



(51) International Patent Classification:

C12N 1/10 (2006.01) C12N 15/79 (2006.01)
C12N 15/11 (2006.01)

(21) International Application Number:

PCT/US2017/041949

(22) International Filing Date:

13 July 2017 (13.07.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/361,741 13 July 2016 (13.07.2016) US

(71) Applicant: DSM IP ASSETS B.V. [NL/NL]; Het Overloon 1, 6411 TE Heerlen (NL).

(72) Inventors; and

(71) Applicants (for US only): LIPPMEIER, James, Casey [US/US]; 5364 Racegate Run, Columbia, MD 21045 (US). BETZ, Yelena [US/US]; 10850 Harmel Drive, Columbia, MD 21044 (US). ROUBOS, Johannes, Andries [NL/NL]; Freule Wittewaall van Stoetwegensingel 45, 2642 DB Pijnacker (NL). VERWAAL, Rene [NL/NL]; Marsdiep 7, 2631 NL Nootdorp (NL).

(74) Agent: CHEN, Xi; 6480 Dobbin Road, Columbia, MD 21045 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available):

AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,

SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available):

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

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: A CRISPR-CAS SYSTEM FOR AN ALGAL HOST CELL

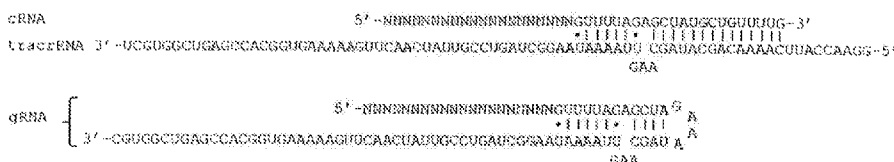


FIG. 1

(57) Abstract: The present invention relates to the field of molecular biology and cell biology. More specifically, the present invention relates to a CRISPR-Cas system for a Labyrinthulomycetes host cell.



A CRISPR-CAS SYSTEM FOR AN ALGAL HOST CELL

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of the filing date of United States Provisional Patent Application No. 62/361,741 filed July 13, 2016, the disclosure of which is hereby incorporated herein by reference in its entirety.

BACKGROUND

Technical Field

[0002] The present disclosure relates to the field of molecular biology and cell biology. More specifically, the present invention relates to a CRISPR-Cas system for an algal Labyrinthulomycetes host cell.

Description of Related Art

[0003] Recent advances in genomic techniques and analysis methods have significantly accelerated the ability to catalog and map genetic factors associated with a diverse range of biological functions and diseases. Precise genome engineering technologies are needed to enable systematic reverse engineering of causal genetic variations by allowing selective perturbation of individual genetic elements, as well as to advance synthetic biology, biotechnological, and medical applications. Although genome-editing techniques such as designer zinc fingers, transcription activator-like effectors nucleases (TALENs), or homing meganucleases are available for producing targeted genome perturbations, a need remains for new genome engineering technologies that are affordable, easy to set up, scalable, and amenable to targeting multiple positions within a genome. The engineering of meganucleases has been challenging for most academic researchers because the DNA recognition and cleavage functions of these enzymes are intertwined in a single domain. Robust construction of engineered zinc finger arrays has also proven to be difficult for many laboratories because of the need to account for context-dependent effects between individual finger domains in an array. There thus exists a pressing need for alternative and robust techniques for targeting of specific sequences within a host cell with a wide

array of applications. The solution to this technical problem is provided by the embodiments characterized in the claims.

BRIEF SUMMARY

[0004] The present application is based on the CRISPR-Cas system, which does not require the generation of customized proteins to target-specific sequences, but rather a single Cas enzyme that can be programmed by a guide-polynucleotide to recognize a specific polynucleotide target; in other words, the Cas enzyme can be recruited to a specific polynucleotide target using said guide-polynucleotide molecule. Adding the CRISPR-Cas system to the repertoire of genomics techniques and analysis methods may significantly simplify existing methodologies in the field of molecular biology.

[0005] The present invention provides a non-naturally occurring or engineered composition comprising a source of a CRISPR-Cas system comprising a guide-polynucleotide and a Cas protein, wherein the guide-polynucleotide comprises a sequence that essentially is the reverse complement of a target-polynucleotide in a host cell and the guide-polynucleotide can direct binding of the Cas protein at the target-polynucleotide in the host cell to form a CRISPR-Cas complex.

[0006] The present invention further relates to a method of modulating expression of a polynucleotide in a cell, comprising contacting a host cell with the composition according to the present invention, wherein the guide-polynucleotide directs binding of the Cas protein at the target-polynucleotide in the host cell to form a CRISPR-Cas complex.

[0007] The present invention further relates to a host cell comprising a composition according to the present invention.

[0008] The present invention further relates to a method of producing a host cell, comprising contacting a host cell with the composition according to the present invention, wherein the guide-polynucleotide directs binding of the Cas protein at the target-polynucleotide in the host cell to form a CRISPR-Cas complex.

[0009] The present invention further relates to a method for the production of a compound of interest, comprising culturing under conditions conducive to the

compound of interest a host cell according to the present invention and optionally purifying or isolating the compound of interest.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] For a further understanding of the nature, objects, and advantages of the present disclosure, reference should be had to the following detailed description, read in conjunction with the following drawings, wherein like reference numerals denote like elements.

[0011] FIG. 1 depicts examples of typical guide-polynucleotides. Both guide-polynucleotide are guide-RNA's comprising a guide-sequence (crRNA) and a guide-polynucleotide structural component. In the upper figure, the guide-polynucleotide structural component is comprised of two separate molecules hybridized to each other; the individual components may be referred to as a tracr sequence and a tracr-mate sequence. In the lower figure, the guide-polynucleotide structural component is comprised of a single molecule with internal hybridization. This figure is adapted from Sander and Joung, 2014 and Mali *et al.*, Nat Methods.10(10):957-63, 2013.

[0012] FIG. 2 depicts how the guide-polynucleotide (guide RNA self-processing ribozymes abbreviated as gRSR) is built. The Hammerhead ribozyme and HDV ribozyme cleave the RNA molecule forming the final and functional mature guide-polynucleotide (guide-RNA).

[0013] FIG. 3 shows digested plasmids and PCR-amplified fragments run on a 1% agarose gel.

[0014] FIG. 4 shows results of colony PCR of transformants resulting from pYB32 and pYB33 bacterial transformations.

[0015] FIG. 5 shows an agarose gel in which amplicons derived from genomic DNA of T188 and T189 transformants were separated, testing for the presence of promoter and partial Cas9.

[0016] FIG. 6 shows amplicons derived from genomic DNA of T188 and T189 transformants run on an agarose gel verifying integration of Cas9 to the OrfA/Pfa1 locus.

[0017] FIG. 7 shows a portion of *Schizochytrium* fatty acid synthase (FAS) gene amplified from genomic DNA and run on a 1% agarose gel.

[0018] FIG. 8 shows the results of PCR cloning of paromomycin and gRNA cassettes into FAS locus in vector pCL400.

[0019] FIG. 9 shows the results of PCR performed to amplify the entire region encompassing the first 3 CS gRNA targets as one amplicon and another PCR to amplify gRNA3 CS4 target separately.

[0020] FIG. 10 shows the results of PCR performed to amplify gRNA3 CS1 target separately from genomic DNA of T206 transformants.

[0021] FIG. 11 shows the results of a digest of pLC122-Cas9 and pYB30 plasmids.

[0022] FIG. 12 shows the results of colony PCR of transformants resulting from pYB61 bacterial transformations.

[0023] FIG. 13 shows the results of a digest of pCL122 plasmid on a 1% agarose gel.

[0024] FIG. 14 shows an amplicon of gRNA3 CS1 cassette for pYB66 cloning on a 1% agarose gel,

[0025] FIG. 15 shows the results of a digestion of gRNA3 CS1 (pYB66 fragment) on a 1% agarose gel.

[0026] FIG. 16 depicts the results of colony PCR of transformants resulting from pYB66 bacterial transformation.

[0027] FIG. 17 shows the result of digestion of pYB61 plasmid on a 1% agarose gel.

[0028] FIG. 18 shows an amplicon of gRNA3 CS1 cassette for pYB73 cloning on a 1% agarose gel.

[0029] FIG. 19 shows the results of colony PCR of transformants resulting from pYB73 bacterial transformation.

[0030] FIG. 20 shows an agarose gel in which amplicons derived from genomic DNA of T212 transformants were separated, testing for the presence of a CarG gene.

[0031] FIG. 21 shows an agarose gel in which amplicons derived from genomic DNA of T280, T281, T285, and T286 transformants were separated, testing for the presence of a partial promoter and Cas9 sequences.

[0032] FIG. 22 shows an agarose gel in which amplicons derived from genomic DNA of T281, and T286 transformants were separated, testing for the presence of a gRNA cassette.

[0033] FIG. 23 shows an agarose gel in which amplicons derived from genomic DNA of T282, and T287 transformants were separated, testing for the presence of a partial gRNA cassette.

[0034] FIG. 24 shows the results of PCR performed to amplify gRNA3 CS1 target separately from genomic DNA of T281, T282, T286, and T287 transformants

DESCRIPTION OF THE SEQUENCE LISTING

[0035] SEQ ID NO:1 sets out the pCL122-Cas9 vector nucleotide sequence.

[0036] SEQ ID NO:2 sets out the pYB31 vector nucleotide sequence.

[0037] SEQ ID NO:3 sets out the pYB32 vector nucleotide sequence.

[0038] SEQ ID NO:4 sets out the pYB33 vector nucleotide sequence.

[0039] SEQ ID NO:5 sets out the pCL399 vector nucleotide sequence.

[0040] SEQ ID NO:6 sets out the pCL400 vector nucleotide sequence.

[0041] SEQ ID NO:7 sets out the pCL401 vector nucleotide sequence.

[0042] SEQ ID NO:8 sets out the pCL402 vector nucleotide sequence.

[0043] SEQ ID NO:9 sets out the pYB36 vector nucleotide sequence.

[0044] SEQ ID NO:10 sets out the pYB37 vector nucleotide sequence.

[0045] SEQ ID NO:11 sets out the pYB38 vector nucleotide sequence.

[0046] SEQ ID NO:12 sets out the pYB39 vector nucleotide sequence.

[0047] SEQ ID NO:13 sets out the 121 Tub seq F primer nucleotide sequence.

[0048] SEQ ID NO:14 sets out the pYB32/3 C R1 primer nucleotide sequence.

[0049] SEQ ID NO:15 sets out the CS pro Kpn IF F1 primer nucleotide sequence.

[0050] SEQ ID NO:16 sets out the CS pro BamH IF R1 primer nucleotide sequence.

[0051] SEQ ID NO:17 sets out the CS pro BamH IF F2 primer nucleotide sequence.

[0052] SEQ ID NO:18 sets out the CS pro Nde IF R2 primer nucleotide sequence.

- [0053]** SEQ ID NO:19 sets out the O A1-KO F primer nucleotide sequence.
- [0054]** SEQ ID NO:20 sets out the pYB32/3 SV40 R1 primer nucleotide sequence.
- [0055]** SEQ ID NO:21 sets out the O A1-KO R primer nucleotide sequence.
- [0056]** SEQ ID NO:22 sets out the pYB32/3 C F1 primer nucleotide sequence.
- [0057]** SEQ ID NO:23 sets out the 5' FAS PmeNde primer nucleotide sequence.
- [0058]** SEQ ID NO:24 sets out the 3' FAS PmeHpa primer nucleotide sequence.
- [0059]** SEQ ID NO:25 sets out the pCL402 IF F primer nucleotide sequence.
- [0060]** SEQ ID NO:26 sets out the pCL402 IF R primer nucleotide sequence.
- [0061]** SEQ ID NO:27 sets out the pYB36 CS1 F primer nucleotide sequence.
- [0062]** SEQ ID NO:28 sets out the pYB36 CS1 R primer nucleotide sequence.
- [0063]** SEQ ID NO:29 sets out the pYB36 CS3 R primer nucleotide sequence.
- [0064]** SEQ ID NO:30 sets out the pYB36 CS4 F primer nucleotide sequence.
- [0065]** SEQ ID NO:31 sets out the pYB36 CS4 R primer nucleotide sequence.
- [0066]** SEQ ID NO:32 sets out the pYB30 vector nucleotide sequence.
- [0067]** SEQ ID NO:33 sets out the pYB61 vector nucleotide sequence.
- [0068]** SEQ ID NO:34 sets out the pYB66 vector nucleotide sequence.
- [0069]** SEQ ID NO:35 sets out the pYB73 vector nucleotide sequence.
- [0070]** SEQ ID NO:36 sets out the pCL310 vector nucleotide sequence.
- [0071]** SEQ ID NO:37 sets out the pCL122 vector nucleotide sequence.
- [0072]** SEQ ID NO:38 sets out the pYB66 BamBgl F primer nucleotide sequence.
- [0073]** SEQ ID NO:39 sets out the pYB66 Nde R primer nucleotide sequence.
- [0074]** SEQ ID NO:40 sets out the pYB66 EF1seq F primer nucleotide sequence.
- [0075]** SEQ ID NO:41 sets out the pCL122 OrfC R primer nucleotide sequence.
- [0076]** SEQ ID NO:42 sets out the pYB73 gRNA Pst Kpn IF F primer nucleotide sequence.
- [0077]** SEQ ID NO:43 sets out the pYB73 gRNA Xho Pst IF R primer nucleotide sequence.
- [0078]** SEQ ID NO:44 sets out the pYB73 seq F primer nucleotide sequence.
- [0079]** SEQ ID NO:45 sets out the pYB73 seq R primer nucleotide sequence.

[0080] SEQ ID NO:46 sets out the pYB13 pYB1 seq F primer nucleotide sequence.

[0081] SEQ ID NO:47 sets out the TT pYB73 HDV R primer nucleotide sequence.

DETAILED DESCRIPTION

[0082] Before the subject disclosure is further described, it is to be understood that the disclosure is not limited to the particular embodiments of the disclosure described below, as variations of the particular embodiments may be made and still fall within the scope of the appended claims. It is also to be understood that the terminology employed is for the purpose of describing particular embodiments, and is not intended to be limiting. Instead, the scope of the present disclosure will be established by the appended claims.

[0083] In a first aspect, the present invention provides a non-naturally occurring or engineered composition comprising a source of a CRISPR-Cas system comprising a guide-polynucleotide and a Cas protein, wherein the guide-polynucleotide comprises a guide-sequence that essentially is the reverse complement of a target-polynucleotide in a host cell and the guide-polynucleotide can direct binding of the Cas protein at the target-polynucleotide in the host cell to form a CRISPR-Cas complex, wherein the guide-sequence is essentially the reverse complement of the (N)y part of a 5'-(N)yPAM-3' polynucleotide sequence target in the genome of the host cell, wherein y is an integer of 8-30, more preferably 10-30, more preferably 15-30, more preferably 17-27, more preferably 17-20, more preferably 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, or 27, wherein PAM is a protospacer adjacent motif, wherein the host cell is of the Labyrinthulomycete class, preferably of the order Thraustochytriales, more preferably of the family Thraustochytriaceae, more preferably a member of a genus selected from the group consisting of *Aurantiochytrium*, *Oblongichytrium*, *Schizochytrium*, *Thraustochytrium*, and *Ulkenia*, even more preferably *Schizochytrium* sp. ATCC# 20888, and wherein PAM is preferably a sequence selected from the group consisting of 5'-XGG-3', 5'-XGGXG-3', 5'-XXAGAAW-3', 5'-XXXXGATT-3', 5'-XXAGAA-3', 5'-XAAAAC-3', wherein X can be any nucleotide or analog thereof, preferably X can be any nucleotide; and W is A or T.

[0084] The composition, source, CRISPR-Cas system, guide-polynucleotide, Cas protein, target-polynucleotide, host cell and CRISPR-Cas complex are herein referred to as a composition, source, CRISPR-Cas system, guide-polynucleotide, Cas protein, target-polynucleotide, host cell and CRISPR-Cas complex according to the present invention. For the sake of completeness, since “a” is defined elsewhere herein as “at least one”, a composition according to the present invention comprises a source of at least one, i.e. one, two, three or more guide-polynucleotides and/or at least one, i.e. one, two, three or more Cas proteins. Accordingly, the present invention conveniently provides for a multiplex CRISPR-Cas system. Such multiplex CRISPR-Cas system can conveniently be used for introduction of a donor polynucleotide, deletion of a polynucleotide and polynucleotide library insertion into the genome of a host cell. Herein, a multiplex CRISPR-Cas system may refer to the use of one or more Cas proteins, one or more guide-polynucleotides and/or one or more donor polynucleotides.

[0085] The terms “CRISPR system”, “CRISPR-Cas system” and “CRISPR enzyme system” are used interchangeably herein and refer in the context of all embodiments of the present invention to a collection of elements required to form, together with a target-polynucleotide, a CRISPR-Cas complex; these elements comprise but are not limited to a Cas protein and a guide-polynucleotide.

[0086] The term “CRISPR-Cas complex” refers in the context of all embodiments of the present invention to a complex comprising a guide-polynucleotide hybridized to a target-polynucleotide and complexed with a Cas protein. In the most straightforward form, where a non-mutated Cas protein is used such as but not limited to the Cas9 protein of *Streptococcus pyogenes*, the formation of the CRISPR-Cas complex results in cleavage of one or both polynucleotide strands in or near (e.g. within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the target-polynucleotide. Typically, a target-polynucleotide according to the present invention (defined below herein) is associated with a PAM sequence (defined below herein) and the PAM sequence is preferably immediately downstream (3') of the target-polynucleotide; the formation of the CRISPR-Cas complex typically results in cleavage of one or both polynucleotide strands 3 base pairs upstream (5') of the PAM sequence.

[0087] The term “non-naturally occurring composition” refers in the context of all embodiments of the present invention to a composition that in its form used in the present invention does not occur in nature. The individual elements may e.g. occur as such or in combinations with other elements in nature, but the non-naturally occurring composition comprises e.g. at least one element more or less than a naturally composition.

[0088] The term “engineered composition” refers in the context of all embodiments of the present invention to a composition wherein at least one of the elements has been engineered, i.e. modified by man, in such a way that resulting element does not occur in nature. It follows that by virtue of comprising at least one engineered element, an engineered composition does not occur in nature.

[0089] The terms “polynucleotide”, “nucleotide sequence” and “nucleic acid” are used interchangeably herein and refer in the context of all embodiments of the present invention to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or mixes or analogs thereof. Polynucleotides may have any three-dimensional structure, and may perform any function, known or unknown. The following are non-limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA), short interfering RNA (siRNA), short-hairpin RNA (shRNA), micro-RNA (miRNA), ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, oligonucleotides and primers. A polynucleotide may comprise one or more modified nucleotides, such as a methylated nucleotide and a nucleotide analogue or nucleotide equivalent wherein a nucleotide analogue or equivalent is defined as a residue having a modified base, and/or a modified backbone, and/or a non-natural internucleoside linkage, or a combination of these modifications. Preferred nucleotide analogues and equivalents are described in the section “General definitions”. As desired, modifications to the nucleotide structure may be introduced before or after assembly of the polynucleotide. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling compound.

[0090] As used herein, a “polynucleotide” can comprise a conventional phosphodiester bond or a non-conventional bond (e.g., an amide bond, such as found in peptide nucleic acids (PNA)). A polynucleotide can contain the nucleotide sequence of the full-length cDNA sequence, including the untranslated 5' and 3' sequences, the coding sequences, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. The polynucleotide can be composed of any polyribonucleotide or polydeoxyribonucleotide, which can be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that can be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotides can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. Polynucleotides can contain ribonucleosides (adenosine, guanosine, uridine, or cytidine; “RNA molecules”) or deoxyribonucleosides (deoxyadenosine, deoxyguanosine, deoxythymidine, or deoxycytidine; “DNA molecules”), or any phosphoester analogs thereof, such as phosphorothioates and thioesters. Polynucleotides can also contain one or more modified bases or DNA or RNA backbones modified for stability or for other reasons. “Modified” bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, “polynucleotide” embraces chemically, enzymatically, or metabolically modified forms. The term nucleic acid molecule refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, inter alia, in linear or circular DNA molecules (e.g., restriction fragments), plasmids, and chromosomes. In discussing the structure of particular double-stranded DNA molecules, sequences can be described herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the non-transcribed strand of DNA (i.e., the strand having a sequence homologous to the mRNA).

[0091] The term “isolated” nucleic acid molecule refers to a nucleic acid molecule, DNA or RNA, which has been removed from its native environment. Further

examples of isolated nucleic acid molecules include nucleic acid molecules comprising recombinant polynucleotides maintained in heterologous host cells or purified (partially or substantially) polynucleotides in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of polynucleotides of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically. In addition, a nucleic acid molecule or polynucleotide can include a regulatory element such as a promoter, ribosome binding site, or a transcription terminator.

[0092] A “gene” refers to an assembly of nucleotides that encode a polypeptide, and includes cDNA and genomic DNA nucleic acids. “Gene” also refers to a nucleic acid fragment that expresses a specific protein, including intervening sequences (introns) between individual coding segments (exons), as well as regulatory sequences preceding (5′ non-coding sequences) and following (3′ non-coding sequences) the coding sequence. “Native gene” refers to a gene as found in nature with its own regulatory sequences.

[0093] In some embodiments, the nucleic acid molecules comprise polynucleotide sequences at least about 80%, 85%, or 90% identical to the polynucleotide sequences reported herein, or at least about 95%, 96%, 97%, 98%, 99%, or 100% identical to the polynucleotide sequences reported herein. The term “percent identity,” as known in the art, is a relationship between two or more amino acid sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, “identity” also means the degree of sequence relatedness between amino acid or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences.

[0094] By a nucleic acid molecule having a polynucleotide sequence at least, for example, 95% “identical” to a reference polynucleotide sequence of the present invention, it is intended that the polynucleotide sequence of the nucleic acid molecule is identical to the reference sequence except that the polynucleotide sequence can include up to five nucleotide differences per each 100 nucleotides of the reference polynucleotide sequence. In other words, to obtain a nucleic acid molecule having a polynucleotide sequence at least 95% identical to a reference polynucleotide sequence, up to 5% of the nucleotides in the reference sequence can be deleted or

substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence can be inserted into the reference sequence.

[0095] As a practical matter, whether any particular polynucleotide sequence or amino acid sequence is at least 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identical to a polynucleotide sequence or amino acid sequence of the present invention can be determined conventionally using known computer programs. A method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence can be determined using the alignment of sequences and calculation of identity scores. The alignments were done using the computer program Geneious (www.geneious.com) with a map to reference algorithm and Geneious reference assembler.

[0096] A guide-polynucleotide according to the present invention comprises at least a guide-sequence that is able to hybridize with the target-polynucleotide and is able to direct sequence-specific binding of the CRISPR-Cas system to the target-polynucleotide to form a CRISPR-Cas complex. In order to enable formation of an active CRISPR-Cas complex, the guide-polynucleotide preferably also comprises a sequence that has a specific secondary structure and allows binding of the Cas protein to the guide-polynucleotide. Such sequence is known in the art as tracr RNA, tracr sequence, tracr scaffold or guide-polynucleotide structural component, these terms are used interchangeably herein; wherein the tracr is the abbreviation for transactivating CRISPR; tracrRNA thus means transactivating CRISPR RNA. The tracrRNA in the original CRISPR-Cas system is the endogenous bacterial RNA that links the crRNA (guide-sequence) to the Cas nuclease, being able to bind any crRNA. A guide-polynucleotide structural component may be comprised of a single polynucleotide molecule or may be comprised of two or more molecules hybridized to each other; or two or more molecules which associate with Cas protein or other nucleases of similar function. Such components of a guide-polynucleotide structure may be referred to as a tracr sequence and a tracr-mate sequence.

[0097] Accordingly, the guide-polynucleotide preferably also comprises a tracr sequence and/or a tracr-mate sequence. The guide-polynucleotide is a polynucleotide according to the general definition of a polynucleotide set out here above; a preferred guide-polynucleotide comprises ribonucleotides, a more preferred guide-

polynucleotide is a RNA (guide-RNA). Two examples of guide-polynucleotide structures are depicted in FIG. 1.

[0098] In the context of the present invention, a sequence is referred to as essentially the reverse complement of a target-sequence or of a target-polynucleotide if the subject sequence is able to hybridize with the target-sequence or target-polynucleotide, preferably under physiological conditions as in a host cell. The degree of complementarity between a guide-sequence and its corresponding target-sequence, when optimally aligned using a suitable alignment algorithm, is preferably higher than 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99% sequence identity. When the target-polynucleotide is a double stranded polynucleotide, the subject sequence, such as a guide-sequence, may be able to hybridize with either strand of the target-polynucleotide e.g. a coding strand or a non-coding strand.

[0099] Preferably, a guide-sequence according to the present invention targets a target-sequence that is unique in the target. Preferably, a guide-sequence according to the present invention has 100% sequence identity with the 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20, more preferably 8, 9, 10, 11 or 12 nucleotides in the target-polynucleotide immediately adjacent to a PAM sequence.

[00100] A guide-sequence according to the present invention preferably is 8-30, more preferably 10-30, more preferably 15-30, more preferably 17-27, more preferably 17-20, more preferably 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, or 27 nucleotides in length. The ability of a guide-sequence to direct sequence-specific binding of a CRISPR-Cas system to a target-sequence to form a CRISPR-Cas complex may be assessed by any suitable assay. For example, the components of a CRISPR system sufficient to form a CRISPR-Cas complex, including the guide-sequence to be tested, may be provided to a host cell having the corresponding target-sequence, such as by transfection with vectors encoding the components of the CRISPR-Cas system, followed by an assessment of preferential cleavage and/or the resulting mutations induced by cellular repair mechanisms within the target-sequence, such as by standard sequence analysis assay. Cleavage of a target-polynucleotide may also be evaluated in a test tube by providing the target-polynucleotide, components of a CRISPR-Cas system, including the guide-sequence to be tested and a control guide-sequence different from the test guide-sequence, and comparing binding or rate of

cleavage at the target-sequence between the test and control guide-sequence reactions. Other assays are possible, and are known to a person skilled in the art.

[00101] A guide-polynucleotide structural component is believed to be necessary for formation of an active CRISPR-Cas complex. The guide-polynucleotide structural component is believed not necessarily to be operably linked to the guide-sequence; however, a guide-polynucleotide structural component may be operably linked to a guide-sequence within a guide-polynucleotide. A guide-polynucleotide structural component according to the present invention, which may comprise or consist of all or a portion of a wild-type guide-polynucleotide structural component (e.g. about or more than about 20, 26, 32, 45, 48, 54, 63, 67, 85, or more nucleotides of a wild-type tracr-sequence) forms part of a CRISPR-Cas complex; e.g. by hybridization of at least a portion of a tracr-sequence according to the present invention to all or a portion of a tracr-mate sequence according to the present invention and preferably operably linked to a guide-sequence according to the present invention. A tracr-sequence according to the present invention has sufficient complementarity to a tracr-mate sequence according to the present invention to hybridize, preferably under physiological condition as in a host cell, and facilitate formation of a CRISPR-Cas complex. As with the target-sequence according to the present invention, it is believed that complete complementarity is not needed, provided there is sufficient complementarity to be functional. Preferably, the tracr-sequence according to the present invention has at least 50%, 60%, 70%, 80%, 90%, 95% or 99% sequence identity along the length of the tracr-mate sequence according to the present invention when optimally aligned. Optimal alignment may be determined using any suitable algorithm for aligning sequences as discussed above.

[00102] In general, a tracr mate sequence according to the present invention includes any sequence that has sufficient complementarity with a tracr sequence according to the present invention to promote formation of a CRISPR-Cas complex at a target-sequence, wherein the CRISPR-Cas complex comprises the tracr mate sequence according to the present invention hybridized to the tracr sequence according to the present invention. The degree of complementarity of the tracr sequence according to the present invention and the tracr mate sequence according to the present invention is preferably defined with respect to optimal alignment of the

tracr mate sequence and tracr sequence along the length of the shorter of the two sequences. Optimal alignment may be determined using any suitable algorithm for aligning sequences as discussed above.

[00103] Preferably, with respect to a tracr mate sequence according to the present invention and a tracr sequence according to the present invention, secondary structures are taken into account, such as self-complementarity within either the tracr sequence or tracr mate sequence. Preferably, the degree of complementarity between the tracr sequence according to the present invention and tracr mate sequence according to the present invention along the length of the shorter of the two sequences when optimally aligned is higher than 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99% sequence identity. Preferably, the tracr mate sequence according to the present invention is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, or more nucleotides in length. Preferably, the tracr sequence according to the present invention is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, or more nucleotides in length. Preferably, the tracr sequence according to the present invention and tracr mate sequence, i.e. the guide-polynucleotide structural component according to the present invention are comprised within a single transcript, such that hybridization between the two produces a hybridization complex comprising a secondary structure, such as a hairpin. Such hybridization complex may also be formed when the tracr sequence and the tracr mate sequence are not comprised in a single transcript. Preferred loop forming sequences in a tracr sequence according to the present invention and/or a tracr mate sequence according to the present invention and/or guide-polynucleotide structural component according to the present invention for formation of hairpin structures are four nucleotides in length, and most preferably have the sequence GAAA; longer or shorter loop sequences may be used, as may alternative sequences. The loop sequences preferably include a nucleotide triplet (for example, AAA), and an additional nucleotide (for example C or G). Examples of loop forming sequences include CAAA and AAAG. Preferably, a tracr sequence according to the present invention and/or tracr mate sequence according to the present invention or hybridization complex thereof and/or guide-polynucleotide structural component according to the present invention comprises or is able to form at least two or more hairpins. More preferably, a tracr sequence according to the present invention and/or

tracr mate sequence according to the present invention or hybridization complex thereof and/or guide-polynucleotide structural component according to the present invention comprises or is able to form two, three, four or five hairpins. Preferably, a tracr sequence according to the present invention and/or tracr mate sequence according to the present invention or hybridization complex thereof and/or guide-polynucleotide structural component according to the present invention comprises or is able to form at most five hairpins. Preferably, the single transcript of a tracr sequence according to the present invention and a tracr-mate sequence according to the present invention or hybridization complex of a tracr sequence according to the present invention and a tracr mate sequence according to the present invention and/or guide-polynucleotide structural component according to the present invention further comprises a transcription termination sequence; preferably this is a polyT sequence, for example six T nucleotides . As said, guide-polynucleotide structural components are known to the person skilled in the art; background information can e.g. be found in Gaj *et al.*, 2013.

[00104] In the context of all embodiments according to the present invention, the term “target-polynucleotide” refers to a target-sequence according to the present invention to which a guide-sequence according to the present invention is designed to have complementarity, where hybridization between a target-sequence according to the present invention and a guide-sequence according to the present invention promotes the formation of a CRISPR-Cas complex. Full complementarity is not necessarily required, provided there is sufficient complementarity to cause hybridization and promote formation of a CRISPR-Cas complex. Preferably, a guide-sequence according to the present invention targets a target-sequence that is unique in the target. Preferably, a guide-sequence according to the present invention has 100% sequence identity with the 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20, more preferably 8, 9, 10, 11 or 12 nucleotides in the target-polynucleotide immediately adjacent to a PAM sequence. A target-polynucleotide according to the present invention may comprise any polynucleotide, such as DNA or RNA polynucleotides and may be single or double stranded. When the target-polynucleotide is a double strand polynucleotide, a guide-sequence according to the present invention, may be able to

hybridize with either strand of the target-polynucleotide e.g. a coding strand or a non-coding strand.

[00105] A target-polynucleotide according to the present invention may be located in the nucleus or cytoplasm of a cell. A target-polynucleotide according to the present invention may be located in an organelle of a host cell, for example in a mitochondrion or plastid. A target-polynucleotide according to the present invention may be comprised in a genome, may be comprised in a chromosome or may be extra-chromosomal, may be comprised in an artificial chromosome, may be present in any chromosomal entity or extra-chromosomal entity such as an autosomal replicating entity such as an episomal plasmid or vector. A target-polynucleotide according to the present invention may be native or foreign to the host cell.

[00106] A target-polynucleotide according to the present invention is preferably associated with a protospacer adjacent motif (PAM), which is a short polynucleotide recognized by the CRISPR-Cas complex. Preferably, the target-polynucleotide and PAM are linked wherein the PAM is preferably immediately downstream (3') of the target-polynucleotide. The exact sequence and length of the PAM may vary, e.g. different Cas proteins and nucleases of similar function may require different PAM sequences. A preferred PAM according to the present invention is a polynucleotide of 2 to 8 nucleotides in length. A preferred PAM is selected from the group consisting of 5'-XGG-3', 5'-XGGXG-3', 5'-XXAGAAW-3', 5'-XXXXGATT-3', 5'-XXAGAA-3', 5'-XAAAAC-3', wherein X can be any nucleotide or analog thereof, preferably any nucleotide; and W is A or T. A more preferred PAM is 5'-XGG-3'. The PAM is preferably matched with the Cas protein. The most widely used CAS/CRISPR system is derived from *S. pyogenes* and the matching PAM sequence 5'-XGG-3' is located immediately downstream (3') of the target-sequence. A preferred PAM for a *Neisseria meningitidis* Cas protein is 5'-XXXXGATT-3'; a preferred PAM for a *Streptococcus thermophilus* Cas protein is 5'-XXAGAA-3'; a preferred PAM for a *Treponema denticola* is 5'-XAAAAC-3'. A preferred PAM matches the Cas protein used. A Cas protein according to the present invention may be engineered to match a different PAM than the native PAM matching the wild-type Cas protein. As such, the CRISPR-Cas system according to the present invention may be used for customized specific targeting.

[00107] The term “hybridization” refers to a reaction in which one or more polynucleotides react to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The hydrogen bonding may occur by Watson Crick base pairing, Hoogsteen binding, or in any other sequence-specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi stranded complex, a single self-hybridizing strand, or any combination of these. A hybridization reaction may constitute a step in a more extensive process, such as the cleavage of a polynucleotide by an enzyme. Preferred hybridization conditions are physiological conditions as within a host cell according to the present invention.

[00108] The term “source” in the context of all embodiments of the present invention refers to any source of a CRISPR-Cas system comprising a guide-polynucleotide and a Cas protein. The guide-polynucleotide and Cas protein may be present in separate sources. In such case, the composition according to the present invention comprises a CRISPR-Cas system comprising a source of a guide-polynucleotide and a source of a Cas-protein. Any source means that the guide-polynucleotide and Cas protein may be present as such in a form that they can function within a CRISPR-Cas system. The guide-polynucleotide and/or the Cas-protein may be provided in its active forms and may e.g. be provided from an inactive form or from another entity. The guide-polynucleotide may e.g. be present on another polynucleotide or may be encoded by a polynucleotide that is transcribed to provide for the actual guide-polynucleotide. The Cas protein may be encoded by a polynucleotide (e.g. DNA or mRNA) that is transcribed and/or translated to provide the actual Cas protein. An encoding polynucleotide may be present in a nucleic acid construct as defined herein and/or in a vector as defined herein. Such nucleic acid construct and vector are herein referred to as a nucleic acid construct according to the present invention and a vector according to the present invention.

[00109] Preferably, in the composition according to the present invention, the Cas protein or nuclease of related function is encoded by a polynucleotide and/or the guide-polynucleotide is encoded by or present on a polynucleotide.

[00110] Preferably, in the composition according to the present invention, the Cas protein or nuclease of related function is encoded by a polynucleotide and/or the

guide-polynucleotide is encoded by or present on another polynucleotide and the polynucleotide or polynucleotides are comprised in a vector.

[00111] Preferably, in a composition according to the invention, the guide-polynucleotide is encoded by a polynucleotide that is transcribed to provide for the actual guide-polynucleotide. Accordingly, in an embodiment, in the composition according to the invention, preferably, the guide polynucleotide is present in the form of a polynucleotide encoding for said guide-polynucleotide and the guide-polynucleotide is obtained upon transcription of said polynucleotide in the host cell.

[00112] Preferably, in the composition according to the present invention, the Cas protein is encoded by a polynucleotide and the guide-polynucleotide is encoded by or present on another polynucleotide and the polynucleotides are comprised in one vector.

[00113] Preferably, in the composition according to the present invention, the Cas protein is encoded by a polynucleotide comprised in a vector and the guide-polynucleotide is encoded by or present on another polynucleotide comprised in another vector. Preferably, the vector encoding the Cas protein is a low copy vector and/or the promoter driving expression of the Cas transcript is a low-strength promoter and the vector encoding the guide-polynucleotide is a high copy vector and/or the promoter driving expression of the gRNA transcript is a high-strength promoter. This allows differential expression of the Cas protein and the guide-polynucleotide; the Cas protein may e.g. be expressed in lower level than the guide-polynucleotide. Promoter strength can be estimated by any means, for example, by RNA sequencing. RNA sequencing (RNAseq) is a highly sensitive and accurate tool for measuring expression across the transcriptome under different conditions. It allows quantitative approximation of gene expression at the transcript level which is reported as RPKM value (Reads Per Kilobase of transcript per Million mapped reads). RPKM values and relative promoter strengths of representative genes from *Schizochytrium* are provided in Table 1.

TABLE 1

Gene	Ave RPKM value	Relative promoter strength
Arginase (EC 3.5.3.1)	3694.34	Strong
Pyruvate kinase (EC 2.7.1.40)	2127.49	Strong
Heat shock protein 70	1857.74	Strong
Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12)	1688.00	Strong
Tubulin alpha chain	696.85	Medium
Protein Translation Elongation Factor 1A (EF-1A)	682.30	Medium
Isocitrate lyase (EC 4.1.3.1)	510.45	Medium
Aconitate hydratase 2 (EC 4.2.1.3)	317.74	Medium
Malate dehydrogenase (EC 1.1.1.37)	192.47	Medium
Vacuolar ATP synthase subunit D (EC 3.6.3.14)	163.74	Medium
Acyl-CoA dehydrogenase (EC 1.3.99.3)	94.43	Weak
Acetylspermidine deacetylase (EC 3.5.1.48)	90.60	Weak
Acetolactate synthase (EC 2.2.1.6)	63.21	Weak
Maleylacetoacetate isomerase (EC 5.2.1.2)	39.46	Weak
Phytoene desaturase (EC 1.14.99.-) (Carotene Synthase)	31.56	Weak
ATP-dependent RNA helicase	28.42	Weak
Acetyl-CoA acetyltransferase (EC 2.3.1.9)	26.74	Weak

[00114] Examples of low-strength (i.e., weak) *Schizochytrium* promoters include, but are not limited to, those driving expression of carotenoid synthase. Examples of medium-strength (i.e., medium) *Schizochytrium* promoters include, but are not limited to, those driving expression of alpha-tubulin. Examples of high-strength (i.e., strong) *Schizochytrium* promoters include, but are not limited to, those driving expression of elongation factor 1 (EF-1). Although RPKM values are considered generally indicative

of relative promoter strength, it is known to those skilled in the art that a promoter in its native genomic context may not have exactly the same strength in the context of an expression vector. Thus, one of skill in the art will understand that the relative promoter strengths provided in Table 1 may vary in the context of the invention.

[00115] The invention thus provides for the possibilities that the guide-polynucleotide and the Cas protein are provided as such, or that they are encoded on or present on a vector. In the latter case, the encoding polynucleotides may each be on a separate vector or may both be on a single vector. Accordingly, in an embodiment, the present invention provides for a composition according to the present invention wherein a polynucleotide encoding a Cas protein according to the present invention, a guide-polynucleotide or a polynucleotide encoding a guide-polynucleotide according to the present invention are present on a single vector, which may further comprise any elements necessary for expressing the encoded products such as promoter and terminator elements. Such single (all-in-one) vector has the advantage that all components necessary for a CRISPR-Cas system are present together; in addition, a single transformation event, optionally in combination with a donor polynucleotide, suffices to introduce the components into a host cell.

Vectors

[00116] In the context of all embodiments of the present invention, a vector may be any vector (e.g., a plasmid or virus), which can conveniently be subjected to recombinant DNA procedures and can mediate expression of a polynucleotide according to the invention. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. Preferred vectors are the vectors used in the examples herein. A vector may be a linear polynucleotide or a linear or closed circular plasmid. A vector may be an integrating or autonomously replicating vector, i.e., a vector, which exists as a chromosomal or an extra-chromosomal entity, the replication of which is dependent on or independent of chromosomal replication, e.g., a plasmid, an extra-chromosomal element, a mini-chromosome, or an artificial chromosome.

[00117] Preferably a vector may be one which, when introduced into the host cell, becomes integrated into the genome and replicated together with the

chromosome(s) into which it has been integrated. An integrative vector may integrate at random or at a predetermined target locus in a chromosome of the host cell. A preferred integrative vector comprises a DNA fragment, which is homologous to a DNA sequence in a predetermined target locus in the genome of the host cell for targeting the integration of the vector to this predetermined locus. In order to promote targeted integration, a vector is preferably linearized prior to transformation of the cell. Linearization is preferably performed such that at least one but preferably either end of the vector is flanked by sequences homologous to the target locus. The length of the homologous sequences flanking the target locus is preferably at least 30 bp, preferably at least 50 bp, preferably at least 0.1 kb, even preferably at least 0.2 kb, more preferably at least 0.5 kb, even more preferably at least 1 kb, most preferably at least 2 kb. Preferably, the efficiency of targeted integration into the genome of the host cell, i.e. integration in a predetermined target locus, is increased by augmented homologous recombination abilities of the host cell.

[00118] The homologous flanking DNA sequences in the vector (which are homologous to the target locus) may be derived from an expressed locus, meaning that they are derived from a gene, which is capable of expression in the host cell. Flanking DNA sequences may be linked with a selectable marker gene, such that transformed cells will grow when culture medium is supplemented with an appropriate selectable agent. The flanking DNA sequences can be designed by any means known to a person skilled in art; one preferred design directs homologous recombination to genes encoding saturated and/or polyunsaturated fatty acid synthases (PUFA synthases), such that mutation or disruption of these synthases creates auxotrophy for saturated and/or polyunsaturated fatty acids. Cells that are auxotrophic for saturated and/or polyunsaturated fatty acids, require saturated and/or polyunsaturated fatty as supplements for growth. Another preferred design enables expression of a dominant selectable marker gene such that transformants of said selectable marker gene are enabled to grow in the presence of an appropriate dominant selectable agent.

[00119] More than one copy of a polynucleotide according to the present invention may be inserted into the microbial host cell to mediate production of the product encoded by said polynucleotide. This can be done, preferably by integrating

multiple copies of the polynucleotide into the genome of the host cell, or by targeting the integration of the polynucleotide at a highly expressed locus in an operable configuration. Alternatively, integration of multiple copies can be achieved by including an amplifiable selectable marker gene with a polynucleotide according to the present invention, such that cells containing amplified copies of the selectable marker gene (and thereby additional copies of the nucleic acid sequence) can be selected for by cultivating the cells in the presence of the appropriate selectable agent. To increase the number of copies of a polynucleotide according to the present invention even more, the technique of gene conversion as described in WO98/46772 may be used.

[00120] When a polynucleotide according to the present invention encoding a Cas protein according to the present invention and/or a guide-polynucleotide according to the present invention is integrated into the genome of the host cell, it may be desirable to excise the polynucleotide from the genome, e.g. when the desired genome editing has taken place. The excision of a polynucleotide can be performed by any means known to the person skilled in art; one preferred means is by secondary transformation with a nucleotide which repairs a gene mutation or disruption which induced an auxotrophy such that cells which have been cured of the auxotrophy can instead be selected by growth in a culture medium in which the nutrient required by the auxotrophs has been omitted. Another means for excision would be to use the CRISPR-Cas system according to the present invention.

[00121] A vector according to the present invention may be a single vector or plasmid or a vector system comprising two or more vectors or plasmids, which together contain the polynucleotides according to the present invention to be introduced into the host cell.

[00122] A vector according to the present invention may contain one or more selectable markers, which permit easy selection of transformed cells. In an embodiment, in a composition according to the invention, one or more or all vectors comprise a selectable marker, preferably each vector comprising a distinct selectable marker. A selectable marker is a gene the product of which provides for biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophs, and the like. The selectable marker may be introduced into the cell on the vector as an expression cassette or may be introduced on a separate vector.

[00123] A selectable marker for use in a Labyrinthulomycete cell may be selected from the group including, but not limited to, nptII (neomycin phosphotransferase II, conferring paromomycin resistance), ALS (acetolactate synthase, conferring sulfometuronmethyl resistance), bsd (blasticidin-S-deaminase, conferring blasticidin resistance), and Sh ble (phleomycin binding, conferring zeocin resistance).

[00124] Alternatively, specific selection markers can be used such as auxotrophic markers which require corresponding mutant host cells harboring inactivating mutations of saturated or polyunsaturated fatty acid synthase genes as previously discussed strains. In a preferred embodiment, the selection marker is deleted from the transformed host cell after introduction of the expression construct so as to obtain transformed host cells capable of producing the polypeptide which are free of selection marker genes.

[00125] The procedures used to ligate elements described above to construct a vector according to the present invention are well known to one skilled in the art (see, e.g. Sambrook & Russell, *Molecular Cloning: A Laboratory Manual, 3rd Ed.*, CSHL Press, Cold Spring Harbor, NY, 2001; and Ausubel *et al.*, *Current Protocols in Molecular Biology*, Wiley InterScience, NY, 1995).

[00126] A Cas protein in the context of all embodiments of the present invention refers to any Cas protein suitable for the purpose of the invention. A Cas protein may comprise enzymatic activity or may not comprise enzymatic activity. Non-limiting examples of Cas proteins include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 and Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx1S, Csf1, Csf2, Csf3, Csf4, homologs thereof or modified versions thereof. These Cas proteins are known to the person skilled in the art; for example, the amino acid sequence of *S. pyogenes* Cas9 protein may be found in the SwissProt database under accession number Q99ZW2. Preferably, an unmodified Cas protein according to the present invention has DNA cleavage activity, such as e.g. Cas9. Preferably, a Cas protein according is Cas9, and may be Cas9 from *S. pyogenes* or *S. pneumoniae*. Preferably, a Cas protein according to the present invention directs cleavage of one or both polynucleotide strands at the location of the target-polynucleotide, such as

within the target-polynucleotide and/or within the reverse complement of the target-polynucleotide. At the location of the target-polynucleotide is herein defined as within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, or more nucleotides from the first or last nucleotide of a target-polynucleotide; more preferably, within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, or more nucleotides from the first or last nucleotide of a target-polynucleotide; even more preferably, within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50 nucleotides from the first or last nucleotide of a target-polynucleotide. Accordingly, a Cas protein according to the present invention preferably directs cleavage of one or both polynucleotide strands within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, or more nucleotides from the first or last nucleotide of a target-polynucleotide; more preferably, within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, or more nucleotides from the first or last nucleotide of a target-polynucleotide; even more preferably, within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50 nucleotides from the first or last nucleotide of a target-polynucleotide. Typically, a target-polynucleotide according to the present invention is associated with a PAM sequence (defined elsewhere herein) and the PAM sequence is preferably immediately downstream (3') of the target-sequence; the formation of the CRISPR-Cas complex typically results in cleavage of one or both polynucleotide strands 3 base pairs upstream (5') of the PAM sequence.

[00127] Preferably, a Cas protein in a composition according to the present invention has activity for directing cleavage of both polynucleotide strands at the location of the target-polynucleotide. Cas nuclease activity is typically performed by two separate catalytic domains, namely RuvC and HNH. Each domain cuts one polynucleotide strand each domain can be inactivated by a single point mutation.

[00128] A Cas protein according to the present invention may thus conveniently be mutated with respect to a corresponding wild-type Cas protein such that the mutated Cas protein has altered nuclease activity and lacks the ability to cleave one or both strands of a target-polynucleotide. For example, an aspartate-to-alanine substitution (D10A) in the RuvC I catalytic domain of Cas9 from *S. pyogenes* converts Cas9 from a nuclease that cleaves both strands to a nickase, which is herein defined as a Cas protein that cleaves a single strand of a target-polynucleotide. Other examples of mutations that render Cas9 into a nickase include, but are not limited to

H840A, N854A, and N863A. In the context of the present invention, a Cas protein having nickase activity may be used for genome editing via homologous recombination, preferably the double nicking technique according to Ran *et al.*, 2013. Accordingly, a preferred Cas protein according to the present invention comprises at least one mutation, such that the protein has altered nuclease activity compared to the corresponding wild-type Cas protein, preferably having activity to direct cleavage of a single polynucleotide strand at the location of the target-sequence. Such so-called nickase mutant can conveniently be used in duplex set-up, i.e. in a composition according to the present invention comprising a Cas protein nickase mutant with RuvC mutated and a Cas protein nickase mutant wherein NHN is mutated, such that the one Cas protein mutant nicks one strand of the polynucleotide target and the other Cas protein mutant nicks the other strand of the polynucleotide target. Depending on the two guide-polynucleotides used, the two different CRISPR-Cas complexes will effectively result in two single-strand nicks in the polynucleotide target; these nicks may be several nucleotides up to 5, 10, 20, 30 or more apart. Such double nicking method greatly enhances specificity of non-homologous end joining (NEJH). Background information on double nicking can be found in e.g. Ran *et al.*, 2013.

[00129] A Cas protein according to the present invention may comprise two or more mutated catalytic domains of Cas9, such as RuvC I, RuvC II and/or RuvC III to result in a mutated Cas9 substantially lacking all DNA cleavage activity. In some embodiments, a D10A mutation is combined with one or more of H840A, N854A, or N863A mutations to produce a Cas9 enzyme substantially lacking all DNA cleavage activity. Preferably, a Cas protein is considered to substantially lack all DNA cleavage activity when the DNA cleavage activity of the mutated enzyme is less than about 25%, 10%, 5%, 1%, 0.1%, 0.01%, or lower with respect to its non-mutated form. A Cas protein lacking substantially all enzyme activity can conveniently be used for gene silencing or down regulation of expression since the CRISPR-Cas complex will hamper transcription from the target-polynucleotide. Other mutations may be useful; where the Cas9 or other Cas protein is from a species other than *S. pyogenes*, mutations in corresponding amino acids may be made to achieve similar effects; the person skilled in the art knows how to identify these corresponding amino acids.

[00130] A Cas protein according to the present invention may be a fusion protein and comprise at least one heterologous functional domain, such domain preferably is a domain comprising *FokI* activity such as described by Aggarwal *et al.* (Aggarwal, A. K.; Wah, D. A.; Hirsch, J. A.; Dorner, L. F.; Schildkraut, I. (1997). "Structure of the multimodular endonuclease *FokI* bound to DNA". *Nature* 388 (6637): 97–100). The enzyme *FokI* is naturally found in *Flavobacterium okeanokoites* and is a bacterial type IIS restriction endonuclease consisting of an N-terminal DNA-binding domain and a non-specific DNA cleavage domain at the C-terminal (Durai *et al.*, 2005). When the *FokI* protein is bound to double stranded DNA via its DNA-binding domain at the 5'-GGATG-3':3'-CATCC-5' recognition site, the DNA cleavage domain is activated and cleaves, without further sequence specificity, the first strand 9 nucleotides downstream and the second strand 13 nucleotides upstream of the nearest nucleotide of the recognition site (Wah *et al.*, 1998). *Cas9-FokI* fusions have been described *inter alia* in Guillinger *et al.*, 2014; and in Tsai *et al.*, 2014.

[00131] A Cas fusion protein according to the present invention may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more domains in addition to the Cas protein. Examples of protein domains that may be fused to a Cas protein include, but are not limited to, epitope tags, reporter gene sequences, and protein domains having one or more of the following activities: methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity and nucleic acid binding activity. Non-limiting examples of epitope tags include histidine (His) tags, V5 tags, FLAG tags, influenza hemagglutinin (HA) tags, Myc tags, VSV-G tags, and thioredoxin (Trx) tags. Examples of reporter genes include, but are not limited to, glutathione-S-transferase (GST), horseradish peroxidase (HRP), chloramphenicol acetyltransferase (CAT) beta-galactosidase, beta-glucuronidase, luciferase, green fluorescent protein (GFP), HcRed, DsRed, cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), and autofluorescent proteins including blue fluorescent protein (BFP). A Cas protein may be fused to a gene sequence encoding a protein or a fragment of a protein that bind DNA molecules or bind other cellular molecules, including but not limited to, maltose binding protein (MBP), S-tag, Lex A DNA binding domain (DBD) fusions, GAL4 DNA binding domain fusions, and herpes simplex virus (HSV) BP 16 protein fusions.

Additional domains that may form part of a fusion protein comprising a CRISPR enzyme are described in US20110059502. A tagged Cas protein may be used to identify the location of a target-polynucleotide. A preferred Cas fusion protein according to the present invention comprises a *FokI* domain as defined here above.

[00132] A preferred Cas protein according to the present invention comprises a nuclear localization sequence, preferably a heterologous nuclear localization sequence. Such nuclear localization sequence is also referred to as a nuclear localization signal. Preferably, such nuclear localization signal confers to the CRISPR-Cas complex sufficient strength to drive accumulation of said CRISPR-Cas complex in a detectable amount in the nucleus of a host cell. Without wishing to be bound by theory, it is believed that a nuclear localization sequence is not necessary for CRISPR-Cas activity in a host cell, but that including such sequences enhances activity of the system, especially as to targeting nucleic acid molecules into the nucleus. Such nuclear localization sequence is preferably present in the Cas protein, but may also be present anywhere else such that targeting of the CRISPR-Cas system to the nucleus is facilitated. A preferred nuclear localization sequence is the SV40 nuclear localization sequence.

[00133] In a composition and in any other embodiment according to the present invention, a Cas protein encoding polynucleotide is preferably codon optimized for the host cell it is to be expressed in, more preferably, the Cas protein encoding polynucleotide is codon pair optimized. In general, codon optimization refers to a process of modifying a nucleic acid sequence for enhanced expression in a host cell of interest by replacing at least one codon (e.g. more than 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more codons) of a native sequence with codons that are more frequently or most frequently used in the genes of that host cell while maintaining the native amino acid sequence. Various species exhibit particular bias for certain codons of a particular amino acid. Codon bias (differences in codon usage between organisms) often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, among other things, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal

gene expression in a given organism based on codon optimization. Codon usage tables are readily available, for example, at the “Codon Usage Database”, and these tables can be adapted in a number of ways. See, e.g., Nakamura, Y., *et al.*, 2000. Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell are also available, such as Gene Forge (Aptagen; Jacobus, PA), are also available. Preferably, one or more codons (e.g. 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more, or all codons) in a sequence encoding a Cas protein correspond to the most frequently used codon for a particular amino acid. Preferred methods for codon optimization are described in WO2006/077258 and WO2008/000632. WO2008/000632 addresses codon-pair optimization. Codon-pair optimization is a method wherein the nucleotide sequences encoding a polypeptide have been modified with respect to their codon-usage, in particular the codon-pairs that are used, to obtain improved expression of the nucleotide sequence encoding the polypeptide and/or improved production of the encoded polypeptide. Codon pairs are defined as a set of two subsequent triplets (codons) in a coding sequence. The amount of Cas protein in a source in a composition according to the present invention may vary and may be optimized for optimal performance. It may be convenient to avoid too high levels of Cas protein in a host cell since high levels of Cas protein may be toxic to the host cell, even without a guide-polynucleotide present (see, e.g., Ryan *et al.*, 2014 and Jacobs *et al.*, 2014). A person skilled in the art knows how to regulate expression levels, such as by choosing a weaker promoter, repressible promoter or inducible promoter for expression of a Cas protein. Examples of promoters suitable for expression of a protein are depicted elsewhere herein.

[00134] In a composition according to the present invention wherein a guide-polynucleotide according to the present invention is encoded by a polynucleotide, expression of the guide-polynucleotide may be facilitated by a promoter operably linked to the encoding polynucleotide. Such promoter may be any suitable promoter known to the person skilled in the art. Several types of promoters can be used. It may be convenient to use an RNA polymerase III promoter or an RNA polymerase II promoter. Background information on RNA polymerase III and its promoters can be found e.g. in Marck *et al.*, 2006. Accordingly, it may be convenient to use an RNA polymerase II promoter; these are known to the person skilled in the art and reviewed

in, e.g., Kornberg, 1999. However, transcripts from an RNA II polymerase often have complex transcription terminators and transcripts are polyadenylated; this may hamper with the requirements of the guide-polynucleotide which because both its 5' and 3' ends need to be precisely defined in order to achieve the required secondary structure to produce a functional CRISPR-Cas system. These drawbacks can however be circumvented. In case a RNA polymerase II promoter is used, the polynucleotide encoding the guide-polynucleotide may also encode self-processing ribozymes and may be operably linked to an RNA polymerase II promoter; as such the polynucleotide encodes a pre-guide-polynucleotide comprising the guide-polynucleotide and self-processing ribozymes, wherein, when transcribed, the guide-polynucleotide is released by the self-processing ribozymes from the pre-guide-polynucleotide transcript. Preferred constructs comprising a polynucleotide encoding a pre-guide-polynucleotide according to the present invention operably linked to an RNA polymerase II promoter are those depicted in examples 1 – 4 herein. Background information on such constructs can be found in e.g. Gao *et al.*, 2014.

[00135] Preferably, in a composition according to the present invention, the guide-polynucleotide is encoded by a polynucleotide.

Preferably, in a composition according to the present invention wherein the guide-polynucleotide is encoded by a polynucleotide, said polynucleotide is operably linked to an RNA polymerase II promoter and encodes a pre-guide-polynucleotide comprising the guide-polynucleotide and self-processing ribozymes, wherein, when transcribed, the guide-polynucleotide is released by the self-processing ribozymes from the pre-guide-polynucleotide transcript. Preferred constructs comprising a polynucleotide encoding a pre-guide-polynucleotide according to the present invention operably linked to an RNA polymerase II promoter are those depicted in examples 1 – 4 herein. Conveniently, multiple pre-guide-polynucleotides and multiple self-processing ribozymes may be encoded by a single polynucleotide, operably linked to one or more RNA polymerase II promoters.

[00136] The composition according to the first aspect of the present invention can conveniently be used to modulate expression of a polynucleotide in a host cell. Accordingly, in a second aspect, the present invention provides a method of

modulating expression of a polynucleotide in a host cell, comprising contacting a host cell with the composition according to the first aspect of the invention, wherein the guide-polynucleotide directs binding of the Cas protein at the target-polynucleotide in the host cell to form a CRISPR-Cas complex.

[00137] The term “expression” in the context of the present invention is herein defined as the process by which a polynucleotide is transcribed from a polynucleotide template (e.g. a DNA template polynucleotide is transcribed into an mRNA polynucleotide transcript or other RNA transcript) and/or the process by which an mRNA transcript is subsequently translated into peptides, polypeptides, or proteins. Transcripts and encoded polypeptides may be collectively referred to as “gene product”. If the polynucleotide transcript is derived from a genomic template DNA, expression may include splicing of the mRNA transcript in a host cell. The term “modulating expression” refers herein to increased or reduced expression compared to a parent host cell wherein expression is not modulated when assayed using the same conditions. Reduced expression may be a reduced amount of transcript such as mRNA and/or a reduced amount of translation product such as a polypeptide. It follows that increased expression may be an enhanced amount of transcript such as mRNA and/or an enhanced amount of translation product such as a polypeptide.

[00138] Preferably, the CRISPR-Cas complex cleaves one or both polynucleotide strands at the location of the target-polynucleotide, resulting in modulated expression of the gene product. The CRISPR-Cas complex may also have altered nuclease activity and substantially lack the ability to cleave one or both strands of a target-polynucleotide; in such case, expression is modulated by the binding of the complex to the target-polynucleotide. A Cas protein lacking substantially all enzyme activity can conveniently be used for gene silencing or down regulation of expression since the CRISPR-Cas complex will hamper transcription from the target-polynucleotide. Alternatively, a Cas protein can be modified into a transcription factor for programmable transcriptional activation or silencing of a gene of interest (Larson, *et al.*, 2013).

[00139] A composition according to the first aspect of the present invention can conveniently be used for the deletion of polynucleotide. In an embodiment, when the composition according to the first aspect of the present invention comprises a source

of at least one or two guide-polynucleotides and/or a source of at least at least one Cas protein, at least one CRISPR-Cas complex or two different CRISPR-Cas complexes are formed that cleave one or both polynucleotide strands at one location or at different locations of the target-polynucleotide, resulting in deletion of a polynucleotide fragment from the target-polynucleotide. Preferably, such composition according to the present invention comprising at least one or two guide-polynucleotides and/or a source of at least at least one Cas protein additionally comprises an exogenous polynucleotide as defined herein below that is at least partly complementary to the at least one or two target-polynucleotides targeted by the guide-polynucleotide(s). Such polynucleotide fragment to be deleted or deleted fragment may be several nucleotides in length up to a few thousand nucleotides in length, an entire gene may be deleted or a cluster of genes may be deleted. Accordingly, the present invention provides for a method of modulating expression of a polynucleotide in a host cell, wherein a polynucleotide fragment is deleted from a target-polynucleotide.

[00140] In an embodiment, the method of modulating expression comprises cleavage of one or both polynucleotide strands in at least one location of the target-polynucleotide followed by modification of the target-polynucleotide by homologous recombination with an exogenous polynucleotide. In such case, the composition according to the first aspect of the present invention preferably further comprises such exogenous polynucleotide. Such modification may result in insertion, deletion or substitution of at least one nucleotide in the target-polynucleotide, wherein the insertion or substitution nucleotide may originate from the exogenous polynucleotide. A modification can also be made when the exogenous polynucleotide is a non-integrating entity such as described in Dong *et al.* and Beetham *et al.*; in this case the target-polynucleotide is modified but no nucleotide of the exogenous polynucleotide is introduced into the target-polynucleotide. Consequently, the resulting host is a non-recombinant host cell when the Cas-protein according to the invention is introduced to the host cell as a protein. The exogenous polynucleotide may be any polynucleotide of interest such as a polynucleotide encoding a compound of interest as defined herein below, or a part of such polynucleotide or a variant thereof. Such exogenous

polynucleotide is herein referred to as an exogenous polynucleotide according to the present invention and may single-stranded or double-stranded.

[00141] Various applications can be considered by the person skilled in the art for the compositions and methods according to the present invention. A polynucleotide (or gene) in a genome may be modified, edited or disrupted using compositions and methods according to the present invention. E.g. when a fully active Cas protein is used that cuts in both strands of the target-polynucleotide and when no exogenous polynucleotide is present as a suitable repair template, the double strand break is repaired by non-homologous end joining repair (NHEJ). During NHEJ insertions and/or deletions (which may be construed as substitution in some cases) of one or several nucleotides may occur, these are randomly inserted or deleted at the repair site; this is characteristic for NHEJ. Such insertions and/or deletions may impact the reading frame of the coding sequence, resulting amino acid changes in the gene product or even a truncated protein in case of genesis of a (premature) stop codon or alteration of a splice site.

[00142] A polynucleotide (or gene) in a genome may be modified, edited or disrupted using compositions and methods according to the present invention using homologous end joining repair (HEJ), also known as homology-directed repair (HDR), when an exogenous polynucleotide is present as repair template. E.g. when an exogenous polynucleotide having sequence identity to the target-polynucleotide (i.e. upstream (5') and downstream (3') of the double strand break) is present together with a CRISPR-Cas system according to the present invention, HDR will introduce (or actually reproduce) the corresponding nucleotides of the exogenous polynucleotide at the double strand break in the target-polynucleotide. Preferably, an exogenous polynucleotide according to the present invention does not contain the target sequence itself followed by a functional PAM sequence to avoid the risk of the exogenous polynucleotide itself or the modified target-polynucleotide being (re)cut by the CRISPR-Cas system.

[00143] In the embodiments of the present invention, when a CRISPR-Cas system according to the present invention comprises an exogenous polynucleotide (donor polynucleotide, donor DNA, repair template), the CRISPR-Cas system according to the present invention preferably comprises two or more guide-

polynucleotides encoded by or present on one or more separate polynucleotides or vectors, and two or more exogenous polynucleotides are provided together with said CRISPR-Cas system enabling the formation of two or more CRISPR-Cas complexes. In a method according to the present invention, such CRISPR-Cas systems according to the present invention can conveniently be used to modulate expression at two or more target-polynucleotides, i.e. a method to target multiple target sites. Such CRISPR-Cas system according to the present invention will by chance form one, two or more CRISPR-Cas complexes at one or more target-polynucleotides. Such method can be used to generate one or more insertions, deletions, substitutions, optionally in combination with the one or more exogenous polynucleotides, in the genome of the host cell, or to modulate expression of genes via the formed CRISPR-Cas complexes.

Host cells

[00144] In a method according to this aspect of the present invention, a preferred host cell comprises a polynucleotide encoding a compound of interest as defined elsewhere herein.

[00145] In a method according to this aspect of the present invention, the host cell may be a recombinant host cell or may be a non-recombinant host cell.

[00146] In some embodiments, the host cell is a Labyrinthulomycete, preferably a member of the order Thraustochytriales, preferably a member of the family Thraustochytriaceae, more preferably a member of a genus selected from the group consisting of *Aurantiochytrium*, *Oblongichytrium*, *Schizochytrium*, *Thraustochytrium*, and *Ulkenia*, even more preferably *Schizochytrium* sp. ATCC# 20888.

[00147] A method of modulating expression of a polynucleotide in a host cell according to this aspect of the present invention, results in a modified host cell that preferably comprises components of the composition according to the first aspect of the present invention. Accordingly, in a third aspect, the present invention provides a host cell comprising a composition according to the first aspect of the present invention. Such host cell may be any host cell as defined herein and may further comprise a polynucleotide encoding a compound of interest as defined elsewhere herein.

[00148] In a fourth aspect, the present invention provides a method of producing a host cell, comprising contacting a host cell with the composition according to the first aspect of the present invention, wherein the guide-polynucleotide directs binding of the Cas protein at the target-polynucleotide in the host cell to form a CRISPR-Cas complex. In an embodiment, the contacting with the composition according to the first aspect of the invention may be performed in two steps, wherein the host cell is first contacted with a source of a Cas protein according to the invention and subsequently the host cell is contacted with a source of a guide-polynucleotide according to the invention and optionally an exogenous polynucleotide according to the invention. A host cell in this embodiment of the present invention may be any type of host cell as defined herein and may comprise a polynucleotide encoding a compound of interest as defined elsewhere herein. A preferred method of producing a host cell according to the present invention comprises a step to produce an offspring host cell, wherein in said offspring host cell no components of a CRISPR-Cas system according to the present invention are present anymore.

[00149] The composition according to the first aspect of the present invention may be any such composition as defined herein. Contacting a host cell with a composition according to the present invention may be performed by any means known to the person skilled in the art. A host cell according to the present invention may simply be brought into a solution comprising a composition according to the present invention. Specific means of delivering a composition according to the present invention into a host cell may be used. The person skilled in the art is aware of such methods (see e.g. Sambrook & Russell; Ausubel, *supra*), which include but are not limited to electroporation methods, particle bombardment or microprojectile bombardment, protoplast methods and *Agrobacterium* mediated transformation (AMT). Labyrinthulomycetes may be transformed using any method known in the art. The general technique for genetic transformation of Thraustochytrids is described in detail in U.S. Patent Nos. 7,001,772 and 8,637,651, and by Cheng *et al.* (2012), all of which are incorporated herein by reference in their entirety.

[00150] Preferably, the CRISPR-Cas complex cleaves one or both polynucleotide strands at the location of the target-polynucleotide, resulting in modulated expression of the gene product. The CRISPR-Cas complex may also have

altered nuclease activity and lack the ability to cleave one or both strands of a target-polynucleotide; in such case, expression is modulated by the binding of the complex to the target-polynucleotide.

[00151] In an embodiment, when the composition according to the first aspect of the present invention comprises a source of at least one or two guide-polynucleotides and/or a source of at least one Cas protein, at least one CRISPR-Cas complex or two different CRISPR-Cas complexes are formed that cleave one or both polynucleotide strands at one location or at different locations of the target-polynucleotide, resulting in deletion of a polynucleotide fragment from the target-polynucleotide. Preferably, such composition according to the present invention comprising at least one or two guide-polynucleotides and/or a source of at least one Cas protein, additionally comprises an exogenous polynucleotide as defined herein below that is at least partly complementary to the at least one or two target-polynucleotides targeted by the guide-polynucleotide(s). Such polynucleotide fragment to be deleted or deleted fragment may be from several nucleotides in length up to a few thousand nucleotides in length, an entire gene may be deleted or a cluster of genes may be deleted. Accordingly, the present invention provides for a method of modulating expression of a polynucleotide in a host cell, wherein a polynucleotide fragment is deleted from a target-polynucleotide.

[00152] In one embodiment, a method of modulating expression of a polynucleotide in a host cell, wherein a polynucleotide fragment is deleted from a target-polynucleotide, comprises contacting a host cell with a composition as described herein, wherein the guide-polynucleotide directs binding of the Cas protein at the target-polynucleotide in the host cell to form a CRISPR-Cas complex. Preferably, a method of modulating expression of a polynucleotide in a host cell, wherein a polynucleotide fragments is deleted from a target-polynucleotide, comprises contacting a host cell with a composition as described herein, wherein the guide-polynucleotide directs binding of the Cas protein at the target-polynucleotide in the host cell to form a CRISPR-Cas complex. In another preferred embodiment, a method of modulating expression of a polynucleotide in a host cell, wherein a polynucleotide fragment is deleted from a target-polynucleotide, comprises contacting a host cell with a composition as described herein, wherein the guide-polynucleotide directs binding

of the Cas protein at the target-polynucleotide in the host cell to form a CRISPR-Cas complex, wherein the composition as described herein does not comprise an exogenous or donor polynucleotide. In another embodiment of the method of modulating expression of a polynucleotide in a host cell, the composition is comprised in an autonomously replicating vector.

[00153] Therefore, the present invention relates in one embodiment to a method of modulating expression of a polynucleotide in a cell, wherein a polynucleotide fragment is deleted from a target-polynucleotide, comprising contacting a host cell with the composition as described herein but preferably not comprising a donor polynucleotide as defined herein, wherein the guide-polynucleotide directs binding of the Cas protein at the target-polynucleotide in the host cell to form a CRISPR-Cas complex.

[00154] In a preferred embodiment, the Cas protein has activity for directing cleavage of both polynucleotide strands at the location of the target-sequence and wherein the cleavage occurs in a region of the genome comprised between two homologous regions which upon cleavage by the Cas protein recombine with each other resulting in the deletion of a polynucleotide comprised between said regions. Preferably, the degree of homology between the two homologous regions is such to allow homologous recombination. Preferably, the two homologous regions have at least 60%, 70%, 80%, 90%, 99% or 100% sequence identity over the whole length of the homologous regions. It has been surprisingly found that the length of homologous region can be very short even in Labyrinthulomycetes, wherein usually a length of at least 1 or several kbp is necessary to allow homologous recombination. Therefore, in a preferred embodiment, the length of the homologous regions is preferably at most 1 kb, at most 0.5 kb, at most 100 bp, at most 50 bp, at most 40 bp, at most 30 bp, at most 20 bp, at most 10 bp.

[00155] Preferably the distance between the two homologous regions is at most 10 kb, at most 9, at most 8 kb, at most 7 kb, at most 6 kb, at most 5 kb, at most 4 kb, at most 3 kb, at most 2 kb, at most 1 kb, at most 0.5 kb, at most 100bp, at most 50bp, at most 40bp, at most 30 bp, at most 20 bp, at most 10 bp.

[00156] In one aspect, the invention relates to a software algorithms able to identify PAM sites in the genome comprised between homology regions of about 7-20

bp in a neighborhood of the PAM site to design a method to target one or more PAM sites and create deletion of polynucleotides without use of a donor DNA.

[00157] The above method can be used for efficient removal of polynucleotide sequences in a designed way. For example, upon introducing a Cas9 expression cassette into the genomic DNA and after several rounds of modifications mediated by the CRISPR/CAS9 system, one can remove the CAS9 expression cassette from the genome by the introduction of a gRNA targeting a site in the Cas9 expression cassette and wherein the Cas9 expression cassette is comprised between two homologous regions as defined above, preferably 100-bp long, more preferably 20-bp, 15-bp long or shorter and cleave out the Cas9 open reading frame or a large part of the expression cassette.

[00158] The above method can also be used for transient inactivation of a gene. One could, for example, make a gene (e.g., a carotenoid synthase or saturated fatty acid synthase or polyunsaturated fatty acid synthase) non-functional by inserting a polynucleotide sequence in the ORF of the carotenoid synthase or saturated fatty acid synthase or polyunsaturated fatty acid synthase gene, comprising two homologous regions at the 5'-end and 3'-end respectively, wherein preferably the homologous regions are 100-bp, more preferably 20-bp, 15-bp long or shorter. An aforementioned synthase gene can be made functional again using a CRISPR-Cas9 system without donor DNA as described above.

[00159] In an embodiment, the method of modulating expression comprises cleavage of one or both polynucleotide strands in at least one location of the target-polynucleotide followed by modification of the target-polynucleotide by homologous recombination with an exogenous polynucleotide. In such case, the composition according to the first aspect of the present invention preferably further comprises such exogenous polynucleotide. Such modification may result in insertion, deletion or substitution of at least one nucleotide in the target-polynucleotide, wherein the insertion or substitution nucleotide may or may not originate from the exogenous polynucleotide. In one embodiment, the exogenous polynucleotide comprises regions of homology with the target-polynucleotide. Preferably, the degree of homology between these homologous regions is such to allow homologous recombination. Preferably, the homologous regions have at least 60%, 70%, 80%, 90%, 99% or 100%

sequence identity over the whole length of the homologous regions. A modification can also be made when the exogenous polynucleotide is a non-integrating entity; in this case the target-polynucleotide is modified but no nucleotide of the exogenous polynucleotide is introduced into the target-polynucleotide. Consequently, the resulting host is a non-recombinant host when the Cas-protein according to the present invention is introduced to the host cell as a protein. In a method according to this aspect of the present invention, the host cell may thus be a recombinant host cell or may be a non-recombinant host cell. The exogenous polynucleotide may be any polynucleotide of interest such as a polynucleotide encoding a compound of interest as defined herein, or a part of such polynucleotide or a variant thereof.

[00160] In another aspect, the present invention provides for a method for the production of a compound of interest, comprising culturing under conditions conducive to the compound of interest a host cell according to the third or fourth aspect of the present invention or a host cell obtained by a method according to the second aspect of the present invention, or a host cell obtainable by a method according to the fourth aspect of the present invention and optionally purifying or isolating the compound of interest.

[00161] A compound of interest in the context of all embodiments of the present invention may be any biological compound. The biological compound may be biomass or a biopolymer or a metabolite. The biological compound may be encoded by a single polynucleotide or a series of polynucleotides composing a biosynthetic or metabolic pathway or may be the direct result of the product of a single polynucleotide or products of a series of polynucleotides, the polynucleotide may be a gene, the series of polynucleotide may be a gene cluster. In all embodiments of the present invention, the single polynucleotide or series of polynucleotides encoding the biological compound of interest or the biosynthetic or metabolic pathway associated with the biological compound of interest, are preferred targets for the compositions and methods according to the present invention. The biological compound may be native to the host cell or heterologous to the host cell.

[00162] The term "heterologous biological compound" is defined herein as a biological compound which is not native to the cell; or a native biological compound in

which structural modifications have been made to alter the native biological compound.

[00163] The term “biopolymer” is defined herein as a chain (or polymer) of identical, similar, or dissimilar subunits (monomers). The biopolymer may be any biopolymer. The biopolymer may for example be, but is not limited to, a nucleic acid, polyamine, polyol, polypeptide (or polyamide), or polysaccharide.

[00164] The biopolymer may be a polypeptide. The polypeptide may be any polypeptide having a biological activity of interest. The term “polypeptide” is not meant herein to refer to a specific length of the encoded product and, therefore, encompasses peptides, oligopeptides, and proteins. The term polypeptide refers to polymers of amino acids of any length. The polymer may be linear or branched, it may comprise modified amino acids, and it may be interrupted by non-amino acids. The terms also encompass an amino acid polymer that has been modified; for example, disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation, such as conjugation with a labeling component. As used herein, the term “amino acid” includes natural and/or unnatural or synthetic amino acids, including glycine and both the D or L optical isomers, and amino acid analogs and peptidomimetics. Polypeptides further include naturally occurring allelic and engineered variations of the above-mentioned polypeptides and hybrid polypeptides. The polypeptide may be native or may be heterologous to the host cell. The polypeptide may be a collagen or gelatine, or a variant or hybrid thereof. The polypeptide may be an antibody or parts thereof, an antigen, a clotting factor, an enzyme, a hormone or a hormone variant, a receptor or parts thereof, a regulatory protein, a structural protein, a reporter, or a transport protein, protein involved in secretion process, protein involved in folding process, chaperone, peptide amino acid transporter, glycosylation factor, transcription factor, synthetic peptide or oligopeptide, intracellular protein. The intracellular protein may be an enzyme such as, a protease, ceramidases, epoxide hydrolase, aminopeptidase, acylases, aldolase, hydroxylase, aminopeptidase, lipase. The polypeptide may also be an enzyme secreted extracellularly. Such enzymes may belong to the groups of oxidoreductase, transferase, hydrolase, lyase, isomerase, ligase, catalase, cellulase, chitinase, cutinase, deoxyribonuclease, dextranase, esterase. The enzyme may be a

carbohydrase, e.g. cellulases such as endoglucanases, β -glucanases, cellobiohydrolases or β -glucosidases, hemicellulases or pectinolytic enzymes such as xylanases, xylosidases, mannanases, galactanases, galactosidases, pectin methyl esterases, pectin lyases, pectate lyases, endo polygalacturonases, exopolygalacturonases rhamnogalacturonases, arabanases, arabinofuranosidases, arabinoxylan hydrolases, galacturonases, lyases, or amylolytic enzymes; hydrolase, isomerase, or ligase, phosphatases such as phytases, esterases such as lipases, proteolytic enzymes, oxidoreductases such as oxidases, transferases, or isomerases. The enzyme may be a phytase. The enzyme may be an aminopeptidase, asparaginase, amylase, a maltogenic amylase, carbohydrase, carboxypeptidase, endo-protease, metallo-protease, serine-protease catalase, chitinase, cutinase, cyclodextrin glycosyltransferase, deoxyribonuclease, esterase, alpha-galactosidase, beta-galactosidase, glucoamylase, alpha-glucosidase, beta-glucosidase, haloperoxidase, protein deaminase, invertase, laccase, lipase, mannosidase, mutanase, oxidase, pectinolytic enzyme, peroxidase, phospholipase, galactolipase, chlorophyllase, polyphenoloxidase, ribonuclease, transglutaminase, or glucose oxidase, hexose oxidase, monooxygenase.

[00165] According to the present invention, a compound of interest can be a polypeptide or enzyme with improved secretion features as described in WO2010/102982. According to the present invention, a compound of interest can be a fused or hybrid polypeptide to which another polypeptide is fused at the N-terminus or the C-terminus of the polypeptide or fragment thereof. A fused polypeptide is produced by fusing a nucleic acid sequence (or a portion thereof) encoding one polypeptide to a nucleic acid sequence (or a portion thereof) encoding another polypeptide.

[00166] Techniques for producing fusion polypeptides are known in the art, and include, ligating the coding sequences encoding the polypeptides so that they are in frame and expression of the fused polypeptide is under control of the same promoter(s) and terminator. The hybrid polypeptides may comprise a combination of partial or complete polypeptide sequences obtained from at least two different polypeptides wherein one or more may be heterologous to the host cell. Examples of

fusion polypeptides and signal sequence fusions are, for example, as described in WO2010/121933.

[00167] The biopolymer may be a polysaccharide. The polysaccharide may be any polysaccharide, including, but not limited to, a mucopolysaccharide (e.g., heparin and hyaluronic acid) and nitrogen-containing polysaccharide (e.g., chitin). In a preferred option, the polysaccharide is hyaluronic acid.

[00168] A polynucleotide coding for the compound of interest or coding for a compound involved in the production of the compound of interest according to the invention may encode an enzyme involved in the synthesis of a primary or secondary metabolite, such as organic acids, carotenoids, (beta-lactam) antibiotics, and vitamins. Such metabolite may be considered as a biological compound according to the present invention.

[00169] The term "metabolite" encompasses both primary and secondary metabolites; the metabolite may be any metabolite. Preferred metabolites are unsaturated fats and lipids (including but not limited to the fatty acids docosahexaenoic acid, docosapentaenoic acid, eucric acid, paullinic acid, vaccenic acid (cis or trans), eicosapentaenoic acid, eicosatetraenoic acid (n-3), arachidonic acid (n-6), octadecapentaenoic acid, stearidonic acid, linolenic acid (n6 or n3), linoleic acid, oleic acid, palmitoleic acid, octacosaoctanoic acid, and lipids comprised thereof), saturated fats and lipids (including, but not limited to, the fatty acids: caproic acid, caprylic acid, capric acid, lauric acid, myristic acid, palmitic acid, stearic acid, arachidic acid, behenic acid, and lipids comprised thereof), sulfated lipids, sphingolipids, lactones, ceramides, phospholipids, fatty alcohols, fatty acid esters, fatty acid ethers, fatty acid aldehydes, sterols, carotenoids, oxylipins, resolvins, leukotrienes, prostaglandins, organic acids (including, but not limited to, acetic acid, butyric acid, citric acid, gluconic acid, adipic acid, fumaric acid, itaconic acid, malic acid, mevalonic acid, and succinic acid), sugar alcohols, and sugar acids.

[00170] A metabolite may be encoded by one or more genes, such as in a biosynthetic or metabolic pathway. Primary metabolites are products of primary or general metabolism of a cell, which are concerned with energy metabolism, growth, and structure. Secondary metabolites are products of secondary metabolism (see, for

example, R. B. Herbert, *The Biosynthesis of Secondary Metabolites*, Chapman and Hall, New York, 1981).

[00171] A primary metabolite may be, but is not limited to, an amino acid, fatty acid, nucleoside, nucleotide, sugar, triglyceride, or vitamin.

[00172] A secondary metabolite may be, but is not limited to, an alkaloid, coumarin, flavonoid, polyketide, quinine, steroid, peptide, or terpene. The secondary metabolite may be an antibiotic, antifeedant, attractant, bactericide, fungicide, hormone, insecticide, or rodenticide.

[00173] The biological compound may also be the product of a selectable marker. A selectable marker is a product of a polynucleotide of interest which product provides for biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophs, and the like. Selectable markers include, but are not limited to, nptII (neomycin phosphotransferase II), ALS (acetolactate synthase), bsd (blasticidin-S-deaminase), and Sh ble (phleomycin binding) as well as equivalents thereof.

[00174] According to the invention, a compound of interest is preferably a polypeptide as described in the list of compounds of interest.

[00175] According to another embodiment of the invention, a compound of interest is preferably a metabolite.

[00176] The host cell according to the present invention may already be capable of producing the compound of interest. The host cell may also be provided with a homologous or heterologous nucleic acid construct that encodes a polypeptide wherein the polypeptide may be the compound of interest or a polypeptide involved in the production of the compound of interest. The person skilled in the art knows how to modify an algal host cell such that it is capable of producing the compound of interest.

General Definitions

[00177] Throughout the present specification and the accompanying claims, the words “comprise”, “include” and “having” and variations such as “comprises”, “comprising”, “includes” and “including” are to be interpreted inclusively. That is, these words are intended to convey the possible inclusion of other elements or integers not specifically recited, where the context allows.

[00178] The terms “a” and “an” are used herein to refer to one or to more than one (i.e. to one or at least one) of the grammatical object of the article. By way of example, “an element” may mean one element or more than one element.

[00179] The word “about” or “approximately” when used in association with a numerical value (e.g. about 10) preferably means that the value may be the given value (of 10) more or less 1% of the value.

[00180] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this disclosure belongs.

[00181] A preferred nucleotide analogue or equivalent comprises a modified backbone. Examples of such backbones are provided by morpholino backbones, carbamate backbones, siloxane backbones, sulfide, sulfoxide and sulfone backbones, formacetyl and thioformacetyl backbones, methyleneformacetyl backbones, riboacetyl backbones, alkene containing backbones, sulfamate, sulfonate and sulfonamide backbones, methyleneimino and methylenehydrazino backbones, and amide backbones. It is further preferred that the linkage between a residue in a backbone does not include a phosphorus atom, such as a linkage that is formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages.

[00182] A preferred nucleotide analogue or equivalent comprises a Peptide Nucleic Acid (PNA), having a modified polyamide backbone (Nielsen, *et al.* (1991) *Science* 254, 1497-1500). PNA-based molecules are true mimics of DNA molecules in terms of base-pair recognition. The backbone of the PNA is composed of N-(2-aminoethyl)-glycine units linked by peptide bonds, wherein the nucleobases are linked to the backbone by methylene carbonyl bonds. An alternative backbone comprises a one-carbon extended pyrrolidine PNA monomer (Govindaraju and Kumar (2005) *Chem. Commun.* 495–497). Since the backbone of a PNA molecule contains no charged phosphate groups, PNA-RNA hybrids are usually more stable than RNA-RNA or RNA-DNA hybrids, respectively (Egholm *et al.* (1993) *Nature* 365, 566-568).

[00183] A further preferred backbone comprises a morpholino nucleotide analog or equivalent, in which the ribose or deoxyribose sugar is replaced by a 6-membered

morpholino ring. A most preferred nucleotide analog or equivalent comprises a phosphorodiamidate morpholino oligomer (PMO), in which the ribose or deoxyribose sugar is replaced by a 6-membered morpholino ring, and the anionic phosphodiester linkage between adjacent morpholino rings is replaced by a non-ionic phosphorodiamidate linkage.

[00184] A further preferred nucleotide analogue or equivalent comprises a substitution of at least one of the non-bridging oxygens in the phosphodiester linkage. This modification slightly destabilizes base-pairing but adds significant resistance to nuclease degradation. A preferred nucleotide analogue or equivalent comprises phosphorothioate, chiral phosphorothioate, phosphorodithioate, phosphotriester, aminoalkylphosphotriester, H-phosphonate, methyl and other alkyl phosphonate including 3'-alkylene phosphonate, 5'-alkylene phosphonate and chiral phosphonate, phosphinate, phosphoramidate including 3'-amino phosphoramidate and aminoalkylphosphoramidate, thionophosphoramidate, thionoalkylphosphonate, thionoalkylphosphotriester, selenophosphate or boranophosphate.

[00185] A further preferred nucleotide analogue or equivalent comprises one or more sugar moieties that are mono- or disubstituted at the 2', 3' and/or 5' position such as a -OH; -F; substituted or unsubstituted, linear or branched lower (C1-C10) alkyl, alkenyl, alkynyl, alkaryl, allyl, aryl, or aralkyl, that may be interrupted by one or more heteroatoms; O-, S-, or N-alkyl; O-, S-, or N-alkenyl; O-, S-, or N-alkynyl; O-, S-, or N-allyl; O-alkyl-O-alkyl, -methoxy, -aminopropoxy; aminoxy, methoxyethoxy; -dimethylaminoxyethoxy; and -dimethylaminoethoxyethoxy. The sugar moiety can be a pyranose or derivative thereof, or a deoxyribose or derivative thereof, preferably a ribose or a derivative thereof, or deoxyribose or derivative thereof. Such preferred derivatized sugar moieties comprise Locked Nucleic Acid (LNA), in which the 2'-carbon atom is linked to the 3' or 4' carbon atom of the sugar ring thereby forming a bicyclic sugar moiety. A preferred LNA comprises 2'-O,4'-C-ethylene-bridged nucleic acid (Morita *et al.* 2001. Nucleic Acid Res Supplement No. 1: 241-242). These substitutions render the nucleotide analogue or equivalent RNase H and nuclease resistant and increase the affinity for the target.

[00186] "Sequence identity" or "identity" in the context of the present invention of an amino acid- or nucleic acid-sequence is herein defined as a relationship between two

or more amino acid (peptide, polypeptide, or protein) sequences or two or more nucleic acid (nucleotide, oligonucleotide, polynucleotide) sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between amino acid or nucleotide sequences, as the case may be, as determined by the match between strings of such sequences. Within the present invention, sequence identity with a particular sequence preferably means sequence identity over the entire length of said particular polypeptide or polynucleotide sequence.

[00187] "Similarity" between two amino acid sequences is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one peptide or polypeptide to the sequence of a second peptide or polypeptide. In a preferred embodiment, identity or similarity is calculated over the whole sequence (SEQ ID NO:) as identified herein. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in Computational Molecular Biology, Lesk, A. M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D. W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A. M., and Griffin, H. G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heine, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., SIAM J. Applied Math., 48:1073 (1988).

[00188] Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include e.g. the GCG program package (Devereux, J., *et al.*, Nucleic Acids Research 12 (1): 387 (1984)), BestFit, BLASTP, BLASTN, and FASTA (Altschul, S. F. *et al.*, J. Mol. Biol. 215:403-410 (1990)). The BLAST X program is publicly available from NCBI and other sources (BLAST Manual, Altschul, S., *et al.*, NCBI NLM NIH Bethesda, MD 20894; Altschul, S., *et al.*, J. Mol. Biol. 215:403-410 (1990)). The well-known Smith Waterman algorithm may also be used to determine identity.

[00189] Preferred parameters for polypeptide sequence comparison is as follows: Algorithm: Needleman and Wunsch, J. Mol. Biol. 48:443-453 (1970); Comparison matrix:

BLOSSUM62 from Hentikoff and Hentikoff, Proc. Natl. Acad. Sci. USA. 89:10915-10919 (1992); Gap Penalty: 12; and Gap Length Penalty: 4. A program useful with these parameters is publicly available as the "Ogap" program from Genetics Computer Group, located in Madison, WI. The aforementioned parameters are the default parameters for amino acid comparisons (along with no penalty for end gaps).

[00190] Preferred parameters for nucleic acid comparison is as follows: Algorithm: Needleman and Wunsch, J. Mol. Biol. 48:443-453 (1970); Comparison matrix: matches=+10, mismatch=0; Gap Penalty: 50; Gap Length Penalty: 3. Available as the Gap program from Genetics Computer Group, located in Madison, Wis. Given above are the default parameters for nucleic acid comparisons.

[00191] Optionally, in determining the degree of amino acid similarity, the skilled person may also take into account so-called "conservative" amino acid substitutions, as will be clear to the skilled person. Conservative amino acid substitutions refer to the interchangeability of residues having similar side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulfur-containing side chains is cysteine and methionine. Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, and asparagine-glutamine. Substitutional variants of the amino acid sequence disclosed herein are those in which at least one residue in the disclosed sequences has been removed and a different residue inserted in its place. Preferably, the amino acid change is conservative. Preferred conservative substitutions for each of the naturally occurring amino acids are as follows: Ala to ser; Arg to lys; Asn to gln or his; Asp to glu; Cys to ser or ala; Gln to asn; Glu to asp; Gly to pro; His to asn or gln; Ile to leu or val; Leu to ile or val; Lys to arg; gln or glu; Met to leu or ile; Phe to met, leu or tyr; Ser to thr; Thr to ser; Trp to tyr; Tyr to trp or phe; and, Val to ile or leu.

[00192] A polynucleotide according to the present invention is represented by a nucleotide sequence. A polypeptide according to the present invention is represented by

an amino acid sequence. A nucleic acid construct according to the present invention is defined as a polynucleotide which is isolated from a naturally occurring gene or which has been modified to contain segments of polynucleotides which are combined or juxtaposed in a manner which would not otherwise exist in nature. Optionally, a polynucleotide present in a nucleic acid construct according to the present invention is operably linked to one or more control sequences, which direct the production or expression of the encoded product in a host cell or in a cell-free system.

[00193] The sequence information as provided herein should not be so narrowly construed as to require inclusion of erroneously identified bases. The skilled person is capable of identifying such erroneously identified bases and knows how to correct for such errors.

[00194] All embodiments of the present invention, i.e. a composition according to the present invention, a method of modulating expression, a host cell comprising a composition according to the present invention, a method of producing a host cell according to the present invention, a host cell according to the present invention and a method for the production of a compound of interest according to the present invention preferably refer to host cell, not to a cell-free *in vitro* system; in other words, the CRISPR-Cas systems according to the present invention are preferably host cell systems, not cell-free *in vitro* systems.

[00195] In all embodiments of the present invention, e.g. a composition according to the present invention, a method of modulating expression, a host cell comprising a composition according to the present invention, a method of producing a host cell according to the present invention, a host cell according to the present invention and a method for the production of a compound of interest according to the present invention, the host cell may be a haploid, diploid or polyploid host cell.

[00196] The host cell according to the present invention is a Labyrinthulomycetes host cell, preferably of the order Thraustochytriales, more preferably of the family Thraustochytriaceae, more preferably a member of a genus selected from the group consisting of *Aurantiochytrium*, *Oblongichytrium*, *Schizochytrium*, *Thraustochytrium*, and *Ulkenia*, even more preferably *Schizochytrium* sp. ATCC# 20888

[00197] A modification, preferably in the genome, is construed herein as one or more modifications. A modification, preferably in the genome of a host cell according to the present invention, can either be effected by

- a) subjecting a parent host cell to recombinant genetic manipulation techniques; and/or
- b) subjecting a parent host cell to (classical) mutagenesis; and/or
- c) subjecting a parent host cell to an inhibiting compound or composition.

[00198] Modification of a genome of a host cell is herein defined as any event resulting in a change in a polynucleotide sequence in the genome of the host cell.

[00199] Preferably, a host cell according to the present invention has a modification, preferably in its genome which results in a reduced or no production of an undesired compound as defined herein if compared to the parent host cell that has not been modified, when analysed under the same conditions.

[00200] A modification can be introduced by any means known to the person skilled in the art, such as but not limited to classical strain improvement, random mutagenesis followed by selection. Modification can also be introduced by site-directed mutagenesis.

[00201] Modification may be accomplished by the introduction (insertion), substitution (replacement) or removal (deletion) of one or more nucleotides in a polynucleotide sequence. A full or partial deletion of a polynucleotide coding for an undesired compound such as a polypeptide may be achieved. An undesired compound may be any undesired compound listed elsewhere herein; it may also be a protein and/or enzyme in a biological pathway of the synthesis of an undesired compound such as a metabolite. Alternatively, a polynucleotide coding for said undesired compound may be partially or fully replaced with a polynucleotide sequence which does not code for said undesired compound or that codes for a partially or fully inactive form of said undesired compound. In another alternative, one or more nucleotides can be inserted into the polynucleotide encoding said undesired compound resulting in the disruption of said polynucleotide and consequent partial or full inactivation of said undesired compound encoded by the disrupted polynucleotide.

[00202] In one embodiment, the mutant microbial host cell according to the invention comprises a modification in its genome selected from

- a) a full or partial deletion of a polynucleotide encoding an undesired compound,
- b) a full or partial replacement of a polynucleotide encoding an undesired compound with a polynucleotide sequence which does not code for said undesired compound or that codes for a partially or fully inactive form of said undesired compound,
- c) a disruption of a polynucleotide encoding an undesired compound by the insertion of one or more nucleotides in the polynucleotide sequence and consequent partial or full inactivation of said undesired compound by the disrupted polynucleotide.

[00203] This modification may, for example, be in a coding sequence or a regulatory element required for the transcription or translation of said undesired compound. For example, nucleotides may be inserted or removed so as to result in the introduction of a stop codon, the removal of a start codon or a change or a frame-shift of the open reading frame of a coding sequence. The modification of a coding sequence or a regulatory element thereof may be accomplished by site-directed or random mutagenesis, DNA shuffling methods, DNA reassembly methods, gene synthesis (see, e.g., Young and Dong (2004), *Nucleic Acids Research* 32(7) electronic access at nar.oupjournals.org/cgi/reprint/32/7/e59 or Gupta *et al.* (1968), *Proc. Natl. Acad. Sci USA*, 60: 1338-1344; Scarpulla *et al.* (1982), *Anal. Biochem.* 121: 356-365; Stemmer *et al.* (1995), *Gene* 164: 49-53), or PCR generated mutagenesis in accordance with methods known in the art. Examples of random mutagenesis procedures are well known in the art, such as, for example, chemical (e.g., NTG), mutagenesis or physical (e.g., UV) mutagenesis. Examples of site-directed mutagenesis procedures are the QuickChange™ site-directed mutagenesis kit (Stratagene Cloning Systems, La Jolla, CA), the 'The Altered Sites® II *in vitro* Mutagenesis Systems (Promega Corporation) or by overlap extension using PCR as described in Ho *et al.* ("Site-directed mutagenesis by overlap extension using the polymerase chain reaction", *Gene*, 1989 Apr 15, 77(1):51-9) or using PCR as described in *Molecular Biology: Current Innovations and Future Trends*. (Eds. A.M. Griffin and H.G. Griffin. ISBN 1-898486-01-8; 1995 Horizon Scientific Press, PO Box 1, Wymondham, Norfolk, U.K.).

[00204] Preferred methods of modification are based on recombinant genetic manipulation techniques such as partial or complete gene replacement or partial or complete gene deletion.

[00205] For example, in case of replacement of a polynucleotide, nucleic acid construct or expression cassette, an appropriate DNA sequence may be introduced at the target locus to be replaced. The appropriate DNA sequence is preferably present on a cloning vector. Preferred integrative cloning vectors comprise a DNA fragment, which is homologous to the polynucleotide and / or has homology to the polynucleotides flanking the locus to be replaced for targeting the integration of the cloning vector to this pre-determined locus. In order to promote targeted integration, the cloning vector is preferably linearized prior to transformation of the cell. Preferably, linearization is performed such that at least one but preferably either end of the cloning vector is flanked by sequences homologous to the DNA sequence (or flanking sequences) to be replaced. This process is called homologous recombination and this technique may also be used in order to achieve (partial) gene deletion.

[00206] For example, a polynucleotide corresponding to the endogenous polynucleotide may be replaced by a defective polynucleotide, that is a polynucleotide that fails to produce a (fully functional) polypeptide. By homologous recombination, the defective polynucleotide replaces the endogenous polynucleotide. It may be desirable that the defective polynucleotide also encodes a marker, which may be used for selection of transformants in which the nucleic acid sequence has been modified.

[00207] A modification which results in decreased or no production of an undesired compound can be obtained by different methods, for example by an antibody directed against such undesired compound or a chemical inhibitor or a protein inhibitor or a physical inhibitor (Tour O. *et al.*, "Genetically targeted chromophore-assisted light inactivation", (2003) Nat. Biotech). Alternatively, or in combination with above-mentioned techniques, decreased or no production of an undesired compound can also be obtained, e.g. by UV or chemical mutagenesis (Lian *et al.*, "Increase of docosahexaenoic acid production by *Schizochytrium* sp. through mutagenesis and enzyme assay" (2010) Appl Biochem Biotechnol 162(4):935-941) or by the use of inhibitors inhibiting enzymatic activity of an undesired polypeptide as described herein (e.g. nojirimycin, which function

as inhibitor for β -glucosidases (Ren *et al.*, "Effect of biotin and cerulenin addition on DHA production by *Schizochytrium sp.*" (2012-01) Chinese J Bioprocess Engineering).

[00208] In an embodiment of the present invention, the modification in the genome of the host cell according to the invention is a modification in at least one position of a polynucleotide encoding an undesired compound.

[00209] A deficiency of a cell in the production of a compound, for example of an undesired compound such as an undesired polypeptide and/or enzyme is herein defined as a mutant microbial host cell which has been modified, preferably in its genome, to result in a phenotypic feature wherein the cell: a) produces less of the undesired compound or produces substantially none of the undesired compound and/or b) produces the undesired compound having a decreased activity or decreased specific activity or the undesired compound having no activity or no specific activity and combinations of one or more of these possibilities as compared to the parent host cell that has not been modified, when analysed under the same conditions.

[00210] Preferably, a modified host cell according to the present invention produces 1% less of the un-desired compound if compared with the parent host cell which has not been modified and measured under the same conditions, at least 5% less of the un-desired compound, at least 10% less of the un-desired compound, at least 20% less of the un-desired compound, at least 30% less of the un-desired compound, at least 40% less of the un-desired compound, at least 50% less of the un-desired compound, at least 60% less of the un-desired compound, at least 70% less of the un-desired compound, at least 80% less of the un-desired compound, at least 90% less of the un-desired compound, at least 91% less of the un-desired compound, at least 92% less of the un-desired compound, at least 93% less of the un-desired compound, at least 94% less of the un-desired compound, at least 95% less of the un-desired compound, at least 96% less of the un-desired compound, at least 97% less of the un-desired compound, at least 98% less of the un-desired compound, at least 99% less of the un-desired compound, at least 99.9% less of the un-desired compound, or most preferably 100% less of the un-desired compound.

[00211] The sequence information as provided herein should not be so narrowly construed as to require inclusion of erroneously identified bases. The skilled person is

capable of identifying such erroneously identified bases and knows how to correct for such errors.

[00212] The disclosure of each reference set forth herein is incorporated herein by reference in its entirety.

[00213] The present invention is further illustrated by the following examples:

EXAMPLES

EXAMPLE 1

[00214] Design and building of Cas9 vectors for transformations

[00215] The *Streptococcus pyogenes* MGAS5005 Cas9 sequence was codon-optimized for expression in *Schizochytrium*. The basic Cas9 peptide sequence used was essentially the same as that used by Jinek *et al.*, 2013, but without the described HA-tag or the GFP fusion (see below). After the Cas9 was synthesized (DNA2.0, Newark, CA), it was cloned into pCL122 vector containing a paromomycin selection cassette, resulting in a vector designated pCL122-Cas9 (SEQ ID). OrfA/Pfa1 flanks were synthesized and cloned (Genscript USA, Piscataway, NJ) into pCL121 (which harbored a zeocin selection cassette) and the final vector was designated pYB31 (SEQ ID NO:2); see, e.g., US 8,940,884. The Cas9 gene was subcloned from pCL122-Cas9 vector into pYB31 vector via digestion and ligation using *Bam*HI and *Nde*I restriction sites and enzymes (New England Biolabs, Ipswich, MA) (FIG. 3), and Rapid DNA Ligation kit (Roche, Risch-Rotkreuz, Switzerland). NEB10 β chemically competent cells were transformed (New England Biolabs) and resulting colonies were screened by colony PCR using selected colonies, GoTaq Green Master mix (Promega, Chicago, IL) with primers 121 Tub seq F (SEQ ID NO:13) and pYB32/3C R1 (SEQ ID NO:14) (Table 2), and 5% DMSO (v:v), and the following cycling conditions were applied to the reaction: 95°C for 5 minutes, (95°C for 30 seconds, 59°C for 30 seconds, 72°C for 1 minute) x35 cycles, 72°C for 5 minutes (FIG. 4). Plasmids were isolated from colonies positive by PCR, sequenced, and one resulting vector was designated pYB32 (SEQ ID NO:3).

TABLE 2

Primer Name	SEQ ID NO	Primer Sequence, 5'→3'
121 Tub seq F	13	GGATCTCATGCTGGAGTTCTTC
pYB32/3C R1	14	GTA CTTCTCGTGGTAGGCAACC

[00216] As previously mentioned, other studies have indicated that the level of Cas9 expression is important for optimizing transformant viability and nuclease activity. In the pYB32 plasmid, the Cas9 gene is under control of the alpha tubulin promoter, a medium strength promoter. A weak promoter from the carotene synthase (CS) gene was selected to be tested as well and was cloned into pYB31 via InFusion PCR Cloning (Clontech/Takara Bio USA, Inc., Mountain View, CA). pYB31 plasmid was digested with *KpnI* and *NdeI* enzymes (New England Biolabs) in preparation for InFusion cloning. CS promoter was amplified with CS pro *KpnI* IF F1 (SEQ ID NO:15) and CS pro *BamHI* IF R1 primers (SEQ ID NO:16) (Table 3), with 5% DMSO (v:v), KOD HotStart Master Mix (EMD Millipore, Billerica, MA), and pTH043, plasmid harboring CS promoter, used as a template. Reaction cycling conditions were as follows: 95°C for 2 minutes, (95°C for 20 seconds, 66°C for 10 seconds, 70°C for 20 seconds) x35 cycles, 70°C for 2 minutes. Cas9 was amplified with CS pro *BamHI* IF F2 (SEQ ID NO:17) and CS pro *NdeI* IF R2 (SEQ ID NO:18) primers (Table 3), with 5% DMSO (v:v), KOD HotStart Master Mix, and pCL122-Cas9 used as a template. Reaction cycling conditions were as follows: 95°C for 2 minutes, (95°C for 20 seconds, 65°C for 10 seconds, 70°C for 1 minute 45 seconds) x35 cycles, 70°C for 2 minutes.

TABLE 3

Primer Name	SEQ ID NO	Primer Sequence, 5'→3'
CS pro <i>KpnI</i> IF F1	15	GTCTGAATTC CCGGGTACCGAGCGGGCGATTCCACCGTC
CS pro <i>BamHI</i> IF R1	16	GTA CTTCTTATCCATGGATCCCTCGGTCTCCGAGCGAGCGAG
CS pro <i>BamHI</i> IF F2	17	TCGCTCGCTCGGAGACCGAGGGATCCATGGATAAGAAGTAC
CS pro <i>NdeI</i> IF R2	18	GATTCACTAGTTTAGATCATATGTTAGACCTTGCGCTTCTTCTTAGGGTCC

[00217] PCR products were run on 1% agarose gel (Lonza, Basel, Switzerland), bands of expected sizes were excised and gel purified using QIAquick Gel Extraction kit (Qiagen, Hilden, Germany) (FIG. 3). Following manufacturer's protocol for InFusion cloning, the Cas9 PCR fragment, CS promoter PCR fragment and linearized pYB31 fragment were ligated, NEB10 β chemically competent cells were transformed, and resulting colonies were screened by colony PCR as described above (FIG. 4) (with 1 minute 30 seconds extension time at 72°C). After sequence verification, one of the resulting vectors containing Cas9 operably linked to the CS promoter was designated pYB33 (SEQ ID NO:4).

[00218] Cloning of pYB32 and pYB33 was done through a combination of digestions, ligations and InFusion PCR cloning. As shown in FIG. 3, all the fragments were produced and had bands of expected sizes after column purification. Digested plasmids and PCR amplified fragments were run on 1% agarose gel. Expected sizes of fragments were observed in all lanes as follows: pYB31-BamHI+NdeI=838 bp + 6555 bp, pYB31-KpnI+NdeI=1289bp + 6104bp, Cas9-BamHI+NdeI=4157bp + 5773bp, CS PCR fragment=1046 bp, Cas9 PCR fragment=4189 bp. The molecular weight markers used in this gel and all subsequent gels was DNA Quanti-Ladder (Origene, Rockville, MD) - fragment sizes are indicated in the panel on the right side of FIG. 3.

[00219] Colonies resulting from pYB32 and pYB33 bacterial transformations were analyzed by colony PCR, and colonies with amplicons of the expected size are marked in boldface type in FIG. 4. These colonies are: pYB32-2, -4, -9, -14, -15, -24, -25, -28, and pYB33-16 and pYB33D-1, -14. pYB32-2 and pYB33-16 were among those verified to have correct sequence and were used for subsequent work. For colony PCR results from cloning of pYB32 and pYB33, expected amplicon sizes are as follows: pYB32=1028 bp, pYB33=1584bp.

EXAMPLE 2

[00220] Transformations and selection of Cas9 transformants

[00221] The wild-type strain of *Schizochytrium* sp., ATCC 20888, was used for transformation with pYB32 and pYB33 using a particle bombardment method and a biolistics instrument (Bio-Rad, Raleigh, NC) as described below. Briefly, 20888 was

grown in 25 mL of M50-20 medium (see, e.g., US 8,003,772) in a 250 mL smooth-bottom Erlenmeyer flask at 27°C, shaking at 200 rpm overnight. Following that, the culture was diluted 1/100 into 50 mL of M2B medium (see, e.g., US 8,003,772) in a 250 mL baffled flask and grown under the same conditions overnight. When the *Schizochytrium* culture was in early log phase (0.6-2 OD units/mL), the culture was harvested by centrifugation at 3,220 x g for 10 minutes. Supernatant was decanted, the pellet was resuspended in M2B to final concentration of 20 OD units/mL, and 100 µL of resulting cell suspension was spread in a circular motion on approximately a third of a non-selective M2B agar plate (approximately 4 cm diameter). Plasmids pYB32 and pYB33 were digested with *Swa*I at 25°C overnight and purified using QIAquick PCR purification kit (Qiagen, Hilden, Germany). Five µg of each purified plasmid was then mixed with 50 µL of 2.5 M CaCl₂, 20 µL of 0.1 M spermidine and 50 µL of prepared M10 Tungsten beads (following the manufacturer's protocol, Bio-Rad), vortexed for 1 minute, then incubated for 10 minutes at room temperature to allow for the beads to settle. DNA-coated beads were washed once with 250 µL of 100% ethanol, and then beads were resuspended in 60 µL of 100% ethanol. Each prepared macrocarrier (following manufacturer's protocol for macrocarrier assembly preparation) had 10 µL of coated beads in ethanol spotted to the center of a macrocarrier disc and ethanol was allowed to dry. Rupture disc holders had 1,100 psi-rated rupture discs placed inside after brief sterilization in 70% isopropanol, and assembled macrocarriers were fitted to the biolistics platform in the top shelf position and M2B agar plates with cell patches of *Schizochytrium* were placed cell-side-up on the third shelf from the top. When vacuum reached ~27 psi inside the biolistics chamber, helium was fired until rupture disc failure, the flow of helium was closed off, the chamber was vented to atmosphere, and the bombarded plate was removed from the chamber. The bombardment process was repeated for all samples and controls, *Schizochytrium* sp. 20888 bombarded with pYB32 was designated T188, and *Schizochytrium* cultures bombarded with pYB33 were designated T189. Bombarded plates were incubated in at 27°C for 4 hours, after which cells were washed off the plate with ~1 mL of M2B and divided equally between four M2B agar plates containing 0.5 mM DHA (Nu-Chek Prep, Waterville, MN) in 0.4% randomly-methylated β-cyclodextrins (CTD Holdings, Inc., High Springs, FL) and 50 µg/mL zeocin (Thermo

Fisher Scientific, Waltham, MA). Cells were spread with 3 mm sterile glass beads, beads removed and plates wrapped and incubated at 27°C for ~7-10 days. When colonies reached 2-4 millimeters in size, each colony was replica-patched on M2B agar plates containing 100 µg/mL zeocin with or without DHA (complexed with randomly-methylated β-cyclodextrins as described above) to select for auxotrophy induced by disruption of the OrfA/Pfa1 gene.

[00222] Colonies confirmed to be auxotrophic for DHA were picked and inoculated into 50 mL of M50-20 supplemented with 500 µL of 40% β-cyclodextrins and 50 mM DHA solution in a 250 mL smooth bottom Erlenmeyer flask and placed in a 27°C shaker at 200 rpm for 48 hrs. After 48 hr incubation time, 2 mL of culture were collected by centrifugation at 5,000 x g for 10 minutes, supernatant was decanted, and pellet was used for genomic DNA isolation following a modified phenol-chloroform extraction. Genomic DNA was extracted and used as PCR template with GoTaq Green mastermix (Promega, Durham, NC) to verify presence of zeocin cassettes and the Cas9 gene using primers, GoTaq Green master mix and cycling conditions as described for colony PCR screening for pYB32 and pYB33 (FIG. 5). Disruption of the OrfA/Pfa1 gene in T188 and T189 transformants was verified by PCR. Specifically, 5' flanking regions were interrogated by PCR with O A1-KO F (SEQ ID NO:19) and pYB32/3 SV40 R1 (SEQ ID NO:20) primers (Table 4), GoTaq Green mastermix, 5% DMSO (v:v) and T188/T189 gDNA as template, and cycling conditions were as follows: 95°C for 2 minutes, (95°C for 30 seconds, 60°C for 30 second, 72°C for 2 minutes 15 seconds) x35 cycles, 72°C for 5 minutes. Furthermore, 3' flanking region were interrogated by PCR with O A1-KO R (SEQ ID NO:21) and pYB32/2 CF 1 (SEQ ID NO:22) primers (Table 4), GoTaq Green mastermix, 5% DMSO (v:v) and T188/T189 gDNA as template, cycling conditions were as follows: 95°C for 2 minutes, (95°C for 30 seconds, 59°C for 30 seconds, 72°C for 2 minutes) x35 cycles, 72°C for 5 minutes (FIG. 6).

TABLE 4

Primer Name	SEQ ID NO	Primer Sequence, 5'→3'
O A1-KO F	19	CCAAGTTCGCCAAGGCTTC
pYB32/3 SV40 R1	20	GTGGAATCGAAATCTCGTAGCAC
O A1-KO R	21	GCTGTTGCAACTTTGCTCCAC
pYB32/3C F1	22	GTTAAGAAGACCGAGGTCCAGAC

[00223] Cultures positive by PCR for Cas9 gene integration and a disruption at OrfA/Pfa1 were inoculated into 25 mL of M50-20 supplemented with randomly methylated β -cyclodextrin and DHA, and grown for 24 hours as described previously. The OD600 of the cultures were measured, and 2 OD units of each culture were inoculated into 50 mL of SPFM, pH 6.75, supplemented with randomly methylated β -cyclodextrin and DHA in 250 mL baffled Erlenmeyer flasks, and grown for 48 hrs in a 27°C shaker at 200 rpm before addition of 5% glycerol (v:v) for cryopreservation. Two clones were selected, one from each transformation group, for further work -T188-1-20 and T189-1-20.

[00224] Wild-type *Schizochytrium* sp. 20888 was transformed with pYB32 and pYB33 plasmids by bombardment resulting in DHA auxotrophy in 20888 due to insertion of Cas9 at OrfA/PFA1 locus of the polyunsaturated fatty acid (PUFA) synthase. Colonies were counted and replica-patched on selective plates with or without DHA. Colonies that were auxotrophic for DHA were picked at random from both transformation groups and were used for further characterization.

[00225] As shown below in Table 5, transformation efficiency with pYB33 plasmid was lower than with pYB32 plasmid, and a similar trend was observed with the fraction of auxotrophic colonies found.

TABLE 5

Transformation	Total Colonies per 5 μ g DNA	Auxotrophic Colonies	% Knock-out
T188: pYB32—Tub pro-Cas9	160	80	50%
T189: pYB33—CS pro-Cas9	96	37	38.5%

[00226] T188 and T189 transformants were interrogated for the presence of promoters linked to Cas9 and for the upstream region of the Cas9 gene. For T188 clones, the expected amplicon size was 1028 bp, and for T189 - 1584 bp. As shown in FIG. 5, all but clone T189-1-30 generated the expected amplicon size and were used for further testing.

[00227] Integration of Cas9 in the OrfA/Pfa1 locus was verified by PCR, both 5' and 3' flanking regions were amplified with a combination of wild-type-specific external primers and Cas9-specific internal primers, such that amplicons were achievable only

if integration occurred (FIG. 6). Expected amplicon sizes for the 5' flank of the Cas9 integration site was 2229 bp, and for the 3' flank was 1973 bp. All of the strains tested were positive for integration of Cas9 at OrfA/Pfa1 locus.

EXAMPLE 3

[00228] Design and building of gRNA vectors

[00229] All of the guide RNA (gRNA) cassettes were designed to be expressed by the elongation factor 1 α (EF-1 α) promoter derived from *Schizochytrium*. In each case, the gRNA sequence was flanked by two ribozyme sequences, Hammerhead and HDV (Gao Y and Zhao Y "Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing" J Integr Plant Biol. 56(4):343-349 (2014)). Four target sequences containing the "NGG" protospacer-adjacent motif were identified within carotene synthase gene and 20 bp immediately upstream of each were chosen as target sequences. One cassette including both ribozymes, target sequence and guide RNA was synthesized and cloned into pCL122 vector using *Bgl*III and *Nde*I sites (DNA2.0) resulting in the vector designated pCL401 - a precursor plasmid for additional cloning (SEQ ID NO:7). To prepare the final gRNA cassette vectors, the plasmid pSP73 (Promega) was digested with *Nde*I and *Hpa*I and the resulting larger fragment was gel purified. A fragment of the *Schizochytrium* Fatty Acid Synthase fragment (FAS) was amplified by PCR off *Schizochytrium* sp. 20888 genomic DNA with 5' FAS PmeNde (SEQ ID NO:23) and 3' FAS PmeHpa (SEQ ID NO:24) primers (Table 6), KOD Hot Start Mastermix, and 5% DMSO, using the following cycling conditions: 95°C for 2 minutes, (95°C for 20 seconds, 62°C for 10 seconds, 70°C for 51 seconds) x40 cycles, 70°C for 5 minutes (FIG. 7). The resulting fragment was digested with *Hpa*I and *Nde*I restriction enzymes and purified.

TABLE 6

Primer Name	SEQ NO	ID	Primer Sequence, 5'→3'
5' FAS PmeNde	23		TAGCATATGTTTAAACTCGCGGCGTCTTTTCGC
3' FAS PmeHpa	24		AGTTAACGTTTAAACAGAGGAGGTGGCTGGC

[00230] The appropriate purified and digested pSP73 and FAS fragments were ligated using Rapid DNA Ligation kit (Roche) and transformed into NEB10 β cells (New England Biolabs), resulting in vector pCL399 (SEQ ID NO:5). The pCL399 vector was then digested with *NdeI* endonuclease alone, the ends were blunted with Mung Bean Nuclease (New England Biolabs), and the vector religated with Rapid DNA Ligation kit, thereby removing *NdeI* site. The resulting vector was designated pCL400 (SEQ ID NO:6). The paromomycin expression cassette and gRNA cassette were amplified using pCL401 (SEQ ID NO:7) as template to be inserted into pCL400 predigested with *XhoI* by InFusion PCR cloning to create vector pCL402 (SEQ ID NO:8). Briefly, pCL402 IF F (SEQ ID NO:25) and PCL402 IF R (SEQ ID NO:26) primers (Table 7), 5% DMSO (v:v), KOD Hot Start Mastermix, pCL401 plasmid template were used to amplify the desired fragment from pCL401 using the following cycling conditions: 95°C for 2 minutes (95°C for 20 seconds, 60°C for 10 seconds, 70°C for 1 minute 25 seconds) x35 cycles, 70°C for 5 minutes (FIG. 8). The resulting vector was designated pCL402.

TABLE 7

Primer Name	SEQ ID NO	Primer Sequence, 5'→3'
pCL402 IF F	25	GAGGCGCTGACCGCCGGCCAAGCTTCCAATTTTAGGCC
pCL402 IF R	26	GCAGGTGCCGAGTTTCTCGAGAAGAATCTGAACTCACGTC

[00231] Four genomic Carotene Synthase target sequences were synthesized (Eurofins, Mebane, NC) with *BglII* and *NdeI* flank sites. These fragments were designated gRNA3 CS1, CS2, CS3 and CS4. All four fragments and the vector pCL402 were digested with *BglII* and *NdeI*. All appropriate fragments were gel or column purified and then ligated together using Rapid DNA Ligation kit (Roche) and resulting vectors were designated pYB36 (gRNA3 CS1; SEQ ID NO:9), pYB37 (gRNA3 CS2; SEQ ID NO:10), pYB38 (gRNA3 CS3; SEQ ID NO:11), and pYB39 (gRNA3 CS4; SEQ ID NO:12).

[00232] In order to remove gene editing components from *Schizochytrium*, both Cas9 and guide RNA cassettes were designed to disrupt (“knock-out”) genes thereby inducing auxotrophies. While the auxotrophy with Cas9 selection was for DHA, the induced auxotrophy for gRNA insertion was for palmitic acid due to disruption of the Fatty Acid Synthase locus (FAS). Part of the FAS gene was amplified from *Schizochytrium* sp. 20888 genomic DNA and then run on a 1% agarose gel (FIG. 8). The expected size of the amplicon was 2530 bp. Two samples were prepared and loaded on the gel, the expected size band was the major product of the PCR, however there were other smaller size bands indicating that PCR conditions might have been less ideal and may need to be optimized in the future. The 2.5 kbp band was cut out, gel purified and used for cloning of the vector pCL399.

[00233] PCR was set up for cloning of paromomycin and gRNA cassettes into FAS locus in vector pCL400. Expected size of the amplicon was 3376 bp, and it was the major band that appeared on the gel (FIG. 8). This fragment was then cloned into pCL400 to result in pCL402.

EXAMPLE 4

[00234] Transformation and selection of gRNA constructs

[00235] T188-1-20 and T189-1-20 strains, both verified by PCR to contain Cas9 expression cassette under regulation of different promoters, Alpha Tubulin and CS, respectively, and inserted at OrfA/PFA1 locus of PUFA synthase, were inoculated into smooth-bottom Erlenmeyer flasks containing 25 mL of M50-20 supplemented with 0.4% randomly methylated β -cyclodextrins with 0.5 mM DHA and grown overnight at 27°C, 200 rpm. Both strains were diluted 1/50 into baffled flasks containing 50 mL of M2B supplemented with β -cyclodextrin and DHA as described above and grown overnight at 27°C. T188 and T189 cultures were harvested the following day as described above and preparation for bombardment was carried out as described above. pYB36-39 plasmids were digested prior to bombardment with *PmeI* and column purified, all four digested plasmids were used to transform the two strains resulting in transformations designated T202-209 (T188 with pYB36-39, T189 with pYB36-39, respectively). Resulting transformants were plated on M2B plates containing 0.4% of β -cyclodextrins with 0.5 mM DHA, 0.5 mM C16:0 (palmitic acid,

Sigma-Aldrich, St. Louis, MO) and 50 µg/mL paromomycin 4 hours after recovery from bombardment. Plates were wrapped and incubated at 27°C for 7-10 days. Once 2-4mm in diameter, colonies were replica patched onto M2B+0.4% β-cyclodextrins+0.5 mM DHA+50 µg/mL paromomycin with or without 0.5 mM palmitic acid to verify auxotrophy for palmitic acid created as a result of gRNA cassette insertion at FAS locus. Ten colonies from each of the 8 transformations verified to be auxotrophic for both DHA and palmitic acid were picked and inoculated into smooth-bottom Erlenmeyer flasks containing 25 mL of M50-20 supplemented with 0.4% β-cyclodextrins, 0.5 mM DHA and 0.5 mM palmitic acid, cultures were grown for 48 hrs at 27°C, 200 rpm, at which time 2 mL of each culture was collected for genomic DNA preparations. After genomic DNA was extracted with phenol:chloroform, PCR was carried out to amplify carotene synthase locus containing regions where gRNA cassette was targeted. gDNA from selected transformants from T202, 203, 204, 207, 208 was amplified by PCR using pYB36 CS1 F (SEQ ID NO:27) and pYB36 CS3 R (SEQ ID NO:29) primers (Table 8), 5% DMSO (v:v), KOD Hot Start Mastermix, respective gDNA as template using following cycling conditions: 95°C for 2 minutes, (95°C for 20 seconds, 61°C for 10 seconds, 70°C for 21 seconds) x35 cycles, 70°C for 5 minutes (FIG. 9). gDNA from T206 was amplified with pYB36 CS1 F and pYB36 CS1 R (SEQ ID NO:28) primers (Table 8), 5% DMSO (v:v), KOD Hot Start Mastermix using following cycling conditions: 95°C for 2 minutes, (95°C for 20 seconds, 61°C for 10 seconds, 70°C for 11 seconds) x35 cycles, 70°C for 5 minutes (FIG. 10). gDNA from T205 and T209 was amplified using pYB36 CS4 F (SEQ ID NO:30) and pYB36 CS4 R (SEQ ID NO:31) primers (Table 8), 5% DMSO (v:v), KOD Hot Start Mastermix using following cycling conditions: 95°C for 2 minutes, (95°C for 20 seconds, 61°C for 10 seconds, 70°C for 21 seconds) x35 cycles, 70°C for 5 minutes (FIG. 9).

TABLE 8

Primer Name	SEQ ID NO	Primer Sequence, 5'→3'
pYB36 CS1 F	27	GAGTCGAAGGAGACGTTGTTCG
pYB36 CS1 R	28	GTCATTGCGAATGATGCGATATG
pYB36 CS3 R	29	GGTCATCATGGAATACAACGCAG
pYB36 CS4 F	30	CGAGCTCATTTGTGCTACACTCTATG
pYB36 CS4 R	31	CACAAGATTTGCAGGATTGATGC

[00236] Resulting PCR amplicons were column purified, cloned into pJet1.2 vector (Thermo Fisher, Waltham, MA), resulting bacterial transformants were minipreped, and the inserts sequenced with primers provided in the kit. Sequences were aligned using Geneious software.

[00237] Transformations of two Cas9 backgrounds (T188 and T189) with vectors encoding gRNAs were carried out successfully and resulted in a number of doubly auxotrophic strains as shown below in Table 9 (designated T202-209). Select strains were later analyzed on a genetic level to detect any mutations of the carotenoid synthase locus, indicating a non-homologous end joining (NHEJ) event.

TABLE 9

Transformation	Total colonies (per 3ug DNA)	Number of 16:0 auxotrophs	Percent KO
T202 (T188+gRNA1)	103	17/32	53%
T203 (T188+gRNA2)	93	15/32	46%
T204 (T188+gRNA3)	65	15/32	46%
T205 (T188+gRNA4)	50	15/32	46%
T206 (T189+gRNA1)	53	12/32	37%
T207 (T189+gRNA2)	28	8/20	40%
T208 (T189+gRNA3)	38	9/29	31%
T209 (T189+gRNA4)	28	9/22	41%

[00238] After gDNA was isolated from gRNA + Cas9 transformants that were shown to be auxotrophic, requiring supplementation with both DHA and palmitic acid, one PCR was used to amplify the entire region encompassing the first 3 CS gRNA targets as one amplicon and another PCR to amplify the gRNA3 CS4 target separately (FIG. 9). The expected size for amplicons from gRNA3 CS1-3 transformant DNA was 1040 bp, and from gRNA 3 CS4 transformants was 715 bp. In the case of amplicons for gRNA CS1-3 targets, samples that came from the T188 lineage had a single band of correct size. However, samples that came from T189 (including gDNA from T189 parental strain itself) had multiple bands with ~1 kbp band being most prominent and

~1.5 kbp band being second most prominent. The 1.5 kbp band was judged to be a product of non-specific amplification. No optimization of the PCR was done in an attempt to reduce the presence of nonspecific bands, and the ~1 kbp band was cut out of the gel for all the transformants and was used for pJet1.2 cloning. All CS4 amplicons were a single band of correct size. Several resulting pJet transformant colonies harboring amplicons from each of the Cas9/gRNA transformants were sent for sequencing to determine whether Cas9/gRNA had any effect on CS sequence.

[00239] Additionally, analyses were performed on the T206 lineage of clones. As described in FIG. 9, T206 transformants had gDNA extracted and used as template for PCR to amplify the gRNA3 CS1 target sequence only and the expected amplicon size was 689 bp. A band of this size was observed on the gel, a major band amidst other, higher molecular weight ones, similar to those observed on the previous figure. The gel fragment of expected ~700 bp size was cut and purified and used for pJet1.2 cloning and sequencing of the fragment to determine whether Cas9-gDNA had induced any sequence changes in CS gene (FIG. 10).

[00240] Upon sequencing multiple pJet vectors containing amplicons of CS locus from clones transformed with both Cas9 and gRNA encoding vectors, it was observed that several kinds of changes occurred (Table 10). Previous work has shown that CAS9 cleaves DNA and makes a double stranded break at a position three base pairs upstream of the PAM sequence (Jinek *et al.* "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity" *Science* 337(6096):816-821, 2012). It is known at the site of the double stranded break mutations might arise from imprecise non-homologous end joining (NHEJ)-mediated repair that can produce insertion and/or deletion mutations of variable length (Sander and Joung, "CRISPR-Cas systems for editing, regulating and targeting genomes" *Nat Biotechnol.* 32(4):347-355, 2014). All of the changes observed occur in the -3 position relative to PAM (AGG in this case), as expected. Exemplary mutations of the carotenoid synthase of *Schizochytrium* sp. 20888 detected in Cas9+gRNA transformants are presented in Table 10, with deletions shown as dashes and insertions shown as underlined text.

TABLE 10

Wild-type nucleotide position #	15	16	17	18	19	20	PAM/NGG
Wild-type	A	C	G	C	G	C	AGG
CS1 target-variant #1	A	C	G	C	-	-	AGG
CS1 target-variant #2	A	C	-	C	G	C	AGG
CS1 target-variant #3	A	C	G	<u>CC</u>	G	C	AGG
CS1 target-variant #4	A	C	G	<u>TC</u>	G	C	AGG

[00241] All references cited in this specification are herein incorporated by reference as though each reference was specifically and individually indicated to be incorporated by reference. The citation of any reference is for its disclosure prior to the filing date and should not be construed as an admission that the present disclosure is not entitled to antedate such reference by virtue of prior invention.

[00242] It will be understood that each of the elements described above, or two or more together may also find a useful application in other types of methods differing from the type described above. Without further analysis, the foregoing will so fully reveal the gist of the present disclosure that others can, by applying current knowledge, readily adapt it for various applications without omitting features that, from the standpoint of prior art, fairly constitute essential characteristics of the generic or specific aspects of this disclosure set forth in the appended claims. The foregoing embodiments are presented by way of example only; the scope of the present disclosure is to be limited only by the following claims.

EXAMPLE 5

[00243] Design and build of non-targeted Cas9 vector (pYB61)

[00244] Previously, Cas9 plasmids were designed for targeted integration at a PUFA synthase gene (OrfA subunit; see pYB32 and pYB33). Integration of a Cas9 cassette designed in this way resulted in a knockout of the OrfA gene and DHA auxotrophy. Induction of auxotrophy enables easier screening of transformants for subsequent integration at OrfA locus, but made the differentiation of CRISPR-mediated carotene synthase knockouts more difficult. A plasmid with the expression

cassette for Cas9 without targeting to any chromosomal locus was built to test the effect of ectopic, random integration. The plasmid pCL122-Cas9 (SEQ ID NO:1) was digested with BamHI-HF (NEB) and NdeI (NEB) resulting in two fragments, 4140 bp and 5780 bp in size. The 4140 bp fragment was separated on a 1% agarose gel, excised, and purified using QIAquick Gel extraction kit (Qiagen) (FIG. 11). The plasmid pYB30 (SEQ ID NO:32) contains a partial fragment of the Alpha-tubulin promoter that drives expression of GFP and Zeocin resistance. This plasmid was digested with BamHI-HF (NEB) and NdeI (NEB), resulting in two fragments, 838 bp and 4724 bp. The 4724 bp fragment was separated on a 1% agarose gel, excised, and purified using QIAquick Gel extraction kit (Qiagen) (FIG. 11). The two fragments of interest were ligated using the Rapid DNA Ligation kit (Roche). NEB10 β chemically competent cells (NEB) were then transformed with some of the ligation reaction. Resulting colonies were used as templates for screening by PCR with: 2x GoTaq Green Master Mix (Promega), primers 121 Tub seq F and pYB32/3C R1, and 5% final DMSO. The following cycling conditions were applied: 95°C for 5 minutes, [95°C for 30 seconds, 59°C for 30 seconds, 72°C for 1 minute] for 35 cycles, and 72°C for 5 minutes (described previously) (FIG. 12). Plasmids were extracted from colonies confirmed as positive transformants by PCR, sequenced, and the resulting vector designated pYB61 (SEQ ID NO:33).

TABLE 11

Primer Name	SEQ ID NO	Primer Sequence, 5'→3'
121 Tub seq F	13	GGATCTCATGCTGGAGTTCTTC
pYB32/3C R1	14	GTA CTTCTCGTGGTAGGCAACC

EXAMPLE 6

[00245] Design and build of non-targeted gRNA vector (pYB66)

[00246] Previously, gRNA plasmids were targeted to the Fatty Acid Synthase (FAS) gene locus, thereby creating an auxotrophic requirement for palmitic acid. This

auxotrophy masked the phenotype that resulted from CRISPR-mediated carotene synthase inactivation. To better discern CRISPR-mediated phenotypic changes at loci involved in carotenoid biosynthesis, a non-targeted gRNA vector was made. The plasmid pCL122 (SEQ ID NO:37) was digested with BamHI-HF (NEB) and NdeI (NEB) to remove the reading frame encoding GFP, resulting in two fragments, sized 838 bp and 5773 bp. The 5773 bp fragment was separated on a 1% agarose gel, excised, and purified using QIAquick Gel extraction kit (Qiagen) (FIG. 13). A cassette encoding gRNA3 CS1 was amplified by PCR using the plasmid pYB36 (SEQ ID NO: 9) as a template with: 2x KOD Hot Start Master Mix (EMD Millipore), primers pYB66 BamBgl F and pYB66 Nde R, and 5% final DMSO. The following cycling conditions were applied: 95°C for 2 minutes, [95°C for 20 seconds, 59°C for 10 seconds, 70°C for 5 seconds] for 35 cycles, and 72°C for 2 minutes (FIG. 14). The resulting PCR fragment was column purified using QIAquick PCR purification kit (Qiagen) and digested with BamHI-HF (NEB) and NdeI (NEB). Following digestion, the resulting DNA fragment was again purified using QIAquick PCR purification kit (FIG. 15). The digested pCL122 and gRNA fragments were ligated using Rapid DNA Ligation kit (Roche). NEB10 β chemically competent cells were then transformed with a portion of the ligation reaction and resulting bacterial colonies were screened by colony PCR using 2x GoTaq Green MasterMix (Promega), primers pYB66 EF1seq F and pCL122 OrfC R, and 5% final volume of DMSO with following cycling conditions: 95°C for 10 minutes, [95°C for 30 seconds, 60°C for 30 seconds, 72°C for 45 seconds] for 35 cycles, and 72°C for 5 minutes (FIG. 16). Plasmids were extracted from colonies positive by PCR, sequenced, and the resulting vector was designated pYB66 (SEQ ID NO:34).

TABLE 12

Primer Name	SEQ ID NO	Primer Sequence, 5'→3'
pYB66 BamBgl F	38	CAAGGGATCCAGATCTTCCGCACTGATGAGTC
pYB66 Nde R	39	AACTCATATGGTCCCATTTCGCCA
pYB66 EF1seq F	40	GAGAGGATAGTATCTTGCGTGCTTG
pCL122 OrfC R	41	GCAAGGTTGGAACATTACGATCAAG

EXAMPLE 7

[00247] Design and build of Cas9 and gRNA vector (pYB73)

[00248] A vector containing both Cas9 and gRNA cassettes on the same vector with a single dominant selectable marker was created to test the efficiency of such configurations on genome editing. Plasmid pYB61 was digested with PstI (NEB) and purified by QIAquick PCR purification kit (Qiagen) (FIG. 17). The gRNA expression cassette was amplified off of pYB36 using PCR with the following: 2x KOD Hot Start Master Mix (Novagen), primers pYB73 gRNA Pst Kpn IF F and pYB73 gRNA Xho Pst IF R, and 5% final DMSO. The cycling conditions were used as follows: 95°C for 2 minutes, [95°C for 20 seconds, 58°C for 10 seconds, 70°C for 38 seconds] for 35 cycles, and 70°C for 2 minutes (FIG. 18). A resulting PCR fragment was column purified using QIAquick PCR clean-up kit (Qiagen). Purified, digested pYB61 fragment was ligated to the purified, digested gRNA PCR fragment using InFusion kit (Clontech) following manufacturer's protocol. NEB10 β chemically competent cells were transformed with a portion of the InFusion reaction, and resulting bacterial colonies were screened by PCR with the following: 2x GoTaq Green Master Mix (Promega), primers pYB73 seq F and pYB73 seq R, and 5% final DMSO. The following cycling conditions were applied: 95°C for 10 minutes, [95°C for 30 seconds, 60°C for 30 seconds, 72°C for 2 minutes] for 35 cycles, and 72°C for 5 minutes (FIG. 19). Plasmids were extracted from colonies positive by PCR, sequenced, and the resulting vector was designated pYB73 (SEQ ID NO:35). Plasmid pYB73 is not designed for targeting to any specific gene.

TABLE 13

Primer Name	SEQ ID NO	Primer Sequence, 5'→3'
pYB73 gRNA Pst Kpn IF F	42	CATACATGGTTCGACCTGCAGGGTACCTCTTA TCTGCCTCGC
pYB73 gRNA Xho Pst IF R	43	ATTAATGCAGGTTCCCTGCAGCTCGAGAAGAA TCTGAACTCACGTC
pYB73 seq F	44	CACCCCAACTTGTTTATTGCAG

Primer Name	SEQ ID NO	Primer Sequence, 5'→3'
pYB73 seq R	45	GAGCGAGGAAGCGGAAGAG

EXAMPLE 8

[00249] Design, build and transformation with CarG strain

[00250] To improve the distinction between transformants with CRIPSR-mediated carotene synthase gene inactivation (white colony phenotype) and without CRIPSR-mediated carotene synthase gene inactivation (yellow-orange colony phenotype), the CarG gene (geranylgeranyl pyrophosphate synthase) from *Mucor circinelloides* was codon-optimized for expression in *Schizochytrium*. The CarG gene, which should increase total amount of carotenoids produced and make transformants more orange, was synthesized by DNA2.0 and cloned into a *Schizochytrium* expression vector containing a blasticidin selection cassette. The resulting vector was named pCL310 (SEQ ID NO:36).

[00251] Wild-type strain of *Schizochytrium* sp., ATCC 20888, was used for transformation with pCL310 via a particle bombardment method (Bio-Rad). ATCC 20888 was grown in 25 mLs of M50-20 medium in a 250 mL smooth-bottom Erlenmeyer flask at +27°C, shaking at 200 rpm overnight. Following that, the 20888 culture was diluted 1/100 into 50 mLs of M2B medium in a 250 mL baffled flask and grown under the previously used conditions overnight. When the 20888 culture was in early log phase (0.6-2 OD units/mL), it was harvested by centrifugation at 3,220 ×g for 10 minutes. Supernatant was decanted and the pellet was resuspended in M2B to a final concentration of 20 OD units/mL. One hundred μL of resulting cell suspension was spread on approximately a third of a non-selective M2B agar plate (approximately 2 inches in diameter). Five μg of plasmid DNA and 5 μg of pCL310 was mixed with 50 μL of 2.5M CaCl₂, 20 μL of 0.1M spermidine and 50 μL of prepared M10 Tungsten beads (following manufacturer's protocol), vortexed for 1 minute, then incubated for 10 minutes at room temperature to allow for the beads to settle. DNA-coated beads were washed once with 250 μL of 100% ethanol, and then beads were resuspended in 60 μL of 100% ethanol. Each prepared macrocarrier (following manufacturer's protocol for macrocarrier assembly preparation) had 10 μL of coated beads in ethanol

spotted to the center under conditions leading to minimal vibrations from the hood and so the ethanol was allowed to dry. Rupture disc holders were affixed with 1,100 psi rupture discs after both were briefly sterilized in 70% isopropanol. Assembled macrocarrier platforms were placed into the top shelf position and M2B agar plates with 20888 cell patch were placed cell-side-up on the third shelf from the top. When vacuum reached approximately 27 inches of mercury inside the chamber of the bombardment machine, helium was fired until rupture disc ruptured. Then, the flow of helium was closed off and the chamber was vented to atmosphere. Bombarded plates were then removed from the chamber. The bombardment process was repeated for all the samples and controls. The 20888 strain bombarded with pCL310 was designated T212. Bombarded plates were incubated in a +27°C incubator for 4 hours, after which cells were washed off the plate with ~1 mL of M2B and divided equally between four M2B agar plates containing 100 µg/mL blasticidin (ThermoFisher). Cells were spread with 3 mm sterile glass beads. After bead removal, plates were wrapped and incubated at +27°C for 5 to 8 days. When colonies reached 2-4 millimeters in size, they were patched on M2B agar plates containing 100 µg/mL blasticidin. Colonies confirmed to be resistant to blasticidin were picked and inoculated into 50 mLs of M50-20 in a 250 mL smooth bottom Erlenmeyer flask and placed in a shaker at +27°C and 200 rpm for 48 hrs. After 48 hr incubation time, two milliliters of culture were collected by centrifugation at 4,000 ×g for 10 minutes. Supernatant was decanted and pellet was used for genomic DNA isolation following a modified phenol-chloroform extraction. Genomic DNA was extracted and used as PCR template with GoTaq Green Master Mix (Promega) to verify presence of Cas9 cassette using primers pYB13 pYB1 seq F and pCL122 OrfC R and DMSO to 5% final concentration. The following cycling conditions were applied: 95°C for 2 minutes, [95°C for 30 seconds, 63°C for 30 seconds, 72°C for 1 minute 45 seconds] for 35 cycles, and 72°C for 5 minutes (FIG. 20). Several transformants that were positive for the presence of pCL130 transforming DNA as determined by PCR and which appeared darker orange compared to the wild type control, were analyzed for total carotenoids by UV-Vis method. Out of this analysis two strains were identified to have higher total carotenoids, T212-3-1 and T212-3-2. T212-3-2 was selected for subsequent work with genome editing elements.

TABLE 14

Primer Name	SEQ ID NO	Primer Sequence, 5'→3'
pYB13 pYB1 seq F	46	GAGAGGATAGTATCTTGCGTGCTTGG
pCL122 OrfC R	41	GCAAGGTTGGAACATTACGATCAAG

EXAMPLE 9

[00252] Transformation of T212 and 20888 with pYB61, pYB66, and pYB73

[00253] To test the effects of non-targeted vectors on editing efficiency in *Schizochytrium* several transformations were performed using the particle bombardment method described above. The transformations were set up as follows: T280—T212-3-2 strain transformed with pYB61, and T285—ATCC 20888 wild-type strain transformed with pYB61. Both transformations were conducted to create a recipient strain pre-expressing Cas9 for gRNA testing work. Subsequent transformations were set up as follows: T281—T212-3-2 strain co-transformed with pYB61 and pYB66, and T286—ATCC 20888 wild-type strain co-transformed with pYB61 and pYB66. T281 and T286 transformations were performed to assess editing efficiencies with plasmids not designed for targeting to any particular locus in the genome. T282—T212-3-2 strain transformed with pYB73, and T287—ATCC 20888 wild-type strain transformed with pYB73, were completed to assess editing efficiencies with a single plasmid bearing both elements for genome editing. T280, T282, T285, T287 transformations were selected on M2B plates supplemented with zeocin (ThermoFisher) at 50 µg/mL. T281 and T286 transformations were selected on M2B plates supplemented with zeocin at 50 µg/mL and paromomycin (Sigma) at 100 µg/mL. Colonies resulting from all of these transformations were patched on selective M2B plates containing either 100 µg/mL of zeocin alone or 100 µg/mL zeocin and 100 µg/mL paromomycin together. Patches that grew robustly on the patch plates and that displayed white colony phenotypes when applicable were chosen for further analysis. Those strains were taken off the patch plate and inoculated into 25 mLs of M50-20 medium in 250 mL Erlenmeyer smooth bottom flasks. Flasks were grown at +27°C, 200 rpm for approximately 24 hrs. A 2 mL aliquot of each inoculation was taken and

placed in a microcentrifuge tube, spun down at 7,500 $\times g$ for 10 minutes. Supernatant was decanted and the pellet was used for preparing genomic DNA using a modified phenol-chloroform extraction method. Resulting gDNA was subjected to analysis by PCR for the presence of the selection cassette and gene of interest. For T280, T281, T285, T286, junctions between the terminator of the selection cassette and the beginning of the Cas9 gene were amplified using the following: 2X GoTaq Green Master Mix, primers 121 Tub seq F and pYB32/3 CR1, DMSO to 5% final concentration, and gDNA. The following conditions were applied: 95°C for 2 minutes, [95°C for 30 seconds, 58°C for 30 seconds, 72°C for 1 minute 2 seconds] for 35 cycles, and 72°C for 5 minutes (FIG. 21). For T281 and T286. The presence of the gRNA cassette was assessed by PCR amplification of the entire gRNA cassette as follows: 2x GoTaq Green Master Mix, primers pYB66 EF1 seq F and pCL122 OrfC R, DMSO to 5% final concentration, and gDNA. The following cycling conditions were applied: 95°C for 2 minutes, [95°C for 30 seconds, 60°C for 30 seconds, 72°C for 47 seconds] for 35 cycles, and 72°C for 5 minutes (FIG. 22). For T282 and T287 the presence of the gRNA cassette was assessed by PCR as follows: 2x GoTaq Green Master Mix, primers pYB66 EF1 seq F and TT pYB73 HDV R, DMSO at 5% final concentration, and gDNA. The following cycling conditions were applied: 95°C for 2 minutes, [95°C for 30 seconds, 60.9°C for 30 seconds, 72°C for 16 seconds] for 35 cycles, and 72°C for 5 minutes (FIG. 23). T281, T282, T286 and T287 transformants that tested positive by PCR for Cas9 and gRNA were subjected to PCR to amplify portions of the CS gene where gRNA affected a change in order to sequence the PCR amplicon and observe the types of indels that resulted from the gene editing experiments. The PCR to amplify the relevant portion of the CS gene was prepared with the following: 2X KOD Hot Star Master Mix, primers pYB36 CS1 F and pYB36 CS1 R, DMSO to 5% final concentration, and gDNA. The following cycling conditions were applied: 95°C for 2 minutes, [95°C for 20 seconds, 61°C for 10 seconds, 70°C for 11 seconds] for 35 cycles, and 70°C for 5 minutes (FIG. 24). PCR fragments were column purified using QIAquick PCR purification kit (Qiagen) and were sequence verified with pYB36 CS1 F and pYB36 CS1 R.

TABLE 15

Primer Name	SEQ ID NO	Primer Sequence, 5'→3'
pYB66 EF1 seq F	40	GAGAGGATAGTATCTTGCGTGCTTG
pCL122 OrfC R	41	GCAAGGTTGGAACATTACGATCAAG
TT pYB73 HDV R	47	GAAGCATGTTGCCCAGCC
pYB36 CS1 F	27	GAGTCGAAGGAGACGTTGTCTG
pYB36 CS1 R	28	GTCATTGCGAATGATGCGATATG

CLAIMS

What is claimed is:

1. A non-naturally occurring or engineered composition comprising a source of a CRISPR-Cas system comprising a guide-polynucleotide and a Cas protein, wherein the guide-polynucleotide comprises a guide-sequence that essentially is the reverse complement of a target-polynucleotide in a host cell and the guide-polynucleotide can direct binding of the Cas protein at the target-polynucleotide in the host cell to form a CRISPR-Cas complex, wherein the guide-sequence is essentially the reverse complement of the (N)_y part of a 5'-(N)_yPAM-3' polynucleotide sequence target in the genome of the host cell, wherein y is an integer of 8-30, wherein PAM is a protospacer adjacent motif, wherein the host cell is a Labyrinthulomycete, preferably a member of the order Thraustochytriales, preferably a member of the family Thraustochytriaceae, more preferably a member of a genus selected from the group consisting of *Aurantiochytrium*, *Oblongichytrium*, *Schizochytrium*, *Thraustochytrium*, and *Ulkenia*, even more preferably *Schizochytrium* sp. ATCC# 20888, and wherein PAM is preferably a sequence selected from the group consisting of 5'-XGG-3', 5'-XGGXG-3', 5'-XXAGAAW-3', 5'-XXXXGATT-3', 5'-XXAGAA-3', 5'-XAAAAC-3', wherein X can be any nucleotide or analog thereof, preferably X can be any nucleotide; and W is A or T.
2. The composition according to claim 1, wherein the Cas protein is encoded by a polynucleotide and/or the guide-polynucleotide is encoded by or present on a polynucleotide.
3. The composition according to claim 1 or 2, wherein the Cas protein is encoded by a polynucleotide and/or the guide-polynucleotide is encoded by or present on another polynucleotide and the polynucleotide or polynucleotides are comprised in a vector.
4. The composition according to any of the preceding claims, wherein the guide polynucleotide is encoded by a polynucleotide that is transcribed to provide for the actual guide-polynucleotide.

5. The composition according to any one of claims 1 - 3, wherein the Cas protein is encoded by a polynucleotide and the guide-polynucleotide is encoded by or present on another polynucleotide and the polynucleotides are comprised in one vector.
6. The composition according to any one of claims 1 - 3, wherein the Cas protein is encoded by a polynucleotide comprised in a vector and the guide-polynucleotide is encoded by or present on another polynucleotide comprised in another vector, wherein preferably the vector encoding the Cas protein is driven by a low strength promoter and the vector encoding the guide-polynucleotide is driven by a high strength promoter.
7. The composition according to claim 6, wherein one or more or all vectors comprise a selectable marker, preferably each vector comprising a distinct selectable marker.
8. The composition according to any of the preceding claims, wherein one or more exogenous polynucleotides are operably linked to the guide-polynucleotide.
9. The composition according to any of the preceding claims, wherein the Cas protein comprises at least one nuclear localization sequence, preferably a heterologous nuclear localization sequence.
10. The composition according to any of the preceding claims, wherein the Cas protein has activity for directing cleavage of both polynucleotide strands at the location of the target-sequence.
11. The composition according to any of the preceding claims, wherein the Cas protein encoding polynucleotide is codon optimized for the host cell, preferably codon pair optimized.

12. The composition according to any of the preceding claims, wherein the guide-polynucleotide is encoded by a polynucleotide that is operably linked to a RNA polymerase II or III promoter, preferably to a Labyrinthulomycete EF-1 promoter.

13. The composition according to any one of claims 1 - 12, wherein a polynucleotide that is operably linked to an RNA polymerase II promoter encodes a pre-guide-polynucleotide comprising the guide-polynucleotide and self-processing ribozymes, wherein, when transcribed, the guide-polynucleotide is released by the self-processing ribozymes from the pre-guide-polynucleotide transcript.

14. A method of modulating expression of a polynucleotide in a cell, comprising contacting a host cell with the composition according to any one of claims 1 - 13, wherein the guide-polynucleotide directs binding of the Cas protein at the target-polynucleotide in the host cell to form a CRISPR-Cas complex.

15. The method according to claim 14, wherein the host cell comprises a polynucleotide encoding a compound of interest.

16. The method according to claim 14 or 15, wherein the host cell is a recombinant host cell.

17. A host cell comprising the composition according to any one of claims 1 - 13.

18. A method of producing a host cell, comprising contacting a host cell with the composition according to any one of claims 1 - 13, wherein the guide-polynucleotide directs binding of the Cas protein at the target-polynucleotide in the host cell to form a CRISPR-Cas complex.

19. The method according to claim 18, wherein the host cell is first contacted with a source of a Cas protein and subsequently is contacted with a source of a guide-polynucleotide and optionally an exogenous polynucleotide.

20. A The method according to claim 19 or the host cell according to claim 17, wherein the host cell comprises a polynucleotide encoding a compound of interest.

21. The method according to any one of claims 18 - 20 or a host cell according to claim 17, wherein the host cell is a recombinant host cell.

22. A method for the production of a compound of interest, comprising culturing under conditions conducive to the production of the compound of interest a host cell obtainable by the method of any one of claims 18 - 20 or a host cell according to claim 17 or a host cell produced according to the method of any one of claims 18 - 20 and optionally purifying or isolating the compound of interest.

23. The method according to any one of claims 14-16 and 18 - 22, wherein the Cas protein is a Cas9 protein.

24. The method according to any one of claims 18 - 20, wherein expression of the Cas protein is driven by a medium strength promoter.

25. A The method according to claim 24, wherein the medium strength promoter is an alpha tubulin promoter.

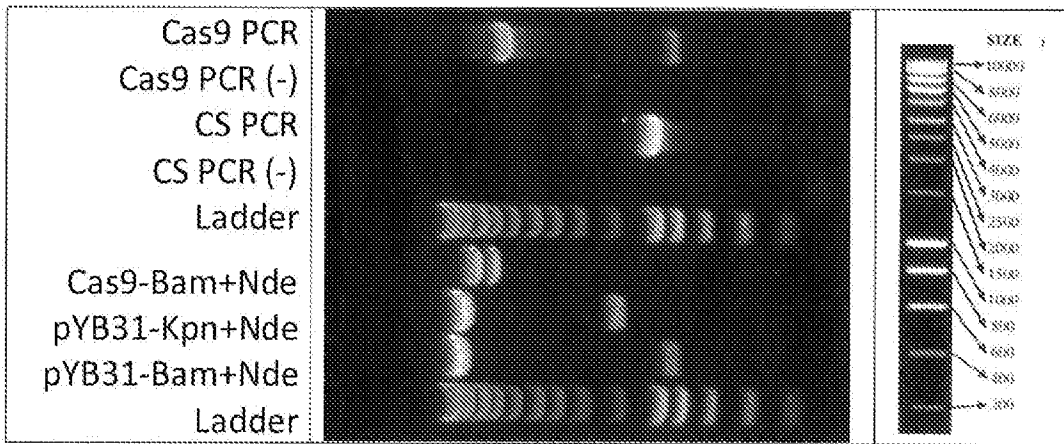


FIG. 3

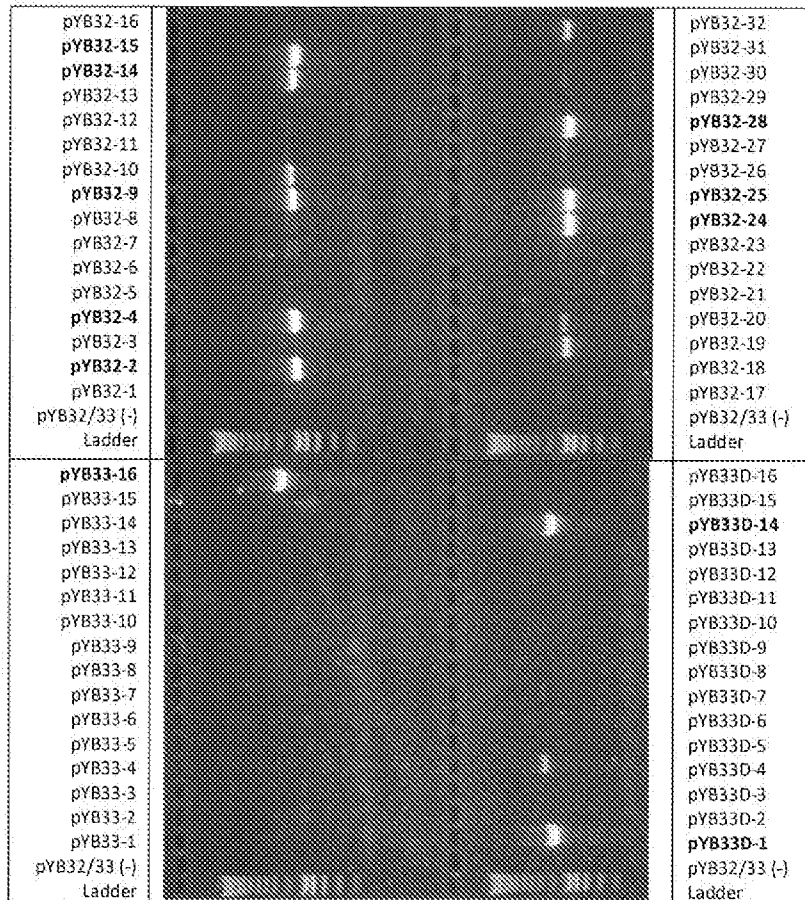


FIG. 4

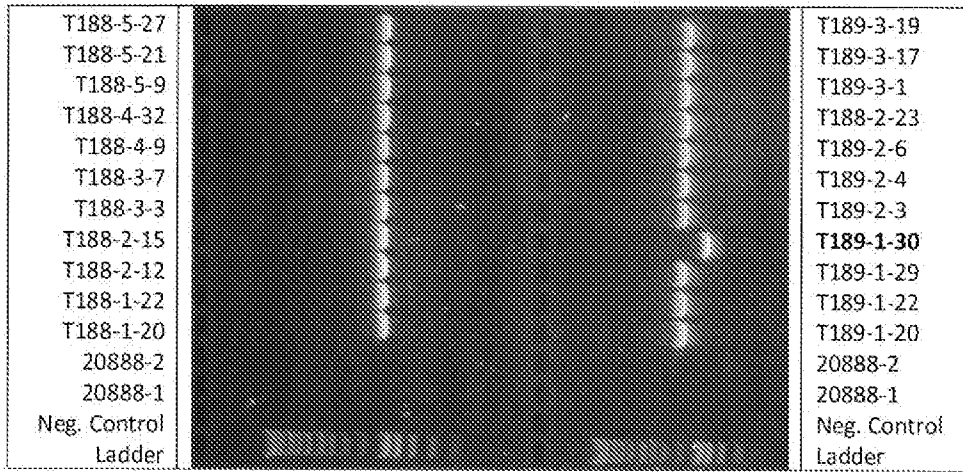


FIG. 5

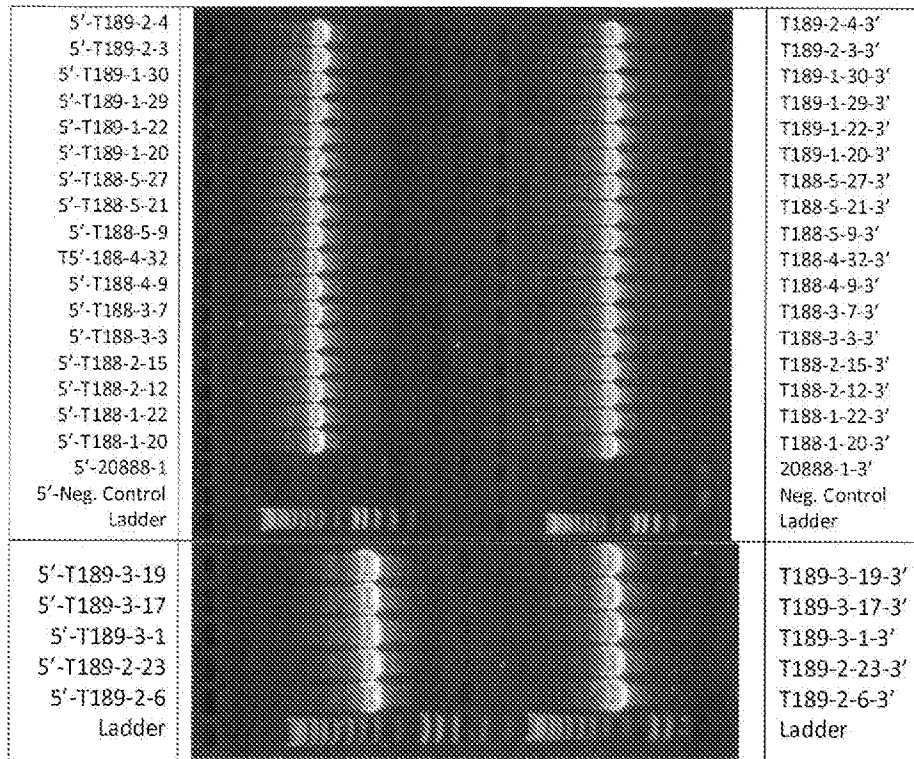


FIG. 6

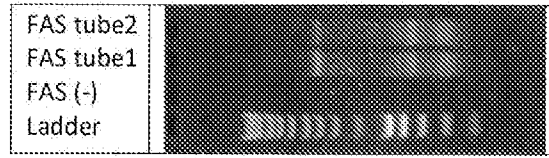


FIG. 7



FIG. 8

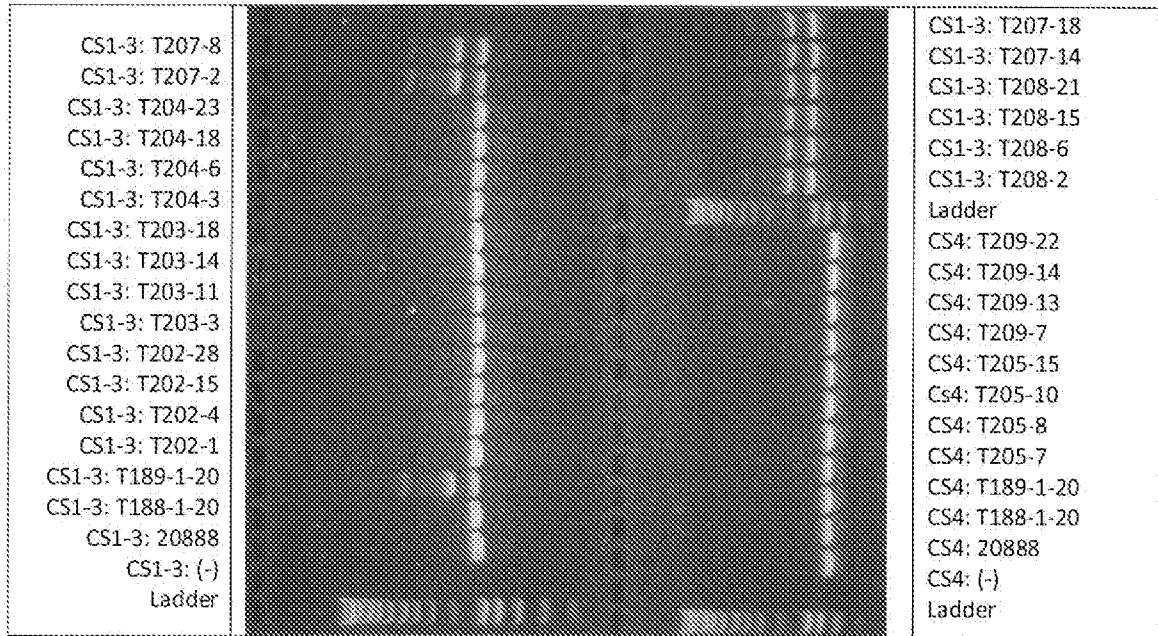


FIG. 9

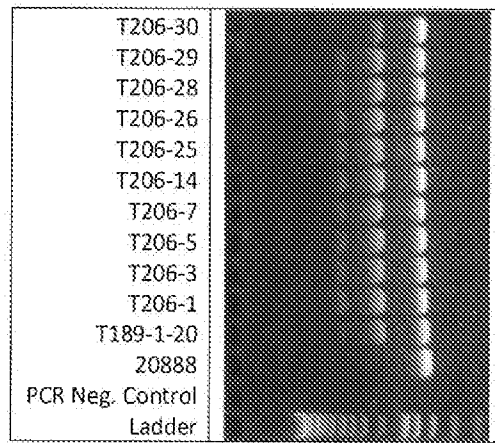


FIG. 10

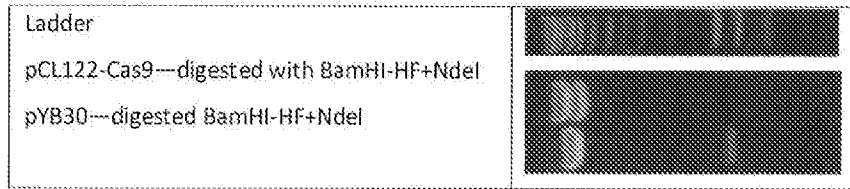


FIG. 11

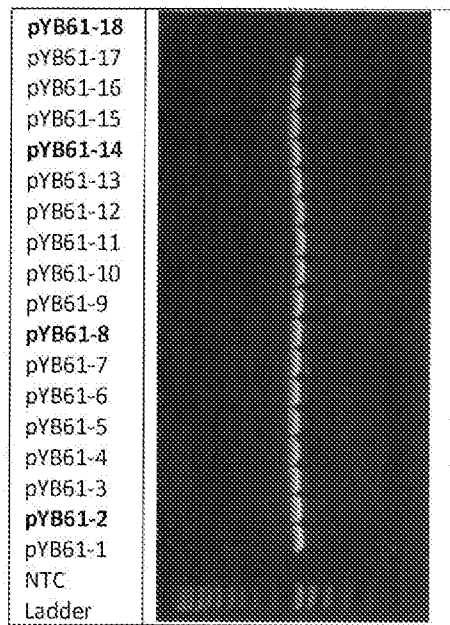


FIG. 12



FIG. 13

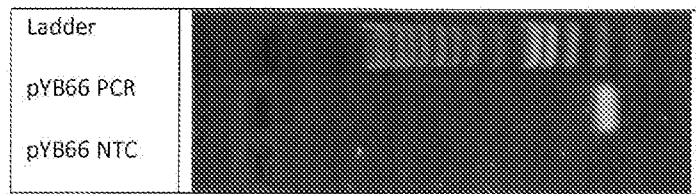


FIG. 14

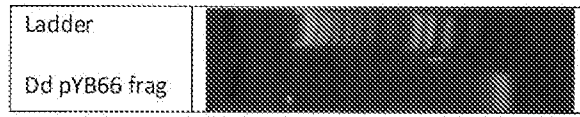


FIG. 15

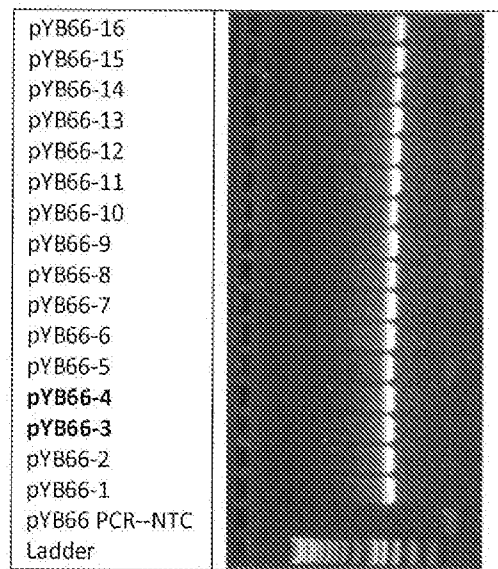


FIG. 16



FIG. 17



FIG. 18

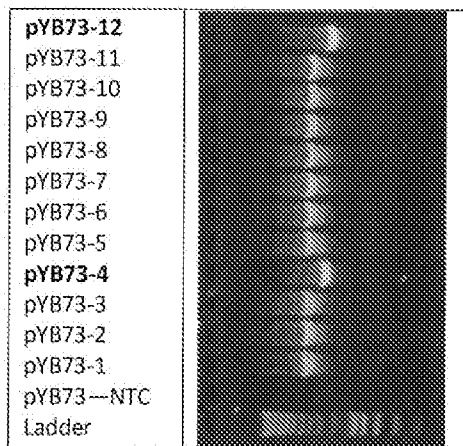


FIG. 19

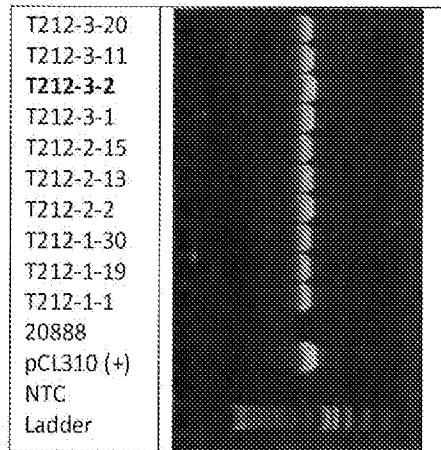


FIG. 20

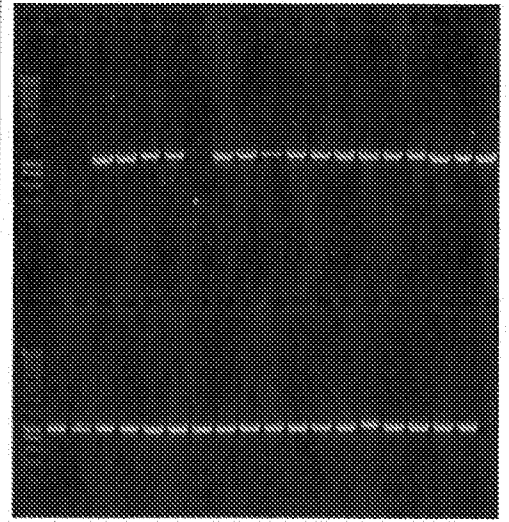
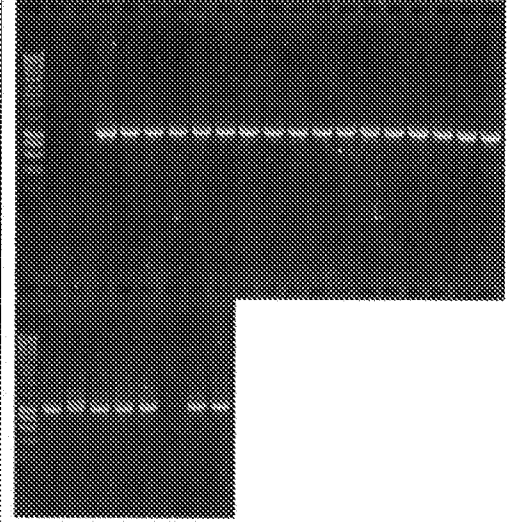
<p>Ladder T281-2-3 T281-2-4 T281-2-6 T281-2-10 T281-2-11 T281-2-25 T284-1-3 T284-1-10 T284-1-11 T284-1-18 T284-1-21 T284-1-30 T284-2-3 T284-2-6 T284-2-11 T284-2-15 T284-2-27 T284-2-31</p>		<p>Ladder NTC 20888 pY861(+) T280-1-3 T280-1-4 T280-1-6 T280-1-20 T280-1-29 T280-2-11 T280-2-20 T280-2-23 T280-2-28 T280-2-31 T281-1-1 T281-1-4 T281-1-9 T281-1-17 T281-1-18 T281-1-26</p>	<p>Ladder NTC 20888 pY861(+) T285-1-2 T285-1-6 T285-1-21 T285-1-23 T285-1-29 T285-1-30 T285-2-2 T285-2-11 T285-2-12 T285-2-21 T285-2-29 T285-2-30 T286-1 T286-2 T286-3 T286-9</p>
<p>Ladder T286-12 T286-16 T286-17 T286-20 T286-28 T286-30 T286-31 T286-32</p>		<p>Ladder NTC 20888 pY861(+) T285-1-2 T285-1-6 T285-1-21 T285-1-23 T285-1-29 T285-1-30 T285-2-2 T285-2-11 T285-2-12 T285-2-21 T285-2-29 T285-2-30 T286-1 T286-2 T286-3 T286-9</p>	<p>Ladder NTC 20888 pY861(+) T285-1-2 T285-1-6 T285-1-21 T285-1-23 T285-1-29 T285-1-30 T285-2-2 T285-2-11 T285-2-12 T285-2-21 T285-2-29 T285-2-30 T286-1 T286-2 T286-3 T286-9</p>

FIG. 21

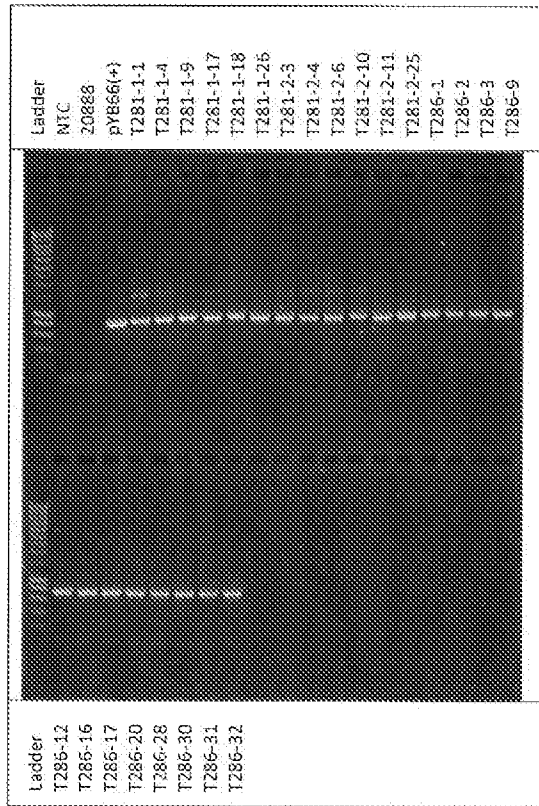


FIG. 22

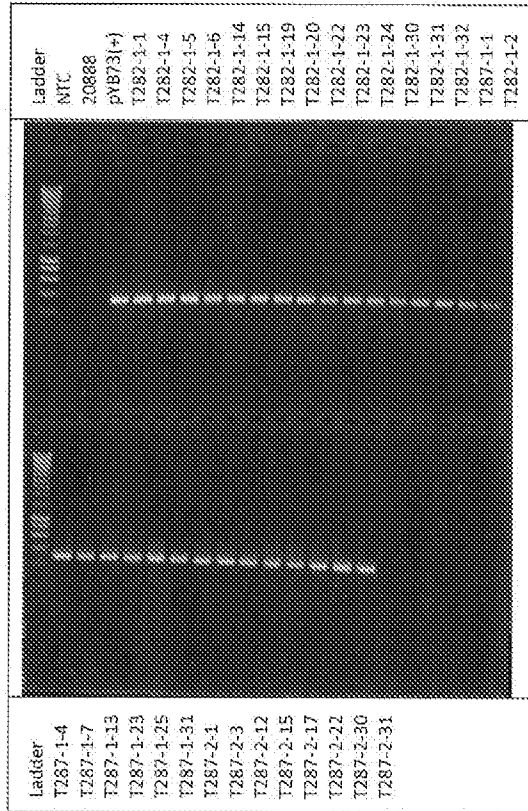


FIG. 23

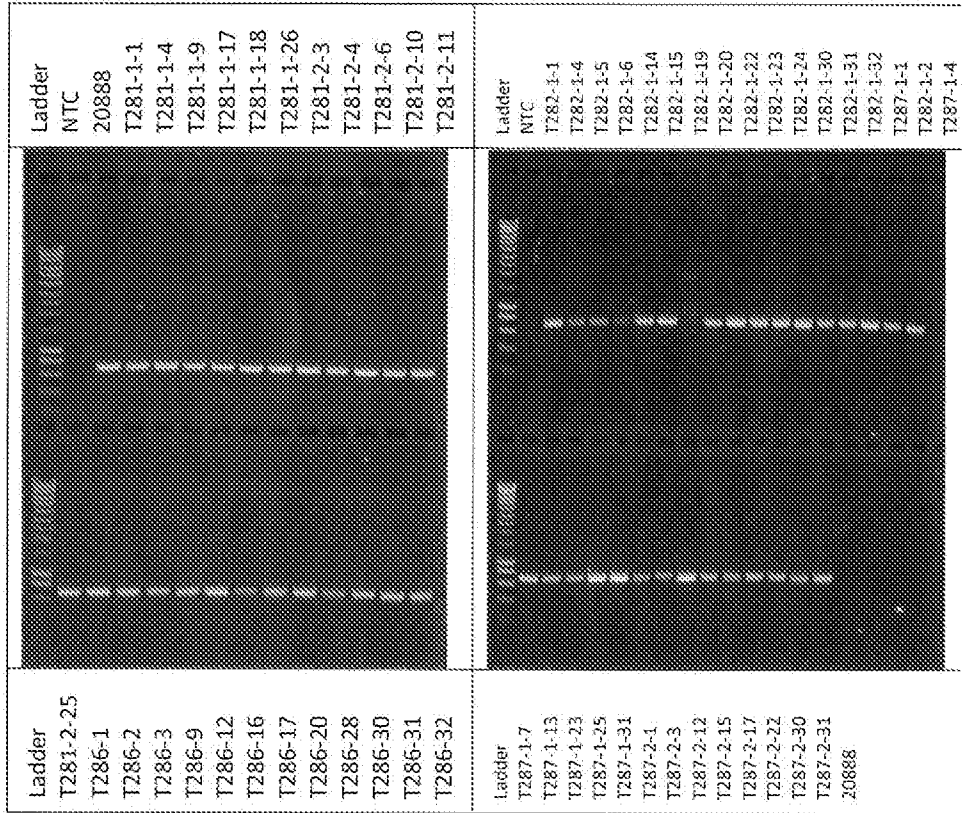


FIG. 24

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2017/041949

A. CLASSIFICATION OF SUBJECT MATTER
 IPC(8) - C12N 1/10; C12N 15/11; C12N 15/79 (2017.01)
 CPC - C12N 1/10; C12N 15/11; C12N 15/79; C12N 2310/11 (2017.08)

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)
 See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
 USPC - 435/471; 435/477; 514/44A; 514/44R (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
 See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2016/109840 A2 (SYNTHETIC GENOMICS, INC.) 07 July 2016 (07.07.2016) entire document	1-3
A	WO 2015/086795 A1 (CELLECTIS) 18 June 2015 (18.06.2015) entire document	1-3
A	WO 2015/004241 A2 (NOVARTIS AG et al) 15 January 2015 (15.01.2015) entire document	1-3
A	US 2014/0186842 A1 (SYNTHETIC GENOMICS, INC.) 03 July 2014 (03.07.2014) entire document	1-3
P, A	WO 2016/197136 A2 (NMC, INC. et al) 08 December 2016 (08.12.2016) entire document	1-3

Further documents are listed in the continuation of Box C. 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	"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
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"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
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search
 13 September 2017

Date of mailing of the international search report

03 OCT 2017

Name and mailing address of the ISA/US
 Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
 P.O. Box 1450, Alexandria, VA 22313-1450
 Facsimile No. 571-273-8300

Authorized officer
 Blaine R. Copenheaver
 PCT Helpdesk: 571-272-4300
 PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2017/041949

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claims Nos.: 4-25
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

- Remark on Protest**
- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- No protest accompanied the payment of additional search fees.