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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 September 2020 (24.09.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/190763 A1

(51) International Patent Classification:
C12P 7/42 (2006.01)

(21) International Application Number:
PCT/US2020/022741

(22) International Filing Date:
13 March 2020 (13.03.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/819,457 15 March 2019 (15.03.2019) US

(71) Applicant: **AMYRIS, INC.** [US/US]; 5885 Hollis Street, Suite 100, Emeryville, California 94608 (US).

(72) Inventor; and
(71) Applicant: **KLEIN, Andrew P.** [US/US]; c/o 5885 Hollis Street, Suite 100, Emeryville, California 94608 (US).

(72) Inventors: **REEVES, Christopher D.;** c/o **5885** Hollis Street, Suite **100,** Emeryville, California **94608** (US). **PADDON, Christopher J.;** c/o **5885** Hollis Street, Suite **100,** Emeryville, California **94608** (US). **HOLMES, Victor F.;** c/o **5885** Hollis Street, Suite **100,** Emeryville, California **94608** (US).

(74) Agent: **BORING, Landin F.** et al.; Kilpatrick, Townsend & Stockton LLP, Mailstop: IP Docketing- 22, Suite 2800, 1100 Peachtree Street NE, Atlanta, Georgia 30309-4528 (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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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, Cl, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) **Title:** MICROBIAL PRODUCTION OF COMPOUNDS

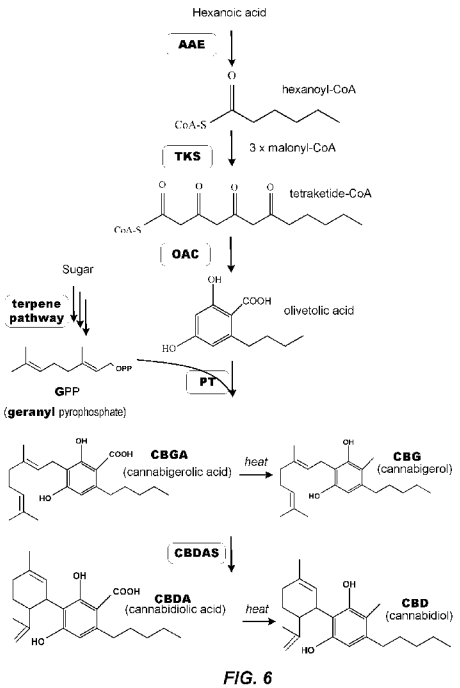

FIG. 6

(57) **Abstract:** Provide are modified host cells that are engineered to decrease expression of a product to undetectable levels in the presence of an exogenous agent, and increase expression of the product in the presence of another exogenous agent. The modified yeast strainshost cells do not express detectable levels of a precursor or substrate used to make the product. The product can be a cannabinoid or precursor thereof, and the substrate can be hexanoate. Also provided are methods for making a product using the modified host cells. The modified host cell can be a yeast strain, such as *S. cerevisiae.*

WO 2020/190763 A1

[Continued on next page]

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(H))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

MICROBIAL PRODUCTION OF COMPOUNDS

CROSS REFERENCES TO RELATED APPLICATIONS

[0001] The present application claims priority to U.S. Provisional Pat. Appl. No. 62/819,457, filed on March 15, 2019, which application is incorporated herein by reference in its entirety.

BACKGROUND OF THE INVENTION

[0002] When organisms are used to produce biomolecules, it is usually beneficial to have an externally-controlled genetic switch to toggle between a high-growth, low-production mode (appropriate for biomass generation and ease of handling) and a low-growth, high-production mode (appropriate for profitable manufacture). For most fermentatively-produced biomolecules, a single switch mechanism is sufficient, even though it allows as much as 10-20% production in the low-production mode. For some biomolecules, such as cannabinoids, regulatory requirements create a need for extremely low or non-detectable production in the low-production state.

[0003] There are many examples of strong single-mechanism switches found in nature, including the galactose regulation system of yeast and the arabinose regulation system in bacteria. Most are based on normal physiological responses where genes are activated when the organism senses a threat or resource. There are also several systems that respond to molecules not usually found in an organisms' environment - tetracycline, IPTG, indigo - that have been used in biotechnological applications.

BRIEF SUMMARY OF THE INVENTION

[0004] In one aspect, a modified, engineered or recombinant host cell is provided, the host cell comprising a heterologous genetic pathway that produces a heterologous product and that is regulated by an exogenous agent, wherein the host cell does not produce a precursor required to make the product. In some embodiments, the exogenous agent comprises a regulator of gene expression.

[0005] In some embodiments, the exogenous agent decreases production of the heterologous product. In some embodiments, the exogenous agent that decreases production

of the heterologous product is glucose and expression of one or more enzymes encoded by the heterologous genetic pathway are under control of a glucose repressed promoter.

[0006] In some embodiments, the exogenous agent increases production of the heterologous product. In some embodiments, the exogenous agent that increases production of the heterologous product is galactose and expression of one or more enzymes encoded by the heterologous genetic pathway are under control of a GAL promoter.

[0007] In some embodiments, the heterologous genetic pathway comprises a galactose-responsive promoter, a maltose-responsive promoter, or a combination of both.

[0008] In some embodiments, the heterologous product is a cannabinoid or cannabinoid precursor. In some embodiments, the cannabinoid or cannabinoid precursor is cannabidiolic acid (CBDA), cannabidiol (CBD), cannabigerolic acid (CBGA), or cannabigerol (CBG).

[0009] In some embodiments, the genetic pathway encodes at least two enzymes selected from the group consisting of hexanoyl-CoA synthase (HCS), tetraketide synthase (TKS) and olivetolic acid cyclase (OAC).

[0010] In some embodiments, the precursor required to make the product is hexanoate.

[0011] In some embodiments, the heterologous genetic pathway comprises a nucleic acid construct comprising at least 3 protein coding regions.

[0012] In some embodiments, the host cell is a yeast cell or yeast strain. In some embodiments, the yeast cell is S. cerevisiae.

[0013] In another aspect, a mixture is provided, the mixture comprising a host cell described herein and a culture media. In some embodiments, the culture media comprises an exogenous agent that decreases production of the heterologous product. In some embodiments, the exogenous agent that decreases production of the heterologous product is glucose, maltose, or lysine.

[0014] In some embodiments, the culture media comprises (i) an exogenous agent that increases production of the heterologous product, and (ii) a precursor required to make the heterologous product. In some embodiments, the exogenous agent that increases production of the heterologous product is galactose. In some embodiments, the precursor required to make the heterologous product is hexanoate.

[0015] In another aspect, a method for decreasing the expression of a heterologous product is provided, the method comprising culturing a host cell described herein in a media comprising the exogenous agent, wherein the exogenous agent decreases the expression of the heterologous product. In some embodiments, the exogenous agent that decreases expression of the heterologous product is glucose, maltose, or lysine. In some embodiments, culturing the host cell strain in the media comprising the exogenous agent results in less than 0.001 mg/L of heterologous product

[0016] In another aspect, a method for increasing the expression of a heterologous product is described, the method comprising culturing a host cell described herein in a media comprising the exogenous agent, wherein the exogenous agent increases expression of the heterologous product. In some embodiments, the exogenous agent that increases expression of the heterologous product is galactose.

[0017] In some embodiments, the method further comprises culturing the host cell with the precursor required to make the heterologous product. In some embodiments, the precursor required to make the heterologous product is hexanoate.

[0018] In some of the embodiments described herein, the heterologous product is a cannabinoid or cannabinoid precursor. In some embodiments, the cannabinoid or cannabinoid precursor is CBDA, CBD, CBGA, or CBG.

[0019] In another aspect, a host cell is provided, the host cell comprising a heterologous genetic pathway that produces a cannabinoid and is regulated by an exogenous agent. In some embodiments, the host cell does not comprise a precursor required to make the cannabinoid, or does not comprise an amount of precursor required to make the cannabinoid above a predetermined level (e.g., greater than 10 mg/L). In some embodiments, the host cell does not comprise hexanoate at a level sufficient to make the cannabinoid in an amount over 10 mg/L. In some embodiments, the cannabinoid is CBDA, CBD, CBGA, or CBG.

[0020] In some embodiments, the exogenous agent downregulates expression of the heterologous genetic pathway. In some embodiments, the exogenous agent that downregulates expression of the heterologous genetic pathway is glucose. In some embodiments, the expression of one or more enzymes encoded by the heterologous genetic pathway are under control of a glucose repressed promoter.

[0021] In some embodiments, the exogenous agent upregulates expression of the heterologous genetic pathway. In some embodiments, the exogenous agent that upregulates expression of the heterologous genetic pathway is galactose. In some embodiments, the expression of one or more enzymes encoded by the heterologous genetic pathway are under control of a GAL promoter.

[0022] In some embodiments, the genetic pathway encodes at least two enzymes selected from the group consisting of hexanoyl-CoA synthase (HCS), tetraketide synthase (TKS) and olivetolic acid cyclase (OAC).

[0023] In some of the aspects or embodiments described herein, the host cell can be a yeast cell or yeast strain. In some of the aspects or embodiments described herein, the yeast cell is S. cerevisiae.

[0024] In another aspect, a method for decreasing expression of a cannabinoid is provided, the method comprising culturing a host cell described herein in a media comprising an exogenous agent, wherein the exogenous agent decreases the expression of the cannabinoid or a precursor thereof. In some embodiments, the exogenous agent that decreases the expression of the cannabinoid or a precursor thereof is glucose, maltose, or lysine. In some embodiments, culturing the host cell in the media comprising the exogenous agent results in less than 0.001 mg/L of cannabinoid or a precursor thereof.

[0025] In another aspect, a method for increasing expression of a cannabinoid is provided, the method comprising culturing a host cell described herein in a media comprising the exogenous agent, wherein the exogenous agent increases the expression of the cannabinoid or a precursor thereof. In some embodiments, the exogenous agent that increases the expression of the cannabinoid or a precursor thereof that is galactose. In some embodiments, the method further comprises culturing the host cell in a media comprising hexanoate.

[0026] In some of the aspects or embodiments described herein, the cannabinoid or cannabinoid precursor is CBDA, CBD, CBGA, or CBG.

BRIEF DESCRIPTION OF THE DRAWINGS

[0027] Fig. 1 shows expression of the cannabinoid precursors olivetol and olivetolic acid by the modified host cells described herein, as described in Example 1.

[0028] Fig. 2 and Fig. 3 show genetic maps of the heterologous nucleic acids transformed into the modified host cells as described in Example 1.

[0029] Fig. 4 shows part the cannabinoid synthetic pathway referred to herein.

[0030] Fig. 5 is a schematic showing the structure and function of a maltose regulated transcriptional switch as used herein.

[0031] Fig. 6 shows the biochemical pathways for the synthesis of cannabigerol (CBG) and cannabidiol (CBD) from sugar derived geranyl pyrophosphate (GPP) and from hexanoic acid.

[0032] Fig. 7 shows the general layout of CBDA synthase (CBDAS) surface-display constructs arranged from the N to the C terminus.

[0033] Fig. 8, Fig. 9, and Fig. 10 are pairs of graphs showing normalized biomass (upper graph each) and normalized CBDA titers (lower graph each) under conditions of no hexanoic acid or 2mM hexanoic acid where each is also measured under conditions of 4% maltose, 2% maltose and 2% sucrose, and 4% sucrose for the strains Y61508 (Fig. 8); Y66316 (Fig. 9); and Y66085 (Fig. 10).

DEFINITIONS

[0034] A “genetic pathway” as used herein refers to a set of at least two different coding sequences, where the coding sequences encode enzymes that catalyze different parts of a synthetic pathway to form a desired product. In a genetic pathway a first encoded enzyme uses a substrate to make a first product which in turn is used as a substrate for a second encoded enzyme to make a second product. In some embodiments, the genetic pathway includes 3 or more members (e.g., 3, 4, 5, 6, 7, 8, 9, etc.), wherein the product of one encoded enzyme is the substrate for the next enzyme in the synthetic pathway. An example of a cannabinoid synthetic pathway is shown in FIG. 4.

[0035] As used herein, the term "endogenous" refers to a substance or process that can occur naturally in a host cell. In contrast, the term “exogenous” refers a substance or compound that originated outside an organism or cell. The exogenous substance or compound can retain its normal function or activity when introduced into an organism or host cell described herein.

[0036] The terms “modified,” “recombinant” and “engineered,” when used to modify a host cell described herein, refer to host cells or organisms that do not exist in nature, or

express compounds, nucleic acids or proteins at levels that are not expressed by naturally occurring cells or organisms.

[0037] As used herein, the term "genetically modified" denotes a host cell that comprises a heterologous nucleotide sequence. The genetically modified host cells described herein typically do not exist in nature.

[0038] The term "heterologous compound" refers to the production of a compound by a cell that does not normally produce the compound, or to the production of a compound at a level at which it is not normally produced by the cell.

[0039] As used herein, the term "heterologous" refers to what is not normally found in nature. The term "heterologous compound" refers to the production of a compound by a cell that does not normally produce the compound, or to the production of a compound at a level not normally produced by the cell. For example a cannabinoid can be a heterologous compound.

[0040] As used herein, the phrase "heterologous enzyme" refers to an enzyme that is not normally found in a given cell in nature. The term encompasses an enzyme that is: (a) exogenous to a given cell (i.e., encoded by a nucleotide sequence 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 is encoded by a nucleotide sequence 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.

[0041] A “heterologous genetic pathway” as used herein refers to a genetic pathway that does not normally or naturally exist in an organism or cell.

[0042] As used herein, the phrase "operably linked" refers to a functional linkage between nucleic acid sequences such that the linked promoter and/or regulatory region functionally controls expression of the coding sequence.

[0043] As used herein, the term "production" generally refers to an amount of compound produced by a genetically modified host cell provided herein. In some embodiments, production is expressed as a yield of the compound by the host cell. In other embodiments, production is expressed as a productivity of the host cell in producing the compound.

[0044] As used herein, the term "productivity" refers to production of a compound by a host cell, expressed as the amount of non-catabolic compound produced (by weight) per amount of fermentation broth in which the host cell is cultured (by volume) over time (per hour).

[0045] As used herein, the term "promoter" refers to a synthetic or naturally-derived nucleic acid that is capable of activating, increasing or enhancing expression of a DNA coding sequence, or inactivating, decreasing, or inhibiting expression of a DNA coding sequence. A promoter may comprise one or more specific transcriptional regulatory sequences to further enhance or repress expression and/or to alter the spatial expression and/or temporal expression of the coding sequence. A promoter may be positioned 5' (upstream) of the coding sequence under its control. A promoter may also initiate transcription in the downstream (3') direction, the upstream (5') direction, or be designed to initiate transcription in both the downstream (3') and upstream (5') directions. The distance between the promoter and a coding sequence to be expressed may be approximately the same as the distance between that promoter and the native nucleic acid sequence it controls. As is known in the art, variation in this distance may be accommodated without loss of promoter function. The term also includes a regulated promoter, which generally allows transcription of the nucleic acid sequence while in a permissive environment (e.g., microaerobic fermentation conditions, or the presence of maltose), but ceases transcription of the nucleic acid sequence while in a non-permissive environment (e.g., aerobic fermentation conditions, or in the absence of maltose). Promoters used herein can be constitutive, inducible or repressible.

[0046] The term "yield" refers to production of a compound by a host cell, expressed as the amount of compound produced per amount of carbon source consumed by the host cell, by weight.

[0047] The term "about" when modifying a numerical value or range herein includes normal variation encountered in the field, and includes plus or minus 1-10% (e.g., 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9% or 10%) of the numerical value or end points of the numerical range. Thus, a value of 10 includes all numerical values from 9 to 11. All numerical ranges described herein include the endpoints of the range unless otherwise noted, and all numerical values in-between the end points, to the first significant digit.

DETAILED DESCRIPTION OF THE INVENTION

[0048] Provided herein are recombinant or modified host cells that are useful for producing a heterologous product, and methods of using the host cells. The recombinant or modified host cells comprise a heterologous genetic pathway that can be differentially regulated by one or more exogenous agents. The recombinant host cells provide the advantage of decreasing expression of the heterologous product to below exceedingly low, and preferably undetectable levels under one set of conditions, while allowing robust expression of the heterologous product under a second set of conditions. In some embodiments, the host cell is engineered to express heterologous enzymes in the cannabinoid pathway. In some embodiments, the host cell is a yeast cell.

MODIFIED HOST CELLS COMPRISING A HETEROLOGOUS GENETIC PATHWAY

[0049] In one aspect, provided herein are host cells comprising a heterologous genetic pathway that produces a heterologous product. In some embodiments, the heterologous genetic pathway comprises a genetic regulatory element, such as a nucleic acid sequence, that is regulated by an exogenous agent. In some embodiments, the exogenous agent acts to regulate expression of the heterologous genetic pathway. Thus, in some embodiments, the exogenous agent can be a regulator of gene expression.

[0050] In some embodiments, the exogenous agent can be used as a carbon source by the host cell. For example, the same exogenous agent can both regulate expression of the heterologous genetic pathway and provide a carbon source for growth of the host cell. In some embodiments, the exogenous agent is glucose. In some embodiments, the exogenous agent is galactose. In some embodiments, the exogenous agent is maltose.

[0051] In some embodiments, the genetic regulatory element is a nucleic acid sequence, such as a promoter. In some embodiments, the genetic regulatory element is a glucose-responsive promoter or a promoter that is repressed by glucose. In some embodiments, glucose negatively regulates expression of the heterologous genetic pathway, thereby decreasing production of the heterologous product. Exemplary glucose repressed promoters include pMALll, pMAL12, pMAL13, pMAL21, pMAL22, pMAL31, pMAL32, pMAL33, pCAT8, pHXT2, pHXT4, pMTHl, and pSUC2.

Table 1: Exemplary Glucose Repressed Promoter Sequences

Promoter	Sequence
pMAL11	SEQ ID NO:
pMAL12	SEQ ID NO:
pMAL13	SEQ ID NO:
pMAL21	SEQ ID NO:
pMAL22	SEQ ID NO:
pMAL31	SEQ ID NO:
pMAL32	SEQ ID NO:
pMAL33	SEQ ID NO:
pCAT8	SEQ ID NO:
pHXT2	SEQ ID NO:
pHXT4	SEQ ID NO:
pMTH1	SEQ ID NO:
pSUC2	SEQ ID NO:

[0052] In some embodiments, the genetic regulatory element is a galactose-responsive promoter. In some embodiments, galactose positively regulates expression of the heterologous genetic pathway, thereby increasing production of the heterologous product. In some embodiments, the galactose-responsive promoter is a GAL1 promoter. In some embodiments, the galactose-responsive promoter is a GAL10 promoter. In some embodiments, the galactose-responsive promoter is a GAL2, GAL3, or GAL7 promoter. In some embodiments, heterologous genetic pathway comprises the galactose-responsive regulatory elements described in Westfall et al. (*PNAS* (2012) vol.109: E1 11-1 18). In some embodiments, the host cell lacks the *gall* gene and is unable to metabolize galactose, but galactose can still induce galactose-regulated genes.

Table 2: Exemplary GAL Promoter Sequences

Promoter	Sequence
pGAL1	SEQ ID NO:
pGAL10	SEQ ID NO:
pGAL2	SEQ ID NO:
pGAL3	SEQ ID NO:
pGAL7	SEQ ID NO:
pGAL4	SEQ ID NO:

[0053] In some embodiments, the galactose regulation system used to control expression of heterologous genes is re-configured such that it is no longer induced by the presence of galactose. Instead, the genes will be expressed unless repressors, which may be lysine in some strains or maltose in other strains, are present in the media.

[0054] In some embodiments, the genetic regulatory element is a maltose-responsive promoter. In some embodiments, maltose negatively regulates expression of the heterologous genetic pathway, thereby increasing production of the heterologous product. In some embodiments, the maltose maltose-responsive promoter is selected from the group consisting of pMALl, pMAL2, pMALll, pMAL12, pMAL31 and pMAL32. The maltose genetic regulatory element can be designed to both activate expression of some genes and repress expression of others, depending on whether maltose is present or absent in the medium. Maltose regulation of gene expression and maltose-responsive promoters are described in U.S. Patent Publication 2016/0177341, which is hereby incorporated by reference. Genetic regulation of maltose metabolism is described in Novak et al., "Maltose Transport and Metabolism in S. cerevisiae," *Food Technol. Biotechnol.* 42 (3) 213-218 (2004).

Table 3: Exemplary MAL Promoter Sequences

Promoter	Sequence
pMAL1	SEQ ID NO:

pMAL2	SEQ ID NO:
pMAL11	SEQ ID NO:
pMAL12	SEQ ID NO:
pMAL31	SEQ ID NO:
pMAL32	SEQ ID NO:

[0055] In some embodiments, the heterologous genetic pathway is regulated by a combination of the maltose and galactose regulons.

[0056] In some embodiments, the heterologous genetic pathway is regulated by lysine. The regulation of LYS genes is described, for example, by Feller et al., *Eur. J. Biochem.* 261, 163-170 (1999).

[0057] In some embodiments, the recombinant host cell does not comprise, or expresses a very low level of (for example, an undetectable amount), a precursor required to make the heterologous product. In some embodiments, the precursor is a substrate of an enzyme in the heterologous genetic pathway.

CANNABINOID PATHWAY

[0058] In another aspect, the host cell comprises a heterologous genetic pathway that produces a cannabinoid or a precursor of a cannabinoid. In some embodiments, the precursor is a substrate in the cannabinoid pathway. In some embodiments, the precursor is a substrate for hexanoyl-CoA synthase (HCS), tetraketide synthase (TKS), or olivetolic acid cyclase (OAC). In some embodiments, the precursor, substrate or intermediate in the cannabinoid pathway is hexanoate, olivetol, or olivetolic acid. In some embodiments, the precursor is hexanoate. In some embodiments, the host cell does not comprise the precursor, substrate or intermediate in an amount sufficient to produce the cannabinoid or a precursor of the cannabinoid. In some embodiments, the host cell does not comprise hexanoate at a level or in an amount sufficient to produce the cannabinoid in an amount over 10 mg/L. In some embodiments, the heterologous genetic pathway encodes at least two enzymes selected from the group consisting of hexanoyl-CoA synthase (HCS), tetraketide synthase (TKS) and olivetolic acid cyclase (OAC). The cannabinoid pathway is described in Keasling et al. (WO 2018/200888).

[0059] In some embodiments, the host cell is a yeast strain. In some embodiments, the yeast strain is a Y27600, Y27602, Y27603, or Y27604 strain.

YEAST STRAINS

[0060] In some embodiments, yeasts useful in the present methods include yeasts that have been deposited with microorganism depositories (e.g. IFO, ATCC, etc.) and belong to the genera Aciculoconidium, Ambrosiozyma, Arthroascus, Arxiozyma, Ashbya, Babjevia, Bensingtonia, Botryoascus, Botryozyma, Brettanomyces, Bullera, Bulleromyces, Candida, Citeromyces, Clavispora, Cryptococcus, Cystofilobasidium, Debaryomyces, Dekkara, Dipodascopsis, Dipodascus, Eeniella, Endomycopsella, Eremascus, Eremothecium, Erythrobasidium, Fellomyces, Filobasidium, Galactomyces, Geotrichum, Guilliermondella, Hanseniaspora, Hansenula, Hasegawaea, Holtermannia, Hormoascus, Hyphopichia, Issatchenkia, Kloeckera, Kloeckeraspora, Kluyveromyces, Kondoa, Kuraishia, Kurtzmanomyces, Leucosporidium, Lipomyces, Lodderomyces, Malassezia, Metschnikowia, Mrakia, Myxozyma, Nadsonia, Nakazawaea, Nematospora, Ogataea, Oosporidium, Pachysolen, Phachytichospora, Phaffia, Pichia, Rhodosporidium, Rhodotorula, Saccharomyces, Saccharomycodes, Saccharomycopsis, Saitoella, Sakaguchia, Satumospora, Schizoblastosporion, chizosaccharomyces, Schwanniomyces, Sporidiobolus, Sporobolomyces, Sporopachydermia, Stephanoascus, Sterigmatomyces, Sterigmatosporidium, Symbiotaphrina, Sympodiomyces, Sympodiomycopsis, Torulaspora, Trichosporiella, Trichosporon, Trigonopsis, Tsuchiyaea, Udeniomyces, Waltomyces, Wickerhamia, Wickerhamiella, Williopsis, Yamadazyma, Yarrowia, Zygoascus, Zygosaccharomyces, Zygowilliopsis, and Zygozyma, among others.

[0061] In some embodiments, the strain is Saccharomyces cerevisiae, Pichia pastoris, Schizosaccharomyces pombe, Dekkera bruxellensis, Kluyveromyces lactis (previously called Saccharomyces lactis), Kluveromyces marxianus, Arxula adeninivorans, or Hansenula polymorphs (now known as Pichia angusta). In some embodiments, the host microbe is a strain of the genus Candida, such as Candida lipolytica, Candida guilliermondii, Candida krusei, Candida pseudotropicalis, or Candida utilis.

[0062] In a particular embodiment, the strain is Saccharomyces cerevisiae. In some embodiments, the host is a strain of Saccharomyces cerevisiae selected from the group consisting of Baker's yeast, CEN.PK, CEN.PK2, CBS 7959, CBS 7960, CBS 7961, CBS 7962, CBS 7963, CBS 7964, IZ-1904, TA, BG-1, CR-1, SA-1, M-26, Y-904, PE-2, PE-5,

VR-1, BR-1, BR-2, ME-2, VR-2, MA-3, MA-4, CAT-1, CB-1, NR-1, BT-1, and AL-1. In some embodiments, the strain of Saccharomyces cerevisiae is selected from the group consisting of PE-2, CAT-1, VR-1, BG-1, CR-1, and SA-1. In a particular embodiment, the strain of Saccharomyces cerevisiae is PE-2. In another particular embodiment, the strain of Saccharomyces cerevisiae is CAT-1. In another particular embodiment, the strain of Saccharomyces cerevisiae is BG-1.

[0063] In some embodiments, the strain is a microbe that is suitable for industrial fermentation. In particular embodiments, the microbe is conditioned to subsist under high solvent concentration, high temperature, expanded substrate utilization, nutrient limitation, osmotic stress due to sugar and salts, acidity, sulfite and bacterial contamination, or combinations thereof, which are recognized stress conditions of the industrial fermentation environment.

[0064] In some embodiments, the yeast strain is a Y27598, Y27599, Y27600, Y27601 Y27602, Y27603, Y27604 or Y25618 strain. Exemplary yeast strains are shown in Table 4 below.

Table 4. Yeast Strains

Strain	Parent	Mega-stitches	Stitch A	Stitch B	locus	SNAP	Genes ex-pressed	HCS	TKS	oAC
Y27599	Y27036	101227	89523	78270	GAS4	CAIB	2xTKS	0	2	0
Y27601	Y27039	101226	85240	89531	GAS2	HC	2x TKS	0	4	0
		101227	89523	78270	GAS4	CAIB	2xTKS			
Y27598	Y27036	96695	85217	78270	GAS4	CAIB	HCS, TKS	1	1	0
Y27600	Y21791	101225	85240	85234	GAS2	HC	HCS, TKS	1	3	0
		101227	89523	78270	GAS4	CAIB	2x TKS			
Y27602	Y27039	101226	85240	89531	GAS2	HC	2x TKS	1	3	0
		96695	85217	78270	GAS4	CAIB	HCS, TKS			
Y25618	Y27039	96692	85221	85231	GAS2	HC	HCS, 2x TKS, OAC	1	2	1
Y27603	Y27039	101225	85240	85234	GAS2	HC	TKS, HCS	2	2	2
		101224	85217	89528	GAS4	CAIB	HCS, TKS, 2x OAC			
Y27604	Y27039	101229	89524	85234	GAS2	HC	2x OAC, TKS, HCS	2	2	4
		101224	85217	89528	GAS4	CAIB	HCS, TKS, 2x OAC			

MIXTURES

[0065] In another aspect, provided are mixtures of the host cells described herein and a culture media described herein. In some embodiments, the culture media comprises an exogenous agent described herein. In some embodiments, the culture media comprises an exogenous agent that decreases production of the heterologous product. In some

embodiments, exogenous agent that decreases production of the heterologous product is glucose or maltose.

[0066] In some embodiments, the culture media comprises an exogenous agent that increases production of the heterologous product. In some embodiments, the exogenous agent that increases production of the heterologous product is galactose. In some embodiments, the culture media comprises a precursor or substrate required to make the heterologous product. In some embodiments, the precursor required to make the heterologous product is hexanoate. In some embodiments, the culture media comprises an exogenous agent that increases production of the heterologous product and a precursor or substrate required to make the heterologous product. In some embodiments, the exogenous agent that increases production of the heterologous product is galactose, and the precursor or substrate required to make the heterologous product is hexanoate.

METHODS OF MAKING THE HOST CELLS

[0067] In another aspect, provided are methods of making the modified host cells described herein. In some embodiments, the methods comprise transforming a host cell with the heterologous nucleic acid constructs described herein encoding the proteins expressed by a heterologous genetic pathway described herein. Methods for transforming host cells are described in "Laboratory Methods in Enzymology: DNA", Edited by Jon Lorsch, Volume 529, (2013); and US Patent No. 9,200,270 to Hsieh, Chung-Ming, et al, and references cited therein.

METHODS FOR PRODUCING A HETEROLOGOUS PRODUCT

[0068] In another aspect, methods are provided for producing a heterologous product described herein. In some embodiments, the method decreases expression of a heterologous product. In some embodiments, the method comprises culturing a host cell comprising a heterologous genetic pathway described herein in a media comprising an exogenous agent, wherein the exogenous agent decreases the expression of the heterologous product. In some embodiments, the exogenous agent is glucose or maltose. In some embodiments, the method results in less than 0.001 mg/L of heterologous product. In some embodiments, the heterologous product is a cannabinoid or a precursor thereof.

[0069] In some embodiments, the method is for decreasing expression of a cannabinoid product or precursor thereof. In some embodiments, the method comprises culturing a host

cell comprising a heterologous cannabinoid pathway described herein in a media comprising an exogenous agent, wherein the exogenous agent decreases the expression of the cannabinoid or a precursor thereof. In some embodiments, the exogenous agent is glucose or maltose. In some embodiments, the method results in the production of less than 0.001 mg/L of cannabinoid or a precursor thereof.

[0070] In some embodiments, the method increases the expression of a heterologous product. In some embodiments, the method comprises culturing a host cell comprising a heterologous genetic pathway described herein in a media comprising the exogenous agent, wherein the exogenous agent increases expression of the heterologous product. In some embodiments, the exogenous agent is galactose. In some embodiments, the method further comprises culturing the host cell with the precursor or substrate required to make the heterologous product.

[0071] In some embodiments, the method increases the expression of a cannabinoid product or precursor thereof. In some embodiments, the method comprises culturing a host cell comprising a heterologous cannabinoid pathway described herein in a media comprising an exogenous agent, wherein the exogenous agent increases the expression of the cannabinoid or a precursor thereof. In some embodiments, the exogenous agent is galactose. In some embodiments, the method further comprises culturing the host cell with a precursor or substrate required to make the heterologous cannabinoid product or precursor thereof. In some embodiments, the precursor required to make the heterologous cannabinoid product or precursor thereof is hexanoate. In some embodiments, the combination of the exogenous agent and the precursor or substrate required to make the heterologous cannabinoid product or precursor thereof produces a higher yield of cannabinoid than the exogenous agent alone.

[0072] In some embodiments, the cannabinoid or a precursor thereof is cannabidiolic acid (CBDA), CBD, cannabigerolic acid (CBGA), or CBG.

NUCLEIC ACIDS

[0073] Due to the inherent degeneracy of the genetic code, other polynucleotides which encode substantially the same or functionally equivalent polypeptides can also be used to clone and express the polynucleotides encoding the protein components of the heterologous genetic pathway described herein.

[0074] As will be understood by those of skill in the art, it can be advantageous to modify a coding sequence to enhance its expression in a particular host. The genetic code is redundant with 64 possible codons, but most organisms typically use a subset of these codons. The codons that are utilized most often in a species are called optimal codons, and those not utilized very often are classified as rare or low-usage codons. Codons can be substituted to reflect the preferred codon usage of the host, in a process sometimes called "codon optimization" or "controlling for species codon bias."

[0075] Optimized coding sequences containing codons preferred by a particular prokaryotic or eukaryotic host (Murray et al, 1989, Nucl Acids Res. 17: 477-508) can be prepared, for example, to increase the rate of translation or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, as compared with transcripts produced from a non-optimized sequence. Translation stop codons can also be modified to reflect host preference. For example, typical stop codons for S. cerevisiae and mammals are UAA and UGA, respectively. The typical stop codon for monocotyledonous plants is UGA, whereas insects and E. coli commonly use UAA as the stop codon (Dalphin et al, 1996, Nucl Acids Res. 24: 216-8).

[0076] Those of skill in the art will recognize that, due to the degenerate nature of the genetic code, a variety of DNA molecules differing in their nucleotide sequences can be used to encode a given enzyme of the disclosure. The native DNA sequence encoding the biosynthetic enzymes described above are referenced herein merely to illustrate an embodiment of the disclosure, and the disclosure includes DNA molecules of any sequence that encode the amino acid sequences of the polypeptides and proteins of the enzymes utilized in the methods of the disclosure. In similar fashion, a polypeptide can typically tolerate one or more amino acid substitutions, deletions, and insertions in its amino acid sequence without loss or significant loss of a desired activity. The disclosure includes such polypeptides with different amino acid sequences than the specific proteins described herein so long as the modified or variant polypeptides have the enzymatic anabolic or catabolic activity of the reference polypeptide. Furthermore, the amino acid sequences encoded by the DNA sequences shown herein merely illustrate embodiments of the disclosure.

[0077] In addition, homologs of enzymes useful for the compositions and methods provided herein are encompassed by the disclosure. In some embodiments, two proteins (or a region of the proteins) are substantially homologous when the amino acid sequences have at

least about 30%, 40%, 50% 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity. To determine the percent identity of two amino acid sequences, or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In one embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, typically at least 40%, more typically at least 50%, even more typically at least 60%, and even more typically at least 70%, 80%, 90%, 100% of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

[0078] When "homologous" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (e.g., charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of homology may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art (See, e.g., Pearson W. R., 1994, Methods in Mol Biol 25: 365-89).

[0079] The following six groups each contain amino acids that are conservative substitutions for one another: 1) Serine (S), Threonine (T); 2) Aspartic Acid (D), Glutamic Acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Alanine (A), Valine (V), and 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

[0080] Sequence homology for polypeptides, which is also referred to as percent sequence identity, is typically measured using sequence analysis software. A typical algorithm used comparing a molecule sequence to a database containing a large number of sequences from different organisms is the computer program BLAST. When searching a database containing sequences from a large number of different organisms, it is typical to compare amino acid sequences.

[0081] Furthermore, any of the genes encoding the foregoing enzymes (or any others mentioned herein (or any of the regulatory elements that control or modulate expression thereol)) may be optimized by genetic/protein engineering techniques, such as directed evolution or rational mutagenesis, which are known to those of ordinary skill in the art. Such action allows those of ordinary skill in the art to optimize the enzymes for expression and activity in a host cell, for example, a yeast.

[0082] In addition, genes encoding these enzymes can be identified from other fungal and bacterial species and can be expressed for the modulation of this pathway. A variety of organisms could serve as sources for these enzymes, including, but not limited to, Saccharomyces spp., including S. cerevisiae and S. uvarum, Kluyveromyces spp., including K. thermotolerans, K. lactis, and K. marxianus, Pichia spp., Hansenula spp., including H. polymorphs, Candida spp., Trichosporon spp., Yamadazyma spp., including Y. spp. stipitis, Torulaspora pretoriensis, Issatchenkia orientalis, Schizosaccharomyces spp., including S. pombe, Cryptococcus spp., Aspergillus spp., Neurospora spp., or Ustilago spp. Sources of genes from anaerobic fungi include, but are not limited to, Piromyces spp., Orpinomyces spp., or Neocallimastix spp. Sources of prokaryotic enzymes that are useful include, but are not limited to, Escherichia coli, Zymomonas mobilis, Staphylococcus aureus, Bacillus spp., Clostridium spp., Corynebacterium spp., Pseudomonas spp., Lactococcus spp., Enterobacter spp., and Salmonella spp.

[0083] Techniques known to those skilled in the art may be suitable to identify additional homologous genes and homologous enzymes. Generally, analogous genes and/or analogous enzymes can be identified by functional analysis and will have functional similarities. Techniques known to those skilled in the art may be suitable to identify analogous genes and analogous enzymes. For example, to identify homologous or analogous ADA genes, proteins, or enzymes, techniques may include, but are not limited to, cloning a gene by PCR using primers based on a published sequence of an ADA gene/enzyme or by degenerate PCR using

degenerate primers designed to amplify a conserved region among ADA genes. Further, one skilled in the art can use techniques to identify homologous or analogous genes, proteins, or enzymes with functional homology or similarity. Techniques include examining a cell or cell culture for the catalytic activity of an enzyme through in vitro enzyme assays for said activity (e.g. as described herein or in Kiritani, K., Branched-Chain Amino Acids Methods Enzymology, 1970), then isolating the enzyme with said activity through purification, determining the protein sequence of the enzyme through techniques such as Edman degradation, design of PCR primers to the likely nucleic acid sequence, amplification of said DNA sequence through PCR, and cloning of said nucleic acid sequence. To identify homologous or similar genes and/or homologous or similar enzymes, analogous genes and/or analogous enzymes or proteins, techniques also include comparison of data concerning a candidate gene or enzyme with databases such as BRENDA, KEGG, or MetaCYC. The candidate gene or enzyme may be identified within the above mentioned databases in accordance with the teachings herein.

[0084] In some embodiments, the nucleic acid sequences encode proteins or polypeptides having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to the amino acid sequence of a protein or enzyme encoded by a heterologous genetic pathway described herein. In some embodiments, the nucleic acid sequences encode proteins or polypeptides having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to the amino acid sequence of HCS, TKS, or OAC.

CULTURE AND FERMENTATION METHODS

[0085] Materials and methods for the maintenance and growth of microbial cultures are well known to those skilled in the art of microbiology or fermentation science (see, for example, Bailey et al., Biochemical Engineering Fundamentals, second edition, McGraw Hill, New York, 1986). Consideration must be given to appropriate culture medium, pH, temperature, and requirements for aerobic, microaerobic, or anaerobic conditions, depending on the specific requirements of the host cell, the fermentation, and the process.

[0086] The methods of producing heterologous products provided herein may be performed in a suitable culture medium (e.g., with or without pantothenate supplementation) in a suitable container, including but not limited to a cell culture plate, a flask, or a fermentor. Further, the methods can be performed at any scale of fermentation known in the art to

support industrial production of microbial products. Any suitable fermentor may be used including a stirred tank fermentor, an airlift fermentor, a bubble fermentor, or any combination thereof. In particular embodiments utilizing Saccharomyces cerevisiae as the host cell, strains can be grown in a fermentor as described in detail by Kosaric, et al, in Ullmann's Encyclopedia of Industrial Chemistry, Sixth Edition, Volume 12, pages 398-473, Wiley-VCH Verlag GmbH & Co. KDaA, Weinheim, Germany.

[0087] In some embodiments, the culture medium is any culture medium in which a genetically modified microorganism capable of producing a heterologous product can subsist, i.e., maintain growth and viability. In some embodiments, the culture medium is an aqueous medium comprising assimilable carbon, nitrogen and phosphate sources. Such a medium can also include appropriate salts, minerals, metals and other nutrients. In some embodiments, the carbon source and each of the essential cell nutrients, are added incrementally or continuously to the fermentation media, and each required nutrient is maintained at essentially the minimum level needed for efficient assimilation by growing cells, for example, in accordance with a predetermined cell growth curve based on the metabolic or respiratory function of the cells which convert the carbon source to a biomass.

[0088] Suitable conditions and suitable media for culturing microorganisms are well known in the art. In some embodiments, the suitable medium is supplemented with one or more additional agents, such as, for example, an inducer (e.g., when one or more nucleotide sequences encoding a gene product are under the control of an inducible promoter), a repressor (e.g., when one or more nucleotide sequences encoding a gene product are under the control of a repressible promoter), or a selection agent (e.g., an antibiotic to select for microorganisms comprising the genetic modifications).

[0089] In some embodiments, the carbon source is a monosaccharide (simple sugar), a disaccharide, a polysaccharide, a non-fermentable carbon source, or one or more combinations thereof. Non-limiting examples of suitable monosaccharides include glucose, galactose, mannose, fructose, ribose, and combinations thereof. Non-limiting examples of suitable disaccharides include sucrose, lactose, maltose, trehalose, cellobiose, and combinations thereof. Non-limiting examples of suitable polysaccharides include starch, glycogen, cellulose, chitin, and combinations thereof. Non-limiting examples of suitable non-fermentable carbon sources include acetate and glycerol.

[0090] The concentration of a carbon source, such as glucose, in the culture medium should promote cell growth, but not be so high as to repress growth of the microorganism used. Typically, cultures are run with a carbon source, such as glucose, being added at levels to achieve the desired level of growth and biomass. Production of heterologous products may also occur in these culture conditions, but at undetectable levels (with detection limits being about <0.1 g/1). In other embodiments, the concentration of a carbon source, such as glucose, in the culture medium is greater than about 1 g/L, preferably greater than about 2 g/L, and more preferably greater than about 5 g/L. In addition, the concentration of a carbon source, such as glucose, in the culture medium is typically less than about 100 g/L, preferably less than about 50 g/L, and more preferably less than about 20 g/L. It should be noted that references to culture component concentrations can refer to both initial and/or ongoing component concentrations. In some cases, it may be desirable to allow the culture medium to become depleted of a carbon source during culture.

[0091] Sources of assimilable nitrogen that can be used in a suitable culture medium include, but are not limited to, simple nitrogen sources, organic nitrogen sources and complex nitrogen sources. Such nitrogen sources include anhydrous ammonia, ammonium salts and substances of animal, vegetable and/or microbial origin. Suitable nitrogen sources include, but are not limited to, protein hydrolysates, microbial biomass hydrolysates, peptone, yeast extract, ammonium sulfate, urea, and amino acids. Typically, the concentration of the nitrogen sources, in the culture medium is greater than about 0.1 g/L, preferably greater than about 0.25 g/L, and more preferably greater than about 1.0 g/L. Beyond certain concentrations, however, the addition of a nitrogen source to the culture medium is not advantageous for the growth of the microorganisms. As a result, the concentration of the nitrogen sources, in the culture medium is less than about 20 g/L, preferably less than about 10 g/L and more preferably less than about 5 g/L. Further, in some instances it may be desirable to allow the culture medium to become depleted of the nitrogen sources during culture.

[0092] The effective culture medium can contain other compounds such as inorganic salts, vitamins, trace metals or growth promoters. Such other compounds can also be present in carbon, nitrogen or mineral sources in the effective medium or can be added specifically to the medium.

[0093] The culture medium can also contain a suitable phosphate source. Such phosphate sources include both inorganic and organic phosphate sources. Preferred phosphate sources include, but are not limited to, phosphate salts such as mono or dibasic sodium and potassium phosphates, ammonium phosphate and mixtures thereof. Typically, the concentration of phosphate in the culture medium is greater than about 1.0 g/L, preferably greater than about 2.0 g/L and more preferably greater than about 5.0 g/L. Beyond certain concentrations, however, the addition of phosphate to the culture medium is not advantageous for the growth of the microorganisms. Accordingly, the concentration of phosphate in the culture medium is typically less than about 20 g/L, preferably less than about 15 g/L and more preferably less than about 10 g/L.

[0094] A suitable culture medium can also include a source of magnesium, preferably in the form of a physiologically acceptable salt, such as magnesium sulfate heptahydrate, although other magnesium sources in concentrations that contribute similar amounts of magnesium can be used. Typically, the concentration of magnesium in the culture medium is greater than about 0.5 g/L, preferably greater than about 1.0 g/L, and more preferably greater than about 2.0 g/L. Beyond certain concentrations, however, the addition of magnesium to the culture medium is not advantageous for the growth of the microorganisms. Accordingly, the concentration of magnesium in the culture medium is typically less than about 10 g/L, preferably less than about 5 g/L, and more preferably less than about 3 g/L. Further, in some instances it may be desirable to allow the culture medium to become depleted of a magnesium source during culture.

[0095] In some embodiments, the culture medium can also include a biologically acceptable chelating agent, such as the dihydrate of trisodium citrate. In such instance, the concentration of a chelating agent in the culture medium is greater than about 0.2 g/L, preferably greater than about 0.5 g/L, and more preferably greater than about 1 g/L. Beyond certain concentrations, however, the addition of a chelating agent to the culture medium is not advantageous for the growth of the microorganisms. Accordingly, the concentration of a chelating agent in the culture medium is typically less than about 10 g/L, preferably less than about 5 g/L, and more preferably less than about 2 g/L.

[0096] The culture medium can also initially include a biologically acceptable acid or base to maintain the desired pH of the culture medium. Biologically acceptable acids include, but are not limited to, hydrochloric acid, sulfuric acid, nitric acid, phosphoric acid and mixtures

thereof. Biologically acceptable bases include, but are not limited to, ammonium hydroxide, sodium hydroxide, potassium hydroxide and mixtures thereof. In some embodiments, the base used is ammonium hydroxide.

[0097] The culture medium can also include a biologically acceptable calcium source, including, but not limited to, calcium chloride. Typically, the concentration of the calcium source, such as calcium chloride, dihydrate, in the culture medium is within the range of from about 5 mg/L to about 2000 mg/L, preferably within the range of from about 20 mg/L to about 1000 mg/L, and more preferably in the range of from about 50 mg/L to about 500 mg/L.

[0098] The culture medium can also include sodium chloride. Typically, the concentration of sodium chloride in the culture medium is within the range of from about 0.1 g/L to about 5 g/L, preferably within the range of from about 1 g/L to about 4 g/L, and more preferably in the range of from about 2 g/L to about 4 g/L.

[0099] In some embodiments, the culture medium can also include trace metals. Such trace metals can be added to the culture medium as a stock solution that, for convenience, can be prepared separately from the rest of the culture medium. Typically, the amount of such a trace metals solution added to the culture medium is greater than about 1 ml/L, preferably greater than about 5 mL/L, and more preferably greater than about 10 mL/L. Beyond certain concentrations, however, the addition of a trace metals to the culture medium is not advantageous for the growth of the microorganisms. Accordingly, the amount of such a trace metals solution added to the culture medium is typically less than about 100 mL/L, preferably less than about 50 mL/L, and more preferably less than about 30 mL/L. It should be noted that, in addition to adding trace metals in a stock solution, the individual components can be added separately, each within ranges corresponding independently to the amounts of the components dictated by the above ranges of the trace metals solution.

[0100] The culture media can include other vitamins, such as pantothenate, biotin, calcium, pantothenate, inositol, pyridoxine-HCl, and thiamine-HCl. Such vitamins can be added to the culture medium as a stock solution that, for convenience, can be prepared separately from the rest of the culture medium. Beyond certain concentrations, however, the addition of vitamins to the culture medium is not advantageous for the growth of the microorganisms.

[0101] The fermentation methods described herein can be performed in conventional culture modes, which include, but are not limited to, batch, fed-batch, cell recycle, continuous

and semi-continuous. In some embodiments, the fermentation is carried out in fed-batch mode. In such a case, some of the components of the medium are depleted during culture, including pantothenate during the production stage of the fermentation. In some embodiments, the culture may be supplemented with relatively high concentrations of such components at the outset, for example, of the production stage, so that growth and/or production is supported for a period of time before additions are required. The preferred ranges of these components are maintained throughout the culture by making additions as levels are depleted by culture. Levels of components in the culture medium can be monitored by, for example, sampling the culture medium periodically and assaying for concentrations. Alternatively, once a standard culture procedure is developed, additions can be made at timed intervals corresponding to known levels at particular times throughout the culture. As will be recognized by those in the art, the rate of consumption of nutrient increases during culture as the cell density of the medium increases. Moreover, to avoid introduction of foreign microorganisms into the culture medium, addition is performed using aseptic addition methods, as are known in the art. In addition, a small amount of anti-foaming agent may be added during the culture.

[0102] The temperature of the culture medium can be any temperature suitable for growth of the genetically modified cells and/or production of compounds of interest. For example, prior to inoculation of the culture medium with an inoculum, the culture medium can be brought to and maintained at a temperature in the range of from about 20.degree. C. to about 45.degree. C., preferably to a temperature in the range of from about 25.degree. C. to about 40.degree. C., and more preferably in the range of from about 28.degree. C. to about 32.degree. C.

[0103] The pH of the culture medium can be controlled by the addition of acid or base to the culture medium. In such cases when ammonia is used to control pH, it also conveniently serves as a nitrogen source in the culture medium. Preferably, the pH is maintained from about 3.0 to about 8.0, more preferably from about 3.5 to about 7.0, and most preferably from about 4.0 to about 6.5.

[0104] In some embodiments, the carbon source concentration, such as the glucose concentration, of the culture medium is monitored during culture. Glucose concentration of the culture medium can be monitored using known techniques, such as, for example, use of the glucose oxidase enzyme test or high pressure liquid chromatography, which can be used

to monitor glucose concentration in the supernatant, e.g., a cell-free component of the culture medium. As stated previously, the carbon source concentration should be kept below the level at which cell growth inhibition occurs. Although such concentration may vary from organism to organism, for glucose as a carbon source, cell growth inhibition occurs at glucose concentrations greater than at about 60 g/L, and can be determined readily by trial. Accordingly, when glucose is used as a carbon source the glucose is preferably fed to the fermentor and maintained below detection limits. Alternatively, the glucose concentration in the culture medium is maintained in the range of from about 1 g/L to about 100 g/L, more preferably in the range of from about 2 g/L to about 50 g/L, and yet more preferably in the range of from about 5 g/L to about 20 g/L. Although the carbon source concentration can be maintained within desired levels by addition of, for example, a substantially pure glucose solution, it is acceptable, and may be preferred, to maintain the carbon source concentration of the culture medium by addition of aliquots of the original culture medium. The use of aliquots of the original culture medium may be desirable because the concentrations of other nutrients in the medium (e.g. the nitrogen and phosphate sources) can be maintained simultaneously. Likewise, the trace metals concentrations can be maintained in the culture medium by addition of aliquots of the trace metals solution.

EXAMPLES

Example 1: Host cells engineered with the cannabinoid synthetic pathway

[0105] Yeast were engineered to express part the cannabinoid synthetic pathway. As shown in Fig. 4, the enzymes hexanoyl-CoA synthase (HCS), tetraketide synthase (TKS) and olivetolic acid cyclase (OAC) synthesize olivetolic acid starting from hexanoate as a substrate. HCS uses hexanoate as a substrate to form hexanoyl-CoA, which in turn is used as a substrate by TKS to for malonyl-CoA, which in turn is used as a substrate by OAC to form olivetolic acid. Coding sequences for each of HCS, TKS, and OAC, each under the control of a GAL promoter, were inserted into *S. cerevisiae* yeast cells. Accordingly, synthesis of each of these enzymes was induced only if the yeast was grown in the presence of galactose. As shown in Table 4, Fig. 2 and Fig. 3, several constructs (and resulting yeast strains) were made, some of which only expressed a subset of HCS, TKS, and OAC whereas other constructs and yeast strains contained at least one copy of each of HCS, TKS, and OAC under control of the GAL promoter. Table 4, Fig. 2 and Fig. 3 are useful for understanding which strains were tested for the data shown in Fig. 1.

[0106] In the case of the cannabinoid pathway, hexanoate can be fed to provide the hexanoyl-coenzyme A substrate required for production of the polyketide precursor to cannabinoids (see Fig. 4). Wild type yeast produces very low levels of hexanoate, so if it is not fed, cannabinoid production is greatly reduced. Fig. 1 shows the level of the cannabinoid precursors olivetol and olivetolic acid produced by various yeast strains engineered for switchable expression of the pathway genes (HCS, TKS, and OAC) and grown under three conditions. In the first two conditions, no hexanoate was fed to the strains and the carbon source was either glucose (glue; turns off pathway expression) or galactose (gal; turns on pathway expression). In the third (right-most) condition, galactose was the carbon source, which activates the pathway genes and hexanoate was fed to the yeast. As can be seen, when galactose was the carbon source and hexanote was fed to the yeast, significant amounts of the cannabinoid precursors were produced. On the other hand when glucose was the carbon source, thereby turning off expression of the cannabinoid pathway, and hexanoate was not fed, cannabinoid production was below the limit of detection of the assay (<0.001 mg/L). This example demonstrates the use of two orthogonal switching systems (galactose-induced pathway expression, and hexanoate addition) to ensure the complete turn-off of production of olivetol and olivetolic acid. Similar orthogonal switching systems in which a precursor of a pathway must be supplied exogenously in combination with a genetic switch (e.g., an induced promoter or alternatively a repressed promoter) can be used to control other heterologous pathways introduced into yeast.

Example 2: Yeast Transformation Methods

[0107] Each DNA construct was integrated into *Saccharomyces cerevisiae* (CEN.PK1 13-7D) using 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 Oü $_{\beta oo}$ of 0.1 in 100 mL YPD, and grown to an $_{OD6OO}$ 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 pL 1 M LiAc, 10 pL boiled salmon sperm DNA, and 74 pL of donor DNA. For transformations that required expression of the endonuclease F-Cphl, the donor DNA included a plasmid carrying the F-Cphl gene expressed under the yeast TDH3 promoter for expression. This will cut the F-Cphl

endonuclease recognition site in the landing pad to facilitate integration of the target gene of interest. 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.

Example 2: Generation of a base strain with a genetic switching system that is suitable for rapid genetic engineering for the production of non-catabolic compounds.

[0108] To generate a strain that can be rapidly engineered to make an arbitrary natural compound, several engineering steps were performed on the original yeast isolate CEN.PK113-7D. First, a meganuclease protein was integrated into the chromosome to enable nuclease-based engineering in subsequent rounds of transformation. Second, seven chromosomal loci were engineered to gain nucleotide sequences that enable high-efficiency integration of future DNA constructs using validated nucleases. Third, a maltose-responsive genetic switch was added to control the expression of genes driven by GAL promoters (pGALx). The resulting strain Y46850 serves as a chassis into which designs for natural compound biosynthesis may be rapidly prototyped.

[0109] The invention and uses of the maltose-responsive genetic switch were previously described in W02016210350; US201615738555; and US201615738918, each of which are incorporated herein by reference in their entireties. In brief, the genetic switch enables a heterologous, non-catabolic pathway to switch between On and Off states in response to maltose and temperature (Fig. 5). When the strain is grown in the presence of maltose and at temperatures <28°C, the expression of all pGALx-driven genes will be Off, allowing cellular resources to instead go toward the generation of biomass, i.e. growth. Conversely, when the strain is grown in the absence of maltose and at temperatures >30°C, the expression of all pGALx-driven genes will be On, enabling high-yield conversion of fed sucrose into a non-catabolic product.

[0110] The maltose switch is a *GAL80* based switch, wherein a maltose-responsive promoter drives expression of *GAL80 (pMALx>GAL80*). A challenge of *GAL80* based switches is that mutations that reactivate Gal80p activity in fermentations will shut down biosynthetic production, an event favored by natural selection. Two major approaches were developed to reduce *GAL80* reactivation. First, a UBRl-targeted degron (D) was fused to a temperature sensitive *GAL80* (GAL80tsl) to speed up Gal80 protein degradation when maltose is depleted and the temperature is >30°C. Second, the GAL80 protein was further

destabilized by fusing a maltose binding protein (MBP) based degron onto the C-terminus. When maltose is present, the GAL80p-MBP mutant fusion protein is stable; however, when maltose is depleted, the GAL80 protein is quickly degraded. Another benefit of using the MBP mutant is that strains with D_GAL80tsl_MBP showed significantly lower "leakiness" of GAL gene expression during growth in OFF-state conditions.

Example 3: Generation of a strain capable of producing cannabigerolic acid (CBGA).

[0111] A set of genes capable of producing the cannabinoid CBGA was engineered into strain Y46850 in three steps (Table 5 and Fig. 6). First, constructs were integrated into chromosomal loci to express three heterologous genes from *Cannibis sativa* AAE, TKS, and OAC, together with the *Zymomonas mobilis* PDC gene and two endogenous *S. cerevisiae* ACS1 and ALD6 genes were, all using pGALx promoters. Second, constructs were integrated into chromosomal loci to express seven endogenous genes of the *S. cerevisiae* mevalonate pathway (ERG10, ERG13, catalytic domain of HMG1, ERG12, ERG8, MVD1, and IDI1). Third, constructs were integrated into chromosomal loci to express *Streptomyces aculeolatus* GPPS and *Cannabis sativa* CBGa synthase (CBGAS) gene. The CBGAS gene required extensive N-terminal engineering to enable its expression in a catalytically active form and that did not inhibit the growth of yeast. This engineering is described elsewhere (forthcoming patent application on DPL1-PT4 engineering and TM78-hop chimeragenesis). The resulting strain Y61508 is capable of producing CBGA when fed a mixture of sucrose and hexanoic acid, as described in the Yeast culturing conditions section below.

[0112] Notably, genes involved in the production of hexanoic acid have not been engineered into this strain. Endogenous yeast metabolism produces a negligible amount of hexanoic acid or hexanoyl-CoA, which means the strains are dependent on the exogenous supply of hexanoic acid to produce cannabinoids (Fig. 6).

Table 5.

Enzyme	**SEQ ID NOs**	**Promoter**
Sc.ACS1		pGAL10
Sc.ALD6		pGAL1
Zm.PDC		pGAL7

Cs.AAE		2 x pGAL10
Cs.TKS		2 x pGAL10
Cs.OAC		4 x pGAL1
Sc.ERG10		pGAL2
Sc.ERG13		pGAL1
Sc.HMG1-truncated		pGAL10
Sc.ERG12		pGAL2
Sc.ERG8		pGAL1
Sc.MVD1		pGAL10
Sc.IDI1		pGAL7
Sa.GPPS		pGAL10
Sc.DPL1–Cs.PT4 fusion		pGAL1
CWP2_CBDAS_12aalink4_SAG1		pGAL1

Example 4: Generation of a strain capable of producing cannabidiolic acid (CBDA).

[0113] Cannabidiolic acid synthase (CBDAS) is an oxidative cyclase that creates a carbon-carbon bond to fold the geranyl moiety of CBGA into a 6-member ring. CBDAS belongs to the Berberine-Bridge Enzyme family that employs a bicovalently bound flavin mononucleotide in the active site to utilize molecular oxygen, and each reaction cycle also produces a molecule of hydrogen peroxide (H_2O_2). CBDAS in *Cannabis sativa* has disulfide bonds, is glycosylated, and is natively secreted into the apoplastic space of trichomes, which is thought to have evolved to prevent auto-toxicity via H_2O_2 generation. A further challenge to functionally expressing CBDAS in yeast is its narrow pH range of ~4.5-5.

[0114] Yeast surface display is a classic molecular biology technique where a protein of interest is hosted on the exterior surface of yeast cells, allowing the protein to interact directly with the media. Surface display fulfills the requirements for CBDAS activity as surface proteins are glycosylated (emanating from the Golgi), and the pH of fermentation media is

low. Surface display is preferable to secretion, as pumping protein into the broth could lead to foaming issues. To design a protein construct for CBDAS surface display, we selected the yeast cell wall mannoprotein CWP2 to supply the signal sequence and SAG1 to serve as a carrier protein (Fig.7 and Table 5). This construct was integrated into a nearly isogenic sibling of strain Y61508 to generate strain Y66085.

Example 5: Yeast culturing conditions.

[0115] For routine strain characterization in a 96-well-plate format, yeast colonies were picked into a 1.1-mL-per-well capacity 96-well 'PreCulture plate' filled with 360 µL per well of Pre-Culture media. Pre-Culture media consists of Bird Seed Media (BSM, originally described by van Hoek et al, (2000), *Biotechnology and Bioengineering*, vol. 68, pp. 517-523) at pH 5.05 with 14 g/L sucrose, 7 g/L maltose, 3.75g/L ammonium sulfate, and 1 g/L lysine. Cells were cultured at 28°C in a high capacity microtiter plate incubator shaking at 1000 rpm and 80% humidity for 3 days until the cultures reached carbon exhaustion.

[0116] The growth-saturated cultures were sub-cultured by taking 14.4 pL from the saturated cultures and diluting into into a 2.2-mL-per-well capacity 96-well 'Production plate' filled with 360 pL per well of Production media. Production media consists of BSM at pH 5.05 with 40 g/L sucrose, 3.75g/L ammonium sulfate, and 2 mM hexanoic acid. 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.

Example 6: Analytical methods for cannabinoid extraction and titer determination.

[0117] At the conclusion of the incubation of the Production plate, methanol is added to each well such that the final concentration is 67% (v/v) methanol. An impermeable seal is added, and the plate is shaken at 1000 rpm for 30 seconds to lyse the cells and extract cannabinoids. The plate is centrifuged for 30 seconds at 200 × g to pellet cell debris. 300 pL of the clarified sample is moved to an empty 1.1-mL-capacity 96-well plate and sealed with a foil seal. The sample plate is stored at -20C until analysis

[0118] Cannabidiolic acid (CBDA) and cannabigerolic acid (CBGA) were separated using a Thermo Vanquish Series UPLC-UV system with an Accupore Polar Premium 2.6pm C18 column (100 x 2.1mm). The mobile phase was a gradient of 5mM Ammonium Formate with 0.1% formic acid aqueous solution and 0.1% formic acid in acetonitrile at a flow rate of

1.2ml/min. Calibration curves were prepared by weight in the extraction solvent using neat standards.

Example 7: Validation of two orthogonal switching systems.

[0119] For some biomolecules, such as cannabinoids, regulatory requirements create a need for extremely low or non-detectable production during the growth phase required to propagate the strain. To this end, the geneticly encoded maltose-responsive switch was combined with the dependency on exogenously supplied hexanoic acid for cannabinoid biosynthesis.

[0120] When strain Y61508 was grown in the absence of maltose and the presence of hexanoic acid, the highest CBGA titer and lowest biomass accumulation was observed (Fig. 8), which is consistent with the channeling of cellular resources into this non-catabolic pathway. As the sucrose was replaced by maltose, the CBGA titer decreases and the biomass accumulation increases. When exogenous hexanoic acid is no longer supplied, the CBGA titer also decreases and the biomass accumulation increases. Highly similar results were observed for the distinct CBGA-producing strain Y66316 (Fig. 9).

[0121] Importantly, the highest biomass accumulation and lowest CBGA titer was observed when these strains were grown in the presence of 4% maltose and without the exogenous supply of hexanoic acid. In this condition, cannabinoid production was below the limit of detection of the assay (<0.001 mg/L). This example demonstrates the use of two orthogonal switching systems to ensure the complete turn-off of cannabinoid production and channeling of cellular resources instead to biomass accumulation, i.e. growth.

[0122] To extend this finding, we tested the CBDA-production strain Y66085 in the same conditions. Once again, the absence of maltose and the exogenous supply of hexanoic acid allowed the cells to switch fully into cannabinoid production at the expense of growth (Fig. 10). By substituting the sucrose with maltose, or by removing the exogenous supply of hexanoic acid, the CBDA titer decreased and biomass increased. This example demonstrates the use of two orthogonal switching systems extends to multiple strains engineered to produce different cannabinoids.

[0123] It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be

suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, including genbank accession numbers, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.

5

INFORMAL SEQUENCE LISTING

MS101225

GACGGCACGGCCACGCGTTTAAACCGCCTACGCCATCATTAAAGACCTGGTCAACTATAA

AATAATACAATCAATACTTGCTTGAACGCTTGATTTTACTGATATTCTATCCAAAAGCAA

GTAGACCAGAAACTCTCAAGATGTTGCAAATACCGTTCGATGTTTTTGGTTTAGATTGTT

TTAATGTTGATGCTTTTTTACTTATTTTTGGAAGCGTCTTTTTAATTTAGTTTTATATTA

TAGGTATATGAATGT GTTTATGCCAATAAGGGTTTTTTTGTACAGTTATGTGATTATAAA

CAGTCTTTTGTCTAGTTTTTTTCACCAGTATCGGCCTCTATTTATAAAAAACGGAGCAGC

TTTCGGT GTCAGTAATTCTGAAAAAATTTGTGTCACTCTGATTGTAAATGAATTAATTTA

GCTAGATAGTTGCGAGCCCCAACGAGAAGATTGTCAGACAAAGACAACATTCAACAACCT

ACATCCGTTACTATTCGTTAACTCGAGGTACTTGAAACTTTTCAGTTAAGTCGCTCGTCC

AACGCCGGCGGACCTGCGAGTAAGCAACTCTGGCGCTGGCATGGCATAACCGGCGACGGC

AATGCGCAAGATGGGATGCTATGGGCAGAGAGCCGTACTTTACTGCTTATGGCACTACAG

CAACAGATGGTTACCCCACTAAGCCTGAAGCGAATCGCCATCAATTCTGCGCAGTGGCGA

GGAGATAAAAGCGCGGAAGTCATTCATCAACTGGCGACGCTACTCAAAGCAGGGTTAACG

CTTTCTGAAGGGCTGGCTCTGCTGGCGGAACAGCATCCCAGTAAGCAATGGCAAGCGTTG

CTGCAATCGCTGGCGCACGATCTCGAACAGGGCATTGCTTTTTCCAATGCCTTATTACCC

TGGTCAGAGGTATTTCCGCCGCTCTATCAGGCGATGATCCGCACGGGTGAACTGACCGGT

AAGCTGGATGAATGCTGCTTTGAACTGGCGCGTCAGCAAAAAGCCCAGCGTCAGTTGACC

GACAAAGTGAAATCAGCGTTACGTTATCCCATCATCATTTTAGCGATGGCAATCATGGTG

GTTGTGGCAATGCTGCATTTTGTTCTGCCGGAGTTTGCCGCTATCTATAAGACCTTCAAC

ACCCCACTACCGGCACTAACGCAGGGGATCATGACGCTGGCAGACTTTAGTGGCGAATGG

AGCTGGCTGCTGGTGTTGTTCGGCTTTCTGCTGGCGATAGCCAATAAGTTGCTGAACGGC

CGGCCAAGCACGCGGGGATCAGTAGGACAAAGGGTTCTCGTAGAGTCCCCGGAAAAAAAA

AAGGACAAAAAGTTTCAAGACGGCAATCTCTTTTTACTGCATCTCGTCAGTTGGCAACTT

GCCAAGAACTTCGCAAATGACTTTGACATATGATAAGACGTCAACTGCCCCACGTACAAT

AACAAAATGGTAGTCATATTATGTCAAGAATAGGTATCCAAAACGCAGCGGTTGAAAGCA

TATCAAGAATTGT GTCCCTGTGTTTCAAAGTTTGTGGATAATCGAAATCTCTTACATTGA

AAACATTATCATACAATCATTTATTAAGTAGTTGAAGCATGTATGAACTATAAAAGTGTT

ACTACTCGTTATTATTGTGTACTTTGTGATGCTAAAGTTATGAGTAGAAAAAAATGAGAA

GTTGTTCTGAACAAAGTAAAAAAAAGAAGTATACTTATTCAAAATGGGAGAATTGTTGAC

GCAAAACTCTACGCATGATCTTGTTGGTGGCAGTTCTAGGCAAAGAAGACAAAGGGACGA

CTCTAGTAACCTTAAACAATGGATTCAACTTCTTTTGCAAACCCAAGTTGAAGGACAATC

TCAATTGGTTCAAGTCGATAGTAGTATCGTTAGAATCCTTCAAGACGAAGAAAATAACCA

ATTGTTCTGGACCACCACCTAATGGTGGAACACCGATAGCAGTGGTTTCGAAAACTCTGT

CATCGACTTCGTTACAAACTCTTTCAATCTCAATGGAAGAGATTTTGATACCACCGATGT

TCATGGTGTCATCAGCACGACCGTGAGCATGGTAGTAACCGTTGGAAGTTAATTCAAAGA

TGTCACCGTGTCTTCTCAAAACTTCACCGTTCAAAGTTGGCATACCTTTGAAGTAGACAT

CGTGGTGGTTACCGTTCAATAAAGTCTTAGAAGCACCGAACATAACTGGACCCAAAGCCA

ATTCACCAATACCTGGCTTGTTCTTTGGCATTGGGTAACCGTTCTTATCCAAAATGTACA

AAGTACAACCCATACATTGGGAAGAAAAGGAGGACAAGGATTGGGCTTGTAAGAAAGAAC

CAGCAGAGAAAGCACCACCGATTTCGGTACCACCACACATTTCGATAACAGGTTTATAGT

TGGCTCTACCCATCAACCACAAGTATTCATCGACGTTAGAAGCTTCACCAGAGGAAGAAA

AGCAACGGATGGTAGACCAGTCATAACCGGAAACGCAGTTGGTGGATTTCCAAGATCTAA

CAATAGATGGAACAACACCTAACATAGTAACCTTAGCGTCTTGGACGAACTTGGCGAAAC

CAGAAACCAATGGGGAACCATTATACAAAGCGATAGAAGCACCGTTCAATAAAGAGGCGT

AAACCAACCATGGACCCATCATCCAACCTAAATTAGTTGGCCAAACAATGACGTCACCTT

TACGAATATCCAAGTGAGACCAACCGTCGGCAGCAGCCTTCAATGGAGTAGCTTGGGTCC

ATGGAATGGCCTTTGGTTCACCAGTGGTACCGGAAGAGAATAAAATGTTGGTGTAGGCAT

CAACTGGTTGTTCACGAGCGGTGAATTCACAGTTCTTGAATTCCTTAGCACGTTCCAAGA

AATAATCCCAGGAAATGTCACCGTCACGCAATTCGGCACCGATGTTGGAACCGGAACATG

GAATGACAATAGCCATTGGAGACTTAGCTTCAACGACTCTAGAATACAATGGAATTCTCT

TCTTACCACGGATGATGTGGTCTTGAGTGAAGATGGCCTTAGCCTTAGACAATCTCAATC

TAGTAGAGATTTCTGGAGCGGAGAAAGAATCAGCGATGGAAACGACGACGTAACCAGCCA

AGACAATGGCTAAGTAGATGACGACAGCGTCAACGTGCATTGGCATATCGATGGCGATAG

CACAACCTTTTTCCAAACCCATTTCTTCCAAGGCATAACCAACCAACCAAACTCTCTTTC

TCAATTGGTCCAAAGTCAACTTGTTCAATGGCAAATCGTCGTTACCCTCATCACGCCAAA

CGATCATAGTATCATTCAACTTTTTGTTAGAGTTAACATTCAAGCAGTTCTTAGCAGAGT

TCAAGTAACCACCTGGCAACCATTCGGAACCACCTGGGTTGTTAATATCGTCTCTACGTA

AGATACATTCTGGATCTTTAGAAAAGGAGATCTTCATTTCATCCATTAAAACAGTTCTCC

AGTAAACTTCTGGGTTTCTGACGGAGAACTCTTGGAAGTGAGAAAAAGAAGAAATTGGAT

CCTTGTATTTAACACCCAAGAATTCCTTACCTCTTTTCTCCAACAAAGCACCCAAGTTGG

TAGACTTGACCTTTTCAGGGTCTGGAATCCAAGCTGGTGGGGCTGGACCAAAGTCCTTGT

AACAACCATAGAATAACATTTGGTGCAAGGAAAATGGCAAGTCTGGGGATAAGATATGGT

TGGCAATGTTAATCCAAGTTTGTGGGGTAGCAGCACCGTAATTACAAACAATTTCAGCTA

ATCTACCATGCAAAGTTTCGGCGACCTCAGAGGTAATACCCAAAGCGATGAAATCGGAAG

CAACAACAGAATCCAAAGATTTGTAGTTCTTACCCATTATAGTTTTTTCTCCTTGACGTT

AAAGTATAGAGGTATATTAACAATTTTTTGTTGATACTTTTATGACATTTGAATAAGAAG

TAATACAAACCGAAAATGTTGAAAGTATTAGTTAAAGTGGTTATGCAGCTTTTGCATTTA

TATATCTGTTAATAGATCAAAAATCATCGCTTCGCTGATTAATTACCCCAGAAATAAGGC

TAAAAAACTAATCGCATTATTATCCTATGGTTGTTAATTTGATTCGTTGATTTGAAGGTT

TGTGGGGCCAGGTTACTGCCAATTTTTCCTCTTCATAACCATAAAAGCTAGTATTGTAGA

ATCTTTATTGTTCGGAGCAGTGCGGCGCGAGGCACATCTGCGTTTCAGGAACGCGACCGG

TGAAGACCAGGACGCACGGAGGAGAGTCTTCCGTCGGAGGGCTGTCGCCCGCTCGGCGGC

TTCTAATCCGTACTTCAATATAGCAATGAGCAGTTAAGCGTATTACTGAAAGTTCCAAAG

AGAAGGTTTTTTTAGGCTAAGATAATGGGGCTCTTTACATTTCCACAACATATAAGTAAG

ATTAGATATGGATATGTATATGGTGGTATTGCCATGTAATATGATTATTAAACTTCTTTG

CGTCCATCCAAAAAAAAAGTAAGAATTTTTGAAAATTCAATATAAATGAACCACTTAAGA

GCTGAAGGTCCAGCTTCCGTTTTGGCCATTGGTACCGCTAACCCAGAAAACATCTTGTTG

CAAGACGAATTTCCAGACTACTACTTCAGAGTCACTAAGTCCGAACACATGACCCAATTG

AAGGAAAAGTTCAGAAAGATTTGTGATAAGTCTATGATCAGAAAAAGAAACTGTTTCTTG

AACGAAGAACACTTGAAACAAAACCCTAGATTAGTTGAACATGAAATGCAAACTTTAGAT

GCCAGACAAGATATGTTGGTCGTCGAAGTCCCAAAGTTGGGTAAGGACGCTTGTGCCAAG

GCTATCAAGGAATGGGGTCAACCAAAGTCTAAGATTACTCATTTGATCTTCACTTCCGCC

TCTACCACCGATATGCCAGGTGCTGATTACCATTGTGCTAAGTTGTTGGGTTTATCCCCA

TCTGTTAAAAGAGTTATGATGTACCAATTGGGTTGTTATGGTGGTGGTACTGTTTTGAGA

ATTGCCAAAGACATCGCTGAAAACAATAAGGGTGCTAGAGTTTTGGCTGTTTGTTGTGAT

ATTATGGCTTGTTTGTTCAGAGGTCCATCCGAGTCTGATTTAGAGTTGTTAGTTGGTCAA

GCTATTTTCGGTGACGGTGCTGCTGCTGTTATTGTTGGTGCTGAACCAGACGAATCTGTT

GGTGAACGTCCAATCTTTGAATTGGTCTCTACCGGTCAAACCATCTTGCCAAACTCTGAA

GGTACCATTGGTGGTCACATCAGAGAAGCTGGTTTGATCTTCGATTTGCATAAAGATGTT

CCTATGTTGATTTCTAATAACATCGAAAAGTGCTTAATCGAAGCTTTCACTCCAATCGGT

ATCTCTGATTGGAATTCCATTTTCTGGATTACCCATCCAGGTGGTAAGGCCATCTTGGAT

AAGGTTGAAGAAAAGTTGCATTTAAAGTCTGATAAGTTCGTTGACTCTCGTCACGTTTTG

TCTGAACATGGTAACATGTCTTCTTCCACTGTTTTGTTTGTTATGGATGAATTGAGAAAA

AGATCCTTGGAAGAAGGTAAGTCTACTACTGGTGATGGTTTTGAATGGGGTGTCTTGTTC

GGTTTTGGTCCAGGTTTGACCGTTGAAAGAGTTGTCGTTAGATCCGTTCCAATCAAGTAC

TAATTTGCCAGCTTACTATCCTTCTTGAAAATATGCACTCTATATCTTTTAGTTCTTAAT

TGCAACACATAGATTTGCTGTATAACGAATTTTATGCTATTTTTTAAATTTGGAGTTCAG

TGATAAAAGTGTCACAGCGAATTTCCTCACATGTAGGGACCGAATTGTTTACAAGTTCTC

TGTACCACCATGGAGACATCAAAGATTGAAAATCTATGGAAAGATATGGACGGTAGCAAC

AAGAATATAGCACGAGCCGCGAAGTTCATTTCGTTACTTTTGATATCGCTCACAACTATT

GCGAAGCGCTTCAGTGAAAAAATCATAAGGAAAAGTTGTAAATATTATTGGTAGTATTCG

TTTGGTAAAGTAGAGGGGGTAATTTTTCCCCTTTATTTTGTTCATACATTCTTAAATTGC

TTTGCCTCTCCTTTTGGAAAGCTAGGTCCGCCGGCGTTGGACGAGCGAAAATTCATTTAA

TATTCAATGAAGTTATAAATTGATAGTTTAAATAAAGTCAGTCTTTTTCCT CATGTTTAG

AATTGTATTAATGTACGCCGTTTACGTTGGAGTGTAAATGTGTCTATTCCAGAACGAAAT

CTAAATGAGCAGTACAGGCAGTTCAGATGGTACTGAAGCGGTACTAAAGATGCATGAATT

GAACAGATGTGGTAGTTATGTATATGAGGAATATGAGTTGTCACATTAAAAATATAATAG

CTATGATCCCATTATTATATTCGTGACAGTTCGTAACGTTTTAATTGGCTTATGTTTTTG

AGAAATGGGTGAATTTTAAGATAATTGTTGGGATTCCATTATTGATAAAGGCTATAATAT

TAGGTATACAGAATATACTGGAAGTTCTCCTCGAGGATATAGGAATCCTCAAAATGGAAT

CTATATTTCTATTTACTAATATCACGATTATTCTTCATTCCGTTTTATATGTTTCATTAT

CCTATTACATTATCAATCCTTGCATTTCAGCTTCCTCTAACTTCGATGACAGCTGGCGGT

TTAAACGCGTGGCCGTGCCGTC

MS101227:

GACGGCACGGCCACGCGTTTAAACCGCCAGAGTATGTCAACTGGCGCAGTAGATACATGT

TTTTCTCTTCCACGTCGAATTTTGTTATATACATAGCATAATCGAGTTGTATGCACCCTT

TTTGTTTATCTCGTTAGTAACTCGGGGTAGGAATAAGACATCCACAAAGGTGACAGAACA

AAATCATCCTAGCCTTGTTCATAATCTACCTCTATATAGCCGCTAAAAAATTAGTAGTAT

TTTGACTCTTTAAGAGCACATTTATTATCAGGCTGCTTTTACATACTTCTTTTGTTTAAA

ACATTTAAAGACGATCACTGCCCTTCCAAAGGACAAATATATATACACAAACACTAGGCC

AAAAGTTCACTTATAATAATTTAGTGGTAATTATGTTGGGTAAAGAAATTGCCAATAGTC

TTTTTTTTTCCGTATTGTAAGGTGAGACTGAGGTAGCGGCACAAAAAAACGACACATAAT

AGGATACTGAGTAAAGCAGTATTAAAATAAAAAGATATATTTTACCTCGAACGCTACAAA

TAAAGCAGAAAAGAACAAAATCGTGAGCCGCTCGTCCAACGCCGGCGGACCTAGCTTTCC

AAAAGGAGAGGCAAAGCAATTTAAGAATGTATGAACAAAATAAAGGGGAAAAATTACCCC

CTCTACTTTACCAAACGAATACTACCAATAATATTTACAACTTTTCCTTATGATTTTTTC

ACTGAAGCGCTTCGCAATAGTTGTGAGCGATATCAAAAGTAACGAAATGAACTTCGCGGC

TCGTGCTATATTCTTGTTGCTACCGTCCATATCTTTCCATAGATTTTCAATCTTTGATGT

CTCCATGGTGGTACAGAGAACTTGTAAACAATTCGGTCCCTACATGTGAGGAAATTCGCT

GTGACACTTTTATCACTGAACTCCAAATTTAAAAAATAGCATAAAATTCGTTATACAGCA

AATCTATGT GTTGCAATTAAGAACTAAAAGATATAGAGTGCATATTTTCAAGAAGGATAG

TAAGCTGGCAAATTAGTACTTGATTGGAACGGATCTAACGACAACTCTTTCAACGGTCAA

ACCTGGACCAAAACCGAACAAGACACCCCATTCAAAACCATCACCAGTAGTAGACTTACC

TTCTTCCAAGGATCTTTTTCTCAATTCATCCATAACAAACAAAACAGTGGAAGAAGACAT

GTTACCATGTTCAGACAAAACGTGACGAGAGTCAACGAACTTATCAGACTTTAAATGCAA

CTTTTCTTCAACCTTATCCAAGATGGCCTTACCACCTGGATGGGTAATCCAGAAAATGGA

ATTCCAATCAGAGATACCGATTGGAGTGAAAGCTTCGATTAAGCACTTTTCGATGTTATT

AGAAATCAACATAGGAACATCTTTATGCAAATCGAAGATCAAACCAGCTTCTCTGATGTG

ACCACCAATGGTACCTTCAGAGTTTGGCAAGATGGTTTGACCGGTAGAGACCAATTCAAA

GATTGGACGTTCACCAACAGATTCGTCTGGTTCAGCACCAACAATAACAGCAGCAGCACC

GTCACCGAAAATAGCTTGACCAACTAACAACTCTAAATCAGACTCGGATGGACCTCTGAA

CAAACAAGCCATAATATCACAACAAACAGCCAAAACTCTAGCACCCTTATTGTTTTCAGC

GATGTCTTTGGCAATTCTCAAAACAGTACCACCACCATAACAACCCAATTGGTACATCAT

AACTCTTTTAACAGATGGGGATAAACCCAACAACTTAGCACAATGGTAATCAGCACCTGG

CATATCGGTGGTAGAGGCGGAAGTGAAGATCAAATGAGTAATCTTAGACTTTGGTTGACC

CCATTCCTTGATAGCCTTGGCACAAGCGTCCTTACCCAACTTTGGGACTTCGACGACCAA

CATATCTTGTCTGGCATCTAAAGTTTGCATTTCATGTTCAACTAATCTAGGGTTTTGTTT

CAAGTGTTCTTCGTTCAAGAAACAGTTTCTTTTTCTGATCATAGACTTATCACAAATCTT

TCTGAACTTTTCCTTCAATTGGGTCATGTGTTCGGACTTAGTGACTCTGAAGTAGTAGTC

TGGAAATTCGTCTTGCAACAAGATGTTTTCTGGGTTAGCGGTACCAATGGCCAAAACGGA

AGCTGGACCTTCAGCTCTTAAGTGGTTCATTTATATTGAATTTTCAAAAATTCTTACTTT

TTTTTTGGATGGACGCAAAGAAGTTTAATAATCATATTACATGGCAATACCACCATATAC

ATATCCATATCTAATCTTACTTATATGTTGTGGAAATGTAAAGAGCCCCATTATCTTAGC

CTAAAAAAACCTTCTCTTTGGAACTTTCAGTAATACGCTTAACTGCTCATTGCTATATTG

AAGTACGGATTAGAAGCCGCCGAGCGGGCGACAGCCCTCCGACGGAAGACTCTCCTCCGT

GCGTCCTGGTCTTCACCGGTCGCGTTCCTGAAACGCAGATGTGCCTCGCGCCGCACTGCT

CCGAACAATAAAGATTCTACAATACTAGCTTTTATGGTTATGAAGAGGAAAAATTGGCAG

TAACCTGGCCCCACAAACCTTCAAATCAACGAATCAAATTAACAACCATAGGATAATAAT

GCGATTAGTTTTTTAGCCTTATTTCTGGGGTAATTAATCAGCGAAGCGATGATTTTTGAT

CTATTAACAGATATATAAATGCAAAAGCTGCATAACCACTTTAACTAATACTTTCAACAT

TTTCGGTTTGTATTACTTCTTATTCAAATGTCATAAAAGTATCAACAAAAAATTGTTAAT

ATACCTCTATACTTTAACGTCAAGGAGAAAAAACTATAATGAACCACTTAAGAGCTGAAG

GTCCAGCTTCCGTTTTGGCCATTGGTACCGCTAACCCAGAAAACATCTTGTTGCAAGACG

AATTTCCAGACTACTACTTCAGAGTCACTAAGTCCGAACACATGACCCAATTGAAGGAAA

AGTTCAGAAAGATTTGTGATAAGTCTATGATCAGAAAAAGAAACTGTTTCTTGAACGAAG

AACACTTGAAACAAAACCCTAGATTAGTTGAACATGAAATGCAAACTTTAGATGCCAGAC

AAGATATGTTGGTCGTCGAAGTCCCAAAGTTGGGTAAGGACGCTTGTGCCAAGGCTATCA

AGGAATGGGGTCAACCAAAGTCTAAGATTACTCATTTGATCTTCACTTCCGCCTCTACCA

CCGATATGCCAGGTGCTGATTACCATTGTGCTAAGTTGTTGGGTTTATCCCCATCTGTTA

AAAGAGTTATGATGTACCAATTGGGTTGTTATGGTGGTGGTACTGTTTTGAGAATTGCCA

AAGACATCGCTGAAAACAATAAGGGTGCTAGAGTTTTGGCTGTTTGTTGTGATATTATGG

CTTGTTTGTTCAGAGGTCCATCCGAGTCTGATTTAGAGTTGTTAGTTGGTCAAGCTATTT

TCGGTGACGGTGCTGCTGCTGTTATTGTTGGTGCTGAACCAGACGAATCTGTTGGTGAAC

GTCCAATCTTTGAATTGGTCTCTACCGGTCAAACCATCTTGCCAAACTCTGAAGGTACCA

TTGGTGGTCACATCAGAGAAGCTGGTTTGATCTTCGATTTGCATAAAGATGTTCCTATGT

TGATTTCTAATAACATCGAAAAGTGCTTAATCGAAGCTTTCACTCCAATCGGTATCTCTG

ATTGGAATTCCATTTTCTGGATTACCCATCCAGGTGGTAAGGCCATCTTGGATAAGGTTG

AAGAAAAGTTGCATTTAAAGTCTGATAAGTTCGTTGACTCTCGTCACGTTTTGTCTGAAC

ATGGTAACATGTCTTCTTCCACTGTTTTGTTTGTTATGGATGAATTGAGAAAAAGATCCT

TGGAAGAAGGTAAGTCTACTACTGGTGATGGTTTTGAATGGGGTGTCTTGTTCGGTTTTG

GTCCAGGTTTGACCGTTGAAAGAGTTGTCGTTAGATCCGTTCCAATCAAGTACTAAGTAT

ACTTCTTTTTTTTACTTTGTTCAGAACAACTTCTCATTTTTTTCTACTCATAACTTTAGC

ATCACAAAGTACACAATAATAACGAGTAGTAACACTTTATAGTTCATACATGCTTCAAC

TACTTAATAAATGATTGTATGATAATGTTTTCAATGTAAGAGATTTCGATTATCCACAAA

CTTTGAAACACAGGGACACAATTCTTGATATGCTTTCAACCGCTGCGTTTTGGATACCTA

TTCTTGACATAATATGACTACCATTTTGTTATTGTACGTGGGGCAGTTGACGTCTTATCA

TATGTCAAAGTCATTTGCGAAGTTCTTGGCAAGTTGCCAACTGACGAGATGCAGTAAAAA

GAGATTGCCGTCTTGAAACTTTTTGTCCTTTTTTTTTTCCGGGGACTCTACGAGAACCCT

TTGTCCTACTGATCCCCGCGTGCTTGGCCGGCCGTGATCATCTACCCATGCCGAAATTCG

GGCCGTTGGCCGGATTGCGCGTTGTCTTCTCCGGTATCGAAATCGCCGGACCGTTTGCCG

GGCAAATGTTCGCAGAATGGGGCGCGGAAGTTATCTGGATCGAGAACGTCGCCTGGGCCG

ACACCATTCGCGTTCAACCGAACTACCCGCAACTCTCCCGCCGCAATTTGCACGCGCTGT

CGTTAAATATTTTCAAAGATGAAGGCCGCGAAGCGTTTCTGAAATTAATGGAAACCACCG

ATATCTTCATCGAAGCCAGTAAAGGTCCGGCCTTTGCCCGTCGTGGCATTACCGATGAAG

TACTGTGGCAGCACAACCCGAAACTGGTTATCGCTCACCTGTCCGGTTTTGGTCAGTACG

GCACCGAGGAGTACACCAATCTTCCGGCCTATAACACTATCGCCCAGGCCTTTAGTGGTT

ACCTGATTCAGAACGGTGATGTTGACCAGCCAATGCCTGCCTTCCCGTATACCGCCGATT

ACTTTTCTGGCCTGACCGCCACCACGGCGGCGCTGGCAGCACTGCATAAAGTGCGTGAAA

CCGGTAAAGGCGAAAGTATCGACATCGCCATGTATGAAGTGATGCTGCGTATGGGCCAGT

ACTTCATGATGGATTACTTCAACGGCGGCGAAATGTGCCCGCGCATGAGCAAAGGTAAAG

ATCCCTACTACGCCGAGGTCCGCCGGCGTTGGACGAGCGACTTTAATGTCGTTCTCCCTT

TTTAAAGAGTAAATACATATTTAAAAAAGTGACTATGGCTATTGCTAAACGTGATAAAAA

TCAGAGCCTATAACACTCTCTGAAATAACGCTATGCAGGAATTTCCAGTTAAGTTCTTCT

TGGGGTGACTTCTTTACTCGGTATGATATGTGTTTTATATGCACAGTACGAGTCCATTAG

GGTAAATTAGTGGCCGAGAAACTTTTGCCGCCGAGCTTTTAAGTATCCTTTTGCCACTTC

TTATTTAGATAAAGACCTGGCAGTAGTAGTCGTAGAAGATAAGATAGACAGAGAATGAAT

ACTAATAAGATAGCACAAGACGAAGTCCAAGATAAGGTTTTGCAAAGAGCAGAACTAGCA

CATTCTGTATGGAACTTAAGGTTCAACCTCAGTAAAGTTGCCAAACGGATTCGCATGGAA

ACAAAGGTATTTCCAGAGATAAAGATAAATGACGCGCAATCACAGTTAGAGCGATCTAGG

TGTAGAATATTTAGCCCTGACCTGGAGGAAGAACATGTGCCCTTGATTCAAGGCGGCGGT

TTAAACGCGTGGCCGTGCCGTC

MS101226：

GACGGCACGGCCACGCGTTTAAACCGCCTACGCCATCATTAAAGACCTGGTCAACTATAA

AATAATACAATCAATACTTGCTTGAACGCTTGATTTTACTGATATTCTATCCAAAAGCAA

GTAGACCAGAAACTCTCAAGATGTTGCAAATACCGTTCGATGTTTTTGGTTTAGATTGTT

TTAATGTTGATGCTTTTTTACTTATTTTTGGAAGCGTCTTTTTAATTTAGTTTTATATTA

TAGGTATATGAATGTGTTTATGCCAATAAGGGTTTTTTTGTACAGTTATGTGATTATAAA

CAGTCTTTTGTCTAGTTTTTTTCACCAGTATCGGCCTCTATTTATAAAAAACGGAGCAGC

TTTCGGT GTCAGTAATTCTGAAAAAATTTGTGTCACTCTGATTGTAAATGAATTAATTTA

GCTAGATAGTTGCGAGCCCCAACGAGAAGATTGTCAGACAAAGACAACATTCAACAACCT

ACATCCGTTACTATTCGTTAACTCGAGGTACTTGAAACTTTTCAGTTAAGTCGCTCGTCC

AACGCCGGCGGACCTGCGAGTAAGCAACTCTGGCGCTGGCATGGCATAACCGGCGACGGC

AATGCGCAAGATGGGATGCTATGGGCAGAGAGCCGTACTTTACTGCTTATGGCACTACAG

CAACAGATGGTTACCCCACTAAGCCTGAAGCGAATCGCCATCAATTCTGCGCAGTGGCGA

GGAGATAAAAGCGCGGAAGTCATTCATCAACTGGCGACGCTACTCAAAGCAGGGTTAACG

CTTTCTGAAGGGCTGGCTCTGCTGGCGGAACAGCATCCCAGTAAGCAATGGCAAGCGTTG

CTGCAATCGCTGGCGCACGATCTCGAACAGGGCATTGCTTTTTCCAATGCCTTATTACCC

TGGTCAGAGGTATTTCCGCCGCTCTATCAGGCGATGATCCGCACGGGTGAACTGACCGGT

AAGCTGGATGAATGCTGCTTTGAACTGGCGCGTCAGCAAAAAGCCCAGCGTCAGTTGACC

GACAAAGTGAAATCAGCGTTACGTTATCCCATCATCATTTTAGCGATGGCAATCATGGTG

GTTGTGGCAATGCTGCATTTTGTTCTGCCGGAGTTTGCCGCTATCTATAAGACCTTCAAC

ACCCCACTACCGGCACTAACGCAGGGGATCATGACGCTGGCAGACTTTAGTGGCGAATGG

AGCTGGCTGCTGGTGTTGTTCGGCTTTCTGCTGGCGATAGCCAATAAGTTGCTGAACGGC

CGGCCAAGCACGCGGGGATCAGTAGGACAAAGGGTTCTCGTAGAGTCCCCGGAAAAAAAA

AAGGACAAAAAGTTTCAAGACGGCAATCTCTTTTTACTGCATCTCGTCAGTTGGCAACTT

GCCAAGAACTTCGCAAATGACTTTGACATATGATAAGACGTCAACTGCCCCACGTACAAT

AACAAAATGGTAGTCATATTATGTCAAGAATAGGTATCCAAAACGCAGCGGTTGAAAGCA

TATCAAGAATTGTGTCCCTGTGTTTCAAAGTTTGTGGATAATCGAAATCTCTTACATTGA

AAACATTATCATACAATCATTTATTAAGTAGTTGAAGCATGTATGAACTATAAAAGTGTT

ACTACTCGTTATTATTGTGTACTTTGTGATGCTAAAGTTATGAGTAGAAAAAAATGAGAA

GTTGTTCTGAACAAAGTAAAAAAAAAGAAGTATACTTAGTACTTGATTGGAACGGATCTAA

CGACAACTCTTTCAACGGTCAAACCTGGACCAAAACCGAACAAGACACCCCATTCAAAAC

CATCACCAGTAGTAGACTTACCTTCTTCCAAGGATCTTTTTCTCAATTCATCCATAACAA

ACAAAACAGTGGAAGAAGACATGTTACCATGTTCAGACAAAACGTGACGAGAGTCAACGA

ACTTATCAGACTTTAAATGCAACTTTTCTTCAACCTTATCCAAGATGGCCTTACCACCTG

GATGGGTAATCCAGAAAATGGAATTCCAATCAGAGATACCGATTGGAGTGAAAGCTTCGA

TTAAGCACTTTTCGATGTTATTAGAAATCAACATAGGAACATCTTTATGCAAATCGAAGA

TCAAACCAGCTTCTCTGATGTGACCACCAATGGTACCTTCAGAGTTTGGCAAGATGGTTT

GACCGGTAGAGACCAATTCAAAGATTGGACGTTCACCAACAGATTCGTCTGGTTCAGCAC

CAACAATAACAGCAGCAGCACCGTCACCGAAAATAGCTTGACCAACTAACAACTCTAAAT

CAGACTCGGATGGACCTCTGAACAAACAAGCCATAATATCACAACAAACAGCCAAAACTC

TAGCACCCTTATTGTTTTCAGCGATGTCTTTGGCAATTCTCAAAACAGTACCACCACCAT

AACAACCCAATTGGTACATCATAACTCTTTTAACAGATGGGGATAAACCCAACAACTTAG

CACAATGGTAATCAGCACCTGGCATATCGGTGGTAGAGGCGGAAGTGAAGATCAAATGAG

TAATCTTAGACTTTGGTTGACCCCATTCCTTGATAGCCTTGGCACAAGCGTCCTTACCCA

ACTTTGGGACTTCGACGACCAACATATCTTGTCTGGCATCTAAAGTTTGCATTTCATGTT

CAACTAATCTAGGGTTTTGTTTCAAGTGTTCTTCGTTCAAGAAACAGTTTCTTTTTCTGA

TCATAGACTTATCACAAATCTTTCTGAACTTTTCCTTCAATTGGGTCATGT GTTCGGACT

TAGTGACTCTGAAGTAGTAGTCTGGAAATTCGTCTTGCAACAAGATGTTTTCTGGGTTAG

CGGTACCAATGGCCAAAACGGAAGCTGGACCTTCAGCTCTTAAGTGGTTCATTATAGTTT

TTTCTCCTTGACGTTAAAGTATAGAGGTATATTAACAATTTTTTGTTGATACTTTTATGA

CATTTGAATAAGAAGTAATACAAACCGAAAATGTTGAAAGTATTAGTTAAAGTGGTTATG

CAGCTTTTGCATTTATATATCTGTTAATAGATCAAAAATCATCGCTTCGCTGATTAATTA

CCCCAGAAATAAGGCTAAAAAACTAATCGCATTATTATCCTATGGTTGTTAATTTGATTC

GTTGATTTGAAGGTTTGTGGGGCCAGGTTACTGCCAATTTTTCCTCTTCATAACCATAAA

AGCTAGTATTGTAGAATCTTTATTGTTCGGAGCAGTGCGGCGCGAGGCACATCTGCGTTT

CAGGAACGCGACCGGTGAAGACCAGGACGCACGGAGGAGAGTCTTCCGTCGGAGGGCTGT

CGCCCGCTCGGCGGCTTCTAATCCGTACTTCAATATAGCAATGAGCAGTTAAGCGTATTA

CTGAAAGTTCCAAAGAGAAGGTTTTTTTAGGCTAAGATAATGGGGCTCTTTACATTTCCA

CAACATATAAGTAAGATTAGATATGGATATGTATATGGTGGTATTGCCATGTAATATGAT

TATTAAACTTCTTTGCGTCCATCCAAAAAAAAAAGTAAGAATTTTTGAAAATTCAATATAA

ATGAACCACTTAAGAGCTGAAGGTCCAGCTTCCGTTTTGGCCATTGGTACCGCTAACCCA

GAAAACATCTTGTTGCAAGACGAATTTCCAGACTACTACTTCAGAGTCACTAAGTCCGAA

CACATGACCCAATTGAAGGAAAAGTTCAGAAAGATTTGTGATAAGTCTATGATCAGAAAA

AGAAACTGTTTCTTGAACGAAGAACACTTGAAACAAAACCCTAGATTAGTTGAACATGAA

ATGCAAACTTTAGATGCCAGACAAGATATGTTGGTCGTCGAAGTCCCAAAGTTGGGTAAG

GACGCTTGTGCCAAGGCTATCAAGGAATGGGGTCAACCAAAGTCTAAGATTACTCATTTG

ATCTTCACTTCCGCCTCTACCACCGATATGCCAGGTGCTGATTACCATTGTGCTAAGTTG

TTGGGTTTATCCCCATCTGTTAAAAGAGTTATGATGTACCAATTGGGTTGTTATGGTGGT

GGTACTGTTTTGAGAATTGCCAAAGACATCGCTGAAAACAATAAGGGTGCTAGAGTTTTG

GCTGTTTGTTGTGATATTATGGCTTGTTTGTTCAGAGGTCCATCCGAGTCTGATTTAGAG

TTGTTAGTTGGTCAAGCTATTTTCGGTGACGGTGCTGCTGCTGTTATTGTTGGTGCTGAA

CCAGACGAATCTGTTGGTGAACGTCCAATCTTTGAATTGGTCTCTACCGGTCAAACCATC

TTGCCAAACTCTGAAGGTACCATTGGTGGTCACATCAGAGAAGCTGGTTTGATCTTCGAT

TTGCATAAAGATGTTCCTAT GTTGATTTCTAATAACATCGAAAAGTGCTTAATCGAAGCT

TTCACTCCAATCGGTATCTCTGATTGGAATTCCATTTTCTGGATTACCCATCCAGGTGGT

AAGGCCATCTTGGATAAGGTTGAAGAAAAGTTGCATTTAAAGTCTGATAAGTTCGTTGAC

TCTCGTCACGTTTTGTCTGAACATGGTAACATGTCTTCTTCCACTGTTTTGTTTGTTATG

GATGAATTGAGAAAAAGATCCTTGGAAGAAGGTAAGTCTACTACTGGTGATGGTTTTGAA

TGGGGTGTCTTGTTCGGTTTTGGTCCAGGTTTGACCGTTGAAAGAGTTGTCGTTAGATCC

GTTCCAATCAAGTACTAATTTGCCAGCTTACTATCCTTCTTGAAAATATGCACTCTATAT

CTTTTAGTTCTTAATTGCAACACATAGATTTGCTGTATAACGAATTTTATGCTATTTTTT

AAATTTGGAGTTCAGTGATAAAAGTGTCACAGCGAATTTCCTCACATGTAGGGACCGAAT

TGTTTACAAGTTCTCTGTACCACCATGGAGACATCAAAGATTGAAAATCTATGGAAAGAT

ATGGACGGTAGCAACAAGAATATAGCACGAGCCGCGAAGTTCATTTCGTTACTTTTGATA

TCGCTCACAACTATTGCGAAGCGCTTCAGTGAAAAAATCATAAGGAAAAGTTGTAAATAT

TATTGGTAGTATTCGTTTGGTAAAGTAGAGGGGGTAATTTTTCCCCTTTATTTTGTTCAT

ACATTCTTAAATTGCTTTGCCTCTCCTTTTGGAAAGCTAGGTCCGCCGGCGTTGGACGAG

CGAAAATTCATTTAATATT CAATGAAGTTATAAATT GATAGTTTAAATAAAGTCAGTCTT

TTTCCTCATGTTTAGAATTGTATTAATGTACGCCGTTTACGTTGGAGTGTAAATGTGTCT

ATT CCAGAACGAAATCTAAATGAGCAGTACAGGCAGTTCAGATGGTACTGAAGCGGTACT

AAAGATGCATGAATTGAACAGATGTGGTAGTTATGTATATGAGGAATATGAGTTGTCACA

TTAAAAATATAATAGCTATGATCCCATTATTATATTCGTGACAGTTCGTAACGTTTTAAT

TGGCTTATGTTTTTGAGAAATGGGTGAATTTTAAGATAATTGTTGGGATTCCATTATTGA

TAAAGGCTATAATATTAGGTATACAGAATATACTGGAAGTTCTCCTCGAGGATATAGGAA

TCCTCAAAATGGAATCTATATTTCTATTTACTAATATCACGATTATTCTTCATTCCGTTT

TATATGTTTCATTATCCTATTACATTATCAATCCTTGCATTTCAGCTTCCTCTAACTTCG

ATGACAGCTGGCGGTTTAAACGCGTGGCCGTGCCGTC

MS96695:

GACGGCACGGCCACGCGTTTAAACCGCCAGAGTATGTCAACTGGCGCAGTAGATACATGT

TTTTCTCTTCCACGTCGAATTTTGTTATATACATAGCATAATCGAGTTGTATGCACCCTT

TTTGTTTATCTCGTTAGTAACTCGGGGTAGGAATAAGACATCCACAAAGGTGACAGAACA

AAATCATCCTAGCCTTGTTCATAATCTACCTCTATATAGCCGCTAAAAAATTAGTAGTAT

TTTGACTCTTTAAGAGCACATTTATTATCAGGCTGCTTTTACATACTTCTTTTGTTTAAA

ACATTTAAAGACGATCACTGCCCTTCCAAAGGACAAATATATATACACAAACACTAGGCC

AAAAGTTCACTTATAATAATTTAGTGGTAATTATGTTGGGTAAAGAAATTGCCAATAGTC

TTTTTTTTTCCGTATTGTA $A_{GGTGAGACTGAGGTAGCGGCACA}$ $AAAAAA_{CGACACATAAT}$

AGGATACTGAGTAAAGCAGTATTAAAATAAAAAGATATATTTTACCTCGAACGCTACAAA

TAAAGCAGAAAAGAACAAAATCGTGAGCCGCTCGTCCAACGCCGGCGGACCTAGCTTTCC

AAAAGGAGAGGCAAAGCAATTTAAGAATGTATGAACAAAATAAAGGGGAAAAATTACCCC

CTCTACTTTACCAAACGAATACTACCAATAATATTTACAACTTTTCCTTATGATTTTTTC

ACTGAAGCGCTTCGCAATAGTTGTGAGCGATATCAAAAGTAACGAAATGAACTTCGCGGC

TCGTGCTATATTCTTGTTGCTACCGTCCATATCTTTCCATAGATTTTCAATCTTTGATGT

CTCCATGGTGGTACAGAGAACTTGTAAACAATTCGGTCCCTACATGTGAGGAAATTCGCT

GTGACACTTTTATCACTGAACTCCAAATTTAAAAAATAGCATAAAATTCGTTATACAGCA

AATCTATGT GTTGCAATTAAGAACTAAAAGATATAGAGTGCATATTTTCAAGAAGGATAG

TAAGCTGGCAAATTATTCAAAATGGGAGAATTGTTGACGCAAAACTCTACGCATGATCTT

GTTGGTGGCAGTTCTAGGCAAAGAAGACAAAGGGACGACTCTAGTAACCTTAAACAATGG

ATTCAACTTCTTTTGCAAACCCAAGTTGAAGGACAATCTCAATTGGTTCAAGTCGATAGT

AGTATCGTTAGAATCCTTCAAGACGAAGAAAATAACCAATTGTTCTGGACCACCACCTAA

TGGTGGAACACCGATAGCAGTGGTTTCGAAAACTCTGTCATCGACTTCGTTACAAACTCT

TTCAATCTCAATGGAAGAGATTTTGAT$_A$CC$_A$CCGATGTTCATGGTGTCATC$_A$GCACGACC

GTGAGCATGGTAGTAACCGTTGGAAGTTAATTCAAAGATGTCACCGTGTCTTCTCAAAAC

TTCACCGTTCAAAGTTGGCATACCTTTGAAGTAGACATCGTGGTGGTTACCGTTCAATAA

AGTCTTAGAAGCACCGAACATAACTGGACCCAAAGCCAATTCACCAATACCTGGCTTGTT

CTTTGGCATTGGGTAACCGTTCTTATCCAAAATGTACAAAGTACAACCCATACATTGGGA

AGAAAAGGAGGACAAGGATTGGGCTTGTAAGAAAGAACCAGCAGAGAAAGCACCACCGAT

TTCGGTACCACCACACATTTCGATAACAGGTTTATAGTTGGCTCTACCCATCAACCACAA

GTATTCATCGACGTTAGAAGCTTCACCAGAGGAAGAAAAGCAACGGATGGTAGACCAGTC

ATAACCGGAAACGCAGTTGGTGGATTTCCAAGATCTAACAATAGATGGAACAACACCTAA

CATAGTAACCTTAGCGTCTTGGACGAACTTGGCGAAACCAGAAACCAATGGGGAACCATT

ATACAAAGCGATAGAAGCACCGTTCAATAAAGAGGCGTAAACCAACCATGGACCCATCAT

CCAACCTAAATTAGTTGGCCAAACAATGACGTCACCTTTACGAATATCCAAGTGAGACCA

ACCGTCGGCAGCAGCCTTCAATGGAGTAGCTTGGGTCCATGGAATGGCCTTTGGTTCACC

AGTGGTACCGGAAGAGAATAAAATGTTGGTGTAGGCATCAACTGGTTGTTCACGAGCGGT

GAATTCACAGTTCTTGAATTCCTTAGCACGTTCCAAGAAATAATCCCAGGAAATGTCACC

GTCACGCAATTCGGCACCGATGTTGGAACCGGAACATGGAATGACAATAGCCATTGGAGA

CTTAGCTTCAACGACTCTAGAATACAATGGAATTCTCTTCTTACCACGGATGATGTGGTC

TTGAGTGAAGATGGCCTTAGCCTTAGACAATCTCAATCTAGTAGAGATTTCTGGAGCGGA

GAAAGAATCAGCGATGGAAACGACGACGTAACCAGCCAAGACAATGGCTAAGTAGATGAC

GACAGCGTCAACGTGCATTGGCATATCGATGGCGATAGCACAACCTTTTTCCAAACCCAT

TTCTTCCAAGGCATAACCAACCAACCAAACTCTCTTTCTCAATTGGTCCAAAGTCAACTT

GTTCAATGGCAAATCGTCGTTACCCTCATCACGCCAAACGATCATAGTATCATTCAACTT

TTTGTTAGAGTTAACATTCAAGCAGTTCTTAGCAGAGTTCAAGTAACCACCTGGCAACCA

TTCGGAACCACCTGGGTTGTTAATATCGTCTCTACGTAAGATACATTCTGGATCTTTAGA

AAAGGAGATCTTCATTTCATCCATTAAAACAGTTCTCCAGTAAACTTCTGGGTTTCTGAC

GGAGAACTCTTGGAAGTGAGAAAAAGAAGAAATTGGATCCTTGTATTTAACACCCAAGAA

TTCCTTACCTCTTTTCTCCAACAAAGCACCCAAGTTGGTAGACTTGACCTTTTCAGGGTC

TGGAATCCAAGCTGGTGGGGCTGGACCAAAGTCCTTGTAACAACCATAGAATAACATTTG

GTGCAAGGAAAATGGCAAGTCTGGGGATAAGATATGGTTGGCAATGTTAATCCAAGTTTG

TGGGGTAGCAGCACCGTAATTACAAACAATTTCAGCTAATCTACCATGCAAAGTTTCGGC

GACCTCAGAGGTAATACCCAAAGCGATGAAATCGGAAGCAACAACAGAATCCAAAGATTT

GTAGTTCTTACCCATTTATATTGAATTTTCAAAAATTCTTACTTTTTTTTTGGATGGACG

CAAAGAAGTTTAATAATCATATTACATGGCAATACCACCATATACATATCCATATCTAAT

CTTACTTATATGTTGTGGAAATGTAAAGAGCCCCATTATCTTAGCCTAAAAAAACCTTCT

CTTTGGAACTTTCAGTAATACGCTTAACTGCTCATTGCTATATTGAAGTACGGATTAGAA

GCCGCCGAGCGGGCGACAGCCCTCCGACGGAAGACTCTCCTCCGTGCGTCCTGGTCTTCA

CCGGTCGCGTTCCTGAAACGCAGATGTGCCTCGCGCCGCACTGCTCCGAACAATAAAGAT

TCTACAATACTAGCTTTTAT GGTTATGAAGAGGAAAAATTGGCAGTAACCTGGCCCCACA

AACCTTCAAATCAACGAATCAAATTAACAACCATAGGATAATAATGCGATTAGTTTTTTA

GCCTTATTTCTGGGGTAATTAATCAGCGAAGCGATGATTTTTGATCTATTAACAGATATA

TAAATGCAAAAGCTGCATAACCACTTTAACTAATACTTTCAACATTTTCGGTTTGTATTA

CTTCTTATTCAAATGTCATAAAAGTATCAACAAAAAATTGTTAATATACCTCTATACTTT

AACGTCAAGGAGAAAAAACTATAATGAACCACTTAAGAGCTGAAGGTCCAGCTTCCGTTT

TGGCCATTGGTACCGCTAACCCAGAAAACATCTTGTTGCAAGACGAATTTCCAGACTACT

ACTTCAGAGTCACTAAGTCCGAACACATGACCCAATTGAAGGAAAAGTTCAGAAAGATTT

GTGATAAGTCTATGATCAGAAAAAGAAACTGTTTCTTGAACGAAGAACACTTGAAACAAA

ACCCTAGATTAGTTGAACATGAAATGCAAACTTTAGATGCCAGACAAGATATGTTGGTCG

TCGAAGTCCCAAAGTTGGGTAAGGACGCTTGTGCCAAGGCTATCAAGGAATGGGGTCAAC

CAAAGTCTAAGATTACTCATTTGATCTTCACTTCCGCCTCTACCACCGATATGCCAGGTG

CTGATTACCATTGTGCTAAGTTGTTGGGTTTATCCCCATCTGTTAAAAGAGTTATGATGT

ACCAATTGGGTTGTTATGGTGGTGGTACTGTTTTGAGAATTGCCAAAGACATCGCTGAAA

ACAATAAGGGTGCTAGAGTTTTGGCTGTTTGTTGTGATATTATGGCTTGTTTGTTCAGAG

GTCCATCCGAGTCTGATTTAGAGTTGTTAGTTGGTCAAGCTATTTTCGGTGACGGTGCTG

CTGCTGTTATTGTTGGTGCTGAACCAGACGAATCTGTTGGTGAACGTCCAATCTTTGAAT

TGGTCTCTACCGGTCAAACCATCTTGCCAAACTCTGAAGGTACCATTGGTGGTCACATCA

GAGAAGCTGGTTTGATCTTCGATTTGCATAAAGATGTTCCTATGTTGATTTCTAATAACA

TCGAAAAGTGCTTAATCGAAGCTTTCACTCCAATCGGTATCTCTGATTGGAATTCCATTT

TCTGGATTACCCATCCAGGTGGTAAGGCCATCTTGGATAAGGTTGAAGAAAAGTTGCATT

TAAAGTCTGATAAGTTCGTTGACTCTCGT CACGTTTTGTCTGAACATGGTAACATGTCTT

CTTCCACTGTTTTGTTTGTTATGGATGAATTGAGAAAAAGATCCTTGGAAGAAGGTAAGT

CTACTACTGGTGATGGTTTTGAATGGGGTGTCTTGTTCGGTTTTGGTCCAGGTTTGACCG

TTGAAAGAGTTGTCGTTAGATCCGTTCCAATCAAGTACTAAGTATACTTCTTTTTTTTAC

TTTGTTCAGAACAACTTCTCATTTTTTTCTACTCATAACTTTAGCATCACAAAGTACACA

ATAATAACGAGTAGTAACACTTTTATAGTTCATACATGCTTCAACTACTTAATAAATGAT

TGTATGATAATGTTTTCAAT GTAAGAGATTTCGATTATCCACAAACTTTGAAACACAGGG

ACACAATTCTTGATATGCTTTCAACCGCTGCGTTTTGGATACCTATTCTTGACATAATAT

GACTACCATTTTGTTATTGTACGTGGGGCAGTTGACGTCTTATCATATGTCAAAGTCATT

TGCGAAGTTCTTGGCAAGTTGCCAACTGACGAGATGCAGTAAAAAGAGATTGCCGTCTTG

AAACTTTTTGTCCTTTTTTTTTTCCGGGGACTCTACGAGAACCCTTTGTCCTACTGATCC

CCGCGTGCTTGGCCGGCCGTGATCATCTACCCATGCCGAAATTCGGGCCGTTGGCCGGAT

TGCGCGTTGTCTTCTCCGGTATCGAAATCGCCGGACCGTTTGCCGGGCAAATGTTCGCAG

AATGGGGCGCGGAAGTTATCTGGATCGAGAACGTCGCCTGGGCCGACACCATTCGCGTTC

AACCGAACTACCCGCAACTCTCCCGCCGCAATTTGCACGCGCTGTCGTTAAATATTTTCA

AAGATGAAGGCCGCGAAGCGTTTCTGAAATTAATGGAAACCACCGATATCTTCATCGAAG

CCAGTAAAGGTCCGGCCTTTGCCCGTCGTGGCATTACCGATGAAGTACTGTGGCAGCACA

ACCCGAAACTGGTTATCGCTCACCTGTCCGGTTTTGGTCAGTACGGCACCGAGGAGTACA

CCAATCTTCCGGCCTATAACACTATCGCCCAGGCCTTTAGTGGTTACCTGATTCAGAACG

GTGATGTTGACCAGCCAATGCCTGCCTTCCCGTATACCGCCGATTACTTTTCTGGCCTGA

CCGCCACCACGGCGGCGCTGGCAGCACTGCATAAAGTGCGTGAAACCGGTAAAGGCGAAA

GTATCGACATCGCCATGTATGAAGTGATGCTGCGTATGGGCCAGTACTTCATGATGGATT

ACTTCAACGGCGGCGAAATGTGCCCGCGCATGAGCAAAGGTAAAGATCCCTACTACGCCG

AGGTCCGCCGGCGTTGGACGAGCGACTTTAATGTCGTTCTCCCTTTTTAAAGAGTAAATA

CATATTTAAAAAAGTGACTATGGCTATTGCTAAACGTGATAAAAATCAGAGCCTATAACA

CTCTCTGAAATAACGCTATGCAGGAATTTCCAGTTAAGTTCTTCTTGGGGTGACTTCTTT

ACTCGGTATGATATGTGTTTTATATGCACAGTACGAGTCCATTAGGGTAAATTAGTGGCC

GAGAAACTTTTGCCGCCGAGCTTTTAAGTATCCTTTTGCCACTTCTTATTTAGATAAAGA

CCTGGCAGTAGTAGTCGTAGAAGATAAGATAGACAGAGAATGAATACTAATAAGATAGCA

CAAGACGAAGTCCAAGATAAGGTTTTTGCAAAGAGCAGAACTAGCACATTCTGTATGGAAC

TTAAGGTTCAACCTCAGTAAAGTTGCCAAACGGATTCGCATGGAAACAAAGGTATTTCCA

GAGATAAAGATAAATGACGCGCAATCACAGTTAGAGCGATCTAGGTGTAGAATATTTAGC

CCTGACCTGGAGGAAGAACATGTGCCCTTGATTCAAGGCGGCGGTTTAAACGCGTGGCCG

TGCCGTC

MS101224 :

GACGGCACGGCCACGCGTTTAAACCGCCAGAGTATGTCAACTGGCGCAGTAGATACATGT

TTTTCTCTTCCACGTCGAATTTTGTTATATACATAGCATAATCGAGTTGTATGCACCCTT

TTTGTTTATCTCGTTAGTAACTCGGGGTAGGAATAAGACATCCACAAAGGTGACAGAACA

AAATCATCCTAGCCTTGTTCATAATCTACCTCTATATAGCCGCTAAAAAATTAGTAGTAT

TTTGACTCTTTAAGAGCACATTTATTATCAGGCTGCTTTTACATACTTCTTTTGTTTAAA

ACATTTAAAGACGATCACTGCCCTTCCAAAGGACAAATATATATACACAAACACTAGGCC

AAAAGTTCACTTATAATAATTTAGTGGTAATTATGTTGGGTAAAGAAATTGCCAATAGTC

TTTTTTTTTCCGTATTGTA A$_{GGTGAGACTGAGGTAGCGGCACA}$ AAAAAA$_{CGACACATAAT}$

AGGATACTGAGTAAAGCAGTATTAAAATAAAAAGATATATTTTACCTCGAACGCTACAAA

TAAAGCAGAAAAGAACAAAATCGTGAGCCGCTCGTCCAACGCCGGCGGACCTAGCTTTCC

AAAAGGAGAGGCAAAGCAATTTAAGAATGTATGAACAAAATAAAGGGGAAAAATTACCCC

CTCTACTTTACCAAACGAATACTACCAATAATATTTACAACTTTTCCTTATGATTTTTTC

ACTGAAGCGCTTCGCAATAGTTGTGAGCGATATCAAAAGTAACGAAATGAACTTCGCGGC

TCGTGCTATATTCTTGTTGCTACCGTCCATATCTTTCCATAGATTTTCAATCTTTGATGT

CTCCATGGTGGTACAGAGAACTTGTAAACAATTCGGTCCCTACATGTGAGGAAATTCGCT

GTGACACTTTTATCACTGAACTCCAAATTTAAAAAATAGCATAAAATTCGTTATACAGCA

AATCTATGT GTTGCAATTAAGAACTAAAAGATATAGAGTGCATATTTTCAAGAAGGATAG

TAAGCTGGCAAATTATTCAAAATGGGAGAATTGTTGACGCAAAACTCTACGCATGATCTT

GTTGGTGGCAGTTCTAGGCAAAGAAGACAAAGGGACGACTCTAGTAACCTTAAACAATGG

ATTCAACTTCTTTTGCAAACCCAAGTTGAAGGACAATCTCAATTGGTTCAAGTCGATAGT

AGTATCGTTAGAATCCTTCAAGACGAAGAAAATAACCAATTGTTCTGGACCACCACCTAA

TGGTGGAACACCGATAGCAGTGGTTTCGAAAACTCTGTCATCGACTTCGTTACAAACTCT

TTCAATCTCAATGGAAGAGATTTTGATACCACCGATGTTCATGGTGTCATCAGCACGACC

GTGAGCATGGTAGTAACCGTTGGAAGTTAATTCAAAGATGTCACCGTGTCTTCTCAAAAC

TTCACCGTTCAAAGTTGGCATACCTTTGAAGTAGACATCGTGGTGGTTACCGTTCAATAA

AGTCTTAGAAGCACCGAACATAACTGGACCCAAAGCCAATTCACCAATACCTGGCTTGTT

CTTTGGCATTGGGTAACCGTTCTTATCCAAAATGTACAAAGTACAACCCATACATTGGGA

AGAAAAGGAGGACAAGGATTGGGCTTGTAAGAAAGAACCAGCAGAGAAAGCACCACCGAT

TTCGGTACCACCACACATTTCGATAACAGGTTTATAGTTGGCTCTACCCATCAACCACAA

GTATTCATCGACGTTAGAAGCTTCACCAGAGGAAGAAAAGCAACGGATGGTAGACCAGTC

ATAACCGGAAACGCAGTTGGTGGATTTCCAAGATCTAACAATAGATGGAACAACACCTAA

CATAGTAACCTTAGCGTCTTGGACGAACTTGGCGAAACCAGAAACCAATGGGGAACCATT

ATACAAAGCGATAGAAGCACCGTTCAATAAAGAGGCGTAAACCAACCATGGACCCATCAT

CCAACCTAAATTAGTTGGCCAAACAATGACGTCACCTTTACGAATATCCAAGTGAGACCA

ACCGTCGGCAGCAGCCTTCAATGGAGTAGCTTGGGTCCATGGAATGGCCTTTGGTTCACC

AGTGGTACCGGAAGAGAATAAAATGTTGGTGTAGGCATCAACTGGTTGTTCACGAGCGGT

GAATTCACAGTTCTTGAATTCCTTAGCACGTTCCAAGAAATAATCCCAGGAAATGTCACC

GTCACGCAATTCGGCACCGATGTTGGAACCGGAACATGGAATGACAATAGCCATTGGAGA

CTTAGCTTCAACGACTCTAGAATACAATGGAATTCTCTTCTTACCACGGATGATGTGGTC

TTGAGTGAAGATGGCCTTAGCCTTAGACAATCTCAATCTAGTAGAGATTTCTGGAGCGGA

GAAAGAATCAGCGATGGAAACGACGACGTAACCAGCCAAGACAATGGCTAAGTAGATGAC

GACAGCGTCAACGTGCATTGGCATATCGATGGCGATAGCACAACCTTTTTCCAAACCCAT

TTCTTCCAAGGCATAACCAACCAACCAAACTCTCTTTCTCAATTGGTCCAAAGTCAACTT

GTTCAATGGCAAATCGTCGTTACCCTCATCACGCCAAACGATCATAGTATCATTCAACTT

TTTGTTAGAGTTAACATTCAAGCAGTTCTTAGCAGAGTTCAAGTAACCACCTGGCAACCA

TTCGGAACCACCTGGGTTGTTAATATCGTCTCTACGTAAGATACATTCTGGATCTTTAGA

AAAGGAGATCTTCATTTCATCCATTAAAACAGTTCTCCAGTAAACTTCTGGGTTTCTGAC

GGAGAACTCTTGGAAGTGAGAAAAAGAAGAAATTGGATCCTTGTATTTAACACCCAAGAA

TTCCTTACCTCTTTTCTCCAACAAAGCACCCAAGTTGGTAGACTTGACCTTTTCAGGGTC

TGGAATCCAAGCTGGTGGGGCTGGACCAAAGTCCTTGTAACAACCATAGAATAACATTTG

GTGCAAGGAAAATGGCAAGTCTGGGGATAAGATATGGTTGGCAATGTTAATCCAAGTTTG

TGGGGTAGCAGCACCGTAATTACAAACAATTTCAGCTAATCTACCATGCAAAGTTTCGGC

GACCTCAGAGGTAATACCCAAAGCGATGAAATCGGAAGCAACAACAGAATCCAAAGATTT

GTAGTTCTTACCCATTTATATTGAATTTTCAAAAATTCTTACTTTTTTTTTGGATGGACG

CAAAGAAGTTTAATAATCATATTACATGGCAATACCACCATATACATATCCATATCTAAT

CTTACTTATATGTTGTGGAAATGTAAAGAGCCCCATTATCTTAGCCTAAAAAAACCTTCT

CTTTGGAACTTTCAGTAATACGCTTAACTGCTCATTGCTATATTGAAGTACGGATTAGAA

GCCGCCGAGCGGGCGACAGCCCTCCGACGGAAGACTCTCCTCCGTGCGTCCTGGTCTTCA

CCGGTCGCGTTCCTGAAACGCAGATGTGCCTCGCGCCGCACTGCTCCGAACAATAAAGAT

TCTACAATACTAGCTTTTAT GGTTATGAAGAGGAAAAATTGGCAGTAACCTGGCCCCACA

AACCTTCAAATCAACGAATCAAATTAACAACCATATAGGATAATAATGCGATTAGTTTTTTA

GCCTTATTTCTGGGGTAATTAATCAGCGAAGCGATGATTTTTGATCTATTAACAGATATA

TAAATGCAAAAGCTGCATAACCACTTTAACTAATACTTTCAACATTTTCGGTTTGTATTA

CTTCTTATTCAAATGTCATAAAAGTATCAACAAAAAATTGTTAATATACCTCTATACTTT

AACGTCAAGGAGAAAAAACTATAATGAACCACTTAAGAGCTGAAGGTCCAGCTTCCGTTT

TGGCCATTGGTACCGCTAACCCAGAAAACATCTTGTTGCAAGACGAATTTCCAGACTACT

ACTTCAGAGTCACTAAGTCCGAACACATGACCCAATTGAAGGAAAAGTTCAGAAAGATTT

GTGATAAGTCTATGATCAGAAAAAGAAACTGTTTCTTGAACGAAGAACACTTGAAACAAA

ACCCTAGATTAGTTGAACATGAAATGCAAACTTTAGATGCCAGACAAGATATGTTGGTCG

TCGAAGTCCCAAAGTTGGGTAAGGACGCTTGTGCCAAGGCTATCAAGGAATGGGGTCAAC

CAAAGTCTAAGATTACTCATTTGATCTTCACTTCCGCCTCTACCACCGATATGCCAGGTG

CTGATTACCATTGTGCTAAGTTGTTGGGTTTATCCCCATCTGTTAAAAGAGTTATGATGT

ACCAATTGGGTTGTTATGGTGGTGGTACTGTTTTGAGAATTGCCAAAGACATCGCTGAAA

ACAATAAGGGTGCTAGAGTTTTGGCTGTTTGTTGTGATATTATGGCTTGTTTGTTCAGAG

GTCCATCCGAGTCTGATTTAGAGTTGTTAGTTGGTCAAGCTATTTTCGGTGACGGTGCTG

CTGCTGTTATTGTTGGTGCTGAACCAGACGAATCTGTTGGTGAACGTCCAATCTTTGAAT

TGGTCTCTACCGGTCAAACCATCTTGCCAAACTCTGAAGGTACCATTGGTGGTCACATCA

GAGAAGCTGGTTTGATCTTCGATTTGCATAAAGATGTTCCTATGTTGATTTCTAATAACA

TCGAAAAGTGCTTAATCGAAGCTTTCACTCCAATCGGTATCTCTGATTGGAATTCCATTT

TCTGGATTACCCATCCAGGTGGTAAGGCCATCTTGGATAAGGTTGAAGAAAAGTTGCATT

TAAAGTCTGATAAGTTCGTTGACTCTCGT CACGTTTTGTCTGAACATGGTAACATGTCTT

CTTCCACTGTTTTGTTTGTTATGGATGAATTGAGAAAAAGATCCTTGGAAGAAGGTAAGT

CTACTACTGGTGATGGTTTTGAATGGGGTGTCTTGTTCGGTTTTGGTCCAGGTTTGACCG

TTGAAAGAGTTGTCGTTAGATCCGTTCCAATCAAGTACTAAGTATACTTCTTTTTTTTAC

TTTGTTCAGAACAACTTCTCATTTTTTTCTACTCATAACTTTAGCATCACAAAGTACACA

ATAATAACGAGTAGTAACACTTTTATAGTTCATACATGCTTCAACTACTTAATAAATGAT

TGTATGATAATGTTTTCAAT GTAAGAGATTTCGATTATCCACAAACTTTGAAACACAGGG

ACACAATTCTTGATATGCTTTCAACCGCTGCGTTTTGGATACCTATTCTTGACATAATAT

GACTACCATTTTGTTATTGTACGTGGGGCAGTTGACGTCTTATCATATGTCAAAGTCATT

TGCGAAGTTCTTGGCAAGTTGCCAACTGACGAGATGCAGTAAAAAGAGATTGCCGTCTTG

AAACTTTTTGTCCTTTTTTTTTTCCGGGGACTCTACGAGAACCCTTTGTCCTACTGATCC

CCGCGTGCTTGGCCGGCCGTGATCATCTACCCATGCCGAAATTCGGGCCGTTGGCCGGAT

TGCGCGTTGTCTTCTCCGGTATCGAAATCGCCGGACCGTTTGCCGGGCAAATGTTCGCAG

AATGGGGCGCGGAAGTTATCTGGATCGAGAACGTCGCCTGGGCCGACACCATTCGCGTTC

AACCGAACTACCCGCAACTCTCCCGCCGCAATTTGCACGCGCTGTCGTTAAATATTTTCA

AAGATGAAGGCCGCGAAGCGTTTCTGAAATTAATGGAAACCACCGATATCTTCATCGAAG

CCAGTAAAGGTCCGGCCTTTGCCCGTCGTGGCATTACCGATGAAGTACTGTGGCAGCACA

ACCCGAAACTGGTTATCGCTCACCTGTCCGGTTTTGGTCAGTACGGCACCGAGGAGTACA

CCAATCTTCCGGCCTATAACACTATCGCCCAGGCCTTTAGTGGTTACCTGATTCAGAACG

GTGATGTTGACCAGCCAATGCCTGCCTTCCCGTATACCGCCGATTACTTTTCTGGCCTGA

CCGCCACCACGGCGGCGCTGGCAGCACTGCATAAAGTGCGTGAAACCGGTAAAGGCGAAA

GTATCGACATCGCCATGTATGAAGTGATGCTGCGTATGGGCCAGTACTTCATGATGGATT

ACTTCAACGGCGGCGAAATGTGCCCGCGCATGAGCAAAGGTAAAGATCCCTACTACGCCG

ACGGCCGGCCAAGCACGCGGGGATCAGTAGGACAAAGGGTTCTCGTAGAGTCCCCGGAAA

AAAAAAAGGACAAAAAGTTTCAAGACGGCAATCTCTTTTTACTGCATCTCGTCAGTTGGC

AACTTGCCAAGAACTTCGCAAATGACTTTGACATATGATAAGACGTCAACTGCCCCACGT

ACAATAACAAAATGGTAGTCATATTATGTCAAGAATAGGTATCCAAAACGCAGCGGTTGA

AAGCATATCAAGAATTGTGTCCCTGT GTTTCAAAGTTTGTGGATAATCGAAATCTCTTAC

ATT GAAAACATTATCATACAATCATTTATTAAGTAGTT GAAGCATGTATGAACTATAAAA

GTGTTACTACTCGTTATTATTGTGTACTTTGTGATGCTAAAGTTATGAGTAGAAAAAAAT

GAGAAGTTGTTCTGAACAAAGTAAAAAAAAGAAGTATACTTACTTTCTAGGGGT GTAATC

AAAGATCAACAACTTTTCCCAGAAAGATCTGTAAACGTCACCGAAACCAACATGAGCTGG

GTGAATAATGTAGTCTTGGATAGTTTCAACAGATTCGAAGGTGACTTCAACAATATGAGT

GTAACCTTCTTCTTTGTTCTTTTGGGTGACGTCTTTACCCCAGTAGACATCCTTCATAGC

TGGAATAATGTTAACCAAGTTAACGTAAGTTTTGAAGAATTCCTCTTTTTGGGCTTCGGT

AATTTCGTCTTTGAACTTTAAGACGATCAAGT GTTTAACAGCCATTATAGTTTTTTCTCC

TTGACGTTAAAGTATAGAGGTATATTAACAATTTTTTGTTGATACTTTTATGACATTTGA

ATAAGAAGTAATACAAACCGAAAATGTTGAAAGTATTAGTTAAAGTGGTTATGCAGCTTT

TGCATTTATATATCTGTTAATAGATCAAAAATCATCGCTTCGCTGATTAATTACCCCAGA

AATAAGGCTAAAAAACTAATCGCATTATTATCCTATGGTTGTTAATTTGATTCGTTGATT

TGAAGGTTTGTGGGGCCAGGTTACTGCCAATTTTTCCTCTTCATAACCATAAAAGCTAGT

ATTGTAGAATCTTTATTGTTCGGAGCAGTGCGGCGCGAGGCACATCTGCGTTTCAGGAAC

GCGACCGGTGAAGACCAGGACGCACGGAGGAGAGTCTTCCGTCGGAGGGCTGTCGCCCGC

TCGGCGGCTTCTAATCCGTACTTCAATATAGCAATGAGCAGTTAAGCGTATTACTGAAAG

TTCCAAAGAGAAGGTTTTTTTAGGCTAAGATAATGGGGCTCTTTACATTTCCACAACATA

TAAGTAAGATTAGATATGGATATGTATATGGTGGTATTGCCATGTAATATGATTATTAAA

CTTCTTTGCGTCCATCCAAAAAAAAAAGTAAGAATTTTTGAAAATTCAATATAAATGGCTG

TTAAACACTTGATCGTCTTAAAGTTCAAAGACGAAATTACCGAAGCCCAAAAAGAGGAAT

TCTTCAAAACTTACGTTAACTTGGTTAACATTATTCCAGCTATGAAGGATGTCTACTGGG

GTAAAGACGTCACCCAAAAGAACAAAGAAGAAGGTTACACTCATATTGTTGAAGTCACCT

TCGAATCTGTTGAAACTATCCAAGACTACATTATTCACCCAGCTCATGTTGGTTTCGGTG

ACGTTTACAGATCTTTCTGGGAAAAGTTGTTGATCTTTGATTACACCCCTAGAAAGTAAT

TTGCCAGCTTACTATCCTTCTTGAAAATATGCACTCTATATCTTTTAGTTCTTAATTGCA

ACACATAGATTTGCTGTATAACGAATTTTATGCTATTTTTTAAATTTGGAGTTCAGTGAT

AAAAGTGTCACAGCGAATTTCCTCACATGTAGGGACCGAATTGTTTACAAGTTCTCTGTA

CCACCATGGAGACATCAAAGATTGAAAATCTATGGAAAGATATGGACGGTAGCAACAAGA

ATATAGCACGAGCCGCGAAGTTCATTTCGTTACTTTTGATATCGCTCACAACTATTGCGA

AGCGCTTCAGTGAAAAAATCATAAGGAAAAGTTGTAAATATTATTGGTAGTATTCGTTTG

GTAAAGTAGAGGGGGTAATTTTTCCCCTTTATTTTGTTCATACATTCTTAAATTGCTTTG

CCTCTCCTTTTGGAAAGCTAGGTCCGCCGGCGTTGGACGAGCGACTTTAATGTCGTTCTC

CCTTTTTAAAGAGTAAATACATATTTAAAAAAGTGACTATGGCTATTGCTAAACGTGATA

AAAATCAGAGCCTATAACACTCTCTGAAATAACGCTATGCAGGAATTTCCAGTTAAGTTC

TTCTTGGGGTGACTTCTTTACTCGGTATGATATGTGTTTTATATGCACAGTACGAGTCCA

TTAGGGTAAATTAGTGGCCGAGAAACTTTTGCCGCCGAGCTTTTAAGTATCCTTTTGCCA

CTTCTTATTTAGATAAAGACCTGGCAGTAGTAGTCGTAGAAGATAAGATAGACAGAGAAT

GAATACTAATAAGATAGCACAAGACGAAGTCCAAGATAAGGTTTTGCAAAGAGCAGAACT

AGCACATTCTGTATGGAACTTAAGGTTCAACCTCAGTAAAGTTGCCAAACGGATTCGCAT

GGAAACAAAGGTATTTCCAGAGATAAAGATAAATGACGCGCAATCACAGTTAGAGCGATC

TAGGTGTAGAATATTTAGCCCTGACCTGGAGGAAGAACATGTGCCCTTGATTCAAGGCGG
CGGTTTAAACGCGTGGCCGTGCCGTC

MS101229 :

GACGGCACGGCCACGCGTTTAAACCGCCTACGCCATCATTAAAGACCTGGTCAACTATAA
AATAATACAATCAATACTTGCTTGAACGCTTGATTTTACTGATATTCTATCCAAAAGCAA
GTAGACCAGAAACTCTCAAGATGTTGCAAATACCGTTCGATGTTTTTGGTTTAGATTGTT
TTAATGTTGATGCTTTTTTACTTATTTTTGGAAGCGTCTTTTTAATTTAGTTTTATATTA
TAGGTATATGAATGTGTTTATGCCAATAAGGGTTTTTTTGTACAGTTATGTGATTATAAA
CAGTCTTTTGTCTAGTTTTTTTCACCAGTATCGGCCTCTATTTATAAAAAACGGAGCAGC
TTTCGGTGTCAGTAATTCTGAAAAAATTTGTGTCACTCTGATTGTAAATGAATTAATTTA
GCTAGATAGTTGCGAGCCCCAACGAGAAGATTGTCAGACAAAGACAACATTCAACAACCT
ACATCCGTTACTATTCGTTAACTCGAGGTACTTGAAACTTTTCAGTTAAGTCGCTCGTCC
AACGCCGGCGGACCTAGCTTTCCAAAAGGAGAGGCAAAGCAATTTAAGAATGTATGAACA
AAATAAAGGGGAAAAATTACCCCCTCTACTTTACCAAACGAATACTACCAATAATATTTA
CAACTTTTCCTTATGATTTTTTCACTGAAGCGCTTCGCAATAGTTGTGAGCGATATCAAA
AGTAACGAAATGAACTTCGCGGCTCGTGCTATATTCTTGTTGCTACCGTCCATATCTTTC
CATAGATTTTCAATCTTTGATGTCTCCATGGTGGTACAGAGAACTTGTAAACAATTCGGT
CCCTACATGTGAGGAAATTCGCTGTGACACTTTTATCACTGAACTCCAAATTTAAAAAAT
AGCATAAAATTCGTTATACAGCAAATCTATGTGTTGCAATTAAGAACTAAAAGATATAGA
GTGCATATTTTCAAGAAGGATAGTAAGCTGGCAAATTACTTTCTAGGGGTGTAATCAAAG
ATCAACAACTTTTCCCAGAAAGATCTGTAAACGTCACCGAAACCAACATGAGCTGGGTGA
ATAATGTAGTCTTGGATAGTTTCAACAGATTCGAAGGTGACTTCAACAATATGAGTGTAA
CCTTCTTCTTTGTTCTTTTGGGTGACGTCTTTACCCCAGTAGACATCCTTCATAGCTGGA
ATAATGTTAACCAAGTTAACGTAAGTTTTGAAGAATTCCTCTTTTTGGGCTTCGGTAATT
TCGTCTTTGAACTTTAAGACGATCAAGTGTTTAACAGCCATTTATATTGAATTTTCAAAA
ATTCTTACTTTTTTTTTTGGATGGACGCAAAGAAGTTTAATAATCATATTACATGGCAATA
CCACCATATACATATCCATATCTAATCTTACTTATATGTTGTGGAAATGTAAAGAGCCCC
ATTATCTTAGCCTAAAAAAACCTTCTCTTTGGAACTTTCAGTAATACGCTTAACTGCTCA
TTGCTATATTGAAGTACGGATTAGAAGCCGCCGAGCGGGCGACAGCCCTCCGACGGAAGA
CTCTCCTCCGTGCGTCCTGGTCTTCACCGGTCGCGTTCCTGAAACGCAGATGTGCCTCGC
GCCGCACTGCTCCGAACAATAAAGATTCTACAATACTAGCTTTTATGGTTATGAAGAGGA
AAAATTGGCAGTAACCTGGCCCCACAAACCTTCAAATCAACGAATCAAATTAACAACCAT
AGGATAATAATGCGATTAGTTTTTTAGCCTTATTTCTGGGGTAATTAATCAGCGAAGCGA

TGATTTTTGATCTATTAACAGATATATAAATGCAAAAGCTGCATAACCACTTTAACTAAT

ACTTTCAACATTTTCGGTTTGTATTACTTCTTATTCAAATGTCATAAAAGTATCAACAAA

AAATTGTTAATATACCTCTATACTTTAACGTCAAGGAGAAAAAACTATAATGGCTGTTAA

ACACTTGATCGTCTTAAAGTTCAAAGACGAAATTACCGAAGCCCAAAAAGAGGAATTCTT

CAAAACTTACGTTAACTTGGTTAACATTATTCCAGCTATGAAGGATGTCTACTGGGGTAA

AGACGTCACCCAAAAGAACAAAGAAGAAGGTTACACTCATATTGTTGAAGTCACCTTCGA

ATCTGTTGAAACTATCCAAGACTACATTATTCACCCAGCTCATGTTGGTTTCGGTGACGT

TTACAGATCTTTCTGGGAAAAGTTGTTGATCTTTGATTACACCCCTAGAAAGTAAGTATA

CTTCTTTTTTTTTACTTTGTTCAGAACAACTTCTCATTTTTTTCTACTCATAACTTTAGCA

TCACAAAGTACACAATAATAACGAGTAGTAACACTTTTATAGTTCATACATGCTTCAACT

ACTTAATAAATGATTGTATGATAATGTTTTCAATGTAAGAGATTTCGATTATCCACAAAC

TTTGAAACACAGGGACACAATTCTTGATATGCTTTCAACCGCTGCGTTTTGGATACCTAT

TCTTGACATAATATGACTACCATTTTGTTATTGTACGTGGGGCAGTTGACGTCTTATCAT

ATGTCAAAGTCATTTGCGAAGTTCTTGGCAAGTTGCCAACTGACGAGATGCAGTAAAAAG

AGATTGCCGTCTTGAAACTTTTTGTCCTTTTTTTTTTCCGGGGACTCTACGAGAACCCTT

TGTCCTACTGATCCCCGCGTGCTTGGCCGGCCGTGCGAGTAAGCAACTCTGGCGCTGGCA

TGGCATAACCGGCGACGGCAATGCGCAAGATGGGATGCTATGGGCAGAGAGCCGTACTTT

ACTGCTTATGGCACTACAGCAACAGATGGTTACCCCACTAAGCCTGAAGCGAATCGCCAT

CAATTCTGCGCAGTGGCGAGGAGATAAAAGCGCGGAAGTCATTCATCAACTGGCGACGCT

ACTCAAAGCAGGGTTAACGCTTTCTGAAGGGCTGGCTCTGCTGGCGGAACAGCATCCCAG

TAAGCAATGGCAAGCGTTGCTGCAATCGCTGGCGCACGATCTCGAACAGGGCATTGCTTT

TTCCAATGCCTTATTACCCTGGTCAGAGGTATTTCCGCCGCTCTATCAGGCGATGATCCG

CACGGGTGAACTGACCGGTAAGCTGGATGAATGCTGCTTTGAACTGGCGCGTCAGCAAAA

AGCCCAGCGTCAGTTGACCGACAAAGTGAAATCAGCGTTACGTTATCCCATCATCATTTT

AGCGATGGCAATCATGGTGGTTGTGGCAATGCTGCATTTTGTTCTGCCGGAGTTTGCCGC

TATCTATAAGACCTTCAACACCCCACTACCGGCACTAACGCAGGGGATCATGACGCTGGC

AGACTTTAGTGGCGAATGGAGCTGGCTGCTGGTGTTGTTCGGCTTTCTGCTGGCGATAGC

CAATAAGTTGCTGAACGGCCGGCCAAGCACGCGGGGATCAGTAGGACAAAGGGTTCTCGT

AGAGTCCCCGGAAAAAAAAAAAGGACAAAAAGTTTCAAGACGGCAATCTCTTTTTACTGCA

TCTCGTCAGTTGGCAACTTGCCAAGAACTTCGCAAATGACTTTGACATATGATAAGACGT

CAACTGCCCCACGTACAATAACAAAATGGTAGTCATATTATGTCAAGAATAGGTATCCAA

AACGCAGCGGTTGAAAGCATATCAAGAATTGTGTCCCTGTGTTTCAAAGTTTGTGGATAA

TCGAAATCTCTTACATTGAAAACATTATCATACAATCATTTATTAAGTAGTTGAAGCATG

TATGAACTATAAAAGTGTTACTACTCGTTATTATTGTGTACTTTGTGATGCTAAAGTTAT

GAGTAGAAAAAAATGAGAAGTTGTTCTGAACAAAGTAAAAAAAAGAAGTATACTTATTCA

AAATGGGAGAATTGTTGACGCAAAACTCTACGCATGATCTTGTTGGTGGCAGTTCTAGGC

AAAGAAGACAAAGGGACGACTCTAGTAACCTTAAACAATGGATTCAACTTCTTTTGCAAA

CCCAAGTTGAAGGACAATCTCAATTGGTTCAAGTCGATAGTAGTATCGTTAGAATCCTTC

AAGACGAAGAAAATAACCAATTGTTCTGGACCACCACCTAATGGTGGAACACCGATAGCA

GTGGTTTCGAAAACTCTGTCATCGACTTCGTTACAAACTCTTTCAATCTCAATGGAAGAG

ATTTTGATACCACCGATGTTCATGGT GTCATCAGCACGACCGTGAGCATGGTAGTAACCG

TTGGAAGTTAATTCAAAGAT GTCACCGT GTCTTCTCAAAACTTCACCGTTCAAAGTTGGC

ATACCTTTGAAGTAGACATCGTGGTGGTTACCGTTCAATAAAGTCTTAGAAGCACCGAAC

ATAACTGGACCCAAAGCCAATTCACCAATACCTGGCTTGTTCTTTGGCATTGGGTAACCG

TTCTTATCCAAAATGTACAAAGTACAACCCATACATTGGGAAGAAAAGGAGGACAAGGAT

TGGGCTTGTAAGAAAGAACCAGCAGAGAAAGCACCACCGATTTCGGTACCACCACACATT

TCGATAACAGGTTTATAGTTGGCTCTACCCATCAACCACAAGTATTCATCGACGTTAGAA

GCTTCACCAGAGGAAGAAAAGCAACGGATGGTAGACCAGTCATAACCGGAAACGCAGTTG

GTGGATTTCCAAGATCTAACAATAGATGGAACAACACCTAACATAGTAACCTTAGCGTCT

TGGACGAACTTGGCGAAACCAGAAACCAATGGGGAACCATTATACAAAGCGATAGAAGCA

CCGTTCAATAAAGAGGCGTAAACCAACCATGGACCCATCATCCAACCTAAATTAGTTGGC

CAAACAATGACGTCACCTTTACGAATATCCAAGTGAGACCAACCGTCGGCAGCAGCCTTC

AATGGAGTAGCTTGGGTCCATGGAATGGCCTTTGGTTCACCAGTGGTACCGGAAGAGAAT

AAAATGTTGGTGTAGGCATCAACTGGTTGTTCACGAGCGGTGAATTCACAGTTCTTGAAT

TCCTTAGCACGTTCCAAGAAATAATCCCAGGAAATGTCACCGTCACGCAATTCGGCACCG

ATGTTGGAACCGGAACATGGAATGACAATAGCCATTGGAGACTTAGCTTCAACGACTCTA

GAATACAATGGAATTCTCTTCTTACCACGGATGATGTGGTCTTGAGTGAAGATGGCCTTA

GCCTTAGACAATCTCAATCTAGTAGAGATTTCTGGAGCGGAGAAAGAATCAGCGATGGAA

ACGACGACGTAACCAGCCAAGACAATGGCTAAGTAGATGACGACAGCGTCAACGTGCATT

GGCATATCGATGGCGATAGCACAACCTTTTTCCAAACCCATTTCTTCCAAGGCATAACCA

ACCAACCAAACTCTCTTTCTCAATTGGTCCAAAGTCAACTTGTTCAATGGCAAATCGTCG

TTACCCTCATCACGCCAAACGATCATAGTATCATTCAACTTTTTGTTAGAGTTAACATTC

AAGCAGTTCTTAGCAGAGTTCAAGTAACCACCTGGCAACCATTCGGAACCACCTGGGTTG

TTAATATCGTCTCTACGTAAGATACATTCTGGATCTTTAGAAAAGGAGATCTTCATTTCA

TCCATTAAAACAGTTCTCCAGTAAACTTCTGGGTTTCTGACGGAGAACTCTTGGAAGTGA

GAAAAAGAAGAAATTGGATCCTTGTATTTAACACCCAAGAATTCCTTACCTCTTTTCTCC

AACAAAGCACCCAAGTTGGTAGACTTGACCTTTTCAGGGTCTGGAATCCAAGCTGGTGGG

GCTGGACCAAAGTCCTTGTAACAACCATAGAATAACATTTGGTGCAAGGAAAATGGCAAG

TCTGGGGATAAGATATGGTTGGCAATGTTAATCCAAGTTTGTGGGGTAGCAGCACCGTAA

TTACAAACAATTTCAGCTAATCTACCATGCAAAGTTTCGGCGACCTCAGAGGTAATACCC

AAAGCGATGAAATCGGAAGCAACAACAGAATCCAAAGATTTGTAGTTCTTACCCATTATA

GTTTTTTCTCCTTGACGTTAAAGTATAGAGGTATATTAACAATTTTTTGTTGATACTTTT

ATGACATTTGAATAAGAAGTAATACAAACGGAAAATGTTGAAAGTATTAGTTAAAGTGGT

TATGCAGCTTTTGCATTTATATATCTGTTAATAGATCAAAAATCATCGCTTCGCTGATTA

ATTACCCCAGAAATAAGGCTAAAAAACTAATCGCATTATTATCCTATGGTTGTTAATTTG

ATTCGTTGATTTGAAGGTTTGTGGGGCCAGGTTACTGCCAATTTTTCCTCTTCATAACCA

TAAAAGCTAGTATTGTAGAATCTTTATTGTTCGGAGCAGTGCGGCGCGAGGCACATCTGC

GTTTCAGGAACGCGACCGGTGAAGACCAGGACGCACGGAGGAGAGTCTTCCGTCGGAGGG

CTGTCGCCCGCTCGGCGGCTTCTAATCCGTACTTCAATATAGCAATGAGCAGTTAAGCGT

ATTACTGAAAGTTCCAAAGAGAAGGTTTTTTTAGGCTAAGATAATGGGGCTCTTTACATT

TCCACAACATATAAGTAAGATTAGATATGGATATGTATATGGTGGTATTGCCATGTAATA

TGATTATTAAACTTCTTTGCGTCCATCCAAAAAAAAAAGTAAGAATTTTTGAAAATTCAAT

ATAAATGAACCACTTAAGAGCTGAAGGTCCAGCTTCCGTTTTGGCCATTGGTACCGCTAA

CCCAGAAAACATCTTGTTGCAAGACGAATTTCCAGACTACTACTTCAGAGTCACTAAGTC

CGAACACATGACCCAATTGAAGGAAAAGTTCAGAAAGATTTGTGATAAGTCTATGATCAG

AAAAAGAAACTGTTTCTTGAACGAAGAACACTTGAAACAAAACCCTAGATTAGTTGAACA

TGAAATGCAAACTTTAGATGCCAGACAAGATATGTTGGTCGTCGAAGTCCCAAAGTTGGG

TAAGGACGCTTGTGCCAAGGCTATCAAGGAATGGGGTCAACCAAAGTCTAAGATTACTCA

TTTGATCTTCACTTCCGCCTCTACCACCGATATGCCAGGTGCTGATTACCATTGTGCTAA

GTTGTTGGGTTTATCCCCATCTGTTAAAAGAGTTATGATGTACCAATTGGGTTGTTATGG

TGGTGGTACTGTTTTGAGAATTGCCAAAGACATCGCTGAAAACAATAAGGGTGCTAGAGT

TTTGGCTGTTTGTTGTGATATTATGGCTTGTTTGTTCAGAGGTCCATCCGAGTCTGATTT

AGAGTTGTTAGTTGGTCAAGCTATTTTCGGTGACGGTGCTGCTGCTGTTATTGTTGGTGC

TGAACCAGACGAATCTGTTGGTGAACGTCCAATCTTTGAATTGGTCTCTACCGGTCAAAC

CATCTTGCCAAACTCTGAAGGTACCATTGGTGGTCACATCAGAGAAGCTGGTTTGATCTT

CGATTTGCATAAAGATGTTCCTATGTTGATTTCTAATAACATCGAAAAGTGCTTAATCGA

AGCTTTCACTCCAATCGGTATCTCTGATTGGAATTCCATTTTCTGGATTACCCATCCAGG

TGGTAAGGCCATCTTGGATAAGGTTGAAGAAAAGTTGCATTTAAAGTCTGATAAGTTCGT

TGACTCTCGTCACGTTTTGTCTGAACATGGTAACATGTCTTCTTCCACTGTTTTGTTTGT

TATGGATGAATTGAGAAAAAGATCCTTGGAAGAAGGTAAGTCTACTACTGGTGATGGTTT

TGAATGGGGTGTCTTGTTCGGTTTTGGTCCAGGTTTGACCGTTGAAAGAGTTGTCGTTAG

ATCCGTTCCAATCAAGTACTAATTTGCCAGCTTACTATCCTTCTT GAAAATATGCACTCT

ATATCTTTTAGTTCTTAATTGCAACACATAGATTTGCTGTATAACGAATTTTATGCTATT
TTTTAAATTTGGAGTTCAGTGATAAAAGTGTCACAGCGAATTTCCTCACATGTAGGGACC
GAATTGTTTACAAGTTCTCTGTACCACCATGGAGACATCAAAGATTGAAAATCTATGGAA
AGATATGGACGGTAGCAACAAGAATATAGCACGAGCCGCGAAGTTCATTTCGTTACTTTT
GATATCGCTCACAACTATTGCGAAGCGCTTCAGTGAAAAAATCATAAGGAAAAGTTGTAA
ATATTATTGGTAGTATTCGTTTGGTAAAGTAGAGGGGGTAATTTTTCCCCTTTATTTTGT
TCATACATTCTTAAATTGCTTTGCCTCTCCTTTTGGAAAGCTAGGTCCGCCGGCGTTGGA
CGAGCGAAAATTCATTTAATATTCAATGAAGTTATAAATTGATAGTTTAAATAAAGTCAG
TCTTTTTCCTCATGTTTAGAATTGTATTAATGTACGCCGTTTACGTTGGAGTGTAAATGT
GTCTATTCCAGAACGAAATCTAAATGAGCAGTACAGGCAGTTCAGATGGTACTGAAGCGG
TACTAAAGATGCATGAATTGAACAGATGTGGTAGTTATGTATATGAGGAATATGAGTTGT
CACATTAAAAATATAATAGCTATGATCCCATTATTATATTCGTGACAGTTCGTAACGTTT
TAATTGGCTTATGTTTTTGAGAAATGGGTGAATTTTAAGATAATTGTTGGGATTCCATTA
TTGATAAAGGCTATAATATTAGGTATACAGAATATACTGGAAGTTCTCCTCGAGGATATA
GGAATCCTCAAAATGGAATCTATATTTCTATTTACTAATATCACGATTATTCTTCATTCC
GTTTTATATGTTTCATTATCCTATTACATTATCAATCCTTGCATTTCAGCTTCCTCTAAC
TTCGATGACAGCTGGCGGTTTAAACGCGTGGCCGTGCCGTC

Sequences of individual cannabinoid pathway genes

HCS> nucleic acid sequence

ATGGGTAAGAACTACAAATCTTTGGATTCTGTTGTTGCTTCCGATTTCATCGCTTTGGGT
ATTACCTCTGAGGTCGCCGAAACTTTGCATGGTAGATTAGCTGAAATTGTTTGTAATTAC
GGTGCTGCTACCCCACAAACTTGGATTAACATTGCCAACCATATCTTATCCCCAGACTTG
CCATTTTCCTTGCACCAAATGTTATTCTATGGTTGTTACAAGGACTTTGGTCCAGCCCCA
CCAGCTTGGATTCCAGACCCTGAAAAGGTCAAGTCTACCAACTTGGGTGCTTTGTTGGAG
AAAAGAGGTAAGGAATTCTTGGGTGTTAAATACAAGGATCCAATTTCTTCTTTTTCTCAC
TTCCAAGAGTTCTCCGTCAGAAACCCAGAAGTTTACTGGAGAACTGTTTTAATGGATGAA
ATGAAGATCTCCTTTTCTAAAGATCCAGAATGTATCTTACGTAGAGACGATATTAACAAC
CCAGGTGGTTCCGAATGGTTGCCAGGTGGTTACTTGAACTCTGCTAAGAACTGCTTGAAT
GTTAACTCTAACAAAAAGTTGAATGATACTATGATCGTTTGGCGTGATGAGGGTAACGAC
GATTTGCCATTGAACAAGTTGACTTTGGACCAATTGAGAAAGAGAGTTTGGTTGGTTGGT
TATGCCTTGGAAGAAATGGGTTTGGAAAAAGGTTGTGCTATCGCCATCGATATGCCAATG
CACGTTGACGCTGTCGTCATCTACTTAGCCATTGTCTTGGCTGGTTACGTCGTCGTTTCC
ATCGCTGATTCTTTCTCCGCTCCAGAAATCTCTACTAGATTGAGATTGTCTAAGGCTAAG

GCCATCTTCACTCAAGACCACATCATCCGTGGTAAGAAGAGAATTCCATTGTATTCTAGA
GTCGTTGAAGCTAAGTCTCCAATGGCTATTGTCATTCCATGTTCCGGTTCCAACATCGGT
GCCGAATTGCGTGACGGTGACATTTCCTGGGATTATTTCTTGGAACGTGCTAAGGAATTC
AAGAACTGTGAATTCACCGCTCGTGAACAACCAGTTGATGCCTACACCAACATTTTATTC
TCTTCCGGTACCACTGGTGAACCAAAGGCCATTCCATGGACCCAAGCTACTCCATTGAAG
GCTGCTGCCGACGGTTGGTCTCACTTGGATATTCGTAAAGGTGACGTCATTGTTTGGCCA
ACTAATTTAGGTTGGATGATGGGTCCATGGTTGGTTTACGCCTCTTTATTGAACGGTGCT
TCTATCGCTTTGTATAATGGTTCCCCATTGGTTTCTGGTTTCGCCAAGTTCGTCCAAGAC
GCTAAGGTTACTATGTTAGGTGTTGTTCCATCTATTGTTAGATCTTGGAAATCCACCAAC
TGCGTTTCCGGTTATGACTGGTCTACCATCCGTTGCTTTTCTTCCTCTGGTGAAGCTTCT
AACGTCGATGAATACTTGTGGTTGATGGGTAGAGCCAACTATAAACCTGTTATCGAAATG
TGTGGTGGTACCGAAATCGGTGGTGCTTTCTCTGCTGGTTCTTTCTTACAAGCCCAATCC
TTGTCCTCCTTTTCTTCCCAATGTATGGGTTGTACTTTGTACATTTTGGATAAGAACGGT
TACCCAATGCCAAAGAACAAGCCAGGTATTGGTGAATTGGCTTTGGGTCCAGTTATGTTC
GGTGCTTCTAAGACTTTATTGAACGGTAACCACCACGATGTCTACTTCAAAGGTATGCCA
ACTTTGAACGGTGAAGTTTTGAGAAGACACGGTGACATCTTTGAATTAACTTCCAACGGT
TACTACCATGCTCACGGTCGTGCTGATGACACCATGAACATCGGTGGTATCAAAATCTCT
TCCATTGAGATTGAAAGAGTTTGTAACGAAGTCGATGACAGAGTTTTCGAAACCACTGCT
ATCGGTGTTCCACCATTAGGTGGTGGTCCAGAACAATTGGTTATTTTCTTCGTCTTGAAG
GATTCTAACGATACTACTATCGACTTGAACCAATTGAGATTGTCCTTCAACTTGGGTTTG
CAAAAGAAGTTGAATCCATTGTTTAAGGTTACTAGAGTCGTCCCTTTGTCTTCTTTGCCT
AGAACTGCCACCAACAAGATCATGCGTAGAGTTTTGCGTCAACAATTCTCCCATTTTGAA
TAA

TKS> nuclei c acid s equence

ATGAACCACTTAAGAGCTGAAGGTCCAGCTTCCGTTTTGGCCATTGGTACCGCTAACCCA
GAAAACATCTTGTTGCAAGACGAATTTCCAGACTACTACTTCAGAGTCACTAAGTCCGAA
CACATGACCCAATTGAAGGAAAAGTTCAGAAAGATTTGTGATAAGTCTATGATCAGAAAA
AGAAACTGTTTCTTGAACGAAGAACACTTGAAACAAAACCCTAGATTAGTTGAACATGAA
ATGCAAACTTTAGATGCCAGACAAGATATGTTGGTCGTCGAAGTCCCAAAGTTGGGTAAG
GACGCTTGTGCCAAGGCTATCAAGGAATGGGGTCAACCAAAGTCTAAGATTACTCATTTG
ATCTTCACTTCCGCCTCTACCACCGATATGCCAGGTGCTGATTACCATTGTGCTAAGTTG
TTGGGTTTATCCCCATCTGTTAAAAGAGTTATGATGTACCAATTGGGTTGTTATGGTGGT
GGTACTGTTTTGAGAATTGCCAAAGACATCGCTGAAAACAATAAGGGTGCTAGAGTTTTG

GCTGTTTGTTGTGATATTATGGCTTGTTTGTTCAGAGGTCCATCCGAGTCTGATTTAGAG

TTGTTAGTTGGTCAAGCTATTTTCGGTGACGGTGCTGCTGCTGTTATTGTTGGTGCTGAA

CCAGACGAATCTGTTGGTGAACGTCCAATCTTTGAATTGGTCTCTACCGGTCAAACCATC

TTGCCAAACTCTGAAGGTACCATTGGTGGTCACATCAGAGAAGCTGGTTTGATCTTCGAT

TTGCATAAAGATGTTCCTAT GTTGATTTCTAATAACATCGAAAAGTGCTTAATCGAAGCT

TTCACTCCAATCGGTATCTCTGATTGGAATTCCATTTTCTGGATTACCCATCCAGGTGGT

AAGGCCATCTTGGATAAGGTTGAAGAAAAGTTGCATTTAAAGTCTGATAAGTTCGTTGAC

TCTCGTCACGTTTTGTCTGAACATGGTAACATGTCTTCTTCCACTGTTTTGTTTGTTATG

GATGAATTGAGAAAAAGATCCTTGGAAGAAGGTAAGTCTACTACTGGTGATGGTTTTGAA

TGGGGTGTCTTGTTCGGTTTTGGTCCAGGTTTGACCGTTGAAAGAGTTGTCGTTAGATCC

GTTCCAATCAAGTACTAA

OAO nucleic acid sequence

ATGGCTGTTAAACACTTGATCGTCTTAAAGTTCAAAGACGAAATTACCGAAGCCCAAAAA

GAGGAATTCTTCAAAACTTACGTTAACTTGGTTAACATTATTCCAGCTATGAAGGATGTC

TACTGGGGTAAAGACGTCACCCAAAAGAACAAAGAAGAAGGTTACACTCATATTGTTGAA

GTCACCTTCGAATCTGTTGAAACTATCCAAGACTACATTATTCACCCAGCTCATGTTGGT

TTCGGTGACGTTTACAGATCTTTCTGGGAAAAGTTGTTGATCTTTGATTACACCCCTAGA

AAGTAA

HCS amino acid sequence:

MGKNYKSLDSWASDFIALGITSEVAETLHGRLAEIVCNYGAATPQTWIN
IANHILS PDLPFSLHQMLFYGCYKDFGPAPPAWI PDPEKVKSTNLGALLE
KRGKEFLGVKYKDPISSFSHFQEFSVRNPEVYWRTVLMDEMKISFSKDPE
CILRRDDINNPGGSEWLPGGYLNSAKNCLNVNSNKKLNDTMIVWRDEGND
DLPLNKLTLDQLRKRVWLVGYALEEMGLEKGCAIAIDMPMHVDAWIYLA
IVLAGYVWSIADSFSAPEISTRLRLSKAKAIFTQDHIIRGKKRIPLYSR
WEARS PMAIVI PCS GSNIGAELRDGDISWDYFLERAKEFKNCEFTAREQ
PVDAYTNILFSSGTTGEPKAIPWTQATPLKAAADGWSHLDIRKGDVIVWP
TNLGWMMGPWLVYASLLNGASIALYNGSPLVSGFAKFVQDAKVTMLGWP
SIVRS WKSTNCVS GYDWSTI RCFSSSGEAS NVDE YLWLMGRAN YKPVI EM
CGGTEIGGAFSAGSFLQAQSLSSFSSQCMGCTLYILDKNGYPMPKNKPGI
GELALGPVMFGASKTLLNGNHHDVYFKGMPTLNGEVLRRHGDI FELTSNG
YYHAHGRADDTMNIGGIKISSIEIERVCNEVDDRVFETTAIGVPPLGGGP
EQLVIFFVLKDSNDTTIDLNQLRLSFNLGLQKKLNPLFKVTRWPLSSLP
RTATNKIMRRVLRQFSHFE

TKS amino acid sequence:

MNHLRAEGPASVLAIGTANPENILLQDEFPDYYFRVTKSEHMTQLKEK
FRKICDKSMIRKRNCFLNEEHLKQNPRLVEHEMQTLDARQDMLWEV
PKLGKDACAKAIKEWGQPKSKITHLI FTSASTTDMPGADYHCAKLLGL
SPSVKRVMMYQLGCYGGGTVLRIAKDIAENNKGARVLAVCCDIMACL
FRGPSES DLELLVGQAI FGDGAAAVIVGAE PDESVGERP IFELVSTGQT I
LPNSEGTIGGHIREAGLIFDLHKDVPMLISNNIEKCLIEAFTPIGISDWNSI
FWITHPGGKAILDKVEEKLHLKSDKFVDSRHVLSEHGNMSSSTVLFVM
DELRKRSLEEGKSTTGDGFEWGVLFGFGPGLTVERVWRSVPIKY

OAC amino acid sequence:

MAVKHLIVLKFKDEITEAQKEEFFKT YVNLVNI IPAMKDVYWGKDVTQKNKEEGYTHIVEVT
FESVETIQDYI IHPAHVGFGDVYRSFWEKLLIFDYTPRK

pGALl

TGGAACTTTCAGTAATACGCTTAACTGCTCATTGCTATATTGAAGTACGGATTAGAAGCC
GCCGAGCGGGCGACAGCCCTCCGACGGAAGACTCTCCTCCGTGCGTCCTGGTCTTCACCG
GTCGCGTTCCTGAAACGCAGATGTGCCTCGCGCCGCACTGCTCCGAACAATAAAGATTCT
ACAATACTAGCTTTTATGGTTATGAAGAGGAAAAATTGGCAGTAACCTGGCCCCACAAAC
CTTCAAATCAACGAATCAAATTAACAACCATAGGATAATAATGCGATTAGTTTTTTAGCC
TTATTTCTGGGGTAATTAATCAGCGAAGCGATGATTTTTGATCTATTAACAGATATATAA
ATGCAAAAGCTGCATAACCACTTTAACTAATACTTTCAACATTTTCGGTTTGTATTACTT
CTTATTCAAATGTCATAAAAGTATCAACAAAAAATTGTTAATATACCTCTATACTTTAAC
GTCAAGGAGAAAAAACT ATA
pGAL!O

CATCG CTTCG CTG ATTAATTACCCCAGAAAT AAG GCTAAAAAACT AAT CG CATTATTATC
CTATGGTTGTTAATTTGATTCGTTGATTTGAAGGTTTGTGGGGCCAGGTTACTGCCAATT
TTTCCTCTTCATAACCATAAAAG CTAGTATT GTA GAAT CTTTATTGTTCGGA GCAGTG CG
GCGCGAGGCACATCTGCGTTTCAGGAACGCGACCGGTGAAGACCAGGACGCACGGAGGAG
AGTCTTCCGTCGGAGGGCTGTCGCCCGCTCGGCGG CTTCTAAT CCGTACTTCAATATAG C
AATGA GCAGTTAAG CGTATTACTGAAAGTT CCAAAGA GAAGG TTTTTTTAGGCTAAGATA
ATGGGGCT CTTTACATTTCCACAACATATAAGTAAG ATTA GATATG GATATGTATATG GT
GGTATTGCCATGTAAT ATGATTATTAAACTT CTTTGCGTCCATCCAAAAAAAAAGT AAGA
ATTTTTGAAAATTCAATATAA

pGAL2

GGCTTAAGTAGGTTG CAATTTCTTTTTCTATTAGTAG CTAAAAAT GGGTCACGTGATCTA
TATTCGAAAGGGGCGGTTGCCTCAGGAAGGCACCGGCGGTCTTTCGTCCGTGCGGAGATA
TCTGCGCCGTTCAGGGGTCCATGTGCCTTGGACGATATTAAGGCAGAAGGCAGTATCGGG
GCGGATCACTCCGAACCGAGATTAGTTAAGCCCTTCCCATCTCAAGATGGGGAGCAAATG
GCATTATACTCCTG CTAGAAAGTTAACTGTG CACATATTCTTAAATTATACAAT GTTCTG
GAGAGCTATTGTTTAAAAAACAAACATTT CG CAG GCTAAAAT GTG GAGATAGGATTAGTT
TTGTAGACATATATAAACAAT CAGTAATT GGATTGAAAATTTGGTGTT GTG AATTGCTCT
TCATTATGCACCTTATTCAATTATCATCAAG AATA GCAATAGTTAAGTAAACACAAG ATT
AACATAATAAAAAAAAT AATT CTTTCATA

pGAL3

TTTT ACTATTATCTTCTACGCTGACAGTAAT ATCAAACAGT GACACATATTAAACACAGT
GGTTTCTTTGCATAAACACCAT CAGCCTCAAGTCGTCAAGTAAAG ATTTCGTGTTCATGC
AGATAG ATAACAAT CTATATGTTGATAATTAGCGTTGCCTCATCAATGCGAGATCCGTTT
AACCGGACCCTAGTG CACTTACCCCACGTTCGGTCCACTGTGTGCCGAACAT GCTCCTTC
ACTATTTTAACAT GTGGAATT CTTGAAAG AATG AAAT CGCCATGCCAAG CCATCACACGG
TCTTTTATGCAATTGATTGACCGCCTG CAACACATAGGCAGTAAAATTTTTACTGAAACG
TATATAAT CATCATAAG CGACAAGTG AGGCAACACCTTTGTTACCACATTGACAACCCCA
GGTATTCATACTTCCTATTAGCGGAATCAGGAGTGCAAAAAGAGAAAATAAAAGTAAAAA
GGTAGGGCAACACATAGT

pGAL7

GGACGGTAGCAACAAGAATATAGCACGAGCCGCGAAGTTCATTTCGTTACTTTTGATATC
GCT CACAACTATTGCGAAG CGCTTCAGTGAAAAAAT CATAAG GAAAAGTT GTAAAT ATTA
TTGGTAGTATTCGTTTGGTAAAGTAGAGGGGGTAATTTTTCCCCTTTATTTTGTTCATAC
ATT CTTAAATT GCTTTGCCTCTCCTTTTGGAAAG CTATACTTCGGAGCACTGTTGAGCGA
AGGCT CATTAGATATATTTT CTGTCATTT CCTTAACCCAAAAAT AAG GGAAAG GGTCCA
AAAAG CGCTCG GACAACTGTTG ACCGTGATCCGAAG GACTG GCTATACAGTGTTCACAAA
ATAG CCAAGCTGAAAAT AAT GTGTAGCTATGTTCAGTT AGTTTG GCTAGCAAAG ATATAA
AAG CAG GTCG GAAAT ATTTATGGGCATTATTATG CAGAGCATCAACATGATAAAAAAAAA
CAGTTGAAT ATTCCCTCAAAA

pGAL4

GCGACACAGAGATGACAGACGGTGGCGCAGGATCCGGTTTAAACGAGGATCCCTTAAGTT
TAAACAACAACAGCAAGCAGGTGTGCAAGACACTAGAGACTCCTAACATGATGTATGCCA
ATAAAACACAAG AGATAAACAACATT GCATGGAGGCCCCAGAGGGG CGATT GGTTTGGGT
GCGTGAGCGGCAAGAAGTTTCAAAACGTCCGCGTCCTTTGAGACAGCATTCGCCCAGTAT
TTTTTTTATTCTACAAACCTTCTATAATTT CAAAGT ATTTACATAATT CTGTATCAGTTT
AAT CACCATAAT ATCGTTTTCTTTGTTTAGTG CAATTAATTTTTCCTATT GTTACTTCGG
GCCTTTTTCT GTTTTATGAGCTATTTTTTCCGTCATCCTTCCCCAGATTTTCAG CTTCAT
CTCCAGATTGTGTCTACGTAAT GCACG CCATCATTTTAAGAGAGGACAGAGAAGCAAGCC
TCCTGAAAG

pMALI

GATGATGGAC ACTAGTGTGT CGAGAATGTA TCAACTATAT ATAGTCCTAA
TGCCACACAA ATATGAAGTG GGGGAAGCCC ATTCTTAATC CGGCTCAATT
TTGGTGCGTG ATCGCGGCCT ATGTTTGCTT CCAGAAAAAG CTTAGAATAA
TATTTCTCAC CTTTGATGGA ATGCTCGCGA GTGCTCGTTT TGATTACCCC
ATATGCATTG TTGCAGCATG CAAGCACTAT TGCAAGCCAC GCATGGAAGA
AATTTGCAAA CACCTATAGC CCCGCGTTGT TGAGGAGGTG GACTTGGTGT
AGGACCATAA AGCTGTGCAC TACTATGGTG AGCTCTGTCG TCTGGTGACC
TTCTATCTCA GGCACATCCT CGTTTTTGTG CATGAGGTTC GAGTCACGCC
CACGGCCTAT TAATCCGCGA AATAAATGCG AAATCTAAAT TATGACGCAA
GGCTGAGAGA TTCTGACACG CCGCATTTGC GGGGCAGTAA TTATCGGGCA
GTTTTCCGGG GTTCGGGATG GGGTTTGGAG AGAAAGTTCA ACACAGACCA

AAACAGCTTG GGACCACTTG GATGGAGGTC CCCGCAGAAG
AGCTCTGGCG CGTTGGACAA ACATTGACAA TCCACGGCAA AATTGTCTAC
AGTTCCGTGT ATGCGGATAG GGATATCTTC GGGAGTATCG CAATAGGATA
CAGGCACTGT GCAGATTACG CGACATGATA GCTTTGTATG TTCTACAGAC
TCTGCCGTAG CAGTCTAGAT ATAATATCGG AGTTTTGTAG CGTCGTAAGG
AAAACTTGGG TTACACAGGT TTCTTGAGAG CCCTTTGACG TTGATTGCTC
TGGCTTCCAT CCAGGCCCTC ATGTGGTTCA GGTGCCTCCG CAGTGGCTGG
CAAGCGTGGGGGTCAATTAC GTCACTTCTA TTCATGTACC CCAGACTCAA
TTGTTGACAG CAATTTCAGC GAGAATTAAA TTCCACAATC AATTCTCGCT
GAAATAATTA GGCCGTGATT TAATTCTCGC TGAAACAGAA TCCTGTCTGG
GGTACAGATA ACAATCAAGT AACTATTATG GACGTGCATA GGAGGTGGAG
TCCATGACGC AAAGGGAAAT ATTCATTTTA TCCTCGCGAA
GTTGGGATGT GTCAAAGCGT CGCGCTCGCT ATAGTGATGA GAATGTCTTT
AGTAAGCTTA AGCCATATAA AGACCTTCCG CCTCCATATT TTTTTTTATC
CCTCTTGACA ATATTAATTC CTT

pMAL2

AAGGAATTAA TATTGTCAAG AGGGATAAAA AAAAATATGG AGGCGGAAGG
TCTTTATATG GCTTAAGCTT ACTAAAGACA TTCTCATCAC TATAGCGAGC
GCGACGCTTT GACACATCCC AACTTCGCGA GGATAAAATG AATATTTCCC
TTTGCGTCAT GGACTCCACC TCCTATGCAC GTCCATAATA GTTACTTGAT
TGTTATCTGT ACCCCAGACA GGATTCTGTT TCAGCGAGAA TTAAATCACG
GCCTAATTAT TTCAGCGAGA ATTGATTGTG GAATTTAATT
CTCGCTGAAA TTGCTGTCAA CAATTGAGTC TGGGGTACAT GAATAGAAGT
GACGTAATTG ACCCCCACGC TTGCCAGCCA CTGCGGAGGC ACCTGAACCA
CATGAGGGCC TGGATGGAAG CCAGAGCAAT CAACGTCAAA GGGCTCTCAA
GAAACCTGTG TAACCCAAGT TTTCCTTACG ACGCTACAAA ACTCCGATAT
TATATCTAGA CTGCTACGGC AGAGTCTGTA GAACATACAA
AGCTATCATG TCGCGTAATC TGCACAGTGC CTGTATCCTA TTGCGATACT
CCCGAAGATA TCCCTATCCG CATACACGGA ACTGTAGACA ATTTTGCCGT
GGATTGTCAA TGTTTGTCCA ACGCGCCAGA GCTCTTCTGC GGGGACCTCC
ATCCAAGTGG TCCCAAGCTG TTTTGGTCTG TGTTGAACTT TCTCTCCAAA
CCCCATCCCG AACCCCGGAA AACTGCCCGA TAATTACTGC
CCCGCAAATG CGGCGTGTCA GAATCTCTCA GCCTTGCGTC ATAATTTAGA
TTTCGCATTT ATTTCGCGGA TTAATAGGCC GTGGGCGTGA CTCGAACCTC
ATGCACAAAA ACGAGGATGT GCCTGAGATA GAAGGTCACC AGACGACAGA
GCTCACCATA GTAGTGCACA GCTTTATGGT CCTACACCAA GTCCACCTCC
TCAACAACGC GGGGCTATAG GTGTTTGCAA ATTTCTTCCA
TGCGTGGCTT GCAATAGTGC TTGCATGCTG CAACAATGCA TATGGGGTAA
TCAAAACGAG CACTCGCGAG CATTCCATCA AAGGTGAGAA ATATTATTCT
AAGCTTTTTC TGGAAGCAAA CATAGGCCGC GATCACGCAC CAAAATTGAG
CCGGATTAAG AATGGGCTTC CCCCACTTCA TATTTGTGTG GCATTAGGAC
TATATATAGT TGATACATTC TCGACACACT AGTGTCCATC ATC

pMALll

GCGCCTCAAG AAAATGATGC TGCAAGAAGA ATTGAGGAAG GAACTATTCA
TCTTACGTTG TTTGTATCAT CCCACGATCC AAATCATGTT ACCTACGTTA
GGTACGCTAG GAACTAAAAA AAGAAAAGAA AAGTATGCGT TATCACTCTT

CGAGCCAATT CTTAATTGTG TGGGGTCCGC GAAAATTTCC GGATAAATCC
TGTAAACTTT AACTTAAACC CCGTGTTTAG CGAAATTTTC AACGAAGCGC
GCAATAAGGA GAAATATTAT CTAAAAGCGA GAGTTTAAGC GAGTTGCAAG
AATCTCTACG GTACAGATGC AACTTACTAT AGCCAAGGTC TATTCGTATT
ACTATGGCAG CGAAAGGAGC TTTAAGGTTT TAATTACCCC ATAGCCATAG
ATTCTACTCG GTCTATCTAT CATGTAACAC TCCGTTGATG CGTACTAGAA
AATGACAACG TACCGGGCTT GAGGGACATA CAGAGACAAT TACAGTAATC
AAGAGTGTAC CCAACTTTAA CGAACTCAGT AAAAAATAAG GAATGTCGAC
ATCTTAATTT TTTATATAAA GCGGTTTGGT ATTGATTGTT TGAAGAATTT
TCGGGTTGGT GTTTCTTTCT GATGCTACAT AGAAGAACAT CAAACAACTA
AAAAAATAGT ATAAT

pMAL12

ATTATACTAT TTTTTTAGTT GTTTGATGTT CTTCTATGTA GCATCAGAAA
GAAACACCAA CCCGAAAATT CTTCAAACAA TCAATACCAA ACCGCTTTAT
ATAAAAAATT AAGATGTCGA CATTCCTTAT TTTTTACTGA GTTCGTTAAA
GTTGGGTACA CTCTTGATTA CTGTAATTGT CTCTGTATGT CCCTCAAGCC
CGGTACGTTG TCATTTTCTA GTACGCATCA ACGGAGTGTT ACATGATAGA
TAGACCGAGT AGAATCTATG GCTATGGGGT AATTAAAACC TTAAAGCTCC
TTTCGCTGCC ATAGTAATAC GAATAGACCT TGGCTATAGT AAGTTGCATC
TGTACCGTAG AGATTCTTGC AACTCGCTTA AACTCTCGCT TTTAGATAAT
ATTTCTCCTT ATTGCGCGCT TCGTTGAAAA TTTCGCTAAA CACGGGGTTT
AAGTTAAAGT TTACAGGATT TATCCGGAAA TTTTCGCGGA CCCCACACAA
TTAAGAATTG GCTCGAAGAG TGATAACGCA TACTTTTCTT TTCTTTTTTT
AGTTCCTAGC GTACCTAACG TAGGTAACAT GATTTGGATC GTGGGATGAT
ACAAACAACG TAAGATGAAT AGTTCCTTCC TCAATTCTTC TTGCAGCATC
ATTTTCTTGA GGCGCTCTGG GCAAGGTATA AAAAGTTCCA TTAATACGTC
TCTAAAAAAT TAAATCATCC ATCTCTTAAG CAGTTTTTTT GATAATCTCA
AATGTACATC AGTCAAGCGT AACTAAATTA CATAA

pMAL31

TTATGTATTT TAGTTACGCT TGACTGATGT ACATTTGAGA TTATCAAAAA
AACTGCTTAA GAGATAGATG GTTTAATTTT TTAGAGACGT ATTAATGGAA
CTTTTTATAC CTTGCCCAGA GCGCCTCAAG AAAATGATGC TGAAAGAAGA
ATTGAGGAAG GAACTACTCA TCTTACGTTG TTTGTATCAT CCCACGATCC
AAATCATGTT ACCTACGTTA GGTACGCTAG GAACTGAAAA AAGAAAAGAA
AAGTATGCGT TATCACTCTT CGAGCCAATT CTTAATTGTG TGGGGTCCGC
GAAAACTTCC GGATAAATCC TGTAAACTTA AACTTAAACC CCGTGTTTAG
CGAAATTTTC AACGAAGCGC GCAATAAGGA GAAATATTAT ATAAAAGCGA
GAGTTTAAGC GAGGTTGCAA GAATCTCTAC GGTACAGATG CAACTTACTA
TAGCCAAGGT CTATTCGTAT TGGTATCCAA GCAGTGAAGC TACTCAGGGG
AAAACATATT TTCAGAGATC AAAGTTATGT CAGTCTCTTT TTCATGTGTA
ACTTAACGTT TGTGCAGGTA TCATACCGGC CTCCACATAA TTTTTGTGGG
GAAGACGTTG TTGTAGCAGT CTCCTTATAC TCTCCAACAG GTGTTTAAAG
ACTTCTTCAG GCCTCATAGT CTACATCTGG AGACAACATT AGATAGAAGT
TTCCACAGAG GCAGCTTTCA ATATACTTTC GGCTGTGTAC ATTTCATCCT
GAGTGAGCGC ATATTGCATA AGTACTCAGT ATATAAAGAG ACACAATATA

CTCCATACTT GTTGTGAGTG GTTTTAGCGT ATTCAGTATA ACAATAAGAA
TTACATCCAA GACTATTAAT TAACT

pMAL32

AGTTAATTAA TAGTCTTGGA TGTAATTCTT ATTGTTATAC TGAATACGCT
AAAACCACTC ACAACAAGTA TGGAGTATAT TGTGTCTCTT TATATACTGA
GTACTTATGC AATATGCGCT CACTCAGGAT GAAATGTACA CAGCCGAAAG
TATATTGAAA GCTGCCTCTG TGGAAACTTC TATCTAATGT TGTCTCCAGA
TGTAGACTAT GAGGCCTGAA GAAGTCTTTA AACACCTGTT GGAGAGTATA
AGGAGACTGC TACAACAACG TCTTCCCCAC AAAAATTATG TGGAGGCCGG
TATGATACCT GCACAAACGT TAAGTTACAC ATGAAAAAGA GACTGACATA
ACTTTGATCT CTGAAAATAT GTTTTCCCCT GAGTAGCTTC ACTGCTTGGA
TACCAATACG AATAGACCTT GGCTATAGTA AGTTGCATCT GTACCGTAGA
GATTCTTGCA ACCTCGCTTA AACTCTCGCT TTTATATAAT ATTTCTCCTT
ATTGCGCGCT TCGTTGAAAA TTTCGCTAAA CACGGGGTTT AAGTTTAAGT
TTACAGGATT TATCCGGAAG TTTTCGCGGA CCCCACACAA TTAAGAATTG
GCTCGAAGAG TGATAACGCA TACTTTTCTT TTCTTTTTTC AGTTCCTAGC
GTACCTAACG TAGGTAACAT GATTTGGATC GTGGGATGAT ACAAACAACG
TAAGATGAGT AGTTCCTTCC TCAATTCTTC TTTCAGCATC ATTTTCTTGA
GGCGCTCTGG GCAAGGTATA AAAAGTTCCA TTAATACGTC TCTAAAAAAT
TAAACCATCT ATCTCTTAAG CAGTTTTTTT GATAATCTCA AATGTACATC
AGTCAAGCGT AACTAAAATA CATAA

pMAL13

AGTACAGGAACGATTGTCTTGATAATATGTGAAAAGTGCACACGAAATTAGAGGGTGTCC
TTTACAAGT ATT CTTA GAAACACATT CAAGA GCACAAAAGT CG ATG CTTTAAG GGT CAAG
GTGGTGGAAAACTTGACTGGAATTCTTGACGAAAAAACAAGAAAAACGTGATTCGAGCAA
TCATAAACAT ACAGCCCCGTTCCAACCG GATCTTGAG GTTTCCCATTTTA GAT GGAAAT A
AGCAGA GCAAAAT AAAAAT CTTG AACAAGT AAT AGTGGTGACTG CAG GTTACGTTGGCAT
ATAAAGT CCGGGTGACCTGG GTTTCCTG CACCACCAGCCCCCATATG CTAGCACAAT GGG
TIIT CTTTATCCCCGGTCATAATT ACTCATTTTGCTATATT CTT CATAACTT AAGTACG C
AGATAGAGAAAATTAATAATCTCGATATATATTAAAGTAAATGAAAAGTAGAAAATTTAG
CCAGAACTC$\overline{\text{TTTTTT}}$GCTTCGAGT

pMAL22

GTAGCTT CACTGCTTGGATACCAAT ACGAAT A GACCTTGGCTATAGT AAGTTG CATCT GT
ACCGTAG A GATT CTTGCAACCTCG CTTAAACT CTCG CTTTTATATAAT ATTT CT CCTTAT
TGCG CG CTTCGTTGAAAATTT CG CTAAACACG GGGTTTAAGTTT AAGTTT ACAG GATTT A
TCCG GAAGTTTT CGCG GACCCCACACAATTAAG AATT GGCTCG AAG AGTG A TAACGCAT A
CTTTTCTTTTCTTTTTTCAGTTCCTAGCGTACCTAACGTAG GTAACAT GATTT GGATCGT
GGGATG ATACAAACAACGT AAG ATG AGTAGTT CCTTCCTCAATTCTTCTTTCAG CATCAT
TTTCTTG A GGCG CTCTG GGCAAG GTAT AAAAAGTT CCATTAAT ACGTCTCTAAAAAATT A
AACCAT CT ATCT CTTAAG CAG$\overline{\text{TTTTTTT}}$GATAAT CT CAAAT GTACAT CAGTCAAG CGTAA
CTAAAATACATAA

pMAL33

AGCTCAGTTGTCAAGATTTAGTCATTAAGAAGGGCCGCAGCAGCTTTTTGTATAATAGAG
CGTCTTTTTTTGTTTGTGAAAAAAATTTTATGGTGAGATATTGTTCGATTCTACGAAGTCA
TTTTACTAGTTTATGGACTCTGATATAAGACAGAGTTGACAAGGAAATGGTGCCGTGATT
GTTTCCGTGTACAGCTTTTGAGAACTTCCTTGAAAACCAAT CATCTAGCACTTTCATTTC
TGGGGAAAAACCT GGAACCAAAT CTTGAAAAAT AAATT CCCCAGAAGTTTT CCTTATTCC
GTGTTCTAATCTTCTCGTTCACTTTGCAGTGACATTCCACGGCCATGCGCAATTTACCCC
GCCCCCGGATTTTATTGTCCGTACCGCCATTTTTTCAATAGATTAAAAAG GAACAAAAAAT
CATTTCAGAAG GTTTCTTTCTCGGGAAAACACT AGAGTGTAAAT ATTGAATATCAAACAT
CGAACGAGAGCATCTTGAAGATATTTATGTTCTAAAT

pCAT8

GTGTTTATTCGCGATATGAGTTGTGATATCAGAGACAGAGAGAGTTTATGTGCGTAACAG
GAACGGAGAAAACCAG AGTAATTGAGTATTATAAG CAATAAAT CATAAAAAG ACATTCTT
TCTCGTGCAATTTTTTGGTATTCGGGATAAT CTTCTACTTGAAACTT CTTTTTTTCGGTG
TTTAATTTGCCTATTGGTAAATATTTTTGCCGCCGAGGTTCTCAGTGATTATATTCGTAT
TAAGCGATAACCGAGACATGCATGGAGCGGCGGGGCTGATATTTTGTGGGGTACGAAAGC
ATGATTGGTCAGTGACACTCAAAAAAAG AAAACAG CCGTAAAAT AGTAG ATTTTGTTAAA
CTCCCCTTTAAACCTGTGATATTGTAAAAAG ACGAAG AATTTAATAATTTAATAATTCAT
TACGGTATTTATTTCTTCATAAACAGTTACAACACCCT AAAG AGAATTTACAAGTTG AGT
AAAAGACAAGACACAAAATT

pHXT2

CGCAGCTTCACTTTTAAGTTTCTTTTT CTCCTCACGGCGCAACCGCTAACTTAAGCTAAT
CCTTATGAATCCGGAGAAAAG CGGGGTCTTTTAACTCAATAAAATTTTCCGAAATCCTTT
TTCCTACGCGTTTTCTTCGGGAACTAGATAGGTGGCT CTTCCACCTGTTTTTCCATCATT
TTAGTmTCGCAAGCCATGCGTGCCTTTTCGTmTGCGATGGCGAAGCAGGGCTGGAA
AAATTAACGGTACGCCGCCTAACGATAGTAATAGGCCACGCAACTGGCGTGGACGACAAC
AATAAGTCGCCCATTTTTTATGTTTTCAAAACCTAGCAACCCCCACCAAACTTGTCATCG
TTCCCGGATTCACAAAT GATATAAAAAG CGATTACAATT CTACATTCTAACCAGATTTGA
GATTTCCTCTTTCTCAATTCCTCTTATATTAGATTATAAG AACAACA AATTAAATT ACAA
AAAG ACTTATAAAG CAACATA

pHXT4

CGTCTCTTTCTGTGGAGAAG AAG ATATTTCCCCGAGCAGTTTTTTTTTCCATGGGGCCCCA
TATTCCCCCGCCTGCAGGAAAACTTGGGGAAAGAGGAAAAACACTTCGGATAAAAACGGT
CAAGAAG CTCTTCGACGATTTAGTG CCACCTTCATGAAAAATT CCAGAGTTTTTTTCCAGC
TGCTTTGATTTTACAGTCCATTATTCGGCGTCTAACGATTCTGATTAAG AAACAACG GAG
GAAAACT CAAATTCTAATATAATATTTTTAAGTTTATGAAGGTGGGGTGGT AAG AAAAG C
AACTAAAATAATCTACAAGTCAATTAGTGGTGAAAAGCTTCAACACTGGGGAATGAATAA
TATGTCATCTAGAAAAAATTTTATATAAAT ACTCAGTGTTTTATTCATTATTCTCGATTC
ATTCACTTCAATTCCTCTTCATGAGTAATAGAAACCAT CAAGAAAAG ATATATTCAAAG C
CTCTTATCAAGGTTTGGTTTTGAAACACTTTT ACAATAAAAT CTGCCAAAA

pMTH I

TCCGAAATTATTCCCTAGAACAAG CGGGAAAAAG GTCCGGGGAAAT GGAGTCCGTGCGAG
TIIT GTTAGGATGACTG CCCCACACATTTCCTCATCTTATAATTTT GTGGAAAAATT CAT
CGTGAGAGAAAAT ACGAGTCCATTTCTCCAGTGAAACT ACCGTAGACATGGAATATCTGC
CATTCTACCCCTTATTCAAGTGCC[T]TTTTTTTTTTTTTTTTCATCCCACATTTTATTGCTGC
CTCAATCTCCATTAAG AAAAAAAATTT ATATAACCAAAT GACATTTTTCCTTTCTTCTCA
AACTTTGTAATGCGCCTGTAACTGCTTCTTTTTTTATTAAAAAACAG CATGGAGTTTTTT
AATAACTTAAGG AAACAT ACAAAAAG ATTTGTTCATTTCACTCCAAGTA[T]TTTTTTAACGT
ATATTGAAAGTTCT CAATAGCGAAACCACAAGCAGCAAT ACAAAG AGAATTTTATTCGAA
CGCATAGAGTACACACACTCAAAGGA

pSUC2

CATTATGAGGG CTTCCATTATTCCCCG CATTTTTATTACTCTGAACAG GAATAAAAAG AA
AAAACCCAGTTTAGG AAATT ATCCGGGGGCGAAGAAATACGCGTAGCGTTAATCGACCCC
ACGTCCAGGGTTTTTCCATGGAGGTTTCTGGAAAAACTGACGAGGAATGTGATTATAAAT
CCCTTTATGTG ATGTCTAAG ACTTTTAAG GTACGCCCGATGTTTG CCTATTACCATCATA
GAGACGTTTCTTTTCGAGGAATGCTTAAACGACTTTGTTTGACAAAAATGTTGCCTAAGG
GCTCTATAGTAAACCATTT GGAAG AAAG ATTTGACGAC[T]TTTTTTTTTTTTGGATTTCGATC
CTATAATCCTTCCTCCTGAAAAG AAACAT ATAAAT AGATATGTATTATTCTTCAAAACAT
TCTCTTGTTCTTGTGC[T]TTTTTTTTACCATATATCTTAC[T]TTTTTTTTTTCTCTCAGAGAA
ACAAG CAAAACAAAAAG CTTTTCTTTTCACTAACGTATATG

WHAT IS CLAIMED IS:

1. A host cell comprising a heterologous genetic pathway that produces a heterologous product and that is regulated by an exogenous agent, wherein the host cell does not produce a precursor required to make the product.

2. The host cell of claim 1, wherein the exogenous agent comprises a regulator of gene expression.

3. The host cell of claim 2, wherein the exogenous agent decreases production of the heterologous product.

4. The host cell of claim 3, wherein the exogenous agent is glucose and expression of one or more enzymes encoded by the heterologous genetic pathway are under control of a glucose repressed promoter.

5. The host cell of claim 2, wherein the exogenous agent increases production of the heterologous product.

6. The host cell of claim 5, wherein the exogenous agent is galactose and expression of one or more enzymes encoded by the heterologous genetic pathway are under control of a GAL promoter.

7. The host cell of claim 1, wherein the heterologous genetic pathway comprises a galactose-responsive promoter, a maltose-responsive promoter, or a combination of both.

8. The host cell of claim 1, wherein the heterologous product is a cannabinoid or cannabinoid precursor.

9. The host cell of claim 8, wherein the cannabinoid or cannabinoid precursor is CBDA, CBD, CBGA, or CBG.

10. The host cell of claim 8, wherein the genetic pathway encodes at least two enzymes selected from the group consisting of hexanoyl-CoA synthase (HCS), tetraketide synthase (TKS) and olivetolic acid cyclase (OAC).

11. The host cell of claim 8, wherein the precursor required to make the product is hexanoate.

12. The host cell of claim 1, wherein the heterologous genetic pathway comprises a nucleic acid construct comprising at least 3 protein coding regions.

13. The host cell of claim 1, wherein the host cell is a yeast cell or yeast strain.

14. The host cell of claim 13, wherein the yeast cell is *S. cerevisiae.*

15. A mixture comprising the host cell of any one of claims 1-14 and a culture media.

16. The mixture of claim 15, wherein the culture media comprises an exogenous agent that decreases production of the heterologous product.

17. The mixture of claim 16, wherein the exogenous agent is glucose, maltose, or lysine.

18. The mixture of claim 15, wherein the culture media comprises (i) an exogenous agent that increases production of the heterologous product, and (ii) a precursor required to make the heterologous product.

19. The mixture of claim 18, wherein the exogenous agent is galactose.

20. The mixture of claim 18, wherein the precursor required to make the heterologous product is hexanoate.

21. A method for decreasing the expression of a heterologous product, comprising culturing the host cell of any one of claims 1-14 in a media comprising the exogenous agent, wherein the exogenous agent decreases the expression of the heterologous product.

22. The method of claim 21, wherein the exogenous agent is glucose, maltose, or lysine.

23. The method of claim 21, wherein culturing the host cell in the media comprising the exogenous agent results in less than 0.001 mg/L of heterologous product

24. A method for increasing the expression of a heterologous product, comprising culturing the host cell of any one of claims 1-14 in a media comprising the exogenous agent, wherein the exogenous agent increases expression of the heterologous product.

25. The method of claim 24, wherein the exogenous agent is galactose.

26. The method of claim 24, further comprising culturing the host cell with the precursor required to make the heterologous product.

27. The method of claim 24, wherein the precursor required to make the heterologous product is hexanoate.

28. A host cell comprising a heterologous genetic pathway that produces a cannabinoid and is regulated by an exogenous agent, wherein the host cell does not comprise hexanoate at a level to make the cannabinoid in an amount over 10 mg/L.

29. The host cell of claim 28, wherein the exogenous agent downregulates expression of the heterologous genetic pathway.

30. The host cell of claim 29, wherein the exogenous agent is glucose and expression of one or more enzymes encoded by the heterologous genetic pathway are under control of a glucose repressed promoter.

31. The host cell of claim 28, wherein the exogenous agent upregulates expression of the heterologous genetic pathway.

32. The host cell of claim 31, wherein the exogenous agent is galactose and expression of one or more enzymes encoded by the heterologous genetic pathway are under control of a GAL promoter.

33. The host cell of claim 28, wherein the genetic pathway encodes at least two enzymes selected from the group consisting of hexanoyl-CoA synthase (HCS), tetraketide synthase (TKS) and olivetolic acid cyclase (OAC).

34. The host cell of claims 28-33, wherein the host cell is a yeast cell or yeast strain.

35. The host cell of claim 34, wherein the yeast is *S. cerevisiae.*

36. A method for decreasing expression of a cannabinoid, comprising culturing the host cell of any one of claims 28-30 and 33-35 in a media comprising the exogenous agent, wherein the exogenous agent decreases the expression of the cannabinoid or a precursor thereof.

37. The method of claim 36, wherein the exogenous agent is glucose, maltose, or lysine.

38. The method of claim 36, wherein culturing the host cell in the media comprising the exogenous agent results in less than 0.001 mg/L of cannabinoid or a precursor thereof.

39. A method for increasing expression of a cannabinoid, comprising culturing the host cell of any one of claims 31-35 in a media comprising the exogenous agent, wherein the exogenous agent increases the expression of the cannabinoid or a precursor thereof.

40. The method of claim 39, wherein the exogenous agent is galactose.

41. The method of claim 39, further comprising culturing the host cell in a media comprising hexanoate.

42. The method of any one of claims 36-41, wherein the cannabinoid is CBDA, CBD, CBGA, or CBG.

43. The method of any one of claims 36-41, wherein the host cell is a yeast cell or yeast strain.

44. The method of claim 43, wherein the yeast is *S. cerevisiae.*

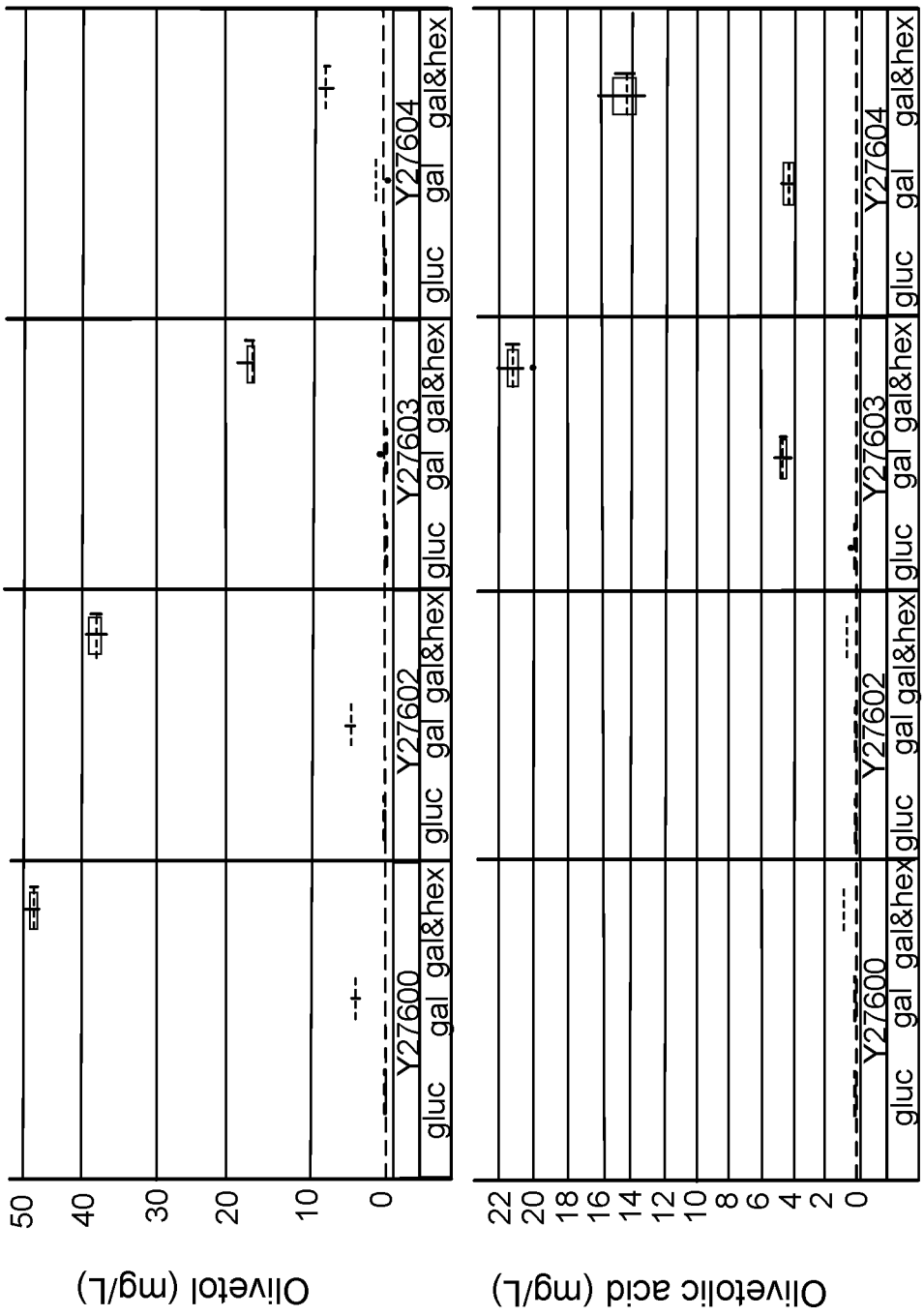
Oliveol (mg/L)
50
40
30
20
10
0
Y27600
gluc gal gal&hex
Y27602
gluc gal gal&hex
Y27603
gluc gal gal&hex
Y27604
gluc gal gal&hex
Oliveolic acid (mg/L)
22
20
18
16
14
12
10
8
6
4
2
0
Y27600
gluc gal gal&hex
Y27602
gluc gal gal&hex
Y27603
gluc gal gal&hex
Y27604
gluc gal gal&hex

FIG. 1

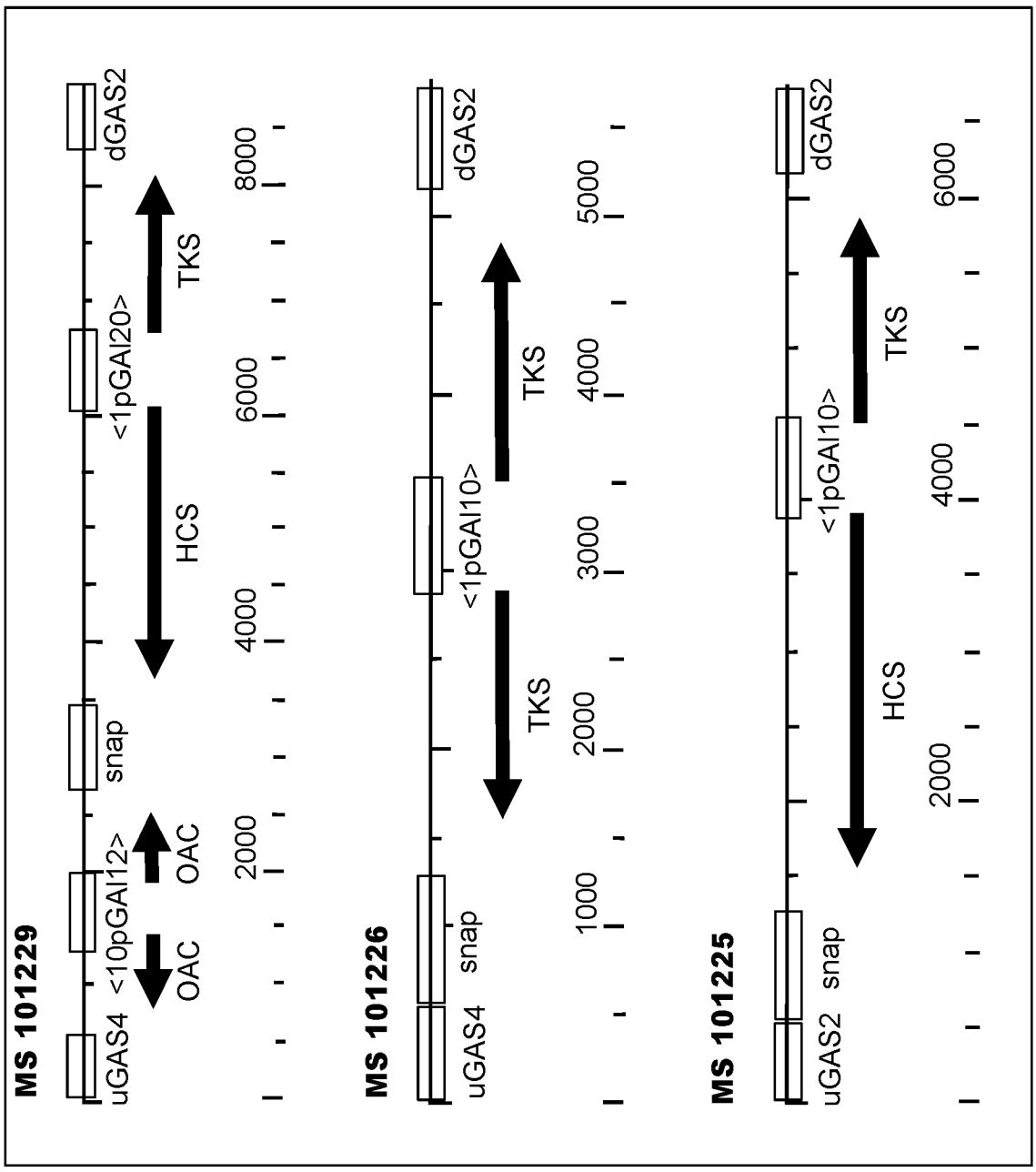
MS 101229
uGAS4
<10pGAI12>
snap
<1pGAI20>
dGAS2
OAC
OAC
HCS
TKS
2000
4000
6000
8000
MS 101226
uGAS4
snap
<1pGAI10>
dGAS2
TKS
TKS
1000
2000
3000
4000
5000
MS 101225
uGAS2
snap
<1pGAI10>
dGAS2
HCS
TKS
2000
4000
6000

FIG. 2

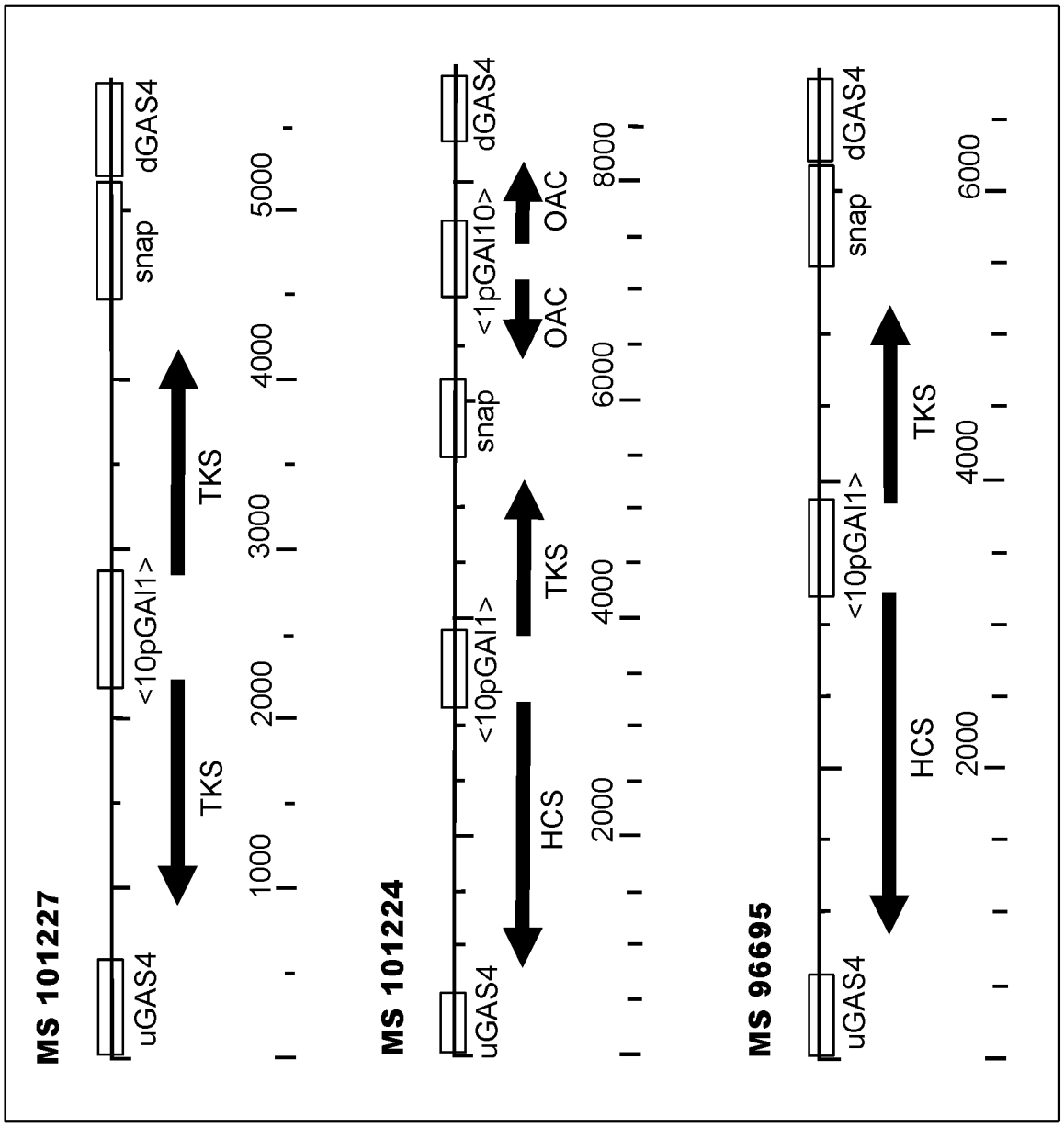
MS 101227
uGAS4
<10pGAl1>
snap
dGAS4
TKS
TKS
1000
2000
3000
4000
5000
MS 101224
uGAS4
<10pGAl1>
snap
<1pGAl10>
dGAS4
HCS
TKS
OAC
OAC
2000
4000
6000
8000
MS 96695
uGAS4
<10pGAl1>
snap
dGAS4
HCS
TKS
2000
4000
6000

FIG. 3

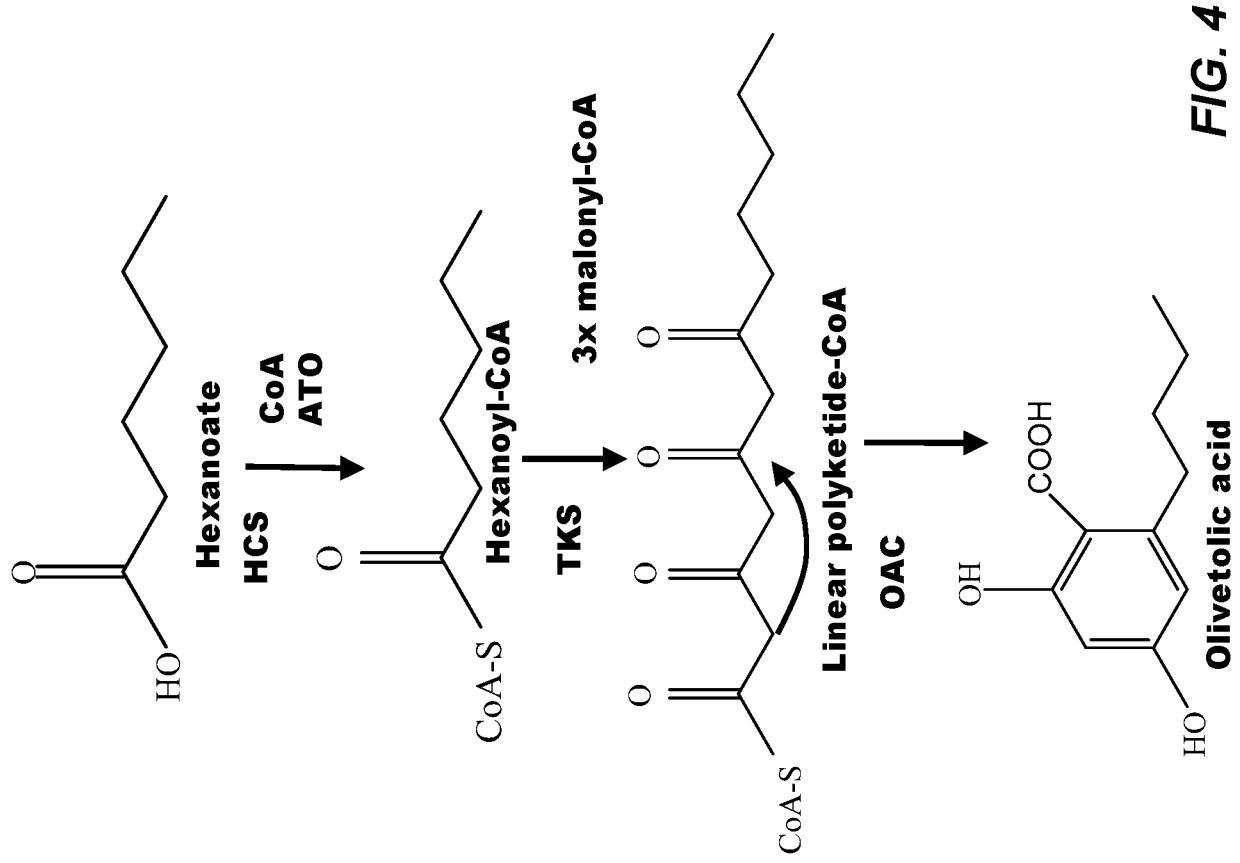
Hexanoate
HCS
CoA
ATO
Hexanoyl-CoA
TKS
3x malonyl-CoA
Linear polyketide-CoA
OAC
Olivetolic acid
HO
O
CoA-S
OH
COOH

FIG. 4

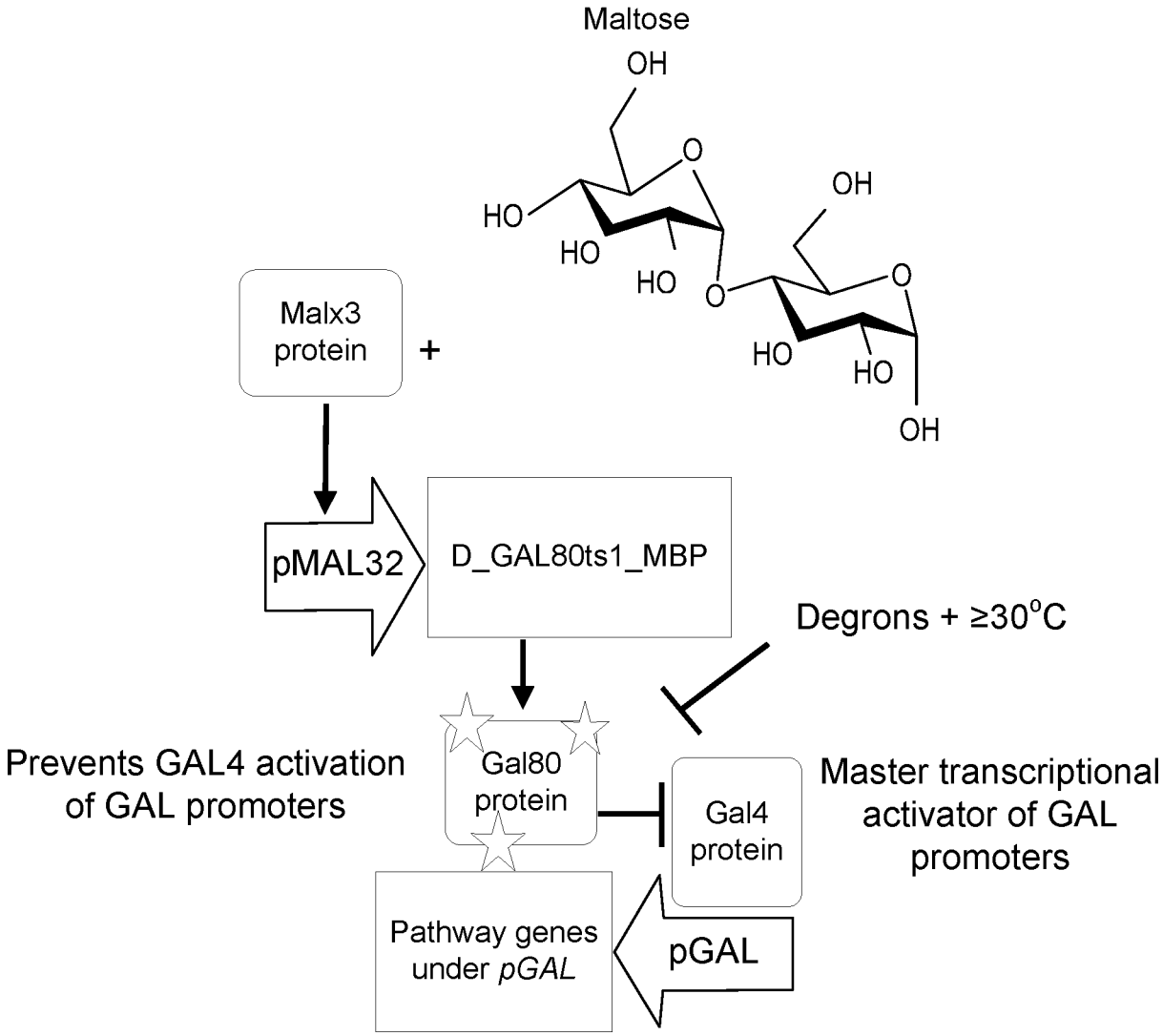
Maltose
OH
O
HO
HO
HO
O
OH
O
HO
HO
OH
Malx3 protein
+
pMAL32
D_GAL80ts1_MBP
Degrons + ≥30°C
Prevents GAL4 activation of GAL promoters
Gal80 protein
Gal4 protein
Master transcriptional activator of GAL promoters
Pathway genes under pGAL
pGAL

FIG. 5

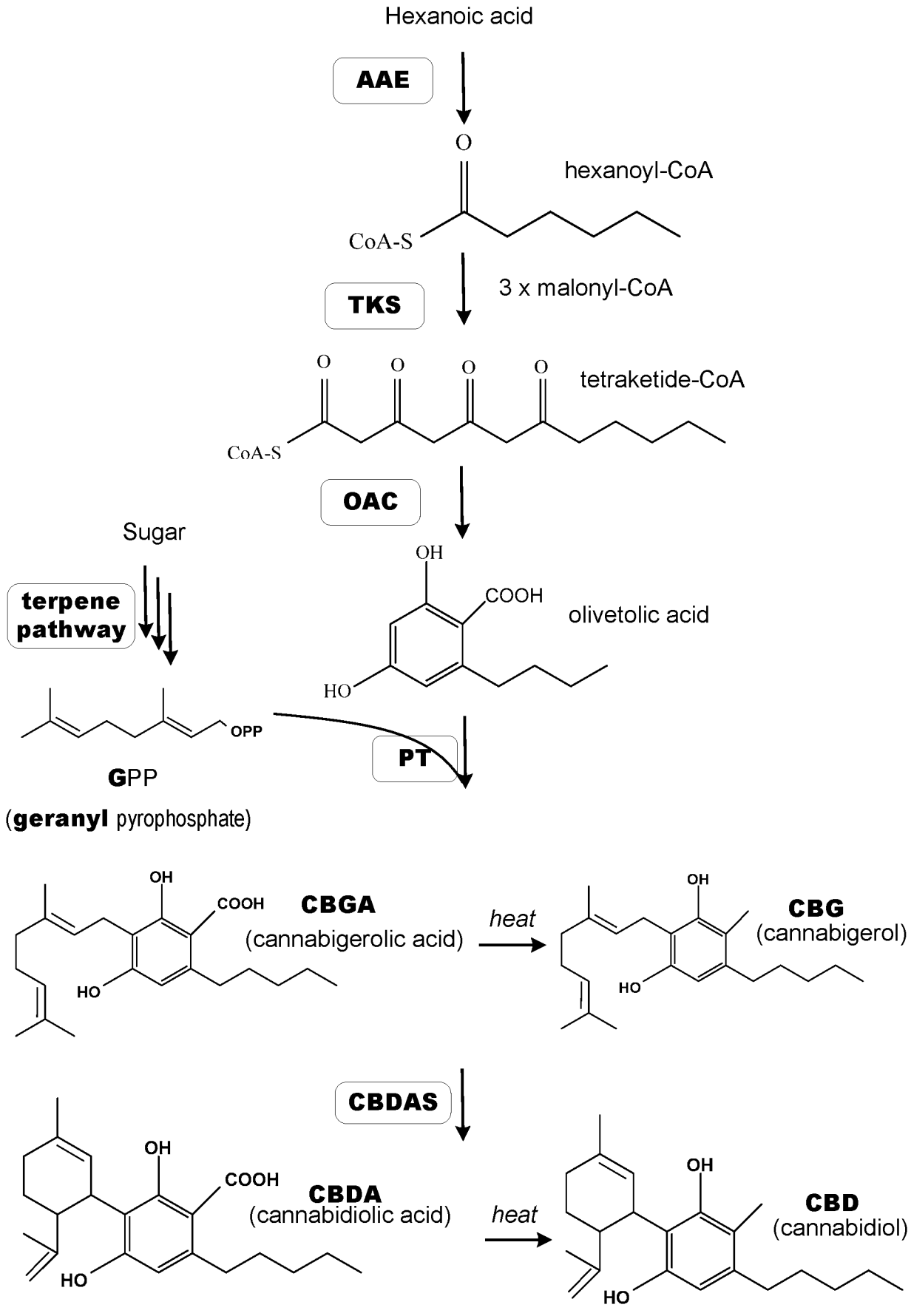
Hexanoic acid
AAE
O
hexanoyl-CoA
CoA-S
TKS
3 x malonyl-CoA
O O O O
tetraketide-CoA
CoA-S
OAC
Sugar
terpene pathway
OH
COOH
olivetolic acid
HO
OPP
PT
GPP
(geranyl pyrophosphate)
OH
COOH
CBGA
(cannabigerolic acid)
heat
HO
OH
CBG
(cannabigerol)
HO
CBDAS
OH
COOH
CBDA
(cannabidiolic acid)
heat
HO
OH
CBD
(cannabidiol)
HO

FIG. 6

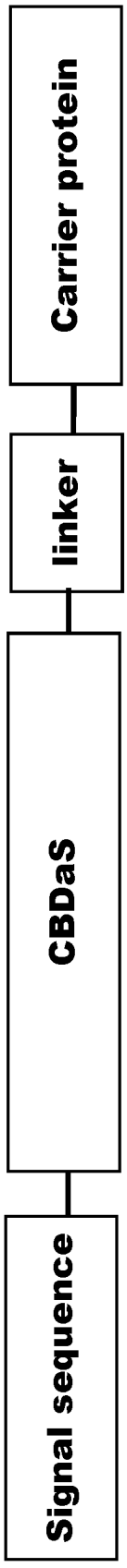

FIG. 7

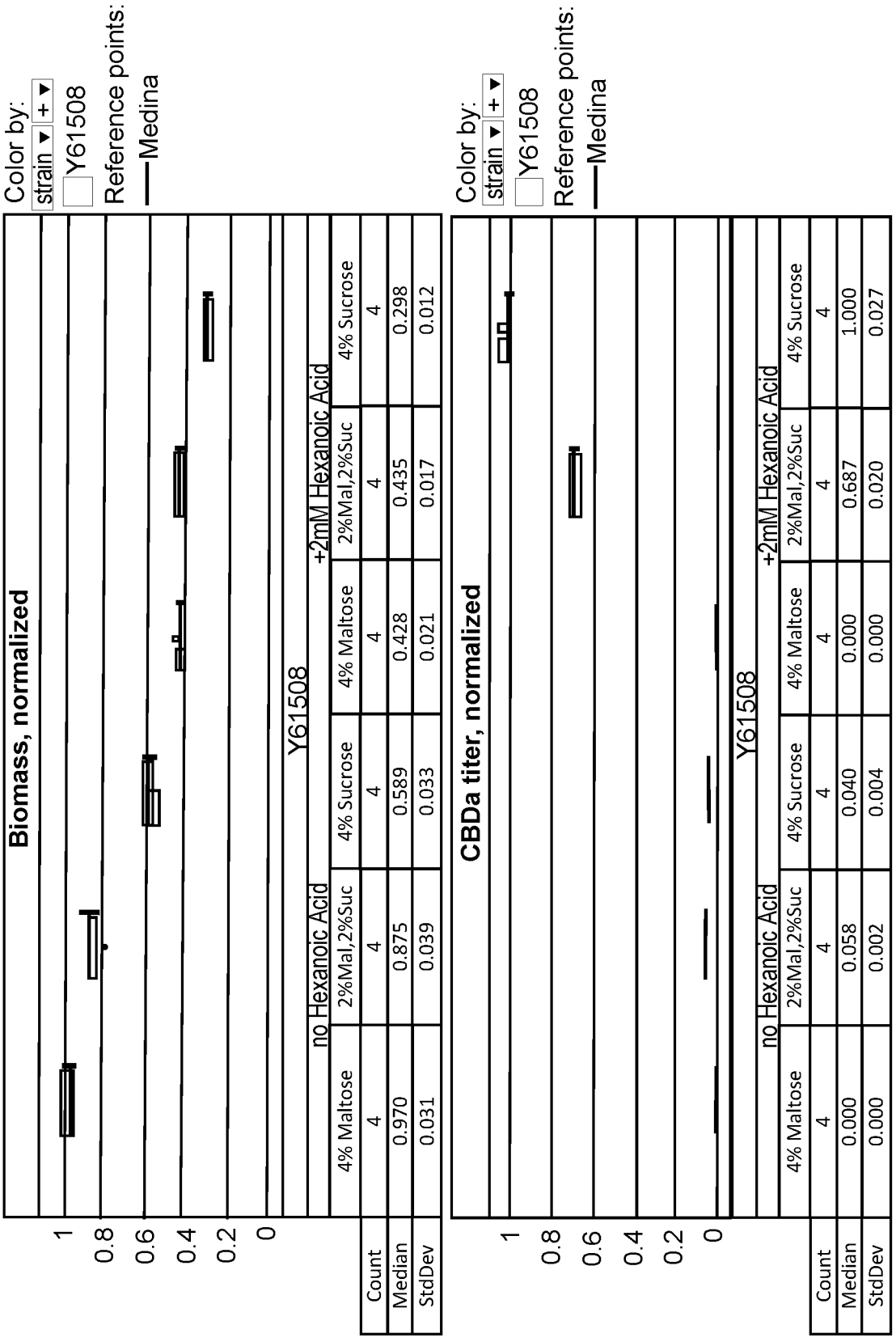
Biomass, normalized
Color by:
strain
Y61508
Reference points:
Medina
1
0.8
0.6
0.4
0.2
0
Y61508
no Hexanoic Acid
+2mM Hexanoic Acid
4% Maltose
2%Mal,2%Suc
4% Sucrose
4% Maltose
2%Mal,2%Suc
4% Sucrose
Count 4 4 4 4 4 4
Median 0.970 0.875 0.589 0.428 0.435 0.298
StdDev 0.031 0.039 0.033 0.021 0.017 0.012
CBDa titer, normalized
Color by:
strain
Y61508
Reference points:
Medina
1
0.8
0.6
0.4
0.2
0
Y61508
no Hexanoic Acid
+2mM Hexanoic Acid
4% Maltose
2%Mal,2%Suc
4% Sucrose
4% Maltose
2%Mal,2%Suc
4% Sucrose
Count 4 4 4 4 4 4
Median 0.000 0.058 0.040 0.000 0.687 1.000
StdDev 0.000 0.002 0.004 0.000 0.020 0.027

FIG. 8

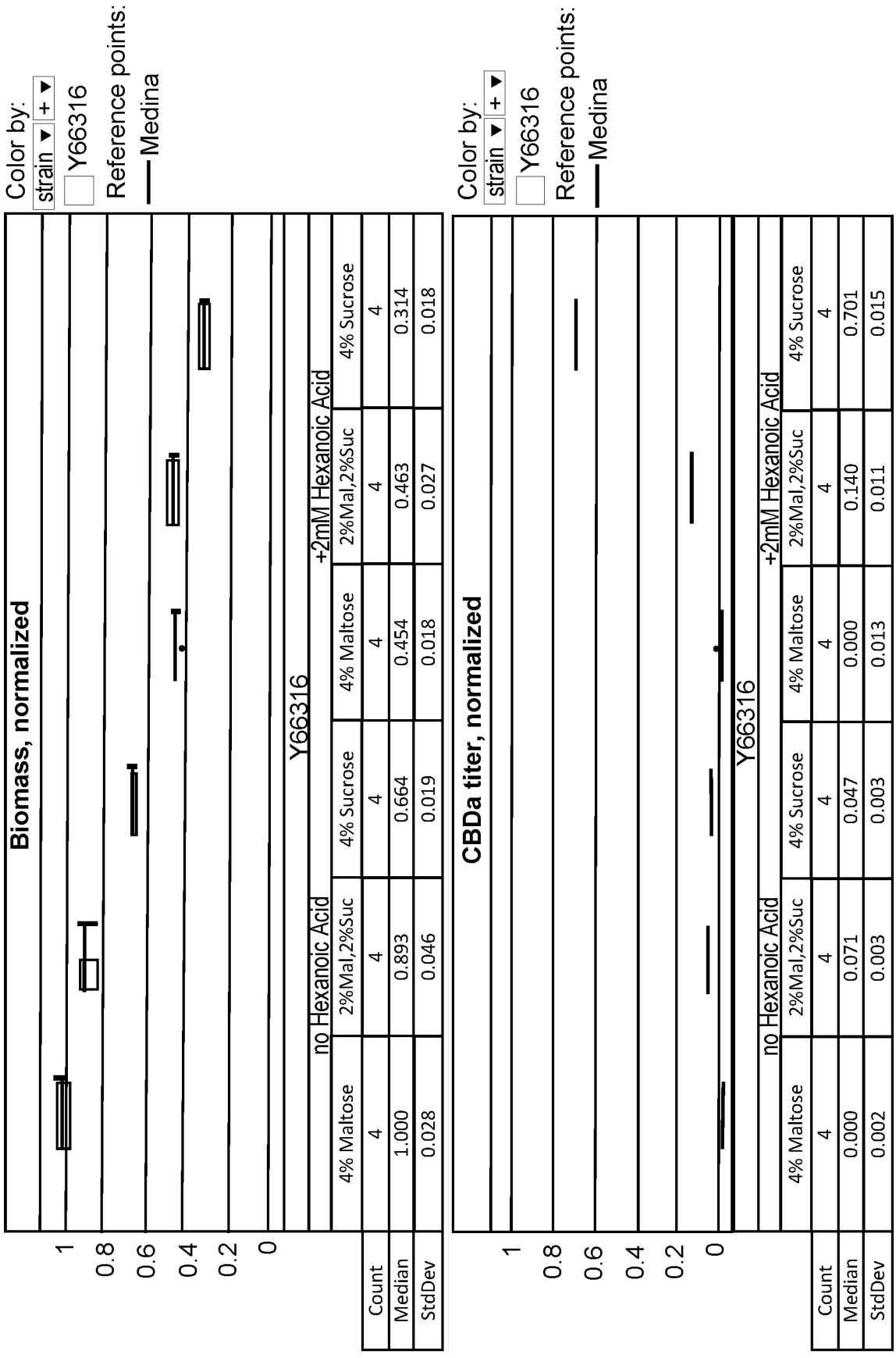
Biomass, normalized
Color by:
strain
Y66316
Reference points:
Medina
1
0.8
0.6
0.4
0.2
0
Y66316
no Hexanoic Acid
+2mM Hexanoic Acid
4% Maltose
2%Mal,2%Suc
4% Sucrose
Count 4 4 4 4 4 4
Median 1.000 0.893 0.664 0.454 0.463 0.314
StdDev 0.028 0.046 0.019 0.018 0.027 0.018
CBDa titer, normalized
Color by:
strain
Y66316
Reference points:
Medina
1
0.8
0.6
0.4
0.2
0
Y66316
no Hexanoic Acid
+2mM Hexanoic Acid
4% Maltose
2%Mal,2%Suc
4% Sucrose
Count 4 4 4 4 4 4
Median 0.000 0.071 0.047 0.000 0.140 0.701
StdDev 0.002 0.003 0.003 0.013 0.011 0.015

FIG. 9

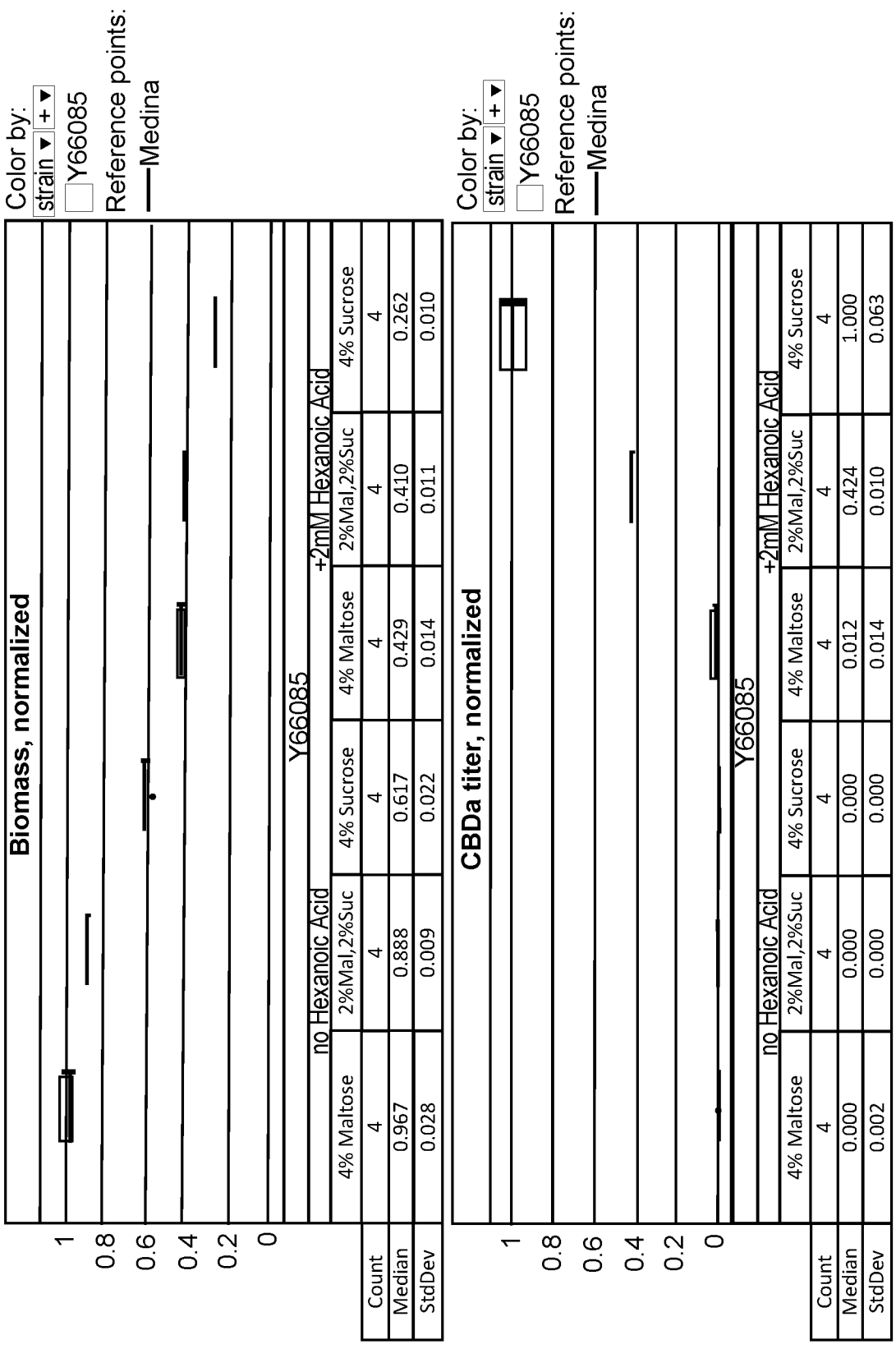
Biomass, normalized
Color by:
strain
Y66085
Reference points:
Medina
1
0.8
0.6
0.4
0.2
0
Y66085
no Hexanoic Acid
+2mM Hexanoic Acid
4% Maltose
2%Mal,2%Suc
4% Sucrose
4% Maltose
2%Mal,2%Suc
4% Sucrose
Count 4 4 4 4 4 4
Median 0.967 0.888 0.617 0.429 0.410 0.262
StdDev 0.028 0.009 0.022 0.014 0.011 0.010
CBDa titer, normalized
Color by:
strain
Y66085
Reference points:
Medina
1
0.8
0.6
0.4
0.2
0
Y66085
no Hexanoic Acid
+2mM Hexanoic Acid
4% Maltose
2%Mal,2%Suc
4% Sucrose
4% Maltose
2%Mal,2%Suc
4% Sucrose
Count 4 4 4 4 4 4
Median 0.000 0.000 0.000 0.012 0.424 1.000
StdDev 0.002 0.000 0.000 0.014 0.010 0.063

FIG. 10

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2020/022741

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12P7/42
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)
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, Sequence Search, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2018/200888 A1 (UNIV CALIFORNIA [US]) 1 November 2018 (2018-11-01) the whole document -----	1-44
T	Jeffrey S Flick ET AL: "Two Systems of Glucose Repression of the GAL] Promoter in Saccharomyces cerevisiae", MOLECULAR AND CELLULAR BIOLOGY, 1 January 1990 (1990-01-01), pages 4757-4769, XP055712941, Retrieved from the Internet: URL:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC361077/ [retrieved on 2020-07-08] the whole document -----	
	-/--	

[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
8 July 2020	21/07/2020

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	Keller, Yves

1

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/US202O/022741

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ÂNGELA CARVALHO ET AL: "Designing microorganisms for heterologous biosynthesis of cannabinoids", FEMS YEAST RESEARCH, vol. 17, no. 4, 1 June 2017 (2017-06-01), XP055658978, DOI: 10.1093/femsyr/fox037 the whole document -----	1-44
T	US 2016/010126 A1 (POULOS JASON L [US] ET AL) 14 January 2016 (2016-01-14) the whole document -----	

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2020/022741

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2018200888 A1	01-11-2018	AU 2018256863 A1	14-11-2019
		BR 112019022500 A2	16-06-2020
		CA 3061718 A1	01-11-2018
		CN 110914416 A	24-03-2020
		EP 3615667 A1	04-03-2020
		JP 2020517293 A	18-06-2020
		SG 11201910019P A	28-11-2019
		US 2019300888 A1	03-10-2019
		US 2020172917 A1	04-06-2020
		WO 2018200888 A1	01-11-2018
US 2016010126 A1	14-01-2016	CA 2990071 A1	21-01-2016
		US 2016010126 A1	14-01-2016
		US 2018073043 A1	15-03-2018
		US 2018371507 A1	27-12-2018
		WO 2016010827 A1	21-01-2016