC
Olivetolic acid
BktB
Erg10
MvaE
Acetoacetyl-CoA
Acetyl-CoA
Sugar
Pyruvate
Erg13 tHMG1 Erg12 IDI1 ERG20*
3a
MvaS MvaE IDI1
Geranyl pyrophosphate
Ethanol
2
ADH1
3b
Pyruvate
Glyceraldehyde-3-P
DXS DXR/ IspC IspD IspE
IspF IspG IspH IspA*
GOT
Δ9-THC
non-enzymatic decarboxylation
Δ9-THCA
THCAS
Cannabigerolic acid
CBD
CBDA
CBDAS

FIG. 1

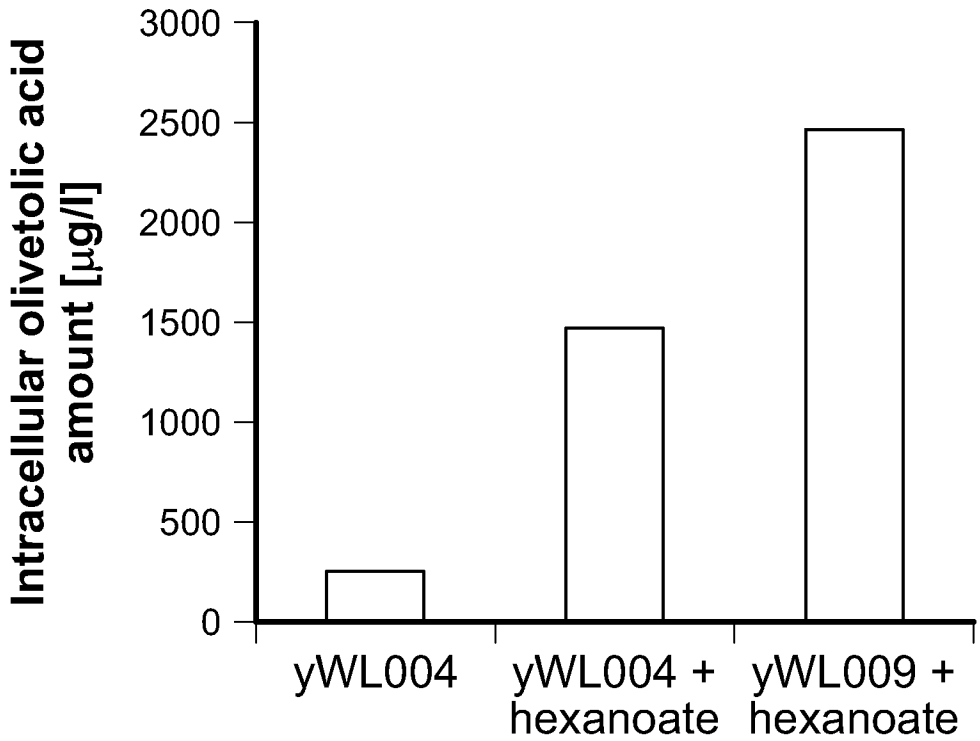

Intracellular olivetolic acid amount [μg/l]
3000
2500
2000
1500
1000
500
0
yWL004
yWL004 + hexanoate
yWL009 + hexanoate

FIG. 2

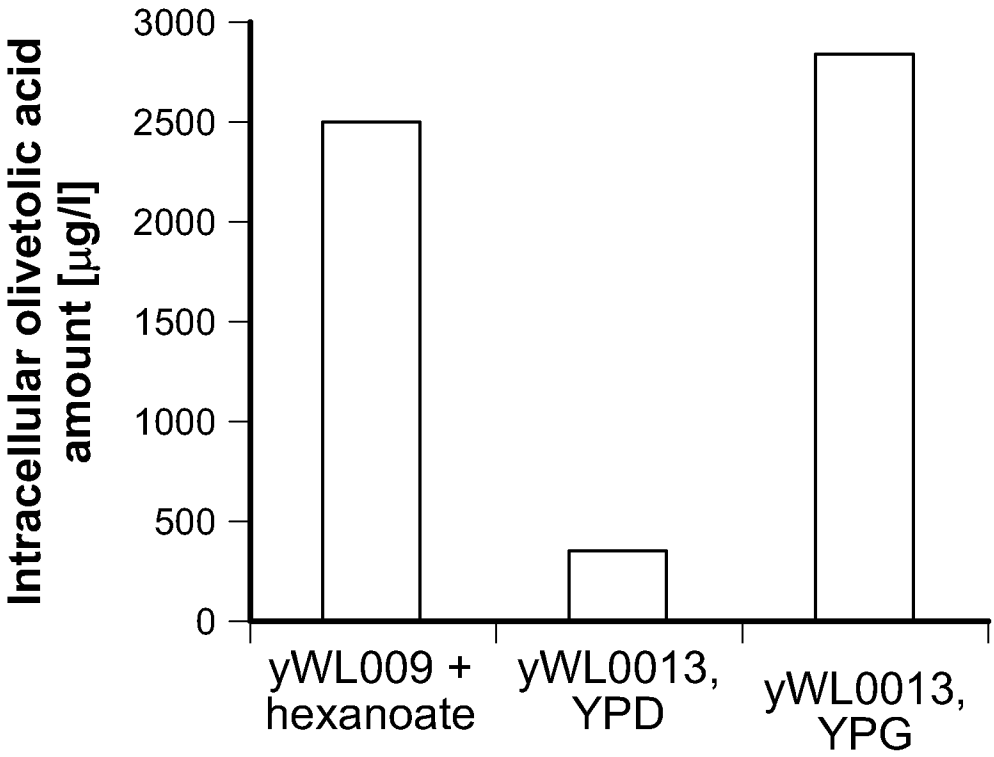

Intracellular olivetolic acid amount [μg/l]
3000
2500
2000
1500
1000
500
0
yWL009 + hexanoate
yWL0013, YPD
yWL0013, YPG

FIG. 3

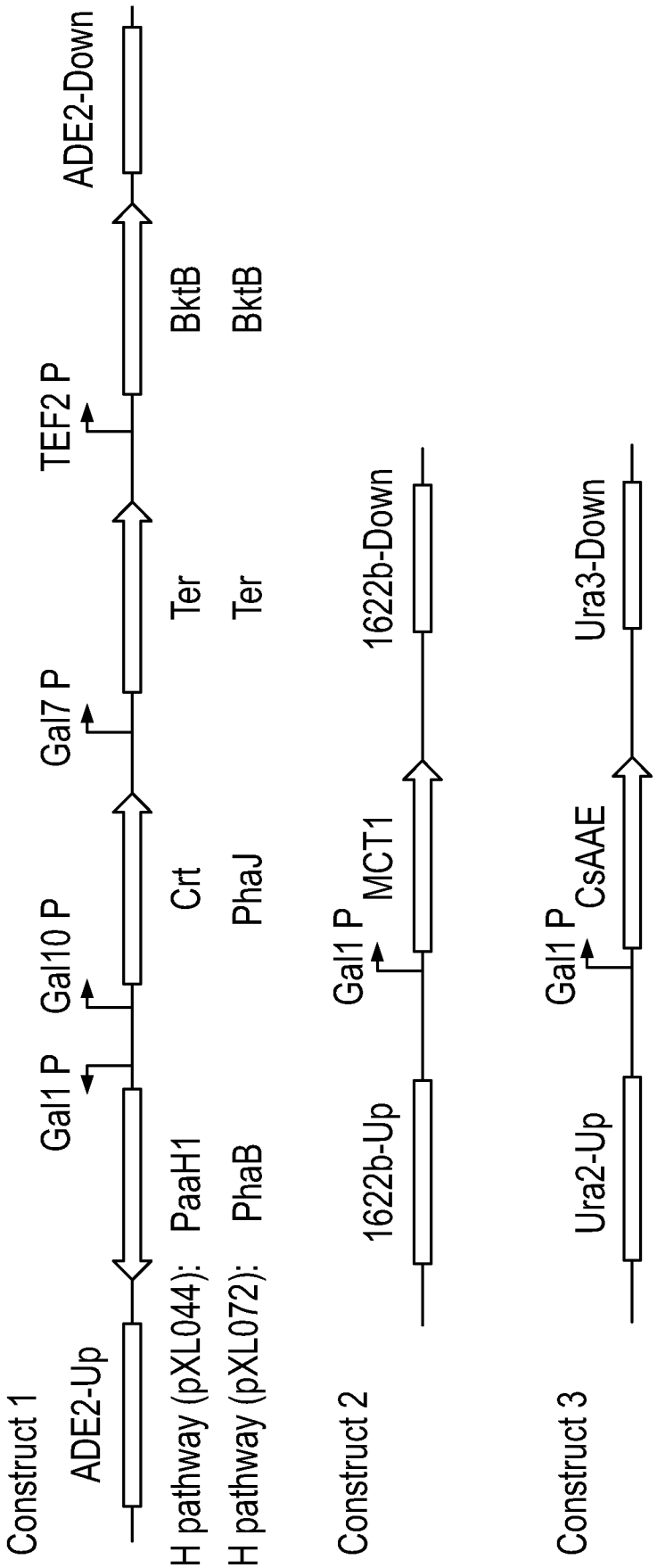

Construct 1
ADE2-Up
Gal1 P
Gal10 P
Gal7 P
TEF2 P
ADE2-Down
NADH pathway (pXL044): PaaH1 Crt Ter BktB
NADPH pathway (pXL072): PhaB PhaJ Ter BktB
Construct 2
1622b-Up
Gal1 P
MCT1
1622b-Down
Construct 3
Ura2-Up
Gal1 P
CsAAE
Ura3-Down

FIG. 4

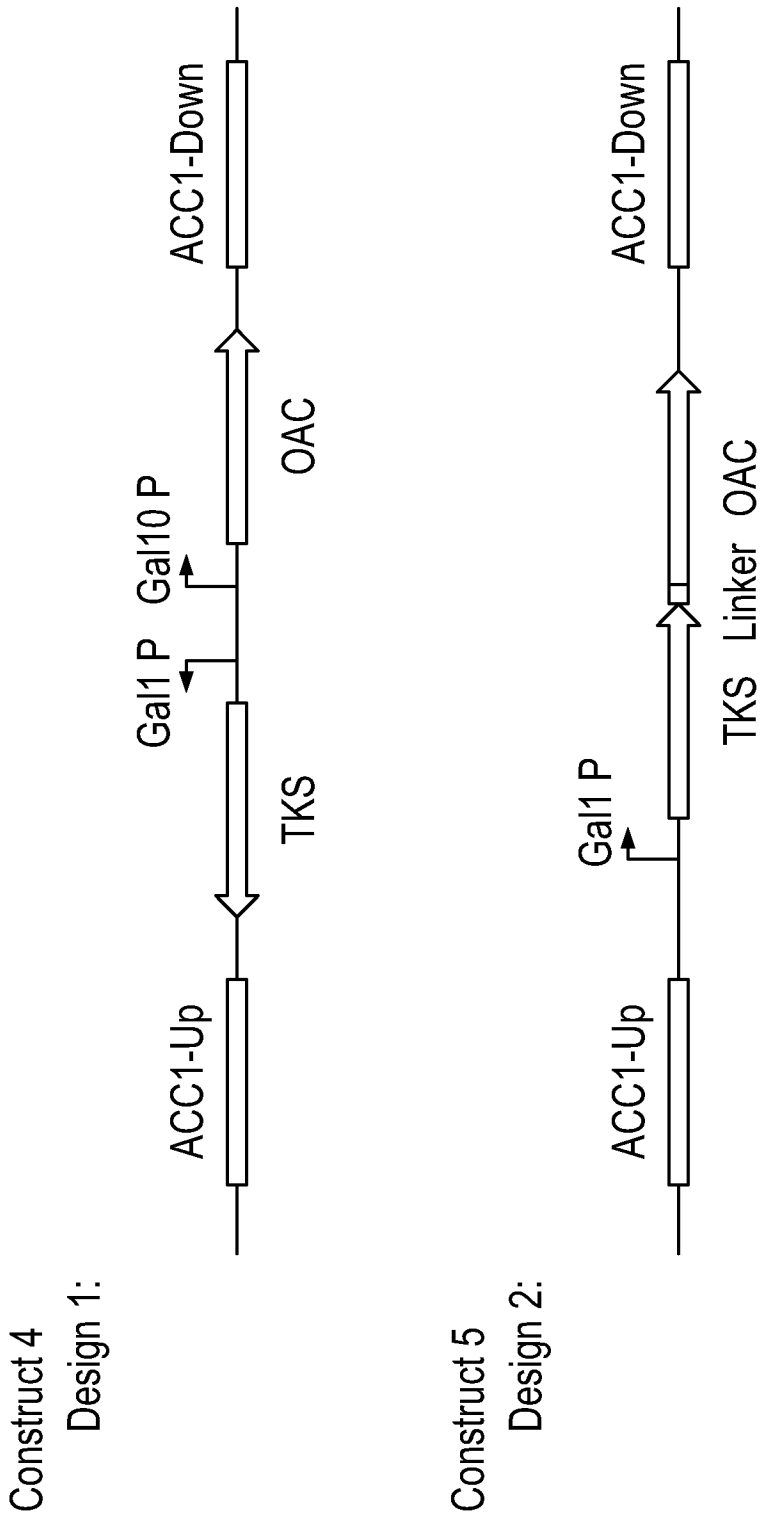

Construct 4
Design 1:
ACC1-Up
TKS
Gal1 P
Gal10 P
OAC
ACC1-Down
Construct 5
Design 2:
ACC1-Up
Gal1 P
TKS
Linker
OAC
ACC1-Down

FIG. 5

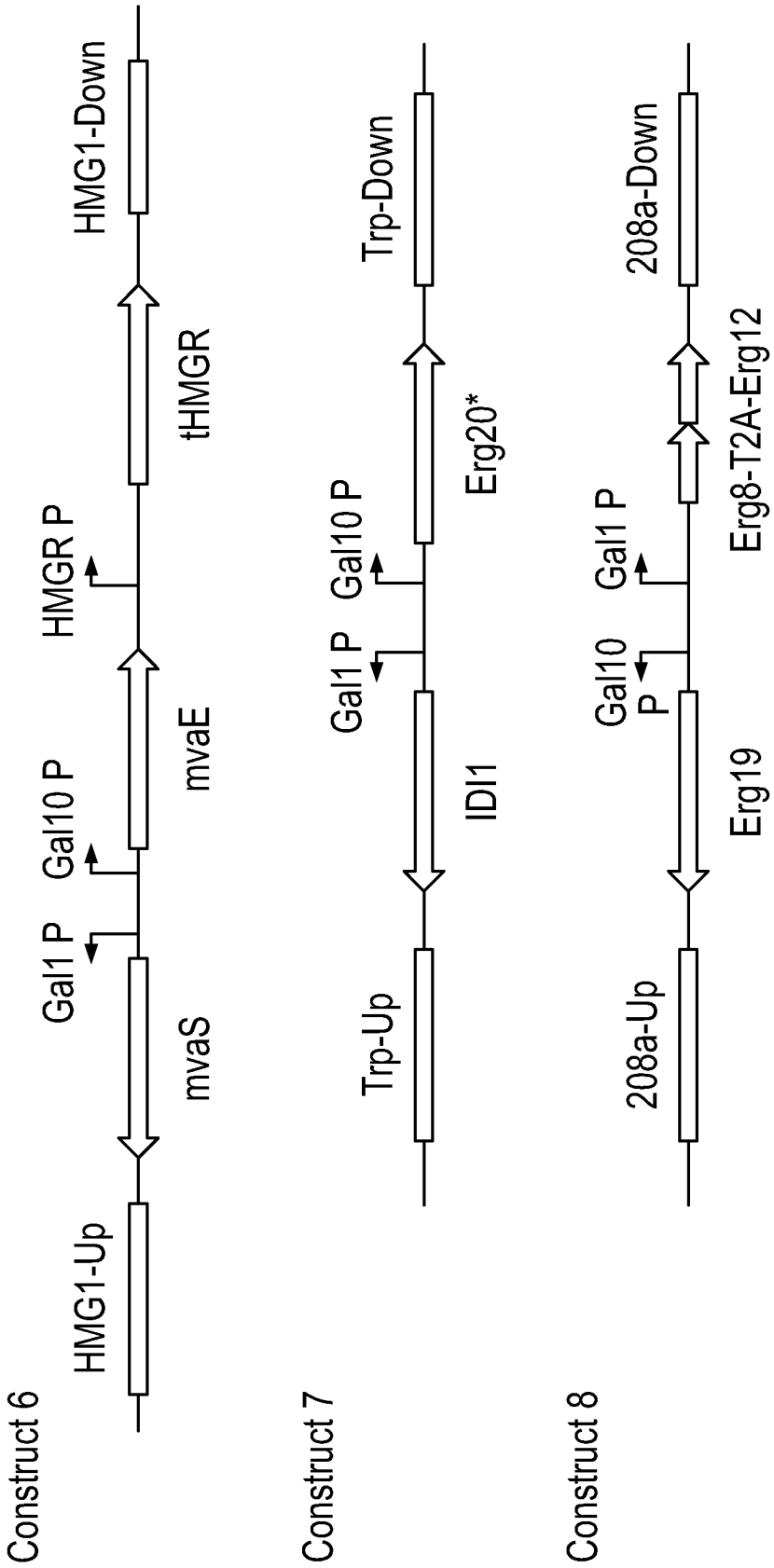

Construct 6
HMG1-Up
Gal1 P
Gal10 P
mvaS
mvaE
HMGR P
tHMGR
HMG1-Down
Construct 7
Trp-Up
Gal1 P
Gal10 P
IDI1
Erg20*
Trp-Down
Construct 8
208a-Up
Gal10 P
Gal1 P
Erg19
Erg8-T2A-Erg12
208a-Down

FIG. 6

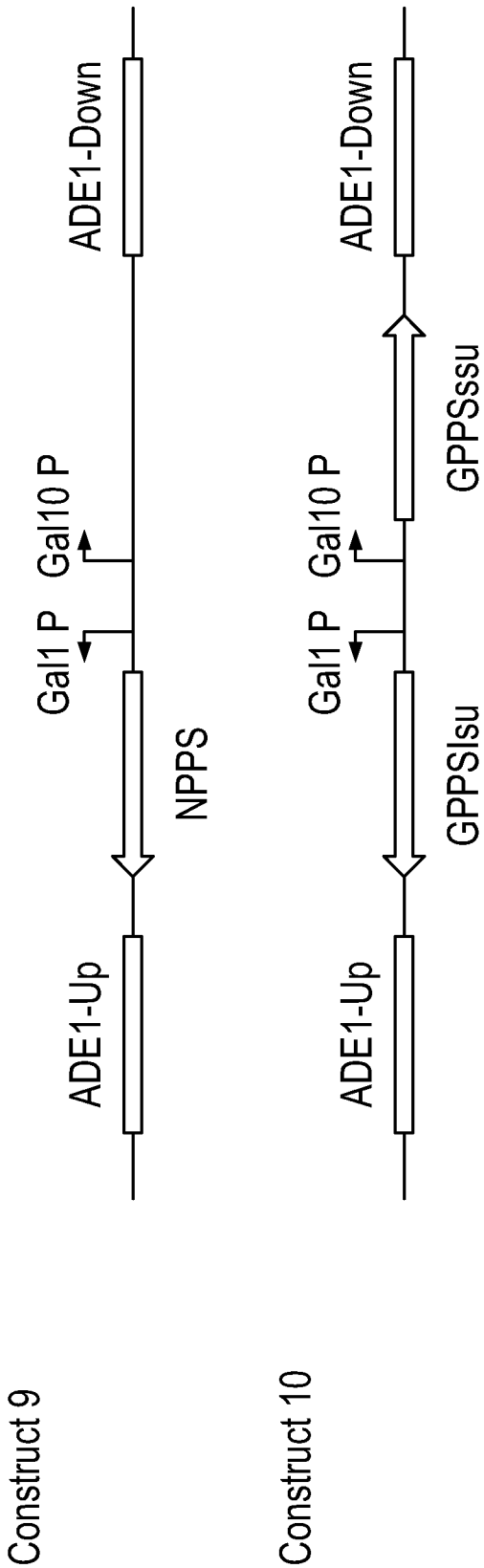

ADE1-Down
Gal10 P
Gal1 P
NPPS
ADE1-Up
Construct 9
ADE1-Down
GPPSssu
Gal10 P
Gal1 P
GPPSlsu
ADE1-Up
Construct 10

FIG. 7

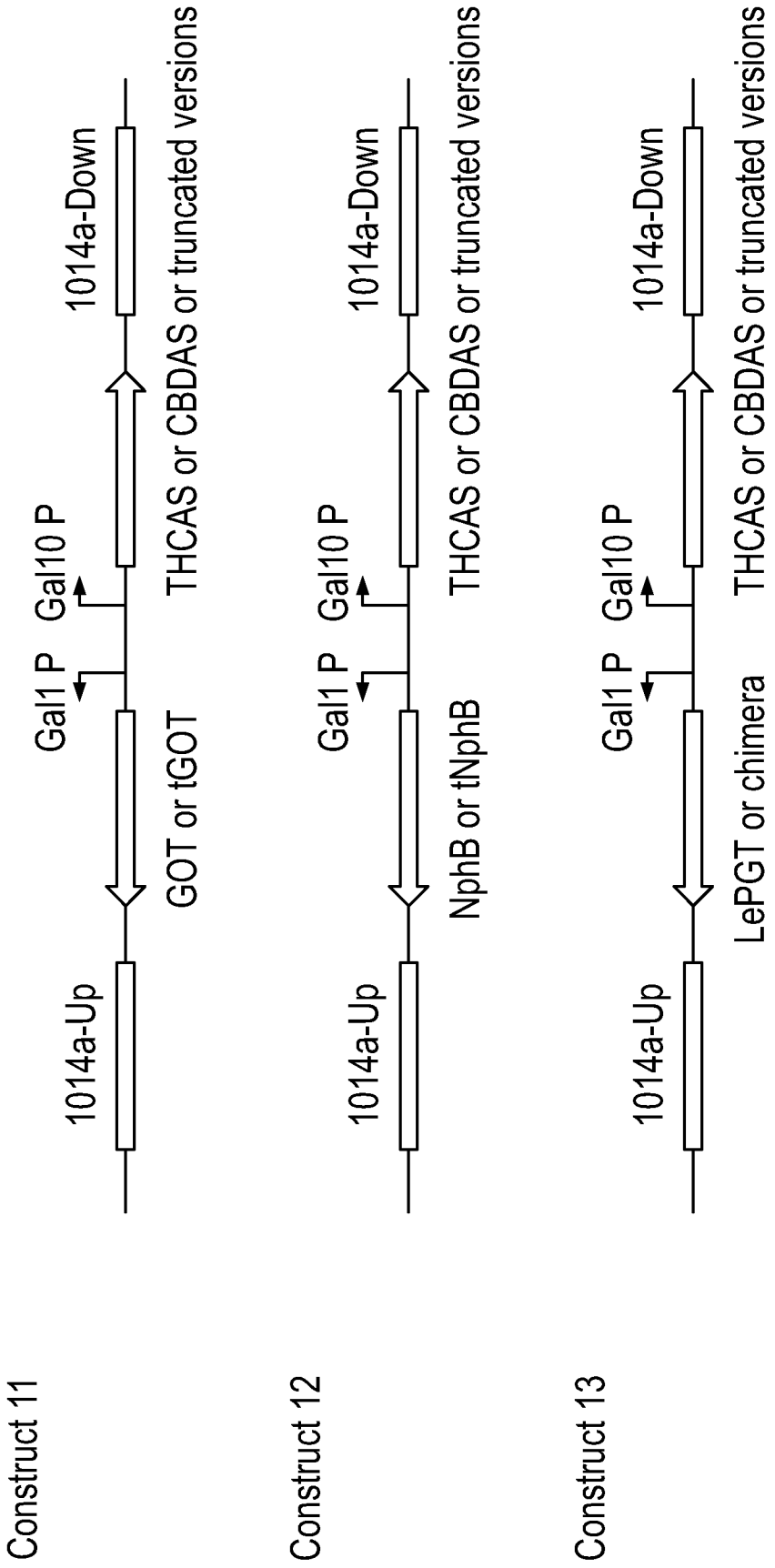

Construct 11
1014a-Up
Gal1 P
GOT or tGOT
Gal10 P
THCAS or CBDAS or truncated versions
1014a-Down
Construct 12
1014a-Up
Gal1 P
NphB or tNphB
Gal10 P
THCAS or CBDAS or truncated versions
1014a-Down
Construct 13
1014a-Up
Gal1 P
LePGT or chimera
Gal10 P
THCAS or CBDAS or truncated versions
1014a-Down

FIG. 8

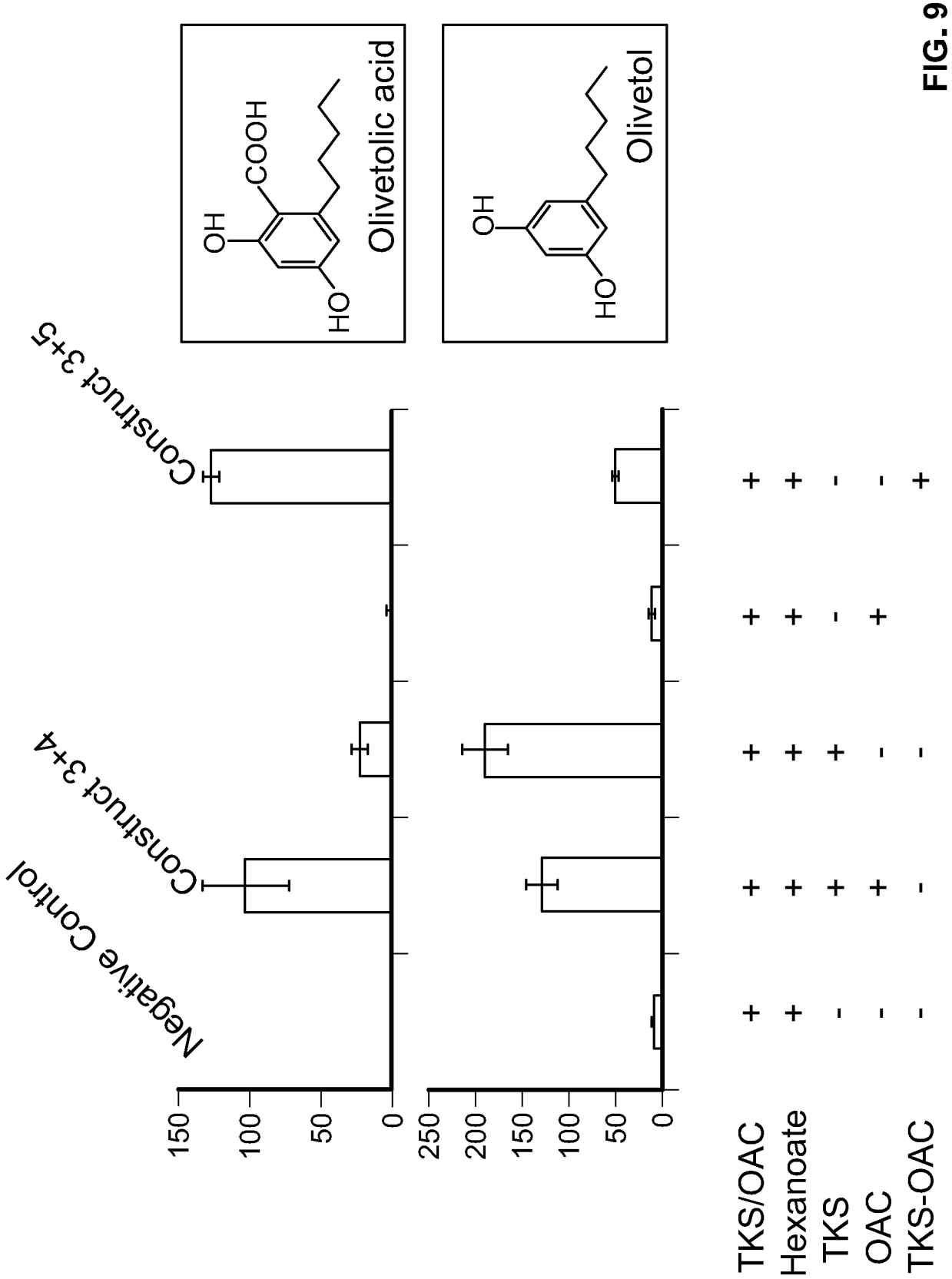

Olivetolic acid
COOH
OH
HO
Olivetol
OH
HO
Negative Control
Construct 3+4
Construct 3+5
150
100
50
0
250
200
150
100
50
0
TKS/OAC
Hexanoate
TKS
OAC
TKS-OAC
+ + - - +
+ + + - -
+ + - + -
+ + - - -

FIG. 9

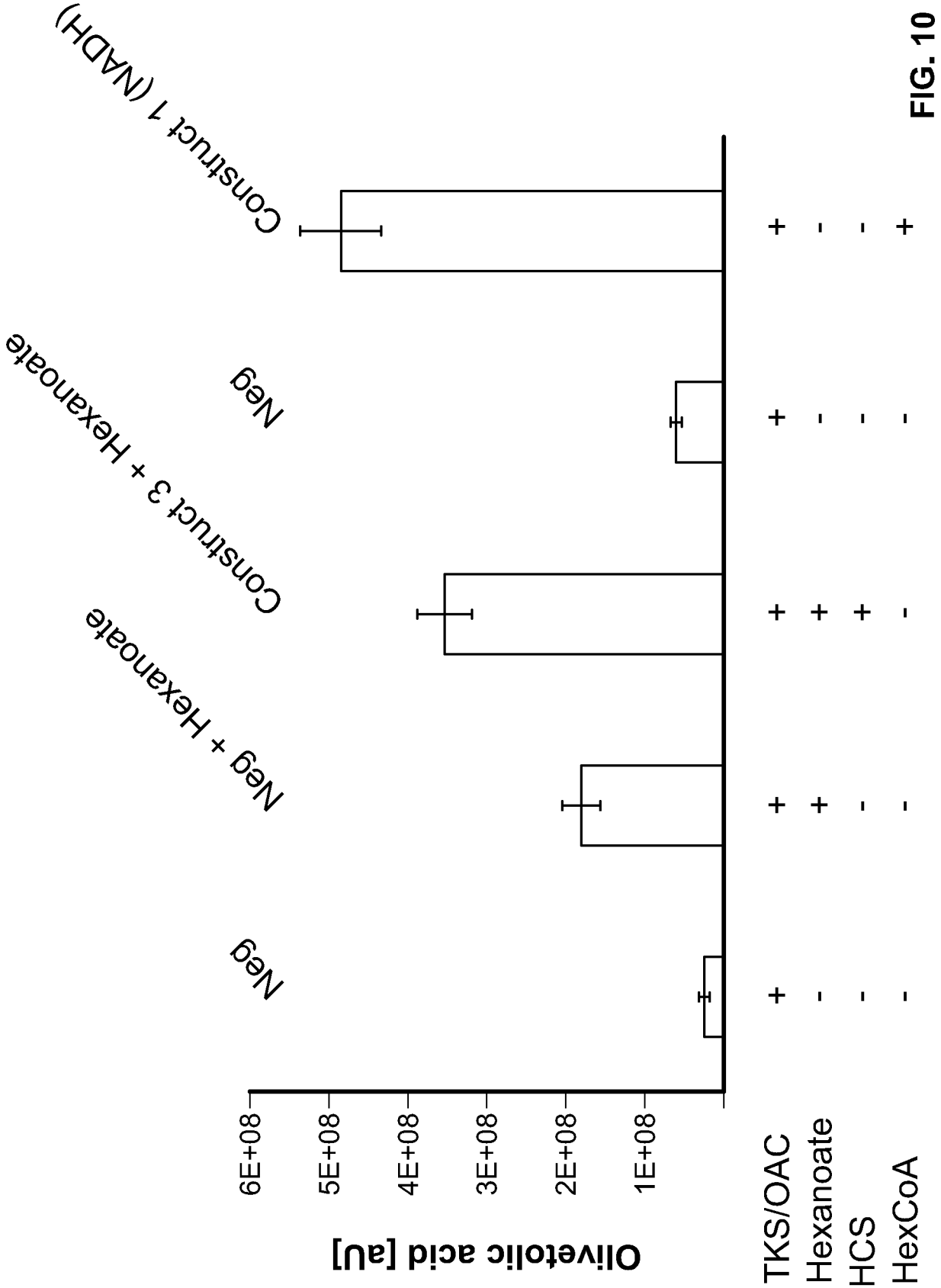

Oliveto lic acid [aU]
6E+08
5E+08
4E+08
3E+08
2E+08
1E+08
TKS/OAC
Hexanoate
HCS
HexCoA
+ - - -
+ + - -
+ + + -
+ - - -
Neg
Neg + Hexanoate
Construct 3 + Hexanoate
Construct 1 (NADH)

FIG. 10

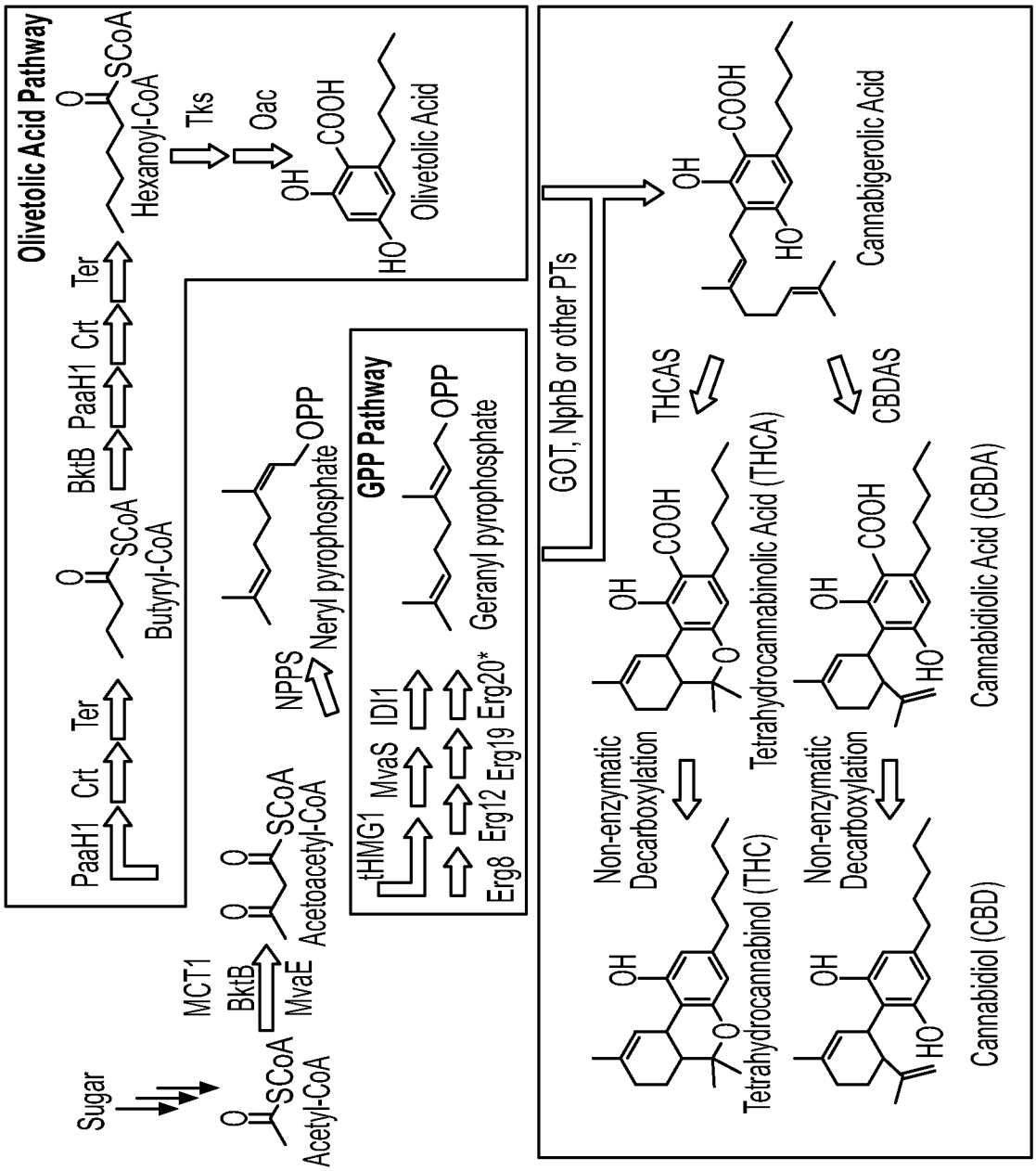

Olivetolic Acid Pathway
PaaH1
Crt
Ter
BktB
PaaH1
Crt
Ter
Butyryl-CoA
Hexanoyl-CoA
Tks
Oac
Olivetolic Acid
GPP Pathway
NPPS
Neryl pyrophosphate
Geranyl pyrophosphate
tHMG1
MvaS
IDI1
Erg8
Erg12
Erg19
Erg20*
Sugar
MCT1
Acetyl-CoA
BktB
MvaE
Acetoacetyl-CoA
GOT, NphB or other PTs
THCAS
CBDAS
Cannabigerolic Acid
Tetrahydrocannabinolic Acid (THCA)
Cannabidiolic Acid (CBDA)
Non-enzymatic Decarboxylation
Tetrahydrocannabinol (THC)
Cannabidiol (CBD)

FIG. 11

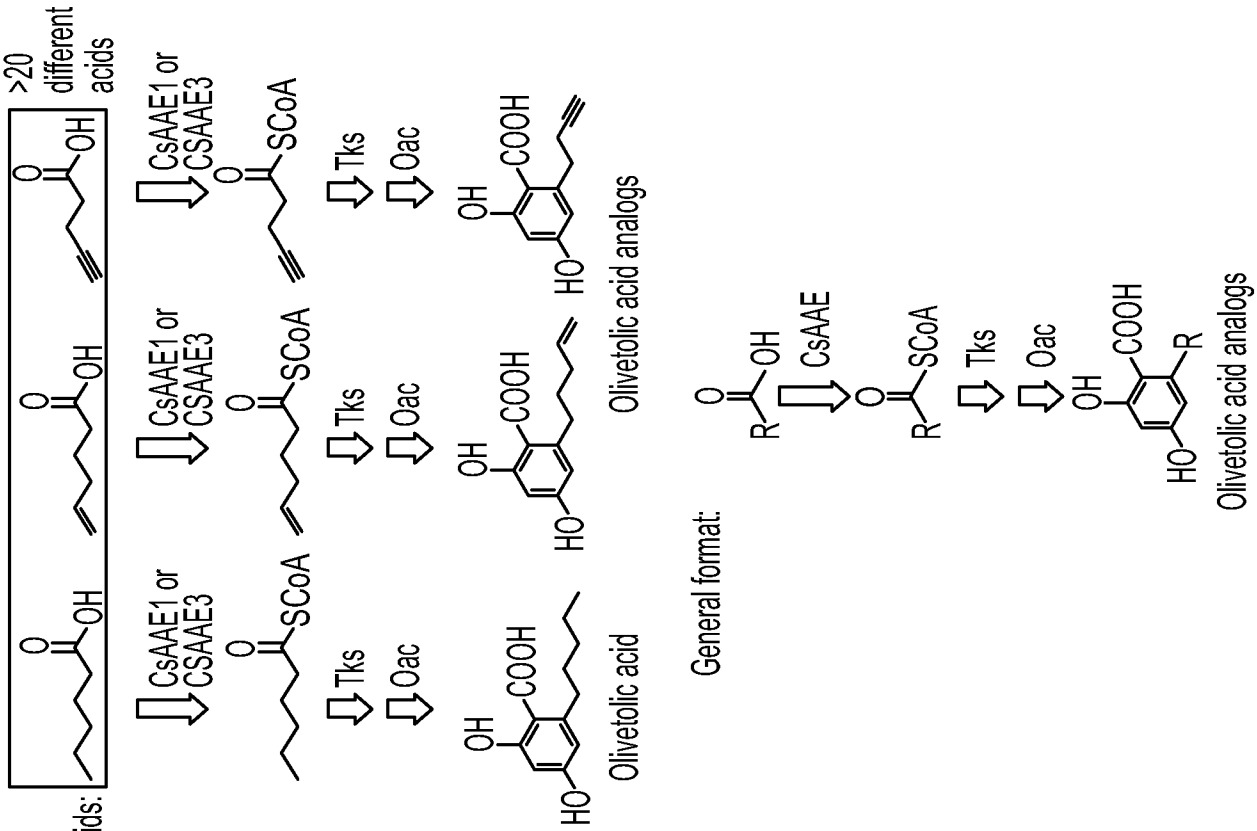

Feeding acids:
>20 different acids
CsAAE1 or CSAAE3
Tks
Oac
Olivetolic acid
Olivetolic acid analogs
General format:
CsAAE
Olivetolic acid analogs

FIG. 11 (Cont.)

12/101

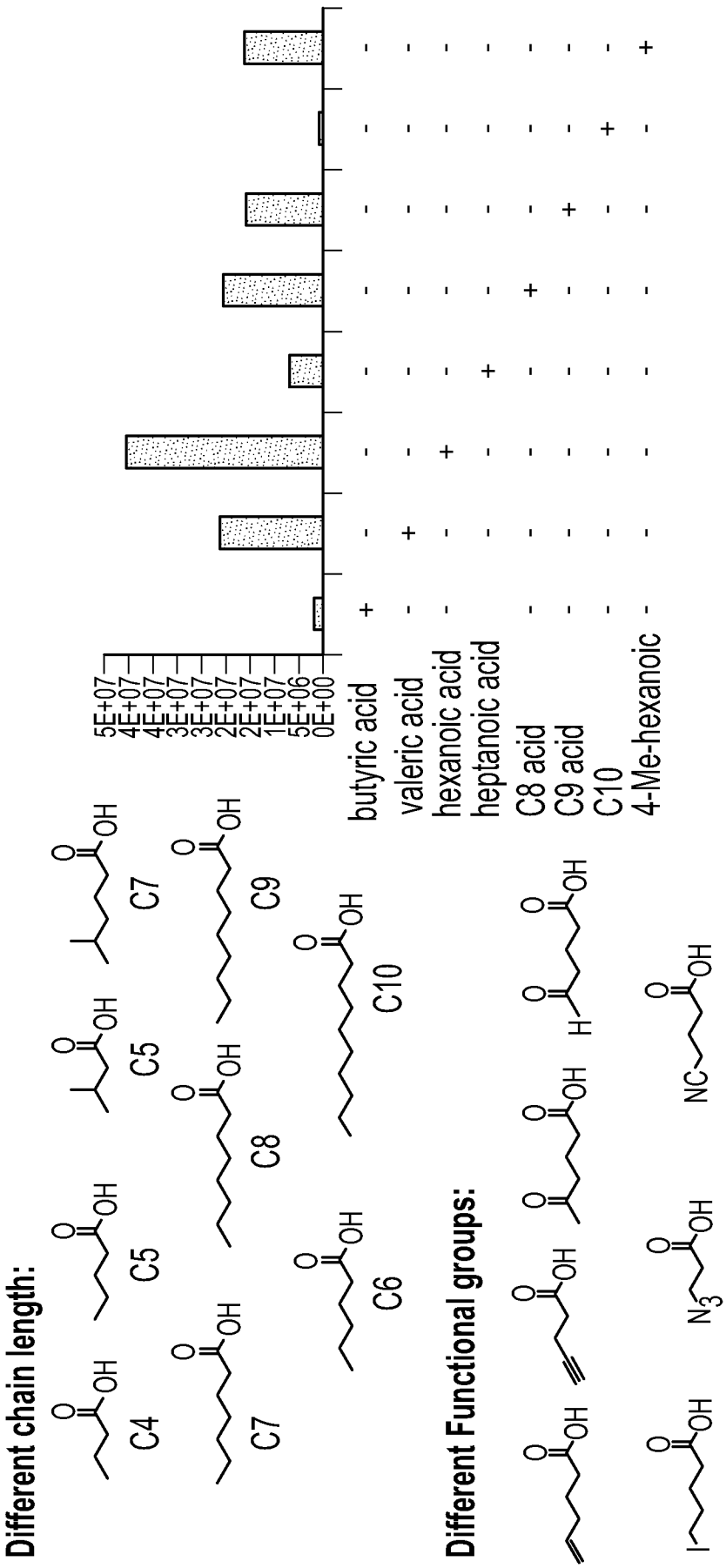

Different chain length:
C4
C5
C5
C7
C7
C8
C9
C6
C10
Different Functional groups:
N3
NC
5E+07
4E+07
4E+07
3E+07
3E+07
2E+07
2E+07
1E+07
5E+06
0E+00
butyric acid
valeric acid
hexanoic acid
heptanoic acid
C8 acid
C9 acid
C10
4-Me-hexanoic

FIG. 12

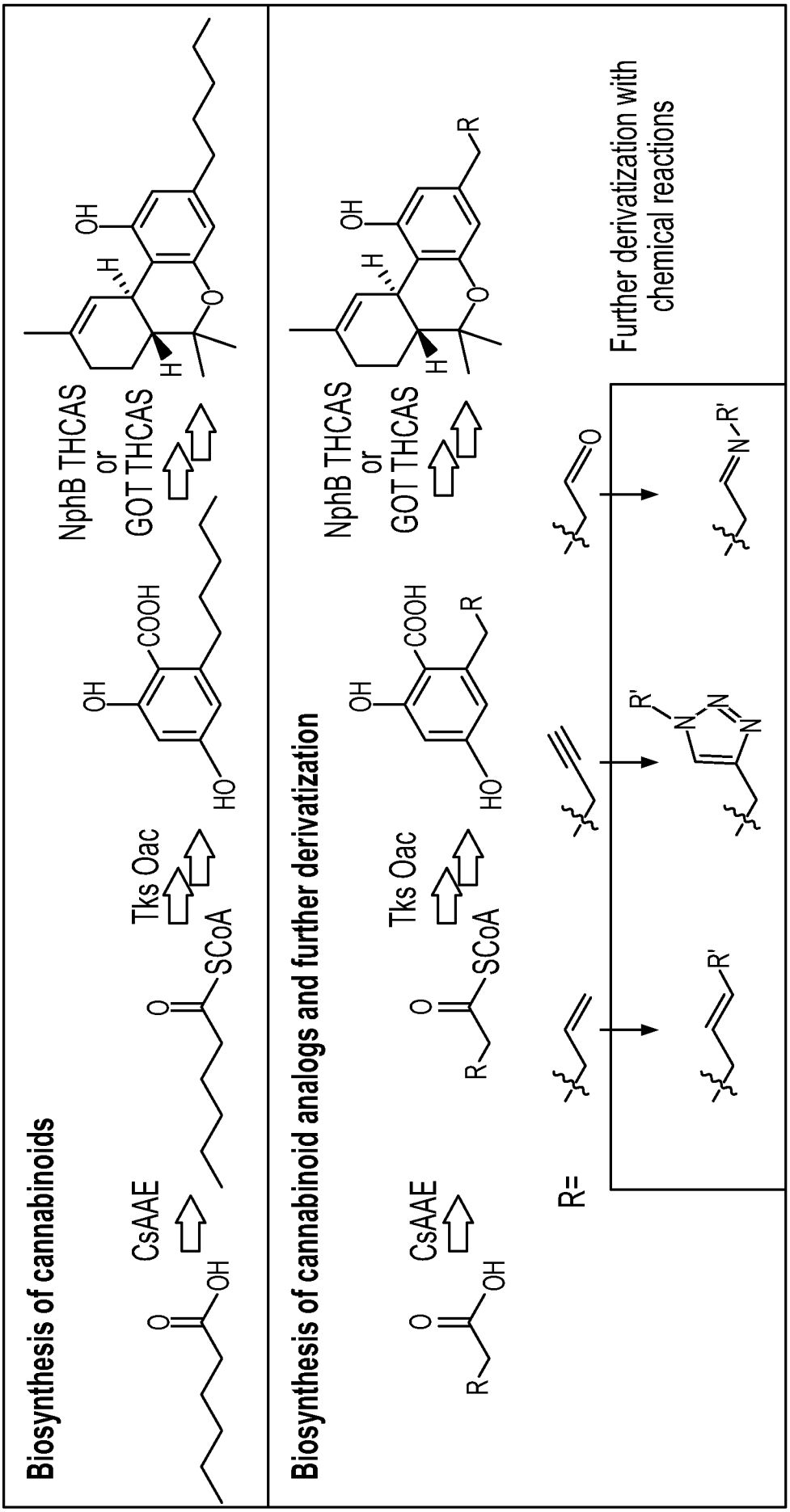

Biosynthesis of cannabinoids
CsAAE
Tks Oac
NphB THCAS
or
GOT THCAS
Biosynthesis of cannabinoid analogs and further derivatization
CsAAE
Tks Oac
NphB THCAS
or
GOT THCAS
Further derivatization with
chemical reactions
R=

FIG. 13

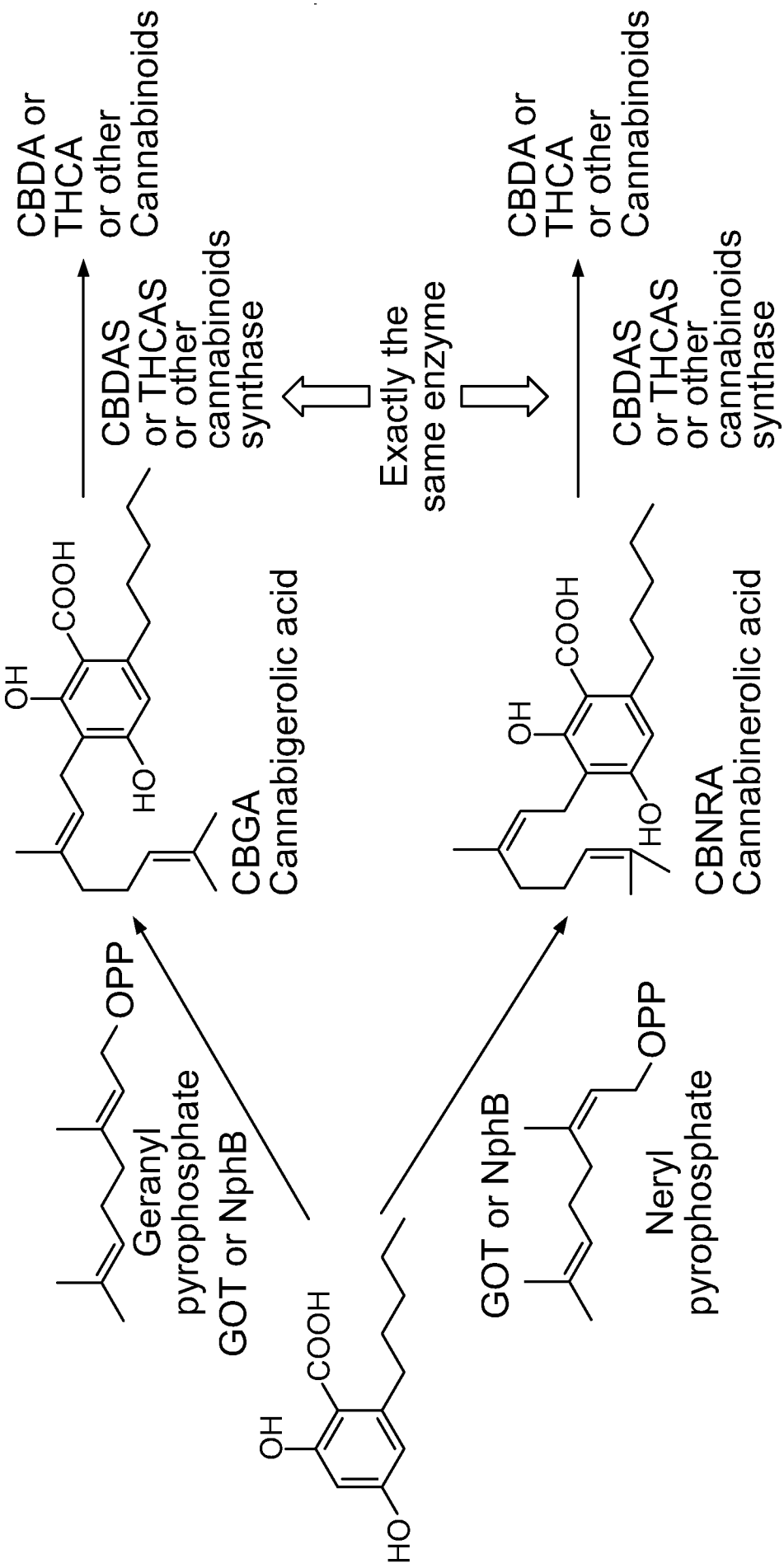

CBDA or
THCA
or other
Cannabinoids
CBDAS
or THCAS
or other
cannabinoids
synthase
Exactly the
same enzyme
CBDA or
THCA
or other
Cannabinoids
CBDAS
or THCAS
or other
cannabinoids
synthase
COOH
OH
HO
CBGA
Cannabigerolic acid
CBNRA
Cannabinerolic acid
OPP
Geranyl
pyrophosphate
GOT or NphB
GOT or NphB
Neryl
pyrophosphate

FIG. 14

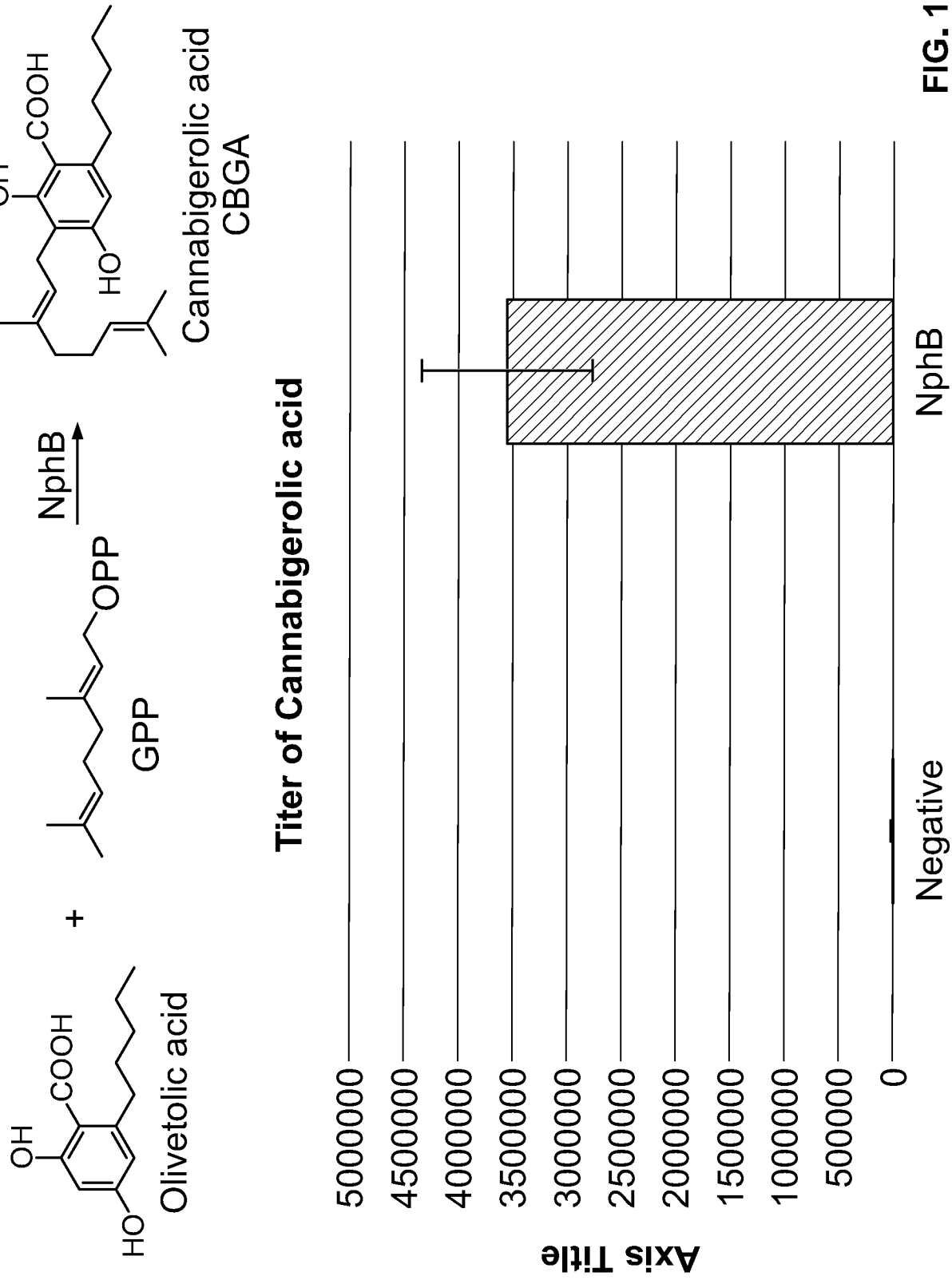

Olivetolic acid
+
GPP
OPP
NphB
Cannabigerolic acid
CBGA
OH
COOH
HO
Titer of Cannabigerolic acid
Axis Title
5000000
4500000
4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
0
Negative
NphB

FIG. 15

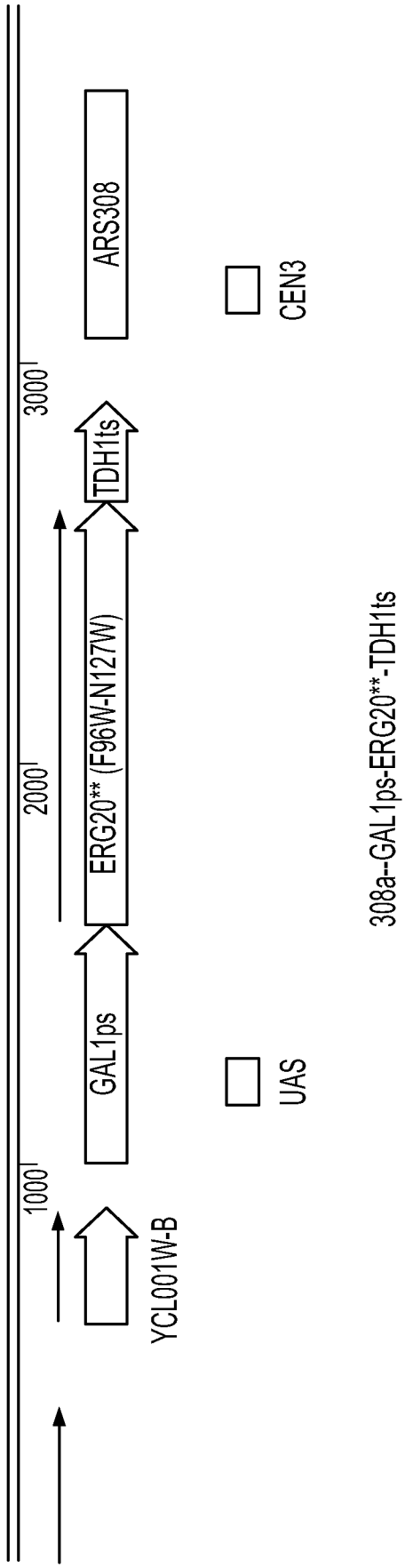

1000
2000
3000
YCL001W-B
GAL1ps
ERG20** (F96W-N127W)
TDH1ts
ARS308
UAS
CEN3
308a--GAL1ps-ERG20**-TDH1ts

FIG. 16

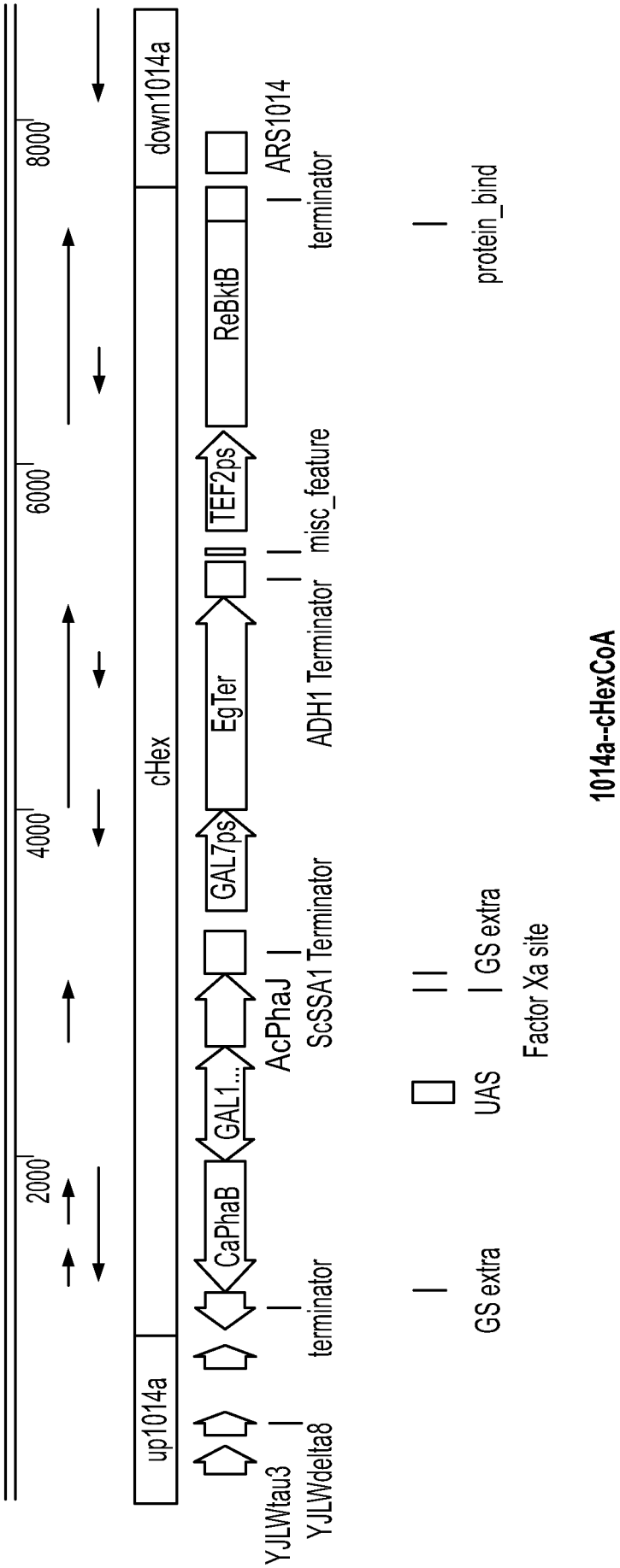

2000
4000
6000
8000
up1014a
cHex
down1014a
YJLWtau3
YJLWdelta8
terminator
CaPhaB
GAL1...
AcPhaJ
ScSSA1 Terminator
GAL7ps
EgTer
ADH1 Terminator
TEF2ps
misc_feature
ReBktB
terminator
ARS1014
GS extra
UAS
GS extra
Factor Xa site
protein_bind
1014a--cHexCoA

FIG. 17

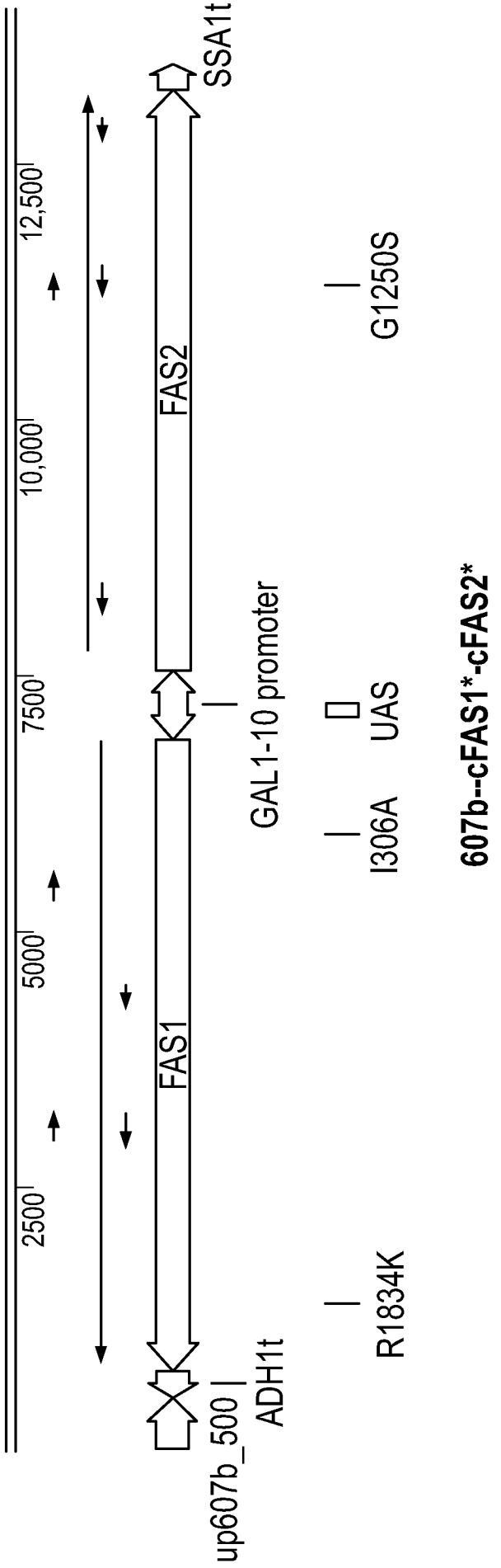
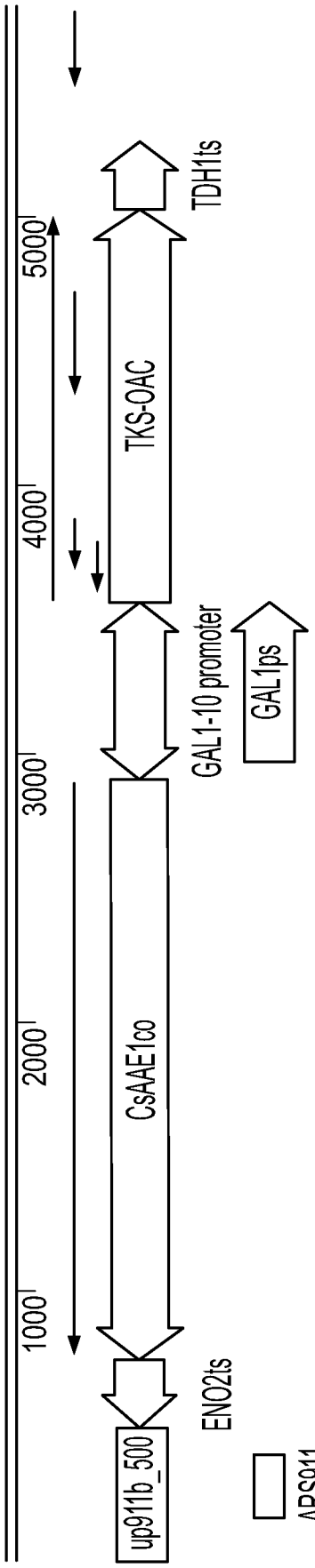
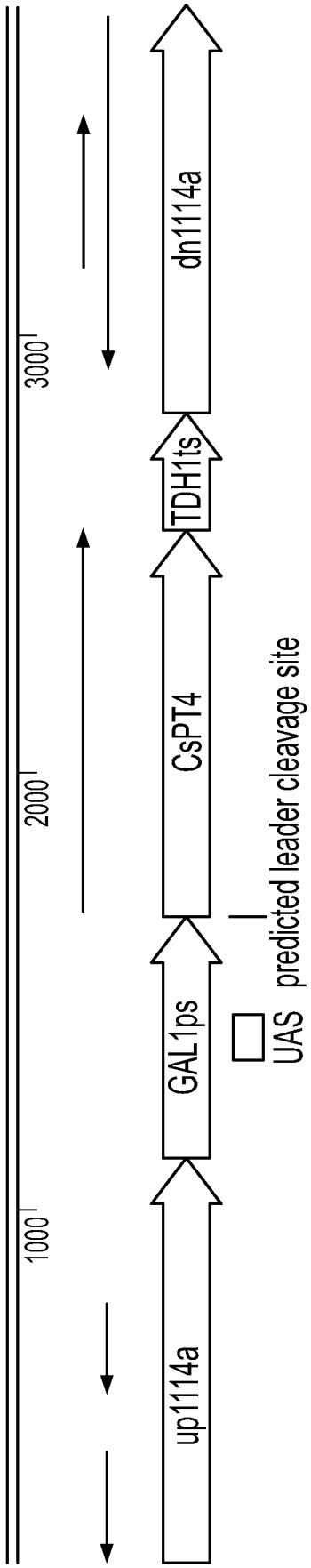

2500
5000
7500
10,000
12,500
FAS1
FAS2
SSA1t
ADH1t
up607b_500
GAL1-10 promoter
UAS
R1834K
I306A
G1250S
607b--cFAS1*-cFAS2*

FIG. 18



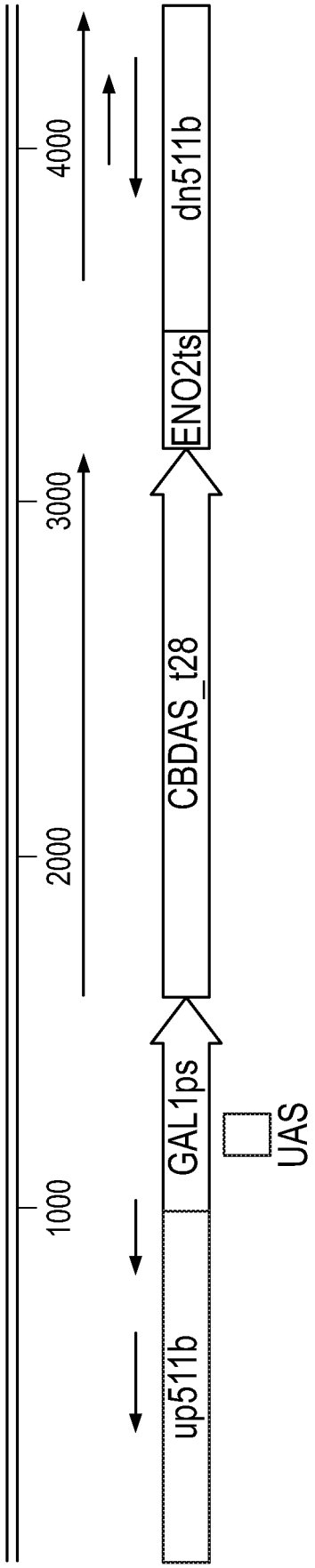
911b--cAAE1--cTKS-OAC

FIG. 19



1114a--tCsPT4

FIG. 20



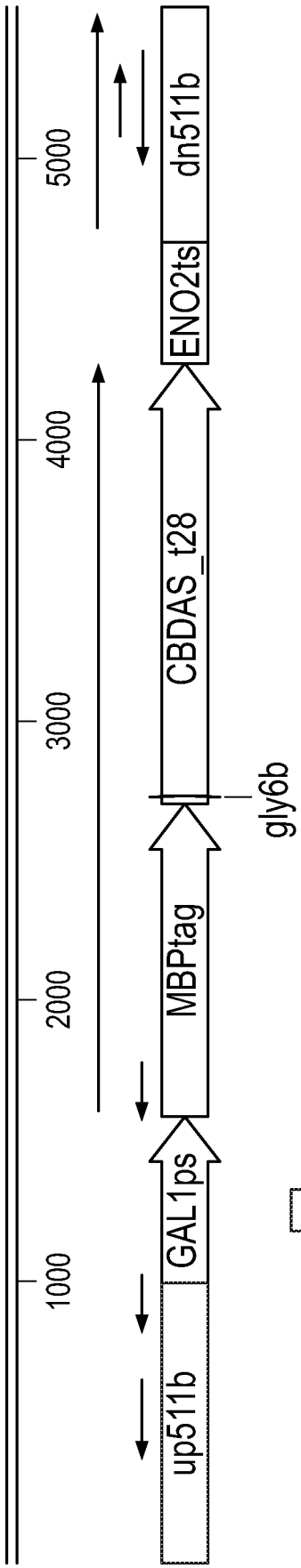
511b-GAL1ps-CBDASt28-ENO2ts

FIG. 21



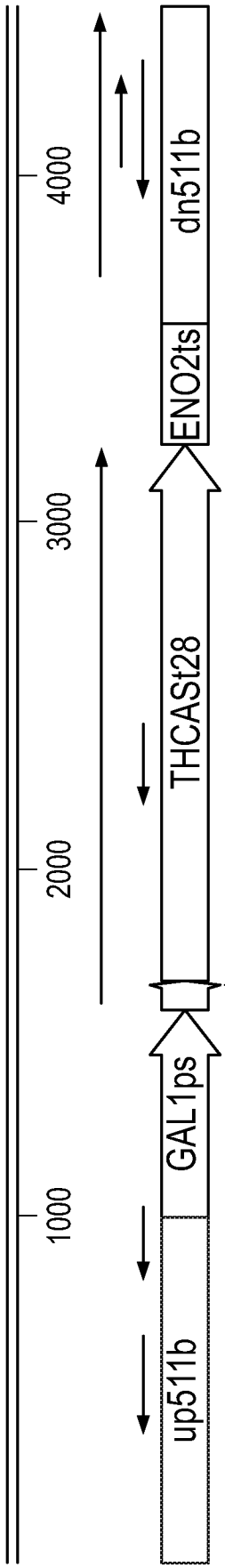
511b--GAL1ps-ProA-CBDAS-ENO2ts

FIG. 22



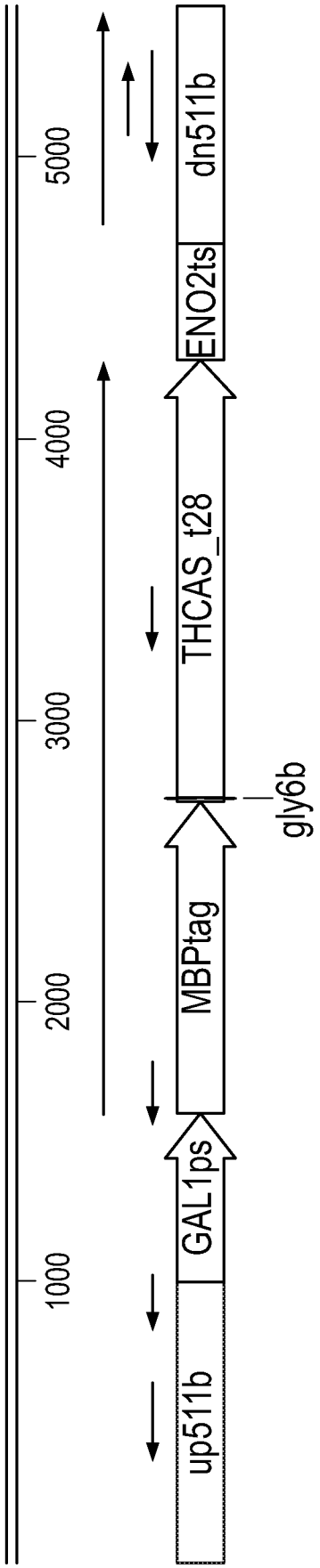
511b--GAL1ps-MBP-CBDAS-ENO2ts

FIG. 23



511b--GAL1ps-ProA-THCAS-ENO2ts

FIG. 24



511b--GAL1ps-MBP-THCAS-ENO2ts

FIG. 25

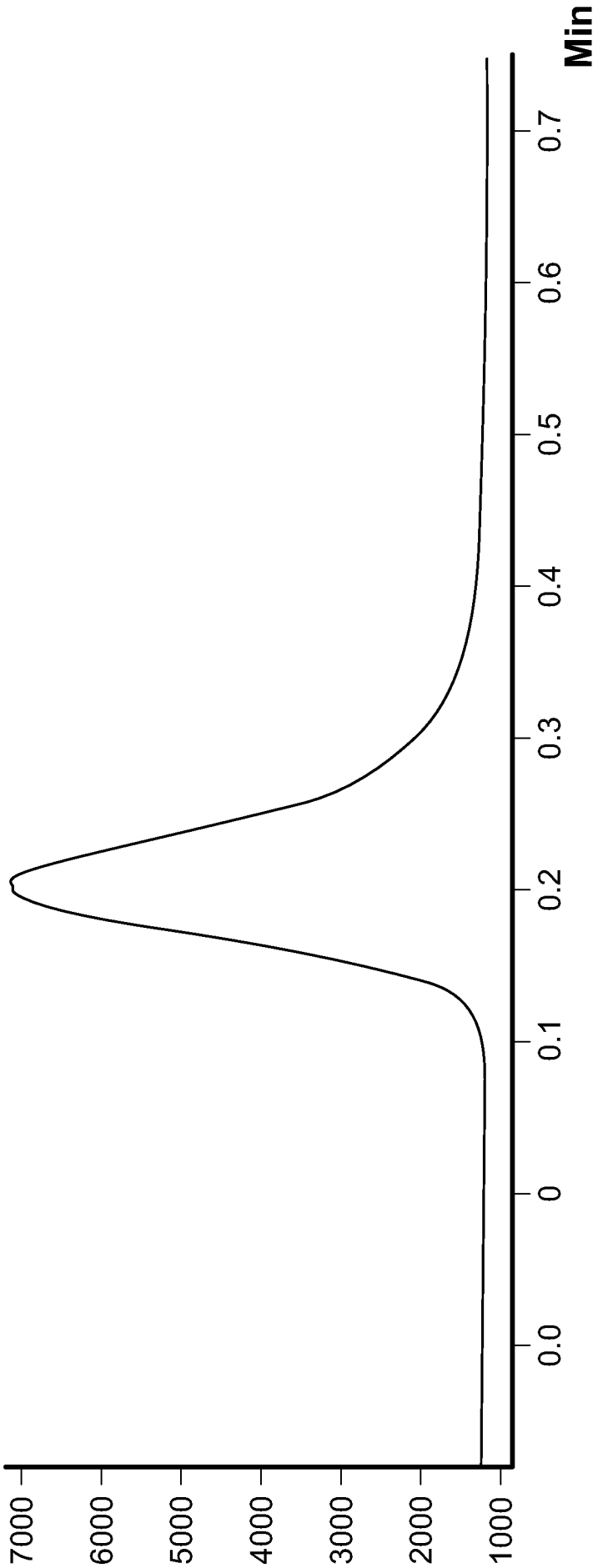


FIG. 26A

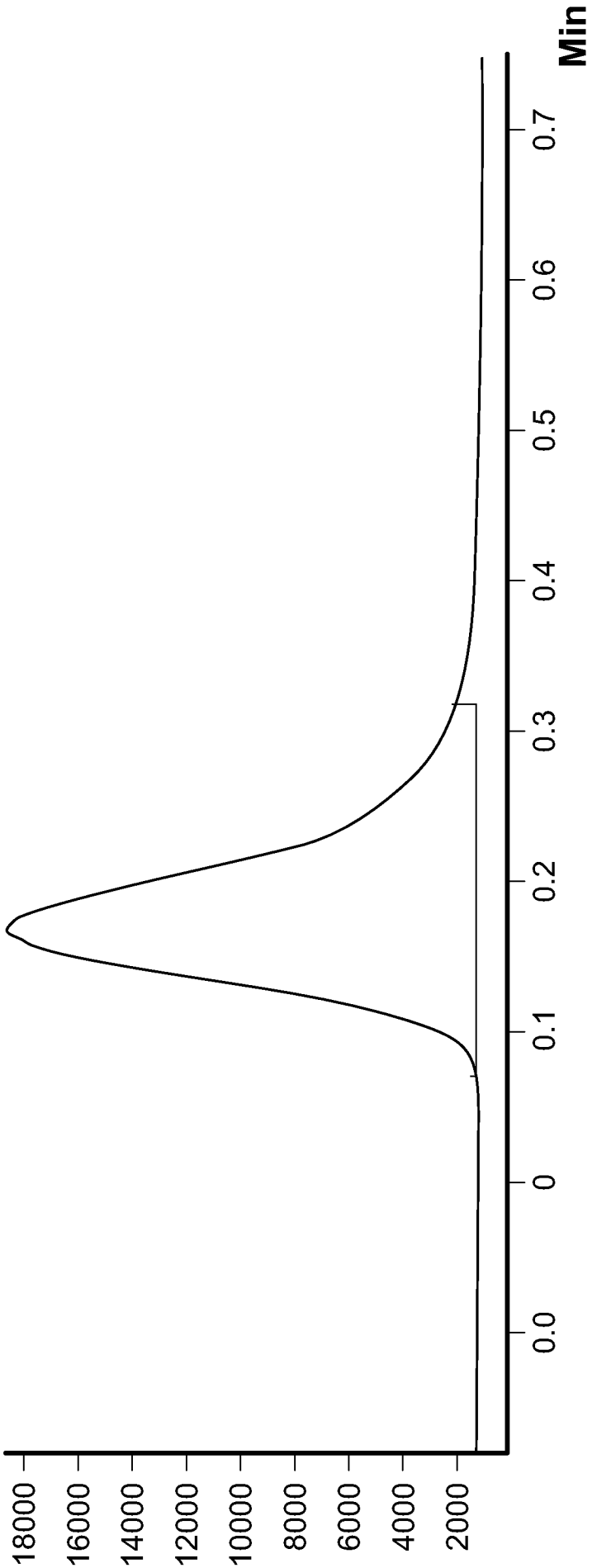

Min
0.0
0.1
0
0.2
0.3
0.4
0.5
0.6
0.7
2000
4000
6000
8000
10000
12000
14000
16000
18000

FIG. 26B

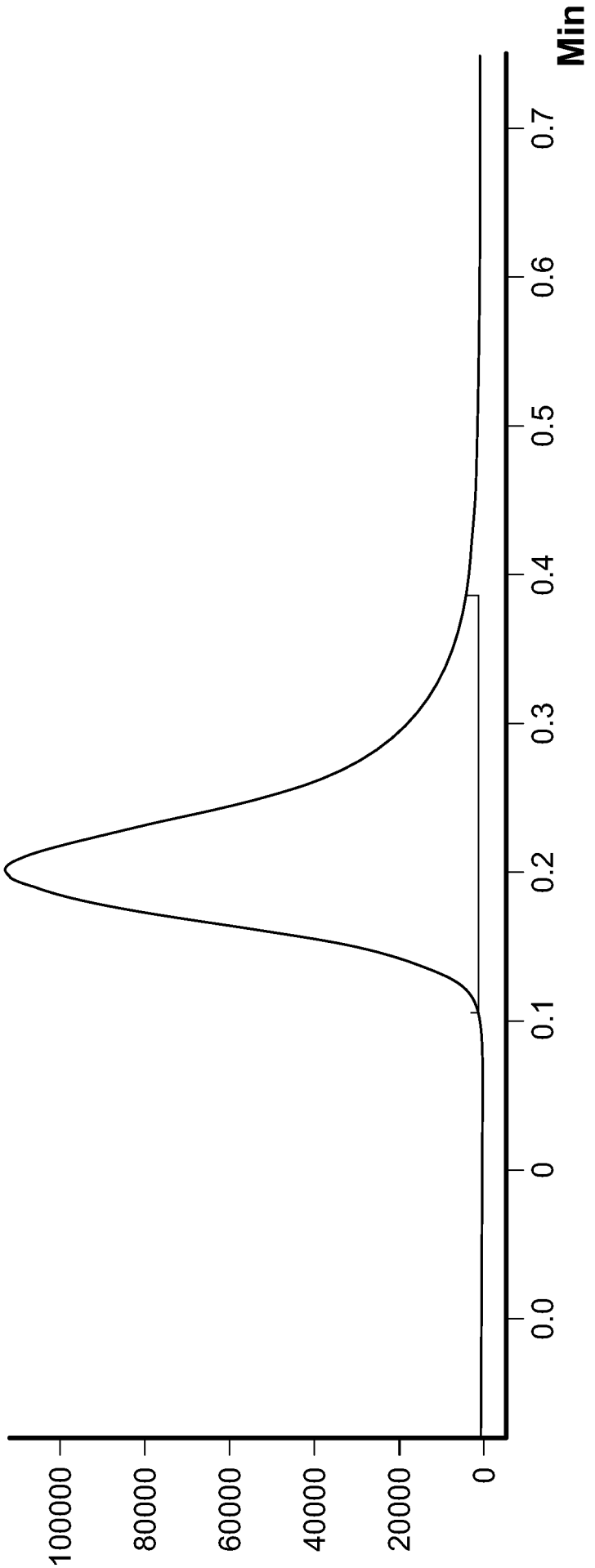


FIG. 26C

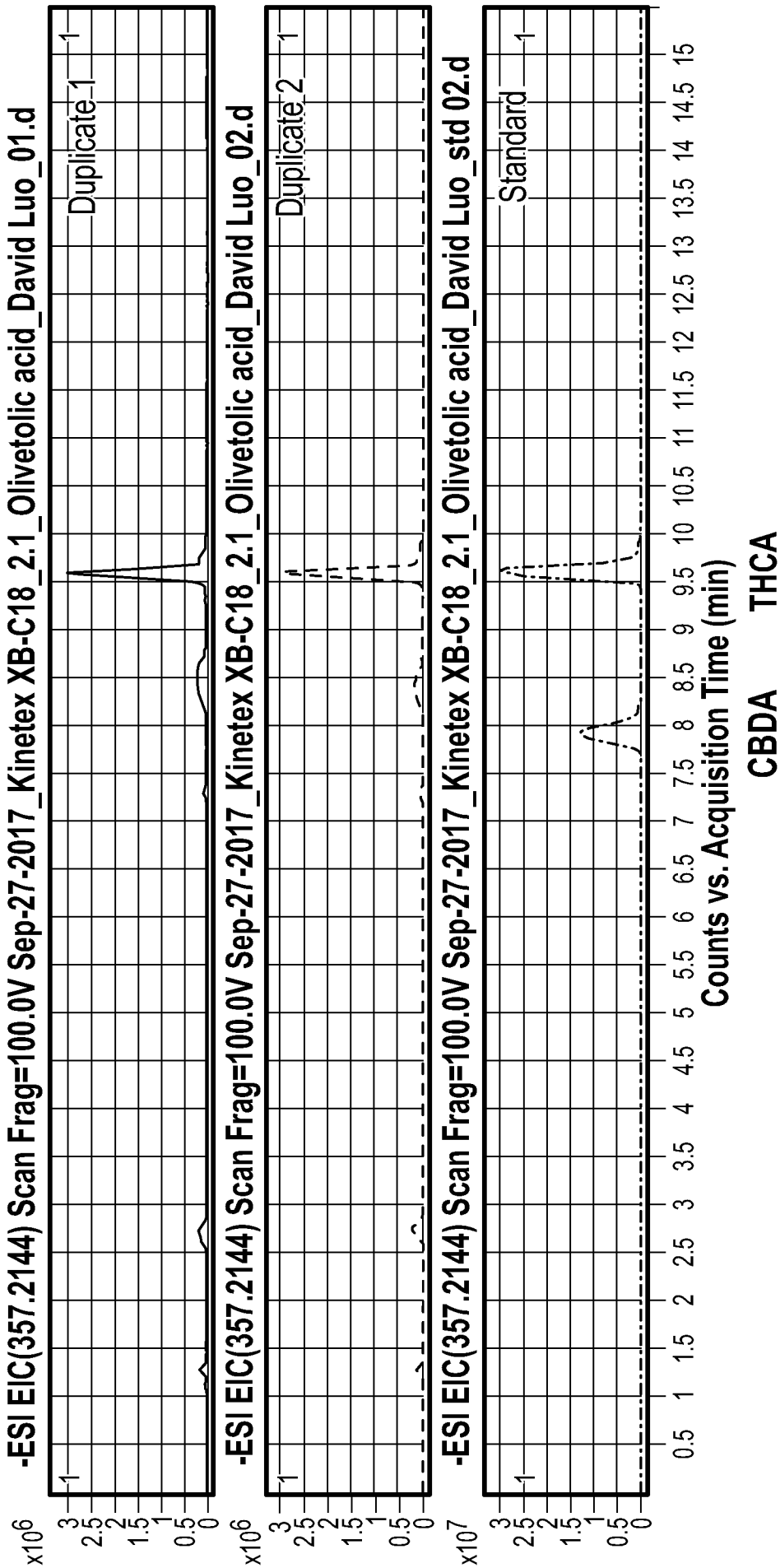

THCA synthase with N-terminal truncation and ProA signal sequence
-ESI EIC(357.2144) Scan Frag=100.0V Sep-27-2017_Kinetex XB-C18_2.1_Olivetolic acid_David Luo_01.d
Duplicate 1
-ESI EIC(357.2144) Scan Frag=100.0V Sep-27-2017_Kinetex XB-C18_2.1_Olivetolic acid_David Luo_02.d
Duplicate 2
-ESI EIC(357.2144) Scan Frag=100.0V Sep-27-2017_Kinetex XB-C18_2.1_Olivetolic acid_David Luo_std 02.d
Standard
x10^6
x10^6
x10^7
Counts vs. Acquisition Time (min)
CBDA
THCA

FIG. 27

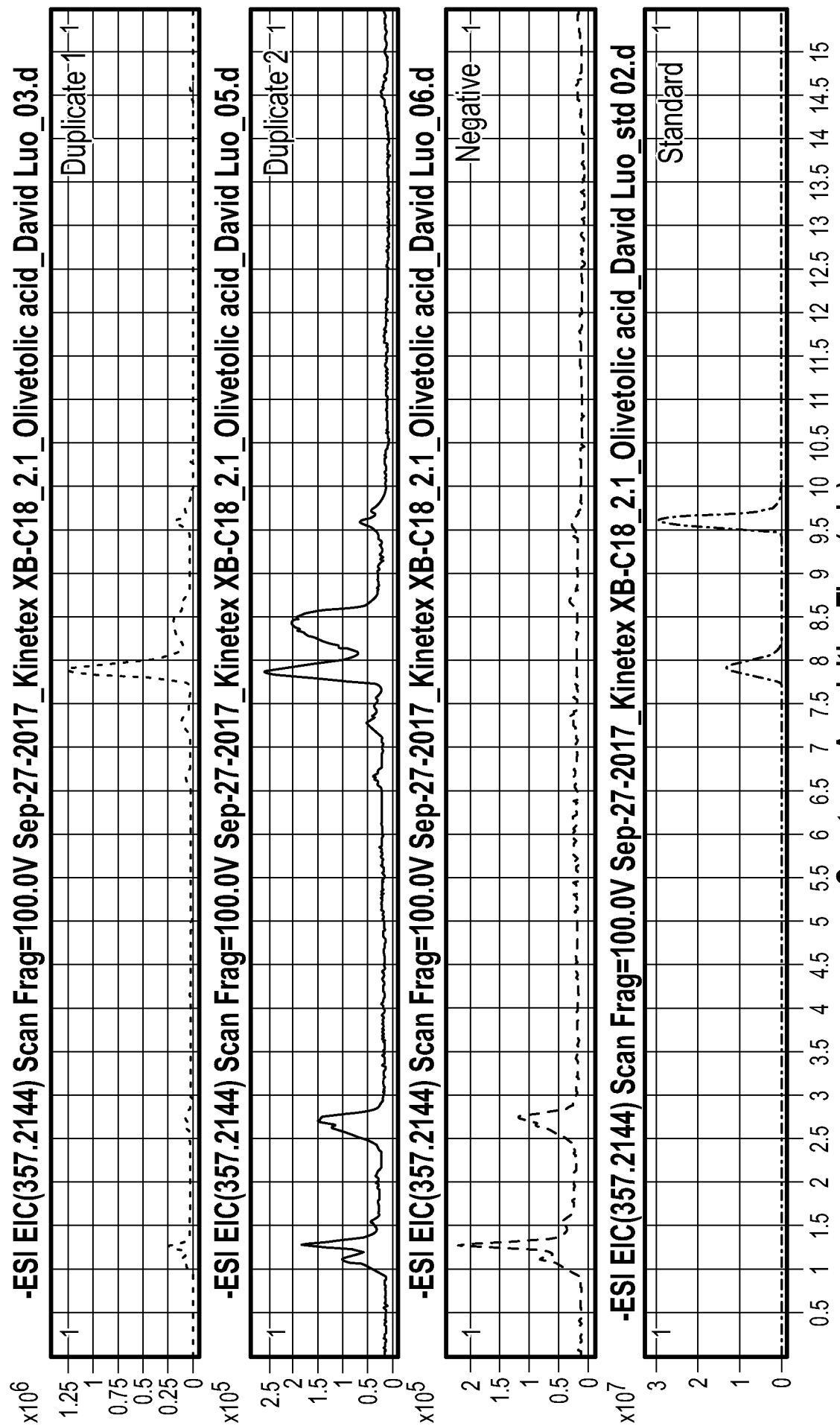

CBDA synthase with N-terminal truncation and ProA signal sequence
-ESI EIC(357.2144) Scan Frag=100.0V Sep-27-2017_Kinetex XB-C18_2.1_Olivetolic acid_David Luo_03.d
Duplicate 1
-ESI EIC(357.2144) Scan Frag=100.0V Sep-27-2017_Kinetex XB-C18_2.1_Olivetolic acid_David Luo_05.d
Duplicate 2
-ESI EIC(357.2144) Scan Frag=100.0V Sep-27-2017_Kinetex XB-C18_2.1_Olivetolic acid_David Luo_06.d
Negative
-ESI EIC(357.2144) Scan Frag=100.0V Sep-27-2017_Kinetex XB-C18_2.1_Olivetolic acid_David Luo_std 02.d
Standard
Counts vs. Acquisition Time (min)
CBDA
THCA

FIG. 28

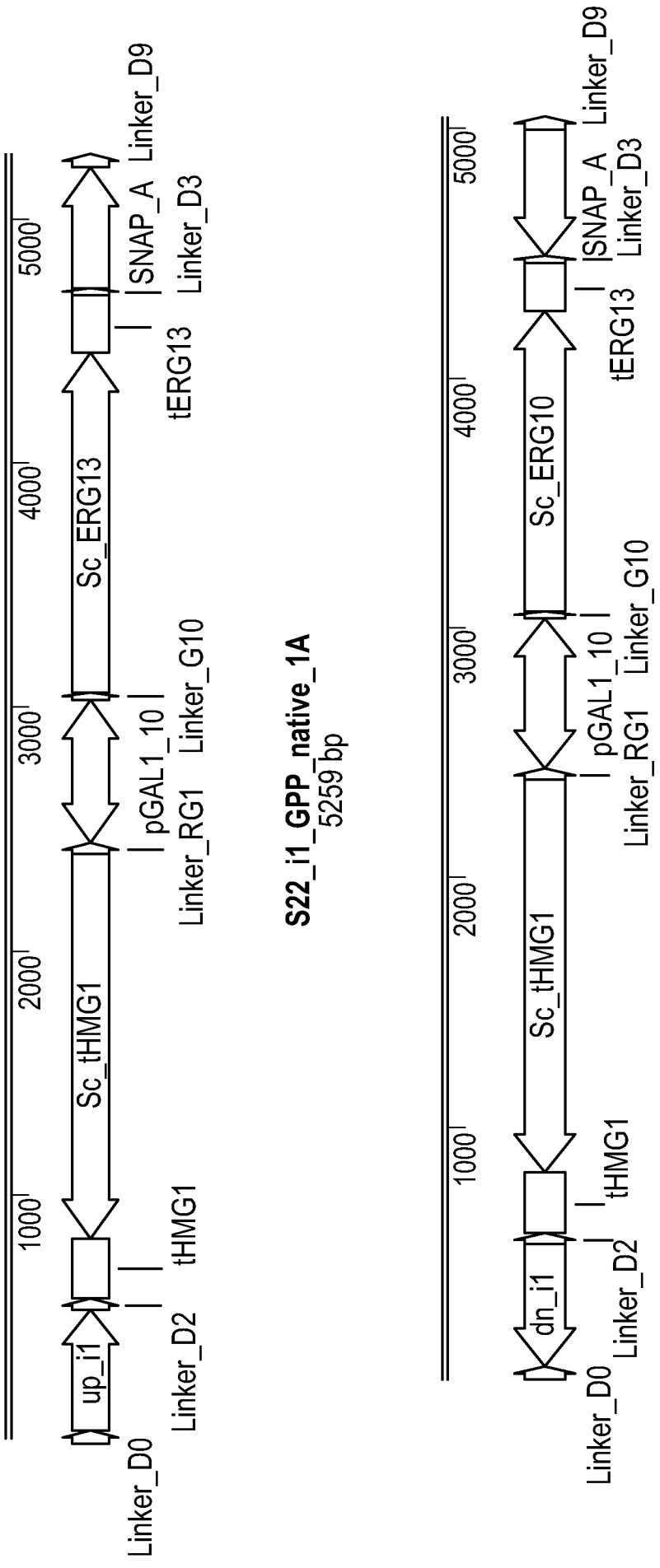

1000
2000
3000
4000
5000
Linker_D0
up_i1
Linker_D2
tHMG1
Sc_tHMG1
Linker_RG1
pGAL1_10
Linker_G10
Sc_ERG13
tERG13
Linker_D3
SNAP_A
Linker_D9
S22_i1_GPP_native_1A
5259 bp
1000
2000
3000
4000
5000
Linker_D0
dn_i1
Linker_D2
tHMG1
Sc_tHMG1
Linker_RG1
pGAL1_10
Linker_G10
Sc_ERG10
tERG13
Linker_D3
SNAP_A
Linker_D9
S23_i1_GPP_native_1B
5028 bp

FIG. 29A

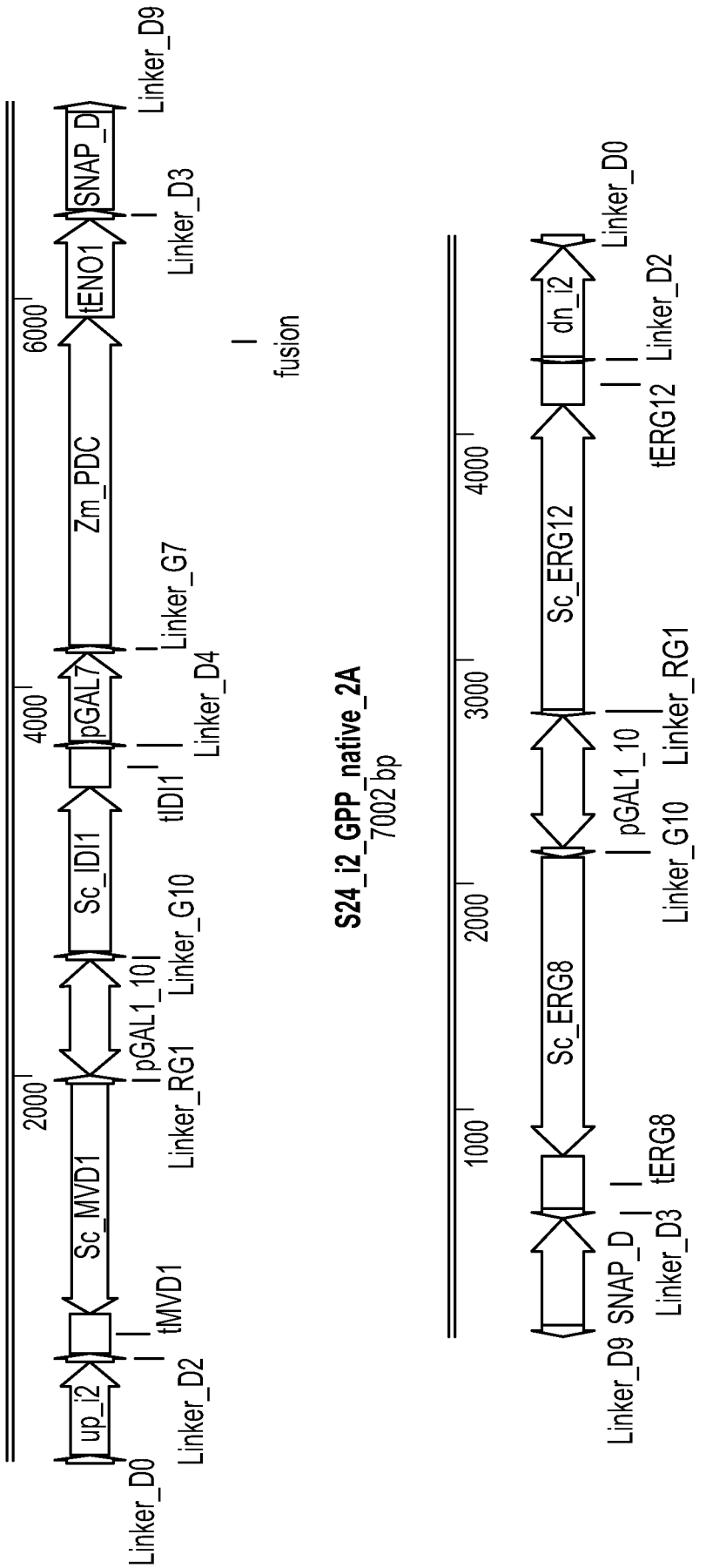

Linker_D0
up_i2
Linker_D2
tMVD1
Sc_MVD1
Linker_RG1
pGAL1_10
Linker_G10
Sc_IDI1
tIDI1
Linker_D4
pGAL7
Linker_G7
Zm_PDC
tENO1
Linker_D3
SNAP_D
Linker_D9
2000
4000
6000
fusion
S24_i2_GPP_native_2A
7002 bp
Linker_D9
SNAP_D
Linker_D3
tERG8
Sc_ERG8
Linker_G10
pGAL1_10
Linker_RG1
Sc_ERG12
tERG12
Linker_D2
dn_i2
Linker_D0
1000
2000
3000
4000
S25_i2_GPP_native_2B
4897 bp

FIG. 29B

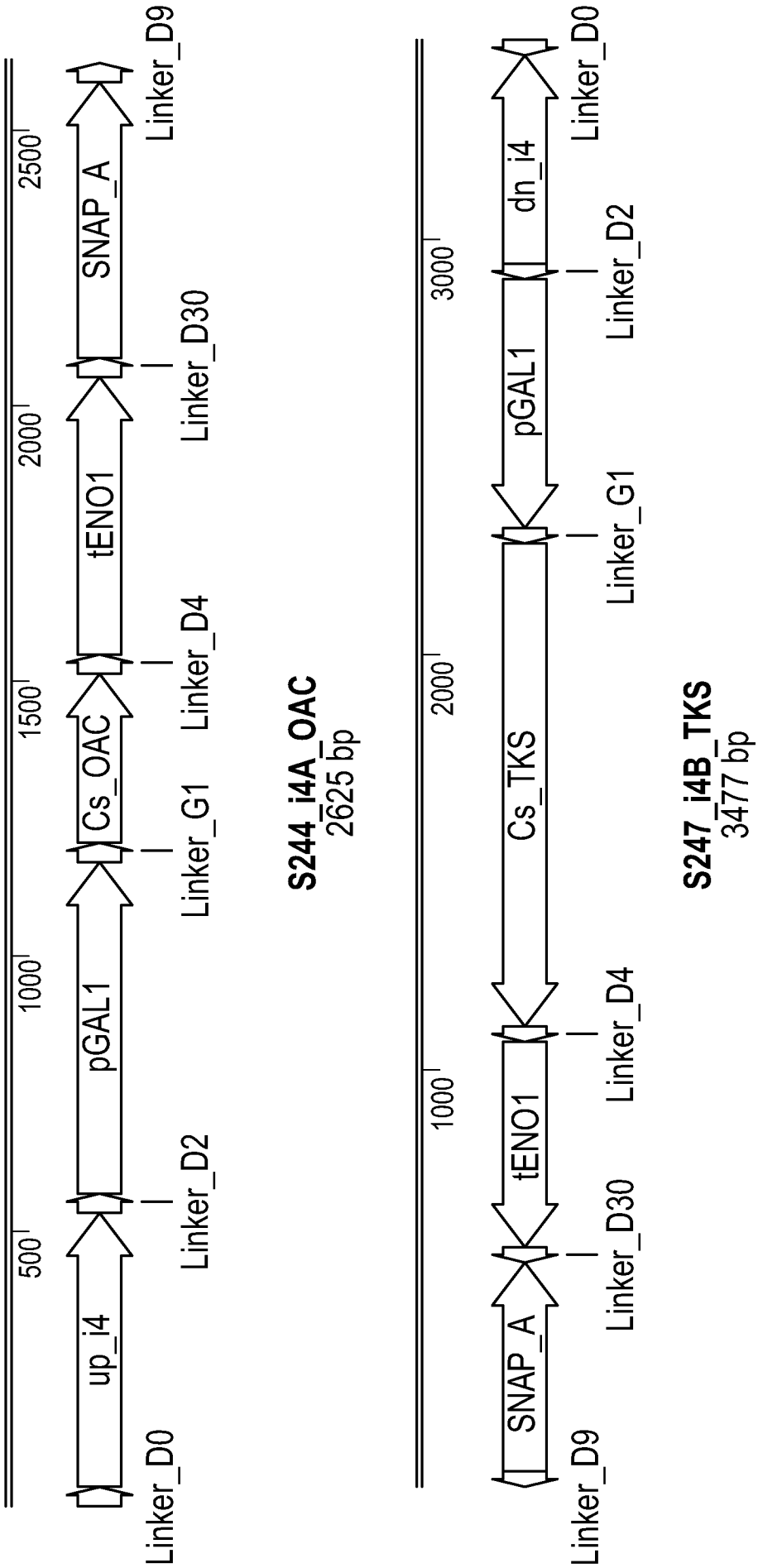

500
1000
1500
2000
2500
Linker_D0
up_i4
Linker_D2
pGAL1
Linker_G1
Cs_OAC
Linker_D4
tENO1
Linker_D30
SNAP_A
Linker_D9
S244_i4A_OAC
2625 bp
1000
2000
3000
Linker_D9
SNAP_A
Linker_D30
tENO1
Linker_D4
Cs_TKS
Linker_G1
pGAL1
Linker_D2
dn_i4
Linker_D0
S247_i4B_TKS
3477 bp

FIG. 30A

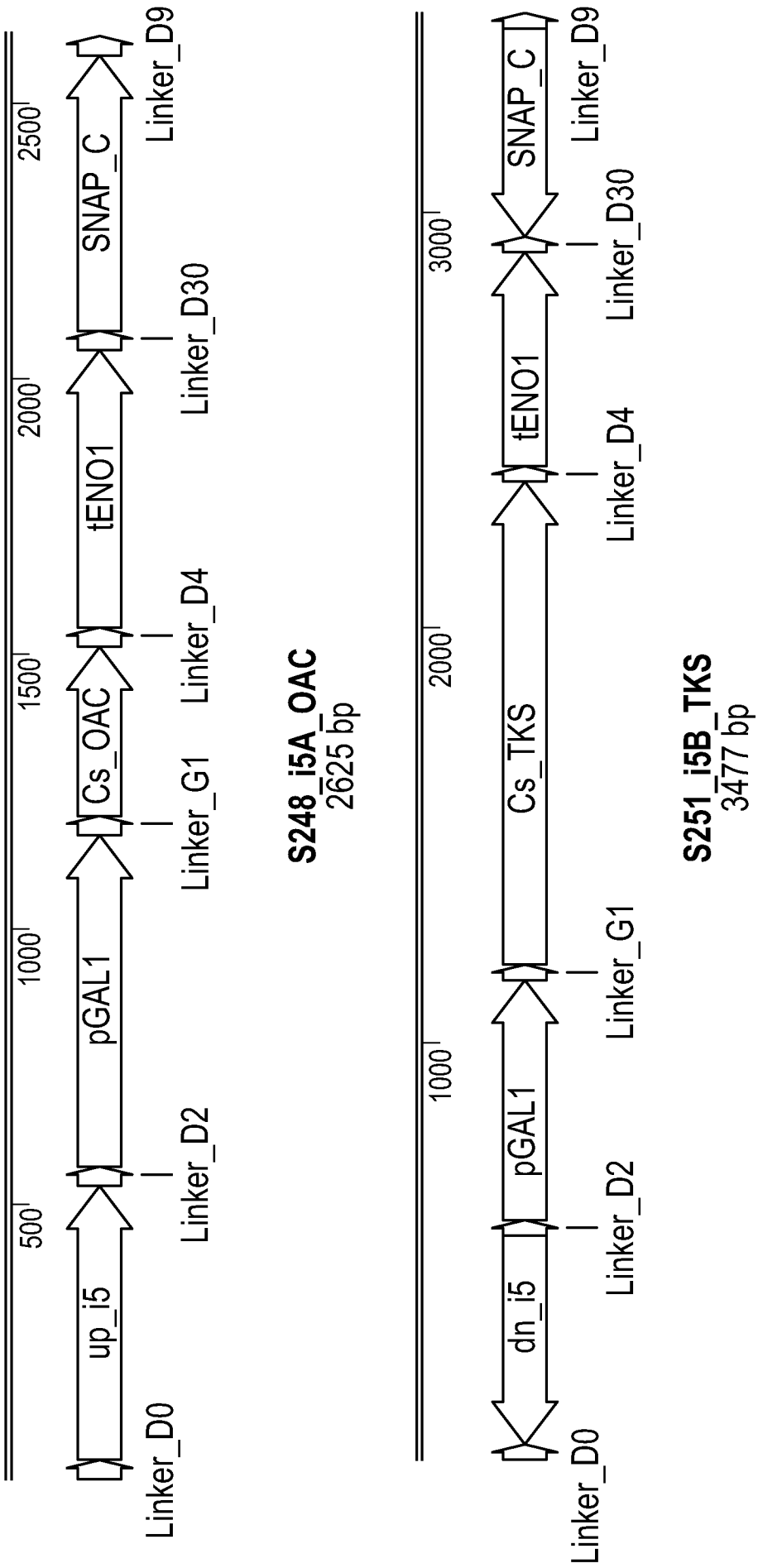

500
1000
1500
2000
2500
Linker_D0
up_i5
Linker_D2
pGAL1
Linker_G1
Cs_OAC
Linker_D4
tENO1
Linker_D30
SNAP_C
Linker_D9
S248_i5A_OAC
2625 bp
1000
2000
3000
Linker_D0
dn_i5
Linker_D2
pGAL1
Linker_G1
Cs_TKS
Linker_D4
tENO1
Linker_D30
SNAP_C
Linker_D9
S251_i5B_TKS
3477 bp

FIG. 30B

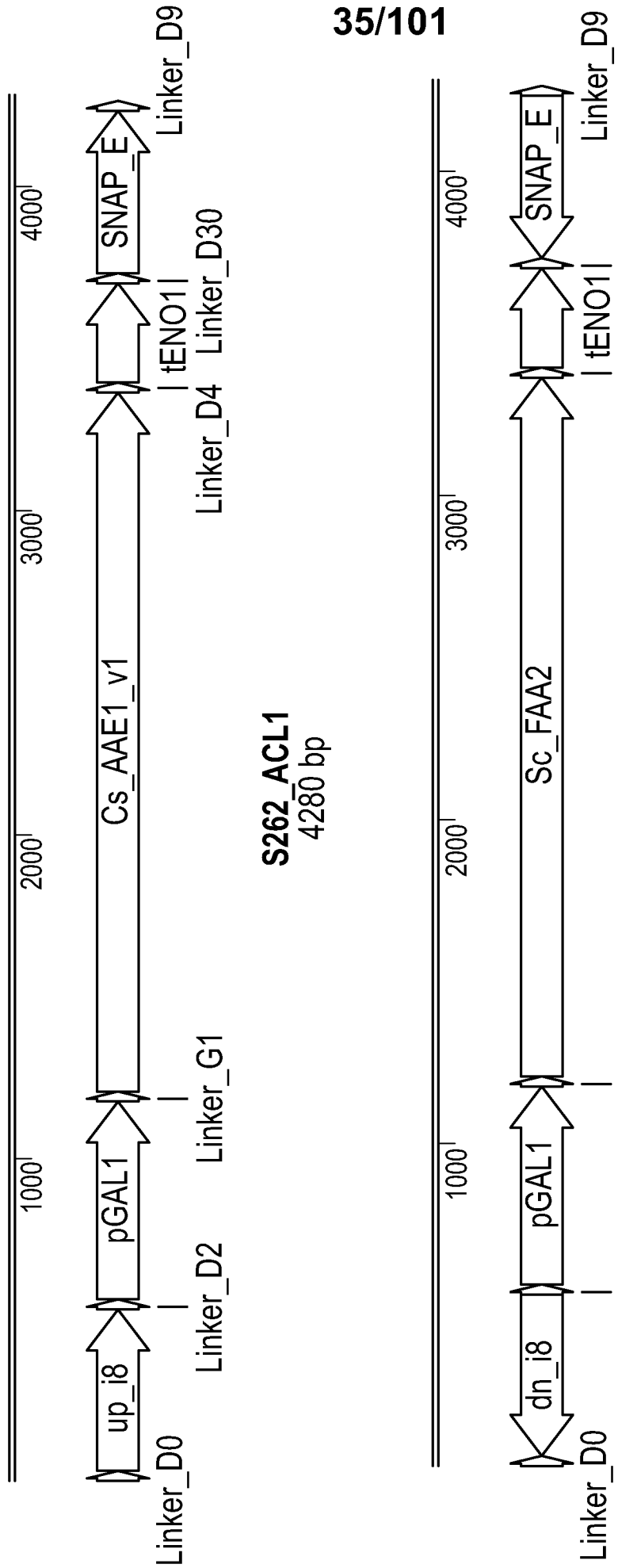

1000
2000
3000
4000
Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Cs_AAE1_v1
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S262_ACL1
4280 bp
1000
2000
3000
4000
Linker_D0
dn_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA2
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S281_ACL20
4352 bp

FIG. 30C

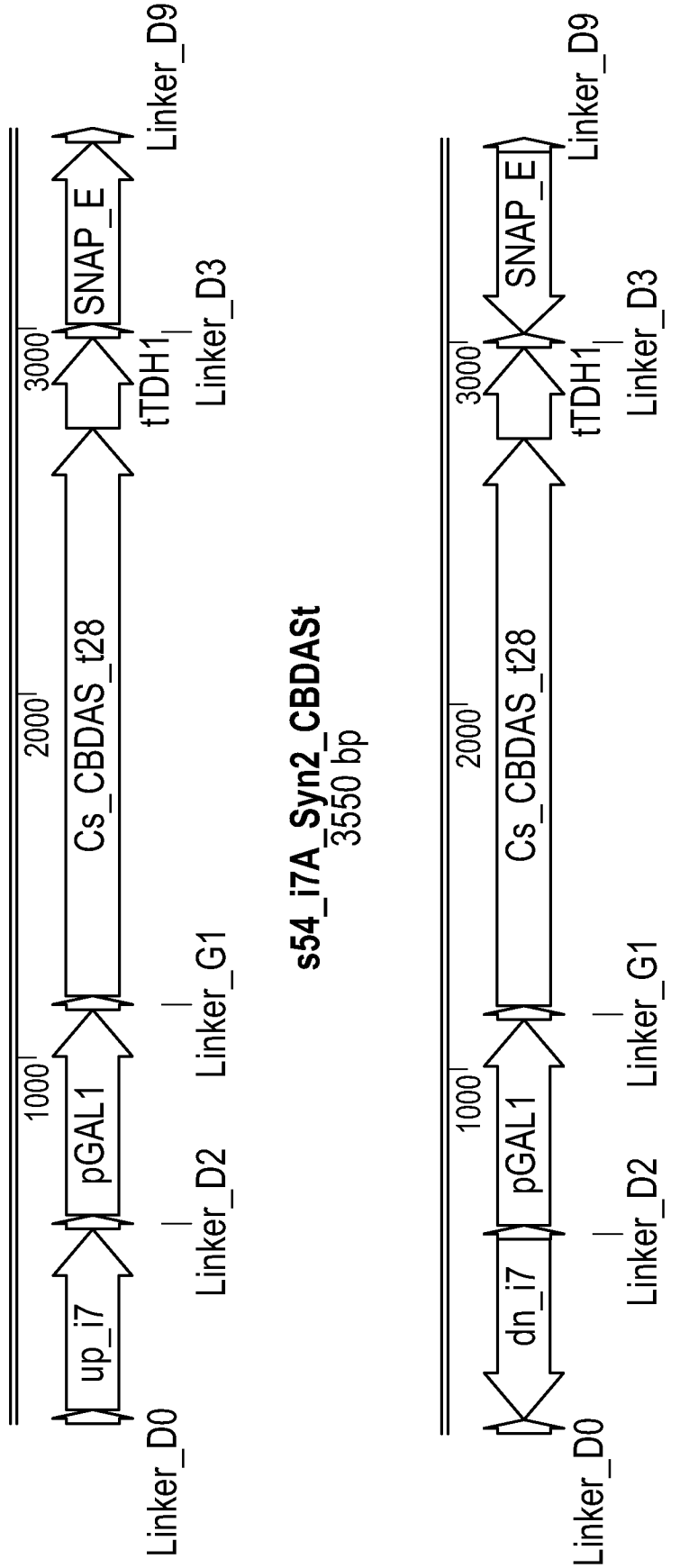
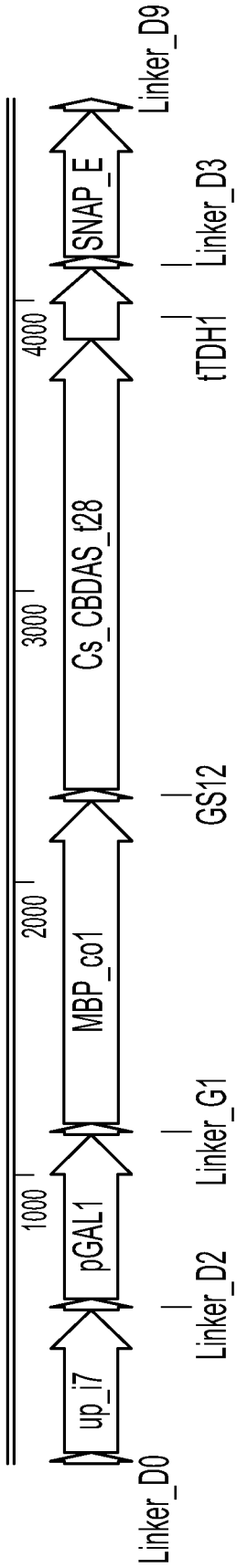

Linker_D0
up_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s54_i7A_Syn2_CBDASt
3550 bp
Linker_D0
dn_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s68_i7B_Syn2_CBDASt
3550 bp

FIG. 31


Linker_D0
up_i7
Linker_D2
pGAL1
Linker_G1
MBP_co1
GS12
Cs_CBDAS_t28
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
4000
s56_i7A_Syn4_MBP-CBDASt
4693 bp

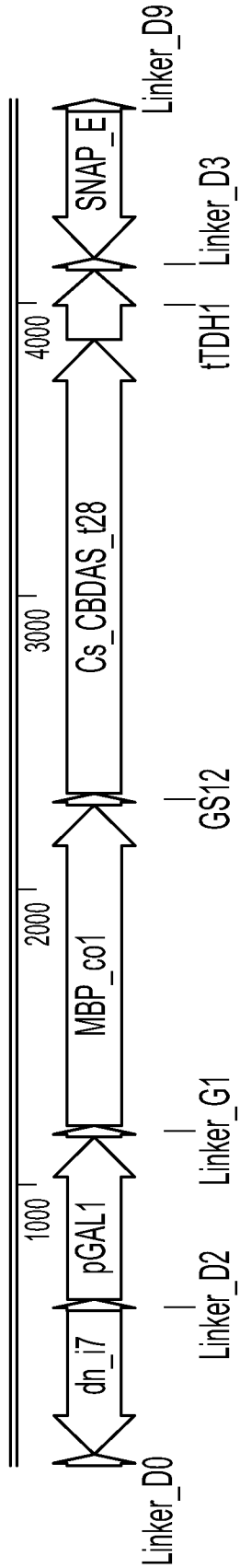

Linker_D0
dn_i7
Linker_D2
pGAL1
Linker_G1
MBP_co1
GS12
Cs_CBDAS_t28
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
4000
s70_i7B_Syn4_MBP-CBDASt
4693 bp

FIG. 32

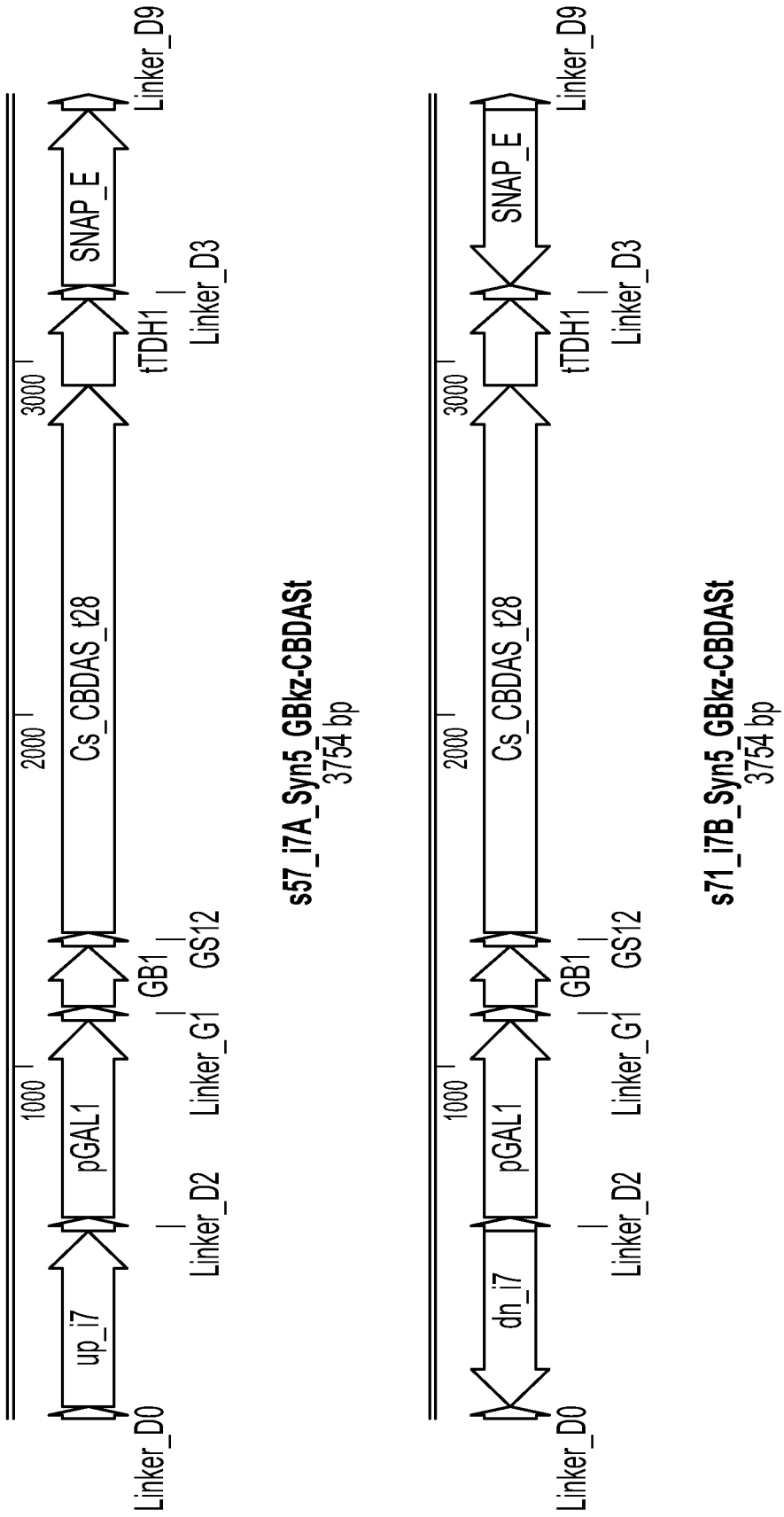
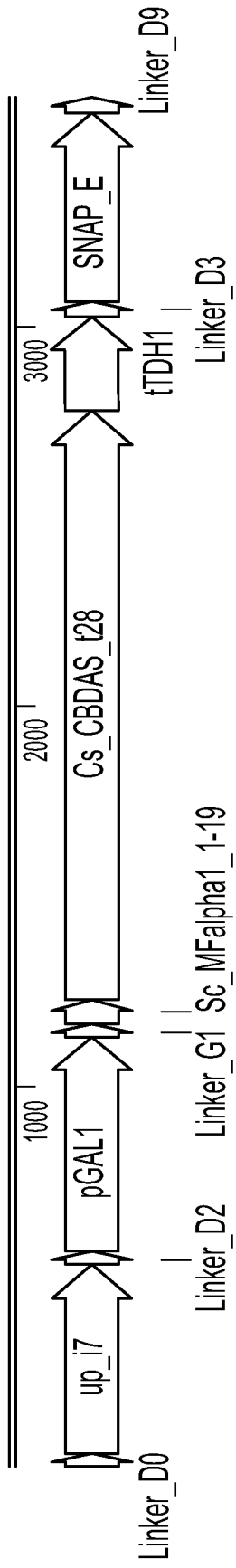
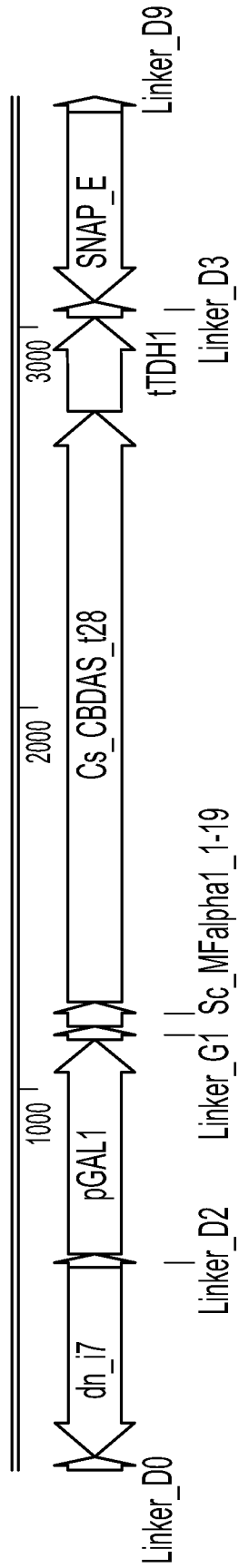

Linker_D0
up_i7
Linker_D2
pGAL1
Linker_G1
GB1
GS12
Cs_CBDAS_t28
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s57_i7A_Syn5_GBkz-CBDASt
3754 bp
Linker_D0
dn_i7
Linker_D2
pGAL1
Linker_G1
GB1
GS12
Cs_CBDAS_t28
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s71_i7B_Syn5_GBkz-CBDASt
3754 bp

FIG. 33



s58_i7A_Syn7_MFer-CBDASt
3601 bp



s72_i7B_Syn7_MFer-CBDASt
3601 bp

FIG. 34

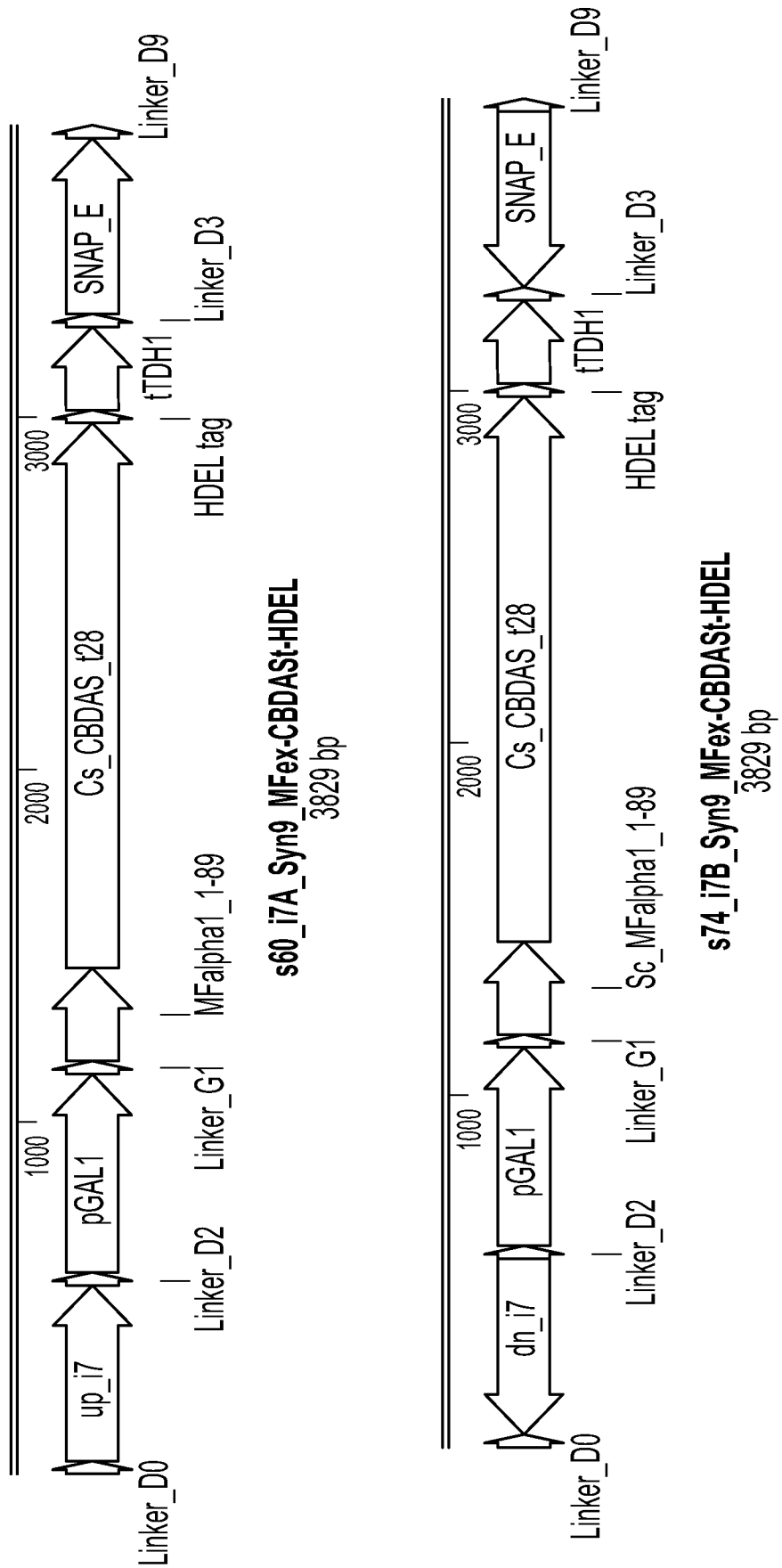

Linker_D0
up_i7
Linker_D2
pGAL1
Linker_G1
MFalpha1_1-89
Cs_CBDAS_t28
HDEL tag
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s60_i7A_Syn9_MFex-CBDASt-HDEL
3829 bp
Linker_D0
dn_i7
Linker_D2
pGAL1
Linker_G1
Sc_MFalpha1_1-89
Cs_CBDAS_t28
HDEL tag
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s74_i7B_Syn9_MFex-CBDASt-HDEL
3829 bp

FIG. 35

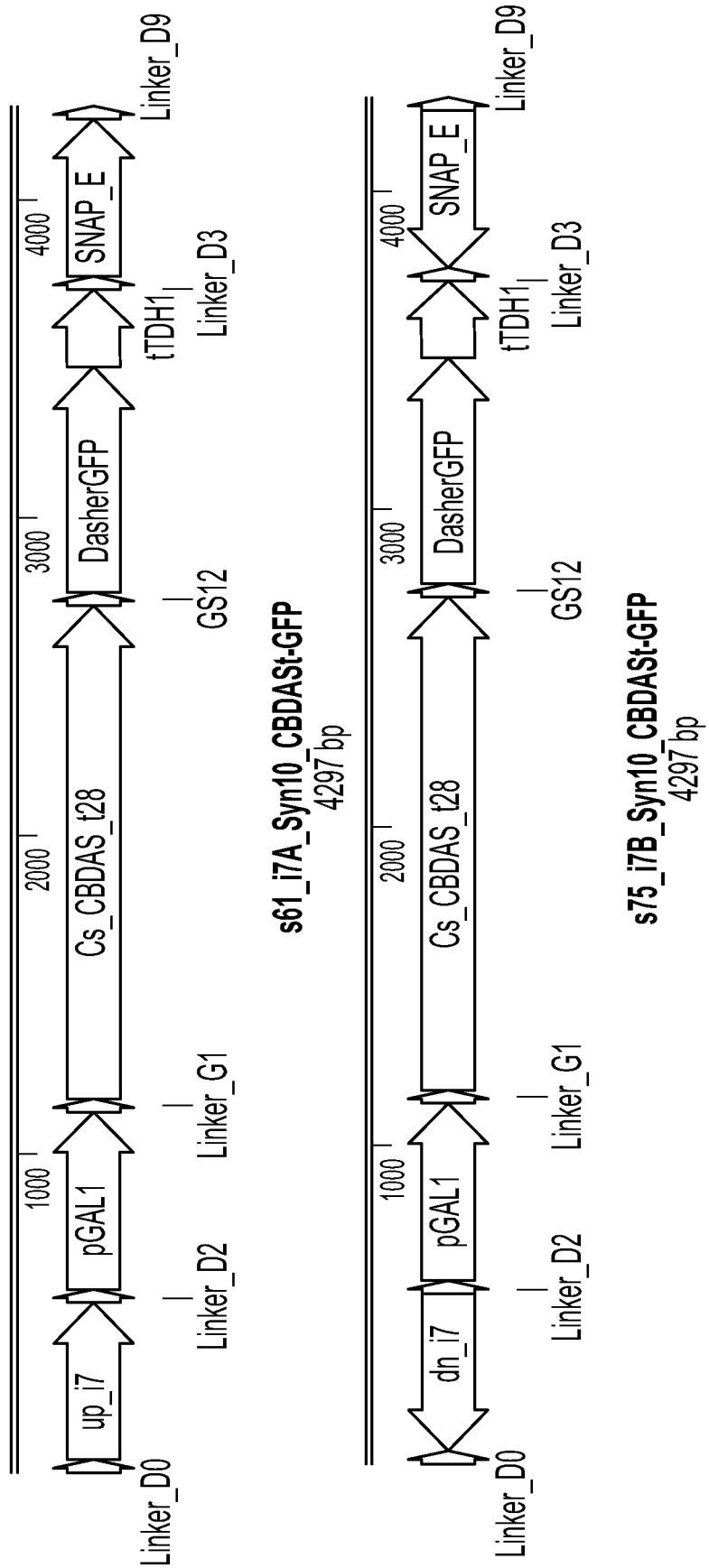

1000
2000
3000
4000
Linker_D0
up_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
GS12
DasherGFP
tTDH1
Linker_D3
SNAP_E
Linker_D9
s61_i7A_Syn10_CBDASt-GFP
4297 bp
Linker_D0
dn_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
GS12
DasherGFP
tTDH1
Linker_D3
SNAP_E
Linker_D9
s75_i7B_Syn10_CBDASt-GFP
4297 bp

FIG. 36

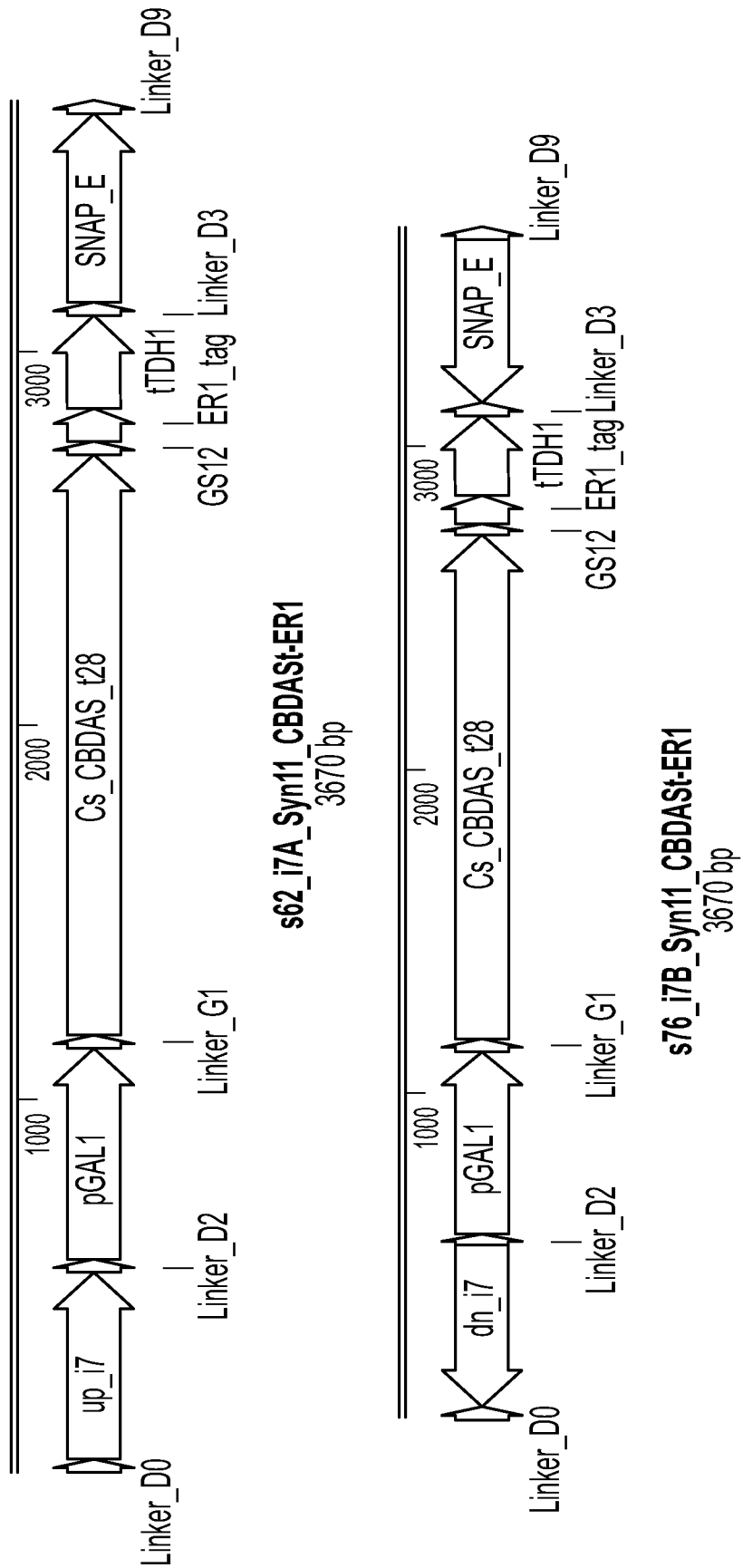

Linker_D0
up_i7
Linker_D2
pGAL1
1000
Linker_G1
Cs_CBDAS_t28
2000
GS12
ER1_tag
tTDH1
3000
Linker_D3
SNAP_E
Linker_D9
s62_i7A_Syn11_CBDASt-ER1
3670 bp
Linker_D0
dn_i7
Linker_D2
pGAL1
1000
Linker_G1
Cs_CBDAS_t28
2000
GS12
ER1_tag
tTDH1
3000
Linker_D3
SNAP_E
Linker_D9
s76_i7B_Syn11_CBDASt-ER1
3670 bp

FIG. 37

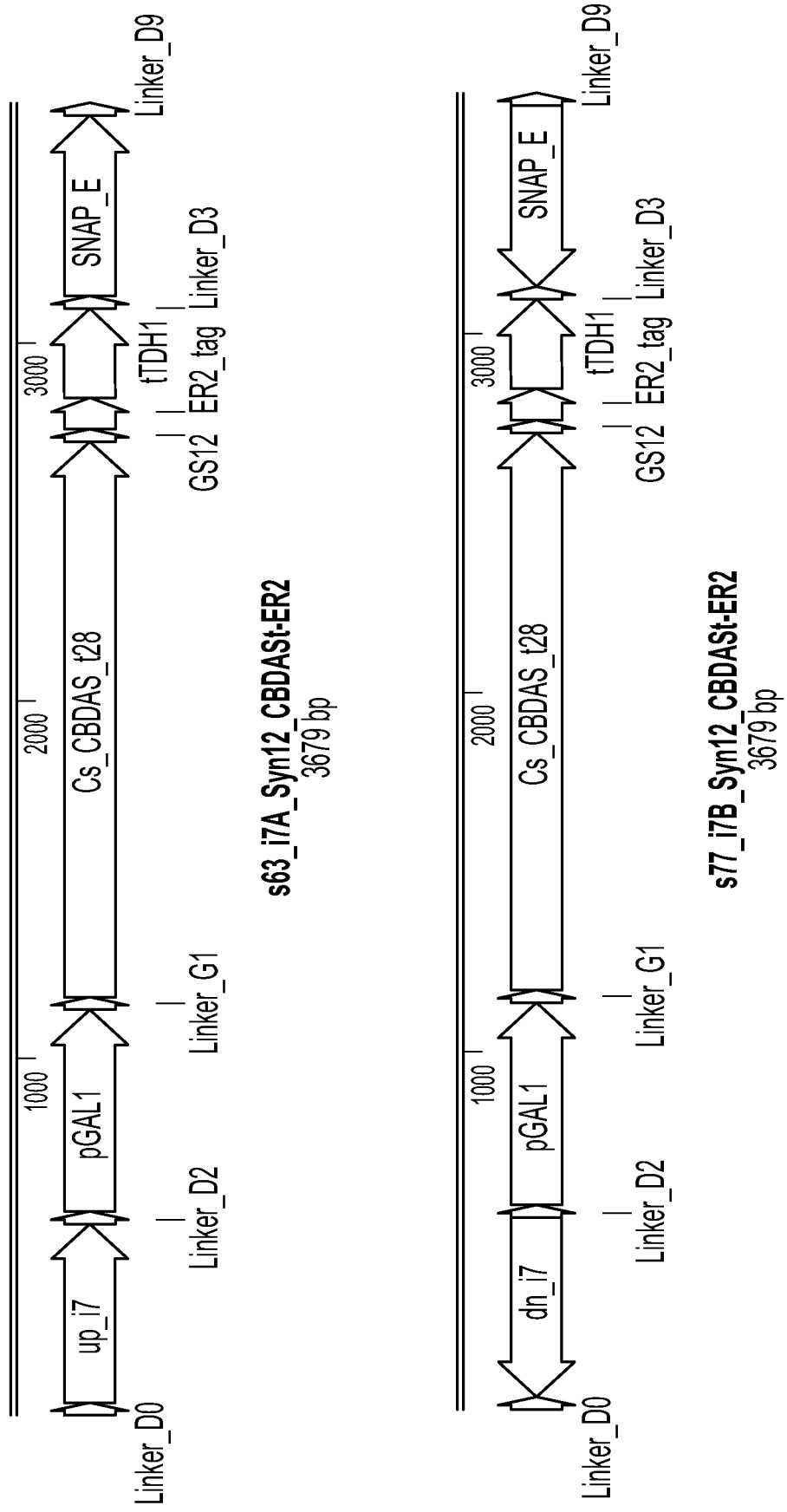

Linker_D0
up_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
GS12
tTDH1
ER2_tag
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s63_i7A_Syn12_CBDASt-ER2
3679 bp
Linker_D0
dn_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
GS12
tTDH1
ER2_tag
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s77_i7B_Syn12_CBDASt-ER2
3679 bp

FIG. 38

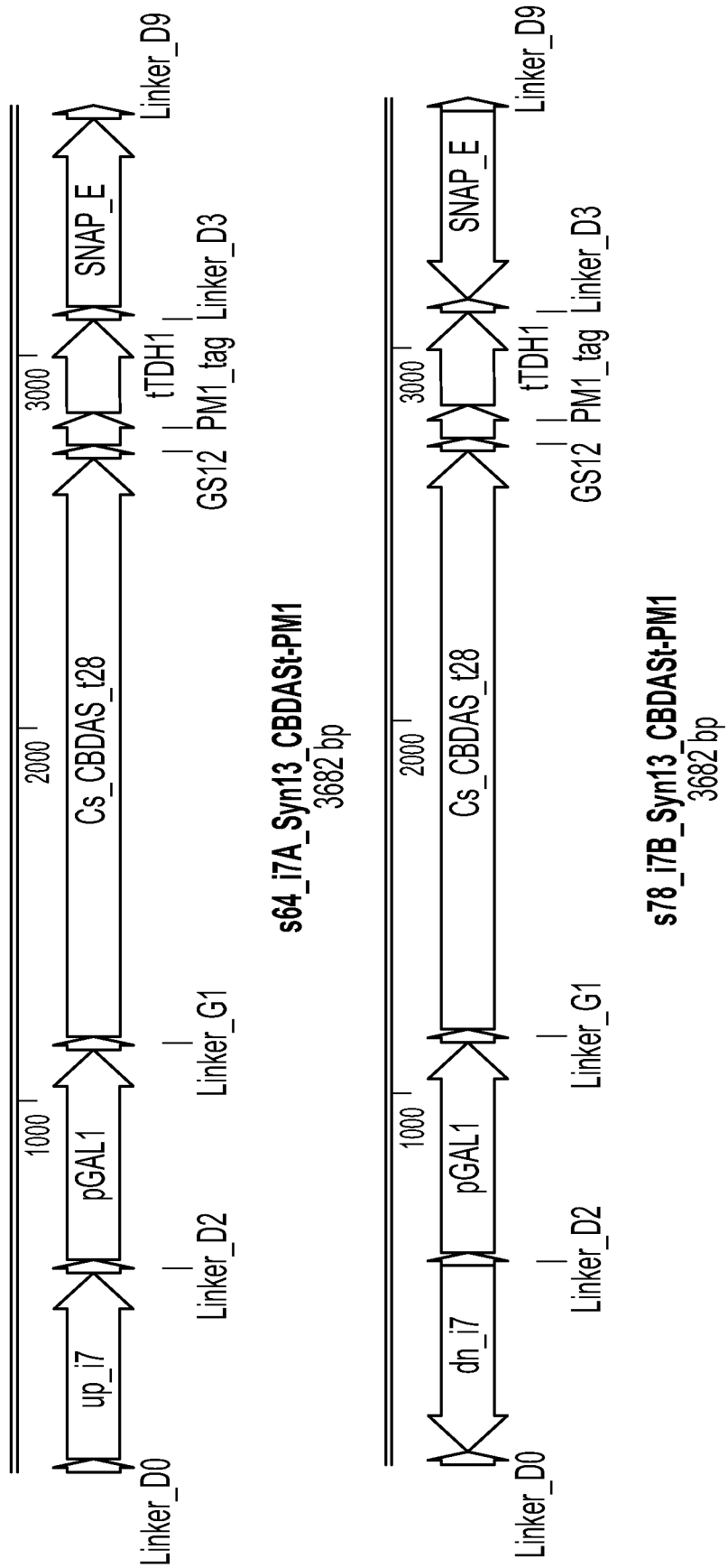

Linker_D0
up_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
tTDH1
GS12
PM1_tag
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s64_i7A_Syn13_CBDASt-PM1
3682 bp
Linker_D0
dn_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
tTDH1
GS12
PM1_tag
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s78_i7B_Syn13_CBDASt-PM1
3682 bp

FIG. 39

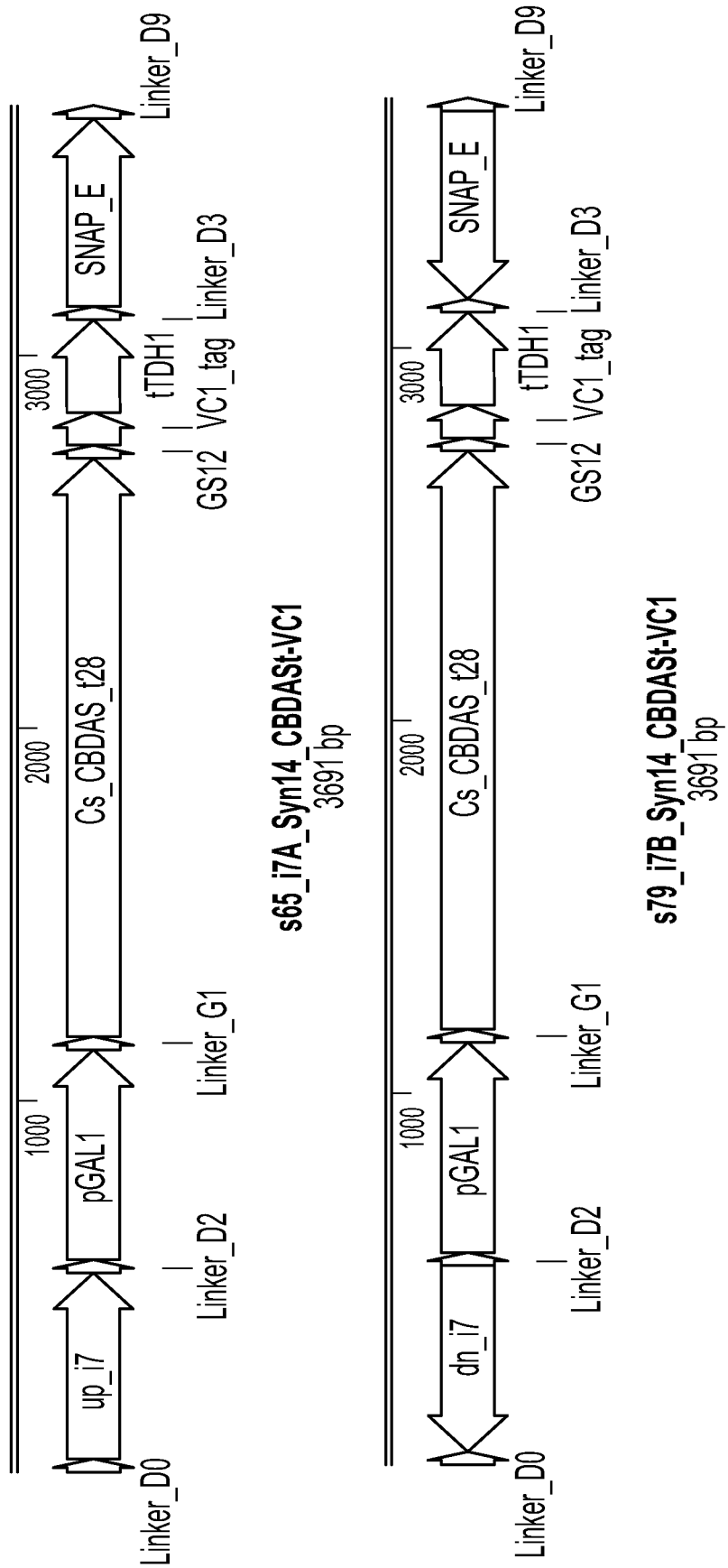

Linker_D0
up_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
tTDH1
GS12
VC1_tag
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s65_i7A_Syn14_CBDASt-VC1
3691 bp
Linker_D0
dn_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
tTDH1
GS12
VC1_tag
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s79_i7B_Syn14_CBDASt-VC1
3691 bp

FIG. 40

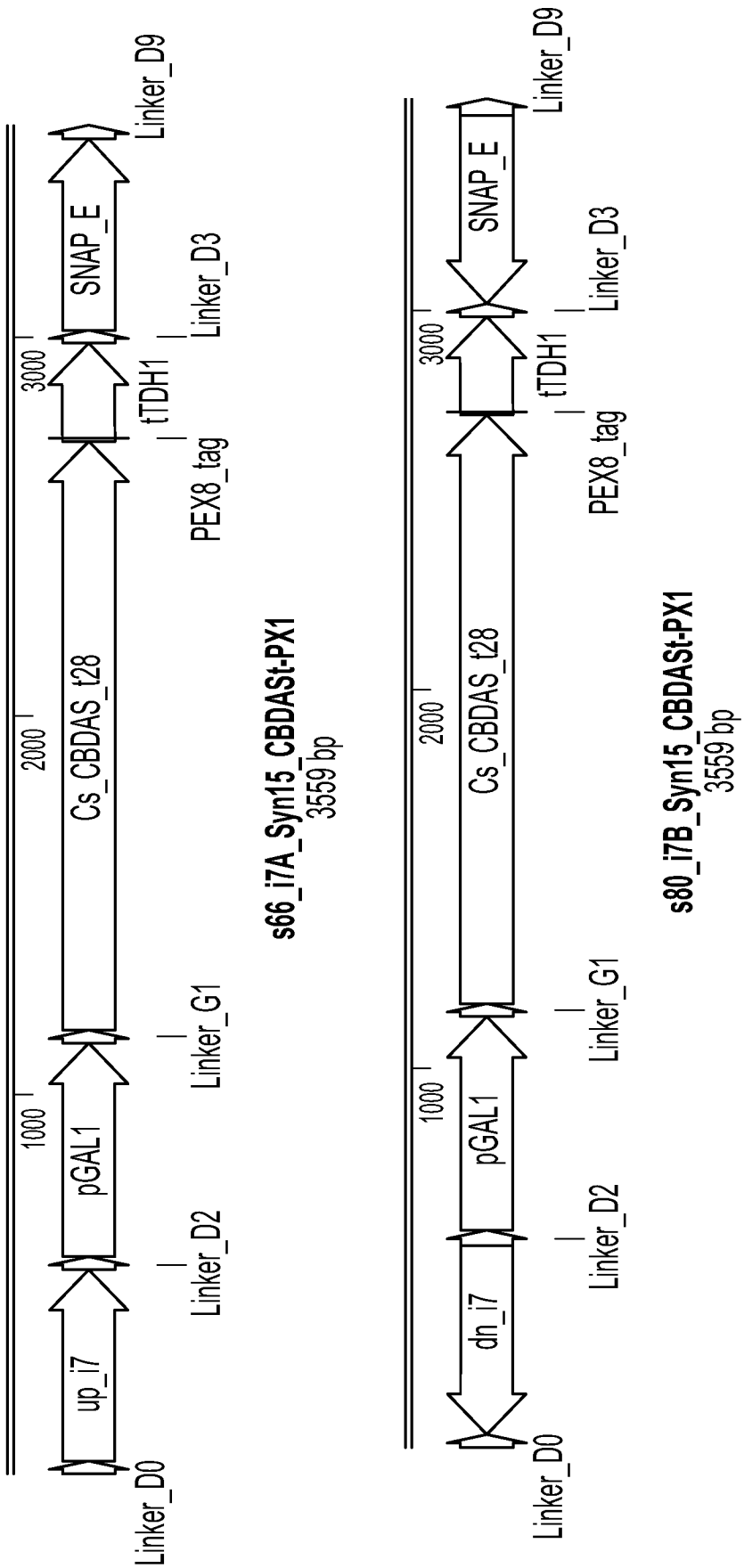

Linker_D0
up_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
PEX8_tag
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s66_i7A_Syn15_CBDASt-PX1
3559 bp
Linker_D0
dn_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
PEX8_tag
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s80_i7B_Syn15_CBDASt-PX1
3559 bp

FIG. 41

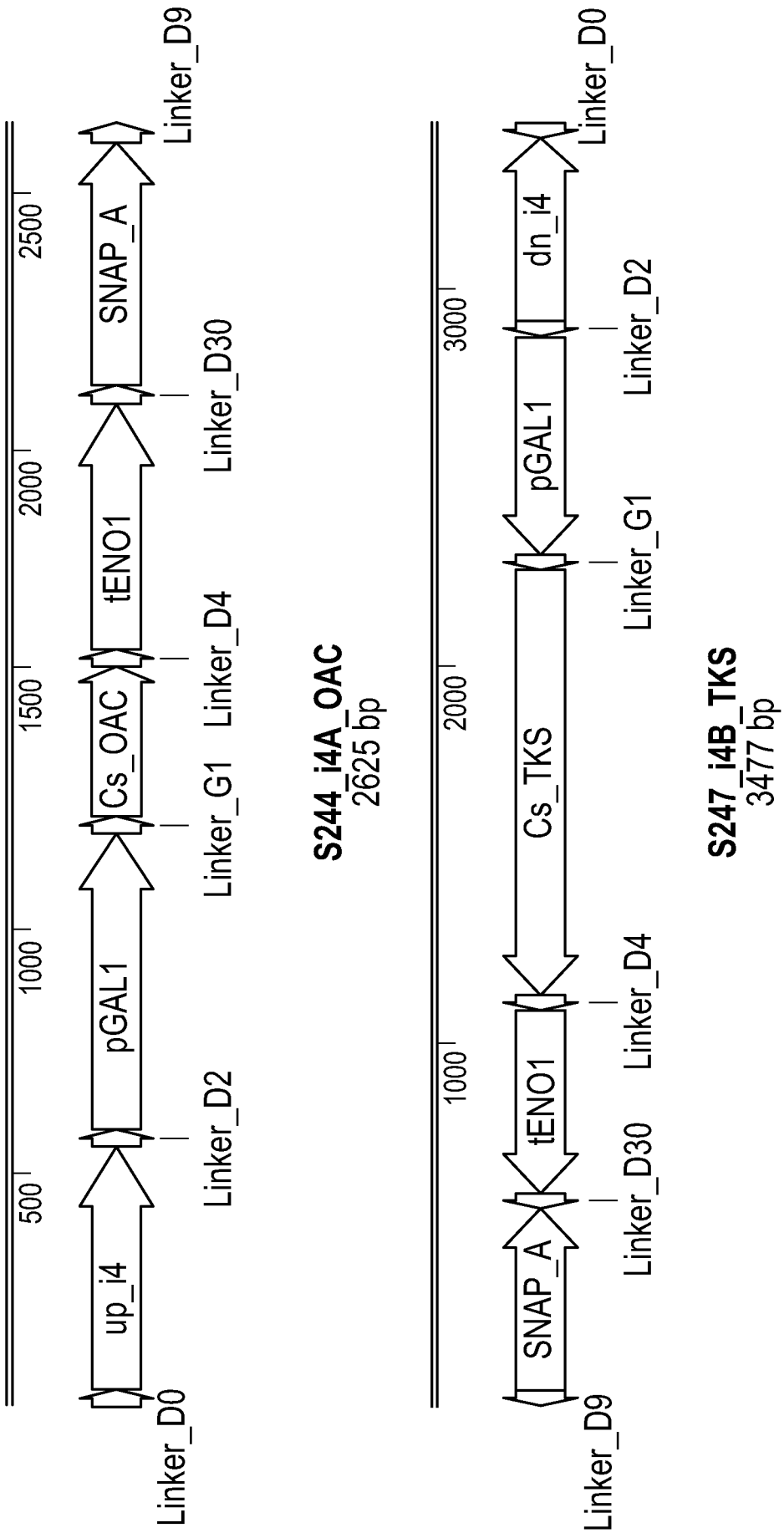

500
1000
1500
2000
2500
Linker_D0
up_i4
Linker_D2
pGAL1
Linker_G1
Cs_OAC
Linker_D4
tENO1
Linker_D30
SNAP_A
Linker_D9
S244_i4A_OAC
2625 bp
1000
2000
3000
Linker_D9
SNAP_A
Linker_D30
tENO1
Linker_D4
Cs_TKS
Linker_G1
pGAL1
Linker_D2
dn_i4
Linker_D0
S247_i4B_TKS
3477 bp

FIG. 42A

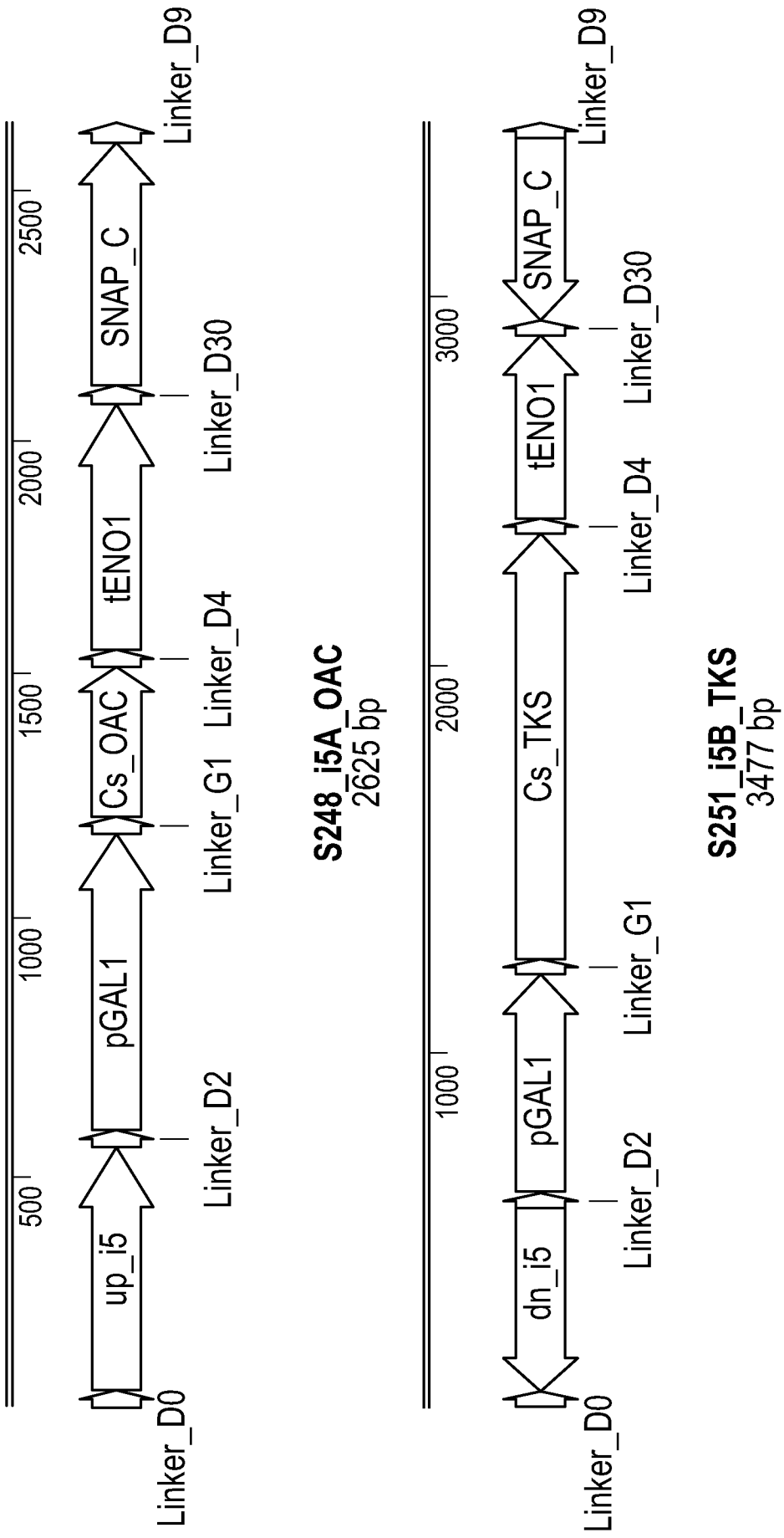

500
1000
1500
2000
2500
Linker_D0
up_i5
Linker_D2
pGAL1
Linker_G1
Cs_OAC
Linker_D4
tENO1
Linker_D30
SNAP_C
Linker_D9
S248_i5A_OAC
2625 bp
1000
2000
3000
Linker_D0
dn_i5
Linker_D2
pGAL1
Linker_G1
Cs_TKS
Linker_D4
tENO1
Linker_D30
SNAP_C
Linker_D9
S251_i5B_TKS
3477 bp

FIG. 42B

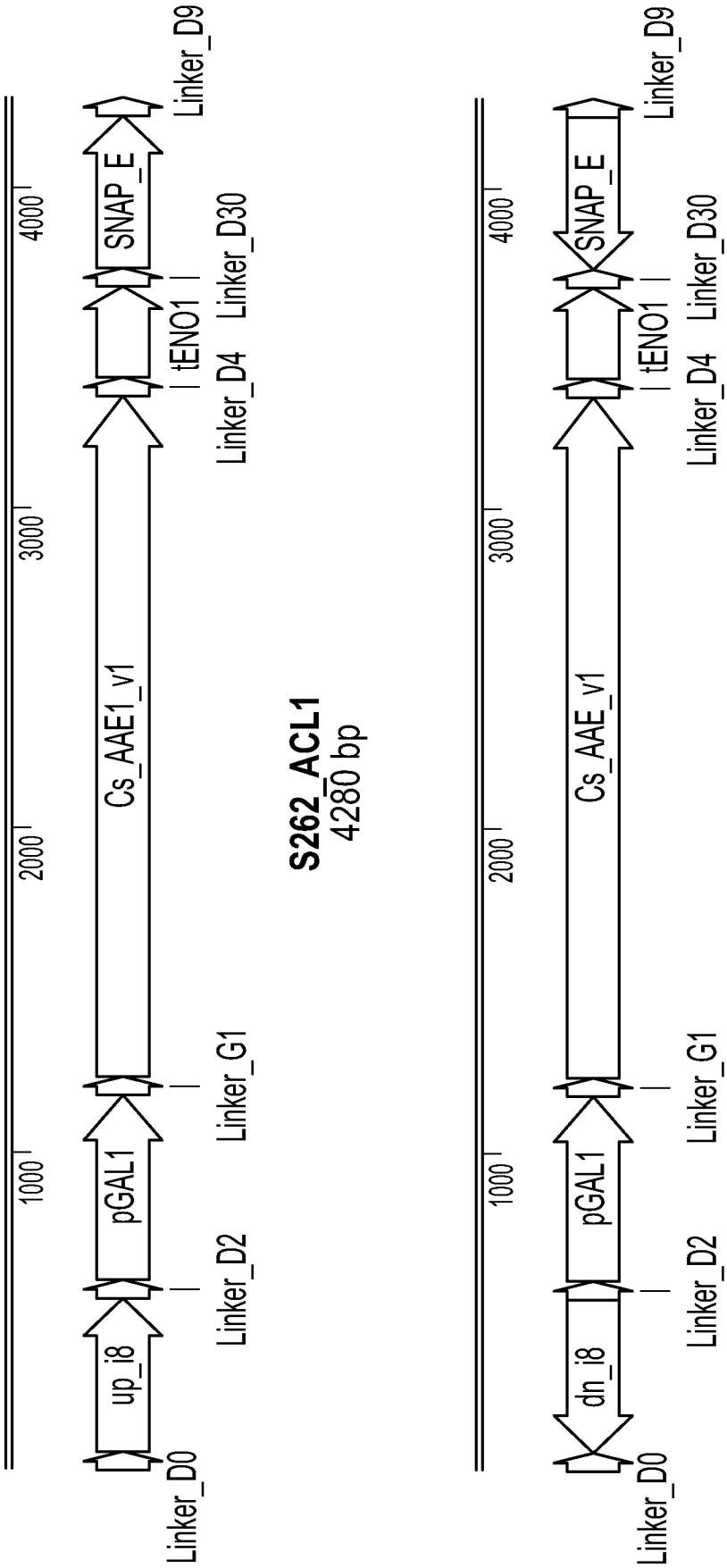

1000
2000
3000
4000
Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Cs_AAE1_v1
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S262_ACL1
4280 bp
1000
2000
3000
4000
Linker_D0
dn_i8
Linker_D2
pGAL1
Linker_G1
Cs_AAE_v1
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S274_ACL13
4280 bp

FIG. 42C

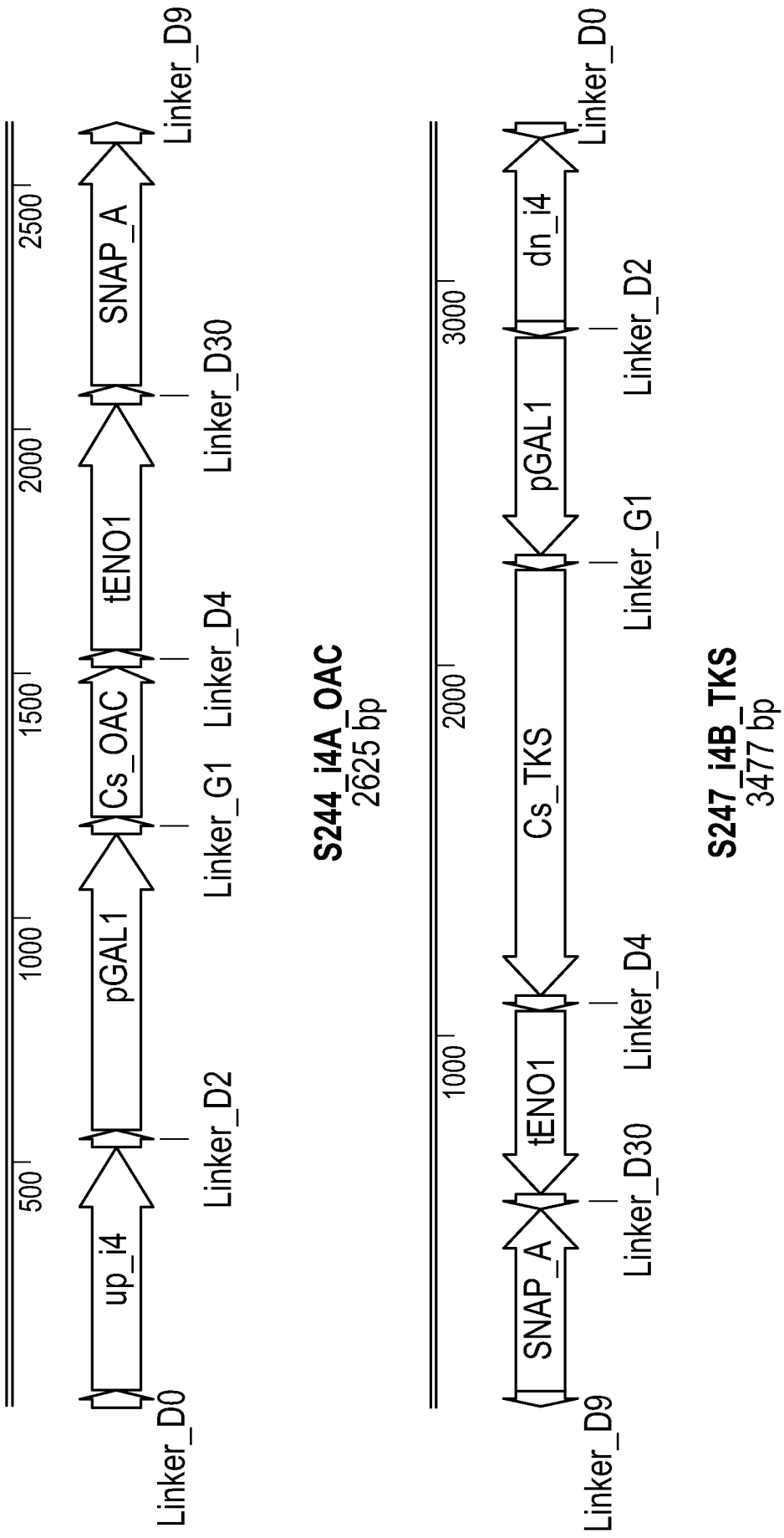

500
1000
1500
2000
2500
Linker_D0
up_i4
Linker_D2
pGAL1
Linker_G1
Cs_OAC
Linker_D4
tENO1
Linker_D30
SNAP_A
Linker_D9
S244_i4A_OAC
2625 bp
1000
2000
3000
Linker_D9
SNAP_A
Linker_D30
tENO1
Linker_D4
Cs_TKS
Linker_G1
pGAL1
Linker_D2
dn_i4
Linker_D0
S247_i4B_TKS
3477 bp

FIG. 43A

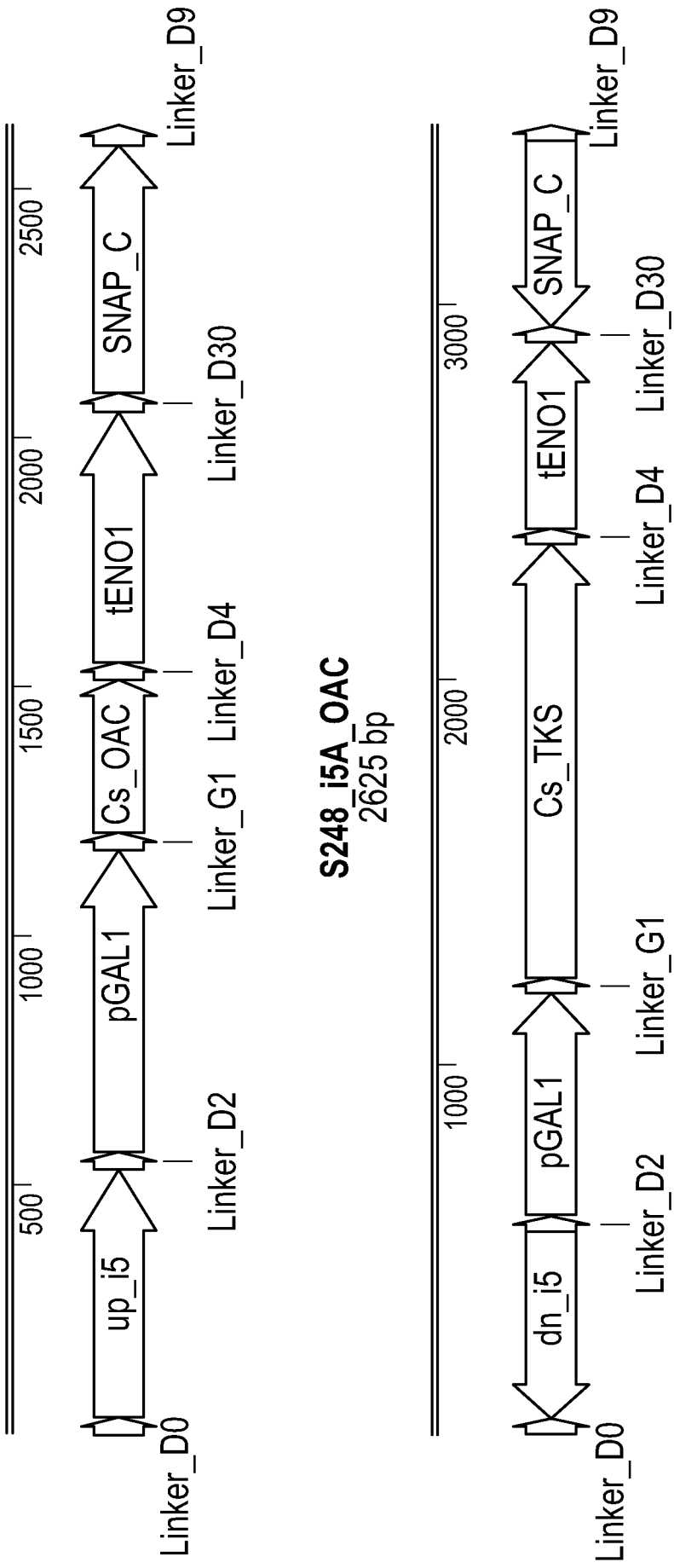

500
1000
1500
2000
2500
Linker_D0
up_i5
Linker_D2
pGAL1
Linker_G1
Cs_OAC
Linker_D4
tENO1
Linker_D30
SNAP_C
Linker_D9
S248_i5A_OAC
2625 bp
1000
2000
3000
Linker_D0
dn_i5
Linker_D2
pGAL1
Linker_G1
Cs_TKS
Linker_D4
tENO1
Linker_D30
SNAP_C
Linker_D9
S251_i5B_TKS
3477 bp

FIG. 43B

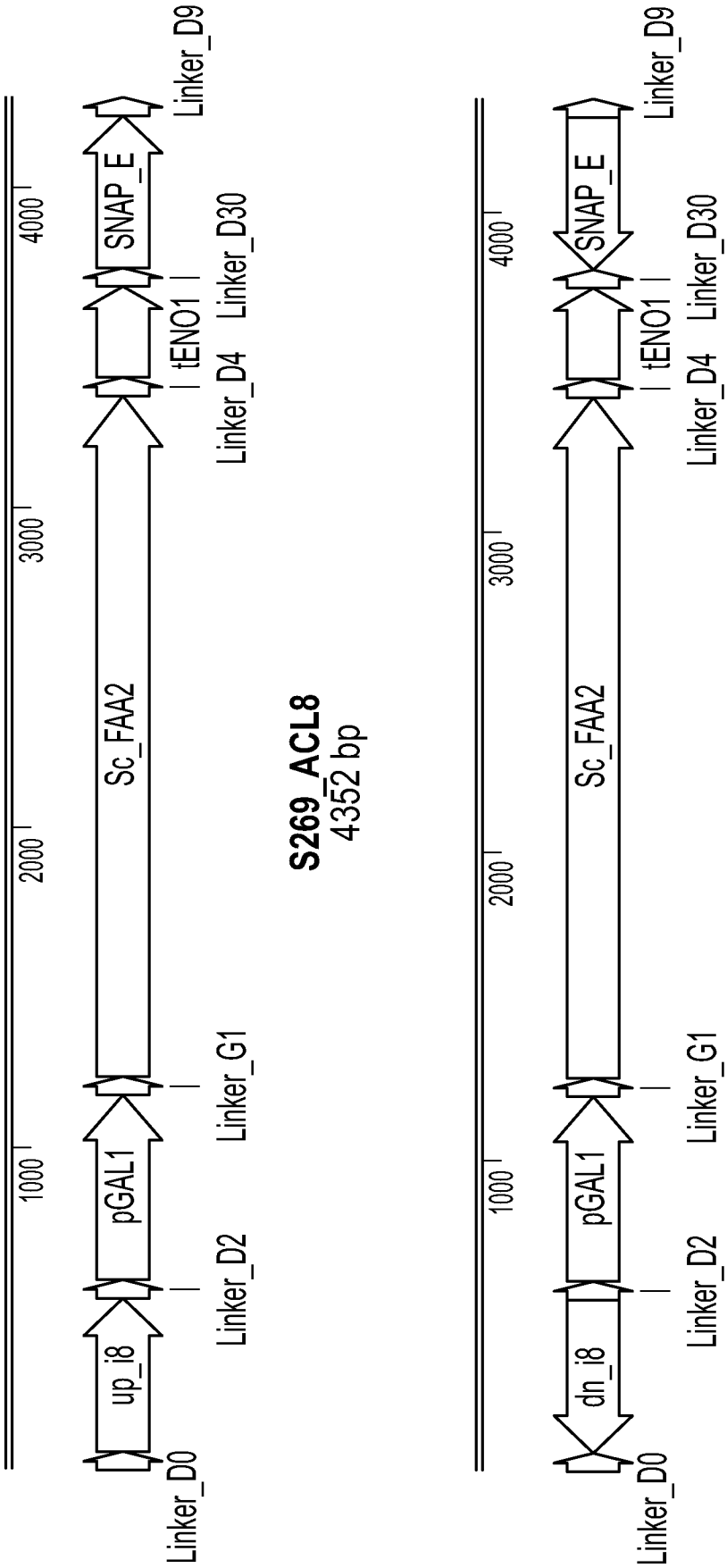

1000
2000
3000
4000
Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA2
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S269_ACL8
4352 bp
1000
2000
3000
4000
Linker_D0
dn_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA2
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S281_ACL20
4352 bp

FIG. 43C

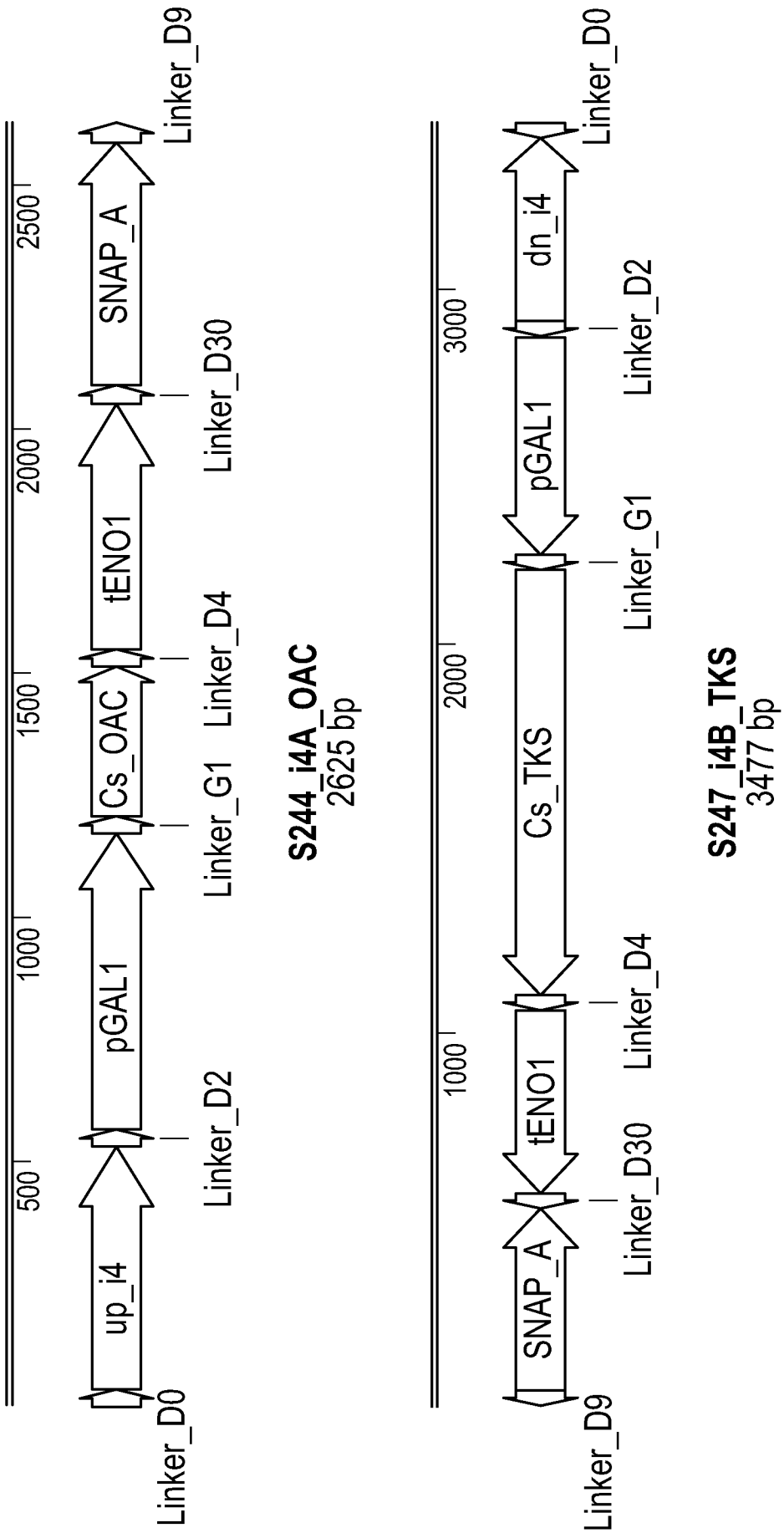

500
1000
1500
2000
2500
Linker_D0
up_i4
Linker_D2
pGAL1
Linker_G1
Cs_OAC
Linker_D4
tENO1
Linker_D30
SNAP_A
Linker_D9
S244_i4A_OAC
2625 bp
1000
2000
3000
Linker_D9
SNAP_A
Linker_D30
tENO1
Linker_D4
Cs_TKS
Linker_G1
pGAL1
Linker_D2
dn_i4
Linker_D0
S247_i4B_TKS
3477 bp

FIG. 44A

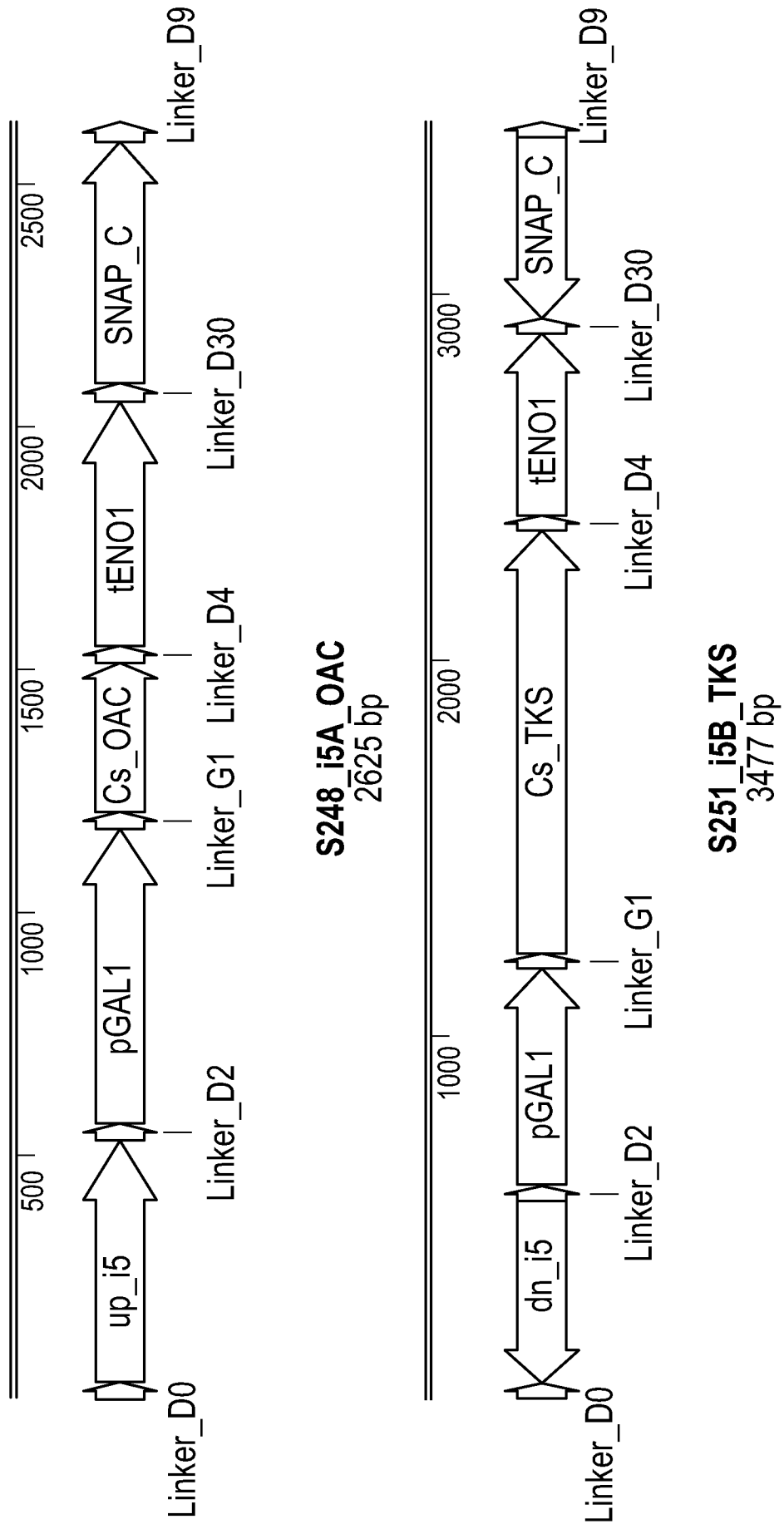

500
1000
1500
2000
2500
up_i5
pGAL1
Cs_OAC
tENO1
SNAP_C
Linker_D0
Linker_D2
Linker_G1
Linker_D4
Linker_D30
Linker_D9
S248_i5A_OAC
2625 bp
1000
2000
3000
dn_i5
pGAL1
Cs_TKS
tENO1
SNAP_C
Linker_D0
Linker_D2
Linker_G1
Linker_D4
Linker_D30
Linker_D9
S251_i5B_TKS
3477 bp

FIG. 44B

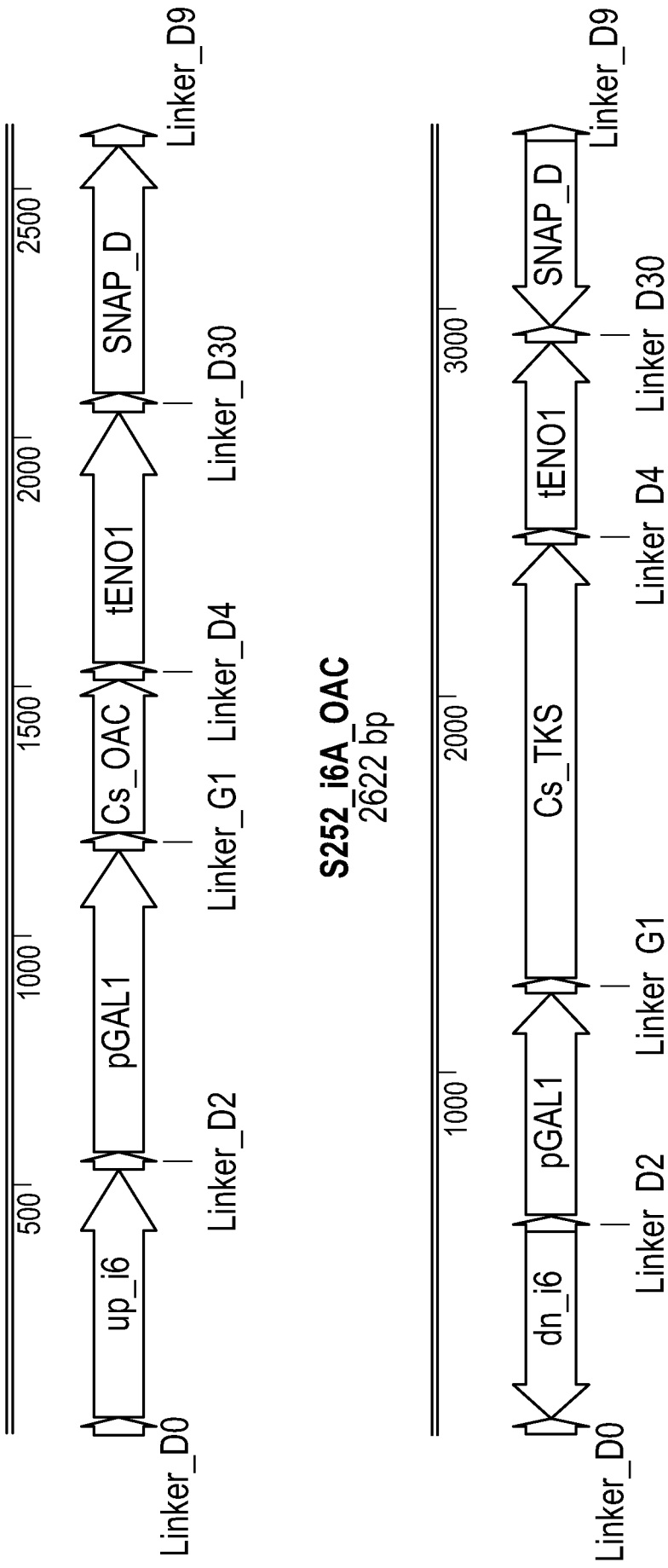

500
1000
1500
2000
2500
Linker_D0
up_i6
Linker_D2
pGAL1
Linker_G1
Cs_OAC
Linker_D4
tENO1
Linker_D30
SNAP_D
Linker_D9
S252_i6A_OAC
2622 bp
1000
2000
3000
Linker_D0
dn_i6
Linker_D2
pGAL1
Linker_G1
Cs_TKS
Linker_D4
tENO1
Linker_D30
SNAP_D
Linker_D9
S255_i6B_TKS
3474 bp

FIG. 44C

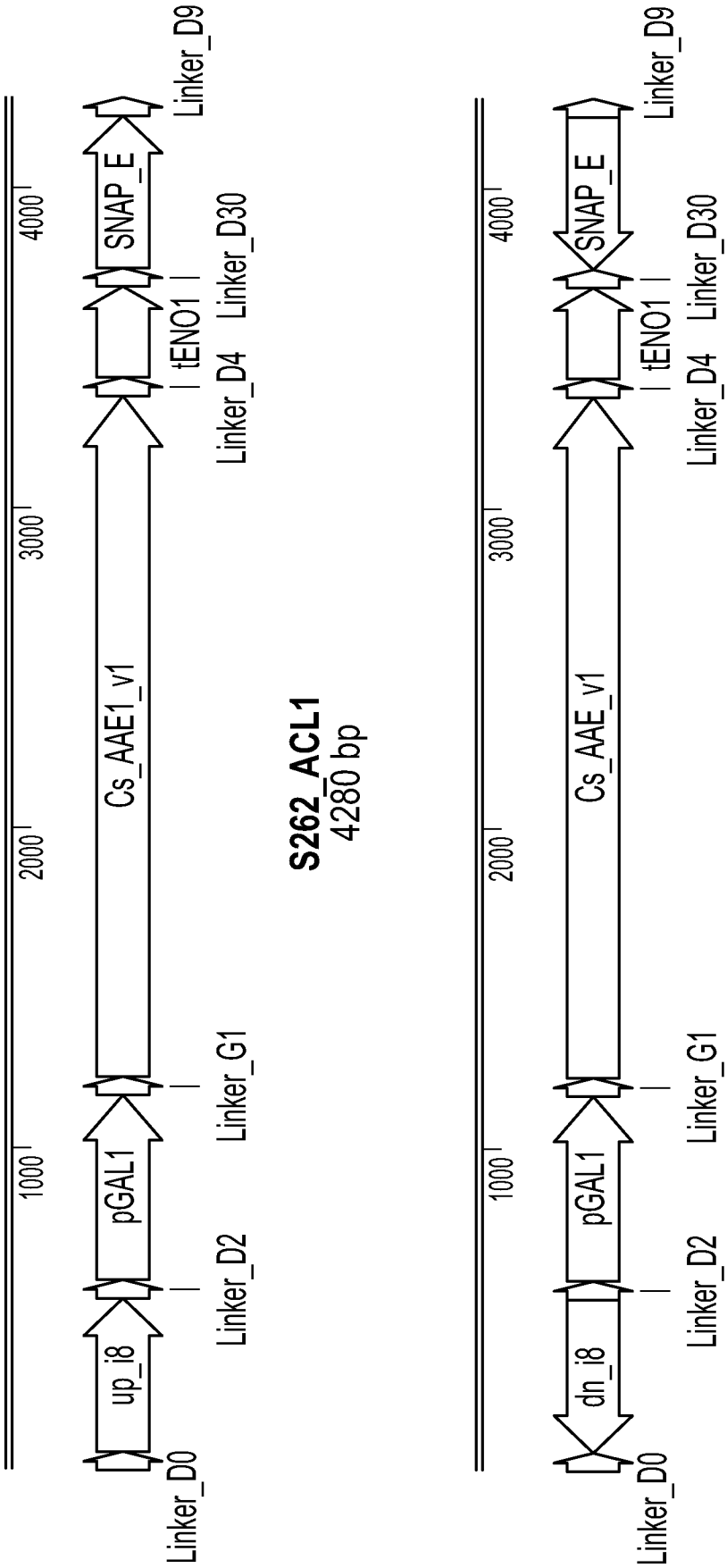

1000
2000
3000
4000
Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Cs_AAE1_v1
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S262_ACL1
4280 bp
1000
2000
3000
4000
Linker_D0
dn_i8
Linker_D2
pGAL1
Linker_G1
Cs_AAE_v1
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S274_ACL13
4280 bp

FIG. 45

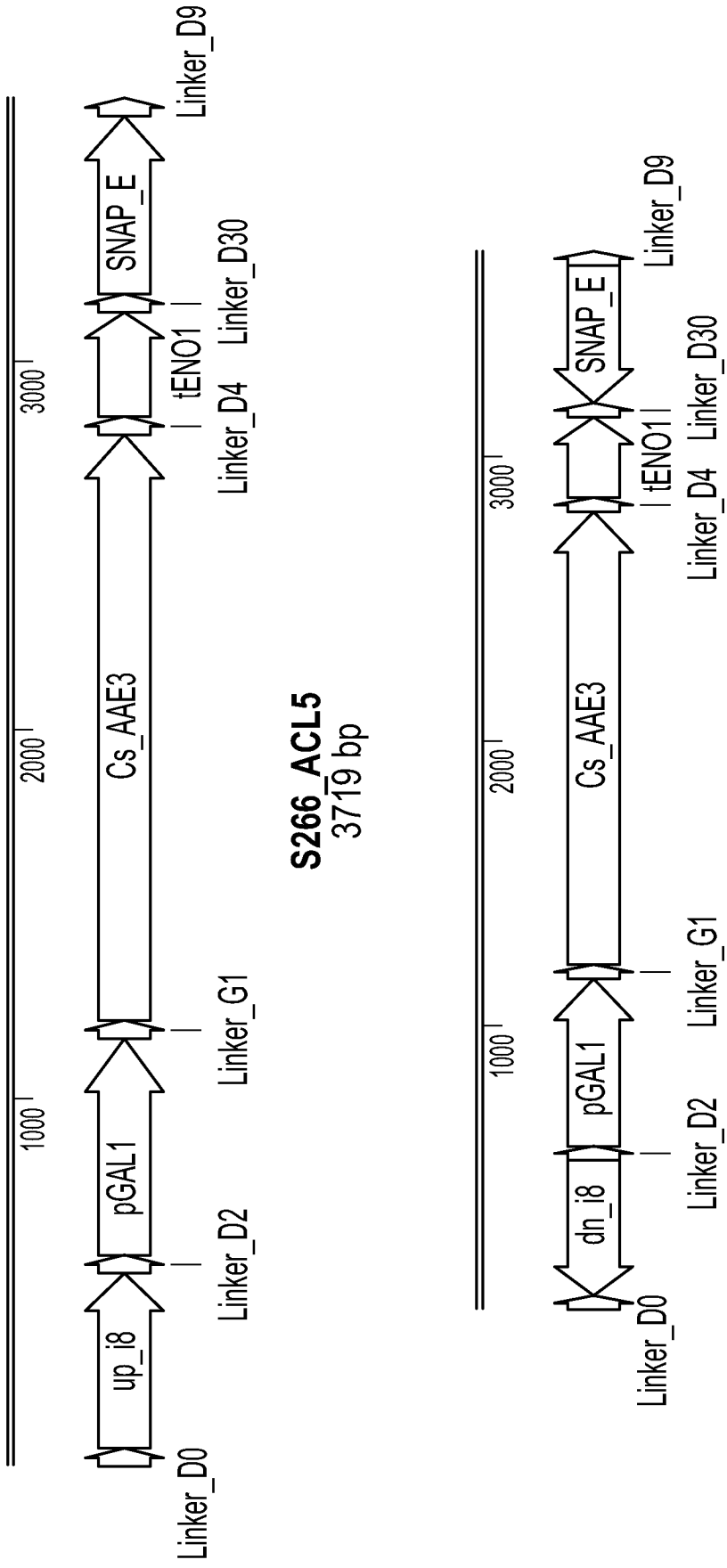

1000
2000
3000
up_i8
pGAL1
Cs_AAE3
SNAP_E
Linker_D0
Linker_D2
Linker_G1
tENO1
Linker_D4
Linker_D30
Linker_D9
S266_ACL5
3719 bp
1000
2000
3000
dn_i8
pGAL1
Cs_AAE3
SNAP_E
Linker_D0
Linker_D2
Linker_G1
tENO1
Linker_D4
Linker_D30
Linker_D9
S278_ACL17
3719 bp

FIG. 46

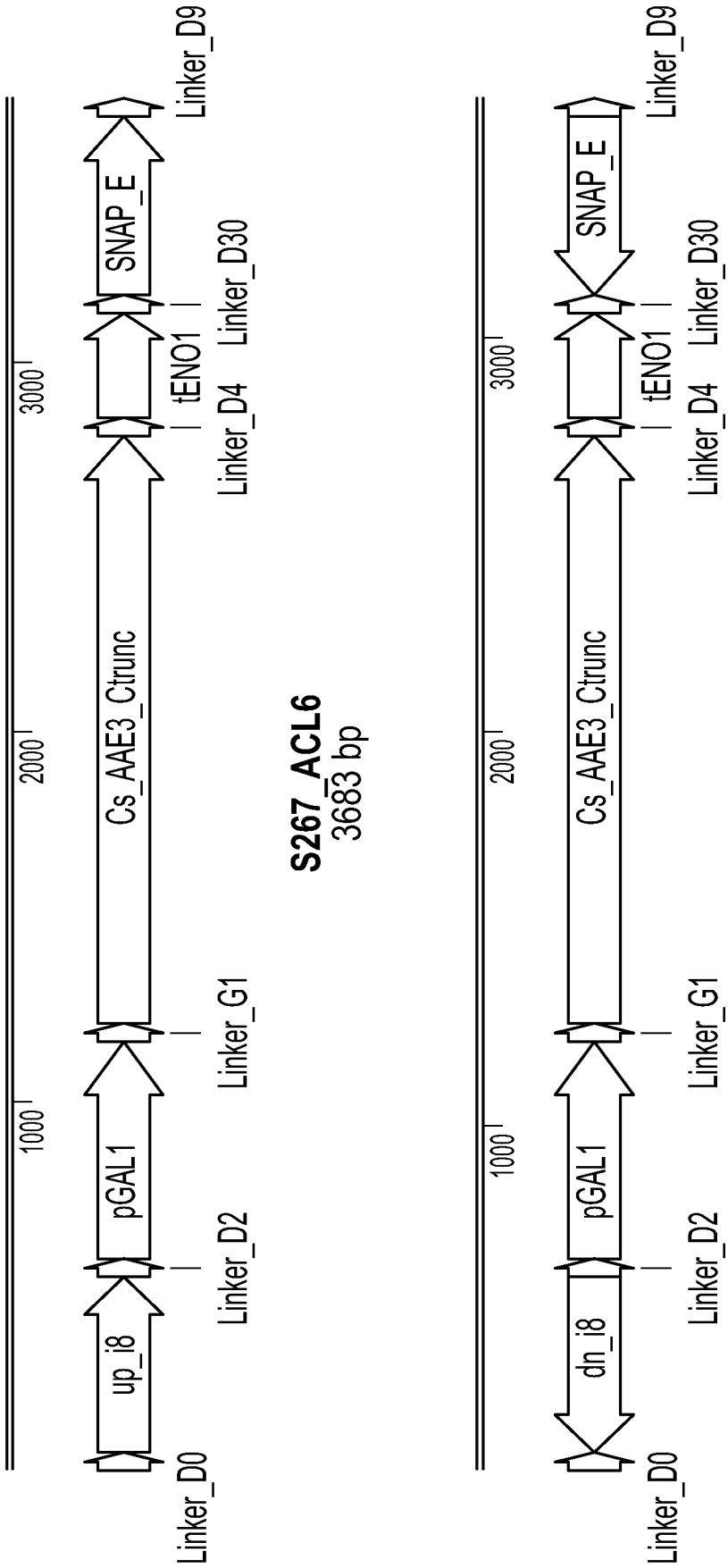

Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Cs_AAE3_Ctrunc
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
1000
2000
3000
S267_ACL6
3683 bp
Linker_D0
dn_i8
Linker_D2
pGAL1
Linker_G1
Cs_AAE3_Ctrunc
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
1000
2000
3000
S279_ACL18
3683 bp

FIG. 47

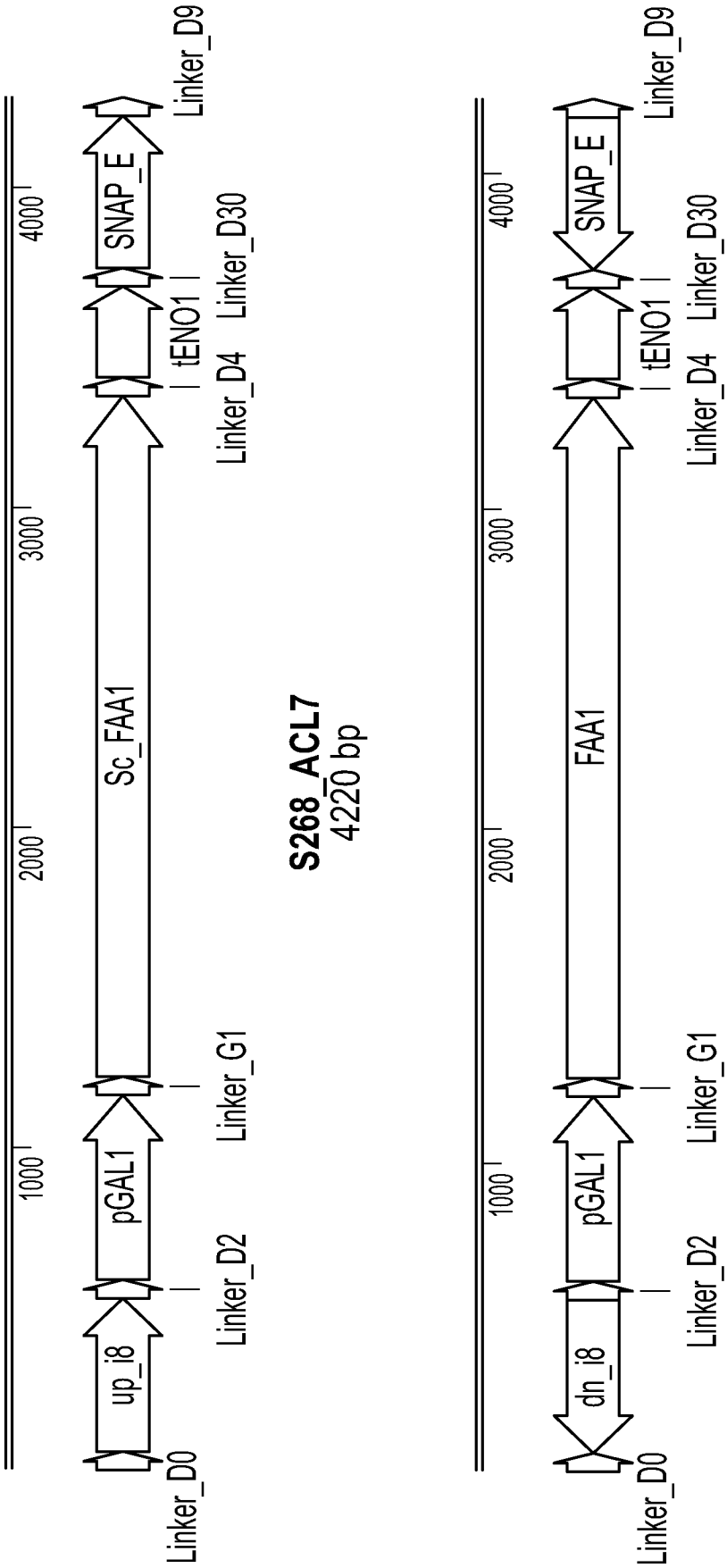

1000
2000
3000
4000
Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA1
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S268_ACL7
4220 bp
1000
2000
3000
4000
Linker_D0
dn_i8
Linker_D2
pGAL1
Linker_G1
FAA1
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S280_ACL19
4220 bp

FIG. 48

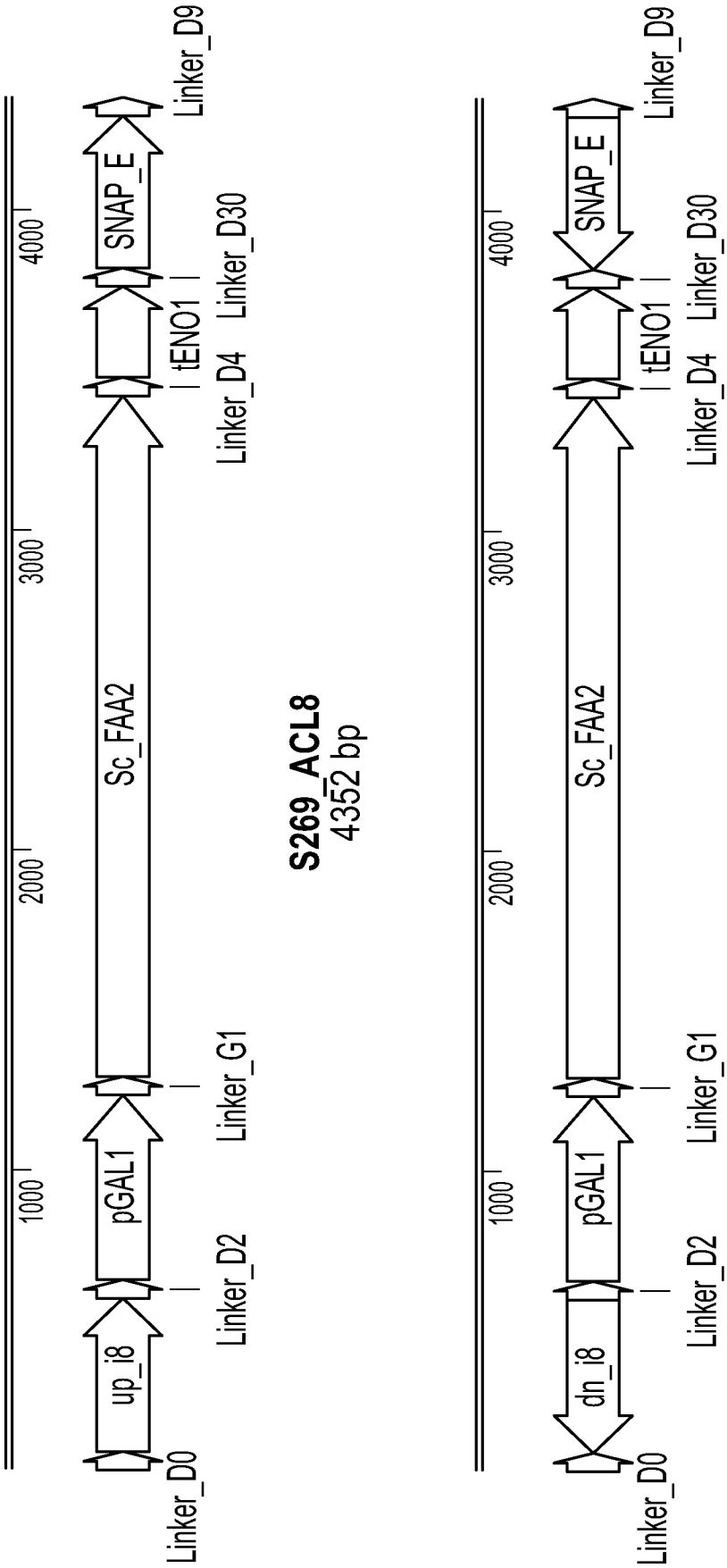

1000
2000
3000
4000
Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA2
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S269_ACL8
4352 bp
1000
2000
3000
4000
Linker_D0
dn_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA2
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S281_ACL20
4352 bp

FIG. 49

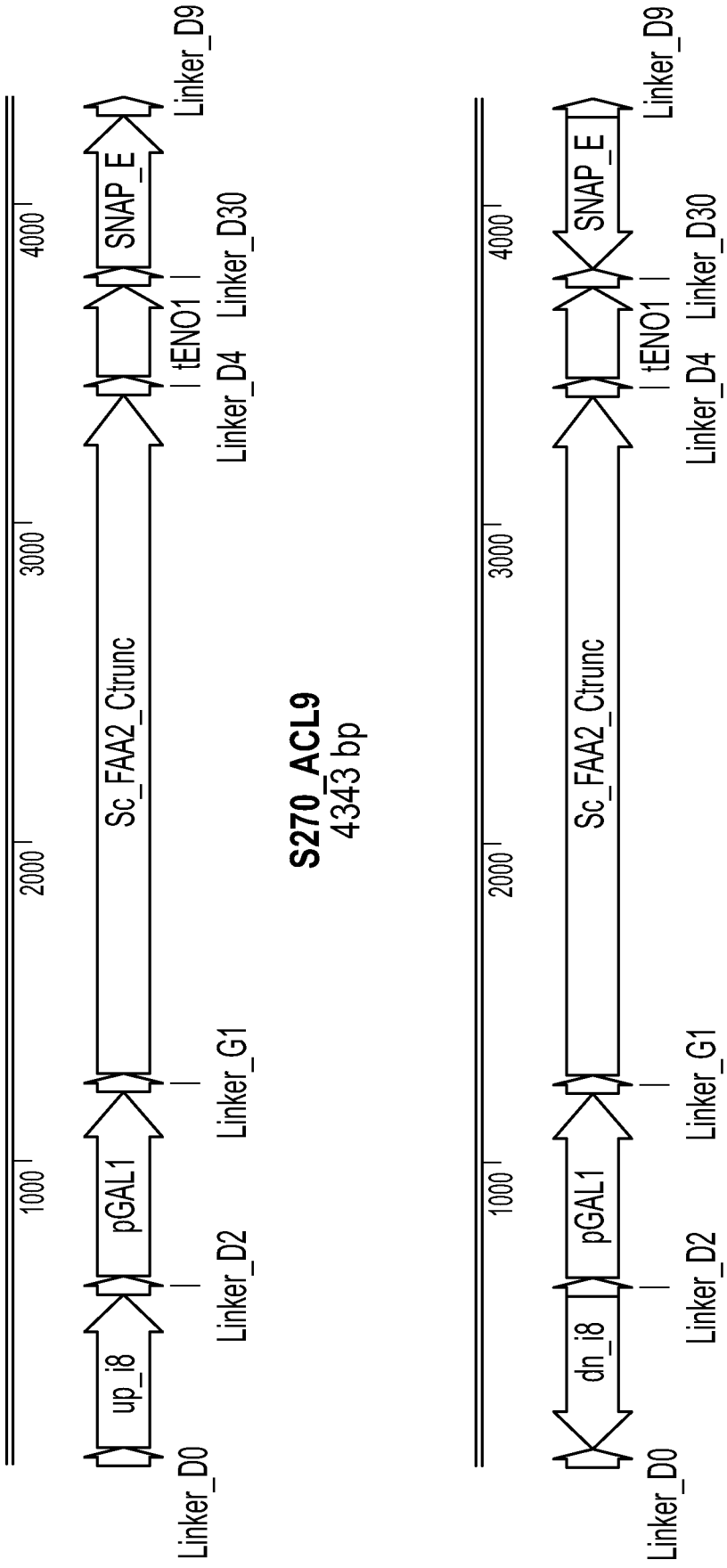

1000
2000
3000
4000
Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA2_Ctrunc
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S270_ACL9
4343 bp
1000
2000
3000
4000
Linker_D0
dn_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA2_Ctrunc
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S282_ACL21
4343 bp

FIG. 50

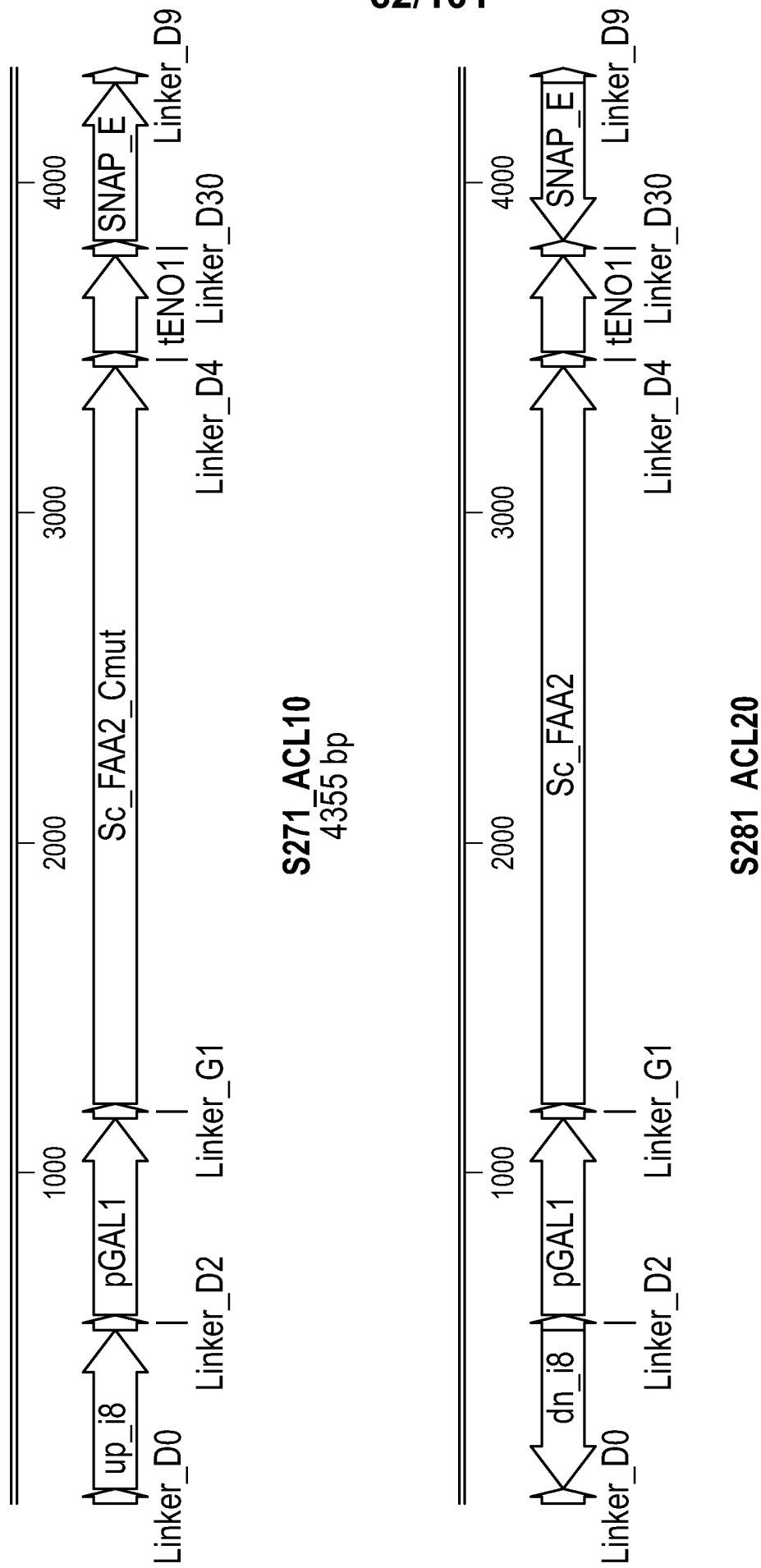

1000
2000
3000
4000
up_i8
pGAL1
Sc_FAA2_Cmut
SNAP_E
Linker_D0
Linker_D2
Linker_G1
Linker_D4
tENO1
Linker_D30
Linker_D9
S271_ACL10
4355 bp
dn_i8
Sc_FAA2
S281_ACL20
4352 bp

FIG. 51

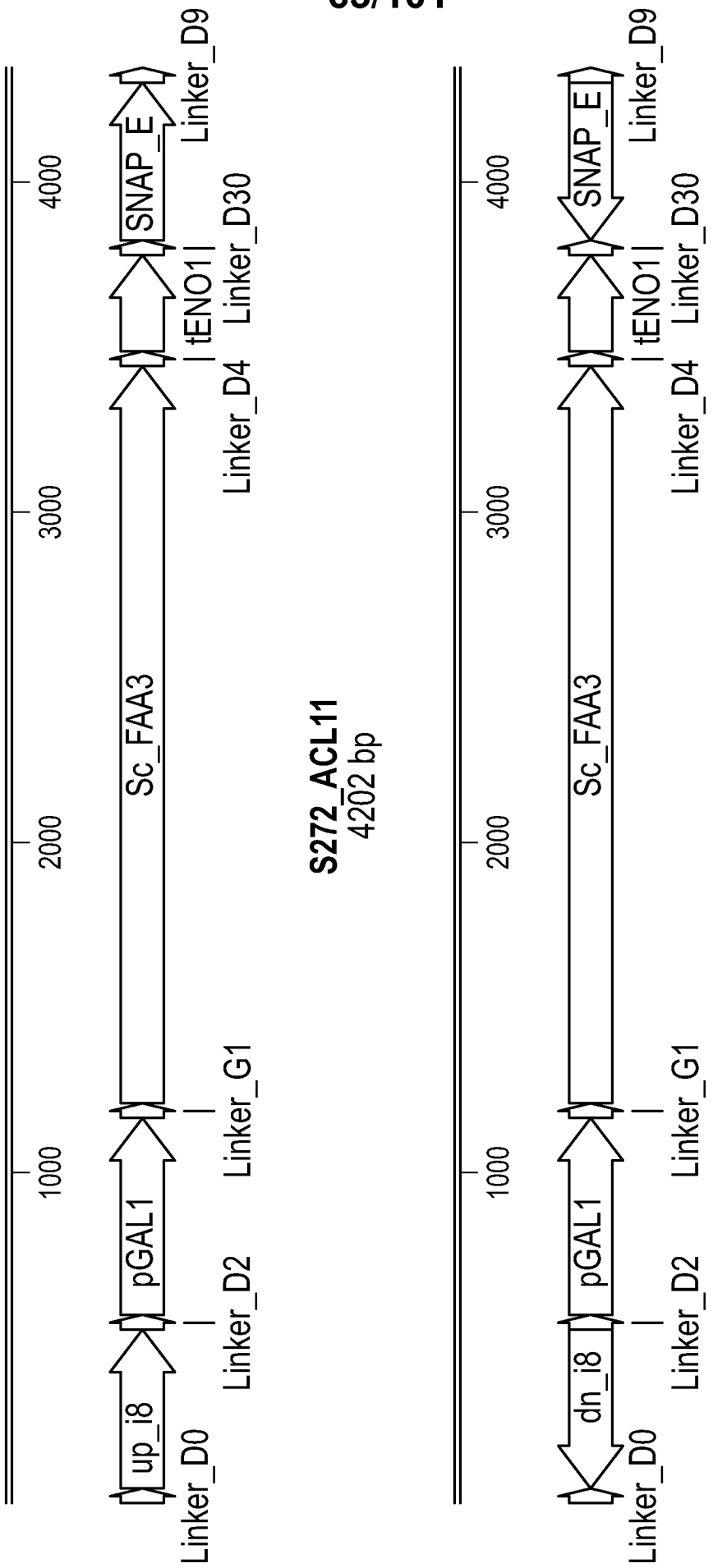

1000
2000
3000
4000
Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA3
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S272_ACL11
4202 bp
dn_i8
S284_ACL23
4202 bp

FIG. 52

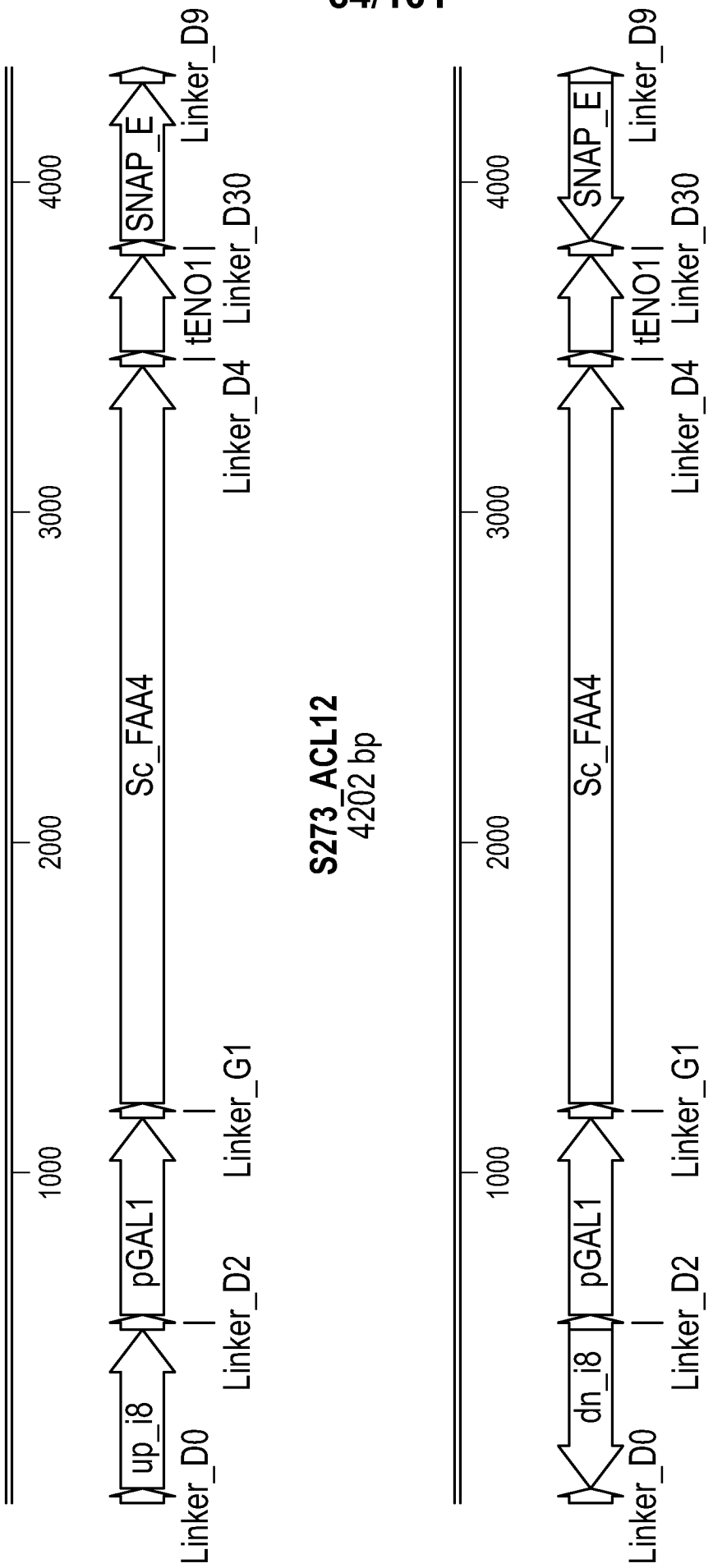

1000
2000
3000
4000
Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA4
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S273_ACL12
4202 bp
1000
2000
3000
4000
Linker_D0
dn_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA4
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S285_ACL24
4202 bp

FIG. 53

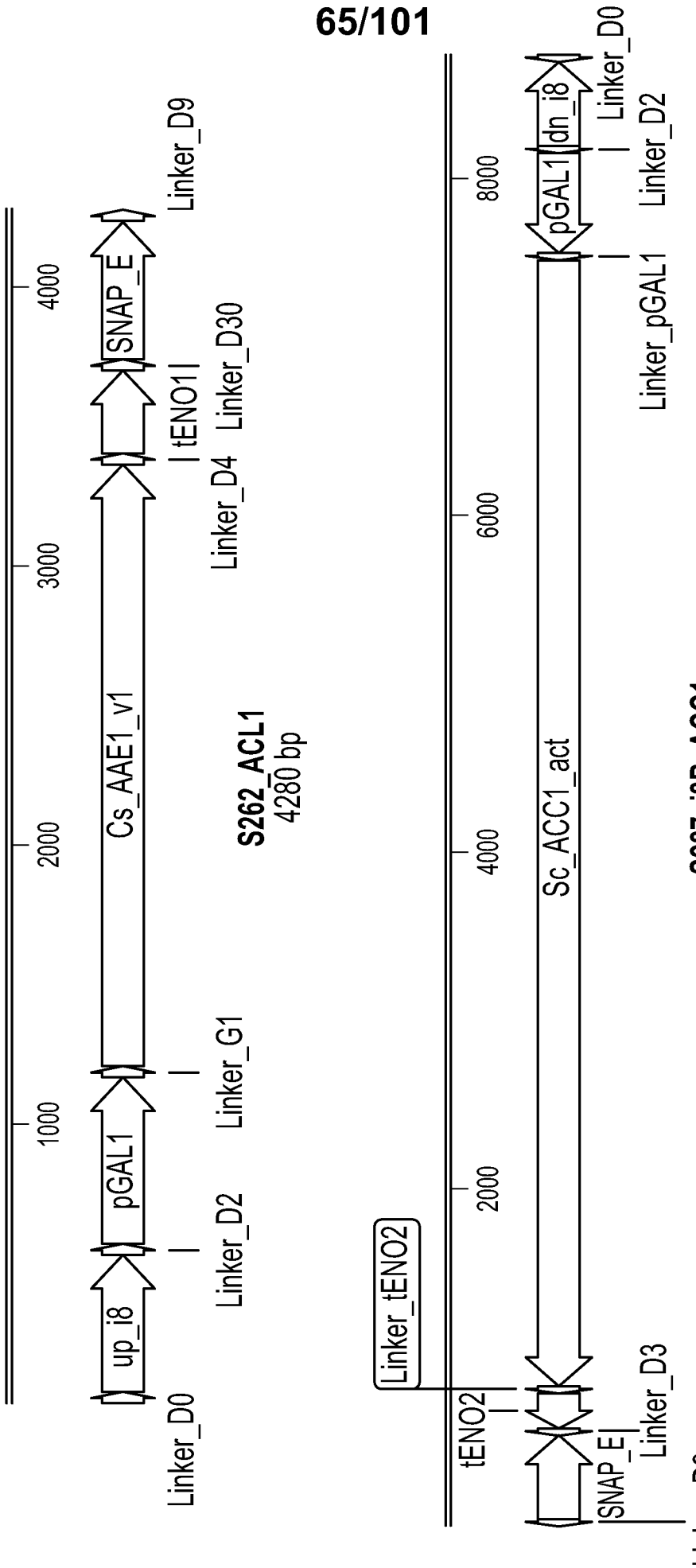

1000
2000
3000
4000
up_i8
pGAL1
Cs_AAE1_v1
SNAP_E
Linker_D0
Linker_D2
Linker_G1
Linker_D4
tENO1
Linker_D30
Linker_D9
S262_ACL1
4280 bp
2000
4000
6000
8000
Linker_tENO2
tENO2
Sc_ACC1_act
SNAP_E
Linker_D3
Linker_D9
pGAL1
dn_i8
Linker_pGAL1
Linker_D2
Linker_D0
S287_i8B_ACC1
8732 bp

FIG. 54

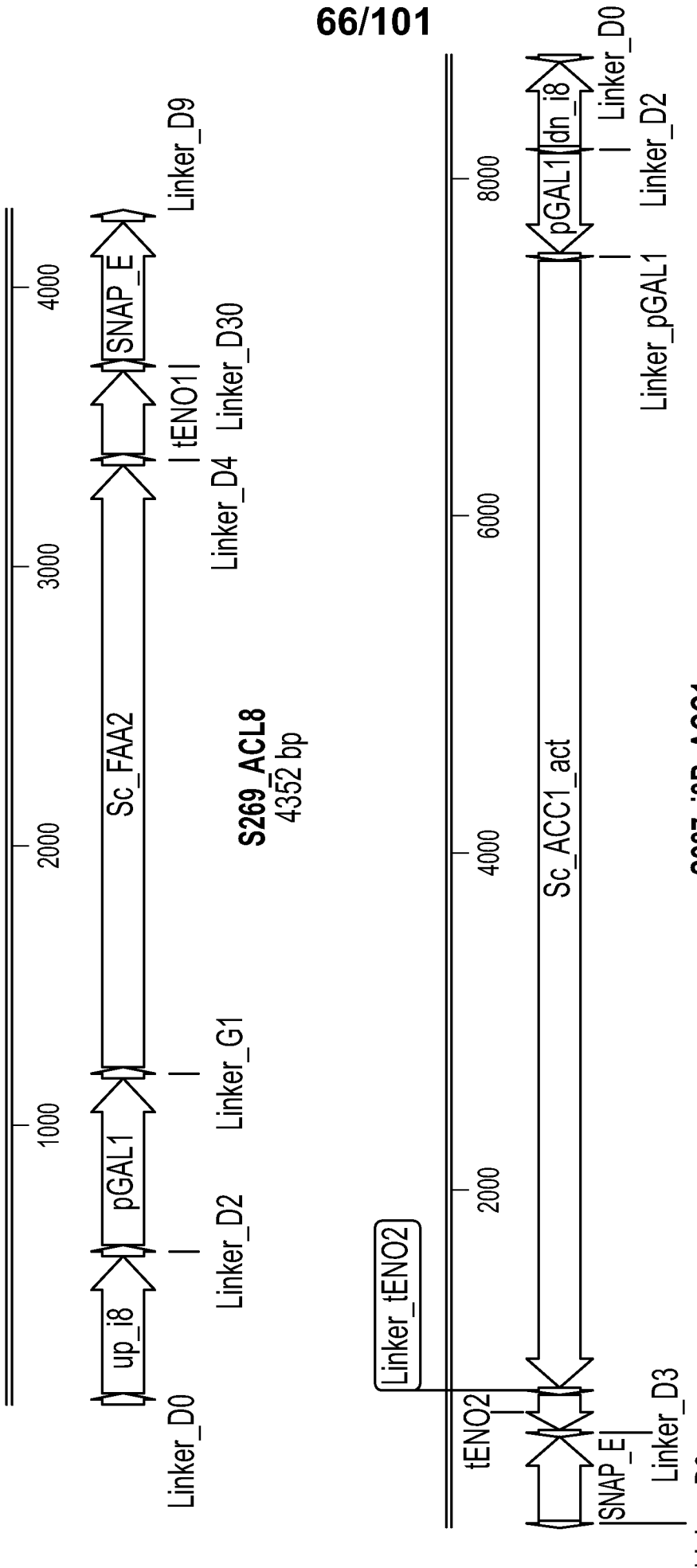

1000
2000
3000
4000
up_i8
Linker_D0
pGAL1
Linker_D2
Sc_FAA2
Linker_G1
SNAP_E
Linker_D9
tENO1
Linker_D4
Linker_D30
S269_ACL8
4352 bp
2000
4000
6000
8000
Linker_tENO2
tENO2
SNAP_E
Linker_D3
Linker_D9
Sc_ACC1_act
pGAL1
Linker_pGAL1
dn_i8
Linker_D0
Linker_D2
S287_i8B_ACC1
8732 bp

FIG. 55

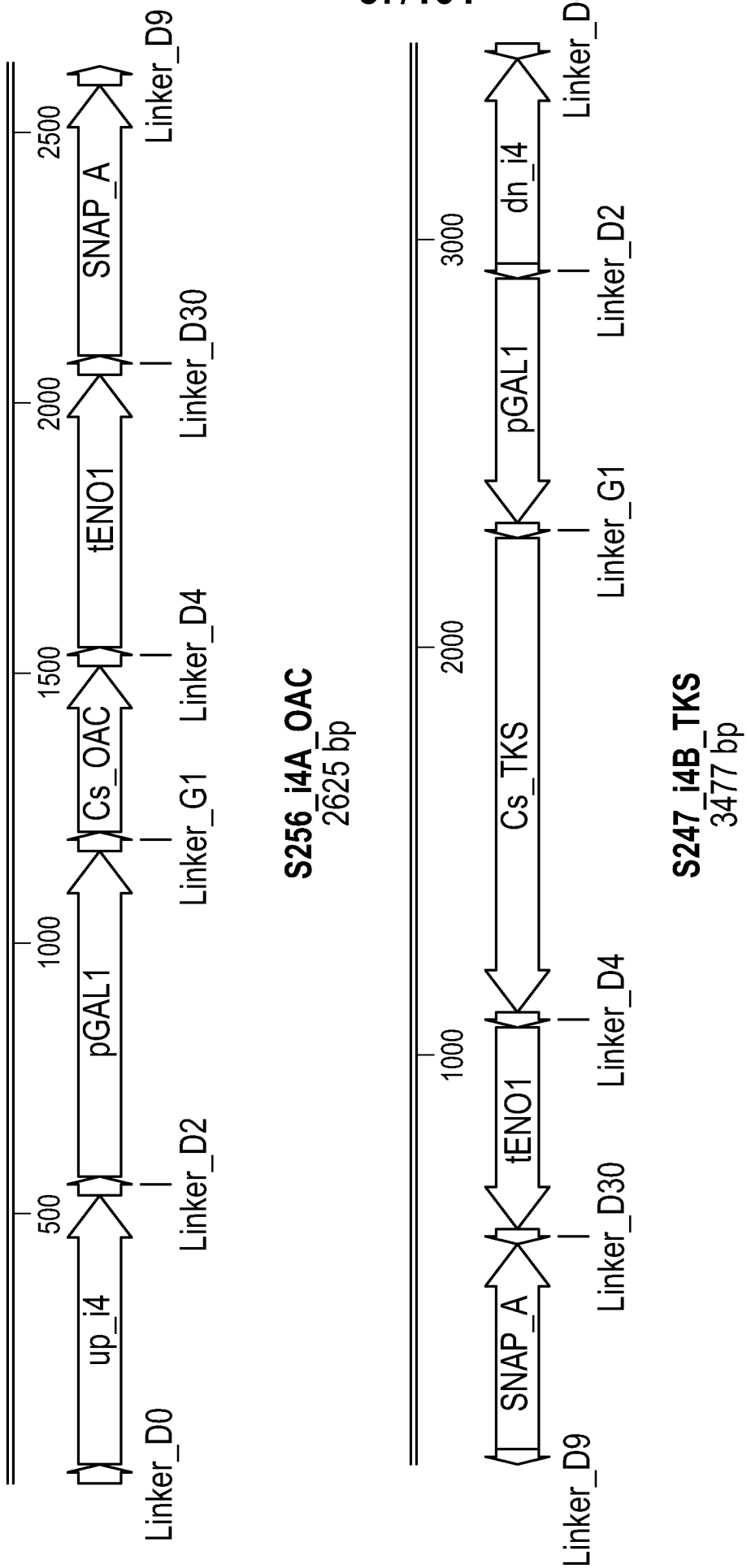

500
1000
1500
2000
2500
Linker_D0
up_i4
Linker_D2
pGAL1
Linker_G1
Cs_OAC
Linker_D4
tENO1
Linker_D30
SNAP_A
Linker_D9
S256_i4A_OAC
2625 bp
1000
2000
3000
Linker_D9
SNAP_A
Linker_D30
tENO1
Linker_D4
Cs_TKS
Linker_G1
pGAL1
Linker_D2
dn_i4
Linker_D0
S247_i4B_TKS
3477 bp

FIG. 56A

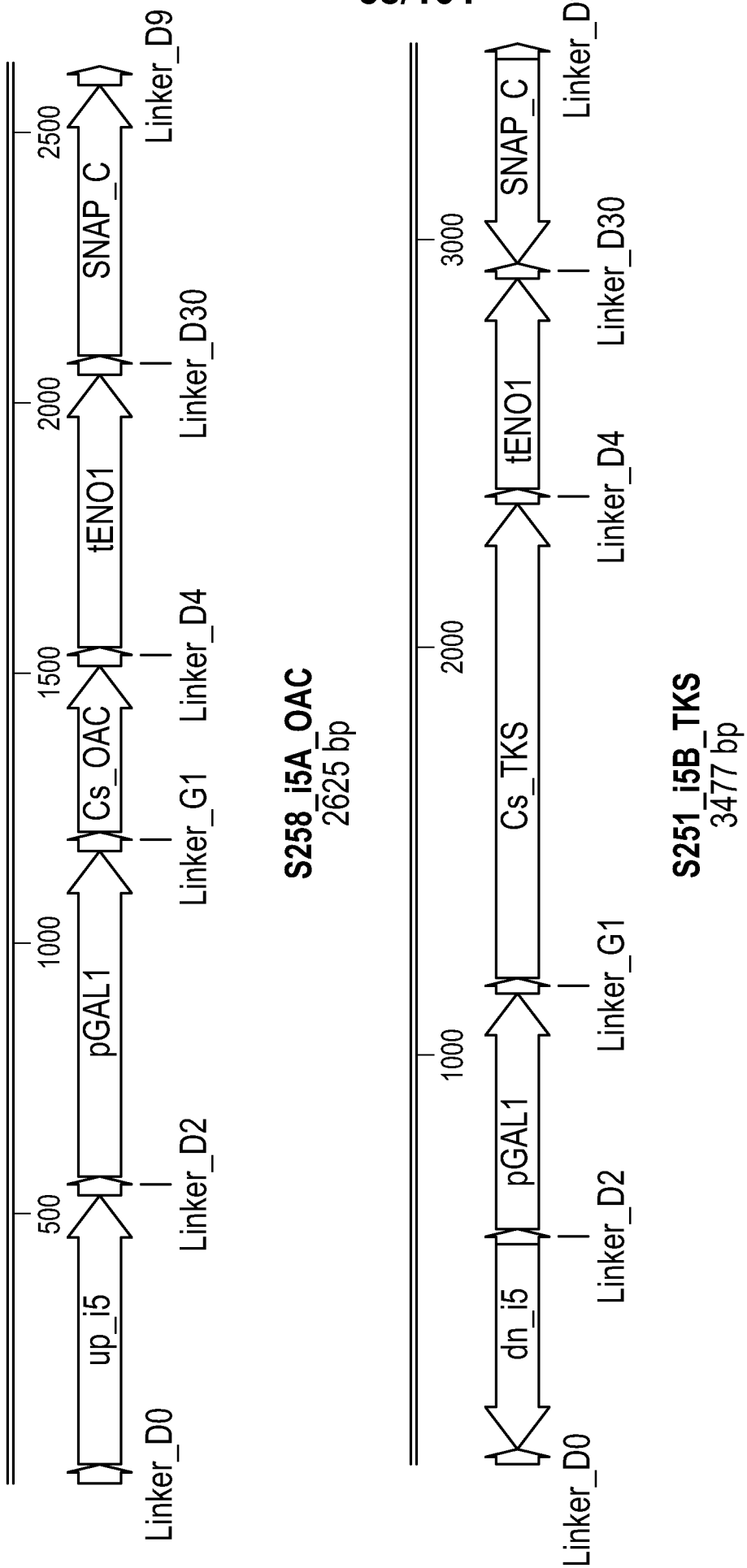

500
1000
1500
2000
2500
up_i5
pGAL1
Cs_OAC
tENO1
SNAP_C
Linker_D0
Linker_D2
Linker_G1
Linker_D4
Linker_D30
Linker_D9
S258_i5A_OAC
2625 bp
1000
2000
3000
dn_i5
pGAL1
Cs_TKS
tENO1
SNAP_C
Linker_D0
Linker_D2
Linker_G1
Linker_D4
Linker_D30
Linker_D9
S251_i5B_TKS
3477 bp

FIG. 56B

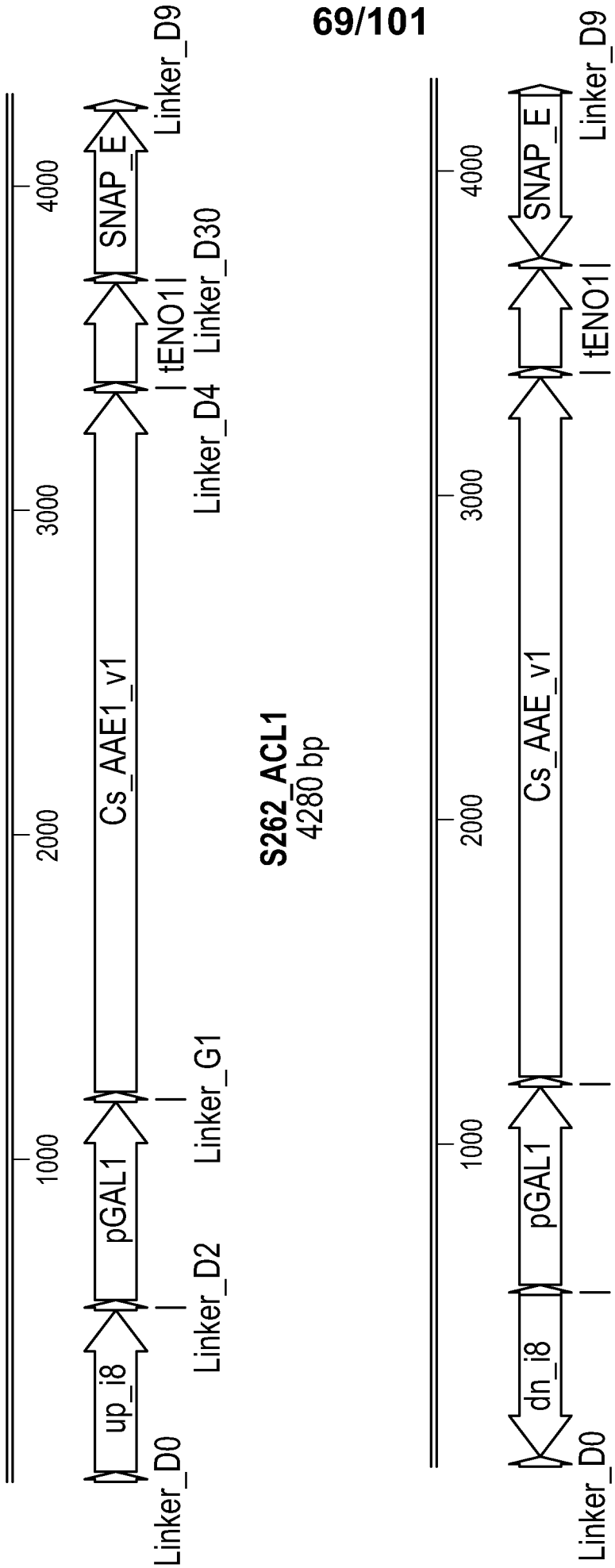

Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Cs_AAE1_v1
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
1000
2000
3000
4000
S262_ACL1
4280 bp
Linker_D0
dn_i8
Linker_D2
pGAL1
Linker_G1
Cs_AAE_v1
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
1000
2000
3000
4000
S274_ACL13
4280 bp

FIG. 56C

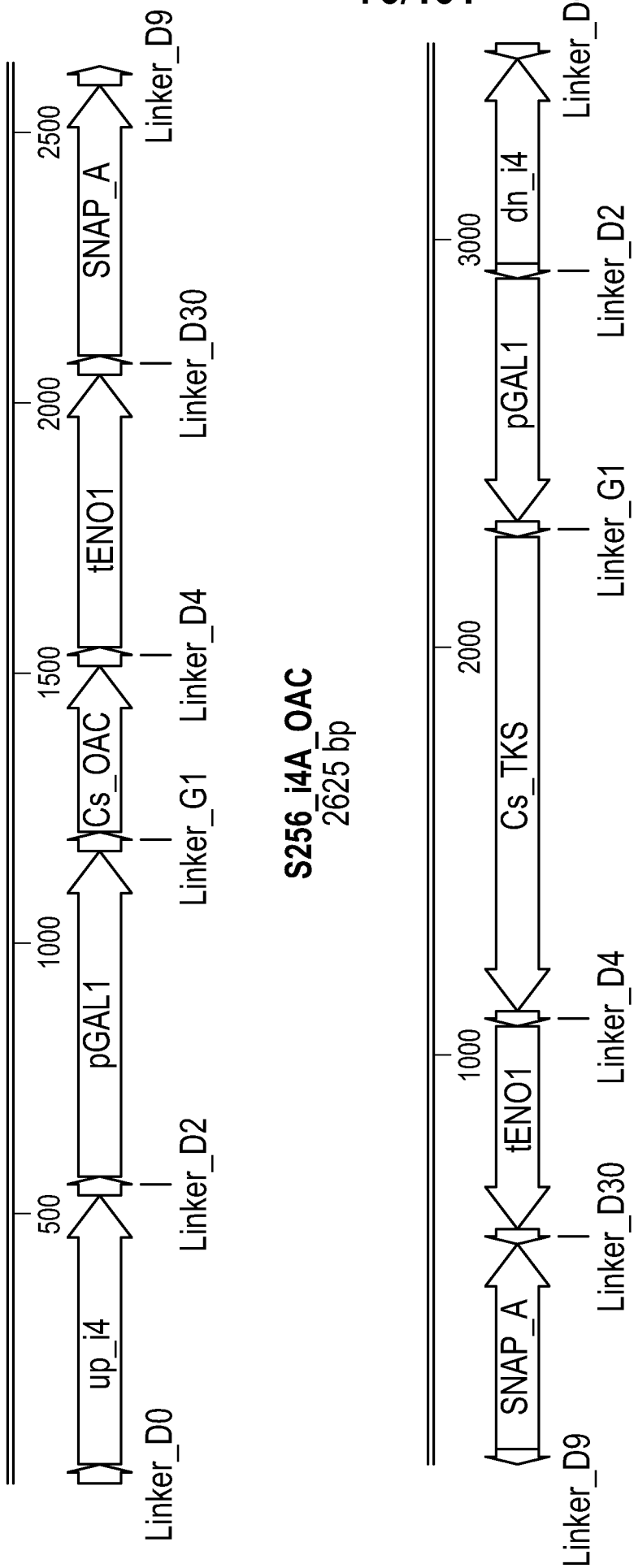

500
1000
1500
2000
2500
up_i4
pGAL1
Cs_OAC
tENO1
SNAP_A
Linker_D0
Linker_D2
Linker_G1
Linker_D4
Linker_D30
Linker_D9
S256_i4A_OAC
2625 bp
1000
2000
3000
SNAP_A
tENO1
Cs_TKS
pGAL1
dn_i4
Linker_D9
Linker_D30
Linker_D4
Linker_G1
Linker_D2
Linker_D0
S247_i4B_TKS
3477 bp

FIG. 57A

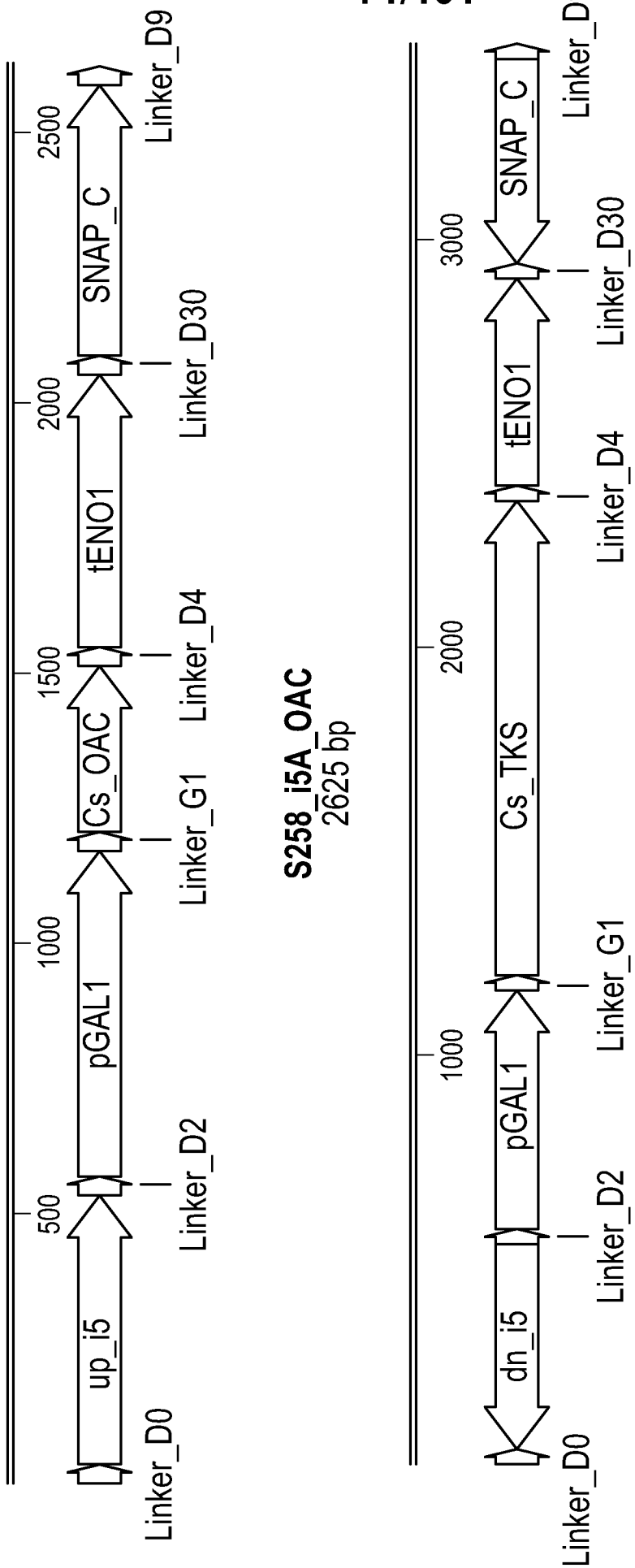

500
1000
1500
2000
2500
Linker_D0
up_i5
Linker_D2
pGAL1
Linker_G1
Cs_OAC
Linker_D4
tENO1
Linker_D30
SNAP_C
Linker_D9
S258_i5A_OAC
2625 bp
1000
2000
3000
Linker_D0
dn_i5
Linker_D2
pGAL1
Linker_G1
Cs_TKS
Linker_D4
tENO1
Linker_D30
SNAP_C
Linker_D9
S251_i5B_TKS
3477 bp

FIG. 57B

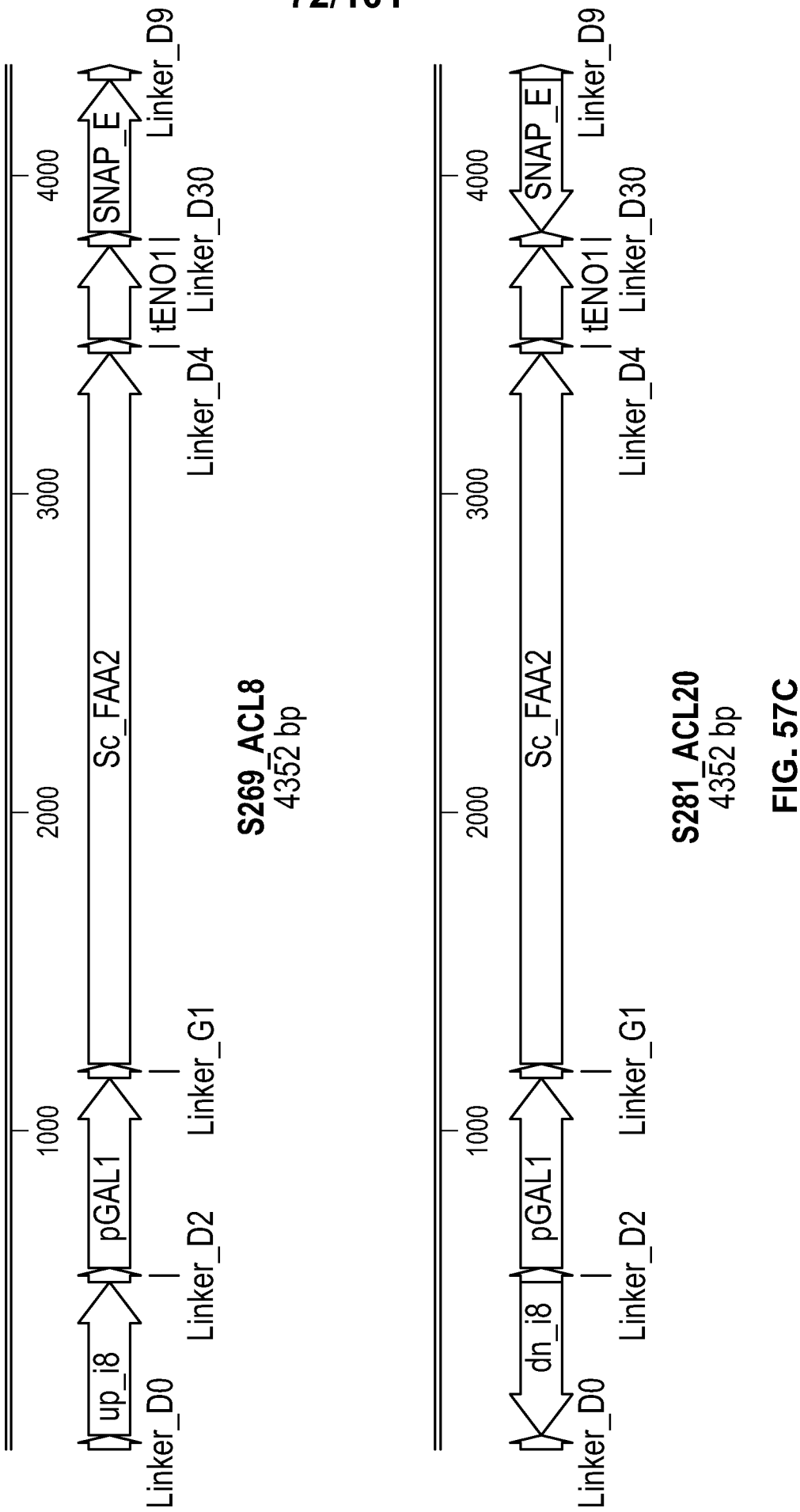
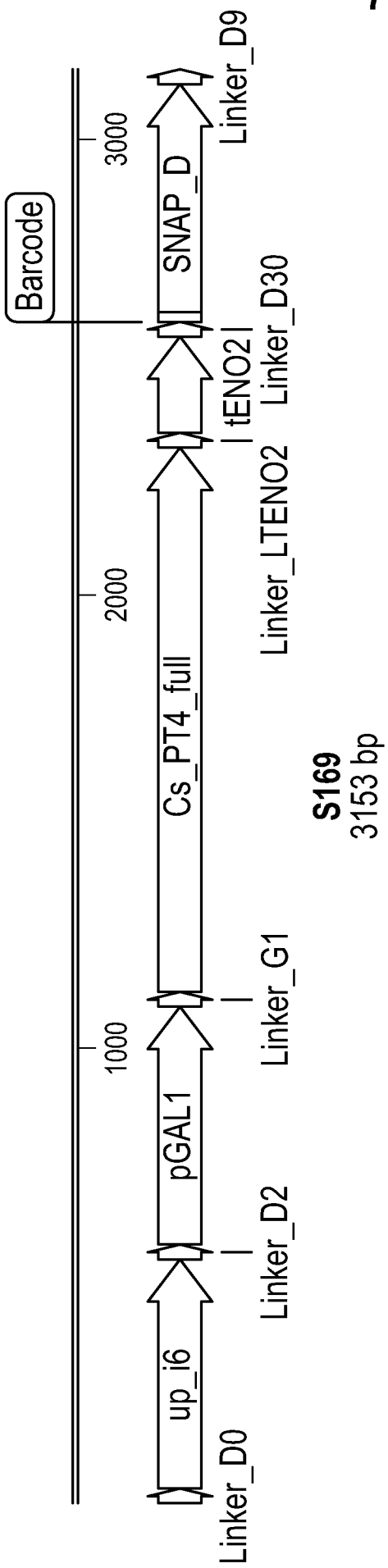
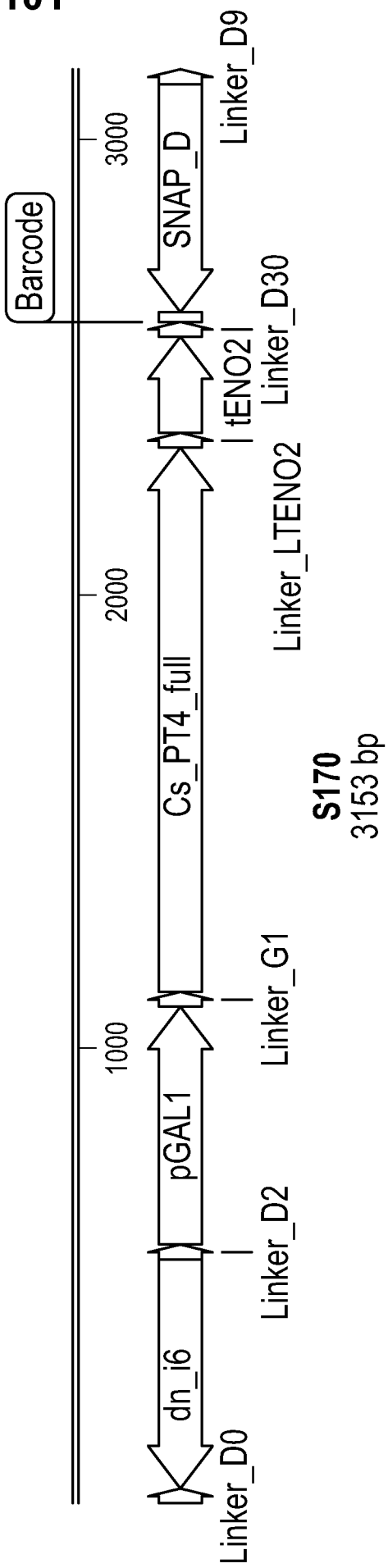

Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA2
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
1000
2000
3000
4000
S269_ACL8
4352 bp
Linker_D0
dn_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA2
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
1000
2000
3000
4000
S281_ACL20
4352 bp

FIG. 57C

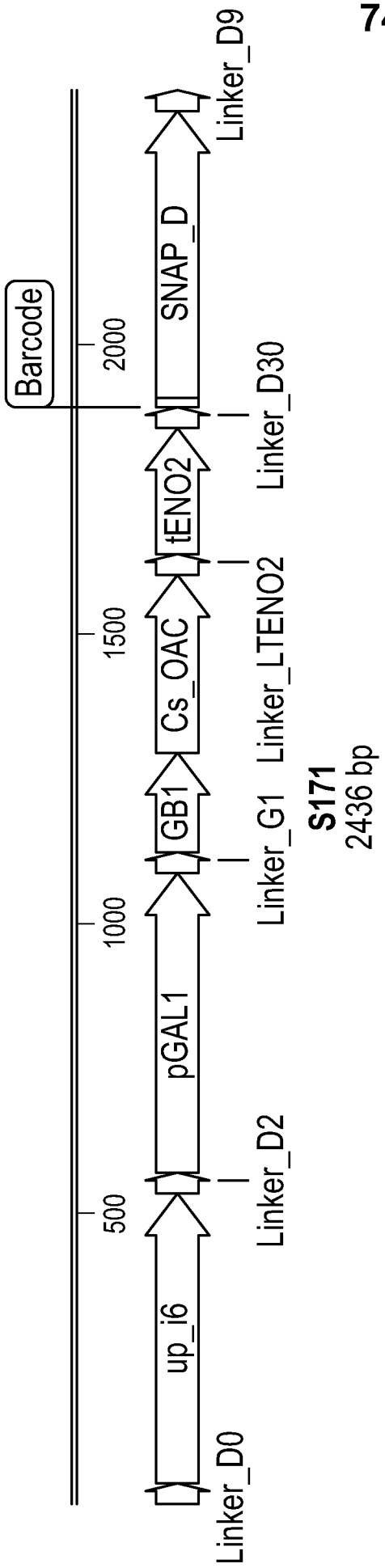


S169
3153 bp



S170
3153 bp

FIG. 58


Barcode
2000
SNAP_D
Linker_D9
1500
tENO2
Linker_D30
Cs_OAC
Linker_LTENO2
1000
GB1
Linker_G1
pGAL1
Linker_D2
500
up_i6
Linker_D0
S171
2436 bp

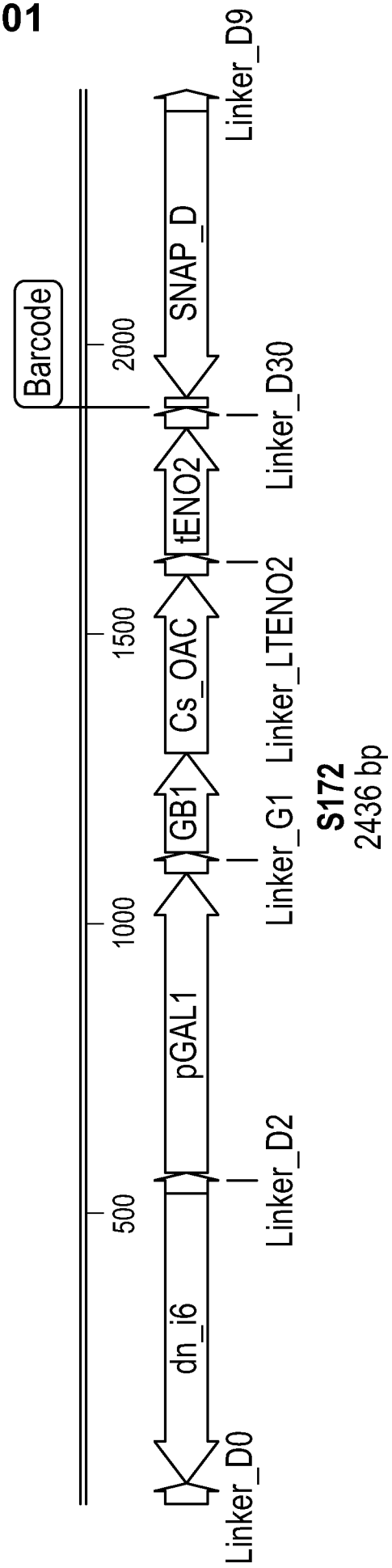

Barcode
2000
SNAP_D
Linker_D9
1500
tENO2
Linker_D30
Cs_OAC
Linker_LTENO2
1000
GB1
Linker_G1
pGAL1
Linker_D2
500
dn_i6
Linker_D0
S172
2436 bp

FIG. 59

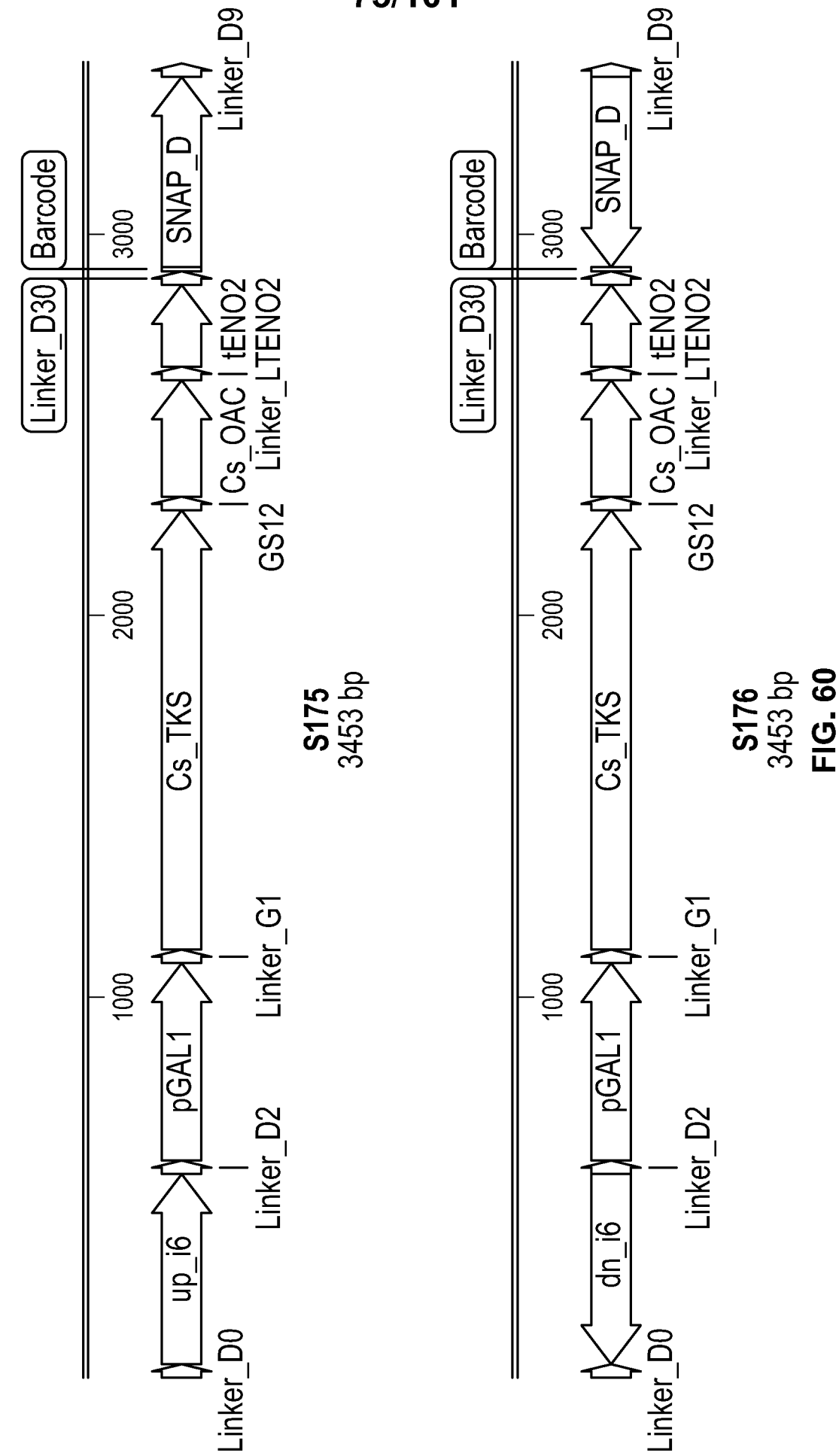

Linker_D0
up_i6
Linker_D2
pGAL1
Linker_G1
Cs_TKS
GS12
Cs_OAC
tENO2
Linker_LTENO2
Linker_D30
Barcode
Linker_D9
SNAP_D
1000
2000
3000
S175
3453 bp
Linker_D0
dn_i6
Linker_D2
pGAL1
Linker_G1
Cs_TKS
GS12
Cs_OAC
tENO2
Linker_LTENO2
Linker_D30
Barcode
Linker_D9
SNAP_D
1000
2000
3000
S176
3453 bp

FIG. 60

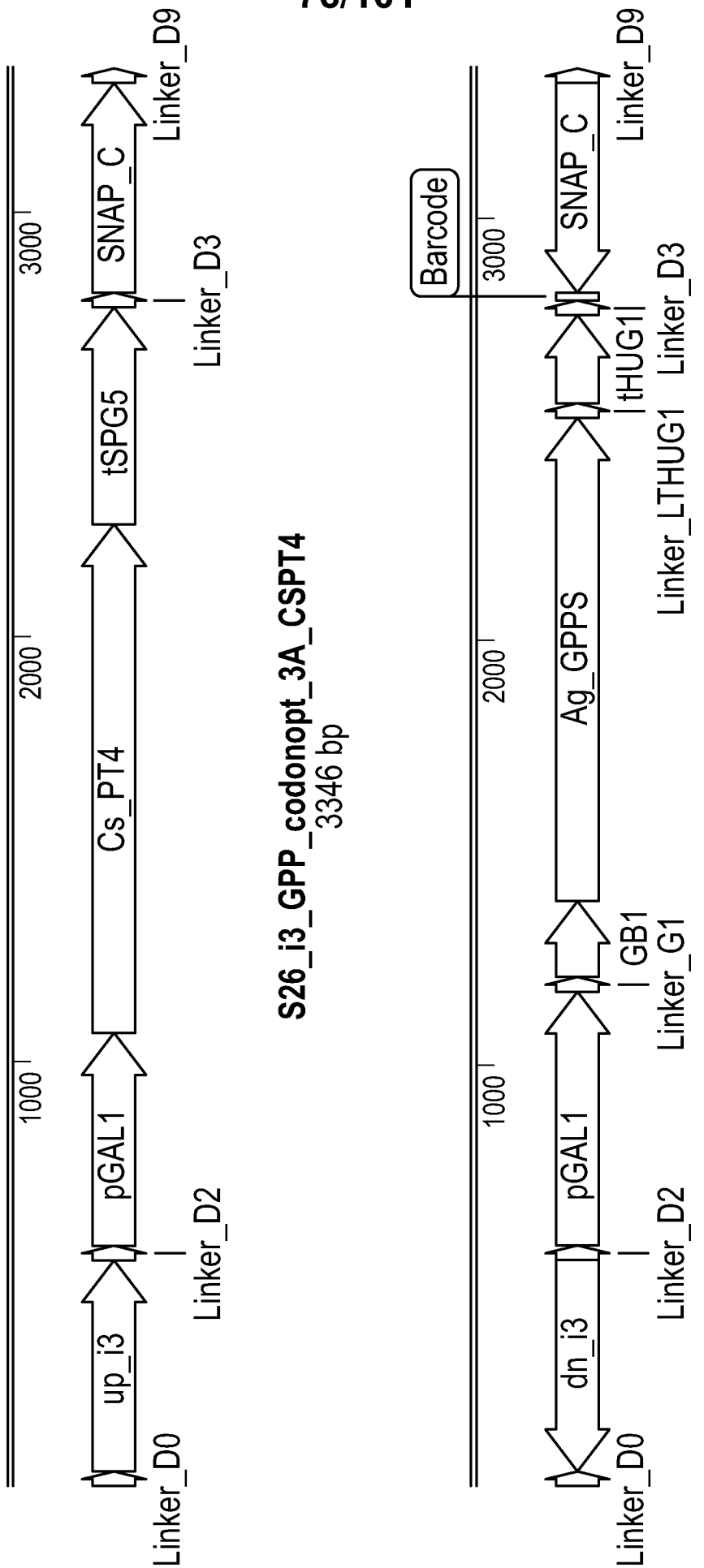
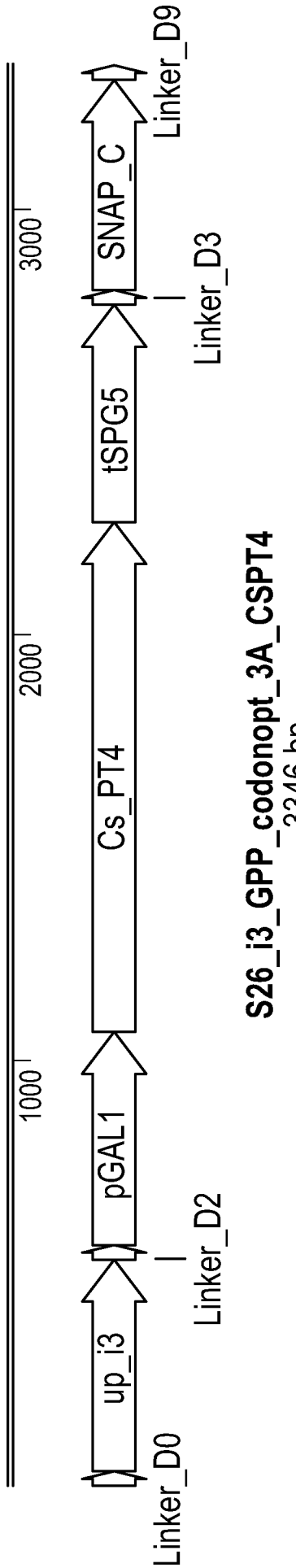
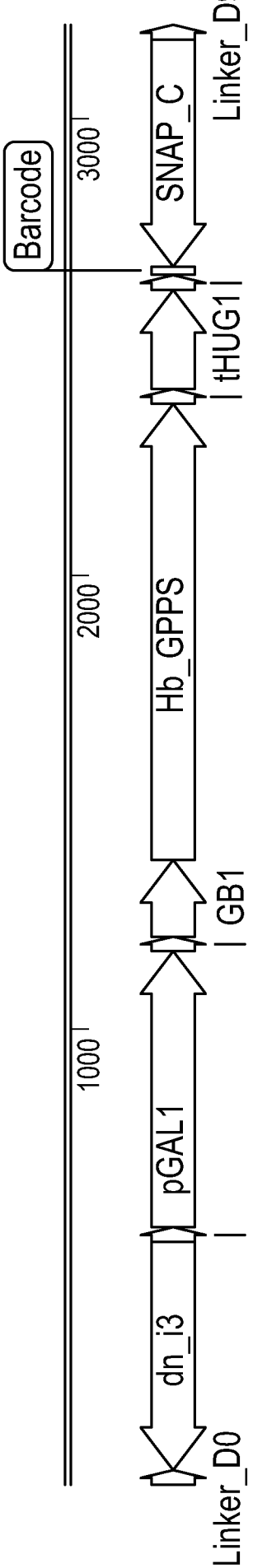

1000
2000
3000
Linker_D0
up_i3
Linker_D2
pGAL1
Cs_PT4
Linker_D3
tSPG5
SNAP_C
Linker_D9
S26_i3_GPP_codonopt_3A_CSPT4
3346 bp
Barcode
dn_i3
GB1
Linker_G1
Ag_GPPS
tHUG1
Linker_LTHUG1
S151
3363 bp

FIG. 61



S26_i3_GPP_codonopt_3A_CSPT4
3346 bp



S156
3210 bp

FIG. 62

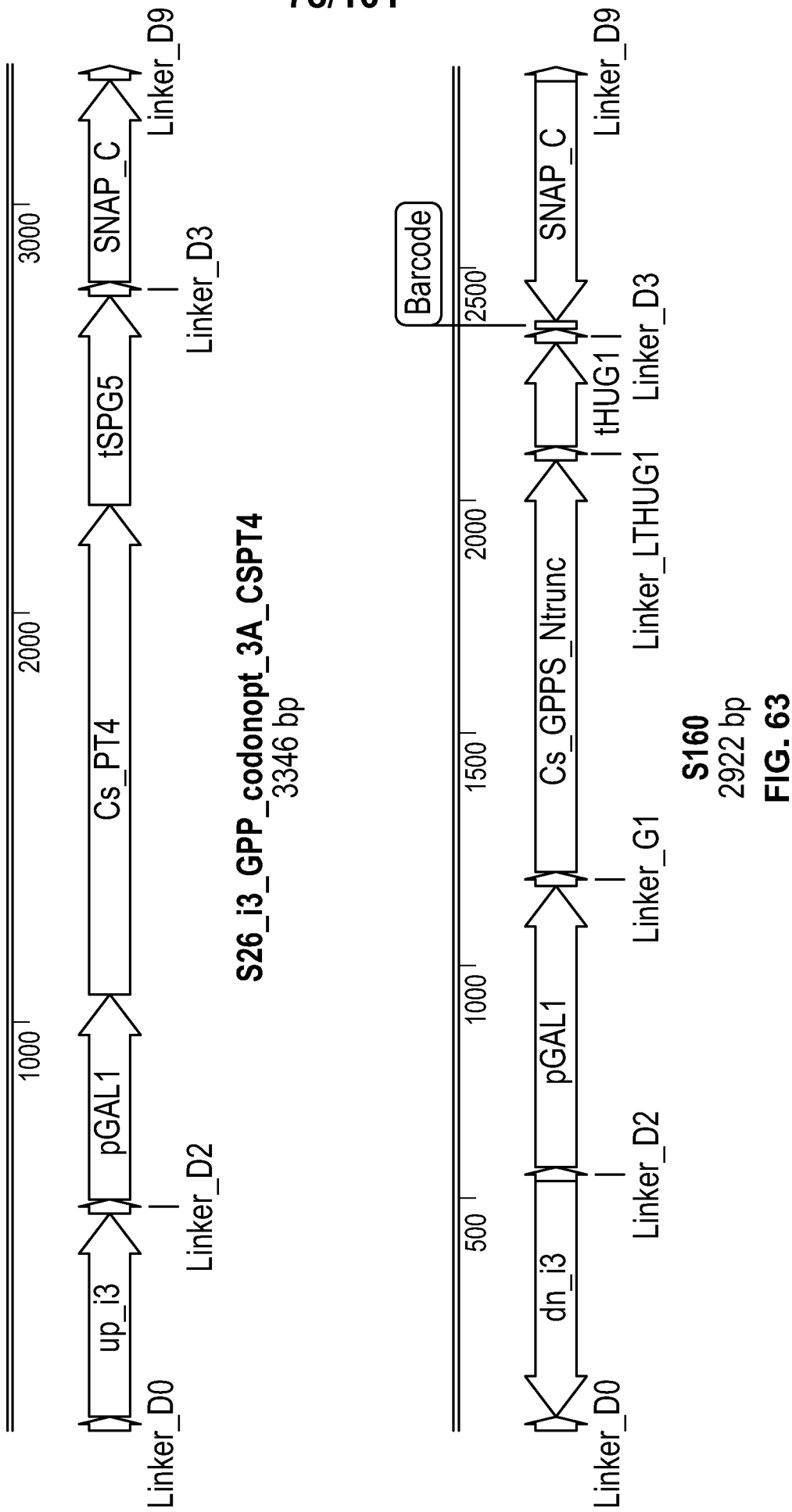

Linker_D0
up_i3
Linker_D2
pGAL1
Cs_PT4
tSPG5
Linker_D3
SNAP_C
Linker_D9
1000
2000
3000
S26_i3_GPP_codonopt_3A_CSPT4
3346 bp
Linker_D0
dn_i3
Linker_D2
pGAL1
Linker_G1
Cs_GPPS_Ntrunc
Linker_LTHUG1
tHUG1
Linker_D3
Barcode
SNAP_C
Linker_D9
500
1000
1500
2000
2500
S160
2922 bp

FIG. 63

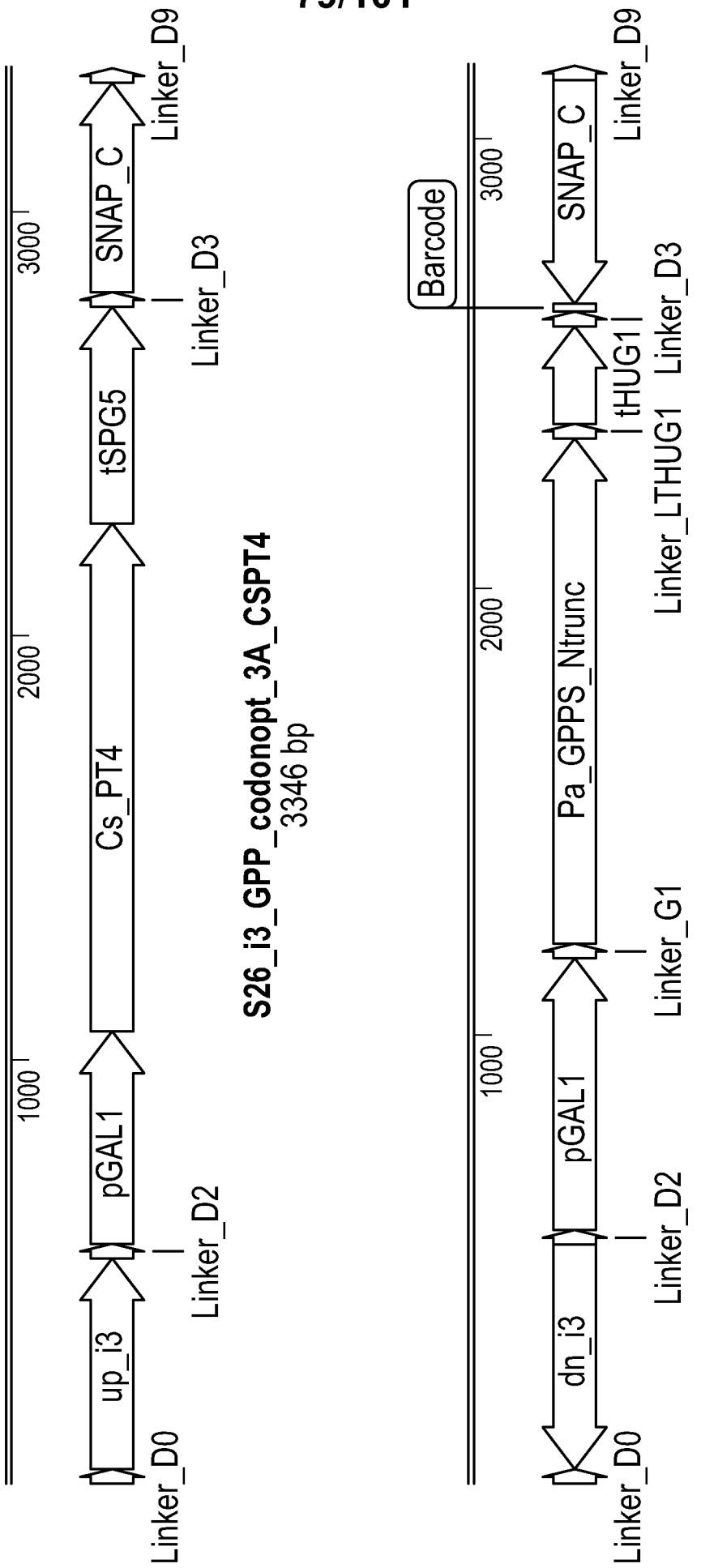

Linker_D0
up_i3
Linker_D2
pGAL1
Cs_PT4
tSPG5
Linker_D3
SNAP_C
Linker_D9
1000
2000
3000
S26_i3_GPP_codonopt_3A_CSPT4
3346 bp
Linker_D0
dn_i3
Linker_D2
pGAL1
Linker_G1
Pa_GPPS_Ntrunc
Linker_LTHUG1
tHUG1
Linker_D3
Barcode
SNAP_C
Linker_D9
1000
2000
3000
S162
3183 bp

FIG. 64

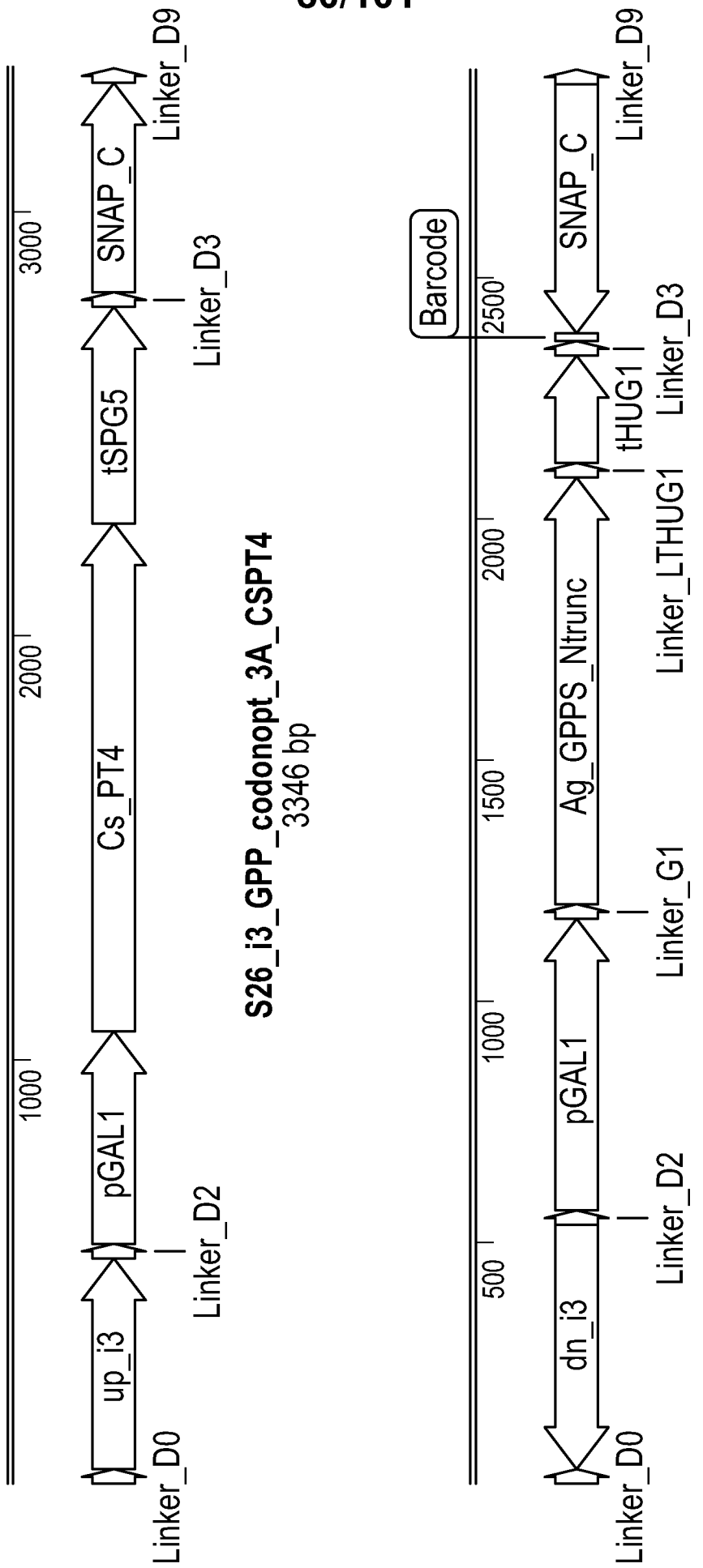

Linker_D0
up_i3
Linker_D2
pGAL1
Cs_PT4
tSPG5
Linker_D3
SNAP_C
Linker_D9
1000
2000
3000
S26_i3_GPP_codonopt_3A_CSPT4
3346 bp
Linker_D0
dn_i3
Linker_D2
pGAL1
Ag_GPPS_Ntrunc
Linker_G1
tHUG1
Linker_LTHUG1
Barcode
Linker_D3
SNAP_C
Linker_D9
500
1000
1500
2000
2500
S163
2937 bp

FIG. 65

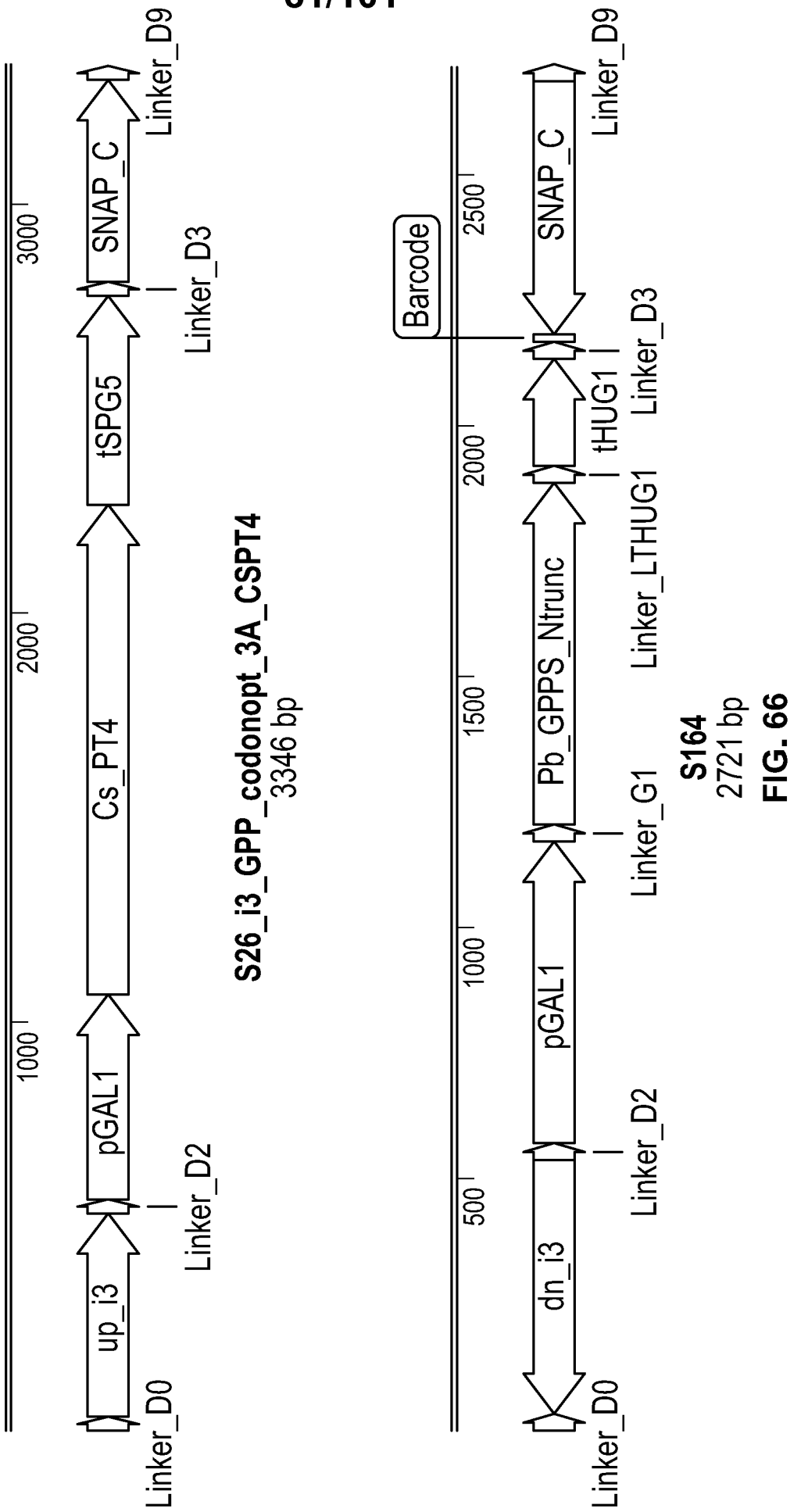

Linker_D0
up_i3
Linker_D2
pGAL1
Cs_PT4
tSPG5
Linker_D3
SNAP_C
Linker_D9
1000
2000
3000
S26_i3_GPP_codonopt_3A_CSPT4
3346 bp
Linker_D0
dn_i3
Linker_D2
pGAL1
Pb_GPPS_Ntrunc
Linker_G1
Barcode
Linker_LTHUG1
tHUG1
Linker_D3
SNAP_C
Linker_D9
500
1000
1500
2000
2500
S164
2721 bp

FIG. 66

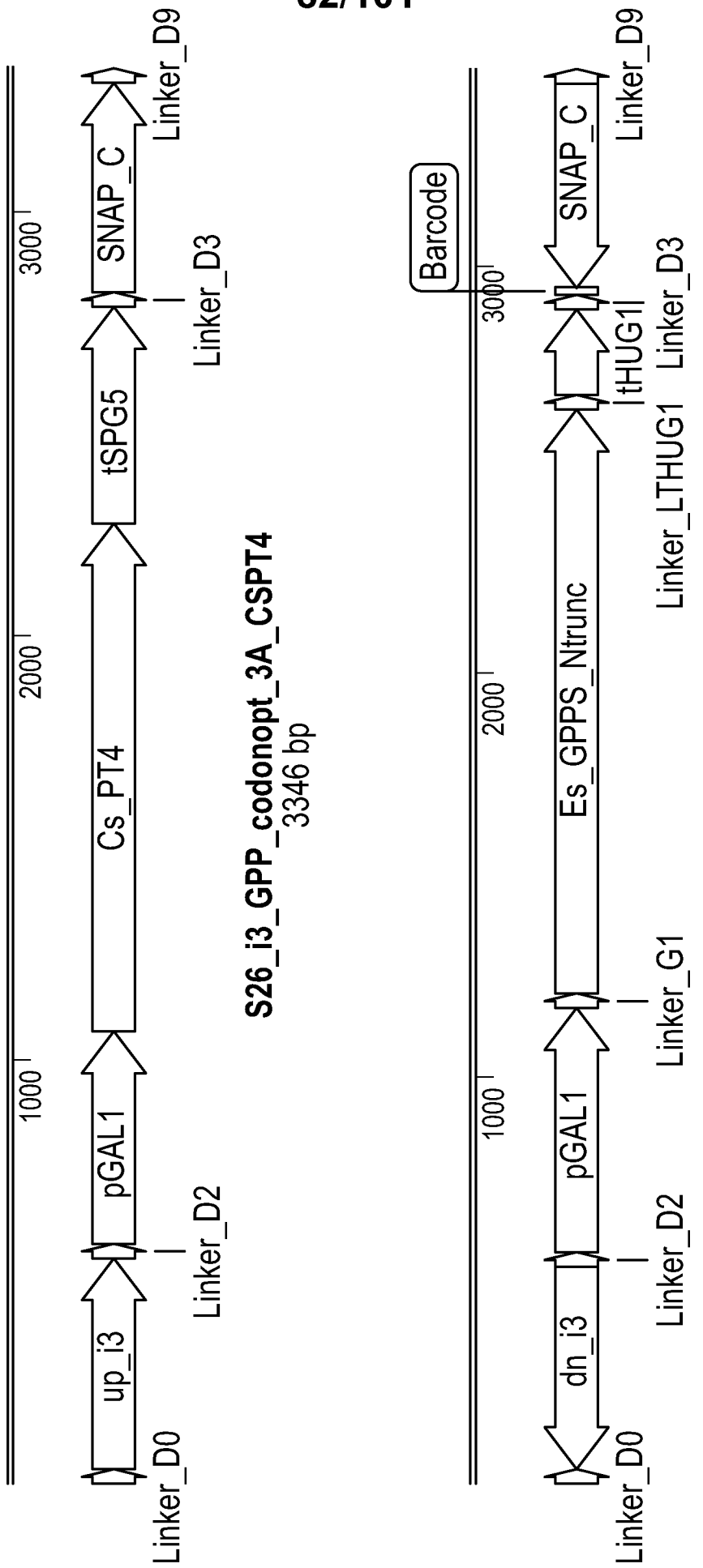

1000
2000
3000
Linker_D0
up_i3
Linker_D2
pGAL1
Cs_PT4
Linker_D3
tSPG5
SNAP_C
Linker_D9
S26_i3_GPP_codonopt_3A_CSPT4
3346 bp
Barcode
dn_i3
Es_GPPS_Ntrunc
tHUG1
Linker_LTHUG1
S166
3483 bp

FIG. 67

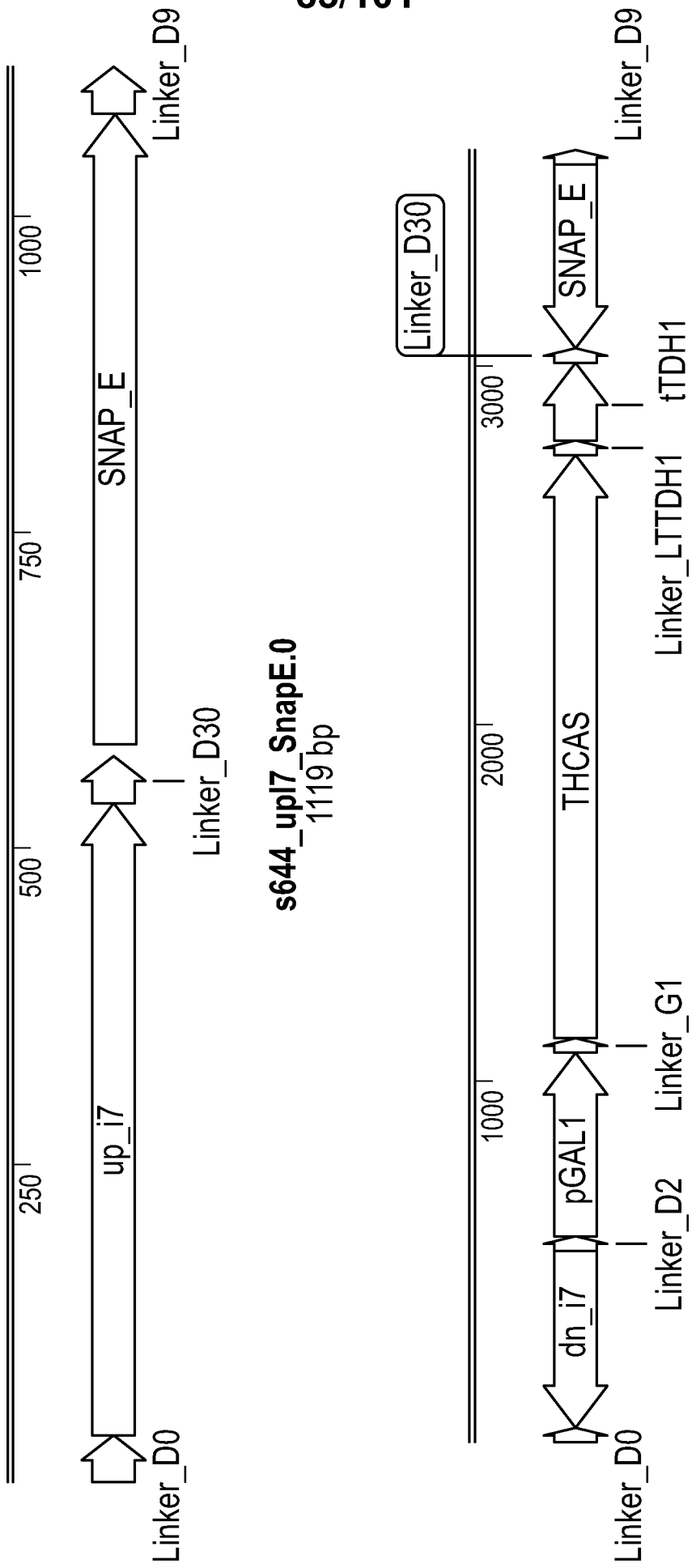
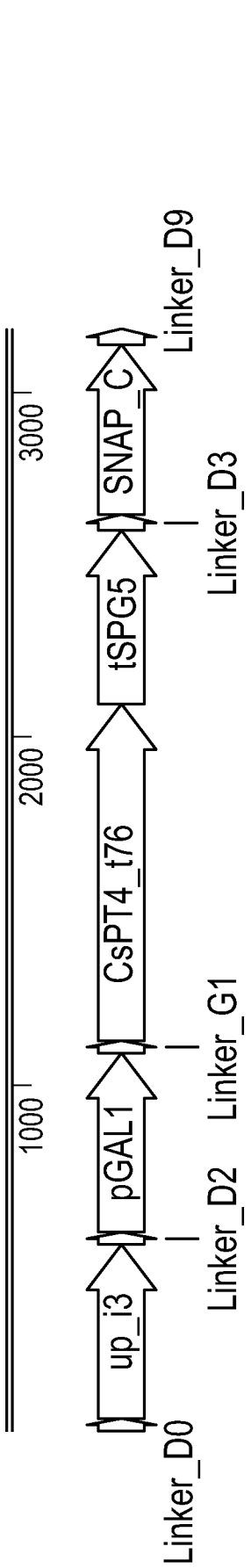
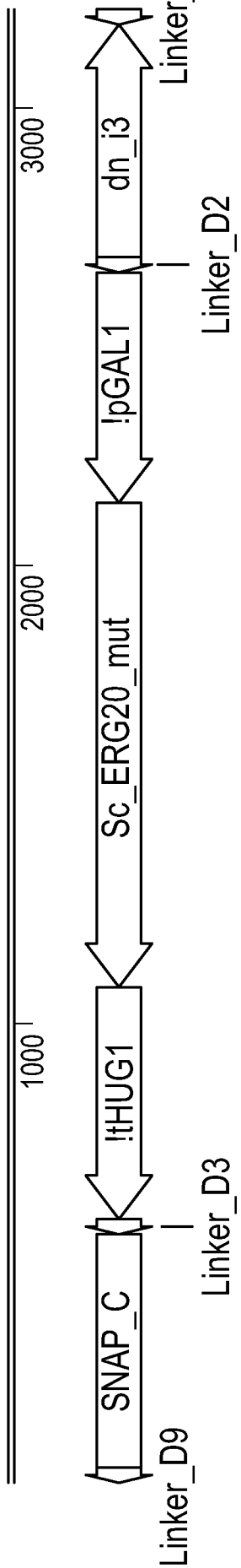

Linker_D0
up_i7
SNAP_E
Linker_D9
Linker_D30
250
500
750
1000
s644_upI7_SnapE.0
1119 bp
Linker_D0
dn_i7
pGAL1
Linker_D2
Linker_G1
THCAS
Linker_LTTDH1
tTDH1
SNAP_E
Linker_D9
Linker_D30
1000
2000
3000
s483_i7b_THCASfull.1
3596 bp

FIG. 68



i3B_CsPT4_t76
3175 bp



S9_i3_GPP_codonopt_3B_0
3212 bp

FIG. 69

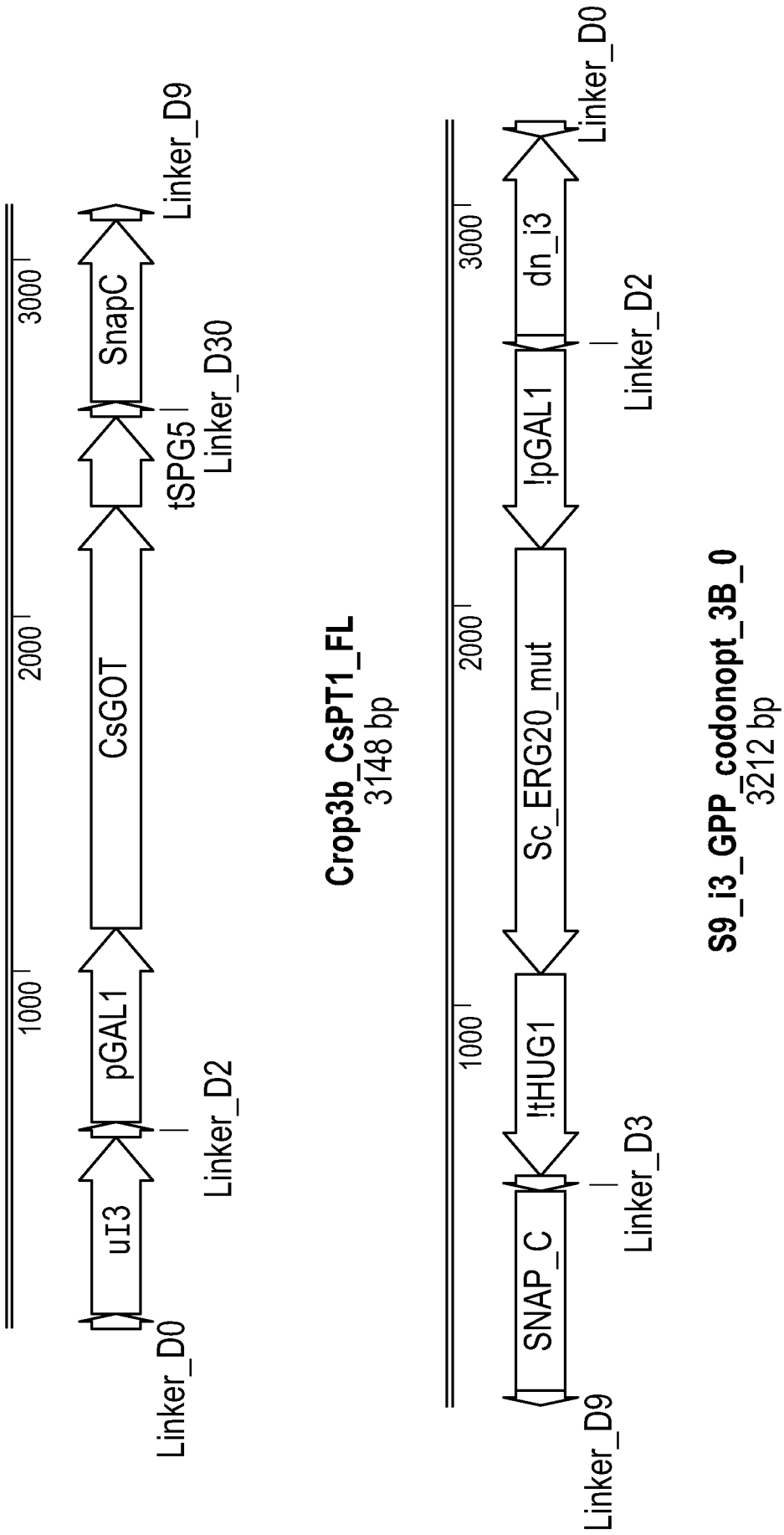

1000
2000
3000
Linker_D0
uI3
Linker_D2
pGAL1
CsGOT
tSPG5
Linker_D30
SnapC
Linker_D9
Crop3b_CsPT1_FL
3148 bp
1000
2000
3000
Linker_D9
SNAP_C
Linker_D3
!tHUG1
Sc_ERG20_mut
!pGAL1
Linker_D2
dn_i3
Linker_D0
S9_i3_GPP_codonopt_3B_0
3212 bp

FIG. 70

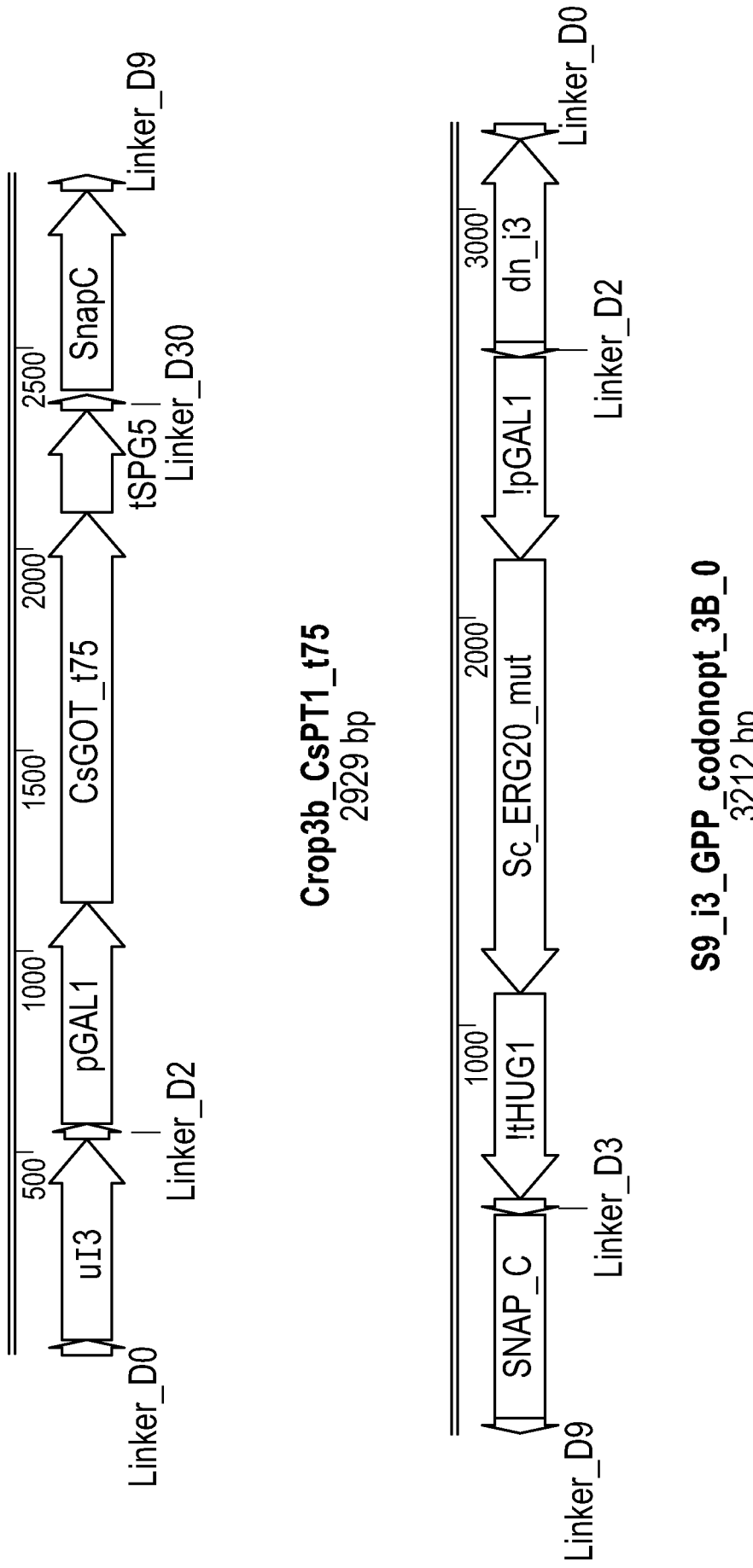

Linker_D0
uI3
500
Linker_D2
pGAL1
1000
CsGOT_t75
1500
tSPG5
2000
Linker_D30
SnapC
2500
Linker_D9
Crop3b_CsPT1_t75
2929 bp
Linker_D9
SNAP_C
Linker_D3
!tHUG1
1000
Sc_ERG20_mut
2000
!pGAL1
Linker_D2
dn_i3
3000
Linker_D0
S9_i3_GPP_codonopt_3B_0
3212 bp

FIG. 71

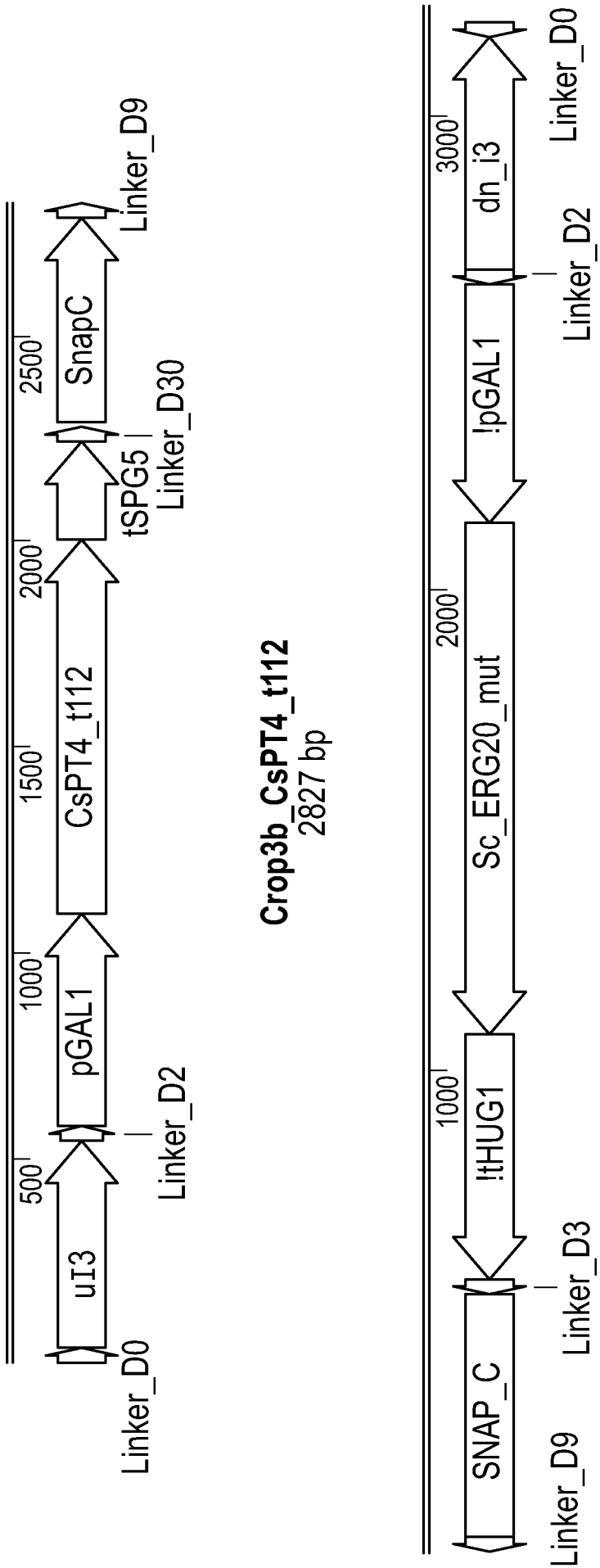

Linker_D0
uI3
500
Linker_D2
pGAL1
1000
CsPT4_t112
1500
tSPG5
Linker_D30
2000
SnapC
2500
Linker_D9
Crop3b_CsPT4_t112
2827 bp
Linker_D9
SNAP_C
Linker_D3
!tHUG1
1000
Sc_ERG20_mut
2000
!pGAL1
Linker_D2
dn_i3
3000
Linker_D0
S9_i3_GPP_codonopt_3B_0
3212 bp

FIG. 72

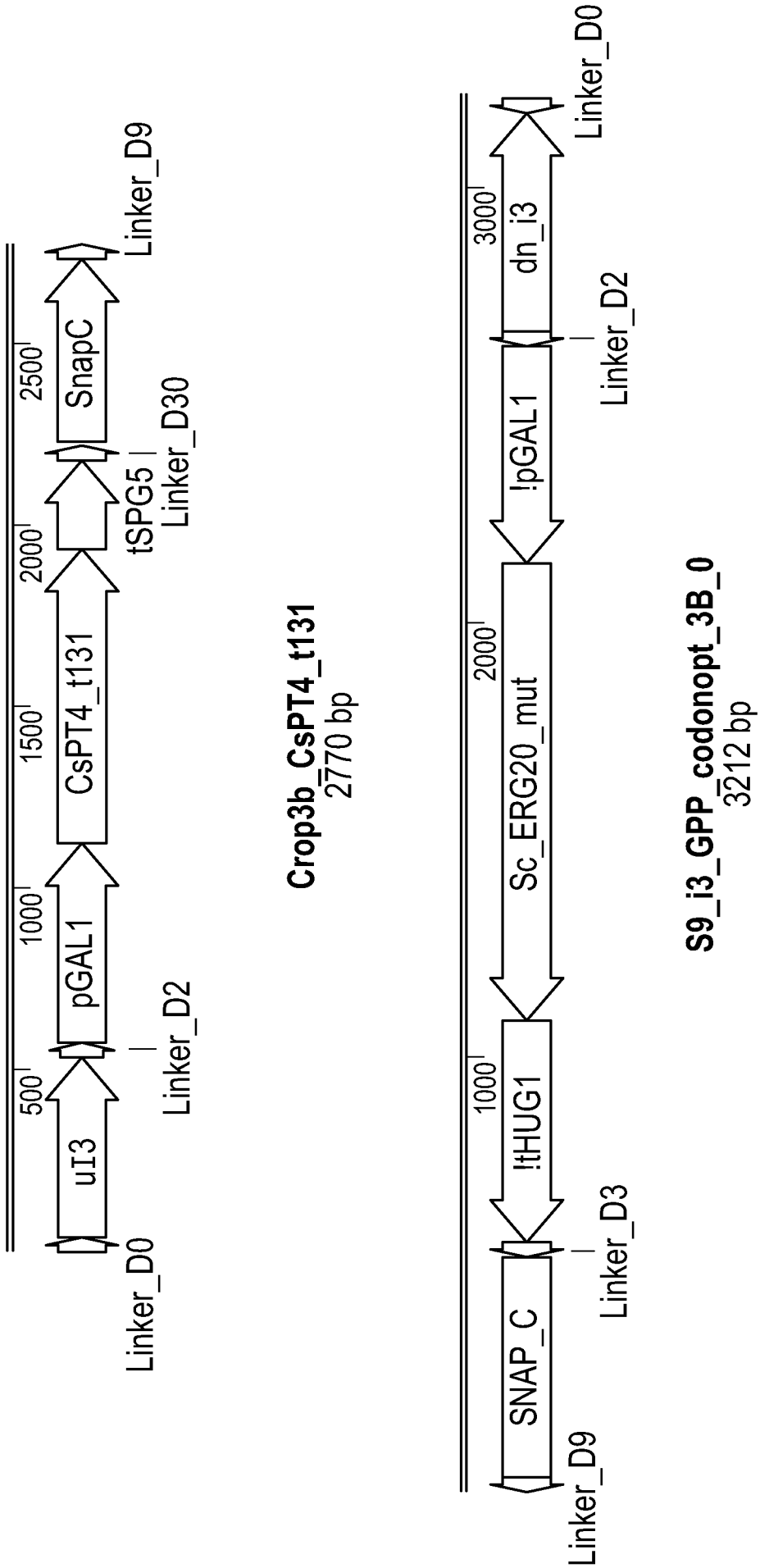

Linker_D0
uI3
500
Linker_D2
pGAL1
1000
CsPT4_t131
1500
tSPG5
2000
Linker_D30
SnapC
2500
Linker_D9
Crop3b_CsPT4_t131
2770 bp
Linker_D9
SNAP_C
Linker_D3
!tHUG1
1000
Sc_ERG20_mut
2000
!pGAL1
Linker_D2
dn_i3
3000
Linker_D0
S9_i3_GPP_codonopt_3B_0
3212 bp

FIG. 73

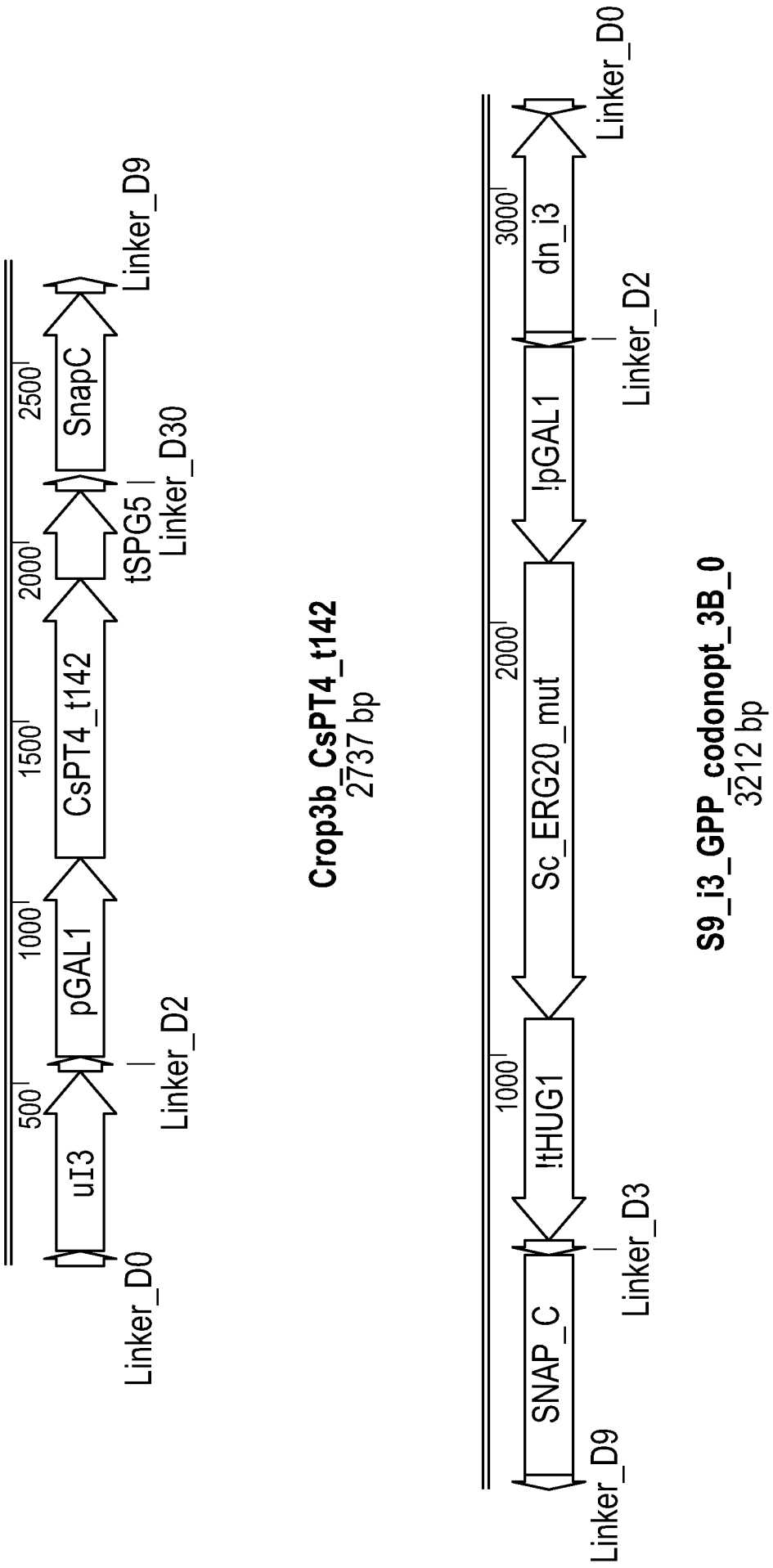

Linker_D0
uI3
500
Linker_D2
pGAL1
1000
CsPT4_t142
1500
tSPG5
2000
Linker_D30
SnapC
2500
Linker_D9
Crop3b_CsPT4_t142
2737 bp
Linker_D9
SNAP_C
Linker_D3
!tHUG1
1000
Sc_ERG20_mut
2000
!pGAL1
Linker_D2
dn_i3
3000
Linker_D0
S9_i3_GPP_codonopt_3B_0
3212 bp

FIG. 74

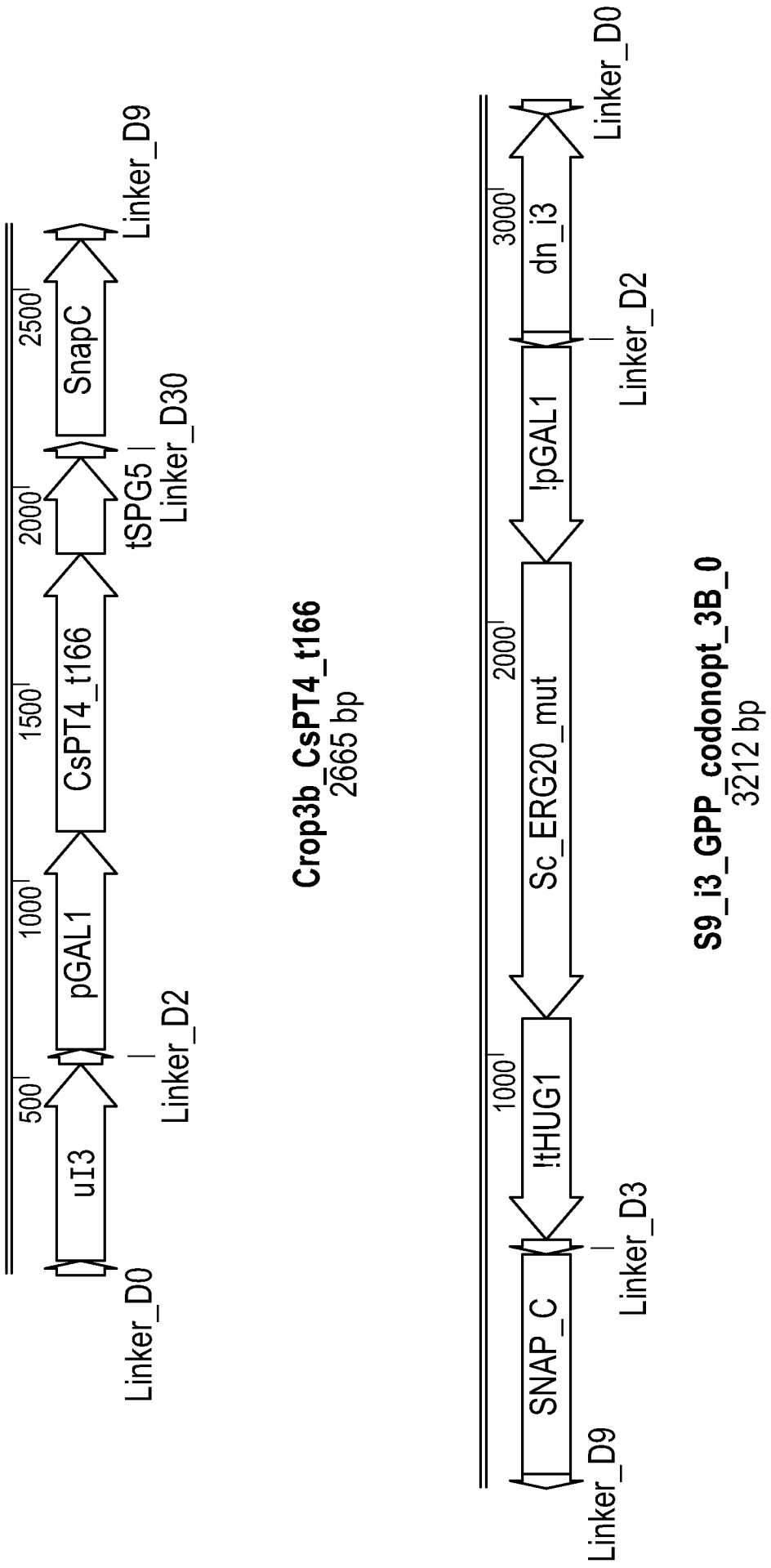

Linker_D0
uI3
500
Linker_D2
pGAL1
1000
CsPT4_t166
1500
tSPG5
Linker_D30
2000
SnapC
2500
Linker_D9
Crop3b_CsPT4_t166
2665 bp
Linker_D0
dn_i3
3000
Linker_D2
!pGAL1
Sc_ERG20_mut
2000
!tHUG1
1000
Linker_D3
SNAP_C
Linker_D9
S9_i3_GPP_codonopt_3B_0
3212 bp

FIG. 75

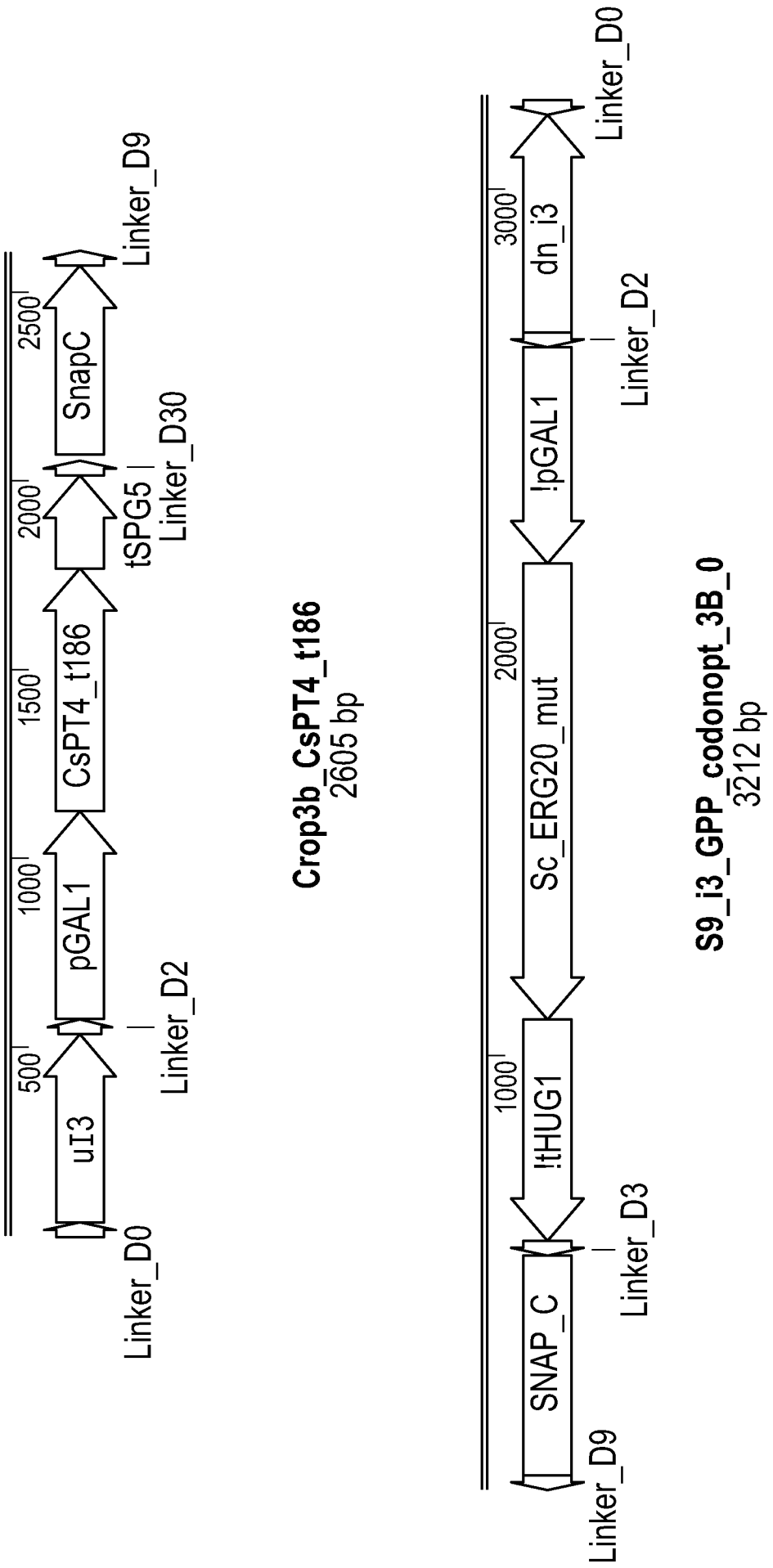

Linker_D0
uI3
500
Linker_D2
pGAL1
1000
CsPT4_t186
1500
tSPG5
Linker_D30
2000
SnapC
2500
Linker_D9
Crop3b_CsPT4_t186
2605 bp
Linker_D0
dn_i3
3000
Linker_D2
!pGAL1
Sc_ERG20_mut
2000
!tHUG1
1000
Linker_D3
SNAP_C
Linker_D9
S9_i3_GPP_codonopt_3B_0
3212 bp

FIG. 76

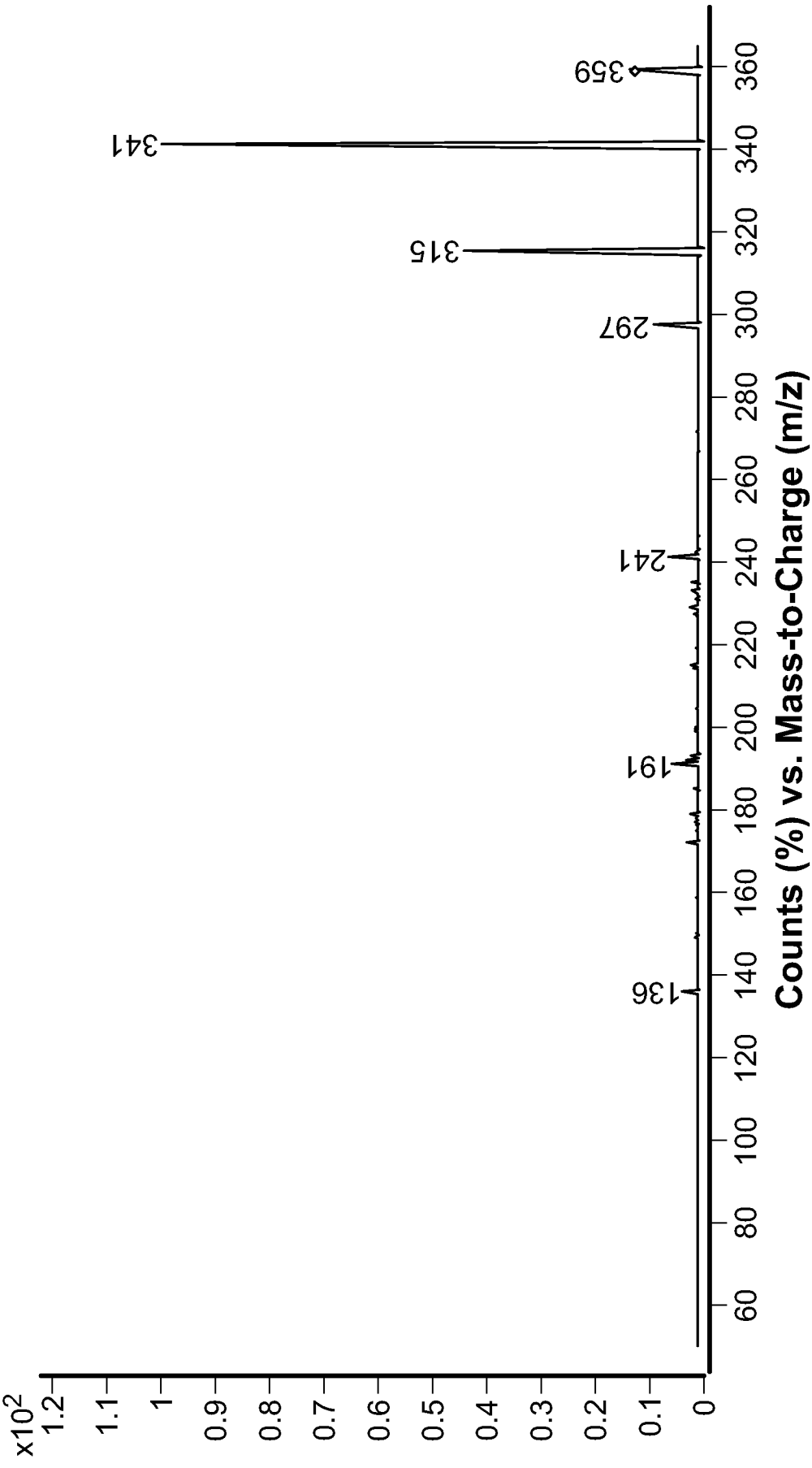

x10^2
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
60
80
100
120
140
160
180
200
220
240
260
280
300
320
340
360
136
191
241
297
315
341
359
Counts (%) vs. Mass-to-Charge (m/z)

FIG. 77

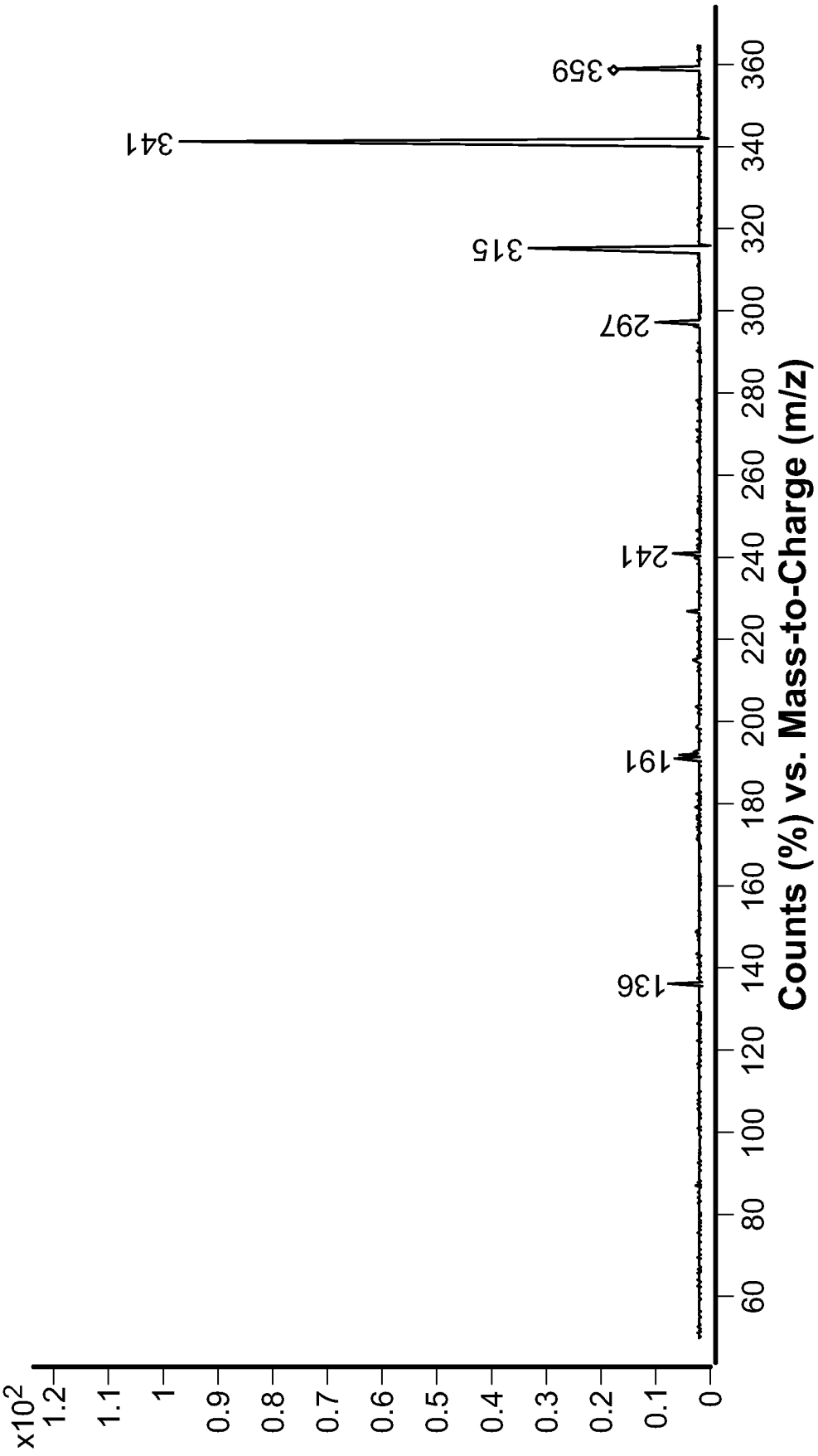

x10^2
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
60
80
100
120
140
160
180
200
220
240
260
280
300
320
340
360
136
191
241
297
315
341
359
Counts (%) vs. Mass-to-Charge (m/z)

FIG. 78

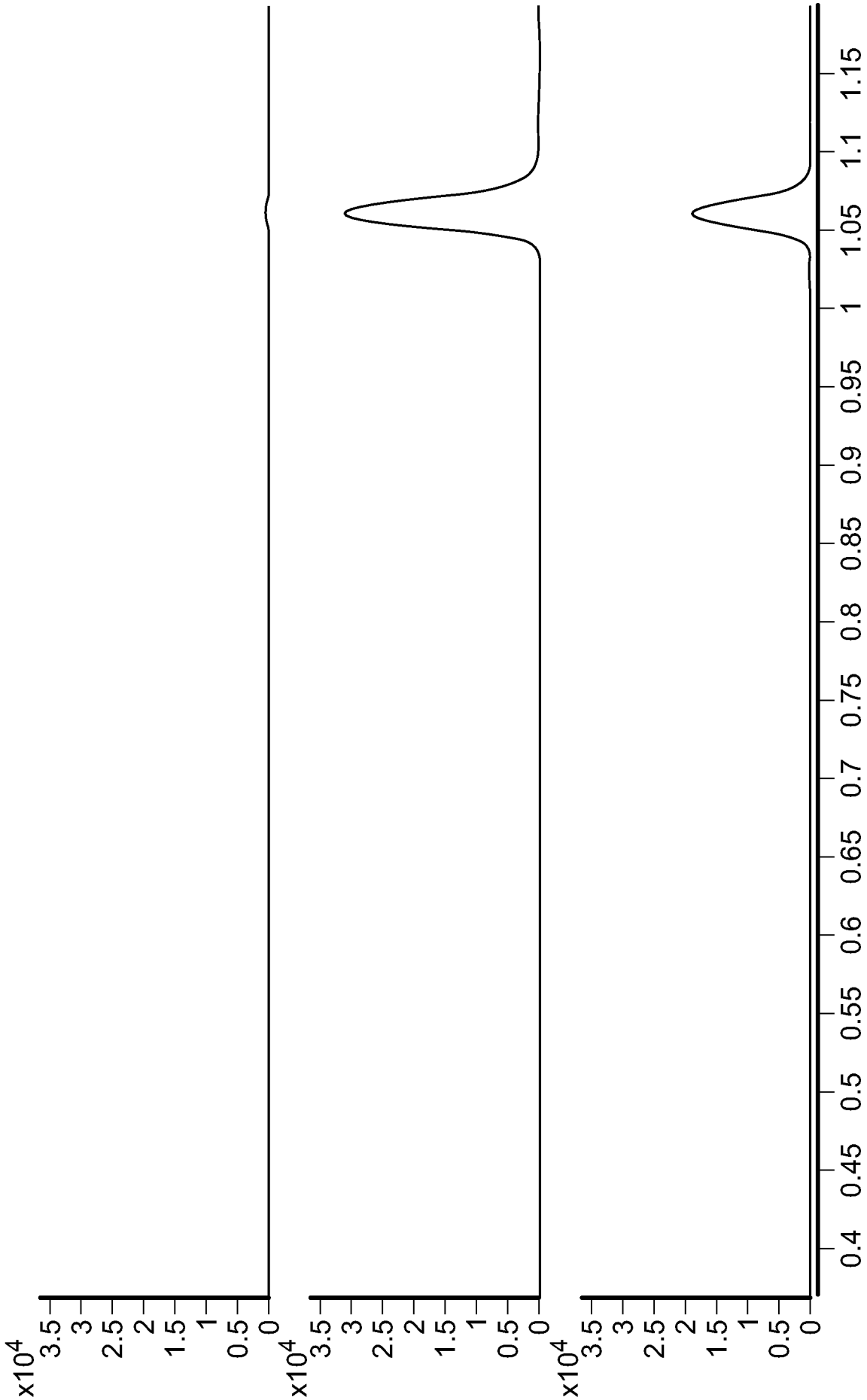

x10^4
3.5
3
2.5
2
1.5
1
0.5
0
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95
1
1.05
1.1
1.15
Counts vs. Acquisition Time (min)

FIG. 79

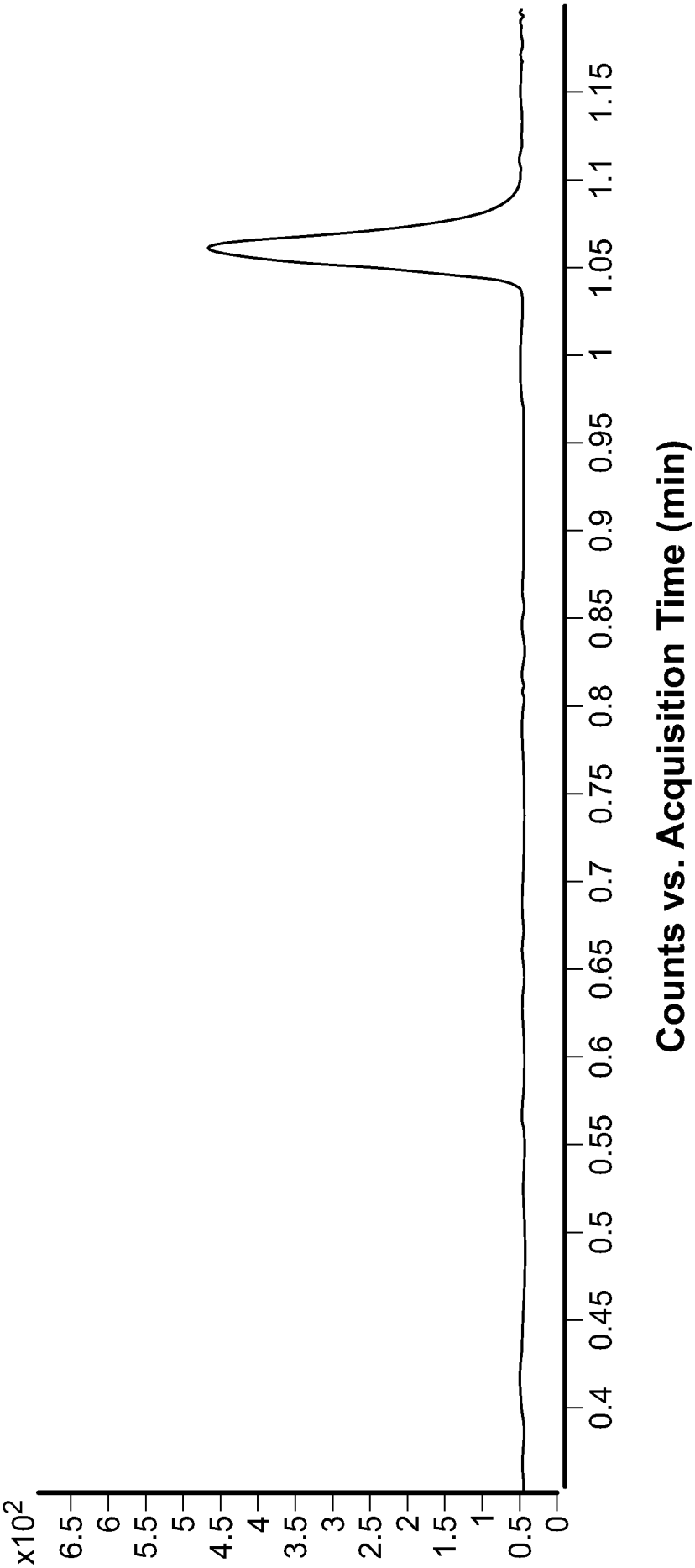

x10^2
6.5
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95
1
1.05
1.1
1.15
Counts vs. Acquisition Time (min)

FIG. 80

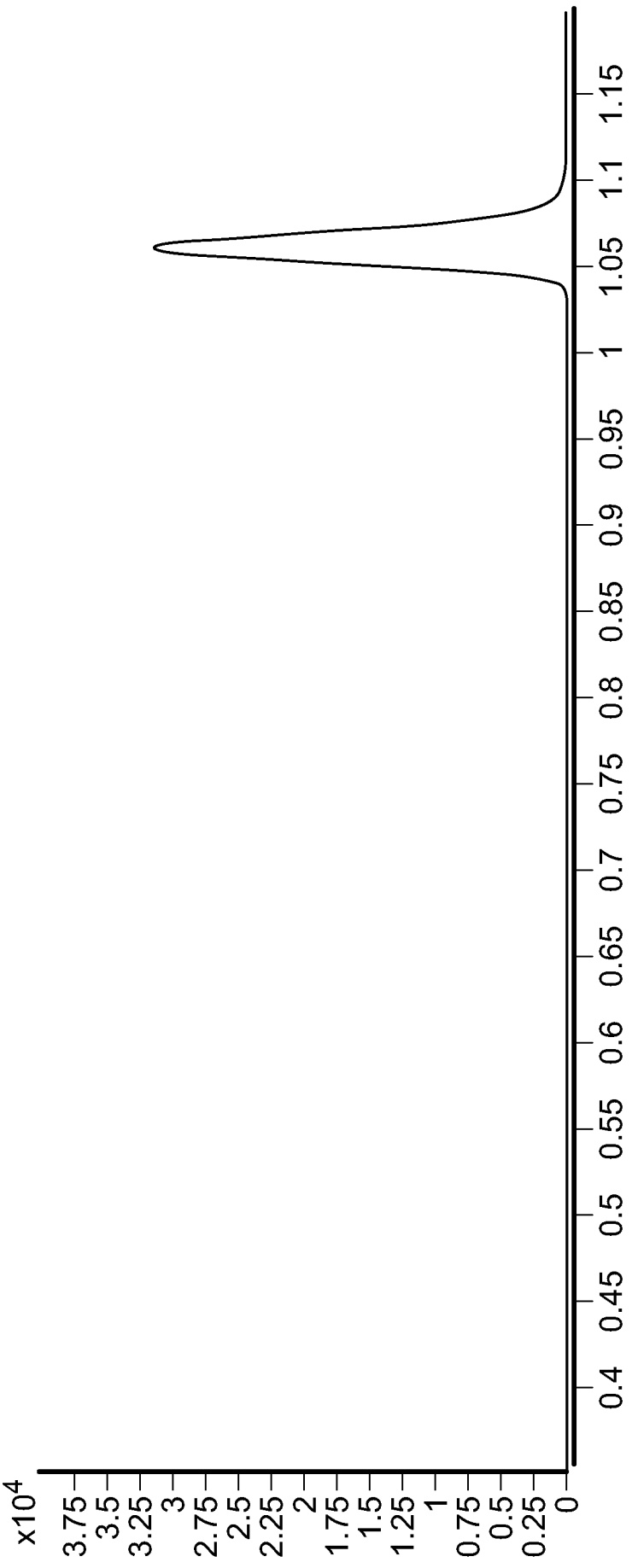

x10^4
3.75
3.5
3.25
3
2.75
2.5
2.25
2
1.75
1.5
1.25
1
0.75
0.5
0.25
0
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95
1
1.05
1.1
1.15
Counts vs. Acquisition Time (min)

FIG. 81

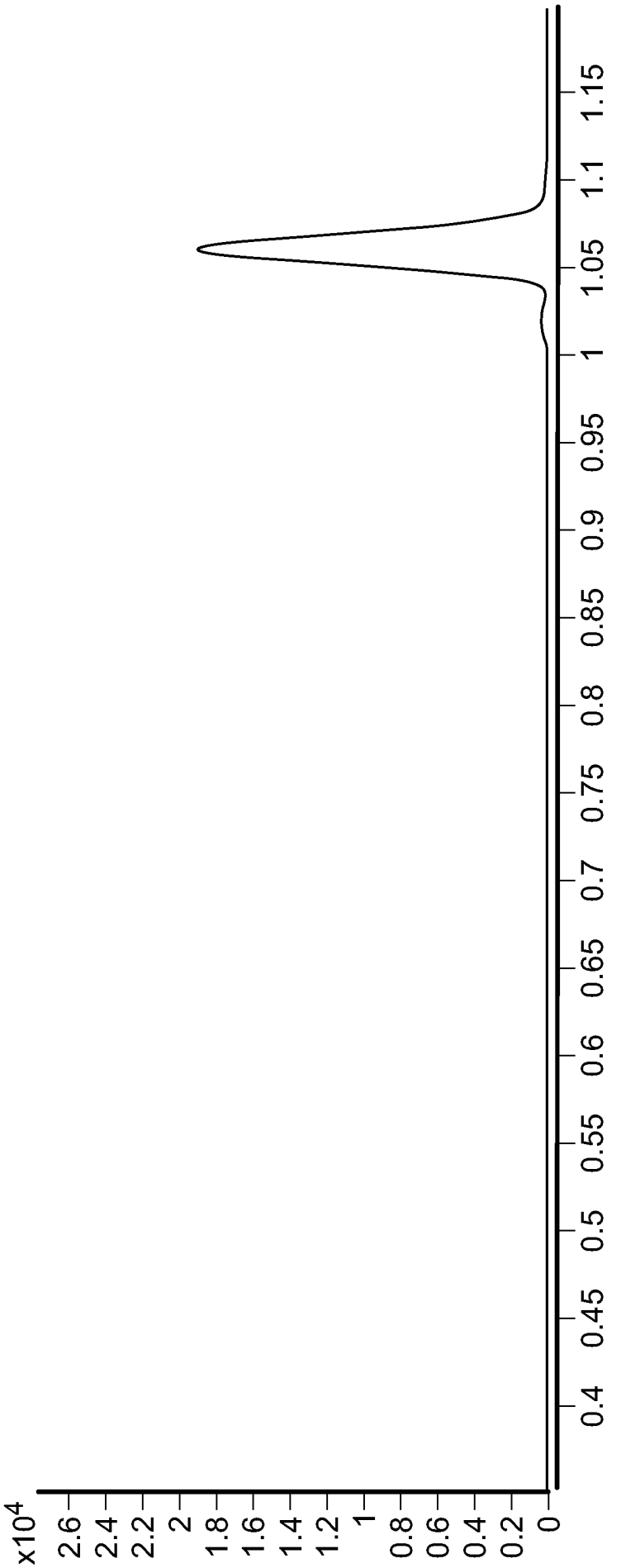

x10^4
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95
1
1.05
1.1
1.15
Counts vs. Acquisition Time (min)

FIG. 82

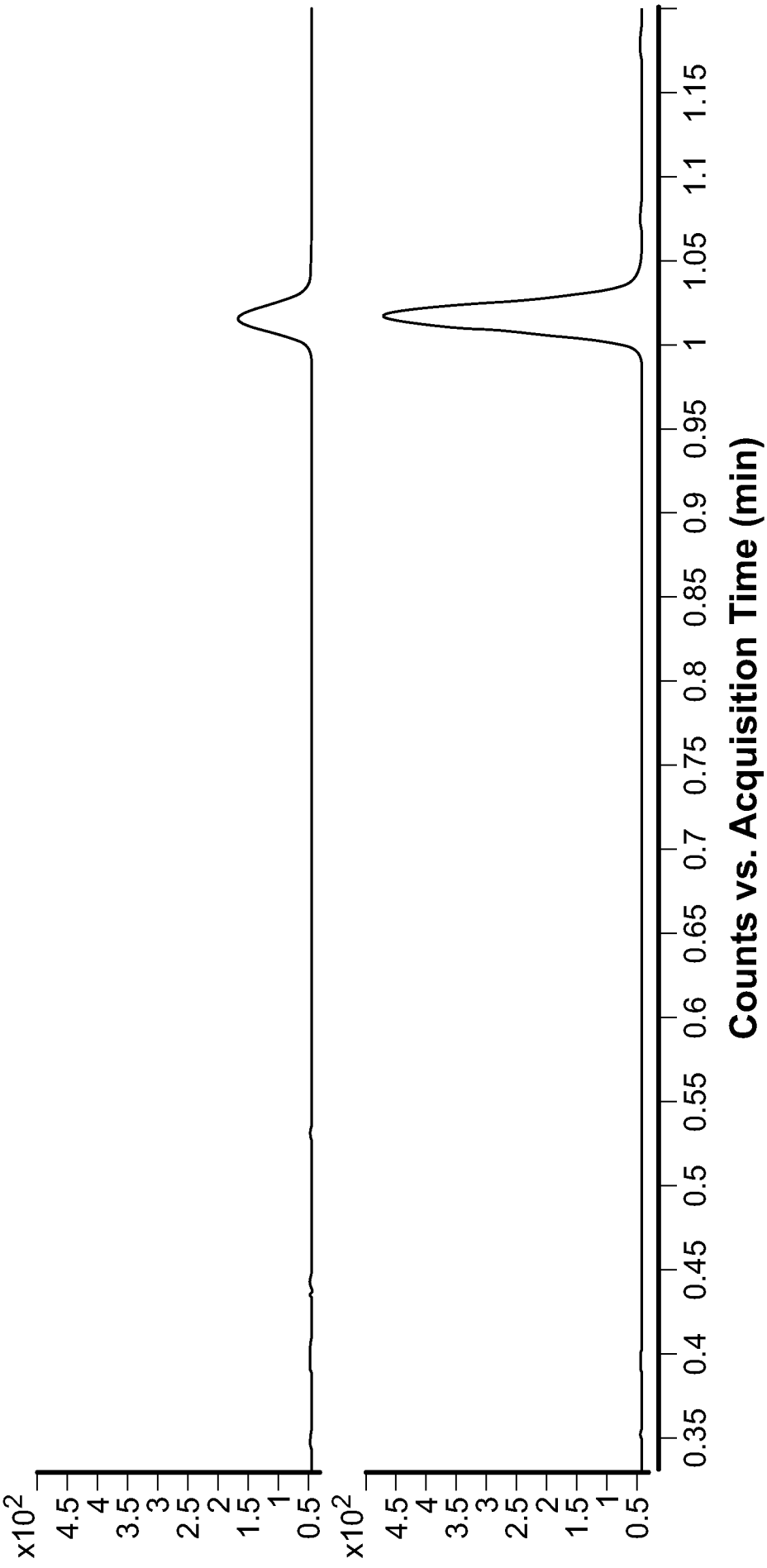

x10^2
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0.35
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95
1
1.05
1.1
1.15
Counts vs. Acquisition Time (min)

FIG. 83

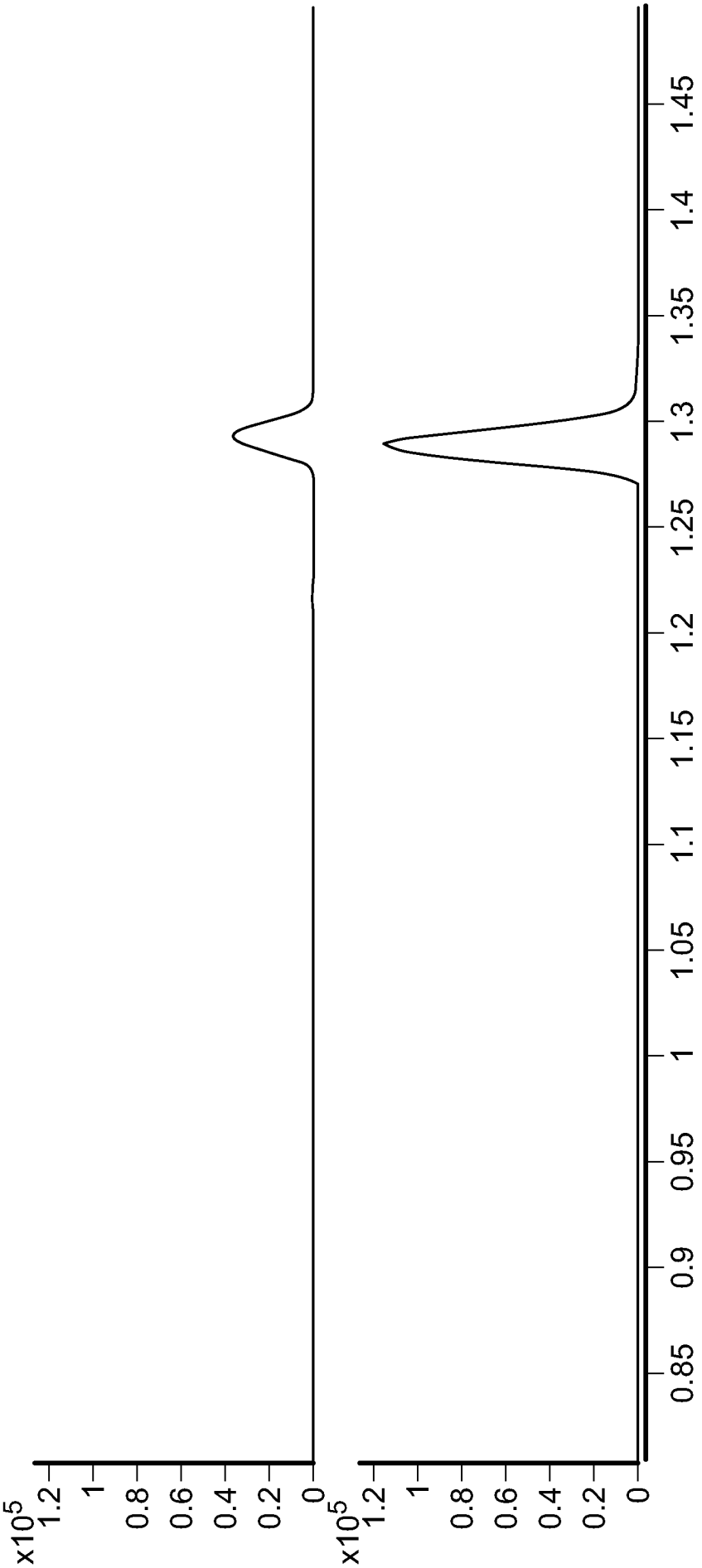

x10^5
1.2
1
0.8
0.6
0.4
0.2
0
0.85
0.9
0.95
1
1.05
1.1
1.15
1.2
1.25
1.3
1.35
1.4
1.45
Counts vs. Acquisition Time (min)

FIG. 84

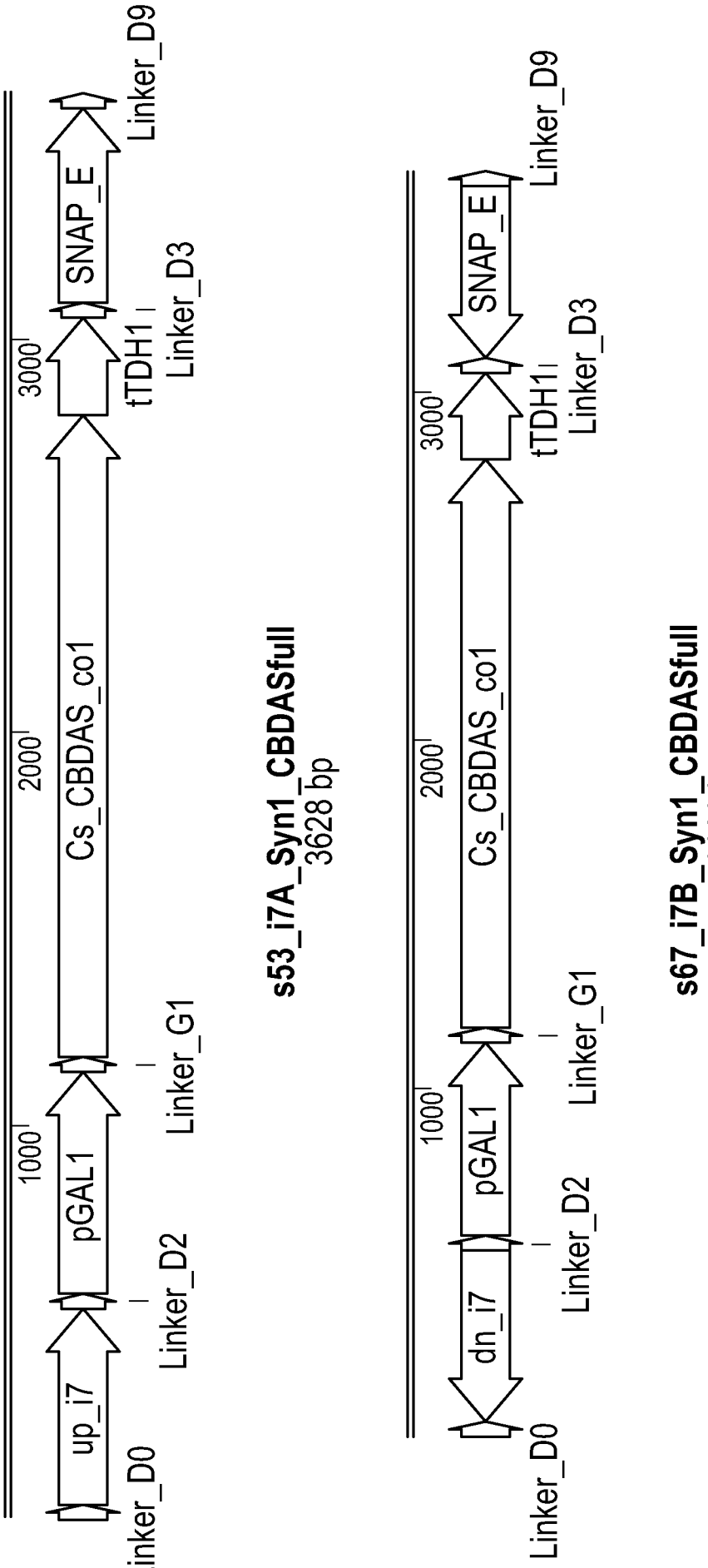

Linker_D0
up_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_co1
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s53_i7A_Syn1_CBDASfull
3628 bp
Linker_D0
dn_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_co1
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s67_i7B_Syn1_CBDASfull
3628 bp

FIG. 85

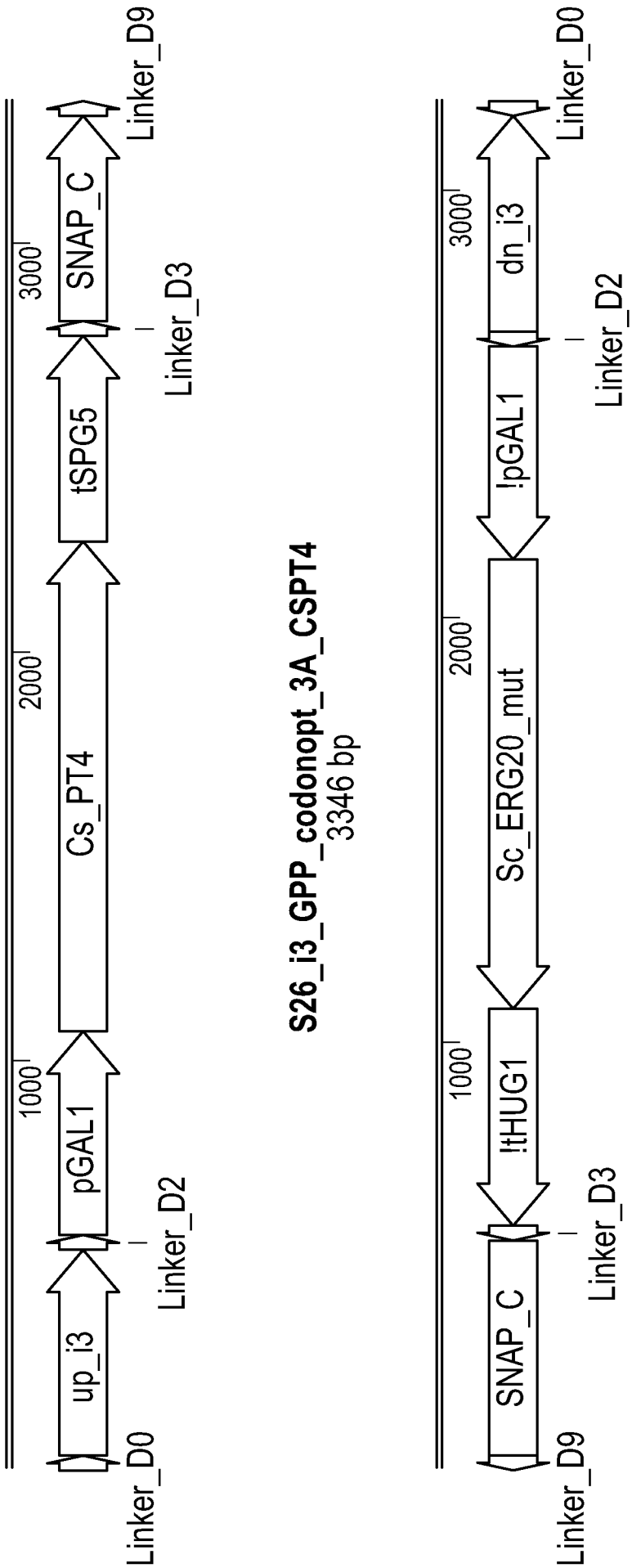

Linker_D0
up_i3
Linker_D2
pGAL1
Cs_PT4
tSPG5
Linker_D3
SNAP_C
Linker_D9
1000
2000
3000
S26_i3_GPP_codonopt_3A_CSPT4
3346 bp
Linker_D9
SNAP_C
Linker_D3
!tHUG1
Sc_ERG20_mut
!pGAL1
Linker_D2
dn_i3
Linker_D0
1000
2000
3000
S9_i3_GPP_codonopt_3B_0
3212 bp

FIG. 86

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2018/029668

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N9/10 C12N15/52 C12P7/22 C12P17/06
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, EMBASE, FSTA, IBM-TDB, Sequence Search

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL [Online] 16 October 2011 (2011-10-16), "TSA: Cannabis sativa PK15523.1_1.CasaPuKu mRNA sequence.", XP002782462, retrieved from EBI accession no. EM_TSA:JP460119 Database accession no. JP460119 the whole document -----	72-77
A	EP 3 067 058 A1 (FARMAGENS HEALTH CARE SRL [IT]) 14 September 2016 (2016-09-14) the whole document in particular paragraphs [0029]-[0043]; Fig. 1, SEQ ID NO:3 ----- -/--	1-81

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
26 June 2018	01/08/2018

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Dumont, Elisabeth

1

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2012/144523 A1 (PAGE JONATHAN E [CA] ET AL) 7 June 2012 (2012-06-07) the whole document in particular [0063], [0065]; Figs 1 and 5; SEQ ID NO:2	1-81
A	US 2016/010126 A1 (POULOS JASON L [US] ET AL) 14 January 2016 (2016-01-14) the whole document in particular paragraphs [0034]-[0049]; Figs 7, 8; SEQ ID NO:27	1-81
A	Sayed Hussein Farag Hussein: "Cannabinoids production in Cannabis sativa L.: An in vitro approach", von der Fakultät Bio- und Chemieingenieurwesen der Technischen Universität Dortmund genehmigte Dissertation , 1 January 2014 (2014-01-01), XP055487626, Retrieved from the Internet: URL:https://eldorado.tu-dortmund.de/bitstream/2003/34350/1/Dissertation.pdf [retrieved on 2018-06-25] page 86, last paragraph-page 88 paragraph 2;pages 94-95, concluding remarks.	1-81
A	ISVETT JOSEFINA FLORES-SANCHEZ ET AL: "Secondary metabolism in cannabis", PHYTOCHEMISTRY REVIEWS, KLUWER ACADEMIC PUBLISHERS, DO, vol. 7, no. 3, 8 April 2008 (2008-04-08), pages 615-639, XP019613387, ISSN: 1572-980X the whole document in particular page 616, col. 2, last paragraph-page 617,col. 2, last paragraph; Fig. 1; Fig. 3.	1-81

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2018/029668

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 3067058 A1	14-09-2016	NONE	
US 2012144523 A1	07-06-2012	CA 2770774 A1	17-02-2011
		US 2012144523 A1	07-06-2012
		US 2015128301 A1	07-05-2015
		WO 2011017798 A1	17-02-2011
US 2016010126 A1	14-01-2016	CA 2990071 A1	21-01-2016
		US 2016010126 A1	14-01-2016
		US 2018073043 A1	15-03-2018
		WO 2016010827 A1	21-01-2016