



US 20190136249A1

(19) **United States**(12) **Patent Application Publication**
SAKAI et al.(10) **Pub. No.: US 2019/0136249 A1**(43) **Pub. Date: May 9, 2019**(54) **ORGANELLE GENOME MODIFICATION
USING POLYNUCLEOTIDE GUIDED
ENDONUCLEASE****Publication Classification**(51) **Int. Cl.****C12N 15/82** (2006.01)**C12N 15/113** (2006.01)(52) **U.S. Cl.****CPC** **C12N 15/8213** (2013.01); **C12N 2310/20**
(2017.05); **C12N 15/113** (2013.01)(71) Applicant: **Napigen, Inc.**, San Francisco, CA (US)(72) Inventors: **Hajime SAKAI**, Newark, DE (US);
Byung-Chun YOO, Newark, DE (US);
Emil Meyer OROZCO, JR.,
Cochranville, PA (US); **Roger WYSE**,
Hillsborough, CA (US); **Ganesh**
KISHORE, St. Louis, MO (US); **Jay**
KEASLING, Berkeley, CA (US);
Narendra Singh YADAV, Wilmington,
DE (US)

(57)

ABSTRACT

Provided herein are methods and systems for altering the genome of an organelle. In some embodiments, the method comprises introducing into an organelle a recombinant DNA construct comprising a first polynucleotide encoding at least one guide RNA and a second polynucleotide encoding a polynucleotide guided polypeptide; and growing a cell comprising the organelle under conditions in which the first polynucleotide and the second polynucleotide are each expressed.

Specification includes a Sequence Listing.(21) Appl. No.: **16/109,523**(22) Filed: **Aug. 22, 2018****Related U.S. Application Data**

(60) Provisional application No. 62/548,723, filed on Aug. 22, 2017.

ctatttcaggcacattcaggaccTAGTGTAGATTTAGCAATTTTGCATTACATTTAACATCAAT
TTCATCATTATTAGGTGCTATTAATTTTCATTGTAACAACATTAAATATGAGAACAAATGGTATG
ACAATGCATAAATTACCATTATTTGTATGATCAATTTTCATTACAGCGTTCTTATTATTATTAT
CATTACCTGTATTATCTGCTGGTATTACAATGTTATTATTAGATAGAACTTCAATACCTTCATT
TTTCGGAGTTTCTGGTGGAGGTGGTGGAATGACACATTTAGAAAGAAGTAGACAAATGTCAAAA
GGTGAAGAATTATTCACTGGAGTAGTACCTATCTTAGTAGAATTAGATGGTGATGTAAATGGTC
ATAAATTCTCAGTATCAGGTGAAGGTGAAGGTGATGCTACATATGGTAAATTAACATTAAAATT
CATCTGTACAACAGGTAAATTACCTGTACCTTGACCTACATTAGTAACAACATTCCGATATGGA
GTACAATGTTTCGCAAGATATCCTGATCATATGAAACAACATGATTTCTTCAAATCAGCAATGC
CTGAAGGTTACGTACAAGAAAGAACAATCTTCTTCAAAGATGATGGTAATTATAAAACAAGAGC
TGAAGTAAATTCGAAGGTGATACATTAGTAAATAGAATCGAGTTAAAGGTATCGATTTCAAA
GAAGATGGTAATATCTTAGGTCATAAATTAGAATATAATTATAATTCACATAATGTATATATCA
TGGCTGATAAACAACAAAAATGGTATCAAAGTAAATTTCAAATCAGACATAATATCGAAGACGG
TTCAGTACAATTAGCAGATCATTATCAACAAAAATACACCTATCGGTGATGGTCCTGTATTATTA
CCTGATAATCATTACTTAAGTACACAATCAGCTTTATCAAAGATCCTAATGAAAAAGAGATC
ATATGGTATTATTAGAATTTGTAACAGCTGCTGGTATCACACATGGTATGGATGAATTATATAA
ATAACAACAGGAATTAAAATTTCTCATGATTAATAAATCCCTTTAGCAAGGATAAAAAATAAAA
ATAAAAAATAAAAGTTGATCAGAAATTATCAAAAAATAAATAATAATAATAATAAAAAACATA
TTAATAATAATAATAATAATTATAATAATAATATATAAAGGTAATTTATATGATATTTATCCA
AGATCAAATAGAAATTATATTCAACCAAATAATATTAATAAAGAATTAGTAGTATATGGTTATA
ATTTAgaatcttggttggtatacctct

ctattcaggcacattcaggaccTAGTGTAGATTTAGCAATTTTGCATTACATTTAACATCAAT
TTCATCATTATTAGGTGCTATTAATTTTCATTGTAACAACATTAAATATGAGAACAAATGGTATG
ACAATGCATAAATTACCATTATTTGTATGATCAATTTTCATTACAGCGTTCTTATTATTATTAT
CATTACCTGTATTATCTGCTGGTATTACAATGTTATTATTAGATAGAACTTCAATACTTCAATT
TTTCGGAGTTTCTGGTGGAGGTGGTGGAATGACACATTTAGAAAGAAGTAGACAAATGTCAAAA
GGTGAAGAATTATTCACTGGAGTAGTACCTATCTTAGTAGAATTAGATGGTGATGTAAATGGTC
ATAAATTCTCAGTATCAGGTGAAGGTGAAGGTGATGCTACATATGGTAAATTAACATTAAAATT
CATCTGTACAACAGGTAAATTACCTGTACCTTGACCTACATTAGTAACAACATTCCGATATGGA
GTACAATGTTTCGCAAGATATCCTGATCATATGAAACAACATGATTTCTTCAAATCAGCAATGC
CTGAAGGTTACGTACAAGAAAGAACAATCTTCTTCAAAGATGATGGTAATTATAAAACAAGAGC
TGAAGTAAATTCGAAGGTGATACATTAGTAAATAGAATCGAGTTAAAAGGTATCGATTTCAAA
GAAGATGGTAATATCTTAGGTCATAAATTAGAATATAATTATAATTCACATAATGTATATATCA
TGGCTGATAAACAACAAAAAATGGTATCAAAGTAAATTTCAAATCAGACATAATATCGAAGACGG
TTCAGTACAATTAGCAGATCATTATCAACAAAATACACCTATCGGTGATGGTCCTGTATTATTA
CCTGATAATCATTACTTAAGTACACAATCAGCTTTATCAAAAGATCCTAATGAAAAAAGAGATC
ATATGGTATTATTAGAATTTGTAACAGCTGCTGGTATCACACATGGTATGGATGAATTATATAA
ATAACAACAGGAATTAAAATTTTCTCATGATTAATAAAATCCCTTTAGCAAGGATAAAAAATAAA
ATAAAAAATAAAAAGTTGATCAGAAATTATCAAAAAATAAATAATAATATAATAAAAACATA
TTTAAATAATAATAATATAATTATAATAAATATATATAAAGGTAATTTATATGATATTTATCCA
AGATCAAATAGAAATTATATTCAACCAAATAATATTAATAAAGAATTAGTAGTATATGGTTATA
ATTTAgaatcttgtgttggtatacctct

FIG. 1

GCTGGTTGGTTCCACTACCAAAAGCTGCTCCAAACTAGAATGGTTCCAAAACG
TTGAATCAATGTTAAACCACCACTTAGGTGGTCTTCTTGGTTTAGGTAGTTTAGC
TTGGGCTGGTCACCAAATTCACGTTTCTTTACCAGTAAACAAATTATTAGATGCT
GGTGTAGATCCAAAAGAAATTCACCTTCCTCATGATTTATTATTAAATCGTGCTA
TTATGGCTGACTTATACCCAAGTTTGTCTAAAGGTATTGCTCCTTTCTTTACTTT
AAACTGGAGTGAATACAGTGATTTCTTAACATTTAAAGGTGGTTTAAACCCTGTT
ACATTATCAGGTTCTGCTGGTTCAGCAGCTGGTATGGCTAAAGGTGAAGAATTAT
TCACAGGTGTTGTACCTATTTTAGTAGAATTAGACGGTGATGTAAACGGTCACAA
ATTTTCAGTTTCTGGTGAAGGTGAAGGTGACGCAACTTATGGTAAATTAACACTT
AAATTCATTTGTACTACAGGTAAATTACCAGTACCTTGGCCATCATTAGTTACAA
CTTTTACATACGGTGTACAATGTTTCAGTCGTTACCCTGATCACATGAAACAACA
TGACTTTTTCAAATCTGCTATGCCAGAAGGTTATGTTCAAGAACGTACTATTTTT
TTCAAAGATGACGGTAATTATAAAACACGTGCTGAAGTAAAATTTGAAGGTGATA
CTTTAGTTAACCGTATTGAATTAAAAGGTATTGACTTCAAAGAAGATGGTAATAT
TTTAGGTCACAACTTGAATATAACTACAATTCACATAACGTATATATTATGGCA
GACAAACAAAAAATGGTATTAAAGTAACTTTAAAATTCGTCATAATATCGAGG
ATGGTTCTGTACAATTAGCTGACCACTATCAACAAAACACACCAATTGGTGATGG
TCCTGTTTTACTTCCAGACAATCATTATTTAAGTACTCAATCTGCTTTATCAAAA
GATCCTAACGAAAAACGTGACCACATGGTATTACTTGAATTTGTTACAGCAGCTG
GTATTACTCACGGTATGGATGAATTATACAAATAATCCATTACAGGTGAAGGTC
ACGTTGGTTTATATGAAATTTTAACAACCTCTTGGCATGCACAATTAGCTATTAA
CTTAGCTTTATTTGGTTCGTTATCAATTATTGTAGCTCACCACATGTACGCAATG
CCTCCATACCCTTATTTAGCTACTGATTACGGTACACAATTATCATTATTTACAC
ACCACACATGGATTGGTGGTTTCTGTATTGTTGGTGCTGGTGCTCACGCAGCTAT
TTTCATGGTTCGTGACTACGATCCTACTAATAACTACAACAACCTATTAGACCGT
GTAATTCGTCACCGTGATGCTATTATTTCTCAC

FIG. 2

ORGANELLE GENOME MODIFICATION USING POLYNUCLEOTIDE GUIDED ENDONUCLEASE

CROSS-REFERENCE

[0001] This application is related to U.S. Provisional Patent Application No. 62/548,723, filed on Aug. 22, 2017, which is entirely incorporated herein by reference.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on Oct. 30, 2018, is named 51090_701_201 SL.txt and is 204,800 bytes in size.

SUMMARY

[0003] In an aspect, a method for altering the genome of an organelle may comprise: (a) introducing into an organelle comprising the following: (i) a first polynucleotide encoding at least one guide polynucleic acid, wherein the at least one guide polynucleic acid directs a polynucleotide guided polypeptide to cleave at least one target sequence present in an organelle genome; (ii) a second polynucleotide encoding a polynucleotide guided polypeptide, wherein the polynucleotide guided polypeptide, when associated with the guide polynucleic acid, cleaves the at least one target sequence; (iii) optionally, a third polynucleotide encoding at least one homologous organelle DNA sequence, wherein the at least one homologous organelle DNA is of sufficient size for homologous recombination, wherein integration of the at least one homologous organelle DNA sequence into the organelle genome results in removal of the at least one target sequence; (iv) optionally, a fourth polynucleotide encoding at least one selectable marker or at least one screenable marker, or both; wherein the fourth polynucleotide is operably linked to a promoter that is functional in the organelle; and (v) optionally, a fifth polynucleotide encoding an origin of replication that is functional in the organelle; and (b) growing a cell comprising the organelle of (a) under conditions in which the first polynucleotide of (i) and the second polynucleotide of (ii) are each expressed.

[0004] In another aspect, a method for altering the genome of an organelle may comprise: (a) introducing into an organelle a recombinant DNA construct comprising the following: (i) a first polynucleotide encoding at least one guide polynucleic acid, wherein the at least one guide polynucleic acid directs a polynucleotide guided polypeptide to cleave at least one target sequence present in an organelle genome; (ii) a second polynucleotide encoding a polynucleotide guided polypeptide, wherein the polynucleotide guided polypeptide, when associated with the guide polynucleic acid, cleaves the at least one target sequence; (iii) optionally, a third polynucleotide encoding at least one homologous organelle DNA sequence, wherein the at least one homologous organelle DNA is of sufficient size for homologous recombination, wherein integration of the at least one homologous organelle DNA sequence into the organelle genome results in removal of the at least one target sequence; (iv) optionally, a fourth polynucleotide encoding at least one selectable marker or at least one screenable marker, or both; wherein the fourth polynucleotide is operably linked to a promoter that is functional in the organelle;

and (v) optionally, a fifth polynucleotide encoding an origin of replication that is functional in the organelle; and (b) growing a cell comprising the organelle of (a) under conditions in which the first polynucleotide of (i) and the second polynucleotide of (ii) are each expressed

[0005] In some embodiments, the method may further comprise a step (c) of selecting a cell having an organelle that comprises an altered genome. In some embodiments, the method may further comprise a step (d) of selecting a cell that is homoplasmic for the altered genome of the organelle.

[0006] In some embodiments, the method may comprise introducing into an organelle the third polynucleotide of (iii), wherein the third polynucleotide of (iii) may comprise a sixth and a seventh polynucleotide, wherein the sixth and the seventh polynucleotides correspond to two adjacent regions of homology in the organelle genome, wherein the sixth and seventh polynucleotides are separated by a sequence that is heterologous to the organelle DNA. In some embodiments, the sequence that is heterologous to the organelle DNA may comprise at least one selected from the group consisting of: the first polynucleotide, the second polynucleotide, the fourth polynucleotide, an eighth polynucleotide, and any combination thereof, wherein the eighth polynucleotide encodes an RNA that is heterologous to the organelle.

[0007] In another embodiment, the at least one guide polynucleic acid may be present on a polycistronic transcription unit. In some embodiments, the at least one guide polynucleic acid may be processed from a polycistronic RNA after transcription of the polycistronic transcription unit by use of at least one selected from the group consisting of: an RNA cleavage site, a Csy4 cleavage site, a ribozyme cleavage site, a polynucleotide guided polypeptide cleavage site, the presence of a tRNA sequence, and any combination thereof. In some embodiments, the polycistronic RNA may comprise a first tRNA sequence 5' to the at least one guide RNA and a second tRNA sequence 3' to the at least one guide RNA.

[0008] In another embodiment, the method may comprise the eighth polynucleotide, wherein the eighth polynucleotide may encode at least one selected from the group consisting of: a herbicide tolerance protein, a pesticidal protein, an accessory protein that binds to a pesticidal protein, a dsRNA, a siRNA, a miRNA, and any combination thereof, wherein the dsRNA, the siRNA and the miRNA suppress at least one target gene present in a plant pest. In some embodiments, the method may comprise the eighth polynucleotide, wherein the eighth polynucleotide may be operably linked to at least one regulatory element that is active in an organelle. In some embodiments, the at least one regulatory element may be a promoter.

[0009] In another embodiment, at least one selected from the group consisting of: the first polynucleotide, the second polynucleotide, the fourth polynucleotide, the fifth polynucleotide, and any combination thereof, may be located outside the region bounded by the sixth and the seventh polynucleotide.

[0010] In another embodiment, the method may comprise the fourth and fifth polynucleotides, wherein both the fourth and the fifth polynucleotides may be located outside the region bounded by the sixth and the seventh polynucleotides.

[0011] In another embodiment, the method may comprise the fourth polynucleotide, wherein the fourth polynucleotide may comprise a first sequence encoding a positive selectable marker and a second sequence encoding a negative selectable marker, wherein the first and the second sequence may be each operably linked to a promoter that is functional in the organelle.

[0012] In another embodiment, the method may comprise the fifth polynucleotide, wherein the fifth polynucleotide may encode an origin of replication that is functional in a plastid (e.g., a chloroplast), wherein the origin of replication functional in a plastid may correspond to DNA sequence from a plastid rRNA intergenic region.

[0013] In another embodiment, the method may comprise the fifth polynucleotide, wherein the fifth polynucleotide may encode an origin of replication that is functional in a mitochondrion.

[0014] In some embodiments, the polynucleotide-guided polypeptide may be selected from the group consisting of: a Cas9 protein, a MAD2 protein, a MAD7 protein, a CRISPR nuclease, a nuclease domain of a Cas protein, a Cpf1 protein, an Argonaute, modified versions thereof, and any combination thereof.

[0015] In some embodiments, the recombinant DNA construct may further comprise a ninth and tenth polynucleotide that have at least 100 nucleotides of 100 percent sequence identity to each other, wherein the ninth and tenth polynucleotides are arranged as direct repeats in the recombinant DNA construct.

[0016] In some embodiments, the recombinant DNA construct may be linear and further wherein the ninth and tenth polynucleotides may be present at the 5' and 3' ends of the recombinant DNA construct.

[0017] In another embodiment, the method may comprise a recombinant DNA construct that comprises at least one selected from the group consisting of: the first polynucleotide, the second polynucleotide, the third polynucleotide, the fourth polynucleotide, the fifth polynucleotide, and any combination thereof. In some embodiments, the method may comprise more than one such recombinant DNA construct.

[0018] In another embodiment, the recombinant DNA construct may further comprise a ninth and tenth polynucleotide, wherein the ninth and tenth polynucleotide may have 100 percent sequence identity to each other, and further wherein the ninth and tenth polynucleotides may be arranged as direct repeats in the recombinant DNA construct. In some embodiments, the ninth and tenth polynucleotides may have at least 20, 21, 22, 23, 24, 25, 30, 40, 50, 60, 70, 80, 90 or 100 nucleotides of 100 percent sequence identity to each other. Optionally, the recombinant DNA construct may be linear and the ninth and tenth polynucleotides are present at the 5' and 3' ends of the recombinant DNA construct.

[0019] In another embodiment, any of the methods herein may further involve introducing into the organelle a polynucleotide encoding at least one selectable marker selected from the group consisting of: a positive selectable marker, a negative selectable marker, and any combination thereof. In some embodiments, the positive selectable marker may be an herbicide tolerance protein. In some embodiments, the herbicide tolerance protein may be at least one selected from the group consisting of: a 4-hydroxyphenylpyruvate dioxygenase (HPPD), a sulfonylurea-tolerant acetolactate synthase (ALS), an imidazolinone-tolerant acetolactate syn-

thase (ALS), a glyphosate-tolerant 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), a glyphosate-tolerant glyphosate oxidoreductase (GOX), a glyphosate N-acetyltransferase (GAT), a phosphinothricin acetyl transferase (PAT), a protoporphyrinogen oxidase (PROTOX), an auxin enzyme or receptor, a P450 polypeptide, an acetyl coenzyme A carboxylase (ACCase), and any combination thereof.

[0020] In some embodiments, the method may further involve growing the cell in the presence of a positive selection agent and selecting a cell that is homoplasmic for the altered genome of the organelle. In some embodiments, the method may further involve growing the cell in the absence of the positive selection agent, followed by selecting a cell that lacks a non-integrated recombinant DNA construct. In some embodiments, the method may further involve growing the cell in the absence of the positive selection agent, followed by growing the cell in the presence of a negative selection agent, followed by selecting a cell that lacks a non-integrated recombinant DNA construct. In some embodiments, the cell may be selected from the group consisting of: a yeast cell, an algal cell, a plant cell, an insect cell, a non-human animal cell, an isolated and purified human cell, and a mammalian tissue culture cell. In some embodiments, in the method for a plant cell, the organelle may be a plastid (e.g., a chloroplast) or a mitochondrion. In some embodiments, the method may further involve regenerating or growing a plant from the plant cell comprising an altered organelle genome. In some embodiments, the plant cell may be monocot cell, e.g., a maize cell. The plant cell may be a dicot cell, e.g., a soybean cell.

[0021] In some embodiments, the cell may be a plant cell, wherein the organelle is a plastid or a mitochondrion, and wherein the method further comprises regenerating a plant from the plant cell comprising an altered organelle genome. In some embodiments, the cell may be a yeast cell or an algal cell. In some embodiments, a plant, seed, root, stem, leaf, flower, fruit, or bean produced by the method disclosed herein may comprise an organelle with an altered genome.

[0022] In another embodiment, the alteration of the genome of the organelle may comprise an insertion of an expression cassette. In some embodiments, the expression cassette may be a polycistronic expression cassette. In some embodiments, the polycistronic expression cassette may encode a selectable marker or a screenable marker, or both.

[0023] In another aspect, a recombinant DNA construct may comprise the following: (i) a first polynucleotide encoding at least one guide polynucleic acid, wherein the at least one guide polynucleic acid directs a polynucleotide guided polypeptide to cleave at least one target sequence present in an organelle genome; (ii) a second polynucleotide encoding a polynucleotide guided polypeptide, wherein the polynucleotide guided polypeptide, when associated with the guide polynucleic acid, cleaves the at least one target sequence; (iii) optionally, a third polynucleotide encoding at least one homologous organelle DNA sequence, wherein the at least one homologous organelle DNA is of sufficient size for homologous recombination, wherein integration of the at least one homologous organelle DNA sequence into the organelle genome results in removal of the at least one target sequence; (iv) optionally, a fourth polynucleotide encoding at least one selectable marker or at least one screenable marker, or both; wherein the fourth polynucleotide is operably linked to a promoter that is functional in the organelle; and (v) optionally, a fifth polynucleotide encoding an origin

of replication that is functional in the organelle. In some embodiments, the third polynucleotide of (iii) may comprise a sixth and a seventh polynucleotide, wherein the sixth and the seventh polynucleotides correspond to two adjacent regions of homology in the organelle genome, wherein the sixth and seventh polynucleotides are separated by a sequence that is heterologous to the organelle DNA. In some embodiments, a yeast cell, algal cell, plant cell, plant, seed, root, stem, leaf, flower, fruit, or bean may comprise the recombinant DNA construct.

[0024] In another aspect, a recombinant DNA construct may comprise the following: (i) a first polynucleotide encoding at least one guide RNA, wherein the at least one guide RNA directs a polynucleotide guided polypeptide to cleave at least one target sequence present in an organelle genome; (ii) a second polynucleotide encoding a polynucleotide guided polypeptide, wherein the polynucleotide guided polypeptide, when associated with the guide RNA, cleaves the at least one target sequence; (iii) a third polynucleotide comprising a sixth and a seventh polynucleotide, wherein the sixth and the seventh polynucleotides correspond to two adjacent regions of homology in the organelle genome, wherein the sixth and seventh polynucleotides are separated by a sequence that is heterologous to the organelle DNA, wherein the sequence that is heterologous to the organelle DNA comprises at least one selected from the group consisting of: the first polynucleotide, the second polynucleotide, the fourth polynucleotide, an eighth polynucleotide, and any combination thereof, wherein the eighth polynucleotide encodes an RNA that is heterologous to the organelle; (iv) optionally, a fourth polynucleotide encoding at least one selectable marker or at least one screenable marker, or both; wherein the fourth polynucleotide is operably linked to a promoter that is functional in the organelle; and (v) optionally, a fifth polynucleotide encoding an origin of replication that is functional in the organelle.

[0025] In another aspect, a method for altering the genome of an organelle may comprise: (a) introducing into a cell: (i) a polynucleotide encoding an RNA sequence comprising an organelle targeting RNA operably linked to a guide polynucleic acid, wherein the guide polynucleic acid directs a polynucleotide guided polypeptide to cleave a target sequence present in an organelle genome, wherein the polynucleotide is operably linked to at least one regulatory element; and (ii) a second polynucleotide encoding a modified polynucleotide guided polypeptide, wherein the second polynucleotide is operably linked to at least one regulatory element, and wherein the modified polynucleotide guided polypeptide comprises a polynucleotide guided polypeptide operably linked to an organelle targeting peptide; wherein the organelle targeting RNA of (i) and the organelle targeting peptide of (ii) each target the same organelle; and (b) growing the cell under conditions in which the polynucleotide of (i) and the second polynucleotide of (ii) are both expressed. In some embodiments, the method may further comprise a step (c) of selecting a cell having an organelle that comprises an altered genome. In some embodiments, the method may further comprise a step (d) of selecting a cell that is homoplasmic for the altered genome of the organelle.

[0026] In another aspect, a method for altering the genome of an organelle may comprise: (a) introducing into a cell: (i) a polynucleotide encoding an RNA sequence comprising an organelle targeting RNA operably linked to a guide poly-

nucleic acid, wherein the guide polynucleic acid directs a polynucleotide guided polypeptide to cleave a target sequence present in an organelle genome, wherein the polynucleotide is operably linked to at least one regulatory element; and (ii) a third polynucleotide, wherein the third polynucleotide is operably linked to at least one regulatory element, wherein the third polynucleotide encodes an RNA molecule comprising an organelle targeting RNA operably linked to an RNA sequence encoding a polynucleotide guided polypeptide; wherein the organelle targeting RNA of (i) and the organelle targeting RNA of (ii) each target the same organelle; and (b) growing the cell under conditions in which the polynucleotide of (i) and the third polynucleotide of (ii) are both expressed. In some embodiments, the method may further comprise a step (c) of selecting a cell having an organelle that comprises an altered genome. In some embodiments, the method may further comprise a step (d) of selecting a cell that is homoplasmic for the altered genome of the organelle.

[0027] In another embodiment, any of the methods herein may further comprise introducing a polynucleotide comprising at least one donor polynucleotide (e.g. donor DNA) into the organelle, wherein the at least one donor polynucleotide (e.g. donor DNA) is bounded by at least one homologous sequence with respect to the organelle genome, wherein integration of all or part of the at least one donor polynucleotide into the organelle genome results in removal of the target site of the guide polynucleic acid. In some embodiments, the at least one donor polynucleotide (e.g. donor DNA) may comprise a first nucleic acid sequence heterologous to the organelle genome, wherein the first nucleic acid sequence is bounded by a second and a third nucleic acid sequence, wherein the second and the third nucleic acid sequences correspond to two adjacent regions of homology in the organelle genome. In some embodiments, the second or the third nucleic acid sequence, or both, may comprise at least one altered sequence, wherein the at least one altered sequence is altered with respect to at least one additional target site in the organelle genome, wherein the at least one altered sequence is not recognized by at least one additional guide polynucleic acid, wherein the at least one additional guide polynucleic acid may direct a polynucleotide guided polypeptide to cleave the at least one additional target site in the organelle genome. In some embodiments, the at least one additional target site in the organelle genome may be present in at least one essential coding region. In some embodiments, the polynucleotide introduced into the organelle may further comprise a fourth nucleic acid sequence, wherein the fourth nucleic acid sequence encodes the at least one additional guide polynucleic acid. In some embodiments, the at least one additional guide polynucleic acid may be operably linked to a promoter that is active in the organelle.

[0028] In some embodiments, the polynucleotide introduced into the organelle further may comprise a fourth nucleic acid sequence, wherein the fourth nucleic acid sequence encodes the at least one additional guide RNA operably linked to a promoter that is active in the organelle. In some embodiments, a cell produced by the method disclosed herein may be selected from the group consisting of: a yeast cell, an algal cell, a plant cell, an insect cell, a non-human animal cell, an isolated and purified human cell, and a mammalian tissue culture cell. In some embodiments,

a plant, seed, root, stem, leaf, flower, fruit, or bean produced by the method disclosed herein may comprise an organelle with an altered genome.

[0029] In another aspect, a method for altering a genome of an organelle may comprise: (a) introducing into an organelle of a cell the following: (i) at least one guide RNA, wherein the at least one guide RNA directs a polynucleotide guided polypeptide to cleave at least one target sequence present in the genome of the organelle; (ii) a polynucleotide guided polypeptide, wherein the polynucleotide guided polypeptide, when associated with the at least one guide RNA, cleaves the at least one target sequence; and (iii) a replacement DNA; and (b) selecting a cell comprising an organelle comprising the replacement DNA. In some embodiments, the replacement DNA of step (a) part (iii) may comprise fragments of organellar DNA or a complete organellar DNA from a cultivar, line, sub-species and other species and is distinct from the genome of the organelle of step (a). In some embodiments, the replacement DNA may be lacking the at least one target sequence. In some embodiments, after step (a) part (ii) and prior to step (a) part (iii), a cell may be selected in which the genome of the organelle has been eliminated. In some embodiments, the at least one target sequence may not be present in the replacement DNA.

[0030] In some embodiments, the guide polynucleic acid in the methods and compositions of matter described herein may comprise the following: i) at least 17 nucleotides that are complementary to at least 17 nucleotides of a target polynucleic acid, wherein said target polynucleic acid is located in the genome of an organelle; and ii) a region that contacts a polynucleotide-guided polypeptide. The guide polynucleic acid may comprise one or more RNA bases. In some embodiments, the guide polynucleic acid may be a guide RNA. The guide polynucleic acid may be a dual guide RNA. In some embodiments, the guide polynucleic acid may be a single guide RNA.

[0031] In another embodiment, the polynucleotide-guided polypeptide in the methods and compositions of matter described herein may be selected from the group consisting of: a Cas9 protein, a MAD2 protein, a MAD7 protein, a CRISPR nuclease, a nuclease domain of a Cas protein, a Cpf1 protein, an Argonaute, modified versions thereof, and any combination thereof. In some embodiments, the sequence encoding the polynucleotide-guided polypeptide may be codon-optimized for a human, a yeast, an alga, or a plant species.

[0032] In another embodiment, the cell may be a plant cell, the organelle may be a plastid (e.g., a chloroplast) or a mitochondrion, and the method may further comprise regenerating or growing a plant from the plant cell comprising an altered organelle genome.

[0033] In another embodiment, a cell produced by any of the methods described herein may be selected from the group consisting of: a yeast cell, an algal cell, a plant cell, an insect cell, a non-human animal cell, an isolated and purified human cell, and a mammalian tissue culture cell.

[0034] In another embodiment, a plant, seed, root, stem, leaf, flower, fruit, or bean produced by any of the methods described herein may comprise an organelle with an altered genome.

[0035] In another embodiment, a cell comprising any of the recombinant DNA constructs described herein may be selected from the group consisting of: a yeast cell, an algal

cell, a plant cell, an insect cell, a non-human animal cell, an isolated and purified human cell, and a mammalian tissue culture cell.

[0036] In another embodiment, a plant, seed, root, stem, leaf, flower, fruit, or bean comprising any of the recombinant DNA constructs described herein may comprise an organelle with an altered genome.

[0037] In one embodiment, a polynucleotide may comprise a) an organelle targeting sequence; and b) a guide polynucleic acid, wherein the guide polynucleic acid comprises i) at least 17 nucleotides that are complementary to at least 17 nucleotides of a target polynucleic acid, wherein said target polynucleic acid is located in the genome of an organelle; and ii) a region that contacts a polynucleotide-guided polypeptide, wherein said organelle targeting sequence and said guide polynucleic acid sequence are operably linked. In another embodiment, the polynucleotide comprises one or more RNA bases. In another embodiment, the polynucleotide further comprises a sequence encoding the polynucleotide-guided polypeptide. In another embodiment, said polynucleotide-guided polypeptide is a Cas9 protein. In another embodiment, said polynucleotide-guided polypeptide is an Argonaute protein. In another embodiment, said polynucleotide-guided polypeptide is a nuclease in a CRISPR family. In another embodiment, said polynucleotide-guided polypeptide is Cpf1. In another embodiment, the sequence encoding said polynucleotide-guided polypeptide is codon-optimized for a human. In another embodiment, the sequence encoding said polynucleotide-guided polypeptide is codon-optimized for a plant species. In another embodiment, said target polynucleic acid comprises a protospacer adjacent motif (PAM) sequence. In another embodiment, said Cas9 has been engineered to associate with an altered PAM sequence. In another embodiment, said polynucleotide-guided polypeptide selectively cleaves the target polynucleic acid. In another embodiment, said polynucleotide-guided polypeptide selectively induces a double-strand break in the target polynucleic acid. In another embodiment, said polynucleotide-guided polypeptide comprises a nuclease domain that induces a nick in the target polynucleic acid. In another embodiment, the polynucleotide comprises two or more different guide polynucleic acids. In another embodiment, the guide polynucleic acid is comprised of a dual-guide RNA. In another embodiment, the guide polynucleic acid is a single guide RNA. In another embodiment, the guide polynucleic acid is comprised of a crRNA and a trRNA, wherein said crRNA and said trRNA are optionally linked. In another embodiment, said guide polynucleic acid comprises a region that is engineered to be complementary to at least 18 nucleotides of the target polynucleic acid in the organelle of a cell. In another embodiment, said guide polynucleic acid is engineered to be substantially complementary to at least 22 nucleic acids of the target polynucleic acid in the organelle of a cell. In another embodiment, said at least 17 nucleotides are contiguous. In another embodiment, said organelle is a mitochondrion. In another embodiment, said organelle is a plastid. In another embodiment, said guide polynucleic acid is engineered to hybridize to a region of a target gene disclosed herein. In another embodiment, the polynucleotide further comprises a modified RNA donor sequence, wherein the modified RNA donor sequence comprises an organelle targeting RNA operably linked to a donor RNA.

[0038] In another embodiment, a DNA sequence that when translated to RNA may result in a polynucleotide of the disclosure.

[0039] In another embodiment, a polynucleotide encoding an RNA sequence may comprise an organelle targeting RNA operably linked to a guide RNA, wherein the guide RNA directs a polynucleotide guided polypeptide to cleave a target sequence present in an organelle genome. The RNA sequence may further comprise a sequence encoding a polynucleotide guided polypeptide, and optionally, an RNA cleavage site between the guide RNA and the sequence encoding a polynucleotide guided polypeptide.

[0040] In another embodiment, an organelle may comprise the polynucleotide of the disclosure. In some embodiments, the organelle is a mitochondrion. In some embodiments, the organelle is a plastid.

[0041] In another embodiment, a cell may comprise any of the polynucleotides of the disclosure. The cell may further comprise a polynucleotide encoding a modified polynucleotide guided polypeptide, wherein the modified polynucleotide guided polypeptide comprises a polynucleotide guided polypeptide operably linked to an organelle targeting peptide.

[0042] In another embodiment, a method for introducing a guide polynucleic acid into an organelle of a cell may comprise: (a) introducing into a cell a polynucleotide encoding an RNA sequence comprising an organelle targeting RNA operably linked to a guide polynucleic acid, wherein the guide polynucleic acid directs a polynucleotide guided polypeptide to cleave a target sequence present in an organelle genome, further wherein the polynucleotide is operably linked to at least one regulatory element; and (b) growing the cell under conditions in which the polynucleotide is expressed.

[0043] In another embodiment, a method for altering the genome of an organelle may comprise: (a) introducing into a cell: (i) a polynucleotide encoding an RNA sequence comprising an organelle targeting RNA operably linked to a guide polynucleic acid, wherein the guide polynucleic acid directs a polynucleotide guided polypeptide to cleave a target sequence present in an organelle genome, wherein the polynucleotide is operably linked to at least one regulatory element; and (ii) a second polynucleotide encoding a modified polynucleotide guided polypeptide, wherein the second polynucleotide is operably linked to at least one regulatory element, and wherein the modified polynucleotide guided polypeptide comprises a polynucleotide guided polypeptide operably linked to an organelle targeting peptide; wherein the organelle targeting RNA of (i) and the organelle targeting peptide of (ii) each target the same organelle; and (b) growing the cell under conditions in which the polynucleotide of (i) and the second polynucleotide of (ii) are both expressed.

[0044] In another embodiment, a method for altering the genome of an organelle may comprise: (a) introducing into a cell: (i) a polynucleotide encoding an RNA sequence comprising an organelle targeting RNA operably linked to a guide polynucleic acid, wherein the guide polynucleic acid directs a polynucleotide guided polypeptide to cleave a target sequence present in an organelle genome, wherein the polynucleotide is operably linked to at least one regulatory element; and (ii) a third polynucleotide, wherein the third polynucleotide is operably linked to at least one regulatory element, wherein the third polynucleotide encodes an RNA

molecule comprising an organelle targeting RNA operably linked to an RNA sequence encoding a polynucleotide guided polypeptide; wherein the organelle targeting RNA of (i) and the organelle targeting RNA of (ii) each target the same organelle; and (b) growing the cell under conditions in which the polynucleotide of (i) and the third polynucleotide of (ii) are both expressed.

[0045] In another embodiment, a method for altering the genome of an organelle may comprise: (a) introducing into a cell a polynucleotide encoding an RNA sequence comprising: (i) an organelle targeting RNA operably linked to a guide polynucleic acid, wherein the guide polynucleic acid directs a polynucleotide guided polypeptide to cleave a target sequence present in an organelle genome, (ii) a sequence encoding a polynucleotide guided polypeptide, and (iii) an RNA cleavage site between the guide polynucleic acid and the sequence encoding a polynucleotide guided polypeptide, wherein the polynucleotide is operably linked to at least one regulatory element; and (b) growing the cell under conditions in which the polynucleotide of (a) is expressed.

[0046] In another embodiment, any of the methods herein may further comprise introducing a polynucleotide comprising at least one donor polynucleotide (e.g. donor DNA) into the organelle, wherein the at least one donor polynucleotide (e.g. donor DNA) is bounded by at least one homologous sequence with respect to the organelle genome, wherein integration of all or part of the at least one donor polynucleotide into the organelle genome results in removal of the target site of the guide polynucleic acid. The at least one donor polynucleotide (e.g. donor DNA) may comprise a first nucleic acid sequence heterologous to the organelle genome, wherein the first nucleic acid sequence is bounded by a second and a third nucleic acid sequence, wherein the second and the third nucleic acid sequences correspond to two adjacent regions of homology in the organelle genome. Additionally, the second or the third nucleic acid sequence, or both, may comprise at least one altered sequence, wherein the at least one altered sequence is altered with respect to at least one additional target site in the organelle genome, wherein the at least one altered sequence is not recognized by at least one additional guide polynucleic acid, wherein the at least one additional guide polynucleic acid directs a polynucleotide guided polypeptide to cleave the at least one additional target site in the organelle genome. The at least one additional target site in the organelle genome may be present in at least one essential coding region. The polynucleotide introduced into the organelle may further comprise a fourth nucleic acid sequence, wherein the fourth nucleic acid sequence encodes the at least one additional guide polynucleic acid operably linked to a promoter that is active in the organelle.

[0047] In another embodiment, a polynucleotide may encode a modified RNA donor sequence, wherein the modified RNA donor sequence may comprise an organelle targeting RNA operably linked to a donor RNA. The modified RNA donor sequence may comprise a reverse transcriptase primer site. Additionally, a cell comprising the polynucleotide, and further comprising a polynucleotide encoding a modified reverse transcriptase, wherein the modified reverse transcriptase comprises a reverse transcriptase operably linked to an organelle targeting peptide.

[0048] In another embodiment, a method of altering the genome of an organelle may further comprise introducing a

donor polynucleotide into the organelle, wherein the donor polynucleotide is introduced into the organelle by: (a) introducing the polynucleotide encoding a modified RNA donor sequence into the cell, wherein the polynucleotide is operably linked to at least one regulatory element; (b) introducing into the cell a polynucleotide encoding a modified reverse transcriptase, wherein the modified reverse transcriptase comprises a reverse transcriptase operably linked to an organelle targeting peptide, wherein the polynucleotide is operably linked to at least one regulatory element, wherein the organelle targeting RNA of (a) and the organelle targeting peptide of (b) each target the same organelle; and (c) growing the cell under conditions wherein the polynucleotides of (a) and (b) are both expressed.

[0049] In another embodiment, a method for altering the genome of an organelle may comprise: (a) introducing into an organelle a recombinant DNA construct comprising the following: (i) a first polynucleotide encoding at least one guide polynucleic acid, wherein the at least one guide polynucleic acid directs a polynucleotide guided polypeptide to cleave at least one target sequence present in an organelle genome; (ii) a second polynucleotide encoding a polynucleotide guided polypeptide, wherein the polynucleotide guided polypeptide, when associated with the guide polynucleic acid, cleaves the at least one target sequence; (iii) a third polynucleotide encoding at least one homologous organelle DNA sequence, wherein the at least one homologous organelle DNA is of sufficient size for homologous recombination, wherein integration of the at least one homologous organelle DNA sequence into the organelle genome results in removal of the at least one target sequence; (iv) optionally, a fourth polynucleotide encoding at least one selectable marker; wherein the fourth polynucleotide is operably linked to a promoter that is functional in the organelle; and (v) optionally, a fifth polynucleotide encoding an origin of replication that is functional in the organelle; and (b) growing a cell comprising the organelle of (a) under conditions in which the first polynucleotide of (i) and the second polynucleotide of (ii) are each expressed. The third polynucleotide of (iii) may comprise a sixth and a seventh polynucleotide, wherein the sixth and the seventh polynucleotides correspond to two adjacent regions of homology in the organelle genome, wherein the sixth and seventh polynucleotides are separated by a sequence that is heterologous to the organelle DNA, wherein the sequence that is heterologous to the organelle DNA comprises at least one selected from the group consisting of: the first polynucleotide, the second polynucleotide, the fourth polynucleotide and an eighth polynucleotide, wherein the eighth polynucleotide encodes an RNA that is heterologous to the organelle.

[0050] In another embodiment, a method wherein at least one selected from the group consisting of: the first polynucleotide, the second polynucleotide, the fourth polynucleotide and the fifth polynucleotide, may be located outside the region bounded by the sixth and the seventh polynucleotide.

[0051] In another embodiment, a method wherein both the fourth and the fifth polynucleotides may be located outside the region bounded by the sixth and the seventh polynucleotides.

[0052] In another embodiment, the fourth polynucleotide comprises a first sequence encoding a positive selectable marker and a second sequence encoding a negative select-

able marker, wherein the first and the second sequence are each operably linked to a promoter that is functional in the organelle.

[0053] In another embodiment, the fifth polynucleotide encodes a plastid origin of replication, wherein the plastid origin of replication corresponds to DNA sequence from a plastid rRNA intergenic region.

[0054] In another embodiment, the fifth polynucleotide encodes a mitochondrial origin of replication.

[0055] In another embodiment, the recombinant DNA construct further comprises an eighth and ninth polynucleotide, wherein the eighth and ninth polynucleotides have at least 100 nucleotides of 100 percent sequence identity to each other, wherein the eighth and ninth polynucleotides are arranged as direct repeats in the recombinant DNA construct. Optionally, the recombinant DNA construct is linear and the eighth and ninth polynucleotides are present at the 5' and 3' ends of the recombinant DNA construct.

[0056] In another embodiment, the recombinant DNA construct is linear and single-stranded, and the recombinant DNA construct is operably linked to a modified VirD2 protein, wherein the modified VirD2 protein comprises a VirD2 protein operably linked to an organelle targeting peptide, wherein the modified VirD2 protein has also been modified such that each native nuclear localization sequence of the VirD2 protein is no longer functional. Optionally, the recombinant DNA construct is operably linked to at least one modified VirE2 protein, wherein the at least one modified VirE2 protein comprises a VirE2 protein operably linked to an organelle targeting peptide, wherein the at least one modified VirE2 protein has also been modified such that each native nuclear localization sequence of the VirE2 protein is no longer functional. Optionally, the recombinant DNA construct is operably linked to at least one modified RecA protein, wherein the at least one modified RecA protein comprises a RecA protein operably linked to an organelle targeting peptide. Optionally, the recombinant DNA construct is operably linked to at least one chimeric polypeptide, wherein the at least one chimeric polypeptide comprises an organelle targeting peptide and a cell penetrating peptide.

[0057] In another embodiment, any of the methods herein may further involve introducing into the organelle a polynucleotide encoding at least one selectable marker selected from the group consisting of: a positive selectable marker, a negative selectable marker, and any combination thereof. The positive selectable marker may be an herbicide tolerance protein. The herbicide tolerance protein may be at least one selected from the group consisting of: a 4-hydroxyphenylpyruvate dioxygenase (HPPD), a sulfonyleurea-tolerant acetolactate synthase (ALS), an imidazolinone-tolerant acetolactate synthase (ALS), a glyphosate-tolerant 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), a glyphosate-tolerant glyphosate oxidoreductase (GOX), a glyphosate N-acetyltransferase (GAT), a phosphinothricin acetyl transferase (PAT), a protoporphyrinogen oxidase (PROTOX), an auxin enzyme or receptor, a P450 polypeptide and an acetyl coenzyme A carboxylase (ACCase). The method may further involve growing the cell in the presence of a positive selection agent and selecting a cell that is homoplasmic for the altered genome of the organelle. Optionally, the method may further involve growing the cell in the absence of the positive selection agent, followed by selecting a cell that lacks a non-integrated recombinant

DNA construct. Alternatively, the method may further involve growing the cell in the absence of the positive selection agent, followed by growing the cell in the presence of a negative selection agent, followed by selecting a cell that lacks a non-integrated recombinant DNA construct. In the method, the cell may be a plant cell, the organelle may be a plastid. The method may further involve regenerating a plant from the plant cell comprising an altered organelle genome. The plant cell may be monocot cell, e.g., a maize cell. The plant cell may be a dicot cell, e.g., a soybean cell.

[0058] In another embodiment, in any of the methods herein for altering the genome of an organelle to contain a heterologous polynucleotide, the heterologous polynucleotide may encode at least one selected from the group consisting of: a herbicide tolerance protein, a pesticidal protein, an accessory protein that binds to a pesticidal protein, a dsRNA, a siRNA and a miRNA, wherein the dsRNA, the siRNA and the miRNA suppress at least one target gene present in a plant pest. The herbicide tolerance protein may be at least one selected from the group consisting of: a 4-hydroxyphenylpyruvate dioxygenase (HPPD), a sulfonyleurea-tolerant acetolactate synthase (ALS), an imidazolinone-tolerant acetolactate synthase (ALS), a glyphosate-tolerant 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), a glyphosate-tolerant glyphosate oxidoreductase (GOX), a glyphosate N-acetyltransferase (GAT), a phosphinothricin acetyl transferase (PAT), a protoporphyrinogen oxidase (PROTOX), an auxin enzyme or receptor, a P450 polypeptide and an acetyl coenzyme A carboxylase (AC-Case). The pesticidal protein may be at least one selected from the group consisting of: Cry1Ac, Cyt1Aa, Cry1Ab, Cry2Aa, CryII, Cry1C, Cry1D, Cry1E, Cry1Be, Cry1Fa and Vip3A. The accessory protein that binds to a pesticidal protein may be at least one selected from the group consisting of: a 20 kDa accessory protein and a 19 kDa accessory protein. The dsRNA, the siRNA and the miRNA can suppress at least one target gene selected from the group consisting of: proteasome A-type subunit peptide (Pas-4), ACT, SHR, EPIC2B and PnPMAl. The heterologous polynucleotide may be operably linked to at least one regulatory element that is active in an organelle. The at least one regulatory element may be selected from the group consisting of: a maize clpP promoter combined with a maize clpP 5'-UTR, a maize clpP promoter combined with a 5'-UTR from gene 10 of bacteriophage T7, a tomato psbA promoter is combined with a 5'-UTR from gene 10 of bacteriophage T7 and a tomato rm16 promoter combined with a modified accD 5'-UTR. The cell may be a plant cell, wherein the organelle is a plastid, and wherein the method further comprises regenerating a plant from the plant cell comprising an altered organelle genome. The plant cell may be a soybean cell.

[0059] In another embodiment, a cell may comprise an organelle with an altered genome, wherein the cell may be produced by any of the above methods. The cell may be selected from the group consisting of: a yeast cell, an algal cell, a plant cell, an insect cell, a non-human animal cell, an isolated and purified human cell, and a mammalian tissue culture cell.

[0060] In another embodiment, a method may comprise altering the genome of an organelle in a cell as described above, wherein the cell is a plant cell and further wherein a plant is regenerated from a plant cell, wherein the plant comprises an organelle with an altered genome. Also, a plant

(e.g., progeny plant) or seed produced from the regenerated plant, wherein the plant or seed comprises an organelle with an altered genome.

[0061] In another embodiment, a plant, seed, root, stem, leaf, flower, fruit, or bean may be produced by a method of the disclosure. In some embodiments, the plant, seed, root, stem, leaf, flower, fruit, or bean comprises an organelle with an altered genome.

[0062] In another embodiment, a plant, seed, root, stem, leaf, flower, fruit, or bean may comprise a polynucleotide of the disclosure.

INCORPORATION BY REFERENCE

[0063] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0064] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings (also "figure" and "FIG." herein), of which:

[0065] FIG. 1 presents the sequence (SEQ ID NO: 144) obtained from PCR amplification of the replaced DNA locus in transformed yeast mitochondrial DNA modified by the Edit Plasmid approach; and

[0066] FIG. 2 presents the sequence (SEQ ID NO: 171) obtained from PCR amplification of the replaced DNA locus in transformed *Chlamydomonas* plastid DNA modified by the Edit Plasmid approach.

BRIEF DESCRIPTION OF THE SEQUENCE LISTING

[0067] The disclosure is more fully understood from the following detailed description and Sequence Listing, which form a part of this application.

[0068] SEQ ID NO: 1 corresponds to the nucleic acid sequence encoding mCas9-A; i.e., a Cas9 comprising ATPase beta mitochondrial targeting peptide.

[0069] SEQ ID NO: 2 corresponds to the nucleic acid sequence encoding mCas9-B; i.e., a Cas9 comprising the 70 kD mitochondrial targeting peptide.

[0070] SEQ ID NO: 3 corresponds to the nucleic acid sequence encoding a guide RNA-tRNA^{Lys} (tRK1) fusion ("N" residues indicate the variable targeting domain of the guide RNA).

[0071] SEQ ID NO: 4 corresponds to the nucleic acid sequence encoding a guide RNA-tRNA^{Lys} fusion (tRK2-2 version for mitochondrial import; "N" residues indicate the variable targeting domain of the guide RNA).

[0072] SEQ ID NO: 5 corresponding to the nucleic acid sequence encoding a guide RNA-tRNA^{Lys} fusion with an altered 5' tRNA end.

[0073] SEQ ID NO: 6 corresponds to the nucleic acid sequence encoding a guide RNA-tRNA^{Lys} fusion (modified tRK2 version with altered 5' end; "N" residues at the 5' end indicate the variable targeting domain of the guide RNA).

[0074] SEQ ID NO: 7 corresponds to the nucleic acid sequence encoding a gRNA embedded in tRK2 intron in the backbone of tRK2-2 (20-mer of “N” residues indicates the variable targeting domain; 3-mer of “N” residues is complementary to the first three nucleotides of the variable targeting domain to preserve the secondary structure for splicing).

[0075] SEQ ID NO: 8 corresponds to the nucleic acid sequence encoding a gRNA embedded in tRK2 type intron in the backbone of tRK1 (20-mer of “N” residues indicates the variable targeting domain; 3-mer of “N” residues is complementary to the first three nucleotides of guide RNA to preserve the secondary structure for splicing).

[0076] SEQ ID NO: 9 corresponds to the nucleic acid sequence encoding a gRNA fused with second half of tRK1 (B form).

[0077] SEQ ID NO: 10 corresponds to the nucleic acid sequence encoding a form of tRK1 to be co-expressed with guide RNA-B form fusion.

[0078] SEQ ID NO: 11 corresponds to the nucleic acid sequence encoding a gRNA constructed between the D arm and F hairpin structures.

[0079] SEQ ID NO: 12 corresponds to the nucleic acid sequence encoding a gRNA fused with the D arm.

[0080] SEQ ID NO: 13 corresponds to the nucleic acid sequence encoding a gRNA fused with F hairpin structure.

[0081] SEQ ID NO: 14 corresponds to the nucleotide sequence for the variable targeting domain of a guide RNA to target the cytochrome b gene in mitochondria.

[0082] SEQ ID NO: 15 corresponds to the nucleotide sequence for the variable targeting domain of a guide RNA to target the COX1 gene in mitochondria.

[0083] SEQ ID NO: 16 corresponds to the nucleotide sequence for the variable targeting domain of a guide RNA to target the COX1 gene in mitochondria.

[0084] SEQ ID NO: 17 corresponds to the nucleotide sequence for the variable targeting domain of a guide RNA to target the COX2 gene in mitochondria.

[0085] SEQ ID NO: 18 corresponds to the nucleic acid sequence that is fused with the 3' end of a variable targeting domain to create a functional guide RNA for Cas9.

[0086] SEQ ID NO: 19 corresponds to the nucleic acid sequence encoding a SNR52 promoter.

[0087] SEQ ID NO: 20 corresponds to the nucleic acid sequence encoding a SUP4 Terminator.

[0088] SEQ ID NO: 21 corresponds to the nucleic acid sequence for an oligonucleotide primer for paromomycin-resistance template DNA

[0089] SEQ ID NO: 22 corresponds to the nucleic acid sequence for a complementary oligonucleotide primer to make template DNA with the primer of SEQ ID NO: 21.

[0090] SEQ ID NO: 23 corresponds to the nucleic acid sequence encoding the variable targeting domain for a guide RNA that targets the 15S rRNA gene in mitochondria.

[0091] SEQ ID NO: 24 corresponds to a nucleic acid sequence encoding a Cas9 gene optimized for expression in yeast mitochondria.

[0092] SEQ ID NO: 25 corresponds to the nucleic acid sequence encoding a COX2 promoter.

[0093] SEQ ID NO: 26 corresponds to the nucleic acid sequence encoding a COX2 terminator.

[0094] SEQ ID NO: 27 corresponds to the nucleotide sequence of the variable targeting domain for a guide RNA to target the mitochondrial 21S rRNA gene in yeast.

[0095] SEQ ID NO: 28 corresponds to the nucleic acid sequence encoding the promoter sequence of the 15S rRNA gene.

[0096] SEQ ID NO: 29 corresponds to the nucleic acid sequence encoding the terminator sequence of the 15S rRNA gene.

[0097] SEQ ID NO: 30 corresponds to the nucleotide sequence for the variable targeting domain of a guide RNA to target the COB gene in mitochondria.

[0098] SEQ ID NO: 31 corresponds to the nucleotide sequence for the variable targeting domain of a guide RNA to target the ATP9 gene in mitochondria.

[0099] SEQ ID NO: 32 corresponds to the amino acid sequence for the NDUFV2 mitochondrial targeting peptide.

[0100] SEQ ID NO: 33 corresponds to the nucleic acid sequence encoding a Cas9 fused with a mitochondrial targeting peptide derived from NDUFV2.

[0101] SEQ ID NO: 34 corresponds to the amino acid sequence of the mitochondrial targeting peptide of citrate synthase.

[0102] SEQ ID NO: 35 corresponds to the nucleic acid sequence encoding a Cas9 fused with the mitochondrial signal peptide derived from human citrate synthase.

[0103] SEQ ID NO: 36 corresponds to the nucleic acid sequence encoding a human 5S rRNA gene for mitochondrial import (the 4-mer “GTCT can be replaced with guide RNA).

[0104] SEQ ID NO: 37 corresponds to the nucleotide sequence of a variable targeting domain for a gRNA sequence targeting the human COX3 gene in mitochondria.

[0105] SEQ ID NO: 38 corresponds to the nucleic acid sequence of an expression cassette for a guide RNA utilizing the promoter and terminator of the human 5S rRNA gene.

[0106] SEQ ID NO: 39 corresponds to the nucleotide sequence of a variable targeting domain for a guide RNA to target the CAPR locus in mouse mitochondrial DNA (CAP^R allele has an A to G substitution at residue 17).

[0107] SEQ ID NO: 40 corresponds to the nucleotide sequence of a polynucleotide modification template with the CAP^R mutation (part of the mouse 16SrRNA).

[0108] SEQ ID NO: 41 corresponds to the nucleotide sequence encoding pcoCas9 without NLS & FLAG domains, but with the potato IV intron. The sequence is codon-optimized for *Arabidopsis* (GenBank ID: KF264451).

[0109] SEQ ID NO: 42 corresponds to the amino acid sequence of pcoCas9.

[0110] SEQ ID NO: 43 corresponds to the amino acid sequence of the transit peptide of AtRbcS (At1g67090). Cleavage occurs after the “N” residue at position 54.

[0111] SEQ ID NO: 44 corresponds to the amino acid sequence of the transit peptide of AtCab (NP_001078288.1). Cleavage occurs after the “P” residue at position 55.

[0112] SEQ ID NO: 45 corresponds to the amino acid sequence of the transit peptide of At DnaJ8 (NP_178207.1). Cleavage occurs after the “V” residue at position 47.

[0113] SEQ ID NO: 46 corresponds to the nucleotide sequence encoding the pcoCas9 with AT-rbcS transit peptide (with potato intron).

[0114] SEQ ID NO: 47 corresponds to the amino acid sequence of pcoCas9 with AT-rbcS chloroplast transit peptide.

[0115] SEQ ID NO: 48 corresponds to the nucleotide sequence encoding the Vd 5'UTR (gi|301016157|gb|HM136583.1|).

[0116] SEQ ID NO: 49 corresponds to the nucleotide sequence encoding the AteIF4E1 full-length cDNA.

[0117] SEQ ID NO: 50 corresponds to the nucleotide sequence encoding a typical gRNA module (5' terminal 20-mer of "N" residues corresponds to the variable targeting domain).

[0118] SEQ ID NO: 51 corresponds to the nucleotide sequence encoding CSY4.

[0119] SEQ ID NO: 52 corresponds to the amino acid sequence of the Csy4 polypeptide.

[0120] SEQ ID NO: 53 corresponds to the nucleotide sequence of the Csy4 recognition site.

[0121] SEQ ID NO: 54 corresponds to the nucleotide sequence encoding a guide RNA flanked by Csy4 recognition sites (multimeric form).

[0122] SEQ ID NO: 55 corresponds to the nucleotide sequence encoding a Nt_ChI_rpoB (*Nicotiana tabacum* RNA polymerase beta chain).

[0123] SEQ ID NO: 56 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting the plastid rpoB gene from *Nicotiana tabacum*.

[0124] SEQ ID NO: 57 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting the plastid rpoB gene from *Nicotiana tabacum*.

[0125] SEQ ID NO: 58 corresponds to the nucleotide sequence encoding a Nt_Cp_psbA (*Nicotiana tabacum* photosystem II protein D1).

[0126] SEQ ID NO: 59 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting the plastid psbA gene from *Nicotiana tabacum*.

[0127] SEQ ID NO: 60 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting the plastid psbA gene from *Nicotiana tabacum*.

[0128] SEQ ID NO: 61 corresponds to the nucleotide sequence encoding a Nt_Cp_rps15 (*Nicotiana tabacum* ribosomal protein S15).

[0129] SEQ ID NO: 62 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting the plastid rps15 gene from *Nicotiana tabacum*.

[0130] SEQ ID NO: 63 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting the plastid rps15 gene from *Nicotiana tabacum*.

[0131] SEQ ID NO: 64 corresponds to the nucleotide sequence encoding a Nt_Cp_rp133 (*Nicotiana tabacum* 50S ribosomal protein L33).

[0132] SEQ ID NO: 65 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting the plastid rp133 gene from *Nicotiana tabacum*.

[0133] SEQ ID NO: 66 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting the plastid rp133 gene from *Nicotiana tabacum*.

[0134] SEQ ID NO: 67 corresponds to the nucleotide sequence encoding a GlmaCp rpoB (*Glycine max* RNA polymerase beta chain).

[0135] SEQ ID NO: 68 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting the plastid rpoB gene from *Glycine max*.

[0136] SEQ ID NO: 69 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting the plastid rpoB gene from *Glycine max*.

[0137] SEQ ID NO: 70 corresponds to the nucleotide sequence encoding a GlmaCp psbA (*Glycine max* photosystem II protein D1).

[0138] SEQ ID NO: 71 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting the plastid psbA gene from *Glycine max*.

[0139] SEQ ID NO: 72 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting the plastid psbA gene from *Glycine max*.

[0140] SEQ ID NO: 73 corresponds to the nucleotide sequence encoding a GlmaCp_rps15 (*Glycine max* ribosomal protein S15).

[0141] SEQ ID NO: 74 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting the plastid rps15 gene from *Glycine max*.

[0142] SEQ ID NO: 75 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting the plastid rps15 gene from *Glycine max*.

[0143] SEQ ID NO: 76 corresponds to the nucleotide sequence encoding a GlmaCp_rp133 (*Glycine max* 50S ribosomal protein L33).

[0144] SEQ ID NO: 77 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting the plastid rp133 gene from *Glycine max*.

[0145] SEQ ID NO: 78 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting the plastid rp133 gene from *Glycine max*.

[0146] SEQ ID NO: 79 corresponds to the nucleotide sequence encoding a *Nicotiana benthamiana* rps16 with intron (ribosomal protein S16, GI: KC495035.1).

[0147] SEQ ID NO: 80 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting the plastid rps16 gene from *Nicotiana benthamiana*.

[0148] SEQ ID NO: 81 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting the plastid rps16 gene from *Nicotiana benthamiana*.

[0149] SEQ ID NO: 82 corresponds to the nucleotide sequence encoding a *Nicotiana benthamiana* matK (maturase K, GI: AB040014).

[0150] SEQ ID NO: 83 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting the plastid matK gene from *Nicotiana benthamiana*.

[0151] SEQ ID NO: 84 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting the plastid matK gene from *Nicotiana benthamiana*.

[0152] SEQ ID NO: 85 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting an intergenic region (NtChrC; 57408 . . . 57389) from *Nicotiana tabacum*.

[0153] SEQ ID NO: 86 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting an intergenic region (NtChrC; 59412 . . . 59393) from *Nicotiana tabacum*.

[0154] SEQ ID NO: 87 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting an intergenic region (NtChrC; 59622 . . . 59603) from *Nicotiana tabacum*.

[0155] SEQ ID NO: 88 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting an intergenic region (NtChrC; 65704 . . . 65723) from *Nicotiana tabacum*.

[0156] SEQ ID NO: 89 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting an intergenic region (GlmaCp_NC_007942.1_59039-59058) from *Glycine max*.

[0157] SEQ ID NO: 90 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting an intergenic region (GlmaCp_NC_007942.1_59100-59119) from *Glycine max*.

[0158] SEQ ID NO: 91 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting an intergenic region (GlmaCp_NC_007942.1_62057-62038) from *Glycine max*.

[0159] SEQ ID NO: 92 corresponds to the nucleotide sequence of variable target region for a guide RNA targeting an intergenic region (GlmaCp_NC_007942.1_62361-62380) from *Glycine max*.

[0160] SEQ ID NO: 93 corresponds to the nucleotide sequence of the target site for the plastid psbA gene.

[0161] SEQ ID NO: 94 corresponds to the nucleotide sequence of the region of the polynucleotide modification template that corresponds to the target site of the plastid psbA gene

[0162] SEQ ID NO: 95 corresponds to the amino acid sequence of the ATPase Beta mitochondrial targeting peptide, which is encoded by SEQ ID NO:1.

[0163] SEQ ID NO: 96 corresponds to the amino acid sequence of the Cas9 polypeptide fused to the ATPase Beta mitochondrial targeting peptide, which is encoded by SEQ ID NO:1.

[0164] SEQ ID NO: 97 corresponds to the amino acid sequence of the 70 kD mitochondrial targeting peptide, which is encoded by SEQ ID NO:2.

[0165] SEQ ID NO: 98 corresponds to the amino acid sequence of the Cas9 polypeptide fused to the 70 kD mitochondrial targeting peptide, which is encoded by SEQ ID NO:2.

[0166] SEQ ID NO: 99 corresponds to the nucleotide sequence of the forward primer ZmPclpP-Forward, for PCR amplification of the maize clpP promoter in combination with the clpP 5'-UTR (ZmPclpP:clpP). This forward primer may also be used for PCR amplification of the maize clpP promoter in combination with the 5'-UTR from gene 10 of bacteriophage T7 (ZmPclpP:G10).

[0167] SEQ ID NO: 100 corresponds to the nucleotide sequence of the reverse primer ZmPclpP-Reverse, for PCR amplification of the maize clpP promoter in combination with the clpP 5'-UTR (ZmPclpP:clpP).

[0168] SEQ ID NO: 101 corresponds to the nucleotide sequence of the reverse primer for PCR amplification of the maize clpP promoter in combination with the 5'-UTR from gene 10 of bacteriophage T7 (ZmPclpP:G10).

[0169] SEQ ID NO: 102 corresponds to the nucleotide sequence of the forward primer for PCR amplification of the tomato psbA promoter in combination with the 5'-UTR from gene 10 of bacteriophage T7 (SIPsbA:T7g10).

[0170] SEQ ID NO: 103 corresponds to the nucleotide sequence of the reverse primer for PCR amplification of the tomato psbA promoter in combination with the 5'-UTR from gene 10 of bacteriophage T7 (SIPsbA:T7g10).

[0171] SEQ ID NO: 104 corresponds to the nucleotide sequence of the forward primer for PCR amplification of the SIPrm16 promoter portion of the tomato rm16 promoter in combination with the accD-mod 5'-UTR.

[0172] SEQ ID NO: 105 corresponds to the nucleotide sequence of the reverse primer for PCR amplification of the SIPrm16 promoter portion of the tomato rm16 promoter in combination with the accD-mod 5'-UTR.

[0173] SEQ ID NO: 106 corresponds to the nucleotide sequence of the forward primer for PCR amplification of the accD-mod 5'-UTR portion of the tomato rm16 promoter in combination with the accD-mod 5'-UTR.

[0174] SEQ ID NO: 107 corresponds to the nucleotide sequence of the reverse primer for PCR amplification of the accD-mod 5'-UTR portion of the tomato rm16 promoter in combination with the accD-mod 5'-UTR.

[0175] SEQ ID NO: 108 corresponds to the nucleotide sequence from *Bacillus thuringiensis* serovar kurstaki HD73 that encodes a Cry1Ac delta-endotoxin (U89872).

[0176] SEQ ID NO: 109 corresponds to the amino acid sequence of the Cry1Ac delta-endotoxin encoded by SEQ ID NO: 108.

[0177] SEQ ID NO: 110 corresponds to the nucleotide sequence from *Bacillus thuringiensis* serovar kurstaki HD73 that encodes a truncated form of a Cry1Ac delta-endotoxin that has insecticidal activity.

[0178] SEQ ID NO: 111 corresponds to the nucleotide sequence from *Bacillus thuringiensis* serovar israelensis that encodes a Cyt1Aa protein (Gene ID: 5759908).

[0179] SEQ ID NO: 112 corresponds to the nucleotide sequence from *Bacillus thuringiensis* serovar israelensis (pBt024) that encodes a 20 kDa accessory protein.

[0180] SEQ ID NO: 113 corresponds to the nucleotide sequence from *Bacillus thuringiensis* serovar israelensis (pBt022) that encodes a 19 kDa accessory protein.

[0181] SEQ ID NO: 114 corresponds to the nucleotide sequence for an open reading frame encoding an *Heterodera glycines* (SCN) specific proteasome A-type subunit peptide referred to herein as Pas-4 (U58067671).

[0182] SEQ ID NO: 115 corresponds to nucleotides 552-699 of SEQ ID NO: 114.

[0183] SEQ ID NO: 116 corresponds to the nucleotide sequence of a first guide RNA target site in the COX1 gene of *Saccharomyces cerevisiae* mitochondrial DNA. The last three nucleotides are the PAM sequence; these three nucleotides are not present in the variable targeting domain of the corresponding guide RNA.

[0184] SEQ ID NO: 117 corresponds to the nucleotide sequence of a second guide RNA target site in the COX1 gene of *Saccharomyces cerevisiae* mitochondrial DNA. The last three nucleotides are the PAM sequence; these three nucleotides are not present in the variable targeting domain of the corresponding guide RNA.

[0185] SEQ ID NO: 118 corresponds to the nucleotide sequence of a third guide RNA target site in the COX1 gene of *Saccharomyces cerevisiae* mitochondrial DNA. The last three nucleotides are the PAM sequence; these three nucleotides are not present in the variable targeting domain of the corresponding guide RNA.

[0186] SEQ ID NO: 119 corresponds to the nucleotide sequence of a fourth guide RNA target site in the COX1 gene of *Saccharomyces cerevisiae* mitochondrial DNA. The last three nucleotides are the PAM sequence; these three nucleotides are not present in the variable targeting domain of the corresponding guide RNA. This target site sequence is present on the reverse complement of the genic sequence.

[0187] SEQ ID NO: 120 corresponds to the nucleotide sequence encoding SpCas9, the Cas9 from *Streptococcus pyogenes*. The coding sequence was optimized for expression in yeast mitochondria.

[0188] SEQ ID NO: 121 corresponds to the nucleotide sequence of the minimal promoter and 5' UTR of the COX2 gene of *Saccharomyces cerevisiae* mitochondrial DNA.

[0189] SEQ ID NO: 122 corresponds to the nucleotide sequence of the minimal terminator of the COX2 gene of *Saccharomyces cerevisiae* mitochondrial DNA.

[0190] SEQ ID NO: 123 corresponds to the nucleotide sequence encoding the tracrRNA, which was used to create guide RNAs targeting the COX2 gene of *Saccharomyces cerevisiae*.

[0191] SEQ ID NO: 124 corresponds to the nucleotide sequence of the minimal promoter of the COX3 gene of *Saccharomyces cerevisiae* mitochondrial DNA.

[0192] SEQ ID NO: 125 corresponds to the nucleotide sequence encoding the tRNA of the tF(GAA) gene from *Saccharomyces cerevisiae* mitochondrial DNA.

[0193] SEQ ID NO: 126 corresponds to the nucleotide sequence encoding the tRNA of the tW(UCA) gene from *Saccharomyces cerevisiae* mitochondrial DNA.

[0194] SEQ ID NO: 127 corresponds to the nucleotide sequence of the minimal terminator of the COX3 gene from *Saccharomyces cerevisiae* mitochondrial DNA.

[0195] SEQ ID NO: 128 corresponds to the nucleotide sequence encoding the tRNA of the tM(CAU) gene from *Saccharomyces cerevisiae* mitochondrial DNA.

[0196] SEQ ID NO: 129 corresponds to the nucleotide sequence encoding GFP. The coding sequence was optimized for expression in yeast mitochondria.

[0197] SEQ ID NO: 130 corresponds to the nucleotide sequence encoding the homologous region from *Saccharomyces cerevisiae*, designated HR1, which is adjacent to the first guide RNA target site (SEQ ID NO: 116) in the COX1 gene.

[0198] SEQ ID NO: 131 corresponds to the nucleotide sequence encoding the homologous region from *Saccharomyces cerevisiae*, designated HR2, which is adjacent to the second guide RNA target site (SEQ ID NO: 117) in the COX1 gene.

[0199] SEQ ID NO: 132 corresponds to the nucleotide sequence encoding the homologous region from *Saccharomyces cerevisiae*, designated HR3, which is adjacent to the third guide RNA target site (SEQ ID NO: 118) in the COX1 gene.

[0200] SEQ ID NO: 133 corresponds to the nucleotide sequence encoding the homologous region from *Saccharomyces cerevisiae*, designated HR4, which is adjacent to the fourth guide RNA target site (SEQ ID NO: 119) in the COX1 gene.

[0201] SEQ ID NO: 134 corresponds to the nucleotide sequence present in the donor DNA that encodes a variant of the first guide RNA target site (SEQ ID NO: 116) in the COX1 gene. Seven nucleotides have been changed in the variant.

[0202] SEQ ID NO: 135 corresponds to the nucleotide sequence present in the donor DNA that encodes a variant of the second guide RNA target site (SEQ ID NO: 117) in the COX1 gene. Sixteen nucleotides at the 5' end have been deleted in the variant.

[0203] SEQ ID NO: 136 corresponds to the nucleotide sequence present in the donor DNA that encodes a variant of

the third guide RNA target site (SEQ ID NO: 118) in the COX1 gene. Five nucleotides at the 3' end have been deleted in the variant.

[0204] SEQ ID NO: 137 corresponds to the nucleotide sequence present in the donor DNA that encodes a variant of the fourth guide RNA target site (SEQ ID NO: 119) in the COX1 gene. Seventeen nucleotides at the 3' end have been deleted in the variant.

[0205] SEQ ID NO: 138 corresponds to the nucleotide sequence of PCR primer C, present in the COX1 gene of *Saccharomyces cerevisiae*.

[0206] SEQ ID NO: 139 corresponds to the nucleotide sequence of PCR primer D, present in the COX1 gene of *Saccharomyces cerevisiae*.

[0207] SEQ ID NO: 140 corresponds to the nucleotide sequence of PCR primer E, present in the COX1 gene of *Saccharomyces cerevisiae*.

[0208] SEQ ID NO: 141 corresponds to the nucleotide sequence of PCR primer F, present in the COX1 gene of *Saccharomyces cerevisiae*.

[0209] SEQ ID NO: 142 corresponds to the nucleotide sequence of PCR primer 11, present in the GFP coding region of the donor DNA.

[0210] SEQ ID NO: 143 corresponds to the nucleotide sequence of PCR primer 12, present in the GFP coding region of the donor DNA.

[0211] SEQ ID NO: 144 corresponds to the nucleotide sequence derived from the PCR amplification products of the GFP integration region in transformed yeast mitochondrial DNA.

[0212] SEQ ID NO: 145 corresponds to the nucleotide sequence of a first guide RNA target site in the *psaA* gene of *Chlamydomonas reinhardtii* plastid DNA. The last three nucleotides are the PAM sequence; these three nucleotides are not present in the variable targeting domain of the corresponding guide RNA.

[0213] SEQ ID NO: 146 corresponds to the nucleotide sequence of a second guide RNA target site in the *psaA* gene of *Chlamydomonas reinhardtii* plastid DNA. The last three nucleotides are the PAM sequence; these three nucleotides are not present in the variable targeting domain of the corresponding guide RNA. This target site sequence is present on the reverse complement of the genic sequence.

[0214] SEQ ID NO: 147 corresponds to the nucleotide sequence of a third guide RNA target site in the *psaA* gene of *Chlamydomonas reinhardtii* plastid DNA. The last three nucleotides are the PAM sequence; these three nucleotides are not present in the variable targeting domain of the corresponding guide RNA.

[0215] SEQ ID NO: 148 corresponds to the nucleotide sequence of a fourth guide RNA target site in the *psaA* gene of *Chlamydomonas reinhardtii* plastid DNA. The last three nucleotides are the PAM sequence; these three nucleotides are not present in the variable targeting domain of the corresponding guide RNA. This target site sequence is present on the reverse complement of the genic sequence.

[0216] SEQ ID NO: 149 corresponds to the nucleotide sequence encoding SpCas9, the Cas9 from *Streptococcus pyogenes*. The coding sequence was codon-optimized for expression in *Chlamydomonas* chloroplasts.

[0217] SEQ ID NO: 150 corresponds to the amino acid sequence of SpCas9, the Cas9 from *Streptococcus pyogenes*, which is encoded by the nucleotide sequences of SEQ ID NO: 150 and SEQ ID NO: 120.

[0218] SEQ ID NO: 151 corresponds to the nucleotide sequence of the promoter and 5' UTR of the *psaA*-exon 1 gene of *Chlamydomonas reinhardtii* plastid DNA.

[0219] SEQ ID NO: 152 corresponds to the nucleotide sequence of the promoter and 5' UTR of the *psbD* gene of *Chlamydomonas reinhardtii* plastid DNA.

[0220] SEQ ID NO: 153 corresponds to the nucleotide sequence of the terminator of the *rbcl* gene of *Chlamydomonas reinhardtii* plastid DNA.

[0221] SEQ ID NO: 154 corresponds to the nucleotide sequence of the promoter of the *trnW* gene of *Chlamydomonas reinhardtii* plastid DNA.

[0222] SEQ ID NO: 155 corresponds to the nucleotide sequence of the 3' UTR of the *trnW* gene of *Chlamydomonas reinhardtii* plastid DNA.

[0223] SEQ ID NO: 156 corresponds to the nucleotide sequence encoding the tRNA of the *trnW* gene of *Chlamydomonas reinhardtii* plastid DNA.

[0224] SEQ ID NO: 157 corresponds to the nucleotide sequence encoding the tRNA of the *trnK* gene of *Chlamydomonas reinhardtii* plastid DNA.

[0225] SEQ ID NO: 158 corresponds to the nucleotide sequence encoding the tRNA of the *trnL* gene of *Chlamydomonas reinhardtii* plastid DNA.

[0226] SEQ ID NO: 159 corresponds to the nucleotide sequence encoding the *aadA* selectable marker.

[0227] SEQ ID NO: 160 corresponds to the nucleotide sequence of the promoter and 5' UTR of the *rbcl* gene of *Chlamydomonas reinhardtii* plastid DNA.

[0228] SEQ ID NO: 161 corresponds to the nucleotide sequence of the 3' UTR of the *psbA* gene of *Chlamydomonas reinhardtii* plastid DNA.

[0229] SEQ ID NO: 162 corresponds to the nucleotide sequence encoding GFP. The coding sequence was codon-optimized for expression in *Chlamydomonas* chloroplasts.

[0230] SEQ ID NO: 163 corresponds to the nucleotide sequence encoding HR1, a homologous region from *Chlamydomonas reinhardtii* plastid DNA, that is present in a donor DNA.

[0231] SEQ ID NO: 164 corresponds to the nucleotide sequence encoding HR2, a homologous region from *Chlamydomonas reinhardtii* plastid DNA, that is present in a donor DNA.

[0232] SEQ ID NO: 165 corresponds to the nucleotide sequence encoding HR3, a homologous region from *Chlamydomonas reinhardtii* plastid DNA, that is present in a donor DNA.

[0233] SEQ ID NO: 166 corresponds to the nucleotide sequence encoding HR4, a homologous region from *Chlamydomonas reinhardtii* plastid DNA, that is present in a donor DNA.

[0234] SEQ ID NO: 167 corresponds to the nucleotide sequence of the forward primer of Primer Set 1 (PS1 FWD Primer), designed to amplify 852 bp of the GFP integration region in the transformed *Chlamydomonas reinhardtii* plastid DNA. PS1 FWD Primer is a chloroplast genomic region-specific primer.

[0235] SEQ ID NO: 168 corresponds to the nucleotide sequence of the reverse primer of Primer Set 1 (PS1 REV Primer), designed to amplify 852 bp of the GFP integration region in the transformed *Chlamydomonas reinhardtii* plastid DNA. PS1 REV Primer is a GFP gene-specific primer.

[0236] SEQ ID NO: 169 corresponds to the nucleotide sequence of the forward primer of Primer Set 2 (PS2 FWD

Primer), designed to amplify 712 bp of the GFP integration region in the transformed *Chlamydomonas reinhardtii* plastid DNA. PS2 FWD Primer is a GFP gene-specific primer.

[0237] SEQ ID NO: 170 corresponds to the nucleotide sequence of the reverse primer of Primer Set 2 (PS2 REV Primer), designed to amplify 712 bp of the GFP integration region in the transformed *Chlamydomonas reinhardtii* plastid DNA. PS2 REV Primer is a chloroplast genomic region-specific primer.

[0238] SEQ ID NO: 171 corresponds to the nucleotide sequence derived from the PCR amplification products of the GFP integration region in transformed *Chlamydomonas reinhardtii* plastid DNA.

[0239] SEQ ID NO: 172 corresponds to the amino acid sequence of a permeant peptide derived from the third alpha helix of *Drosophila melanogaster* transcription factor Antennapedia.

DETAILED DESCRIPTION

[0240] The present disclosure now will be described more fully hereinafter but should not be construed as limited to the embodiments set forth herein.

[0241] The meaning of abbreviations can be as follows: “sec” can mean second(s), “min” can mean minute(s), “h” can mean hour(s), “d” can mean day(s), “μL” can mean microliter(s), “mL” can mean milliliter(s), “L” can mean liter(s), “μM” can mean micromolar, “mM” can mean millimolar, “M” can mean molar, “mmol” can mean millimole(s), “μmole” can mean micromole(s), “g” can mean gram(s), “μg” can mean microgram(s), “ng” can mean nanogram(s), “U” can mean unit(s), “nt” can mean nucleotide(s); “bp” can mean base pair(s), “kb” can mean kilobase(s) and “kbp” can mean kilobase pair(s).

[0242] “Transgenic” can refer to any cell, cell line, callus, tissue, organism part or whole organism (e.g., plant), the genome of which has been altered by the presence of a heterologous nucleic acid, such as a recombinant DNA construct. Transgenic events can include those created by sexual crosses or asexual propagation. In some embodiments, the term “transgenic” may not encompass the alteration of the genome (e.g., chromosomal or extra-chromosomal) by breeding methods or by naturally occurring events such as random cross-fertilization, non-recombinant viral infection, non-recombinant bacterial transformation, non-recombinant transposition, or spontaneous mutation. In some embodiments, the term “transgenic” may encompass the alteration of the genome (e.g., chromosomal or extra-chromosomal) by breeding methods or by naturally occurring events such as random cross-fertilization, non-recombinant viral infection, non-recombinant bacterial transformation, non-recombinant transposition, or spontaneous mutation.

[0243] “Genome”, for example, of a cell or whole organism can encompass chromosomal DNA found within the nucleus (nuclear DNA), and organellar DNA (e.g., mitochondrial DNA, plastid DNA) found within subcellular components of the cell. Methods and compositions of the disclosure can be used for editing of the nuclear genome, organellar genome (e.g., mitochondria, chloroplasts), or both.

[0244] The terms “full complement” and “full-length complement” can be used interchangeably herein, and can refer to a complement of a given nucleotide sequence. In some aspects, the complement and the nucleotide sequence

comprise of the same number of nucleotides. In some aspects, the complement and the nucleotide sequence can comprise 100% complementary. The complement and the nucleotide sequence can differ in the number of nucleotides. Complementarity (e.g., between the complement and the nucleotide sequence) can be at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or 100%. Complementarity (e.g., between the complement and the nucleotide sequence) can be at most about 10%, at most about 20%, at most about 30%, at most about 40%, at most about 50%, at most about 60%, at most about 65%, at most about 70%, at most about 75%, at most about 80%, at most about 85%, at most about 90%, at most about 95%, at most about 97%, at most about 98%, at most about 99%, or 100%.

[0245] “Polynucleotide”, “nucleic acid”, “nucleic acid sequence”, “nucleotide sequence”, or “nucleic acid fragment”, which can be used interchangeably, can refer to a polymer of a nucleic acid (e.g., RNA, DNA, or both, and analogs thereof) that can be single-stranded or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. Nucleotides (e.g., in their 5'-monophosphate form) can be referred to by their single letter designation as follows (for RNA or DNA, respectively): “A” for adenylate or deoxyadenylate, “C” for cytidylate or deoxycytidylate, “G” for guanylate or deoxyguanylate, “U” for uridylate, “T” for deoxythymidylate, “R” for purine-based nucleotides (A or G), “Y” for pyrimidine-based nucleotides (C or T), “K” for G or T, “H” for A or C or T, “I” for inosine, and “N” for any nucleotide.

[0246] “Polypeptide”, “peptide”, “amino acid sequence” and “protein”, which can be used interchangeably herein, can refer to a polymer of amino acid residues. The terms can apply to amino acid polymers in which one or more amino acid residue can be, for example, an artificial chemical analogue of a corresponding naturally occurring amino acid and/or to naturally occurring amino acid polymers. The terms “polypeptide”, “peptide”, “amino acid sequence”, and “protein” can be inclusive of modifications including, but not limited to, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation.

[0247] A “functional fragment” of a polynucleotide or polypeptide can refer to any subset of contiguous nucleotides or contiguous amino acids, respectively, in which the original (e.g., wild type) activity (or substantially similar activity) of the polynucleotide or polypeptide can be retained. The terms “functional fragment”, “functional subfragment”, “fragment that is functionally equivalent”, “subfragment that is functionally equivalent”, “functionally equivalent fragment” and “functionally equivalent subfragment” can be used interchangeably herein.

[0248] The terms “functional variant”, “variant that is functionally equivalent” and “functionally equivalent variant” can be used interchangeably herein. In the context of a polynucleotide or a polypeptide, these terms can refer to a variant of the nucleic acid sequence or the amino acid sequence, respectively, in which the original activity (or substantially similar activity) of the polynucleotide or poly-

peptide can be retained. Fragments and variants can be obtained via methods such as site-directed mutagenesis and synthetic construction.

[0249] The activity of the functional fragment or function variant can be, for example, about: 100%, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 50%, 40%, 30%, 20%, 10%, or less than 10% of that of the original (e.g., wild type) activity.

[0250] “RNA transcript” can refer to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it can be referred to as the primary transcript. A RNA transcript can be referred to as the mature RNA, for example, when it is a RNA sequence derived from post-transcriptional processing of the primary transcript.

[0251] “Messenger RNA” or “mRNA” can refer to the RNA that is without introns and that can be translated into protein by the cell.

[0252] “Sense” RNA can refer to the RNA transcript that includes the mRNA. Sense RNA can be translated into protein within a cell or in vitro.

[0253] “Antisense RNA” can refer to an RNA transcript that can be complementary to all or part of a target RNA (e.g., a primary transcript or mRNA). Antisense RNA can be used to block expression of a target gene. The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. “Functional RNA” can refer to antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet can have an effect on cellular processes. The terms “complement” and “reverse complement” can be used interchangeably herein, for example, with respect to mRNA transcripts and may be used to define the antisense RNA of the message.

[0254] “cDNA” can refer to a DNA that can be complementary to and synthesized from a mRNA template using the enzyme reverse transcriptase. The cDNA can be single-stranded or converted into the double-stranded form using the Klenow fragment of DNA polymerase I.

[0255] “Coding region” can refer to the portion of a messenger RNA (or the corresponding portion of another nucleic acid molecule such as a DNA molecule) which can encode a protein or polypeptide. “Non-coding region” can refer to a portion of a messenger RNA or other nucleic acid molecule that are not a coding region, including but not limited to, for example, the promoter region, 5' untranslated region (“UTR”), 3' UTR, intron and terminator. The terms “coding region” and “coding sequence” can be used interchangeably herein. The terms “non-coding region” and “non-coding sequence” can be used interchangeably herein.

[0256] “Coding sequence” can be abbreviated “CDS”. “Open reading frame” can be abbreviated “ORF”.

[0257] An “Expressed Sequence Tag” (“EST”) can be a DNA sequence derived from a cDNA library. An EST can be a sequence which has been transcribed. An EST can be obtained by a single sequencing pass of a cDNA insert. The sequence of an entire cDNA insert can be termed the “Full-Insert Sequence” (“FIS”). A “Contig” sequence can be a sequence assembled from two or more sequences that can be selected from, but not limited to, the group consisting of an EST, FIS and PCR sequence. A sequence encoding an

entire or functional protein can be termed a “Complete Gene Sequence” (“CGS”). A CGS can be derived from an FIS or a contig.

[0258] “Gene” can refer to a nucleic acid fragment that can express a functional molecule such as, but not limited to, a specific protein, including: introns, exons, regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. “Native gene” can refer to a gene as found in nature, for example, with its own regulatory sequences.

[0259] A “mutated gene” can be a gene that has been altered relative to the corresponding naturally occurring gene; e.g., through human intervention. Such a “mutated gene” can have a sequence that differs from the sequence of the corresponding non-mutated gene by at least one nucleotide addition, deletion, or substitution. In certain embodiments of the disclosure, the mutated gene can comprise an alteration that results from a polynucleotide guided polypeptide system as disclosed herein. A mutated organism can be an organism comprising a mutated gene; e.g., a mutated plant with an organellar genome comprising a mutated gene. The terms “mutated gene” and “mutant gene” can be used interchangeably herein.

[0260] A “silent mutation” can refer to a mutated sequence that has the same functionality as the wild-type sequence; e.g., replacement of a codon in a protein-coding region with a synonymous codon that can encode the same amino acid.

[0261] As used herein, a “targeted mutation” can be a DNA modification made at or near a specific target site in the genome. The targeted mutation may be as small as a single nucleotide change in a native gene. The targeted mutation may involve a larger DNA modification such as the insertion of one or more heterologous DNAs; e.g., a heterologous regulatory element, a heterologous protein-coding sequence, or an expression cassette coding for a heterologous protein or functional RNA. The targeted mutation may also involve a change in the sequence of a target site.

[0262] The term “SDN” can refer to “site-directed nuclease”. The following are non-limiting examples of SDN-induced mutations: (1) induction of site-specific random mutations; (2) the induction of mutations in a predefined sequence of a particular gene; and (3) the replacement or the insertion of an entire gene. These SDN-induced mutations can be referred to as SDN-1, SDN-2 and SDN-3, respectively.

[0263] A “codon-modified gene” or “codon-preferred gene” or “codon-optimized gene” can be a gene having its frequency of codon usage designed to mimic the frequency of preferred codon usage of the host cell in the compartment of interest, e.g., the nucleus, the mitochondria or the chloroplast.

[0264] “Mature” protein can refer to a post-translationally processed polypeptide; for example, one from which any pre- or pro-peptides present in the primary translation product have been removed.

[0265] “Precursor” protein can refer to the primary product of translation of an mRNA; for example, with pre- and pro-peptides still present. Pre- and pro-peptides may, for example, comprise intracellular localization signals.

[0266] “Isolated” can refer to materials, such as nucleic acid molecules, proteins, and cells that may be substantially free or otherwise removed from components that normally accompany or interact with the materials in a naturally occurring environment. Isolated polynucleotides may be

purified from a host cell in which they naturally occur. Nucleic acid purification methods can be used to obtain isolated polynucleotides. Isolated polynucleotides can include, for example, recombinant polynucleotides and chemically synthesized polynucleotides.

[0267] “Heterologous”, for example, with respect to sequence, can mean a sequence that originates from a foreign species, or, if from the same species, is substantially modified from its native form in composition and/or genomic locus by deliberate human intervention. The terms “heterologous nucleotide sequence”, “heterologous sequence”, “heterologous nucleic acid fragment”, and “heterologous nucleic acid sequence” can be used interchangeably herein.

[0268] “Recombinant” can refer to an artificial combination of two or more otherwise separated segments of sequence, e.g., by chemical synthesis or by the manipulation of isolated segments of nucleic acids by genetic engineering techniques. “Recombinant” can also include reference to a cell or vector, for example, that has been modified by the introduction of a heterologous nucleic acid or a cell derived from a cell so modified.

[0269] “Recombinant DNA construct” can refer to a combination of nucleic acid fragments that may not normally be found together in nature. A recombinant DNA construct may comprise, for example, regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source. The sequences in a recombinant DNA construct can be arranged in a manner different than that normally found in nature. The terms “recombinant DNA construct”, “recombinant DNA molecule”, “recombinant construct”, “DNA construct” and “construct” can be used interchangeably herein.

[0270] “Expression” can refer to the production of a functional product. For example, expression of a nucleic acid fragment may refer to transcription of the nucleic acid fragment (e.g., transcription resulting in mRNA or functional RNA) and/or translation of mRNA into a precursor or mature protein.

[0271] “Expression cassette” can refer to a construct containing, for example, a polynucleotide, a regulatory element (s), and a polynucleotide that allow for expression of the polynucleotide in a host. The terms “expression cassette” and “expression construct” can be used interchangeably herein.

[0272] The terms “entry clone” and “entry vector” can be used interchangeably herein.

[0273] “Regulatory sequences” can refer to nucleotide sequences, for example, located upstream (e.g., 5' non-coding sequences), within (e.g., in introns), or downstream (e.g., 3' non-coding sequences) of a coding sequence. Regulatory sequences can influence, for example, the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include, but are not limited to, promoters, translation leader sequences, 5' untranslated sequences, 3' untranslated sequences, introns, polyadenylation target sequences, RNA processing sites, effector binding sites, and stem-loop structures. A regulatory sequence may act in “cis” or “trans”. The nucleic acid molecule regulated by a regulatory sequence may not necessarily have to encode a functional peptide or polypeptide, e.g., the regulatory sequence can modulate the expression of a short interfering RNA or an anti-sense RNA.

The terms “regulatory sequence” and “regulatory element” can be used interchangeably herein.

[0274] “Promoter” can refer to a nucleic acid fragment that can control transcription of another nucleic acid fragment. A promoter can include a core promoter (also known as minimal promoter) sequence. A core promoter can be a minimal sequence for direct transcription initiation. A core promoter can optionally include enhancers or other regulatory elements. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. Different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions.

[0275] “Promoter functional in a plant” can be a promoter that can control transcription in plant cells. The promoter can be from any suitable origin, which can include plant cells and non-plant cells.

[0276] “Tissue-specific promoter” and “tissue-preferred promoter” can be used interchangeably, and can refer to a promoter that can be expressed predominantly in one tissue, one organ or one cell type. A tissue-specific promoter may not be necessarily exclusive in one tissue, one organ or one cell type. Root-preferred promoters include, for example, the following: soybean root-specific glutamine synthase gene; cytosolic glutamine synthase (GS); root-specific control element in the GRP 1.8 gene of French bean; root-specific promoter of *A. tumefaciens* mannopine synthase (MAS); root-specific promoters isolated from *Parasponia andersonii* and *Trema tomentosa*; *A. rhizogenes* rolC and rolD root-inducing genes; *Agrobacterium* wound-induced TR1' and TR2' genes; VtENOD-GRP3 gene promoter; and rolB promoter. Seed-preferred promoters include both seed-specific promoters active during seed development, as well as seed-germinating promoters active during seed germination. Seed-preferred promoters include, but are not limited to, the following: Cim1 (cytokinin-induced message); cZ19B1 (maize 19 kDa zein); milps (myo-inositol-1-phosphate synthase); END1; and END2. For dicots, seed-preferred promoters include, but are not limited to, the following: bean β -phaseolin; napin; β -conglycinin; soybean lectin; cruciferin; and the like. For monocots, seed-preferred promoters include, but are not limited to, the following: maize 15 kDa zein; 22 kDa zein; 27 kDa gamma zein; waxy; shrunken 1; shrunken 2; globulin 1; oleosin; nud; and *Zea mays*-Rootmet2 promoter. Leaf-preferred promoters include, but are not limited to, the following: plant rbcS promoters, such as the soybean rbcS promoter and the maize rbcS promoter; *Zea mays* PEPC1 promoter.

[0277] “Developmentally regulated promoter” can refer to a promoter whose activity can be determined by developmental events.

[0278] “Inducible promoter” can refer to a promoter that selectively expresses an operably linked DNA sequence in response to the presence of an endogenous or exogenous stimulus, for example by chemical compounds (e.g., chemical inducers) or in response to environmental, hormonal, chemical, and/or developmental signals. Inducible or regulated promoters include, for example, promoters regulated by light, heat, stress, flooding or drought, phytohormones, wounding, or chemicals such as ethanol, jasmonate, salicylic acid, or safeners. Pathogen-inducible promoters induced following infection by a pathogen include, but are not

limited to those regulating expression of PR proteins, SAR proteins, beta-1,3-glucanase, chitinase, etc. Stress-inducible promoters include plant RAB17 promoters, such as the maize RAB17 promoter. Chemical-inducible promoters include, but are not limited to, the following: the maize ln2-2 promoter, activated by benzene sulfonamide herbicide safeners; the maize GST promoter, activated by hydrophobic electrophilic compounds used as pre-emergent herbicides; and the tobacco PR-1a promoter, activated by salicylic acid. Other chemical-regulated promoters include steroid-responsive promoters, for example, the glucocorticoid-inducible promoter, and tetracycline-inducible and tetracycline-repressible promoters.

[0279] “Constitutive promoter” can refer to promoters active in all or most tissues or cell types of an organism at all or most developing stages. As with other promoters classified as “constitutive” (e.g. ubiquitin), some variation in absolute levels of expression can exist among different tissues or stages. The term “constitutive promoter” or “tissue-independent promoter” can be used interchangeably herein. Constitutive promoters include the following: the core promoter of the Rsyn7 promoter; the core CaMV 35S promoter; plant actin promoter, such as a rice actin promoter and a maize actin promoter; plant ubiquitin promoter, such as a maize ubiquitin promoter and a soybean ubiquitin promoter; pEMU; MAS promoter; ALS promoter; plant GOS2 promoter, such as a maize GOS2 promoter; soybean GM-EF1 A2 promoter; plant U6 polymerase III promoter, such as a maize U6 polymerase III promoter and a soybean U6 polymerase III promoter (GM-U6-9.1 and GM-U6-13.1).

[0280] An enhancer element can be any nucleic acid molecule that increases transcription of a nucleic acid molecule when functionally linked to a promoter regardless of its relative position. An enhancer may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter.

[0281] A repressor (also sometimes called herein silencer) can be defined as any nucleic acid molecule which inhibits the transcription when functionally linked to a promoter regardless of relative position.

[0282] “Translation leader sequence” can refer to a polynucleotide sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence can be present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency.

[0283] “Transcription terminator”, “termination sequence”, or “terminator” can refer to DNA sequences that, when operably linked to the 3' end of a polynucleotide sequence that is to be expressed, can terminate transcription from the polynucleotide sequence. Transcription termination can refer to the process by which RNA synthesis by RNA polymerase can be stopped and both the RNA and the enzyme are released from the DNA template.

[0284] “Operably linked” can refer to the association of fragments in a single fragment (e.g., a polynucleotide or polypeptide), or in a single complex, so that the function of one can be regulated by the other. The linkage may be covalent or non-covalent. For example, with respect to nucleic acid fragments, a promoter can be operably linked with a nucleic acid fragment if the promoter can regulate the transcription of that nucleic acid fragment. For example,

with respect to a polypeptide, an organelle targeting peptide can be operably linked with a polypeptide if the organelle targeting peptide can transport that polypeptide into the relevant organelle. For example, with respect to a complex, a guide RNA can be operably linked to a Cas polypeptide if the guide RNA/Cas polypeptide complex can cleave a target sequence as directed by the guide RNA.

[0285] “Phenotype” can refer to the detectable characteristics of a cell or organism.

[0286] The term “introduced” can mean providing a polynucleic acid (e.g., expression construct) or protein into a cell. Introduced can include reference to the incorporation of a nucleic acid into a eukaryotic or prokaryotic cell, for example, where the nucleic acid may be incorporated into the genome of the cell. Introduced can include reference to the transient provision of a nucleic acid or protein to the cell. Introduced can include reference to stable or transient transformation methods. Introduced can include sexually crossing. Introduced, for example, in the context of inserting a nucleic acid fragment (e.g., a recombinant DNA construct) into a cell, can include “transfection” or “transformation” or “transduction”. Introduced can include reference to the incorporation of a nucleic acid fragment into a eukaryotic or prokaryotic cell where the nucleic acid fragment may be incorporated into the genome of the cell (e.g., chromosome, plasmid, plastid or mitochondrial DNA), converted into an autonomous replicon, or transiently expressed (e.g., transfected mRNA).

[0287] A “transformed cell” can be any cell into which a nucleic acid fragment (e.g., a recombinant DNA construct) has been introduced.

[0288] “Transformation” as used herein can refer to stable transformation. Transformation can refer to transient transformation.

[0289] “Stable transformation” can refer to the introduction of a nucleic acid fragment into a genome of a host organism resulting in genetically stable inheritance. Once stably transformed, the nucleic acid fragment can be stably integrated in the genome of the host organism and any subsequent generation.

[0290] “Transient transformation” can refer to the introduction of a nucleic acid fragment into the nucleus, or DNA-containing organelle, of a host organism resulting in gene expression without genetically stable inheritance.

[0291] Host organisms containing the transformed nucleic acid fragments can be referred to as “transgenic” organisms.

[0292] “Transformation cassette” can refer to a construct having elements that facilitates transformation of a particular host cell. The terms “transformation cassette” and “transformation construct” can be used interchangeably herein.

[0293] “Allele” can be one of several alternative forms of a gene occupying a given locus on a chromosome. When the alleles present at a given locus on a pair of homologous chromosomes in a diploid plant are the same that plant can be homozygous at that locus. If the alleles present at a given locus on a pair of homologous chromosomes in a diploid plant differ, that plant can be heterozygous at that locus. If a transgene is present on one of a pair of homologous chromosomes in a diploid plant that plant can be hemizygous at that locus.

[0294] A “chloroplast transit peptide” can be an amino acid sequence that can direct a protein to the chloroplast or other plastid types present in the cell. The chloroplast transit peptide can be translated in conjunction with the protein in

the cell in which the protein can be made. The terms “chloroplast transit peptide”, “plastid transit peptide”, “chloroplast targeting peptide” and “plastid targeting peptide” can be used interchangeably herein. “Chloroplast transit sequence” can refer to a nucleotide sequence that can encode a chloroplast transit peptide.

[0295] A “signal peptide” can be an amino acid sequence that can direct a protein to the secretory system. The signal peptide can be translated in conjunction with a protein. For example, if the protein is to be directed to a vacuole, a vacuolar targeting signal (supra) can further be added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (supra) may be added. If the protein is to be directed to the nucleus, any signal peptide present may be removed and a nuclear localization signal can be included.

[0296] A “mitochondrial signal peptide” can be an amino acid sequence which can direct a precursor protein into the mitochondria. The terms “mitochondrial signal peptide”, “mitochondrial transit peptide” and “mitochondrial targeting peptide” can be used interchangeably herein.

[0297] An “organelle targeting polynucleotide” can be a nucleotide sequence which can direct import of the polynucleotide into an organelle. The terms “organelle targeting polynucleotide”, “organelle targeting nucleic acid” and “organelle targeting nucleic acid sequence” can be used interchangeably herein. An organelle targeting polynucleotide may be directed to, for example, the plastid (“plastid targeting polynucleotide”) or the mitochondria (“mitochondria targeting polynucleotide”). The polynucleotide may be RNA (“organelle targeting RNA”), DNA (“organelle targeting DNA”) or a combination of RNA and DNA. An organelle targeting RNA directed to the plastid can be termed a “plastid targeting RNA”. The terms “plastid targeting RNA”, “chloroplast targeting RNA” and “transit RNA” are used interchangeably herein. An organelle targeting RNA directed to the mitochondria can be termed a “mitochondria targeting RNA”.

[0298] RNAs can be imported into mitochondria. One such mitochondrial targeting RNA can be the yeast tRNA^{Lys}. The yeast tRNA^{Lys} and its variants can be imported into human mitochondria. Another RNA that can be imported into mitochondria can be 5S rRNA. 5S rRNA can function as a vector for delivering heterologous RNA sequences into, for example, mitochondria (e.g., human). Such RNAs can be used with the compositions and methods of the disclosure for example, for targeting to an organelle (e.g., the mitochondria).

[0299] RNAs can be imported into plastids. Plastid targeting RNAs that can mediate import of attached heterologous RNA can include vd-5'UTR (e.g., viroid-derived ncRNA sequence acting as 5'UTR and eIF4E1 mRNA). Such RNAs can be used with the compositions and methods of the disclosure for targeting to an organelle (e.g., the plastid).

[0300] As used herein, “fusion” can refer to a protein and/or nucleic acid comprising one or more non-native sequences (e.g., moieties). Any of the molecules described herein (e.g., nucleic acids, proteins, polypeptides, polynucleic acid, Cas protein, guide polynucleotide) can be engineered as fusions. A fusion can comprise one or more of the same non-native sequences. A fusion can comprise one or more of different non-native sequences. A fusion can be a chimera. A fusion can comprise a nucleic acid affinity tag. A fusion can comprise a barcode. A fusion can comprise a peptide affinity tag. A fusion can provide for subcellular

localization of the site-directed polypeptide. A fusion can provide a non-native sequence (e.g., affinity tag) that can be used to track or purify. A fusion can be a small molecule such as biotin or a dye such as alexa fluor dyes, Cyanine3 dye, and Cyanine5 dye.

[0301] A fusion can refer to any protein with a functional effect. For example, a fusion protein can comprise deaminase activity, cytidine deaminase activity (US Patent Publication No. US20150166980, herein incorporated by reference), adenine deaminase activity (US Patent Publication No. US20180073012, herein incorporated by reference), uracil glycosylase inhibitor activity (US Patent Publication No. US20170121693, herein incorporated by reference), methyltransferase activity, demethylase activity, dismutase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity, transposase activity, recombinase activity, polymerase activity, ligase activity, helicase activity, photolyase activity or glycosylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity, remodeling activity, protease activity, oxidoreductase activity, transferase activity, hydrolase activity, lyase activity, isomerase activity, synthase activity, synthetase activity, or demyristoylation activity. An effector protein can modify a genomic locus. A fusion protein can be a fusion in a Cas protein. The Cas protein may be a modified form that has nickase activity or that has no substantial nucleic acid-cleaving activity. A fusion protein can be a non-native sequence in a Cas protein.

[0302] As used herein, a “nucleic acid” can refer to a polynucleotide sequence, or fragment thereof. A nucleic acid can comprise nucleotides. A nucleic acid can be exogenous or endogenous to a cell. A nucleic acid can exist in a cell-free environment. A nucleic acid can be a gene or fragment thereof. A nucleic acid can be DNA. A nucleic acid can be RNA. A nucleic acid can comprise one or more analogs (e.g. altered backbone, sugar, or nucleobase). Some non-limiting examples of analogs include: 5-bromouracil, peptide nucleic acid, xeno nucleic acid, morpholinos, locked nucleic acids, glycol nucleic acids, threose nucleic acids, dideoxynucleotides, cordycepin, 7-deaza-GTP, fluorophores (e.g. rhodamine or fluorescein linked to the sugar), thiol containing nucleotides, biotin linked nucleotides, fluorescent base analogs, CpG islands, methyl-7-guanosine, methylated nucleotides, inosine, thiouridine, pseudouridine, dihydrouridine, queuosine, and wyosine.

Suppression of Gene Expression

[0303] “Suppression DNA construct” can be a recombinant DNA construct which when transformed or stably integrated into the genome of the plant, can result in “silencing” of a target gene (e.g., in a plant). The target gene may be endogenous or transgenic to a target cell (e.g., plant).

[0304] “Silencing,” as used herein with respect to the target gene, can refer to the suppression of levels of mRNA or protein/enzyme expressed by the target gene, and/or the level of the enzyme activity or protein functionality. The terms “suppression”, “suppressing” and “silencing”, which can be used interchangeably herein, can include lowering, reducing, declining, decreasing, inhibiting, eliminating or preventing. “Silencing” or “gene silencing” can occur by

any suitable mechanism. Non-limiting examples of silencing can include anti-sense, cosuppression, viral-suppression, hairpin suppression, stem-loop suppression, RNAi-based approaches, and small RNA-based approaches

[0305] A suppression DNA construct may comprise a region derived from a target gene of interest. A suppression DNA construct may comprise all or part of the nucleic acid sequence of the sense strand (or antisense strand, or both) of the target gene of interest. The region may be 100% identical or less than 100% identical (e.g., at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical) to all or part of the sense strand (or antisense strand, or both) of the gene of interest. A suppression DNA construct may comprise 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 or 1000 contiguous nucleotides of the sense strand (or antisense strand, or both) of the gene of interest, and combinations thereof.

[0306] Suppression DNA constructs can be readily constructed, for example, once the target gene of interest is selected. A suppression DNA construct can include, without limitation, cosuppression constructs, antisense constructs, viral-suppression constructs, hairpin suppression constructs, stem-loop suppression constructs, double-stranded RNA-producing constructs, and more generally, RNAi (RNA interference) constructs and small RNA constructs such as siRNA (short interfering RNA) constructs and miRNA (microRNA) constructs.

[0307] Suppression of gene expression may also be achieved by, for example, use of artificial miRNA precursors, ribozyme constructs and gene disruption. A modified plant miRNA precursor may be used, wherein the precursor has been modified, for example, to replace the miRNA encoding region with a sequence designed to produce a miRNA directed to the nucleotide sequence of interest. Gene disruption may be achieved by use of transposable elements or by use of chemical agents that cause site-specific mutations.

[0308] “Antisense inhibition” can refer to the production of antisense RNA transcripts that can suppress the expression of the target gene or gene product. “Antisense RNA” can refer to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA. Antisense RNA can block the expression of a target isolated nucleic acid fragment. The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence.

[0309] “Cosuppression” can refer to the production of sense RNA transcripts that can suppress the expression of the target gene or gene product. “Sense” RNA can refer to RNA transcript that can include the mRNA. Sense RNA can be translated into protein within a cell or in vitro. Cosuppression constructs in plants can be designed, for example, by focusing on overexpression of a nucleic acid sequence having homology to a native mRNA, in the sense orientation, which can result in the reduction of RNA having homology to the overexpressed sequence.

[0310] Plant viral sequences can be used to direct the suppression of proximal mRNA encoding sequences.

[0311] RNA interference can refer to the process of sequence-specific post-transcriptional gene silencing (e.g., in animals) mediated by, for example, short interfering RNAs (siRNAs). The corresponding process in plants can be referred to as post-transcriptional gene silencing (PTGS) or RNA silencing and can also be referred to as quelling in fungi. The process of post-transcriptional gene silencing can be an evolutionarily-conserved cellular defense mechanism used to prevent the expression of foreign genes. Post-transcriptional gene silencing can be shared by diverse flora and phyla.

[0312] Small RNAs can play an important role in controlling gene expression. Small RNAs can function by base-pairing to complementary RNA or DNA target sequences. When bound to RNA, small RNAs can trigger either RNA cleavage or translational inhibition of the target sequence. When bound to DNA target sequences, small RNAs can mediate DNA methylation of the target sequence. Small RNAs can lead to inhibition of gene expression.

[0313] MicroRNAs (miRNAs) can be noncoding RNAs with a length of, for example, about 19 to about 24 nucleotides (nt). MicroRNAs can occur in animals and plants. miRNAs can be processed from longer precursor transcripts that can range in size, for example, from approximately 70 to 200 nt. The precursor transcripts can form stable hairpin structures.

[0314] MicroRNAs (miRNAs) can regulate target genes, for example, by binding to complementary sequences located in the transcripts produced by these genes. miRNAs can enter, for example, at least two pathways of target gene regulation: (1) translational inhibition; and/or (2) RNA cleavage. MicroRNAs entering the RNA cleavage pathway can be analogous to the 21-25 nt short interfering RNAs (siRNAs) generated during RNA interference (RNAi) in animals and posttranscriptional gene silencing (PTGS) in plants. These microRNAs entering the RNA cleavage pathway can be incorporated into an RNA-induced silencing complex (RISC) that can be similar or identical to that seen for RNAi.

[0315] The terms “miRNA-star sequence” and “miRNA* sequence” can be used interchangeably herein and can refer to a sequence in the miRNA precursor that can be highly complementary to the miRNA sequence. The miRNA and miRNA* sequences can form part of the stem region of the miRNA precursor hairpin structure.

Sequence Identity, Similarity and Variation

[0316] Sequence alignments and percent identity or similarity calculations may be determined using a variety of comparison methods designed to detect homologous sequences including, but not limited to, the MEGALIGN™ program of the LASERGENE™ bioinformatics computing suite (DNASTAR™ Inc., Madison, Wis.). In some embodiments, where sequence analysis software is used for analysis, the results of the analysis can be based on the “default values” of the program referenced. As used herein “default values” can mean any set of values or parameters that originally load with the software when first initialized.

[0317] The “Clustal V method of alignment” can correspond to the alignment method labeled Clustal V and, for example, found in the MEGALIGN™ program of the LASERGENE™ bioinformatics computing suite (DNASTAR™ Inc., Madison, Wis.). For multiple alignments, the default values can correspond to GAP PENALTY=10 and

GAP LENGTH PENALTY=10. Default parameters for pairwise alignments and calculation of percent identity of protein sequences using the Clustal method can be, for example, KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. For nucleic acids these parameters can be for example KTUPLE=2, GAP PENALTY=5, WINDOW=4 and DIAGONALS SAVED=4. After alignment of the sequences using the Clustal V program, “percent identity” and “divergence” values can be obtained by viewing the “sequence distances” table in the same program.

[0318] The “Clustal W method of alignment” can correspond to the alignment method labeled Clustal W and, for example, found in the MEGALIGN™ v6.1 program of the LASERGENE™ bioinformatics computing suite (DNASTAR™ Inc., Madison, Wis.). Default parameters for multiple alignment can correspond to for example: GAP PENALTY=10, GAP LENGTH PENALTY=0.2, Delay Divergence Sequences=30%, DNA Transition Weight=0.5, Protein Weight Matrix=Gonnet Series, DNA Weight Matrix=IUB. After alignment of the sequences using the Clustal W program, “percent identity” values can be obtained by viewing the “sequence distances” table in the same program.

[0319] Sequence identity/similarity values can also be obtained using GAP Version 10 (GCG, Accelrys, San Diego, Calif.) using for example the following parameters: % identity and % similarity for a nucleotide sequence using a gap creation penalty weight of 50 and a gap length extension penalty weight of 3, and the nws gapdna.cmp scoring matrix; % identity and % similarity for an amino acid sequence using a GAP creation penalty weight of 8 and a gap length extension penalty of 2, and the BLOSUM62 scoring matrix. GAP can use an algorithm to find an alignment of two complete sequences that can maximize the number of matches and minimizes the number of gaps. GAP can consider all possible alignments and gap positions. GAP can create the alignment with the largest number of matched bases and the fewest gaps, using, for example, a gap creation penalty and a gap extension penalty in units of matched bases.

[0320] “BLAST” can be a searching algorithm provided by the National Center for Biotechnology Information (NCBI) that can be used to find regions of similarity between biological sequences. The program can compare nucleotide or protein sequences to sequence databases. The program can calculate the statistical significance of matches to identify sequences having sufficient similarity to a query sequence such that the similarity may not be predicted to have occurred randomly. BLAST can report the identified sequences and their local alignment to the query sequence.

[0321] The term “conserved domain” or “motif” can mean a set of amino acids conserved at specific positions along an aligned sequence of evolutionarily related proteins. While amino acids at other positions can vary between homologous proteins, amino acids that are highly conserved at specific positions can indicate, for example, amino acids that are essential to the structure, the stability, or the activity of a protein.

[0322] Conserved domains or motifs can be identified by their high degree of conservation in aligned sequences of a family of protein homologues. Conserved domains can be used as identifiers, or “signatures”, for example, to determine if a protein with a newly determined sequence belongs to a previously identified protein family.

[0323] Polynucleotide and polypeptide sequences, variants thereof, and the structural relationships of these sequences can be described by the terms “homology”, “homologous”, “substantially identical”, “substantially similar” and “corresponding substantially” which are used interchangeably herein. These can refer to polypeptide or nucleic acid fragments wherein changes in one or more amino acids or nucleotide bases may not affect the function of the molecule, such as the ability to mediate gene expression or to produce a certain phenotype. These terms can also refer to modification(s) of nucleic acid fragments that may not substantially alter the functional properties of the resulting nucleic acid fragment relative to the initial, unmodified fragment. These modifications can include deletion, substitution, and/or insertion of one or more nucleotides in the nucleic acid fragment.

[0324] Substantially similar nucleic acid sequences encompassed may be defined by their ability to hybridize (for example, under moderately stringent conditions, e.g., 0.5×SSC, 0.1% SDS, 60° C.) with the sequences exemplified herein, or to any portion of the nucleotide sequences disclosed herein. Substantially similar nucleic acid sequences can be functionally equivalent to any of the nucleic acid sequences disclosed herein. Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. Post-hybridization washes can determine stringency conditions.

[0325] The term “selectively hybridizes” can include reference to hybridization, for example under stringent hybridization conditions, of a nucleic acid sequence to a specified nucleic acid target sequence to a detectably greater degree (e.g., at least 2-fold over background) than its hybridization to non-target nucleic acid sequences and to the substantial exclusion of non-target nucleic acids. Selectively hybridizing sequences can have, for example, about at least 80% sequence identity, or 90% sequence identity, up to and including 100% sequence identity (i.e., fully complementary) with each other.

[0326] The term “stringent conditions” or “stringent hybridization conditions” can include reference to conditions under which a probe can selectively hybridize to its target sequence in an in vitro hybridization assay. Stringent conditions can be sequence-dependent. Stringent conditions can be different in different circumstances. By controlling the stringency of the hybridization and/or washing conditions, target sequences can be identified which are 100% complementary to the probe (homologous probing).

[0327] Alternatively, stringency conditions can be adjusted to allow some mismatching in sequences so that lower degrees of similarity are detected (heterologous probing). In some embodiments, a probe can be less than about 1000 nucleotides in length, optionally less than 500 nucleotides in length.

[0328] In some embodiments, stringent conditions can be those in which the salt concentration is less than about 1.5 M Na ion, for example, about 0.01 to 1.0 M Na ion concentration (or other salt(s)) at pH 7.0 to 8.3, and, for example, at least about 30° C. for short probes (e.g., 10 to 50 nucleotides) and, for example, at least about 60° C. for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of desta-

bilizing agents such as formamide. Exemplary low stringency conditions can include hybridization with a buffer solution of, for example, 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulphate) at 37° C., and a wash in 1× to 2×SSC (20×SSC=3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55° C. Exemplary moderate stringency conditions can include hybridization in 40 to 45% formamide, 1 M NaCl, 1% SDS at 37° C., and a wash in 0.5× to 1×SSC at 55 to 60° C. Exemplary high stringency conditions can include hybridization in, for example, 50% formamide, 1 M NaCl, 1% SDS at 37° C., and a wash in 0.1×SSC at 60 to 65° C.

[0329] “Sequence identity” or “identity” in the context of nucleic acid or polypeptide sequences can refer to the nucleic acid bases or amino acid residues in two sequences that are the same when aligned for maximum correspondence over a specified comparison window.

[0330] The term “percentage of sequence identity” can refer to the value determined by comparing two optimally aligned sequences over a comparison window. The portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which may or may not comprise additions or deletions) for optimal alignment of the two sequences. The percentage can be calculated by, for example, determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the results by 100 to yield the percentage of sequence identity. Percent sequence identities can include, but are not limited to, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% or 95%, or any percentage from 50% to 100%. Sequence identity can include an integer percentage from 50% to 100%. These identities can be determined using any of the programs described herein.

[0331] Sequence identity can be useful in identifying polypeptides from other species or modified naturally or synthetically wherein such polypeptides have the same or similar function or activity. Percent identities can include, but are not limited to, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% or 95%. Sequence identity (e.g., amino acid sequence identity) can include an integer percentage from 50% to 100%. Sequence (e.g., amino acid) identity can include, for example, about: 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100%.

Definitions, Traits and Processes Relevant to Plants

[0332] “Plant” can include reference to whole plants, plant organs, plant tissues, plant propagules, seeds and plant cells and progeny of same. Plant cells include, without limitation, cells from seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen, and microspores.

[0333] “Propagule” can include products of meiosis and/or mitosis able to propagate a new plant. Propagule can include seeds, spores and parts of a plant that can serve as a means of vegetative reproduction, such as corms, tubers, offsets, or runners. Propagule can include grafts where one portion of a plant can be grafted to another portion of a

different plant (even one of a different species) to create a living organism. Propagule can include plants and seeds produced by cloning or by bringing together meiotic products, or allowing meiotic products to come together to form an embryo or fertilized egg (naturally or with human intervention).

[0334] “Progeny” can comprise any subsequent generation of a plant.

[0335] The terms “monocot” and “monocotyledonous plant” can be used interchangeably herein. A monocot can include the Gramineae.

[0336] The terms “dicot” and “dicotyledonous plant” can be used interchangeably herein. A dicot can include, for example, the following families: Brassicaceae, Leguminosae, and Solanaceae.

[0337] “Transgenic plant” can include reference to a plant which comprises within its genome a heterologous polynucleotide. For example, the heterologous polynucleotide may be stably integrated within the genome (e.g., nuclear, plastid, mitochondrial) such that the polynucleotide can be passed on to successive generations. The heterologous polynucleotide may be integrated into the genome alone or as part of a recombinant DNA construct.

[0338] “Transgenic plant” can include reference to plants which can comprise more than one heterologous polynucleotide within their genome. Each heterologous polynucleotide may confer a different trait to the transgenic plant.

[0339] Multiple traits can be introduced into crop plants, and can be referred to as a gene stacking approach. Gene stacking can be used, for example, for development of genetically improved germplasm. In this approach, multiple genes conferring different characteristics of interest can be introduced into a plant. Gene stacking can be accomplished by many means including but not limited to co-transformation, retransformation, and crossing lines with different transgenes. As used herein, the term “stacked” can include having multiple traits present in the same plant (e.g., both traits are incorporated into the nuclear genome, one trait is incorporated into the nuclear genome and one trait is incorporated into the genome of an organelle, or both traits are incorporated into the genome of an organelle).

[0340] The term “crossed” or “cross” or “crossing” in the context of the disclosure can mean the fusion of gametes (e.g., via pollination) to produce progeny (e.g., cells, seeds, or plants). The term can encompass both sexual crosses (e.g., the pollination of one plant by another) and selfing (e.g., self-pollination; when the pollen and ovule are from the same plant or genetically identical plants).

[0341] The term “maternal inheritance” can refer to the transmission of traits that can be solely dependent on properties of the genome of the female gamete.

[0342] The term “paternal inheritance” can refer to the transmission of traits that are solely dependent on properties of the genome of the male gamete.

[0343] The term “introgression” can refer to the transmission of a desired allele of a genetic locus from one genetic background to another. For example, introgression of a desired allele at a specified locus can be transmitted to at least one progeny plant via a sexual cross between two parent plants, where at least one of the parent plants has the desired allele within its genome. Alternatively, for example, transmission of an allele can occur by recombination between two donor genomes, e.g., in a fused protoplast, where at least one of the donor protoplasts has the desired

allele in its genome. The desired allele can be, e.g., a transgene or a selected allele of a marker or QTL.

[0344] “A plant-optimized nucleotide sequence” can be a nucleotide sequence that has been optimized for increased expression in plants, particularly for increased expression in plants or in one or more plants of interest. For example, a plant-optimized nucleotide sequence can be synthesized by modifying a nucleotide sequence encoding a protein such as, for example, a double-strand-break-inducing agent (e.g., an endonuclease) as disclosed herein, using one or more plant-preferred codons for improved expression. A host-preferred codon usage can be utilized for codon optimization.

[0345] Plant-preferred genes can be synthesized. Additional sequence modifications can enhance gene expression in a plant host. These can include, for example, elimination of: one or more sequences encoding spurious polyadenylation signals, one or more exon-intron splice site signals, one or more transposon-like repeats, and sequences that may be deleterious to gene expression. The G-C content of the sequence may be adjusted, for example, to levels average for a given plant host, as calculated by reference to genes expressed in the host plant cell. When possible, the sequence can be modified to avoid one or more predicted hairpin secondary mRNA structures. Thus, “a plant-optimized nucleotide sequence” of the present disclosure can comprise one or more of such sequence modifications.

[0346] A “trait” can refer to, for example, a physiological, morphological, biochemical, or physical characteristic of a plant or particular plant material or cell. In some instances, this characteristic can be visible to the human eye, such as seed or plant size, or can be measured by biochemical techniques, such as detecting the protein, starch, or oil content of seed or leaves, or by observation of a metabolic or physiological process, e.g. by measuring tolerance to water deprivation or particular salt or sugar concentrations, or by the observation of the expression level of a gene or genes, or by agricultural observations such as osmotic stress tolerance or yield.

[0347] “Agronomic characteristic” can be a measurable parameter including but not limited to, abiotic stress tolerance, greenness, yield, growth rate, biomass, fresh weight at maturation, dry weight at maturation, fruit yield, seed yield, total plant nitrogen content, fruit nitrogen content, seed nitrogen content, nitrogen content in a vegetative tissue, total plant free amino acid content, fruit free amino acid content, seed free amino acid content, free amino acid content in a vegetative tissue, total plant protein content, fruit protein content, seed protein content, protein content in a vegetative tissue, drought tolerance, nitrogen uptake, root lodging, harvest index, stalk lodging, plant height, ear height, ear length, salt tolerance, early seedling vigor and seedling emergence under low temperature stress.

[0348] Particular phenotypes may include, but are not limited to kernel number, kernel area, grain weight, and predicted weight of the grain on the ear (based on the calibration of kernel area to grain weight).

[0349] Abiotic stress may be at least one condition selected from the group consisting of: drought, water deprivation, flood, high light intensity, high temperature, low temperature, salinity, etiolation, defoliation, heavy metal toxicity, anaerobiosis, nutrient deficiency, nutrient excess, UV irradiation, atmospheric pollution (e.g., ozone) and exposure to chemicals (e.g., paraquat) that induce production of reactive oxygen species (ROS).

[0350] “Increased stress tolerance” of a plant can be measured relative to a reference or control plant, and can be a trait of the plant to survive under stress conditions over prolonged periods of time, without exhibiting the same degree of physiological or physical deterioration relative to the reference or control plant grown under similar stress conditions.

[0351] A plant with “increased stress tolerance” can exhibit increased tolerance to one or more different stress conditions.

[0352] “Stress tolerance activity” of a polypeptide can indicate that over-expression of the polypeptide in a transgenic plant can confer increased stress tolerance to the transgenic plant relative to a reference or control plant.

[0353] Increased biomass can be measured, for example, as an increase in plant height, plant total leaf area, plant fresh weight, plant dry weight or plant seed yield, as compared with control plants.

[0354] The ability to increase the biomass or size of a plant can have several important commercial applications. Crop species may be generated that can produce larger cultivars, generating higher yield in, for example, plants in which the vegetative portion of the plant can be useful as food, biofuel or both.

[0355] Increased leaf size can be produced by the methods and composition of the disclosure. Increasing leaf biomass can be used to increase production of plant-derived pharmaceutical or industrial products. An increase in total plant photosynthesis can be achieved by, for example, increasing leaf area of the plant. Additional photosynthetic capacity may be used to increase the yield derived from particular plant tissue, including the leaves, roots, fruits or seed, or permit the growth of a plant under decreased light intensity or under high light intensity.

[0356] Modification of the biomass of a tissue, such as root tissue, may be useful to improve a plant's ability to grow under harsh environmental conditions, including drought or nutrient deprivation. Larger roots may better reach water or nutrients or take up water or nutrients.

[0357] The ability to provide larger varieties can be highly desirable, for example, for some ornamental plants. For many plants, including fruit-bearing trees, trees that are used for lumber production, or trees and shrubs that serve as view or wind screens, increased stature can provide improved benefits in the forms of greater yield or improved screening.

Herbicide Resistance in Plants

[0358] An “herbicide resistance protein” or a protein resulting from expression of an “herbicide resistance-encoding nucleic acid molecule” can include proteins that can confer upon a cell the ability to tolerate a higher concentration of an herbicide, for example, compared with cells that do not express the protein. An herbicide resistance protein or a protein resulting from expression of a herbicide resistance-encoding nucleic acid molecule can include proteins that can confer upon a cell the ability to tolerate a concentration of a herbicide for a longer period of time than cells that do not express the protein. Herbicide resistance traits may be introduced into plants by, for example, genes coding for resistance to herbicides. Genes coding for resistance to herbicides include, for example, genes that act to inhibit the action of acetolactate synthase (ALS), such as the sulfonylurea-type herbicides, genes that act to inhibit the action of glutamine synthase, such as phosphinothricin or

basta (e.g., the bar gene), glyphosate (e.g., the EPSP synthase gene), HPPD inhibitors (e.g., the HPPD gene).

[0359] Herbicide resistance proteins can include the following: a 4-hydroxyphenylpyruvate dioxygenase (HPPD), a sulfonylurea-tolerant acetolactate synthase (ALS), an imidazolinone-tolerant acetolactate synthase (ALS), a glyphosate-tolerant 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), a glyphosate-tolerant glyphosate oxidoreductase (GOX), a glyphosate N-acetyltransferase (GAT), a phosphinothricin acetyl transferase (PAT), a protoporphyrinogen oxidase (PROTOX), an auxin enzyme or receptor, a P450 polypeptide and an acetyl coenzyme A carboxylase (AC-Case). Non-limiting examples of genes useful for conferring herbicide resistance in plants can include genes that encode the above proteins.

[0360] As used herein, “Hydroxyphenylpyruvate dioxygenase” and “HPPD”, “4-hydroxy phenyl pyruvate (or pyruvic acid) dioxygenase (4-HPPD)” and “p-hydroxy phenyl pyruvate (or pyruvic acid) dioxygenase (p-OHPP)” can be synonymous and can refer to a non-heme iron-dependent oxygenase that catalyzes the conversion of 4-hydroxyphenylpyruvate to homogentisate. In organisms that degrade tyrosine, the reaction catalyzed by HPPD can be the second step in the pathway. In plants, formation of homogentisate can be necessary for the synthesis of plastoquinone, which can serve as a redox cofactor, and tocopherol. A polynucleotide molecule encoding hydroxyphenylpyruvate dioxygenase (HPPD) can provide tolerance to HPPD inhibitors.

[0361] As used herein, an “HPPD inhibitor” can comprise any compound or combinations of compounds which can decrease the ability of HPPD to catalyze the conversion of 4-hydroxyphenylpyruvate to homogentisate. In specific embodiments, the HPPD inhibitor can comprise an herbicidal inhibitor of HPPD. Non-limiting examples of HPPD inhibitors include, triketones (such as, mesotrione, sulcotrione, topramezone, and tembotrione); isoxazoles (such as, pyrasulfotole and isoxaflutole); pyrazoles (such as, benzo-fenap, pyrazoxyfen, and pyrazolynate); and benzobicyclon. Agriculturally acceptable salts of the various inhibitors can include salts (e.g., the cations or anions) for the formation of salts for agricultural or horticultural use.

[0362] An “ALS inhibitor-tolerant polypeptide” can comprise any polypeptide which when expressed in a plant can confer tolerance to at least one ALS inhibitor. ALS inhibitors include, for example, sulfonylurea, imidazolinone, triazolopyrimidines, pyrimidinyoxy(thio)benzoates, and/or sulfonylaminocarbonyltriazolinone herbicides. ALS mutations can fall into different classes with regard to tolerance to, for example, sulfonylureas, imidazolinones, triazolopyrimidines, and pyrimidinyl(thio)benzoates. ALS mutations can include mutations having one or more of the following characteristics: (1) broad tolerance to all four of these groups (e.g., sulfonylureas, imidazolinones, triazolopyrimidines, and pyrimidinyl(thio)benzoates); (2) tolerance to imidazolinones and pyrimidinyl(thio)benzoates; (3) tolerance to sulfonylureas and triazolopyrimidines; and (4) tolerance to sulfonylureas and imidazolinones.

[0363] Polynucleotide molecules encoding proteins involved in herbicide resistance can include a polynucleotide molecule encoding 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) for example, for imparting glyphosate tolerance.

[0364] Glyphosate tolerance can also be obtained by expression of polynucleotide molecules encoding a glyphosate oxidoreductase (GOX) or a glyphosate-N-acetyl transferase (GAT).

[0365] Polynucleotides encoding an exogenous phosphinothricin acetyltransferase can be used for herbicide resistance. Plants containing an exogenous phosphinothricin acetyltransferase can exhibit improved tolerance to glufosinate herbicides, which can inhibit, for example, the enzyme glutamine synthase.

[0366] Polynucleotides conferring altered protoporphyrinogen oxidase (protox) activity can be used for herbicide resistance. Plants containing such polynucleotides can exhibit improved tolerance to any of a variety of herbicides which can target, for example, the protox enzyme (also referred to as “protox inhibitors”).

[0367] Dicamba monooxygenase can be used for providing dicamba tolerance.

[0368] A polynucleotide molecule encoding AAD12 or encoding AAD1 can be used for providing resistance to, for example, auxin herbicides.

[0369] A P450 sequence can be used for conferring herbicide resistance. A P450 sequence can provide tolerance to HPPD inhibitors by, for example, metabolism of the herbicide. Such sequences include, but are not limited to, the NSF1 gene.

Pest Resistance in Plants by Gene Silencing

[0370] A “plant pest” can mean any living stage of an entity that can directly or indirectly injure, cause damage to, or cause disease in any plant or plant product. A plant pest can include a protozoan, a nonhuman animal, a parasitic plant, a bacterium, a fungus, a virus, a viroid, an infectious agent, a pathogen, or any article similar to or allied thereof.

[0371] Double-stranded RNA (dsRNA) can be used to provide resistance to plant pests.

[0372] Plant pest invertebrates can include, but are not limited to, pest nematodes, pest mollusks (slugs and snails), and pest insects. Plant pathogens can include fungi and nematodes.

[0373] The plant pathogen can be a eukaryotic plant pathogen. This includes for example, a fungal pathogen, such as a phytopathogenic fungus.

[0374] Non-limiting examples of fungal plant pathogens include, e.g., the fungi that cause powdery mildew, rust, leaf spot and blight, damping-off, root rot, crown rot, cotton boll rot, stem canker, twig canker, vascular wilt, smut, or mold, including, but not limited to, *Fusarium* spp., *Phakospora* spp., *Rhizoctonia* spp., *Aspergillus* spp., *Gibberella* spp., *Pyricularia* spp., *Alternaria* spp., and *Phytophthora* spp. Specific examples of fungal plant pathogens include *Phakospora pachirhizi* (Asian soy rust), *Puccinia sorghi* (corn common rust), *Puccinia polysora* (corn Southern rust), *Fusarium oxysporum* and other *Fusarium* spp., *Alternaria* spp., *Penicillium* spp., *Pythium aphanidermatum* and other *Pythium* spp., *Rhizoctonia solani*, *Exserohilum turcicum* (Northern corn leaf blight), *Bipolaris maydis* (Southern corn leaf blight), *Ustilago maydis* (corn smut), *Fusarium graminearum* (*Gibberella zeae*), *Fusarium verticilliodes* (*Gibberella moniliformis*), *F. proliferatum* (*G. fujikuroi* var. *intermedia*), *F. subglutinans* (*G. subglutinans*), *Diplodia maydis*, *Sporisorium holci-sorghii*, *Colletotrichum gramini-*

cola, *Setosphaeria turcica*, *Aureobasidium zeae*, *Phytophthora infestans*, *Phytophthora sojae*, *Sclerotinia sclerotiorum*, and fungal species.

[0375] Non-limiting examples of invertebrate pests can include cyst nematodes *Heterodera* spp. such as soybean cyst nematode *Heterodera glycines*, root knot nematodes *Meloidogyne* spp., lance nematodes *Hoplolaimus* spp., stunt nematodes *Tylenchorhynchus* spp., spiral nematodes *Helicotylenchus* spp., lesion nematodes *Pratylenchus* spp., ring nematodes *Criconea* spp., foliar nematodes *Aphelenchus* spp. or *Aphelenchoides* spp., corn rootworms, *Lygus* spp., aphids and similar sap-sucking insects such as *phylloxera* (*Daktulosphaira vitifoliae*), corn borers, cutworms, armyworms, leafhoppers, Japanese beetles, grasshoppers, and other pest coleopterans, dipterans, and lepidopterans. Additional examples of invertebrate pests can include pests that can infest the root systems of crop plants, e.g., northern corn rootworm (*Diabrotica barberi*), southern corn rootworm (*Diabrotica undecimpunctata*), Western corn rootworm (*Diabrotica virgifera*), corn root aphid (*Anuraphis maidiradicis*), black cutworm (*Agrotis ipsilon*), glassy cutworm (*Crymodes devastator*), dingy cutworm (*Feltia ducens*), claybacked cutworm (*Agrotis gladiaria*), wireworm (*Melanotus* spp., *Aeolus mellillus*), wheat wireworm (*Aeolus mancus*), sand wireworm (*Horistonotus uhlerii*), maize billbug (*Sphenophorus maidis*), timothy billbug (*Sphenophorus zeae*), bluegrass billbug (*Sphenophorus parvulus*), southern corn billbug (*Sphenophorus callosus*), white grubs (*Phyllophaga* spp.), seedcorn maggot (*Delia platura*), grape colaspis (*Colaspis brunnea*), seedcorn beetle (*Stenolophus lecontei*), and slender seedcorn beetle (*Clivinia impressifrons*), and parasitic nematodes.

[0376] A target gene of interest (e.g., for gene silencing) may include any coding or non-coding sequence from any species (including, but not limited to, eukaryotes such as fungi; plants, including monocots and dicots, such as crop plants, ornamental plants, and non-domesticated or wild plants; invertebrates such as arthropods, annelids, nematodes, and mollusks; and vertebrates such as amphibians, fish, birds, and mammals). Non-limiting examples of a non-coding sequence (e.g., that can be expressed by a gene expression element such as a regulatory sequence) include, but not limited to, 5' untranslated regions, promoters, enhancers, or other non-coding transcriptional regions, 3' untranslated regions, terminators, introns, microRNAs, microRNA precursor DNA sequences, small interfering RNAs, RNA components of ribosomes or ribozymes, small nucleolar RNAs, and other non-coding RNAs. Non-limiting examples of a gene of interest further include, but are not limited to, translatable (coding) sequence, such as genes encoding transcription factors and genes encoding enzymes involved in the biosynthesis or catabolism of molecules of interest (such as amino acids, fatty acids and other lipids, sugars and other carbohydrates, biological polymers, and secondary metabolites including alkaloids, terpenoids, polyketides, non-ribosomal peptides, and secondary metabolites of mixed biosynthetic origin).

[0377] The target gene (e.g., for gene silencing) may be an essential gene of the plant pest or plant pathogen. Essential genes can include genes that may be required for development of the pest or pathogen to a fertile reproductive adult. Essential genes can include genes that, when silenced or suppressed, can result in the death of the organism (e.g., as an adult or at any developmental stage, including gametes)

or in the organism's inability to successfully reproduce (e.g., sterility in a male or female parent or lethality to the zygote, embryo, or larva). Non-limiting examples of nematode essential genes include major sperm protein, RNA polymerase II, and chitin synthase. Additional soybean cyst nematode essential genes are provided in U.S. Patent Publication US20070271630, incorporated by reference herein. The gene can be a *Drosophila* essential gene. The gene can be a fungal essential gene.

[0378] Target genes (e.g., from pests) can include invertebrate genes for major sperm protein, alpha tubulin, beta tubulin, vacuolar ATPase, glyceraldehyde-3-phosphate dehydrogenase, PvNA polymerase TT, chitin synthase, cytochromes, miRNAs, miRNA precursor molecules and miRNA promoters. Target genes (e.g., from pathogens) can include genes for miRNAs, miRNA precursor molecules, fungal tubulin, fungal vacuolar ATPase, fungal chitin synthase, fungal MAP kinases, fungal Pad Tyr/Thr phosphatase, enzymes involved in nutrient transport (e.g., amino acid transporters or sugar transporters), enzymes involved in fungal cell wall biosynthesis, cutinases, melanin biosynthetic enzymes, polygalacturonases, pectinases, pectin lyases, cellulases, proteases, genes that interact with plant avirulence genes, and genes involved in invasion and replication of the pathogen in the infected plant.

[0379] Plants may be transformed (e.g., in the nucleus, an organelle, or both) with an expression cassette encoding, for example, a dsRNA, a siRNA or a miRNA. The dsRNA, siRNA, or miRNA can suppress (e.g., expression of) at least one (e.g., at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, or at least 10) target gene present in a plant pest. The dsRNA, siRNA, or miRNA can suppress, for example, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 or more target genes of a plant pest. Suppression of a target gene present in the plant pest can provide complete or nearly complete protection from the plant pest. "Complete protection" can mean that no (e.g., substantial) damage can be caused to the plant by the plant pest.

[0380] The dsRNA, the siRNA or the miRNA may be designed for suppression of a gene selected from the group consisting of: proteasome A-type subunit peptide (Pas-4), ACT, SHR, EPIC2B and PnPMAL.

[0381] SEQ ID NO:114 corresponds to an open reading frame encoding an *Heterodera glycines* (SCN) specific proteasome A-type subunit peptide that can be referred to herein as Pas-4. SEQ ID NO: 115 corresponds to nucleotides 552-699 of SEQ ID NO: 114. SEQ ID NO: 115 or SEQ ID NO: 114 can be useful for dsRNA-mediated suppression of Pas-4. ACT can encode β -actin, which can be an essential cytoskeletal protein. SHR can encode Shrub (also known as Vps32 or Snf7), which can be an essential subunit of a protein complex involved in membrane remodeling for vesicle transport. EPIC2B can encode a *Phytophthora infestans* protein that can interact with and/or inhibit a novel papain-like extracellular Cys protease, for example, *Phytophthora* Inhibited Protease 1. The PnPMA gene from *Phytophthora parasitica* can encode a plasma membrane H⁺ ATPase.

Resistance to Plant Pests

[0382] Resistance to pests in plants can be achieved by, for example, transgenic control. In-plant transgenic control of, for example, insect pests, can be achieved through, for example, plant expression of crystal (Cry) delta endotoxin

genes and/or Vegetative Insecticidal Proteins (VIP) such as from *Bacillus thuringiensis*. Non-limiting examples of Cry toxins include, for example, the 60 main groups of "Cry" toxins (e.g., Cry1-Cry59) and VIP toxins. Cry toxins can include subgroups of Cry toxins, for example, Cry 1a.

[0383] An expression cassette for use in transformation (e.g., into an organelle) may be constructed using, for example, a Cry sequence. The Cry sequence can include, for example, the wild-type (e.g., native) nucleic acid sequence encoding at least one protein selected from the group consisting of: Cry1Ac, Cyt1Aa, Cry1Ab, Cry2Aa, Cry1I, Cry1C, Cry1D, Cry1E, Cry1Be, Cry1Fa and Vip3A. The Cry sequence can include, for example, a modified (e.g., truncated or fusion) nucleic acid sequence encoding at least one protein selected from the group consisting of: Cry1Ac, Cyt1Aa, Cry1Ab, Cry2Aa, Cry1I, Cry1C, Cry1D, Cry1E, Cry1Be, Cry1Fa and Vip3A. A modified such as a truncated nucleic acid sequence can encode a modified such as a truncated protein fragment that can retain insecticidal activity. The nucleic acid sequence encoding the full-length, or modified (e.g., truncated) protein may be codon-optimized for the organelle of interest. The Cry protein can be a Cyt1Aa protein (e.g., from *Bacillus thuringiensis* serovar israelensis; Gene ID: 5759908; SEQ ID NO:111).

[0384] Accessory proteins, for example, for a Cry protein, can be introduced into a cell (e.g., into an organelle). An accessory protein can, for example, increase expression, stability, and/or function of, for example, a Cry protein. Non-limiting examples of accessory proteins include 20 kDa accessory proteins (e.g., from *Bacillus thuringiensis* serovar israelensis) and 19 kDa accessory proteins (e.g., from *Bacillus thuringiensis* serovar israelensis). The accessory protein can be the 20 kDa accessory protein from *Bacillus thuringiensis* serovar israelensis (pBt024; SEQ ID NO:112). The accessory protein can be the 19 kDa accessory protein from *Bacillus thuringiensis* serovar israelensis, (pBt022; SEQ ID NO:113). Accessory proteins can be included in an expression cassette as a polycistronic unit. Accessory proteins can be expressed from separate expression cassettes.

[0385] Polynucleotides that encode proteins useful in conferring insect resistance to a plant may be included in an expression cassette as a polycistronic unit, or may be expressed from separate expression cassettes. In some embodiments, these polynucleotides can encode the following: (a) the Cyt1Aa protein from *Bacillus thuringiensis* serovar israelensis (Gene ID: 5759908; SEQ ID NO:111); (b) the 20 kDa accessory protein from *Bacillus thuringiensis* serovar israelensis (pBt024; SEQ ID NO:112); and (c) the 19 kDa accessory protein from *Bacillus thuringiensis* serovar israelensis, (pBt022; SEQ ID NO:113).

Genome Modification

[0386] The disclosure provides compositions and methods that can be used for, for example, genome modification of a target sequence in the genome (e.g., a plastid or a mitochondrial genome) of an organism or cell (e.g., a plant or plant cell), for selecting the modified organism or cell, for gene editing, and for inserting a donor polynucleotide into the genome of an organism or cell. The methods can employ a polynucleotide guided polypeptide system; e.g., a guide polynucleotide/Cas protein system. The Cas protein can be guided by the guide polynucleotide to recognize a target polynucleic acid. The Cas protein can introduce a single strand or double strand break at a specific target site into the

genome of a cell. The guide polynucleotide/Cas polypeptide system can provide for an effective system for modifying target sites within the genome of a plant, plant cell or seed.

[0387] A variety of methods can be employed to further modify a target site to introduce a donor polynucleotide of interest. The nucleotide sequence to be edited (e.g., the nucleotide sequence of interest) can be located within or outside a target site that is recognized by a polynucleotide guided polypeptide.

[0388] Further provided are methods and compositions employing a polynucleotide guided polypeptide system for modification of multiple target sites within the genome of an organelle. Modification of multiple target sites within the genome of an organelle can facilitate the creation of homoplastic transformation events.

Polynucleotide Guided Polypeptide Systems

[0389] A polynucleotide-guided polypeptide can be a polypeptide that can bind to a target nucleic acid. A polynucleotide-guided polypeptide can be a nuclease. A polynucleotide-guided polypeptide can be an endonuclease. A polynucleotide-guided polypeptide can be a Cas protein. A polynucleotide-guided polypeptide can be an Argonaut protein. A polynucleotide guided polypeptide can form a complex with a guide polynucleotide. A polynucleotide guided polypeptide can be directed to a target nucleic acid by a guide polynucleotide. A polynucleotide guided polypeptide can complex with a guide polynucleotide to recognize a target nucleic acid. A polynucleotide guided polypeptide can introduce a single strand or double strand break at a specific target site (e.g., the genome of a cell).

[0390] a. CRISPR Loci

[0391] CRISPR loci (Clustered Regularly Interspaced Short Palindromic Repeats) (also known as SPIDRs-SPacer Interspersed Direct Repeats) can constitute a family of DNA loci. CRISPR loci can consist of short and highly conserved DNA repeats (e.g., 24 to 40 bp, repeated from 1 to 140 times-also referred to as CRISPR-repeats). CRISPR DNA repeats can be partially palindromic. The repeated sequences (e.g., usually specific to a species) can be interspaced by variable sequences of constant length (e.g., 20 to 58 by depending on the CRISPR locus).

[0392] CRISPR loci can occur in, for example, *E. coli*, *Haloferax mediterranei*, *Streptococcus pyogenes*, *Anabaena*, and *Mycobacterium tuberculosis*. The CRISPR loci can comprise short regularly spaced repeats (SRSRs). The repeats can be short elements that can occur in clusters. The repeats can be regularly spaced by variable sequences of constant length.

[0393] CRISPR systems can belong to different classes, with different repeat patterns, sets of genes, and species ranges. The number of Cas genes at a given CRISPR locus can vary between species.

[0394] b. Cas Protein

[0395] A Cas protein can be a protein of a CRISPR/Cas system. A Cas protein can be a Class 1 or a Class 2 Cas protein. A Cas protein can be a Type I, Type II, Type III, Type IV, Type V, or Type VI Cas protein.

[0396] "Cas gene" can refer to a gene that encodes a Cas protein. The terms Cas protein and Cas polypeptide can be used interchangeably herein. Cas gene can be coupled, associated or close to or in the vicinity of flanking CRISPR loci. The terms "Cas gene", "CRISPR-associated (Cas) gene" can be used interchangeably herein.

[0397] A Cas protein can bind to a target nucleic acid. A Cas protein can be a Cas nuclease. A Cas protein can be a Cas endonuclease. A Cas protein can complex with a guide polynucleotide. A Cas protein can be directed to a target nucleic acid by a guide polynucleotide. A Cas protein can complex with a guide polynucleotide to recognize a target nucleic acid. A Cas protein can introduce a single strand or double strand break at a target nucleic acid sequence (e.g., DNA or RNA). A Cas protein can be enabled by the guide polynucleotide to recognize and introduce a single strand or double strand break at a specific target site into the genome of a cell.

[0398] A Cas protein can comprise one or more domains. Non-limiting examples of domains include, guide nucleic acid recognition and/or binding domain, nuclease domains (e.g., DNase or RNase domains, RuvC, HNH), DNA binding domain, RNA binding domain, helicase domains, protein-protein interaction domains, and dimerization domains. A guide nucleic acid recognition and/or binding domain can interact with a guide nucleic acid. A nuclease domain can comprise catalytic activity for nucleic acid cleavage. A nuclease domain can lack catalytic activity to prevent nucleic acid cleavage. A Cas protein can be a chimeric Cas protein that is fused to other proteins or polypeptides. A Cas protein can be a chimera of various Cas proteins, for example, comprising domains from different Cas proteins (e.g., homologues).

[0399] Non-limiting examples of Cas proteins include c2c1, C2c2, c2c3, Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5e (CasD), CasH, Cas6e, Cas6f, Cas7, Cas8a, Cas8a1, Cas8a2, Cas8b, Cas8c, Cas9 (Csn1 or Csx12), Cas10, Cas10d, Cas10, Cas10d, CasF, CasG, CasH, Cpf1, Csy1, Csy2, Csy3, Cse1 (CasA), Cse2 (CasB), Cse3 (CasE), Cse4 (CasC), Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, and Cul966, and homologs or modified versions thereof.

[0400] A Cas protein may be from any suitable organism. Non-limiting examples include *Streptococcus pyogenes*, *Streptococcus thermophilus*, *Streptococcus* sp., *Staphylococcus aureus*, *Nocardiosis dassonvillei*, *Streptomyces pristinae spiralis*, *Streptomyces viridochromogenes*, *Streptomyces viridochromogenes*, *Streptosporangium roseum*, *Streptosporangium roseum*, *Alicyclobacillus acidocaldarius*, *Bacillus pseudomyoides*, *Bacillus selenitireducens*, *Exiguobacterium sibiricum*, *Lactobacillus delbrueckii*, *Lactobacillus salivarius*, *Microscilla marina*, *Burkholderiales bacterium*, *Polaromonas naphthalenivorans*, *Polaromonas* sp., *Crocospaera watsonii*, *Cyanothece* sp., *Microcystis aeruginosa*, *Pseudomonas aeruginosa*, *Synechococcus* sp., *Acetohalobium arabaticum*, *Ammonifex degenii*, *Caldicellulosiruptor beccsii*, *Candidatus Desulfurudis*, *Clostridium botulinum*, *Clostridium difficile*, *Finexgoldia magna*, *Natronaerobius thermophilus*, *Pelotomaculum thermopropionicum*, *Acidithiobacillus caldus*, *Acidithiobacillus ferrooxidans*, *Allochromatium vinosum*, *Marinobacter* sp., *Nitrosococcus halophilus*, *Nitrosococcus watsoni*, *Pseudalteromonas haloplanktis*, *Ktedonobacter racemifer*, *Methanohalobium evestigatum*, *Anabaena variabilis*, *Nodularia spumigena*, *Nostoc* sp., *Arthrospira maxima*, *Arthrospira platensis*, *Arthrospira* sp., *Lyngbya* sp., *Microcoleus chthonoplastes*, *Oscillatoria* sp., *Petrotoga mobilis*, *Thermosiphon africanus*, *Acaryochloris marina*, *Leptotrichia shahii*, and

Francisella novicida. In some aspects, the organism can be *Streptococcus pyogenes* (*S. pyogenes*).

[0401] A Cas protein as used herein can be a wildtype or a modified form of a Cas protein. A Cas protein can be an active variant, inactive variant, or fragment of a wild type or modified Cas protein. A Cas protein can comprise an amino acid change such as a deletion, insertion, substitution, variant, mutation, fusion, chimera, or any combination thereof relative to a wild-type version of the Cas protein. A Cas protein can be a polypeptide with at least about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity or sequence similarity to a wild type exemplary Cas protein (e.g., Cas9 from *S. pyogenes*). A Cas protein can be a polypeptide with at most about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% sequence identity and/or sequence similarity to a wild type exemplary Cas protein. Variants or fragments can comprise at least about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity or sequence similarity to a wild type or modified Cas protein or a portion thereof. Variants or fragments can be targeted to a nucleic acid locus in complex with a guide nucleic acid while lacking nucleic acid cleavage activity.

[0402] A Cas protein can comprise one or more nuclease domains, such as DNase domains. For example, a Cas9 protein can comprise a RuvC-like nuclease domain and/or an HNH-like nuclease domain. The RuvC and HNH domains can each cut a different strand of double-stranded DNA to make a double-stranded break in the DNA. A Cas protein can comprise only one nuclease domain (e.g., Cpf1 comprises RuvC domain but lacks HNH domain).

[0403] A Cas protein can comprise an amino acid sequence having at least about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity or sequence similarity to a nuclease domain (e.g., RuvC domain, HNH domain) of a wild-type Cas protein.

[0404] A Cas protein can be modified to optimize activity e.g., cleavage, regulation of gene expression. A Cas protein can be modified to increase or decrease nucleic acid binding affinity, nucleic acid binding specificity, and/or enzymatic activity. Cas proteins can also be modified to change any other activity or property of the protein, such as stability. For example, one or more nuclease domains of the Cas protein can be modified, deleted, or inactivated, or a Cas protein can be truncated to remove domains that are not essential for the function of the protein or to optimize (e.g., enhance or reduce) the activity of the Cas protein.

[0405] A Cas protein can be a fusion protein. For example, a Cas protein can be fused to a cleavage domain, an epigenetic modification domain, a transcriptional activation domain, or a transcriptional repressor domain. A Cas protein can also be fused to a heterologous polypeptide providing increased or decreased stability. The fused domain or heterologous polypeptide can be located at the N-terminus, the C-terminus, or internally within the Cas protein.

[0406] A Cas protein can comprise a heterologous polypeptide for ease of tracking or purification, such as a fluorescent protein, a purification tag, or an epitope tag. Examples of fluorescent proteins include green fluorescent proteins (e.g., GFP, GFP-2, tagGFP, turboGFP, eGFP, Emerald, Azami Green, Monomeric Azami Green, CopGFP, AceGFP, ZsGreen1), yellow fluorescent proteins (e.g., YFP,

eYFP, Citrine, Venus, YFPet, PhiYFP, ZsYellow1), blue fluorescent proteins (e.g. eBFP, eBFP2, Azurite, mKalamal, GFPuv, Sapphire, T-sapphire), cyan fluorescent proteins (e.g. eCFP, Cerulean, CyPet, AmCyan1, Midoriishi-Cyan), red fluorescent proteins (mKate, mKate2, mPlum, DsRed monomer, mCherry, mRFP1, DsRed-Express, DsRed2, DsRed-Monomer, HcRed-Tandem, HcRed1, AsRed2, eqFP611, mRaspberry, mStrawberry, Jred), orange fluorescent proteins (mOrange, mKO, Kusabira-Orange, Monomeric Kusabira-Orange, mTangerine, tdTomato), and any other suitable fluorescent protein. Examples of tags include glutathione-S-transferase (GST), chitin binding protein (CBP), maltose binding protein, thioredoxin (TRX), poly (NANP), tandem affinity purification (TAP) tag, myc, AcV5, AU1, AUS, E, ECS, E2, FLAG, hemagglutinin (HA), nus, Softag 1, Softag 3, Strep, SBP, Glu-Glu, HSV, KT3, S, SI, T7, V5, VSV-G, histidine (His), biotin carboxyl carrier protein (BCCP), and calmodulin.

[0407] A Cas protein can be provided in any form. For example, a Cas protein can be provided in the form of a protein, such as a Cas protein alone or complexed with a guide nucleic acid. A Cas protein can be provided in the form of a nucleic acid encoding the Cas protein, such as an RNA (e.g., messenger RNA (mRNA)) or DNA.

[0408] The nucleic acid encoding the Cas protein can be codon optimized for efficient translation into protein in a particular cell, organelles, or organism.

[0409] Nucleic acids encoding Cas proteins can be stably integrated in the genome of an organelle or a cell. Nucleic acids encoding Cas proteins can be operably linked to a promoter active in the cell. Nucleic acids encoding Cas proteins can be operably linked to a promoter in an expression construct. Expression constructs can include any nucleic acid constructs that can direct expression of a gene or other nucleic acid sequence of interest (e.g., a Cas gene). Expression constructs can include any nucleic acid constructs that can transfer such a nucleic acid sequence of interest to a target cell (e.g., into an organelle).

[0410] In some aspects, a Cas protein can be a Class 2 Cas protein. In some aspects, a Cas protein can be a type II Cas protein. In some aspects, the Cas protein can be a Cas9 protein, a modified version of a Cas9 protein, or derived from a Cas9 protein.

[0411] Cas9 can refer to a polypeptide with at least about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% sequence identity and/or sequence similarity to a wild type exemplary Cas9 polypeptide (e.g., Cas9 from *S. pyogenes*). Cas9 can refer to a polypeptide with at most about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% sequence identity and/or sequence similarity to a wild type exemplary Cas9 polypeptide (e.g., from *S. pyogenes*). Cas9 can refer to the wildtype or a modified form of the Cas9 protein that can comprise an amino acid change such as a deletion, insertion, substitution, variant, mutation, fusion, chimera, or any combination thereof.

[0412] In one embodiment, the polynucleotide guided polypeptide gene can be a Cas9 protein, such as but not limited to, Cas9 sequences listed in SEQ ID NOs: 462, 474, 489, 494, 499, 505, and 518 of WO2007/025097 and incorporated herein by reference. The Cas9 protein can unwind the DNA duplex in close proximity of the genomic target site. The Cas9 protein can cleave for example both DNA strands upon recognition of a target sequence by a guide polynucleic acid. In some aspects, the Cas9 endonuclease

can cleave only if the correct protospacer-adjacent motif (PAM) is approximately oriented at the 3' end of the target sequence. Mutagenesis of *Streptococcus pyogenes* Cas9 catalytic domains can produce “nicking” enzymes (Cas9n) that can induce single-strand nicks rather than double-strand breaks.

[0413] In another embodiment, the polynucleotide guided polypeptide coding sequence can be modified to use codons preferred by the target organism, e.g., a plant, maize or soybean codon-optimized sequence encoding a Cas (e.g., Cas9) protein. In another embodiment, the sequence that encodes a polynucleotide guided polypeptide can be operably linked to one or more sequences encoding nuclear localization signals; e.g., to a SV40 nuclear targeting signal upstream of the Cas protein coding region and a bipartite VirD2 nuclear localization signal downstream of the Cas protein coding region.

[0414] In another embodiment, the polynucleotide guided polypeptide may be an Argonaute protein such as *Natronobacterium gregoryi* Argonaute (“NgAgo”). The Argonaute protein can be a DNA-guided endonuclease. Argonaute proteins can bind a guide DNA such as a 5'-phosphorylated single-stranded guide DNA (gDNA) of for example, 24 nucleotides. Argonaute proteins can create site-specific target nucleic acid (e.g., DNA) breaks (e.g., double-stranded breaks) when loaded with the gDNA. The Argonaute protein—gDNA system may not require a protospacer-adjacent motif (PAM) for recognition of a target nucleic acid.

[0415] In some aspects, the polynucleotide guided polypeptide can be a dead Cas protein. A Cas protein can be a dead Cas protein. A dead Cas protein can be a protein that lacks nucleic acid cleavage activity.

[0416] A Cas protein can comprise a modified form of a wild type Cas protein. The modified form of the wild type Cas protein can comprise an amino acid change (e.g., deletion, insertion, or substitution) that reduces the nucleic acid-cleaving activity of the Cas protein. For example, the modified form of the Cas protein can have less than less than 90%, less than 80%, less than 70%, less than 60%, less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, less than 5%, or less than 1% of the nucleic acid-cleaving activity of the wild-type Cas protein (e.g., Cas9 from *S. pyogenes*). The modified form of Cas protein can have no substantial nucleic acid-cleaving activity. When a Cas protein is a modified form that has no substantial nucleic acid-cleaving activity, it can be referred to as enzymatically inactive and/or “dead” (abbreviated by “d”). A dead Cas protein (e.g., dCas, dCas9) can bind to a target polynucleotide but may not cleave the target polynucleotide. In some aspects, a dead Cas protein can be a dead Cas9 protein.

[0417] Enzymatically inactive can refer to a polypeptide that can bind to a nucleic acid sequence in a polynucleotide in a sequence-specific manner, but may not cleave a target polynucleotide. An enzymatically inactive site-directed polypeptide can comprise an enzymatically inactive domain (e.g. nuclease domain). Enzymatically inactive can refer to no activity. Enzymatically inactive can refer to substantially no activity. Enzymatically inactive can refer to essentially no activity. Enzymatically inactive can refer to an activity less than 1%, less than 2%, less than 3%, less than 4%, less than 5%, less than 6%, less than 7%, less than 8%, less than 9%, or less than 10% activity compared to a wild-type exemplary activity (e.g., nucleic acid cleaving activity, wild-type Cas9 activity).

[0418] One or a plurality of the nuclease domains (e.g., RuvC, HNH) of a Cas protein can be deleted or mutated so that they are no longer functional or comprise reduced nuclease activity. For example, in a Cas protein comprising at least two nuclease domains (e.g., Cas9), if one of the nuclease domains is deleted or mutated, the resulting Cas protein, known as a nickase, can generate a single-strand break at a CRISPR RNA (crRNA) recognition sequence within a double-stranded DNA but not a double-strand break. Such a nickase can cleave the complementary strand or the non-complementary strand, but may not cleave both. If all of the nuclease domains of a Cas protein (e.g., both RuvC and HNH nuclease domains in a Cas9 protein; RuvC nuclease domain in a Cpf1 protein) are deleted or mutated, the resulting Cas protein can have a reduced or no ability to cleave both strands of a double-stranded DNA. An example of a mutation that can convert a Cas9 protein into a nickase can be a D10A (aspartate to alanine at position 10 of Cas9) mutation in the RuvC domain of Cas9 from *S. pyogenes*. H939A (histidine to alanine at amino acid position 839) or H840A (histidine to alanine at amino acid position 840) in the HNH domain of Cas9 from *S. pyogenes* can convert the Cas9 into a nickase. An example of a mutation that can convert a Cas9 protein into a dead Cas9 is a D10A (aspartate to alanine at position 10 of Cas9) mutation in the RuvC domain and H939A (histidine to alanine at amino acid position 839) or H840A (histidine to alanine at amino acid position 840) in the HNH domain of Cas9 from *S. pyogenes*.

[0419] A dead Cas protein can comprise one or more mutations relative to a wild-type version of the protein. The mutation can result in less than 90%, less than 80%, less than 70%, less than 60%, less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, less than 5%, or less than 1% of the nucleic acid-cleaving activity in one or more of the plurality of nucleic acid-cleaving domains of the wild-type Cas protein. The mutation can result in one or more of the plurality of nucleic acid-cleaving domains retaining the ability to cleave the complementary strand of the target nucleic acid but reducing its ability to cleave the non-complementary strand of the target nucleic acid. The mutation can result in one or more of the plurality of nucleic acid-cleaving domains retaining the ability to cleave the non-complementary strand of the target nucleic acid but reducing its ability to cleave the complementary strand of the target nucleic acid. The mutation can result in one or more of the plurality of nucleic acid-cleaving domains to lack the ability to cleave the complementary strand and the non-complementary strand of the target nucleic acid. The residues to be mutated in a nuclease domain can correspond to one or more catalytic residues of the nuclease. For example, residues in the wild type exemplary *S. pyogenes* Cas9 polypeptide such as Asp10, His840, Asn854 and Asn856 can be mutated to inactivate one or more of the plurality of nucleic acid-cleaving domains (e.g., nuclease domains). The residues to be mutated in a nuclease domain of a Cas protein can correspond to residues Asp10, His840, Asn854 and Asn856 in the wild type *S. pyogenes* Cas9 polypeptide, for example, as determined by sequence and/or structural alignment.

[0420] As non-limiting examples, residues D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or A987 (or the corresponding mutations of any of the Cas proteins) can be mutated. For example, e.g., D10A, G12A, G17A, E762A, H840A, N854A, N863A, H982A,

H983A, A984A, and/or D986A. Mutations other than alanine substitutions can be suitable.

[0421] A D10A mutation can be combined with one or more of H840A, N854A, or N856A mutations to produce a polynucleotide guided polypeptide (e.g., Cas9 protein) substantially lacking DNA cleavage activity (e.g., a dead Cas9 protein).

[0422] In another embodiment, the polynucleotide guided polypeptide can be a polypeptide moiety (e.g., a chimeric polypeptide) that can form a programmable nucleoprotein molecular complex with a specificity conferring nucleic acid (SCNA). The programmable nucleoprotein molecular complex can assemble in-vivo, in a target cell, or in an organelle. The programmable nucleoprotein molecular complex can interact with a predetermined target nucleic acid sequence. The programmable nucleoprotein molecular complex may comprise a polynucleotide molecule encoding a chimeric polypeptide. The chimeric polypeptide can comprise a functional domain that can modify a target nucleic acid site. The functional domain can be devoid of a specific nucleic acid binding site. The chimeric polypeptide can comprise a linking domain that can interact with a SCNA. The linking domain can be devoid of a specific target nucleic acid binding site. A SCNA can comprise a nucleotide sequence complementary to a region of a target nucleic acid flanking the target site. A SCNA can comprise a recognition region that can specifically attach to the linking domain of a chimeric polypeptide. Assembly of the chimeric polypeptide and the SCNA within the target cell can form a functional nucleoprotein complex. The nucleoprotein complex can specifically modify a target nucleic acid at the target site.

[0423] In another embodiment, the polynucleotide guided endonuclease gene can be a full-length polynucleotide guided endonuclease (e.g., Cas endonuclease, Cas9 endonuclease), or any functional fragment or functional variant thereof.

[0424] The terms “functional fragment”, “fragment that is functionally equivalent” and “functionally equivalent fragment” can be used interchangeably herein. In the context of a sequence encoding a polynucleotide guided polypeptide, these terms can refer to a portion or subsequence of the polynucleotide guided polypeptide sequence. The portion or subsequence of the polynucleotide guided polypeptide sequence can comprise the ability to create a single-strand or double-strand break.

[0425] The terms “functional variant”, “variant that is functionally equivalent” and “functionally equivalent variant” can be used interchangeably herein. In the context of a polynucleotide guided polypeptide, these terms can refer to a variant of the polynucleotide guided polypeptide. The variant can comprise the ability to create a single-strand or double-strand break. Fragments and variants can be obtained via methods such as site-directed mutagenesis and synthetic construction.

[0426] In one embodiment, the polynucleotide guided polypeptide coding sequence can be a plant codon-optimized *Streptococcus pyogenes* Cas9 coding sequence. The codon optimized Cas9 sequence can recognize any genomic sequence, for example, of the form N(12-30)NGG.

[0427] In one embodiment, the polynucleotide guided polypeptide can be introduced directly into a cell by any suitable method, for example, but not limited to transient introduction methods, transfection and/or topical application.

[0428] Compositions and methods of the disclosure can use endonucleases. Endonucleases can be enzymes that cleave the phosphodiester bond within a polynucleotide chain. Endonucleases can include restriction endonucleases that cleave DNA at specific sites without damaging the bases. Restriction endonucleases can include Type I, Type II, Type III, and Type IV endonucleases, which can further include subtypes. In the Type I and Type III systems, both the methylase and restriction activities can be contained in a single complex. Endonucleases can also include meganucleases, also known as homing endonucleases (HEases). Meganucleases can bind and cut at a specific recognition site, which can be about 18 bp or more. Meganucleases can be classified into four families based on conserved sequence motifs. The meganuclease families can be LAGLIDADG, GIY-YIG, H—N—H, and His-Cys box families. These motifs can participate in the coordination of metal ions and hydrolysis of phosphodiester bonds. HEases can have long recognition sites, and can tolerate sequence polymorphisms in their DNA substrates. The naming convention for meganuclease can be similar to the convention for other restriction endonuclease.

[0429] Meganucleases can also be characterized by prefix F-, I-, or PI- for enzymes encoded by free-standing ORFs, introns, and inteins, respectively. One step in the recombination process can involve polynucleotide cleavage at or near the recognition site. This cleaving activity can be used to produce a double-strand break. In some examples the recombinase can be from the Integrase or Resolvase families.

[0430] Compositions and methods of the disclosure can use Transcription activator-like effector nucleases (TALENs; TAL effector nucleases) can be a class of sequence-specific nucleases. TALENs can be used to cleave (e.g., double-strand breaks) at specific target sequences (e.g., in the genome of a plant or other organism). TAL effector nucleases can be created by fusing a native or engineered transcription activator-like (TAL) effector, or functional part thereof, to the catalytic domain of an endonuclease, such as, for example, FokI. The unique, modular TAL effector DNA binding domain can allow for the design of proteins with potentially any given DNA recognition specificity.

[0431] Compositions and methods of the disclosure can use zinc finger nucleases (ZFNs). ZFNs can be engineered cleavage (e.g., double-strand break) inducing agents comprised of a zinc finger DNA binding domain and a double-strand-break-inducing agent domain. Recognition site specificity can be conferred by the zinc finger domain, which can comprise two, three, or four zinc fingers, for example having a C2H2 structure. Zinc finger domains can be amenable for designing polypeptides which specifically bind a selected polynucleotide recognition sequence. ZFNs can consist of an engineered DNA-binding zinc finger domain linked to a non-specific endonuclease domain, for example, a nuclease domain from a Type IIS endonuclease such as FokI. Additional functionalities can be fused to the zinc-finger binding domain, including transcriptional activator domains, transcription repressor domains, and methylases. In some examples, dimerization of nuclease domain may be required for cleavage activity. Each zinc finger can recognize, for example, three consecutive base pairs in the target DNA. For example, a 3 finger domain recognized a sequence of 9 contiguous nucleotides, with a dimerization requirement of

the nuclease, two sets of zinc finger triplets can be used to bind an 18 nucleotide recognition sequence.

[0432] c. Guide Polynucleic Acid

[0433] Bacteria and archaea can have evolved adaptive immune defenses termed clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) systems that can use short RNA to direct degradation of foreign nucleic acids. The type II CRISPR/Cas system from bacteria can employ a crRNA and tracrRNA to guide the Cas polypeptide to a nucleic acid target. The crRNA (CRISPR RNA) can contain the region complementary to one strand of the double strand DNA target. The crRNA can base pair with the tracrRNA (trans-activating CRISPR RNA) to form a RNA duplex that can direct the Cas polypeptide to recognize and optionally cleave the DNA target.

[0434] As used herein, the term “guide polynucleotide”, can refer to a polynucleotide sequence that can form a complex with a polynucleotide guided polypeptide (e.g., a Cas protein). The guide polynucleotide can direct the polynucleotide guided polypeptide to recognize and optionally cleave (or nick) a DNA target site. The terms “guide polynucleotide” and “guide polynucleic acid” can be used interchangeably herein. The guide polynucleotide can be comprised of a single molecule (unimolecular) or two molecules (bimolecular). The guide polynucleotide sequence can be a RNA sequence, a DNA sequence, or a combination thereof (a RNA-DNA combination sequence).

[0435] Optionally, the guide polynucleotide can comprise at least one nucleotide, phosphodiester bond or linkage modification such as, but not limited, to Locked Nucleic Acid (LNA), 5-methyl dC, 2,6-Diaminopurine, 2'-Fluoro A, 2'-Fluoro U, 2'-O-Methyl RNA, phosphorothioate bond, linkage to a cholesterol molecule, linkage to a polyethylene glycol molecule, linkage to a spacer 18 (hexaethylene glycol chain) molecule, or 5' to 3' covalent linkage resulting in circularization. A guide polynucleotide that solely comprises ribonucleic acids can also be referred to as a “guide RNA” (gRNA). In some embodiments, the guide polynucleic acid can be a guide RNA.

[0436] As used herein, the term “single guide RNA” (sgRNA) can refer to a synthetic fusion of two RNA molecules, for example, a crRNA (CRISPR RNA) comprising a variable targeting domain, and a tracrRNA. In one embodiment, the guide RNA can comprise a variable targeting domain of 12 to 30 nucleotide sequences and a RNA fragment that can interact with a Cas protein.

[0437] As used herein, “crRNA” can refer to a nucleic acid with at least about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% sequence identity and/or sequence similarity to a wild type exemplary crRNA (e.g., a crRNA from *S. pyogenes*). crRNA can refer to a nucleic acid with at most about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% sequence identity and/or sequence similarity to a wild type exemplary crRNA (e.g., a crRNA from *S. pyogenes*). crRNA can refer to a modified form of a crRNA that can comprise a nucleotide change such as a deletion, insertion, or substitution, variant, mutation, or chimera. A crRNA can be a nucleic acid having at least about 60% identical to a wild type exemplary crRNA (e.g., a crRNA from *S. pyogenes*) sequence over a stretch of at least 6 contiguous nucleotides. For example, a crRNA sequence can be at least about 60% identical, at least about 65% identical, at least about 70% identical, at least about 75%

identical, at least about 80% identical, at least about 85% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, or 100% identical, to a wild type exemplary crRNA sequence (e.g., a crRNA from *S. pyogenes*) over a stretch of at least 6 contiguous nucleotides

[0438] As used herein, “tracrRNA” can refer to a nucleic acid with at least about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% sequence identity and/or sequence similarity to a wild type exemplary tracrRNA sequence (e.g., a tracrRNA from *S. pyogenes*). tracrRNA can refer to a nucleic acid with at most about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% sequence identity and/or sequence similarity to a wild type exemplary tracrRNA sequence (e.g., a tracrRNA from *S. pyogenes*). tracrRNA can refer to a modified form of a tracrRNA that can comprise a nucleotide change such as a deletion, insertion, or substitution, variant, mutation, or chimera. A tracrRNA can refer to a nucleic acid that can be at least about 60% identical to a wild type exemplary tracrRNA (e.g., a tracrRNA from *S. pyogenes*) sequence over a stretch of at least 6 contiguous nucleotides. For example, a tracrRNA sequence can be at least about 60% identical, at least about 65% identical, at least about 70% identical, at least about 75% identical, at least about 80% identical, at least about 85% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, or 100% identical, to a wild type exemplary tracrRNA (e.g., a tracrRNA from *S. pyogenes*) sequence over a stretch of at least 6 contiguous nucleotides.

[0439] A guide polynucleotide can be bimolecular (i.e., two molecules; also referred to as “double molecule”, “dual” or “duplex” guide polynucleotide) comprising, for example, a first nucleotide sequence domain (referred to as Variable Targeting domain or VT domain) that is complementary to a nucleotide sequence in a target polynucleic acid (e.g., target DNA) and a second nucleotide sequence domain (referred to as Cas endonuclease recognition domain or CER domain) that interacts with a Cas polypeptide. The VT domain can refer to the spacer region of a guide polynucleic acid. The VT domain can comprise a spacer region of a guide polynucleic acid. The spacer region can interact with a protospacer region of a target nucleic acid in a sequence-specific manner via hybridization (e.g., base pairing). The CER domain of the bimolecular guide polynucleotide can comprise two separate molecules that can be hybridized along a region of complementarity to form, for example, a duplex or a partial duplex. The two separate molecules can be RNA, DNA, and/or RNA-DNA-combination sequences. In some embodiments, the first molecule of the duplex guide polynucleotide comprising a VT domain linked to a CER domain can be referred to as “crDNA” (when composed of a contiguous stretch of DNA nucleotides) or “crRNA” (when composed of a contiguous stretch of RNA nucleotides), or “crDNA-RNA” (when composed of a combination of DNA and RNA nucleotides). The crNucleotide can comprise a fragment of the crRNA naturally occurring in bacteria and archaea. In one embodiment, the size of the fragment of the crRNA naturally occurring in bacteria and archaea that can be present in a crNucleotide disclosed herein can range from, but is not limited to, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more nucleotides. In some embodiments, the second molecule of the duplex guide polynucleotide comprising a CER domain can be referred to

as “tracrRNA” (when composed of a contiguous stretch of RNA nucleotides) or “tracrDNA” (when composed of a contiguous stretch of DNA nucleotides) or “tracrDNA-RNA” (when composed of a combination of DNA and RNA nucleotides). In one embodiment, the RNA that guides the RNA/Cas9 polypeptide complex, can be a duplexed RNA comprising a duplex crRNA-tracrRNA.

[0440] Complementarity between a guide polynucleic acid (e.g., the VT domain, spacer region) and a target polynucleic acid (e.g., protospacer) can be perfect, substantial, or sufficient. Perfect complementarity between two nucleic acids can mean that the two nucleic acids can form a duplex in which every base in the duplex can be bonded to a complementary base by Watson-Crick pairing. Substantial or sufficient complementarity can mean that a sequence in one strand may not be completely and/or perfectly complementary to a sequence in an opposing strand, but that sufficient bonding occurs between bases on the two strands to form a stable hybrid complex in a set of hybridization conditions (e.g., salt concentration and temperature).

[0441] A guide polynucleotide can also be a single molecule (i.e., unimolecular), comprising a first nucleotide sequence domain (referred to as Variable Targeting domain or VT domain) that can be complementary to a nucleotide sequence in a target polynucleic acid (e.g., target DNA) and a second nucleotide domain (referred to as Cas endonuclease recognition domain or CER domain) that interacts with a Cas polypeptide. For a single molecule guide polynucleotide, the CER domain can be formed from a contiguous stretch of nucleotides that can be RNA, DNA, and/or RNA-DNA-combination sequence. The VT domain and/or the CER domain of a single guide polynucleotide can comprise a RNA sequence, a DNA sequence, or a RNA-DNA-combination sequence. In some embodiments, the single guide polynucleotide comprises a crNucleotide (comprising a VT domain linked to a CER domain) linked to a tracrNucleotide (comprising a CER domain), wherein the linkage can be a nucleotide sequence comprising a RNA sequence, a DNA sequence, or a RNA-DNA combination sequence. The single guide polynucleotide being comprised of sequences from the crNucleotide and tracrNucleotide may be referred to as “single guide RNA” (sgRNA; when composed of a contiguous stretch of RNA nucleotides) or “single guide DNA” (sgDNA; when composed of a contiguous stretch of DNA nucleotides) or “single guide RNA-DNA” (sgDNA-RNA; when composed of a combination of DNA and RNA nucleotides). In one embodiment of the disclosure, the single guide RNA (sgRNA) comprises a crRNA or crRNA fragment and a tracrRNA or tracrRNA fragment of the type II CRISPR/Cas system that can form a complex with a type II Cas polypeptide, wherein said guide RNA/Cas polypeptide complex can direct the Cas polypeptide to a plant genomic target site, enabling the Cas polypeptide to introduce a double strand break into the genomic target site.

[0442] The term “variable targeting domain” or “VT domain” can be used interchangeably herein and can refer to a nucleotide sequence that can be present in the guide polynucleotide. VT domain can be complementary to one strand of a double stranded DNA target site. The percent complementation between the first nucleotide sequence domain (VT domain) and the target sequence can be at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 63%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%,

80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100%. The variable target domain can be at least 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 or 30 nucleotides in length. In some embodiments, the variable target domain can comprise at least 17 nucleotides that are complementary to at least 17 nucleotides of a target polynucleic acid. In some embodiments, the variable targeting domain can comprise a contiguous stretch of nucleotides that are complementary to the target polynucleic acid. In some embodiments, the nucleotides of the guide polynucleic acid that are complementary to the target polynucleic acid can be non-contiguous. In some embodiments, the variable targeting domain can comprise a contiguous stretch of 12 to 30 nucleotides. The variable targeting domain can be composed of a DNA sequence, a RNA sequence, a modified DNA sequence, a modified RNA sequence, or any combination thereof.

[0443] A target polynucleotide can be identified by identifying a protospacer adjacent motif (PAM) within a region of interest and selecting a region of a desired size upstream or downstream of the PAM as the protospacer. A corresponding spacer sequence can be designed by determining the complementary sequence of the protospacer region.

[0444] The term “Cas endonuclease recognition domain” or “CER domain” of a guide polynucleotide can be used interchangeably herein and can refer to a nucleotide sequence (such as a second nucleotide sequence domain of a guide polynucleotide), that interacts with a Cas polypeptide. The CER domain can be composed of a DNA sequence, a RNA sequence, a modified DNA sequence, a modified RNA sequence (see for example modifications described herein), or any combination thereof.

[0445] The nucleotide sequence linking the crNucleotide and the tracrNucleotide of a single guide polynucleotide can comprise a RNA sequence, a DNA sequence, or a RNA-DNA combination sequence. In one embodiment, the nucleotide sequence linking the crNucleotide and the tracrNucleotide of a single guide polynucleotide can be at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100 nucleotides in length. In another embodiment, the nucleotide sequence linking the crNucleotide and the tracrNucleotide of a single guide polynucleotide can comprise a tetranucleotide loop sequence, such as, but not limiting to a GAAA tetranucleotide loop sequence. Nucleotide sequence modification of the guide polynucleotide, VT domain and/or CER domain can be selected from, but not limited to, the group consisting of a 5' cap, a 3' polyadenylated tail, a riboswitch sequence, a stability control sequence, a sequence that forms a dsRNA duplex, a modification or sequence that targets the guide polynucleotide to a subcellular location, a modification or sequence that provides for tracking, a modification or sequence that provides a binding site for proteins, a Locked Nucleic Acid (LNA), a 5-methyl-2'-deoxycytidine (5mdC), a 2,6-Diaminopurine nucleotide, a 2'-Fluoroadenosine nucleotide, a 2'-Fluorouridine nucleotide; a 2'-O-Methyl RNA nucleotide, a phosphorothioate (PS) bond, linkage to a cholesterol molecule, linkage to a polyethylene glycol molecule, linkage to a spacer 18 mol-

ecule, a 5' to 3' covalent linkage, or any combination thereof. These modifications can result in at least one additional beneficial feature, wherein the additional beneficial feature can be selected from the group consisting of: modified or regulated stability, subcellular targeting, tracking, a fluorescent label, a binding site for a protein or protein complex, modified binding affinity to complementary target sequence, modified resistance to cellular degradation, and increased cellular permeability.

[0446] In one embodiment, the guide RNA and Cas polypeptide can form a complex that can enable the Cas polypeptide to introduce a single strand or double strand break at a DNA target site.

[0447] In one embodiment, the variable target domain can be 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 or 30 nucleotides in length.

[0448] In one embodiment, the guide RNA can comprise a crRNA (or crRNA fragment) and a tracrRNA (or tracrRNA fragment) of the type II CRISPR/Cas system that can form a complex with a type II Cas polypeptide. The guide RNA/Cas polypeptide complex can direct the Cas polypeptide to a target nucleic acid site (e.g., DNA target). The Cas polypeptide can introduce a double strand break into the DNA target site.

[0449] In one embodiment the guide polynucleic acid can be introduced into a cell directly using any suitable method such as, but not limited to, particle bombardment or topical applications.

[0450] In another embodiment the guide polynucleic acid can be introduced indirectly by introducing a recombinant DNA molecule comprising a polynucleotide encoding the guide polynucleic acid operably linked to a nuclear or organellar promoter that can transcribe the polynucleotide in said nucleus or organelle, respectively.

[0451] In some embodiments, the guide polynucleic acid can be introduced into a plant cell via particle bombardment or *Agrobacterium* transformation of a recombinant DNA construct comprising a polynucleotide encoding the guide polynucleic acid operably linked to a promoter functional in a plant; e.g., a plant U6 polymerase III promoter, a CaMV 35S polymerase II promoter.

[0452] In one embodiment, the guide polynucleic acid can be a duplexed RNA comprising a duplex crRNA-tracrRNA. A single guide polynucleic acid (e.g., single guide RNA) can require one expression cassette to express the single guide RNA. A duplexed crRNA-tracrRNA can require one or more expression cassette needs to express the duplexed crRNA-tracrRNA.

[0453] A plurality of polynucleic acids can be multiplexed to target multiple target nucleic acids. For example, 2, 3, 4, 5, 6, 7, 9, 10, or more than 10 target nucleic acids can be targeted simultaneously or iteratively. Multiplexing can be used, as non-limiting examples, to generate large genomic deletions, modify multiple different sequences at once, and/or in conjunction with dual-nickases to target a gene. In some examples, more than one CRISPR/Cas system can be delivered to target two or more nucleic acid sequence targets. Homologous Cas proteins can be used for multiplexing applications.

Target Sites for Genome Modification

[0454] The terms “target site”, “target sequence”, “target polynucleotide”, “target polynucleic acid”, “target locus”, “genomic target site”, “genomic target sequence”, and

“genomic target locus” can be used interchangeably herein. Target polynucleic acid can refer to a polynucleotide sequence in the genome (e.g., plastid or mitochondrial genome) of, for example, a plant cell. Target polynucleic acid can refer to the site (e.g., in a genome) recognized by a guide polynucleic acid. Target polynucleic acid can refer to the site (e.g., in a genome) at which a single-strand or double-strand break can be induced (e.g., by a Cas polypeptide). The target site can be an endogenous site in the genome. The target site can be heterologous to the organism and thereby not be naturally occurring in the genome. Target site can be found in a heterologous genomic location compared to where it occurs in nature. As used herein, terms “endogenous target sequence” and “native target sequence” can be used interchangeably herein and can refer to a target sequence that can be endogenous or native to the genome of the organism. Endogenous target sequence can occur at the endogenous or native position of that target sequence in the genome of the organism.

[0455] A target polynucleic acid can be DNA, RNA, or both. In some embodiments, the target polynucleic acid can be DNA (e.g., target DNA). In some embodiments, the target polynucleic acid can be genomic DNA. In some embodiments, the target polynucleic acid can be nuclear genomic DNA. In some embodiments, the target polynucleic acid can be organelle genomic DNA. In some embodiments, the target polynucleic acid can be nuclear genomic DNA and organelle genomic DNA.

[0456] The terms “artificial target site” and “artificial target sequence” can be used interchangeably herein and can refer to a target sequence that has been introduced into the genome of a plant. Such an artificial target sequence can be identical in sequence to an endogenous or native target sequence in the genome of an organism but may be located in a different position (i.e., a non-endogenous or non-native position) in the genome of the organism.

[0457] An “altered target site”, “altered target sequence”, “modified target site”, “modified target sequence” can be used interchangeably herein and can refer to a target sequence as disclosed herein that can comprise at least one alteration when compared to the non-altered target sequence. Such “alterations” can include, for example: (i) replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, or (iv) any combination of (i)-(iii).

[0458] Methods for modifying an organellar genomic target site are disclosed herein.

[0459] In one embodiment, a method for modifying a target site in the genome of an organelle can comprise introducing a guide polynucleic acid (e.g., guide RNA, single guide RNA) into a plant cell. The plant cell can comprise a polynucleotide guided polypeptide (e.g., a Cas polypeptide). The guide polynucleic acid and polynucleotide guided polypeptide can form a complex that can direct the polynucleotide guided polypeptide to introduce a single strand or double strand break at the target site.

[0460] Also provided is a method for modifying a target site in the genome of an organelle. The method can comprise introducing a guide polynucleic acid and a polynucleotide guided polypeptide (e.g., a Cas polypeptide) into the organelle. The guide polynucleic acid and polynucleotide guided polypeptide can form a complex. The complex can direct the

polynucleotide guided polypeptide to introduce a single strand or double strand break at the target site in the genome of the organelle.

[0461] Further provided is a method for modifying a target site in the genome of an organelle. The method can comprise introducing a guide polynucleic acid and a donor polynucleotide (e.g., donor DNA) into an organelle. The organelle can comprise a polynucleotide guided polypeptide (e.g., a Cas polypeptide). The guide polynucleic acid and polynucleotide guided polypeptide can form a complex that can direct the polynucleotide guided polypeptide to introduce a single strand or double strand break at the target site. The donor polynucleotide can be inserted into the site of cleavage in the genome.

[0462] Further provided is a method for modifying a target site in the genome of an organelle. The method can comprise: a) introducing into an organelle a guide polynucleic acid comprising a variable targeting domain and a polynucleotide guided polypeptide (e.g., a Cas polypeptide), wherein said guide polynucleic acid and said polynucleotide guided polypeptide can form a complex that can enable the polynucleotide guided polypeptide to introduce a single strand or double strand break at said target site; and, b) identifying at least one organelle that has a modification at said target site, wherein the modification includes at least one deletion or substitution of one or more nucleotides in said target site.

[0463] Further provided, a method for modifying a target polynucleic acid (e.g., target DNA) sequence in the genome of an organelle, the method comprising: a) introducing into an organelle a first recombinant DNA construct that can express a guide polynucleic acid and a second recombinant DNA construct that can express a polynucleotide guided polypeptide (e.g., a Cas polypeptide), wherein said guide polynucleic acid and said polynucleotide guided polypeptide can form a complex that can enable the polynucleotide guided polypeptide to introduce a single strand or double strand break at said target site; and, b) identifying at least one organelle that has a modification at said target site, wherein the modification includes at least one deletion or substitution of one or more nucleotides in said target site.

[0464] The length of the target site can vary and includes, for example, target sites that are at least 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 or more nucleotides in length. The target site can be palindromic, that is, the sequence on one strand reads the same in the opposite direction on the complementary strand. The nick/cleavage site can be within the target sequence. The nick/cleavage site can be outside of the target sequence. In another variation, the cleavage could occur at nucleotide positions immediately opposite each other to produce a blunt end cut or, in other cases, the incisions could be staggered to produce single-stranded overhangs, also called “sticky ends”, which can be either 5' overhangs, or 3' overhangs.

[0465] The target nucleic acid sequence can be 5' or 3' of the PAM. The target nucleic acid sequence can be, for example, 16, 17, 18, 19, 20, 21, 22, or 23 bases immediately 5' of the first nucleotide of the PAM. The target nucleic acid sequence can be, for example, 16, 17, 18, 19, 20, 21, 22, or 23 bases immediately 3' of the last nucleotide of the PAM. The target nucleic acid sequence can be 20 bases immediately 5' of the first nucleotide of the PAM. The target nucleic acid sequence can be 20 bases immediately 3' of the last nucleotide of the PAM.

[0466] Site-specific cleavage of a target nucleic acid by a polynucleotide guided polypeptide (e.g., Cas protein) can occur at locations determined by base-pairing complementarity between the guide nucleic acid and the target nucleic acid. Site-specific cleavage of a target nucleic acid by a polynucleotide guided polypeptide (e.g., Cas protein) can occur at locations determined by the protospacer adjacent motif (PAM). For example, the cleavage site of Cas (e.g., Cas9) can be about 1 to about 25, or about 2 to about 5, or about 19 to about 23 base pairs (e.g., 3 base pairs) upstream or downstream of the PAM sequence. In some embodiments, the cleavage site of Cas (e.g., Cas9) can be 3 base pairs upstream of the PAM sequence. In some embodiments, the cleavage site of Cas (e.g., Cpf1) can be 19 bases on the (+) strand and 23 base on the (−) strand, producing a 5' overhang 5 nt in length. In some cases, the cleavage can produce blunt ends. In some cases, the cleavage can produce staggered or sticky ends with 5' overhangs. In some cases, the cleavage can produce staggered or sticky ends with 3' overhangs.

[0467] Different organisms can comprise different PAM sequences. Different Cas proteins can recognize different PAM sequences. For example, in *S. pyogenes*, the PAM can be a sequence in the target nucleic acid that comprises the sequence 5'-XRR-3', where R can be either A or G, where X can be any nucleotide and X can be immediately 3' of the target nucleic acid sequence targeted by the spacer sequence. The PAM sequence of *S. pyogenes* Cas9 (SpyCas9) can be 5'-XGG-3', where X can be any DNA nucleotide and can be immediately 3' of the CRISPR recognition sequence of the non-complementary strand of the target DNA. The PAM of Cpf1 can be 5'-TTX-3', where X can be any DNA nucleotide and can be immediately 5' of the CRISPR recognition sequence.

[0468] Active variants of genomic target sites can also be used. Such active variants can comprise at least 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to the given target site. The active variants can retain biological activity. The active variants can be recognized by a polynucleotide guided polypeptide (e.g., Cas protein). The active variants can be cleaved by a polynucleotide guided polypeptide (e.g., Cas protein). Assays can be used to measure the double-strand break of a target site by an endonuclease. Assays can measure the overall activity and/or specificity of an endonuclease on DNA substrates containing recognition sites (e.g., target sites, active variants).

Methods for Integrating a Donor Polynucleotide

[0469] The disclosure provides methods to obtain an organelle comprising a donor polynucleotide. Such methods can employ homologous recombination to provide integration of the polynucleotide at the target site. A polynucleotide of interest can be provided to the organelle in a donor DNA molecule.

[0470] A donor polynucleotide can be a nucleic acid sequence (e.g., DNA, RNA, or both) that can be integrated into a target nucleic acid, for example, the genome of an organelle. The donor polynucleotide can be inserted into a genome e.g., at a cleavage site of a polynucleotide guided polypeptide. The donor polynucleotide can be inserted into a genome by homologous recombination. In some embodiments, the donor polynucleotide can comprise DNA and can be referred to as donor DNA.

[0471] A donor polynucleotide of any suitable size can be integrated into a genome. In some embodiments, the donor polynucleotide integrated into a genome can be less than 3, about 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, 10.5, 11, 11.5, 12, 12.5, 13, 13.5, 14, 14.5, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500 or more than 500 kilobases (kb) in length. In some embodiments, the donor polynucleotide integrated into a genome can be at least about 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, 10.5, 11, 11.5, 12, 12.5, 13, 13.5, 14, 14.5, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500 or more than 500 (kb) in length. In some embodiments, the donor polynucleotide integrated into a genome can be up to about 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, 10.5, 11, 11.5, 12, 12.5, 13, 13.5, 14, 14.5, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500 or more than 500 (kb) in length.

[0472] A donor polynucleotide can comprise a polynucleotide of interest, a polynucleotide modification template, a heterologous expression cassette, or both. A donor polynucleotide (e.g. donor DNA) can be flanked by a first and a second region of homology. The polynucleotide modification template can be, for example, a single nucleotide change to create a different allele in the organellar genome. The first and second regions of homology of the donor polynucleotide (e.g. donor DNA) can share homology to a first and a second genomic region, respectively, present in or flanking the target site (e.g., of the organellar genome).

[0473] “Homology” can mean DNA sequences that are similar. Homology can mean, for example, nucleic acid sequences with about: 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% homology or identity. For example, a “region of homology to a genomic region” can be a region of DNA that has a similar sequence to a given “genomic region” in the organellar genome. A region of homology can be of any length that can be sufficient to promote homologous recombination at the cleaved target site. For example, the region of homology can comprise at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100 or more bases in length such that the region of homology has sufficient homology to undergo homologous recombination with the corresponding genomic region. “Sufficient homology” can indicate that two polynucleotide sequences can have sufficient structural similarity to act as substrates for a homologous recombination reaction.

[0474] The donor polynucleotide (e.g., donor DNA) may comprise an expression cassette (e.g., encoding a heterologous polynucleotide of interest). The donor polynucleotide may comprise multiple expression cassettes. The expression cassette may be a polycistronic expression cassette; e.g., where multiple protein-coding regions, functional RNAs, or a combination of both, are expressed under control of a single promoter.

[0475] A “donor RNA” can be a corresponding RNA molecule that comprises, for example, the same nucleic acid sequence as a donor DNA; i.e., with uridylyate (“U”) in place of deoxythymidylate (“T”). A “donor polynucleotide” may be either a donor DNA or a donor RNA, or a combination

of DNA and RNA. The donor polynucleotide may be either single-stranded or double-stranded.

[0476] An alternative method for modification of an organellar genome can be the replacement of part or all of the organelle DNA with a “replacement DNA”. Endogenous organellar DNA can be reduced or eliminated by use of site-specific endonucleases such as polynucleotide guided polypeptides (e.g., Cas polypeptide, Cas9 polypeptide). At the same time or subsequently, a replacement DNA may be introduced. The term “replacement DNA” can refer to fragments of organellar DNA or complete organellar DNA that can convey a new genotype and corresponding trait(s) when transformed into the organelle. The terms “replacement DNA” and “replacement organellar DNA” can be used interchangeably herein. In the case of organellar DNA fragments, they can be integrated into the remaining endogenous organellar DNA by homologous recombination. In the case of complete organellar DNA replacement, the replacement DNA can be isolated from cultivars, lines, sub species and other species which possess DNA compositions distinct from the endogenous organellar DNA of recipient cells. In some embodiments, the replacement DNA can comprise a DNA element functioning as a DNA replication origin in the recipient organelles.

[0477] A sequence functional as an origin of replication can be included with the compositions (e.g., polynucleotides, constructs, cassettes) of the disclosure. Such sequences can include origin of replication for an organelle. The origin of replication sequence can be a plastid origin of replication (e.g., plastid rRNA intergenic region) sequence. The origin of replication sequence can be a mitochondrial origin of replication sequence.

[0478] As used herein, a “genomic region” can refer to a segment of a chromosome in the genome of, for example, an organelle. Genomic region can be present on either side of the target site. Genomic region can comprise a portion of the target site. The genomic region can comprise at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100 or more bases. The genomic region can comprise sufficient homology to undergo homologous recombination with the corresponding region of homology.

[0479] Donor polynucleotides, polynucleotides of interest and/or traits can be stacked together in a complex trait locus. The guide polynucleotide/polypeptide system can be used to generate double strand breaks and for stacking traits in a complex trait locus.

[0480] Two or more polynucleotides encoding RNA and/or proteins can be included in a cassette as a polycistronic unit. Polynucleotides encoding RNA can be expressed from separate cassettes.

[0481] In one embodiment, the guide polynucleotide/polypeptide system can be used for introducing one or more donor polynucleotides or one or more traits of interest into one or more target sites by providing one or more guide polynucleotides, one or more polynucleotide guided polypeptides (e.g., Cas polypeptides), and optionally one or more donor polynucleotides (e.g. donor DNA) to a plant cell. An organism can be produced from that cell that comprises an alteration at said one or more target sites of the organellar DNA, wherein the alteration can be selected from the group consisting of (i) replacement of at least one nucleotide, (ii)

a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, and (iv) any combination of (i)-(iii).

[0482] The structural similarity between a given genomic region and the corresponding region of homology flanking the donor polynucleotide (e.g. donor DNA) can be any degree of sequence identity that allows for homologous recombination to occur. For example, the amount of homology or sequence identity shared by the “region of homology” flanking the donor polynucleotide (e.g. donor DNA) and the “genomic region” of the plant genome can be at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity, such that the sequences undergo homologous recombination.

[0483] The region of homology flanking the donor polynucleotide (e.g. donor DNA) can have homology to any sequence flanking the target site. While in some embodiments, the regions of homology share significant sequence homology to the genomic sequence immediately flanking the target site, the regions of homology can be designed to have sufficient homology to regions that may be further 5' or 3' to the target site. In still other embodiments, the regions of homology can also have homology with a fragment of the target site along with downstream genomic regions. In one embodiment, the first region of homology further comprises a first fragment of the target site and the second region of homology comprises a second fragment of the target site, wherein the first and second fragments are dissimilar.

[0484] As used herein, “homologous recombination” can refer to the exchange of DNA fragments between two DNA molecules at the sites of homology. The frequency of homologous recombination can be influenced by a number of factors. The length of the region of homology can affect the frequency of homologous recombination events, for example, the longer the region of homology, the greater the frequency. The length of the homology region needed to observe homologous recombination may vary among species.

[0485] Intermolecular recombination can occur in plastids, for example, transplastomic plants can arise through site-specific integration of foreign sequences by homologous recombination with the flanking sequence on the transformation vector.

[0486] The generation of novel plastome genotypes by transformation can rely on integration of foreign sequence by intermolecular homologous recombination (HR). Mechanistically similar to gene conversion, HR and repair pathways can participate in the subsequent events that yield homoplasmic transplastomic cells and eventually stable transplastomic plants. Intra- or intermolecular recombination between repeated sequences, both in wild-type plastomes, can generate, for example, inversions when repeats are palindromic or deletions when direct. The role of HR proteins in damage repair may be compromised, for example, when foreign DNA is introduced, and through associated tissue culture and selective pressure, as these manipulations can place additional stress on recombination machinery leading to unintended events.

[0487] Among the DNA repair and recombination genes identified in the nuclear genomes of *Oryza* and *Arabidopsis*, about 19 and 17%, respectively, can be targeted to plastids.

[0488] Plastid-localized RecA (e.g., from *P. sativum*) can comprise DNA strand transfer activity. RecA can be impli-

cated in recombination-mediated repair of damaged ptDNA. Reduced RecA1 (AT1G79050) activity can lead to a destabilization and reduction in ptDNA. The reduction in plastome copy number in mutant lines relative to wild type can suggest that RecA1 may participate in recombination-mediated replication.

[0489] Methods of the disclosure can use any suitable plastid enzymes for homologous DNA recombination pathway. The predominance of homologous recombination in plastids can result from suppression of illegitimate recombination by plastid-localized members of the whirly family of single-stranded DNA binding proteins. HR activity in a cell can be optimized by increasing HR pathway members.

[0490] To achieve efficient foreign sequence integration by homologous recombination endogenous plastome sequences can be used to target insertions. A positive correlation can be present between the rate of recombination and the length and/or degree of sequence homology.

[0491] The minimum flanking sequence length for plastid transformation can be as little as 400 bp on either side of the expression cassette and can be sufficient to obtain transformation at a reasonable frequency. Targeting sequences can extend from 1 to 1.5 kb on either side of the expression cassette.

[0492] Non-homologous end-joining (NHEJ) can be a major DNA repair pathway in the eukaryotic nucleus. NHEJ can also be active in bacteria and in plant mitochondria. In some cases, NHEJ may not occur in angiosperm plastids. NHEJ products can be produced in *Arabidopsis*. In some cases, repair of DSBs by NHEJ following I-CreII activity can be detected at low frequency. NHEJ repair events can represent 17% of the rearranged products in Whirly knockout lines. NHEJ can occur in plastids. NHEJ can be a quantitatively minor pathway.

[0493] The methods of the disclosure can use homology-directed repair (HDR) or NHEJ. In some embodiments, HDR can be used. In some embodiments, the efficiency of HDR can be increased by, for example, increasing expression of proteins and enzymes involved in HDR. In some embodiments, the efficiency of NHEJ can be reduced, by for example, targeting genes and/or proteins (e.g., DNA ligase) involved in NHEJ.

[0494] In some embodiments, the efficiency of the disclosed methods for genome engineering or modification can be about 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100%.

[0495] In one embodiment provided herein, the method can comprise contacting an organelle of a plant cell with the donor polynucleotide (e.g. donor DNA), the guide polynucleic acid and the polynucleotide guided polypeptide. At least one one single-strand or double-strand break can be introduced in the target site by the polynucleotide guided polypeptide, the first and second regions of homology flanking the donor polynucleotide (e.g. donor DNA) can undergo homologous recombination with their corresponding genomic regions of homology resulting in exchange of DNA between the donor and the genome. As such, the provided methods can result in the integration of the donor polynucleotide (e.g. donor DNA) into the single-strand or double-strand break(s) in the target site in the organellar genome, thereby altering the original target site and producing an altered genomic target site.

[0496] The donor polynucleotide (e.g. donor DNA) may be introduced by any suitable means. For example, a plant having a target site can be provided. The donor polynucleotide (e.g. donor DNA) may be provided by any suitable transformation method including, for example, *Agrobacterium*-mediated transformation or biolistic particle bombardment. The donor polynucleotide (e.g. donor DNA) may be present transiently in the cell or it could be introduced via a viral replicon. In the presence of the guide polynucleotide (e.g., guide RNA), the polynucleotide guided polypeptide (e.g., Cas polypeptide) and the target site, the donor polynucleotide (e.g. donor DNA) can be inserted into the organellar genome.

[0497] Donor polynucleotides can be reflective of the commercial markets. Donor polynucleotides can be reflective of traits for the development of the crop. Crops and markets of interest can change, and as developing nations open up world markets, new crops and technologies can emerge also. In addition, as the understanding of agronomic traits and characteristics such as yield and heterosis increase, the choice of genes for transformation can change accordingly.

Methods for Modulating Gene Expression

[0498] In some aspects are provided methods for modulating expression (e.g., transcription) of a target nucleic acid (e.g., a gene) in a host cell or organelle. The methods can involve contacting the target nucleic acid with an enzymatically inactive Cas protein (e.g., dead Cas) and a guide polynucleic acid.

[0499] In some aspects, the present disclosure provides a method of selectively modulating transcription of a target nucleic acid in a host cell. The method can involve introducing into the host cell an enzymatically inactive Cas protein (e.g., dead Cas) and a guide polynucleic acid. The guide nucleic acid and the dead Cas protein can form a complex in the host cell. The complex can selectively modulate transcription of a target polynucleic acid (e.g., target DNA) in the host cell or organelle.

[0500] In some aspects, the present disclosure provides for selective transcription modulation (e.g., reduction or increase) of a target nucleic acid in a host cell. Selective modulation of transcription of a target nucleic acid can reduce or increase transcription of the target nucleic acid, but may not substantially modulate transcription of a non-target nucleic acid or off-target nucleic acid, e.g., transcription of a non-target nucleic acid may be modulated by less than 1%, less than 5%, less than 10%, less than 20%, less than 30%, less than 40%, or less than 50% compared to the level of transcription of the non-target nucleic acid in the absence of the guide nucleic acid/enzymatically inactive or enzymatically reduced Cas protein complex. For example, selective modulation (e.g., reduction or increase) of transcription of a target nucleic acid can reduce or increase transcription of the target nucleic acid by at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or greater than 90%, compared to the level of transcription of the target nucleic acid in the absence of a guide nucleic acid/enzymatically inactive or enzymatically reduced Cas protein complex.

[0501] In some aspects, the disclosure provides methods for increasing transcription of a target nucleic acid. The transcription of a target nucleic acid can increase by at least

about 1.1 fold, at least about 1.2 fold, at least about 1.3 fold, at least about 1.4 fold, at least about 1.5 fold, at least about 1.6 fold, at least about 1.7 fold, at least about 1.8 fold, at least about 1.9 fold, at least about 2 fold, at least about 2.5 fold, at least about 3 fold, at least about 3.5 fold, at least about 4 fold, at least about 4.5 fold, at least about 5 fold, at least about 6 fold, at least about 7 fold, at least about 8 fold, at least about 9 fold, at least about 10 fold, at least about 12 fold, at least about 15 fold, at least about 20-fold, at least about 50-fold, at least about 70-fold, or at least about 100-fold compared to the level of transcription of the target polynucleic acid (e.g., target DNA) in the absence of a guide nucleic acid/enzymatically inactive or enzymatically reduced Cas protein complex. Selective increase of transcription of a target nucleic acid increases transcription of the target nucleic acid, but may not substantially increase transcription of a non-target polynucleic acid, e.g., transcription of a non-target nucleic acid can be increased, if at all, by less than about 5-fold, less than about 4-fold, less than about 3-fold, less than about 2-fold, less than about 1.8-fold, less than about 1.6-fold, less than about 1.4-fold, less than about 1.2-fold, or less than about 1.1-fold compared to the level of transcription of the non-targeted DNA in the absence of the guide nucleic acid/enzymatically inactive or enzymatically reduced Cas protein complex.

[0502] In some aspects, the disclosure provides methods for decreasing transcription of a target nucleic acid. The transcription of a target nucleic acid can decrease by at least about 1.1 fold, at least about 1.2 fold, at least about 1.3 fold, at least about 1.4 fold, at least about 1.5 fold, at least about 1.6 fold, at least about 1.7 fold, at least about 1.8 fold, at least about 1.9 fold, at least about 2 fold, at least about 2.5 fold, at least about 3 fold, at least about 3.5 fold, at least about 4 fold, at least about 4.5 fold, at least about 5 fold, at least about 6 fold, at least about 7 fold, at least about 8 fold, at least about 9 fold, at least about 10 fold, at least about 12 fold, at least about 15 fold, at least about 20-fold, at least about 50-fold, at least about 70-fold, or at least about 100-fold compared to the level of transcription of the target polynucleic acid (e.g., target DNA) in the absence of a guide nucleic acid/enzymatically inactive or enzymatically reduced Cas protein complex. Selective decrease of transcription of a target nucleic acid decreases transcription of the target nucleic acid, but may not substantially decrease transcription of a non-target DNA, e.g., transcription of a non-target nucleic acid can be decreased, if at all, by less than about 5-fold, less than about 4-fold, less than about 3-fold, less than about 2-fold, less than about 1.8-fold, less than about 1.6-fold, less than about 1.4-fold, less than about 1.2-fold, or less than about 1.1-fold compared to the level of transcription of the non-targeted DNA in the absence of the guide nucleic acid/enzymatically inactive or enzymatically reduced Cas protein complex.

[0503] Transcription modulation can be achieved by fusing the enzymatically inactive Cas protein to a heterologous sequence. The heterologous sequence can be a suitable fusion partner, e.g., a polypeptide that provides an activity that indirectly increases, decreases, or otherwise modulates transcription by acting directly on the target nucleic acid or on a polypeptide (e.g., a histone or other DNA-binding protein) associated with the target nucleic acid. Non-limiting examples of suitable fusion partners include a polypeptide that provides for methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity,

kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity, or demyristoylation activity.

[0504] A suitable fusion partner can include a polypeptide that directly provides for increased transcription of the target nucleic acid. For example, a transcription activator or a fragment thereof, a protein or fragment thereof that recruits a transcription activator, or a small molecule/drug-responsive transcription regulator. A suitable fusion partner can include a polypeptide that directly provides for decreased transcription of the target nucleic acid. For example, a transcription repressor or a fragment thereof, a protein or fragment thereof that recruits a transcription repressor, or a small molecule/drug-responsive transcription regulator.

[0505] The heterologous sequence or fusion partner can be fused to the C-terminus, N-terminus, or an internal portion (i.e., a portion other than the N- or C-terminus) of the dead Cas protein.

Methods for Delivery

[0506] Any suitable delivery method can be used for introducing the compositions and molecules of the disclosure into a host cell or organelle. The compositions (e.g., Cas protein, polynucleotide-guided polypeptide, guide polynucleic acid, donor polynucleotide) can be delivered simultaneously or temporally separated. The choice of method of genetic modification can be dependent on the type of cell being transformed and/or the circumstances under which the transformation is taking place (e.g., in vitro, ex vivo, or in vivo).

[0507] A method of delivery can involve contacting a target polynucleotide or introducing into a cell (or a population of cells) one or more nucleic acids comprising nucleotide sequences encoding the compositions of the disclosure. Suitable nucleic acids comprising nucleotide sequences encoding the compositions of the disclosure can include expression vectors, where an expression vector comprising a nucleotide sequence encoding one or more compositions of the disclosure can be a recombinant expression vector.

[0508] Non-limiting examples of delivery methods or transformation include, for example, viral or bacteriophage infection, transfection, conjugation, protoplast fusion, lipofection, electroporation, calcium phosphate precipitation, polyethyleneimine (PEI)-mediated transfection, DEAE-dextran mediated transfection, liposome-mediated transfection, particle gun technology, calcium phosphate precipitation, direct micro injection, and nanoparticle-mediated nucleic acid delivery.

[0509] In some aspects, the present disclosure provides methods comprising delivering one or more polynucleotides, or one or more vectors as described herein, or one or more transcripts thereof, and/or one or proteins transcribed therefrom, to a host cell or organelle. In some aspects, the disclosure further provides cells produced by such methods, and organisms (such as animals, plants, or fungi) and organelles comprising or produced from such cells. In some embodiments, a Cas protein in combination with, and optionally complexed with, a guide sequence can be delivered to a cell or organelle.

[0510] Viral and non-viral based gene transfer methods can be used to introduce nucleic acids. Such methods can be used to administer nucleic acids encoding compositions of

the disclosure to cells in culture, or in a host organism. Non-viral vector delivery systems can include DNA plasmids, RNA (e.g. a transcript of a vector described herein), naked nucleic acid, and nucleic acid complexed with a delivery vehicle, such as a liposome. Viral vector delivery systems can include DNA and RNA viruses, which can have either episomal or integrated genomes after delivery to the cell.

[0511] Methods of non-viral delivery of nucleic acids can include lipofection, nucleofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides can be used. Delivery can be to cells (e.g. in vitro or ex vivo administration) or target tissues (e.g. in vivo administration). The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, can be used.

[0512] RNA or DNA viral based systems can be used to target specific cells and trafficking the viral payload to an organelle of the cell. Viral vectors can be administered directly (in vivo) or they can be used to treat cells in vitro, and the modified cells can optionally be administered (ex vivo). Viral based systems can include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Integration in the host genome can occur with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, which can result in long term expression of the inserted transgene. High transduction efficiencies can be observed in many different cell types and target tissues.

[0513] The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that can transduce or infect non-dividing cells and produce high viral titers. Selection of a retroviral gene transfer system can depend on the target tissue. Retroviral vectors can comprise cis-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum cis-acting LTRs can be sufficient for replication and packaging of the vectors, which can be used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Retroviral vectors can include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immunodeficiency virus (SIV), human immunodeficiency virus (HIV), and combinations thereof.

[0514] An adenoviral-based systems can be used. Adenoviral-based systems can lead to transient expression of the transgene. Adenoviral based vectors can have high transduction efficiency in cells and may not require cell division. High titer and levels of expression can be obtained with adenoviral based vectors. Adeno-associated virus ("AAV") vectors can be used to transduce cells with target nucleic acids, e.g., in the in vitro production of nucleic acids and peptides, and for in vivo and ex vivo gene therapy procedures.

[0515] Packaging cells can be used to form virus particles that can infect a host cell. Such cells can include 293 cells, (e.g., for packaging adenovirus), and psi.2 cells or PA317 cells (e.g., for packaging retrovirus). Viral vectors can be generated by producing a cell line that packages a nucleic acid vector into a viral particle. The vectors can contain the

minimal viral sequences required for packaging and subsequent integration into a host. The vectors can contain other viral sequences being replaced by an expression cassette for the polynucleotide(s) to be expressed. The missing viral functions can be supplied in trans by the packaging cell line. For example, AAV vectors can comprise ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA can be packaged in a cell line, which can contain a helper plasmid encoding the other AAV genes, namely rep and cap, while lacking ITR sequences. The cell line can also be infected with adenovirus as a helper. The helper virus can promote replication of the AAV vector and expression of AAV genes from the helper plasmid. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus can be more sensitive than AAV. Additional methods for the delivery of nucleic acids to cells can be used, for example, as described in US20030087817, incorporated herein by reference.

[0516] A host cell can be transiently or non-transiently transfected with one or more vectors described herein. A cell can be transfected as it naturally occurs in a subject. A cell can be taken or derived from a subject and transfected. A cell can be derived from cells taken from a subject, such as a cell line. In some embodiments, a cell transfected with one or more vectors described herein can be used to establish a new cell line comprising one or more vector-derived sequences. In some embodiments, a cell transiently transfected with the compositions of the disclosure (such as by transient transfection of one or more vectors, or transfection with RNA), and modified through the activity of a CRISPR complex, can be used to establish a new cell line comprising cells containing the modification but lacking any other exogenous sequence.

[0517] Any suitable vector compatible with the host cell can be used with the methods of the disclosure. Non-limiting examples of vectors include pXT1, pSG5 (Stratagene™), pSVK3, pBPV, pMSG, and pSVLSV40 (Pharmacia™).

[0518] In some embodiments, a nucleotide sequence encoding a guide nucleic acid and/or Cas protein can be operably linked to a control element, e.g., a transcriptional control element, such as a promoter. In some embodiments, a nucleotide sequence encoding a guide nucleic acid and/or a Cas protein can be operably linked to multiple control elements that allow expression of the nucleotide sequence encoding a guide nucleic acid and/or a Cas protein or chimera.

[0519] Depending on the host/vector system utilized, any of a number of suitable transcription and translation control elements, including constitutive and inducible promoters, transcription enhancer elements, transcription terminators, etc. may be used in the expression vector (e.g., U6 promoter, H1 promoter, etc.; see above).

[0520] In some embodiments, compositions of the disclosure can be provided as RNA. In such cases, the compositions of the disclosure can be produced by direct chemical synthesis or may be transcribed in vitro from a DNA. The compositions of the disclosure can be synthesized in vitro using an RNA polymerase enzyme (e.g., T7 polymerase, T3 polymerase, SP6 polymerase, etc.). Once synthesized, the RNA can directly contact a target polynucleic acid (e.g., target DNA) or can be introduced into a cell using any suitable technique for introducing nucleic acids into cells (e.g., microinjection, electroporation, transfection, etc.).

[0521] Nucleotides encoding a guide nucleic acid (introduced either as DNA or RNA) and/or a Cas protein (introduced as DNA or RNA) can be provided to the cells using a suitable transfection technique. Nucleic acids encoding the compositions of the disclosure may be provided on vectors or cassettes (e.g., DNA vectors). Many vectors, e.g. plasmids, cosmids, minicircles, phage, viruses, etc., useful for transferring nucleic acids into target cells are available. The vectors comprising the nucleic acid(s) can be maintained episomally, e.g. as plasmids, minicircle DNAs, viruses such as cytomegalovirus, adenovirus, etc., or they may be integrated into the target cell genome, through homologous recombination or random integration, e.g. retrovirus-derived vectors such as MMLV, HIV-1, and ALV.

[0522] A Cas protein can be provided to cells as a polypeptide. Such a protein may optionally be fused to a polypeptide domain that increases solubility of the product. The domain may be linked to the polypeptide through a defined protease cleavage site, e.g. a TEV sequence, which can be cleaved by TEV protease. The linker may also include one or more flexible sequences, e.g. from 1 to 10 glycine residues. In some embodiments, the cleavage of the fusion protein can be performed in a buffer that maintains solubility of the product, e.g. in the presence of from 0.5 to 2 M urea, in the presence of polypeptides and/or polynucleotides that increase solubility, and the like. Domains of interest include endosomolytic domains, e.g. influenza HA domain; and other polypeptides that aid in production, e.g. IF2 domain, GST domain, GRPE domain, and the like. The polypeptide may be formulated for improved stability. For example, the peptides may be PEGylated, where the polyethyleneoxy group provides for enhanced lifetime in the blood stream.

[0523] The compositions of the disclosure may be fused to a polypeptide permeant domain to promote uptake by the cell. A number of permeant domains can be used in the non-integrating polypeptides of the present disclosure, including peptides, peptidomimetics, and non-peptide carriers. For example, a permeant peptide may be derived from the third alpha helix of *Drosophila melanogaster* transcription factor Antennapedia, referred to as penetratin, which comprises the amino acid sequence RQIKIWFQNRRMK-WKK (SEQ ID NO: 172). As another example, the permeant peptide can comprise the HIV-1 tat basic region amino acid sequence, which may include, for example, amino acids 49-57 of naturally-occurring tat protein. Other permeant domains can include poly-arginine motifs, for example, the region of amino acids 34-56 of HIV-1 rev protein, nona-arginine, and octa-arginine. The nona-arginine (R9) sequence can be used. The site at which the fusion can be made may be selected in order to optimize the biological activity, secretion or binding characteristics of the polypeptide.

[0524] The compositions of the disclosure may be produced in vitro or by host cells, and it may be further processed by unfolding, e.g. heat denaturation, DTT reduction, etc. and may be further refolded.

[0525] The compositions of the disclosure may be prepared by in vitro synthesis. Various commercial synthetic apparatuses can be used. By using synthesizers, naturally occurring amino acids can be substituted with unnatural amino acids. The particular sequence and the manner of preparation can be determined by convenience, economics, and purity required.

[0526] The compositions of the disclosure may also be isolated and purified in accordance with recombinant synthesis methods. A lysate may be prepared of the expression host and the lysate purified using HPLC, exclusion chromatography, gel electrophoresis, affinity chromatography, or other purification technique. The compositions can comprise, for example, at least 20% by weight of the desired product, at least about 75% by weight, at least about 95% by weight, and for therapeutic purposes, for example, at least about 99.5% by weight, in relation to contaminants related to the method of preparation of the product and its purification. The percentages can be based upon total protein.

[0527] The compositions of the disclosure, whether introduced as nucleic acids or polypeptides, can be provided to the cells for about 30 minutes to about 24 hours, e.g., 1 hour, 1.5 hours, 2 hours, 2.5 hours, 3 hours, 3.5 hours, 4 hours, 5 hours, 6 hours, 7 hours, 8 hours, 12 hours, 16 hours, 18 hours, 20 hours, or any other period from about 30 minutes to about 24 hours, which can be repeated with a frequency of about every day to about every 4 days, e.g., every 1.5 days, every 2 days, every 3 days, or any other frequency from about every day to about every four days. The compositions may be provided to the subject cells one or more times, e.g. one time, twice, three times, or more than three times, and the cells allowed to incubate with the agent(s) for some amount of time following each contacting event e.g. 16-24 hours, after which time the media can be replaced with fresh media and the cells can be cultured further.

[0528] In cases in which two or more different targeting complexes are provided to the cell (e.g., two different guide nucleic acids that are complementary to different sequences within the same or different target polynucleic acid (e.g., target DNA)), the complexes may be provided simultaneously (e.g. as two polypeptides and/or nucleic acids), or delivered simultaneously. Alternatively, they may be provided consecutively, e.g. the targeting complex being provided first, followed by the second targeting complex, etc. or vice versa.

[0529] An effective amount of the compositions of the disclosure can be provided to the target polynucleic acid (e.g., target DNA) or cells. An effective amount can be the amount to induce, for example, at least about a 2-fold change (increase or decrease) or more in the amount of target nucleic acid modulation (e.g., expression) observed between two homologous sequences relative to a negative control, e.g. a cell contacted with an empty vector or irrelevant polypeptide. An effective amount or dose can induce, for example, about 2-fold change, about 3-fold change, about 4-fold change, about a 7-fold, about 8-fold increase, about 10-fold, about 50-fold, about 100-fold, about 200-fold, about 500-fold, about 700-fold, about 1000-fold, about 5000-fold, or about 10,000-fold change in target gene modulation (e.g., expression). The amount of target gene modulation may be measured by any suitable method.

[0530] Contacting the cells with a composition of the disclosure can occur in any culture media and under any culture conditions that promote the survival of the cells. For example, cells may be suspended in any appropriate nutrient medium. The culture may contain growth factors to which the cells are responsive. Growth factors can be molecules that can promote survival, growth and/or differentiation of cells (e.g., in culture, in the intact tissue), for example,

through specific effects on a transmembrane receptor. Growth factors can include polypeptides and non-polypeptide factors.

[0531] In numerous embodiments, the chosen delivery system can be targeted to specific cell types. In some cases, tissue- or cell-targeting of the delivery system can be achieved by binding the delivery system to tissue- or cell-specific markers, such as cell surface proteins. Viral and non-viral delivery systems can be customized to target tissue or cell-types of interest.

Genome Editing Using a Polynucleotide Guided Polypeptide System

[0532] As described herein, the polynucleotide guided polypeptide system can be used in combination with a co-delivered polynucleotide modification template to allow for editing of an organellar nucleotide sequence of interest. Also, as described herein, for each embodiment that uses an RNA guided polypeptide system, a similar polynucleotide guided polypeptide system can be deployed where the guide polynucleotide may not solely comprise ribonucleic acids but wherein the guide polynucleotide comprises a combination of RNA-DNA molecules or solely comprises DNA molecules.

[0533] Genome modification methods can rely on the homologous recombination system. Homologous recombination (HR) can provide molecular means for finding genomic DNA sequences of interest and modifying them according to the experimental specifications. Homologous recombination can be enhanced by introducing double-strand breaks (DSBs) at selected endonuclease target sites. Described herein is the use of a polynucleotide guided polypeptide system which can provide flexible genome cleavage specificity and can result in a high frequency of double-strand breaks at an organellar DNA target site. This specific cleavage can enable efficient gene editing of a nucleotide sequence of interest. The nucleotide sequence of interest to be edited can be located within or outside the target site recognized and/or cleaved by a polynucleotide guided polypeptide (e.g., a Cas polypeptide).

[0534] The term “polynucleotide modification template” can refer to a polynucleotide that can comprise at least one nucleotide modification when compared to the nucleotide sequence to be edited. A nucleotide modification can be at least one nucleotide substitution, addition or deletion. Examples of minor genome modifications created by use of a polynucleotide modification template include creation of a mutant allele (e.g., antibiotic resistant rRNA gene) and removal of a target site for a polynucleotide guided polypeptide. Optionally, the polynucleotide modification template can be flanked by homologous nucleotide sequences, wherein the flanking homologous nucleotide sequences can provide sufficient homology to the desired nucleotide sequence to be edited. The polynucleotide modification template can be a donor polynucleotide.

[0535] In one embodiment, the disclosure provides a method for editing a nucleotide sequence in the organellar genome of a cell. The method can comprise providing a guide polynucleotide (e.g., guide RNA), a polynucleotide modification template, and at least one polynucleotide guided polypeptide (e.g., Cas polypeptide) to an organelle. The polynucleotide guided polypeptide can introduce a single-strand or double-strand break at a target sequence in the organellar genome of the cell. The polynucleotide modi-

modification template can include at least one nucleotide modification of said nucleotide sequence. Cells include, but are not limited to, human, animal, bacterial, fungal, insect, and plant cells as well as organisms and tissues, e.g., plants and seeds, produced by the methods described herein. Cell can be an isolated and purified human cell. The nucleotide to be edited can be located within or outside a target site recognized and cleaved by a polynucleotide guided polypeptide. In one embodiment, the at least one nucleotide modification may not be a modification at a target site recognized and cleaved by a polynucleotide guided polypeptide. In another embodiment, there can be at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 900 or 1000 nucleotides between the at least one nucleotide to be edited and the organellar DNA target site.

[0536] In another embodiment, the disclosure provides a method for editing a nucleotide sequence in the organellar genome of a cell. The method can comprise providing a guide polynucleotide (e.g., guide RNA), a polynucleotide modification template and at least one polynucleotide guided polypeptide (e.g., Cas polypeptide) to an organelle, wherein said guide polynucleotide and said polynucleotide guided polypeptide can form a complex that can enable the polynucleotide guided polypeptide to introduce a single-strand or double-strand break at an organellar target site, wherein said polynucleotide modification template comprises at least one nucleotide modification of said nucleotide sequence.

[0537] In another embodiment, the disclosure provides a method for editing a nucleotide sequence in the organellar genome of a plant cell. The method can comprise introducing a guide polynucleotide (e.g., guide RNA), a polynucleotide modification template, and at least one organelle codon-optimized polynucleotide guided polypeptide (e.g., Cas9 polypeptide) into an organelle, wherein the organelle optimized polynucleotide guided polypeptide can introduce a single-strand or double-strand break at an organellar target sequence, wherein said polynucleotide modification template includes at least one nucleotide modification of said nucleotide sequence.

[0538] The nucleotide sequence to be edited can be a sequence that can be endogenous, artificial, pre-existing, or transgenic to the cell that is being edited. For example, the nucleotide sequence in the organellar genome of a cell can be a transgene that is stably incorporated into the organellar genome of a cell. Editing of such transgene may result in a further desired phenotype or genotype. The nucleotide sequence in the genome of a cell can also be a mutated or pre-existing sequence that was either endogenous or artificial from origin such as an endogenous gene or a mutated gene of interest.

[0539] In one embodiment, the region of interest can be flanked by two independent guide polynucleotide/polypeptide target sequences. Cutting can be done concurrently. The deletion event can be the repair of the two chromosomal ends without the region of interest. Alternative results can include inversions of the region of interest, mutations at the cut sites and duplication of the region of interest.

[0540] Methods for Identifying at Least One Plant Cell Comprising in its Organellar Genome a Polynucleotide of Interest Integrated at the Target Site.

[0541] Further provided are methods for identifying at least one plant cell comprising in its organellar genome a polynucleotide of interest integrated at the target site. A

donor polynucleotide can comprise a polynucleotide of interest. A polynucleotide of interest can be integrated at a target site in a cell (e.g., genome). A variety of methods can be used for identifying those plant cells with insertion into the genome at or near to the target site without using a screenable marker phenotype. Such methods can be viewed as directly analyzing a target sequence to detect any change in the target sequence, including but not limited to PCR methods, sequencing methods, nuclease digestion, Southern blots, and any combination thereof.

[0542] The method can also comprise recovering a plant from the plant cell comprising a polynucleotide of interest integrated into its organellar genome. The plant may be sterile or fertile. Any polynucleotide of interest can be provided, integrated into the plant organellar genome at the target site, and expressed in a plant.

[0543] Polynucleotides of interest can be reflective of the commercial markets and interests of those involved in the development of the crop. Crops and markets of interest change, and as developing nations open up world markets, new crops and technologies can emerge also. In addition, as our understanding of agronomic traits and characteristics such as yield, stress tolerance and heterosis increase, the choice of genes for transformation can change accordingly.

[0544] Polynucleotides/polypeptides of interest include, but are not limited to, herbicide-tolerance coding sequences, insecticidal coding sequences, nematocidal coding sequences, antimicrobial coding sequences, antifungal coding sequences, antiviral coding sequences, abiotic and biotic stress tolerance coding sequences, or sequences modifying plant traits such as yield, grain quality, nutrient content, starch quality and quantity, nitrogen fixation and/or utilization, and oil content and/or composition. polynucleotides of interest can include, but are not limited to, genes that improve crop yield, polypeptides that improve desirability of crops, genes encoding proteins conferring resistance to abiotic stress, such as drought, nitrogen, temperature, salinity, toxic metals or trace elements, or those conferring resistance to toxins such as pesticides and herbicides, or to biotic stress, such as attacks by fungi, viruses, bacteria, insects, and nematodes, and development of diseases associated with these organisms. Genes of interest can include, for example, those genes involved in information, such as zinc fingers, those involved in communication, such as kinases, and those involved in housekeeping, such as heat shock proteins. Polynucleotides of interest can include genes encoding important traits for agronomics, insect resistance, disease resistance, herbicide resistance, fertility or sterility, grain characteristics, and commercial products. Genes of interest can include, generally, those involved in oil, starch, carbohydrate, or nutrient metabolism as well as those affecting photosynthesis, photorespiration and ATP metabolism.

[0545] Commercial traits can also be obtained by expression of proteins encoded on a polynucleotide. A commercial use of transformed plants can be the production of polymers and bioplastics. Polynucleotides of interest can include genes such as β -ketothiolase, PHBase (polyhydroxybutyrate synthase), and acetoacetyl-CoA reductase can facilitate expression of polyhydroxyalkanoates (PHAs).

[0546] Polynucleotides/polypeptides that can influence amino acid biosynthesis include, for example, anthranilate synthase (AS; EC 4.1.3.27) which can catalyze the first reaction branching from the aromatic amino acid pathway to the biosynthesis of tryptophan in plants, fungi, and bacteria.

In plants, the chemical processes for the biosynthesis of tryptophan can be compartmentalized in the chloroplast. Additional donor sequences of interest can include Chorismate Pyruvate Lyase (CPL) which can refer to a gene encoding an enzyme can which catalyze the conversion of chorismate to pyruvate and pHBA. One example of CPL gene is from *E. coli* and bears the GenBank accession number M96268.

[0547] Polynucleotide sequences of interest may encode proteins involved in providing disease or pest resistance. By “disease resistance” or “pest resistance” can be intended that the plants can avoid the harmful symptoms that are the outcome of the plant-pathogen interactions. Pest resistance genes may encode resistance to pests that have great yield drag such as rootworm, cutworm, European Corn Borer, and the like. Disease resistance and insect resistance genes such as lysozymes or cecropins for antibacterial protection, or proteins such as defensins, glucanases or chitinases for antifungal protection, or *Bacillus thuringiensis* endotoxins, protease inhibitors, collagenases, lectins, or glycosidases for controlling nematodes or insects are all examples of useful gene products. Genes encoding disease resistance traits include detoxification genes, such as against fumonisin; avirulence (avr) and disease resistance (R) genes; and the like. Insect resistance genes may encode resistance to pests that have great yield drag such as rootworm, cutworm, European Corn Borer, and the like. Such genes include, for example, *Bacillus thuringiensis* toxic protein genes; and the like.

[0548] An “herbicide resistance protein” or a protein resulting from expression of an “herbicide resistance-encoding nucleic acid molecule” can include proteins that confer upon a cell the ability to tolerate a higher concentration of an herbicide than cells that do not express the protein, or to tolerate a certain concentration of an herbicide for a longer period of time than cells that do not express the protein. Herbicide resistance traits may be introduced into plants by genes coding for resistance to herbicides that act to inhibit the action of acetolactate synthase (ALS), for example, the sulfonylurea-type herbicides, genes coding for resistance to herbicides that can act to inhibit the action of glutamine synthase, such as phosphinothricin or basta (e.g., the bar gene), glyphosate (e.g., the EPSP synthase gene and the GAT gene), HPPD inhibitors (e.g., the HPPD gene) or other such genes. The bar gene can encode resistance to the herbicide basta, the aadA can encode resistance to spectinomycin and streptomycin, the nptII gene can encode resistance to the antibiotics kanamycin and geneticin, and certain ALS-gene mutants can encode resistance to the herbicide chlorsulfuron.

[0549] Sterility genes can also be encoded in an expression cassette or integrated into the genome. Sterility genes can provide an alternative to physical detasseling. Examples of genes used in such ways include male fertility genes such as MS26, MS45, or MSCA1. Maize plants (*Zea mays* L.) can be bred by both self-pollination and cross-pollination techniques. Maize can have male flowers, located on the tassel, and female flowers, located on the ear, on the same plant. It can self-pollinate (“selfing”) or cross pollinate. Natural pollination can occur in maize when wind blows pollen from the tassels to the silks that protrude from the tops of the incipient ears. Pollination may be readily controlled by suitable methods. The development of maize hybrids can require the development of homozygous inbred lines, the

crossing of these lines, and the evaluation of the crosses. Pedigree breeding and recurrent selections are two of the breeding methods that can be used to develop inbred lines from populations. Breeding programs can combine desirable traits from two or more inbred lines or various broad-based sources into breeding pools from which new inbred lines are developed by selfing and selection of desired phenotypes. A hybrid maize variety can be a cross of two such inbred lines, each of which may have one or more desirable characteristics lacked by the other or which complement the other. The new inbreds can be crossed with other inbred lines and the hybrids from these crosses can be evaluated to determine which have commercial potential. The hybrid progeny of the first generation can be designated F1. The F1 hybrid can be more vigorous than its inbred parents. This hybrid vigor, or heterosis, can be manifested in many ways, including increased vegetative growth and increased yield.

[0550] Hybrid maize seed can be produced by a male sterility system incorporating manual detasseling. To produce hybrid seed, the male tassel can be removed from the growing female inbred parent, which can be planted in various alternating row patterns with the male inbred parent. Consequently, providing that there is sufficient isolation from sources of foreign maize pollen, the ears of the female inbred can be fertilized only with pollen from the male inbred. The resulting seed can therefore be hybrid (F1) and can form hybrid plants.

[0551] Field variation impacting plant development can result in plants tasseling after manual detasseling of the female parent is completed. Or, a female inbred plant tassel may not be completely removed during the detasseling process. In any event, the result can be that the female plant can successfully shed pollen and some female plants can be self-pollinated. This can result in seed of the female inbred being harvested along with the hybrid seed which can be normally produced. Female inbred seed may not exhibit heterosis and therefore may not be as productive as F1 seed. In addition, the presence of female inbred seed can represent a germplasm security risk for the company producing the hybrid.

[0552] Alternatively, the female inbred can be mechanically detasseled by machine. Mechanical detasseling can be approximately as reliable as hand detasseling, but may be faster and less costly. However, most detasseling machines can produce more damage to the plants than hand detasseling. Thus, no form of detasseling may be presently entirely satisfactory, and a need continues to exist for alternatives which further reduce production costs and to eliminate self-pollination of the female parent in the production of hybrid seed.

[0553] One method to convey male sterility without mechanical detasseling can be the use of cytoplasmic male sterility (CMS) genes. Chimeric mitochondrial ORFs can be found to lead to male sterility, producing unisex-female plants. The methods described herein could be used to introduce custom-designed, CMS ORFs into mitochondria of maize elite inbred lines. Additionally, these methods can provide a means to introduce the CMS system into other crops; e.g., rice, wheat and soybean.

[0554] The donor polynucleotide may also encode an RNA or double-stranded RNA that can be complementary to a target gene from a plant pest or plant pathogen. A method of alleviating pest infestation of plants can comprise, for example, a) identifying a DNA sequence from said pest

which can be critical either for its survival, growth, proliferation or reproduction, b) cloning said sequence or a fragment thereof in a suitable vector relative to one or more promoters that can transcribe said sequence to RNA or dsRNA upon binding of an appropriate transcription factor to said promoters, and/or c) introducing said vector into the plant. The plant pest can be a nematode. Another method for alleviating pest infestation can include, for example, providing: a) DNA sequences which when transcribed yield a double-stranded RNA molecule that can reduce the expression of an essential gene of a plant sap-sucking insect; b) methods of using such DNA sequences and plants or plant cells transformed with such DNA sequences; and c) the use of cationic oligopeptides that facilitate the entry of dsRNA or siRNA molecules in insect cells, such as plant sap-sucking insect cells.

[0555] The donor polynucleotide may comprise and/or lead to expression of antisense sequences complementary to at least a portion of the messenger RNA (mRNA) for a targeted gene sequence of interest; e.g., a target gene from a plant pest or plant pathogen. Antisense nucleotides can be constructed to hybridize with the corresponding mRNA. Antisense nucleotides can be targeted to bind a splicing site on a pre-mRNA and modify the exon content of an mRNA, thereby modulating (e.g., disrupting) expression of a target gene.

[0556] Modifications of the antisense sequences may be made as long as the sequences hybridize to and interfere with expression of the corresponding mRNA. In this manner, antisense constructions having 70%, 80%, or 85% sequence identity to the corresponding antisense sequences may be used. Furthermore, portions of the antisense nucleotides may be used to disrupt the expression of the target gene. Generally, sequences of at least 50 nucleotides, 100 nucleotides, 200 nucleotides, or greater may be used.

[0557] The donor polynucleotide can also be a phenotypic marker. A phenotypic marker can be screenable or a selectable marker that includes visual markers and selectable markers whether it is a positive or negative selectable marker. Any phenotypic marker can be used. Specifically, a selectable or screenable marker can comprise a DNA segment that can allow one to identify, or select for or against a molecule or a cell that contains it, e.g., under particular conditions. These markers can encode an activity, such as, but not limited to, production of RNA, peptide, or protein, or can provide a binding site for RNA, peptides, proteins, inorganic and organic compounds or compositions and the like.

[0558] Examples of selectable markers include, but are not limited to, DNA segments that comprise restriction enzyme sites; DNA segments that encode products which provide resistance against otherwise toxic compounds including antibiotics, such as, spectinomycin, ampicillin, kanamycin, tetracycline, Basta, neomycin phosphotransferase II (NEO) and hygromycin phosphotransferase (HPT); DNA segments that encode products which are otherwise lacking in the recipient cell (e.g., tRNA genes, auxotrophic markers); DNA segments that encode products which can be readily identified (e.g., phenotypic markers such as β -galactosidase, GUS; fluorescent proteins such as green fluorescent protein (GFP), cyan (CFP), yellow (YFP), red (RFP), and cell surface proteins); the generation of new primer sites for PCR (e.g., the juxtaposition of two DNA sequence not previously juxtaposed), the inclusion of DNA sequences not acted upon

or acted upon by a restriction endonuclease or other DNA modifying enzyme, chemical, etc.; and, the inclusion of a DNA sequences required for a specific modification (e.g., methylation) that allows its identification.

[0559] Additional selectable markers include genes that can confer resistance to herbicidal compounds, such as glyphosate, sulfonylureas, glufosinate ammonium, bromoxynil, imidazolinones, and 2,4-dichlorophenoxyacetate (2,4-D).

[0560] Commercial traits can also be encoded on a gene or genes that could increase for example, starch for ethanol production, or provide expression of proteins. Another important use of transformed plants can be the production of polymers and bioplastics. Genes such as β -Ketothiolase, PHBase (polyhydroxybutyrate synthase), and acetoacetyl-CoA reductase can facilitate expression of polyhydroxyalkanoates (PHAs).

[0561] Exogenous products include plant enzymes and products as well as those from other sources including prokaryotes and other eukaryotes. Such products include enzymes, cofactors, hormones, and the like. The level of proteins, particularly modified proteins having improved amino acid distribution to improve the nutrient value of the plant, can be increased. This can be achieved by the expression of such proteins having enhanced amino acid content.

[0562] The transgenes, recombinant DNA molecules, DNA sequences of interest, and donor polynucleotides can comprise one or more DNA sequences for gene silencing of a target gene; e.g., a target gene in a plant pest or plant pathogen. Methods for gene silencing involving the expression of DNA sequences in plant can include, but are not limited to, cosuppression, antisense suppression, double-stranded RNA (dsRNA) interference, hairpin RNA (hpRNA) interference, intron-containing hairpin RNA (ihpRNA) interference, transcriptional gene silencing, and microRNA (miRNA) interference.

[0563] In one embodiment, the targeted mutation can involve use of a double-strand-break-inducing agent that can induce a double-strand break in the DNA of the target sequence.

[0564] In one embodiment, the targeted mutation can be the result of a guide polynucleotide/polypeptide induced gene editing as described herein. The guide polynucleotide/polypeptide induced targeted mutation can occur in a nucleotide sequence that can be located within or outside a genomic target site that can be recognized and cleaved by a polynucleotide guided polypeptide.

[0565] In certain embodiments, a fertile plant can be a plant that can produce viable male and female gametes and can be self-fertile. Such a self-fertile plant can produce a progeny plant without the contribution from any other plant of a gamete and the genetic material contained therein. Other embodiments may involve the use of a plant that may not be self-fertile, for example, because the plant may not produce male gametes, or female gametes, or both, that are viable or otherwise capable of fertilization. As used herein, a "male sterile plant" can be a plant that does not produce male gametes that are viable or otherwise capable of fertilization. As used herein, a "female sterile plant" can be a plant that does not produce female gametes that are viable or otherwise capable of fertilization. Male-sterile and female-sterile plants can be female-fertile and male-fertile, respectively. A male fertile (but female sterile) plant can produce viable progeny when crossed with a female fertile plant and that a

female fertile (but male sterile) plant can produce viable progeny when crossed with a male fertile plant.

Breeding Methods and Methods for Selecting Plants Utilizing a Two Component RNA Guide and Cas Polypeptide System

[0566] The present disclosure can find use in the breeding of plants comprising one or more transgenic traits. Transgenic traits can be randomly inserted throughout the plant genome as a consequence of transformation systems based on *Agrobacterium*, biolistics, or other suitable procedures. Directed transgene insertion can be used. Site-specific integration (SSI) can enable the targeting of a transgene to the same chromosomal location as a previously inserted transgene. Custom-designed meganucleases and custom-designed zinc finger meganucleases can be used to design nucleases to target specific chromosomal locations, and these reagents can allow the targeting of transgenes at the chromosomal site cleaved by these nucleases.

[0567] Genetic engineering of eukaryotic genomes, e.g. plant genomes, using homing endonucleases, meganucleases, zinc finger nucleases, and transcription activator-like effector nucleases (TALENs) can require de novo protein engineering for every new target locus. The highly specific, polynucleotide guided polypeptide system (e.g., guide RNA/Cas polypeptide system) described herein, can be more easily customizable and can be more useful when modification of many different target sequences is the goal. The polynucleotide guided polypeptide system can be a two component system, for example, with its constant protein component, the polynucleotide guided polypeptide (e.g., Cas polypeptide), and its variable and easily reprogrammable targeting component, the guide polynucleotide (e.g., guide RNA or crRNA).

[0568] The polynucleotide guided polypeptide system described herein can be especially useful for genome engineering in circumstances where endonuclease off-target cutting can be toxic to the targeted cells. In one embodiment of the polynucleotide guided polypeptide system described herein, the constant component, a polynucleotide encoding an organelle targeted polynucleotide guided polypeptide, can be stably integrated into the nuclear genome of the cell. The polynucleotide can encode a modified polynucleotide guided polypeptide comprising an enzymatically active polynucleotide guided polypeptide (e.g., Cas polypeptide) fused to an organellar transport sequence (e.g., a mitochondrial targeting peptide or a chloroplast targeting peptide). Expression of the polynucleotide encoding the modified polynucleotide guided polypeptide can be under control of a promoter. The promoter can be a constitutive promoter, a tissue-specific promoter or an inducible promoter, e.g. a temperature-inducible, stress-inducible, developmental stage inducible, or chemically inducible promoter. In the absence of the variable component (e.g., the guide RNA or crRNA), the polynucleotide guided polypeptide may not cut the target nucleic acid. In the absence of the variable component (e.g., the guide RNA or crRNA) the presence of the polynucleotide guided polypeptide in the plant cell may have little or no consequence. A polynucleotide guided polypeptide system can be used to create and/or maintain a cell line or transgenic organism capable of efficient expression of the polynucleotide guided polypeptide. Expression of the polynucleotide guided polypeptide in the cell line or transgenic organism may have little or no consequence to

cell viability. In order to induce cutting at desired genomic sites to achieve targeted genetic modifications, guide polynucleotides (e.g., guide RNAs or crRNAs) can be introduced by a variety of methods into cells containing the stably-integrated and expressed expression cassette for the polynucleotide guided polypeptide. For example, guide polynucleotide (e.g., guide RNAs or crRNAs) can be chemically or enzymatically synthesized, and introduced into the polynucleotide guided polypeptide expressing cells via direct delivery methods such as particle bombardment or electroporation. A guide polynucleic acid may be fused to an RNA molecule that allows for transport into an organelle. Alternatively, a guide polynucleic acid may be fused to an RNA molecule that allows for binding to a protein that facilitates transport into the organelle.

[0569] Alternatively, genes that can efficiently express guide polynucleotides (e.g., guide RNAs or crRNAs) in the target cells can be synthesized chemically, enzymatically or in a biological system. These genes can be introduced into the polynucleotide guided polypeptide expressing cells, for example, via direct delivery methods such as particle bombardment, electroporation or biological delivery methods such as *Agrobacterium*-mediated DNA delivery.

[0570] One embodiment of the disclosure can be a method for selecting a plant comprising an altered target site in its organellar genome. The method can comprise a) obtaining a first plant that can comprise at least one polynucleotide guided polypeptide (e.g., Cas polypeptide) that can be transported into an organelle and can introduce a single-strand or double strand break at a target site in the organellar genome. In some cases, the polynucleotide guided polypeptide (e.g., dead Cas) may not cleave a target site. The method can further comprise b) obtaining a second plant comprising a guide polynucleotide (e.g., guide RNA) that can be transported into an organelle and can form a complex with the polynucleotide guided polypeptide of (a). The method can further comprise c) crossing the first plant of (a) with the second plant of (b). The method can further comprise d) evaluating the progeny of (c) for an alteration in the target site. The method can further comprise e) selecting a progeny plant that possesses the desired alteration of said target site. When an enzymatically inactive polynucleotide guided polypeptide is used, the method can comprise evaluating and selecting a progeny with altered target gene regulation or expression.

[0571] Another embodiment of the disclosure can be a method for selecting a plant comprising an altered target site in its organellar genome. The method can comprise: a) obtaining a first plant comprising at least one polynucleotide guided polypeptide (e.g., Cas polypeptide) that can be transported into an organelle and can introduce a single-strand or double strand break at a target site in the organellar genome. The method can further comprise b) obtaining a second plant comprising a guide polynucleotide (e.g., guide RNA) and a donor polynucleotide (e.g. donor DNA). The guide polynucleotide and donor polynucleotide (e.g. donor DNA) can be transported into the organelle. The guide polynucleotide can form a complex with the polynucleotide guided polypeptide of (a). The method can further comprise c) crossing the first plant of (a) with the second plant of (b). The method can further comprise d) evaluating the progeny of (c) for an alteration in the target site. The method can further comprise e) selecting a progeny plant that comprises the donor polynucleotide inserted at said target site.

[0572] Another embodiment of the disclosure can be a method for selecting a plant comprising an altered target site in its organellar genome. The method can comprise selecting at least one progeny plant that comprises an alteration at a target site in its organellar genome. The progeny plant can be a plant, for example, obtained by crossing a first plant expressing at least one polynucleotide guided polypeptide (e.g., Cas polypeptide) that can be transported into an organelle to a second plant comprising a guide polynucleotide (e.g., guide RNA) and optionally a donor polynucleotide (e.g. donor DNA), wherein said guide polynucleotide and said donor polynucleotide (e.g. donor DNA) can be transported into an organelle, wherein said polynucleotide guided polypeptide can introduce a single-strand or double strand break at said target site.

[0573] A suitable method can be used to identify those cells having an altered genome at or near a target site without using a screenable marker phenotype. Such methods can be viewed as directly analyzing a target sequence to detect any change in the target sequence, including but not limited to PCR methods, sequencing methods, nuclease digestion, Southern blots, and any combination thereof.

[0574] Proteins may be altered in various ways including amino acid substitutions, deletions, truncations, and insertions. For example, amino acid sequence variants of the protein(s) can be prepared by mutations in the DNA. Methods for mutagenesis and nucleotide sequence alterations can be used.

[0575] Guidance regarding amino acid substitutions not likely to affect biological activity of the protein can be determined.

[0576] Conservative substitutions, such as exchanging one amino acid with another having similar properties, can be carried out. Conservative deletions, insertions, and amino acid substitutions may not produce radical changes in the characteristics of the protein. The effect of any substitution, deletion, insertion, or combination thereof can be evaluated by screening assays. Assays for double-strand-break-inducing activity can measure, for example, the overall activity and specificity of the agent on DNA substrates containing target sites.

[0577] Sufficient homology or sequence identity can indicate that two polynucleotide sequences have sufficient structural similarity to act as substrates for a homologous recombination reaction. The structural similarity can include overall length of each polynucleotide fragment, and the sequence similarity of the polynucleotides. Sequence similarity can be described by the percent sequence identity over the whole length of the sequences, and/or by conserved regions comprising localized similarities such as contiguous nucleotides having 100% sequence identity, and percent sequence identity over a portion of the length of the sequences.

[0578] The amount of homology or sequence identity shared by a target and a donor polynucleotide can vary. For example, the length of sequence homology may be at least one of the following: 20 bp, 50 bp, 100 bp, 150 bp, 250 bp, 300 bp, 400 bp, 500 bp, 600 bp, 700 bp, 800 bp, 900 bp, 1000 bp, 1250 bp, 1500 bp, 1750 bp, 2000 bp, 2.5 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb or 10 kb. The amount of homology can also be described by percent sequence identity over the full aligned length of the two polynucleotides which includes percent sequence identity of at least any of the following: 50%, 55%, 60%, 65%, 70%, 71%, 72%, 73%,

74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100%. Sufficient homology can include any combination of polynucleotide length, global percent sequence identity, and optionally conserved regions of contiguous nucleotides or local percent sequence identity, for example sufficient homology can be described as a region of 75-150 bp having at least 80% sequence identity to a region of the target locus. Sufficient homology can also be described by the predicted ability of two polynucleotides to specifically hybridize under high stringency conditions.

[0579] A variety of methods can be used for the introduction of nucleotide sequences and polypeptides into an organism, including, for example, transformation, sexual crossing, and the introduction of the polypeptide, DNA, or mRNA into the cell. Methods for contacting, providing, and/or introducing a composition into various organisms can include but are not limited to, stable transformation methods, transient transformation methods, virus-mediated methods, and sexual breeding. Stable transformation can indicate that the introduced polynucleotide can integrate into the genome of the organism and can be inherited by progeny thereof. Transient transformation can indicate that the introduced composition can only temporarily be expressed or present in the organism.

[0580] Protocols for introducing polynucleotides and polypeptides into plants may vary depending on the type of plant or plant cell targeted for transformation, such as monocot or dicot. Suitable methods of introducing polynucleotides and polypeptides into plant cells and subsequent insertion into the plant genome include microinjection, meristem transformation, electroporation, *Agrobacterium*-mediated transformation, direct gene transfer, and ballistic particle acceleration.

[0581] Alternatively, polynucleotides may be introduced into plants by contacting plants with a virus or viral nucleic acids. Such methods can involve incorporating a polynucleotide within a viral DNA or RNA molecule. In some examples a polypeptide of interest may be initially synthesized as part of a viral polypeptide, which can be later processed by proteolysis *in vivo* or *in vitro* to produce the desired recombinant protein. Methods for introducing polynucleotides into plants and expressing a protein encoded therein, can involve viral DNA or RNA molecules. Transient transformation methods include, but are not limited to, the introduction of polypeptides, such as a double-strand break inducing agent, directly into the organism, the introduction of polynucleotides such as DNA and/or RNA polynucleotides, and the introduction of the RNA transcript, such as an mRNA encoding a double-strand break inducing agent, into the organism. Such methods include, for example, microinjection or particle bombardment.

[0582] DNA transformation of organellar genomes can be performed in, for example, plastids and mitochondria (e.g., yeast). Selectable marker genes can include, for example, photosynthesis (*atpB*, *tscA*, *psaA/B*, *petB*, *petA*, *ycf3*, *rpoA*, *rbcL*), antibiotic resistance (*rrnS*, *rrnL*, *aadA*, *nptII*, *aphA-6*), herbicide resistance (*psbA*, *bar*, *AHAS* (ALS), *EPSPS*, *HPPD*) and metabolism (*BADH*, *codA*, *ARG9*, *ASA2*) genes.

[0583] DNA transformation of, for example, the yeast nuclear genome can be facilitated by the development of shuttle vectors that can replicate in *E. coli* and yeast as

autonomous plasmids. Vector systems can include low-copy-number plasmids and integrative DNA through homologous recombination.

[0584] Methods of the invention can provide transformation efficiency into an organelle (e.g., mitochondria, plastids) of, for example, at least about: 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% transformation efficiency.

[0585] In one embodiment, an expression construct of the current disclosure may comprise a promoter operably linked to a nucleotide sequence encoding a Cas gene and a promoter operably linked to a guide RNA. The promoter can drive expression of an operably linked nucleotide sequence in a cell.

[0586] The cells having the introduced sequence may be grown or regenerated into plants. These plants may then be grown, and either pollinated with the same transformed strain or with a different transformed or untransformed strain, and the resulting progeny having the desired characteristic and/or comprising the introduced polynucleotide or polypeptide identified. Two or more generations may be grown to ensure that the polynucleotide can be stably maintained and inherited, and seeds harvested.

[0587] Any plant can be used, including monocot and dicot plants. Examples of monocot plants that can be used include, but are not limited to, corn (*Zea mays*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), millet (e.g., pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*)), maize, wheat (*Triticum aestivum*), sugarcane (*Saccharum* spp.), oats (*Avena*), barley (*Hordeum*), switchgrass (*Panicum virgatum*), pineapple (*Ananas comosus*), banana (*Musa* spp.), palm, ornamentals, turfgrasses, and other grasses. Examples of dicot plants that can be used include, but are not limited to, soybean (*Glycine max*), canola (*Brassica napus* and *B. campestris*), alfalfa (*Medicago sativa*), tobacco (*Nicotiana tabacum*), *Arabidopsis* (*Arabidopsis thaliana*), sunflower (*Helianthus annuus*), cotton (*Gossypium arboreum*), and peanut (*Arachis hypogaea*), tomato (*Solanum lycopersicum*), potato (*Solanum tuberosum*) etc.

[0588] The transgenes, recombinant DNA molecules, DNA sequences of interest, and donor polynucleotides can comprise one or more genes of interest. Such genes of interest can encode, for example, a protein that can provide an agronomic advantage to the plant.

[0589] Also, as described herein, for each example or embodiment that cites a guide RNA, a similar guide polynucleotide can be designed wherein the guide polynucleotide does not solely comprise ribonucleic acids but wherein the guide polynucleotide comprises a combination of RNA-DNA molecules or solely comprises DNA molecules.

[0590] In order to edit organellar genomes with polynucleotide guided (e.g., RNA guided) methodologies, two molecular components, a polynucleotide guided polypeptide (e.g., Cas protein, Cas9) and a guide polynucleotide (e.g., guide RNA), can be introduced into organelles. The introduction of these components may be accomplished by a combination of a suitable approach. One approach can be to create a modified polynucleotide guided polypeptide by a translational fusion of the polynucleotide guided polypeptide with an organelle targeting peptide that can allow protein import into an organelle. Another approach can be to create a transcriptional fusion of a guide polynucleic acid

with an RNA molecule that can be imported into an organelle. For the latter, the configuration of imported guide polynucleic acid (e.g., guide RNA) can be designed to enable appropriate function, i.e., the 5' end of guide RNA can be accessible to bind with the target site on the organellar DNA. The combination of these two components can be sufficient to edit organellar genomes to create small deletions (e.g., SDN1 modifications) and additions of a few nucleotides at the cleavage sites (e.g., SDN2 modifications). To achieve organellar genome editing with more extensive SDN2 and SDN3 modifications, a polynucleotide modification template can be introduced into the corresponding organelle.

[0591] After creating a designed change in organellar DNA, the next step can be to maintain the edited organellar DNA in the pool of unmodified organellar DNA and to shift the balance among organellar DNA to favor the maintenance of genome edited organellar DNA. This can be achieved by reducing the amplification of unmodified organellar DNA. In one approach, guide polynucleic acids can be designed for multiple target sites in the unmodified organelle genome. The donor polynucleotide (e.g. donor DNA) can be designed such that these target sites have been altered to no longer be recognized by the relevant polynucleotide guided polypeptide system(s). Expression of the polynucleotide guided polypeptides can result in the introduction of single-strand or double-strand breaks into the unmodified organellar DNA and can thereby increase the proportion of modified genomes. In one variation, cells may be pretreated with relevant polynucleotide guided polypeptide systems to introduce cleavages in organellar DNA. The pretreatment can reduce the number of organelle DNA molecules available for homologous recombination.

[0592] Embodiments can involve a single guide RNA (sgRNA), i.e., where the variable targeting domain can be fused to a polynucleotide that contains a tracrRNA sequence. Alternatively, embodiments may involve a duplex guide RNA, i.e., where the variable targeting domain and the tracrRNA sequence are present on separate RNA molecules. The terms “duplex guide RNA” and “dual guide RNA” are used interchangeably herein.

[0593] In some cases, protein and/or RNA expression levels can be higher when transformed into an organelle (e.g., plastid, mitochondria) compared with that in nucleus. For example, protein expression level can be at least about: 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% higher with organelle transformation when compared with nuclear transformation. The expression stability of a transcript can be higher with organelle transformation compared with nuclear transformation.

EMBODIMENTS

[0594] In one embodiment, a polynucleotide encoding an RNA sequence may comprise an organelle targeting RNA operably linked to a guide polynucleic acid (e.g., single guide RNA), wherein the guide polynucleic acid can direct a polynucleotide guided polypeptide (e.g., Cas polypeptide; Cas9 polypeptide) to cleave a target sequence present in an organelle genome. The guide polynucleic acid may be single guide RNA or a duplex guide RNA; for a duplex RNA, each component RNA is operably linked to an organelle targeting RNA. The RNA sequence may further comprise a sequence encoding a polynucleotide guided polypeptide (e.g., Cas polypeptide; Cas9 polypeptide). The RNA sequence may

further comprise an RNA cleavage site between the guide polynucleic acid and the sequence encoding a polynucleotide guided polypeptide. The RNA cleavage site may be at least one selected from the group consisting of: a Csy4 cleavage site, a C2c2 cleavage site, a ribozyme cleavage site, an RNase III cleavage site, and any combination thereof.

[0595] In another embodiment, a cell may comprise any of the polynucleotides of the disclosure.

[0596] In another embodiment, a cell may comprise any of the above polynucleotide, wherein the cell further comprising a polynucleotide encoding a modified polynucleotide guided polypeptide, wherein the modified polynucleotide guided polypeptide comprises a polynucleotide guided polypeptide (e.g., Cas polypeptide; Cas9 polypeptide) operably linked to an organelle targeting peptide.

[0597] In another embodiment, a method for introducing a guide polynucleic acid into an organelle of a cell may comprise: (a) introducing into a cell any of the above polynucleotides, wherein the polynucleotide is operably linked to at least one regulatory element; and (b) growing the cell under conditions in which the polynucleotide is expressed. The method may further comprise (c) selecting a cell having an organelle that comprises a guide polynucleic acid.

[0598] In another embodiment, a method for altering the genome of an organelle may comprise: (a) introducing into a cell: (i) a first polynucleotide encoding an RNA sequence comprising an organelle targeting RNA operably linked to a guide polynucleic acid, wherein the guide polynucleic acid can direct a polynucleotide guided polypeptide (e.g., Cas polypeptide; Cas9 polypeptide) to cleave a target sequence present in an organelle genome, wherein the polynucleotide is operably linked to at least one regulatory element; and (ii) a second polynucleotide encoding a modified polynucleotide guided polypeptide, wherein the second polynucleotide is operably linked to at least one regulatory element, and wherein the modified polynucleotide guided polypeptide comprises a polynucleotide guided polypeptide (e.g., Cas polypeptide; Cas9 polypeptide) operably linked to an organelle targeting peptide; wherein the organelle targeting RNA of (i) and the organelle targeting peptide of (ii) each target the same organelle; and (b) growing the cell under conditions in which the first polynucleotide of (i) and the second polynucleotide of (ii) are both expressed. The method may further comprise (c) selecting a cell having an organelle that comprises an altered genome.

[0599] In another embodiment, a method for altering the genome of an organelle may comprise: (a) introducing into a cell: (i) a first polynucleotide encoding an RNA sequence comprising an organelle targeting RNA operably linked to a guide polynucleic acid, wherein the guide polynucleic acid can direct a polynucleotide guided polypeptide (e.g., Cas polypeptide; Cas9 polypeptide) to cleave a target sequence present in an organelle genome, wherein the polynucleotide is operably linked to at least one regulatory element; and (ii) a third polynucleotide, wherein the third polynucleotide is operably linked to at least one regulatory element, wherein the third polynucleotide encodes an RNA molecule comprising an organelle targeting RNA operably linked to an RNA sequence encoding a polynucleotide guided polypeptide (e.g., Cas polypeptide; Cas9 polypeptide); wherein the organelle targeting RNA of (i) and the organelle targeting RNA of (ii) each target the same organelle; and (b) growing the cell under conditions in which the polynucleotide of (i)

and the third polynucleotide of (ii) are both expressed. The method may further comprise (c) selecting a cell having an organelle that comprises an altered genome.

[0600] In another embodiment, a method for altering the genome of an organelle may comprise: (a) introducing into a cell a polynucleotide encoding an RNA sequence comprising an organelle targeting RNA operably linked to a guide polynucleic acid, wherein the guide polynucleic acid can direct a polynucleotide guided polypeptide (e.g., Cas polypeptide; Cas9 polypeptide) to cleave a target sequence present in an organelle genome, wherein the RNA sequence further comprises a second RNA sequence encoding a polynucleotide guided polypeptide (e.g., Cas polypeptide; Cas9 polypeptide), wherein the polynucleotide is operably linked to at least one regulatory element; and (b) growing the cell under conditions in which the polynucleotide of (a) is expressed. The method may further comprise (c) selecting a cell having an organelle that comprises an altered genome.

[0601] In any of the above methods for altering the genome of an organelle, the method may further comprise introducing a polynucleotide comprising at least one donor polynucleotide (e.g., donor DNA) into the organelle, wherein the at least one donor polynucleotide is bounded by at least one homologous sequence with respect to the organelle genome, wherein integration of all or part of the at least one donor polynucleotide into the organelle genome results in removal of the target site of the guide polynucleic acid. The at least one donor polynucleotide may comprise a first nucleic acid sequence that is heterologous to the organelle genome, wherein the first nucleic acid sequence is bounded by a second and a third nucleic acid sequence, wherein the second and the third nucleic acid sequences correspond to two adjacent regions of homology in the organelle genome. The first nucleic acid sequence that is heterologous to the organelle genome may encode a selectable marker. The selectable marker may be *aadA* and the selection agent may be spectinomycin or streptomycin. The first nucleic acid sequence that is heterologous to the organelle genome may be operably linked to at least one regulatory element that is active in the organelle. The second or the third nucleic acid sequence, or both, may comprise at least one altered sequence, wherein the at least one altered sequence is altered with respect to at least one additional target site in the organelle genome, wherein the at least one altered sequence is not cleavable by at least one additional guide polynucleic acid, wherein the at least one additional guide polynucleic acid can direct a polynucleotide guided polypeptide (e.g., Cas polypeptide; Cas9 polypeptide) to cleave the at least one additional target site in the organelle genome. The at least one additional target site in the organelle genome may be present in at least one essential coding region. The polynucleotide introduced into the organelle may further comprise a fourth nucleic acid sequence, wherein the fourth nucleic acid sequence encodes the at least one additional guide polynucleic acid operably linked to a promoter that is active in the organelle.

[0602] In another embodiment, a polynucleotide may encode a modified RNA donor sequence, wherein the modified RNA donor sequence may comprise an organelle targeting RNA operably linked to a donor RNA. The modified RNA donor sequence may comprise a reverse transcriptase primer site.

[0603] In another embodiment, the cell may comprise the polynucleotide encoding the modified RNA donor sequence,

and further comprise a polynucleotide encoding a modified reverse transcriptase, wherein the modified reverse transcriptase comprises a reverse transcriptase operably linked to an organelle targeting peptide.

[0604] In any of the above methods for altering the genome of an organelle, the method may further comprise introducing a polynucleotide comprising at least one donor polynucleotide (e.g., donor DNA) into the organelle, wherein the donor polynucleotide is introduced into the organelle by: (a) introducing into a cell a polynucleotide encoding a modified RNA donor sequence, wherein the modified RNA donor sequence comprises an organelle targeting RNA operably linked to a donor RNA, wherein the modified RNA donor sequence comprises a reverse transcriptase primer site, and wherein the polynucleotide is operably linked to at least one regulatory element; (b) introducing into the cell a polynucleotide encoding a modified reverse transcriptase, wherein the modified reverse transcriptase comprises a reverse transcriptase operably linked to an organelle targeting peptide, wherein the polynucleotide is operably linked to at least one regulatory element, wherein the organelle targeting RNA of (a) and the organelle targeting peptide of (b) each target the same organelle; and (c) growing the cell under conditions wherein the polynucleotides of (a) and (b) are both expressed. The method may further comprise (d) selecting a cell having an organelle that comprises an altered genome.

[0605] In another embodiment, a method for altering the genome of an organelle may comprise: (a) introducing into an organelle the following: (i) a first polynucleotide encoding at least one guide polynucleic acid, wherein the at least one guide polynucleic acid can direct a polynucleotide guided polypeptide (e.g., Cas polypeptide; Cas9 polypeptide) to cleave at least one target sequence present in an organelle genome; (ii) a second polynucleotide encoding a polynucleotide guided polypeptide (e.g., Cas polypeptide; Cas9 polypeptide), wherein the polynucleotide guided polypeptide, when associated with the guide polynucleic acid (e.g., guide RNA), can cleave the at least one target sequence; (iii) optionally, a third polynucleotide encoding at least one homologous organelle DNA sequence, wherein the at least one homologous organelle DNA is of sufficient size for homologous recombination, wherein integration of the at least one homologous organelle DNA sequence into the organelle genome results in removal of the at least one target sequence; (iv) optionally, a fourth polynucleotide encoding at least one selectable marker or at least one screenable marker, or both, wherein the sequence encoding the at least one selectable marker, or at least one screenable marker, or both, is operably linked to a promoter that is functional in the organelle; and (v) optionally, a fifth polynucleotide encoding an origin of replication that is functional in the organelle; and (b) growing a cell comprising the organelle of (a) under conditions in which the first polynucleotide of (i) and the second polynucleotide of (ii) are each expressed. The method may further comprise a step (c) of selecting a cell having an organelle that comprises an altered genome. The method may further comprise a step (d) of selecting a cell that is homoplasmic for the altered genome of the organelle. The third polynucleotide of (iii) may comprise a sixth and a seventh polynucleotide, wherein the sixth and the seventh polynucleotides correspond to two adjacent regions of homology in the organelle genome, wherein the sixth and seventh polynucleotides are separated by a sequence that is

heterologous to the organelle DNA. The sequence that is heterologous to the organelle DNA may comprise at least one selected from the group consisting of: the first polynucleotide of (i), the second polynucleotide of (ii), the fourth polynucleotide of (iv), an eighth polynucleotide, and any combination thereof, wherein the eighth polynucleotide encodes an RNA that is heterologous to the organelle or comprises a non-coding sequence (e.g., a regulatory sequence, such as a promoter) that is heterologous to the organelle, or both. The RNA that is heterologous to the organelle may be at least one selected from the group consisting of: an mRNA, a functional RNA, and any combination thereof. The functional RNA may be at least one selected from the group consisting of: guide RNA, siRNA, miRNA, dsRNA, tRNA, rRNA, and any combination thereof. At least one selected from the group consisting of: the first polynucleotide of (i), the second polynucleotide of (ii), the fourth polynucleotide of (iv), the fifth polynucleotide of (v), and any combination thereof, may be located outside the region bounded by the sixth and the seventh polynucleotide. The fifth polynucleotide of (v) may encode a plastid origin of replication, a mitochondrial origin of replication, or both. The plastid origin of replication may correspond to DNA sequence from a plastid rRNA intergenic region.

[0606] In any of the methods described herein, one or more of the polynucleotides described herein may be present on a recombinant DNA construct.

[0607] In any of the methods described herein, the method may comprise more than one such recombinant DNA construct.

[0608] In any of the methods described herein, the recombinant DNA construct may further comprise a ninth and tenth polynucleotide, wherein the ninth and tenth polynucleotides have 100 percent sequence identity to each other, and further wherein the ninth and tenth polynucleotides are arranged as direct repeats in the recombinant DNA construct. The ninth and tenth polynucleotides may have at least 20, 21, 22, 23, 24, 25, 30, 40, 50, 60, 70, 80, 90 or 100 nucleotides of 100 percent sequence identity to each other. The recombinant DNA construct may be linear, and the ninth and tenth polynucleotides may be present at the 5' and 3' ends of the recombinant DNA construct, respectively.

[0609] In any of the methods described herein for altering the genome of an organelle, the recombinant DNA construct may be linear, single-stranded and operably linked to a modified VirD2 protein. The modified VirD2 protein may comprise a VirD2 protein operably linked to an organelle targeting peptide, wherein the modified VirD2 protein has also been modified such that at least one native nuclear localization sequence of the VirD2 protein is no longer functional.

[0610] In the above methods for altering the genome of an organelle, the recombinant DNA construct may be operably linked to at least one modified VirE2 protein. The at least one modified VirE2 protein may comprise a VirE2 protein operably linked to an organelle targeting peptide, wherein the at least one modified VirE2 protein has also been modified such that at least one native nuclear localization sequence of the VirE2 protein is no longer functional.

[0611] In any of the methods described herein for altering the genome of an organelle, the recombinant DNA construct may be operably linked to at least one modified RecA

protein. The at least one modified RecA protein may comprise a RecA protein operably linked to an organelle targeting peptide.

[0612] In any of the methods described herein for altering the genome of an organelle, the recombinant DNA construct may be operably linked to at least one chimeric polypeptide. The at least one chimeric polypeptide may comprise an organelle targeting peptide and a cell penetrating peptide and optionally, a DNA-binding polypeptide.

[0613] In another embodiment, a method for altering the genome of an organelle may comprise using of both a site-directed nuclease (e.g., TALENS, Zinc-Finger Nuclease or Meganuclease) and a polynucleotide guided polypeptide. The initial cleavage of the organelle genome may be done by a site-directed nuclease (e.g., TALENS, Zinc-Finger Nuclease, Meganuclease), to facilitate homologous recombination with a donor polynucleotide. The donor polynucleotide may contain modified target sites that are not recognized by a polynucleotide guided polypeptide. A homoplasmic state may be facilitated by cleavage of the unmodified organelle genomes at the target sites by treatment with a polynucleotide guided polypeptide. In another embodiment, any of the above methods may further comprise introducing into the organelle a polynucleotide encoding at least one marker selected from the group consisting of: a positive selectable marker, a negative selectable marker, a screenable marker, and any combination thereof. The positive selectable marker may be an herbicide tolerance protein. The herbicide tolerance protein may be at least one selected from the group consisting of: a 4-hydroxyphenylpyruvate dioxygenase (HPPD), a sulfonylurea-tolerant acetolactate synthase (ALS), an imidazolinone-tolerant acetolactate synthase (ALS), a glyphosate-tolerant 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), a glyphosate-tolerant glyphosate oxidoreductase (GOX), a glyphosate N-acetyltransferase (GAT), a phosphinothricin acetyl transferase (PAT), a protoporphyrinogen oxidase (PROTOX), an auxin enzyme or receptor, a P450 polypeptide, an acetyl coenzyme A carboxylase (ACCase), and any combination thereof. The method may further involve growing the cell in the presence of a positive selection agent and selecting a cell that is homoplasmic for the altered genome of the organelle. Optionally, the method may further involve growing the cell in the absence of the positive selection agent, followed by selecting a cell that lacks a non-integrated recombinant DNA construct. Alternatively, the method may further involve growing the cell in the absence of the positive selection agent, followed by growing the cell in the presence of a negative selection agent, followed by selecting a cell that lacks a non-integrated recombinant DNA construct. In the method, the cell may be a plant cell, the organelle may be a plastid, and the method may further involve regenerating a plant from the plant cell comprising an altered organelle genome. The plant cell may be monocot cell, e.g., a maize cell. The plant cell may be a dicot cell, e.g., a soybean cell.

[0614] In another embodiment, a method for altering a genome of an organelle may comprise: (a) introducing into an organelle of a cell the following: (i) at least one guide RNA, wherein the at least one guide RNA directs a polynucleotide guided polypeptide to cleave at least one target sequence present in the genome of the organelle; (ii) a polynucleotide guided polypeptide, wherein the polynucleotide guided polypeptide, when associated with the at least

one guide RNA, cleaves the at least one target sequence; and (iii) a replacement DNA; and (b) selecting a cell comprising an organelle comprising the replacement DNA. The replacement DNA of step (a) part (iii) may comprise fragments of organellar DNA or a complete organellar DNA from a cultivar, line, sub-species and other species and is distinct from the genome of the organelle of step (a). The replacement DNA may be lacking the at least one target sequence. Additionally, after step (a) part (ii) and prior to step (a) part (iii), a cell may be selected in which the genome of the organelle has been eliminated.

[0615] In another embodiment, the guide polynucleic acid in the methods and compositions of matter described herein may comprise the following: i) at least 17 nucleotides that are complementary to at least 17 nucleotides of a target polynucleic acid, wherein said target polynucleic acid is located in the genome of an organelle; and ii) a region that contacts a polynucleotide-guided polypeptide. The guide polynucleic acid may comprise one or more RNA bases. The guide polynucleic acid may be a guide RNA. The guide polynucleic acid may be a dual guide RNA. The guide polynucleic acid may be a single guide RNA.

[0616] In another embodiment, the polynucleotide-guided polypeptide in the methods and compositions of matter described herein may be selected from the group consisting of: a Cas9 protein, a MAD2 protein (U.S. Pat. No. 10,011, 849; herein incorporated by reference), a MAD7 protein (U.S. Pat. No. 9,982,279; herein incorporated by reference), a CRISPR nuclease, a nuclease domain of a Cas protein, a Cpf1 protein, an Argonaute, modified versions thereof, and any combination thereof. The sequence encoding the polynucleotide-guided polypeptide may be codon-optimized for a human, a yeast, an alga, or a plant species.

[0617] In any of the methods described herein for altering the genome of an organelle, the method may further involve growing the cell in the presence of a positive selection agent and selecting a cell that is homoplasmic for the altered genome of the organelle. The method may further involve: (i) growing the cell in the absence of the positive selection agent, followed by selecting a cell that lacks a non-integrated recombinant DNA construct; or (ii) growing the cell in the absence of the positive selection agent, followed by growing the cell in the presence of a negative selection agent, followed by selecting a cell that lacks a non-integrated recombinant DNA construct.

[0618] In any of the methods described herein that involve a guide polynucleic acid and a polynucleotide guided polypeptide, the method may comprise an increase in transformation efficiency of at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, 400%, or 500%, as compared to the corresponding method lacking the guide polynucleic acid, the polynucleotide guided polypeptide, or lacking both.

[0619] In any of the methods described herein that involve a guide polynucleic acid and a polynucleotide guided polypeptide, the method may comprise a decrease in the amount of time required to achieve a homoplasmic state, wherein the decrease is at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% or 90%, as compared to the amount of time required for the corresponding method lacking the guide polynucleic acid, the polynucleotide guided polypeptide, or lacking both.

[0620] In another embodiment, a recombinant DNA construct (e.g., for use in any of the methods described herein) may comprise any one or more of the polynucleotides described herein.

[0621] In another embodiment, a cell may comprise an organelle, wherein the organelle may comprise at least one of the above recombinant DNA constructs. The cell may be selected from the group consisting of: a yeast cell, an algal cell, a plant cell, an insect cell, a non-human animal cell and a mammalian tissue culture cell.

[0622] In another embodiment, a plant or seed may comprise any of the above organelles, cells or recombinant DNA constructs.

[0623] In another embodiment, a cell comprising an organelle with an altered genome may be produced by any of the above methods. The cell may be selected from the group consisting of: a yeast cell, an algal cell, a plant cell, an insect cell, a non-human animal cell and a mammalian tissue culture cell.

[0624] In another embodiment, a method may alter the genome of an organelle in a cell, wherein the cell is a plant cell. Furthermore, a plant may be regenerated from the plant cell comprising an organelle with an altered genome, wherein the regenerated plant comprises an organelle with an altered genome. Also, a plant (e.g., progeny plant) or seed may be produced from the regenerated plant, wherein the plant or seed comprises an organelle with an altered genome.

[0625] In any of the above embodiments involving guide polynucleic acid (e.g., guide RNA), the guide polynucleic acid may be a single guide RNA (unimolecular) or a duplex guide RNA (bimolecular). In any embodiment involving multiple guide RNAs, the multiple guide RNAs may be single guide RNAs, duplex guide RNAs, or both.

[0626] In any of the above embodiments, multiple guide RNAs (and/or other heterologous RNAs) may be encoded on separate transcription units or may be encoded on a polycistronic transcription unit. A guide RNA may be processed from a polycistronic RNA after transcription; e.g., by use of an RNA cleavage site (e.g., Csy4; C2c2), a ribozyme cleavage site, a polynucleotide guided polypeptide cleavage site or the presence of a tRNA sequence. A guide RNA may be processed from a polycistronic RNA by having a first tRNA sequence 5' to the guide RNA and a second tRNA sequence 3' to the guide RNA. Multiple guide RNAs may be arrayed with multiple tRNA sequences (at each guide RNA 5' and 3' end) for processing from a polycistronic RNA.

[0627] In any of the above embodiments, the polynucleotide (e.g., donor DNA, donor RNA) that can be introduced into the organelle may comprise at least one selected from the group consisting of: an expression cassette encoding a polynucleotide of interest and an expression cassette encoding a polycistronic transcript that comprises multiple polynucleotides of interest; e.g., a polycistronic transcript comprising multiple protein-coding regions, multiple functional RNAs, or a combination of both. The polynucleotide of interest may be heterologous with respect to the genome of the organelle.

[0628] In any of the above methods for altering the genome of an organelle to contain a heterologous polynucleotide, the heterologous polynucleotide may encode at least one selected from the group consisting of: an herbicide tolerance protein, a pesticidal protein, an accessory protein that binds to a pesticidal protein, a dsRNA, a siRNA, a miRNA, and any combination thereof, wherein the dsRNA,

the siRNA and the miRNA can suppress at least one target gene present in a plant pest. The herbicide tolerance protein may be at least one selected from the group consisting of: a 4-hydroxyphenylpyruvate dioxygenase (HPPD), a sulfonyleurea-tolerant acetolactate synthase (ALS), an imidazolinone-tolerant acetolactate synthase (ALS), a glyphosate-tolerant 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), a glyphosate-tolerant glyphosate oxidoreductase (GOX), a glyphosate N-acetyltransferase (GAT), a phosphinothricin acetyl transferase (PAT), a protoporphyrinogen oxidase (PROTOX), an auxin enzyme or receptor, a P450 polypeptide, an acetyl coenzyme A carboxylase (ACCase), and any combination thereof. The pesticidal protein may be at least one selected from the group consisting of: Cry1Ac, Cyt1Aa, Cry1Ab, Cry2Aa, Cry1l, Cry1C, Cry1D, Cry1E, Cry1Be, Cry1Fa and Vip3A. The accessory protein that binds to a pesticidal protein may be at least one selected from the group consisting of: a 20 kDa accessory protein and a 19 kDa accessory protein. The dsRNA, the siRNA and the miRNA can suppress at least one target gene selected from the group consisting of: proteasome A-type subunit peptide (Pas-4), ACT, SHR, EPIC2B, PnPMAL, and any combination thereof. The heterologous polynucleotide may be operably linked to at least one regulatory element that is active in an organelle. The at least one regulatory element may be selected from the group consisting of: a maize clpP promoter combined with a maize clpP 5'-UTR, a maize clpP promoter combined with a 5'-UTR from gene 10 of bacteriophage T7, a tomato psbA promoter is combined with a 5'-UTR from gene 10 of bacteriophage T7, a tomato rm16 promoter combined with a modified accD 5'-UTR, and any combination thereof. The cell may be a plant cell, wherein the organelle is a plastid (e.g., a chloroplast), and wherein the method further comprises regenerating a plant from the plant cell comprising an altered organelle genome. The plant cell may be a soybean cell.

[0629] In any of the above methods for altering the genome of an organelle to contain a heterologous polynucleotide, the heterologous polynucleotide may be flanked by direct repeat sequences. The direct repeat sequences may have at least 20, 21, 22, 23, 24, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500 or 600 nucleotides of 100 percent sequence identity to each other. The direct repeat sequences may comprise a site-specific recombinase site (e.g., loxP, attP, attB). The heterologous polynucleotide may encode at least one marker selected from the group consisting of: a positive selectable marker, a negative selectable marker, a screenable marker, and any combination thereof. Optionally, the method may further involve growing the cell in the absence of the positive selection agent, followed by selecting a cell that is homoplasmic for organelles that lack the heterologous polynucleotide. Alternatively, the method may further involve growing the cell in the presence of a negative selection agent, followed by selecting a cell that is homoplasmic for organelles that lack the heterologous polynucleotide. Optionally, the method may involve growing the cell under conditions in which a heterologous site-specific recombinase (e.g., Cre, phiC31, Bxb1) is expressed in the organelle.

[0630] In the above embodiments, the target organelle may be a plastid (e.g., chloroplast) or a mitochondrion. The organelle targeting polynucleotide may be tRNA, viroid RNA or eIF4E RNA.

[0631] In the above embodiments, expression of an antibiotic marker gene may be used in conjunction with antibiotic selection for obtaining (and selecting) a plastid or mitochondrial transformation event (e.g., a homoplasmic event). The polynucleotide comprising the donor polynucleotide (e.g., donor DNA) may also comprise an expression cassette for the antibiotic marker gene; the expression cassette may be within the donor polynucleotide region (i.e., for integration into the organelle genome) or outside the donor polynucleotide region.

EXAMPLES

[0632] The present disclosure is further defined in the following Examples, in which parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating embodiments, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this disclosure, and without departing from the spirit and scope thereof, can make various changes and modifications of the disclosure to adapt it to various usages and conditions. Such modifications are also intended to fall within the scope of the appended claims.

[0633] Experiments typically involve a single guide RNA (sgRNA), i.e., where the variable targeting domain is fused to a polynucleotide that contains a tracrRNA sequence. Alternatively, experiments may involve a duplex guide RNA, i.e., where the variable targeting domain and the tracrRNA sequence are present on separate RNA molecules.

Example 1

Targeting Cas9 and Guide RNA into Yeast Mitochondria

[0634] To create the Cas9 protein for mitochondrial genome editing, a protein functional in nuclear genome editing is modified by fusing a mitochondrial targeting peptide at the amino terminal end and by deleting any NLS (nuclear localization signal) elements. The organelle targeting peptides of the ATPase beta subunit and the 70 KD protein are used for the modification, creating mCas9-A (encoded by SEQ ID NO: 1) and mCas9-B (encoded by SEQ ID NO: 2), respectively. Each polynucleotide encoding a modified Cas9 is cloned into a yeast shuttle vector with expression of the polynucleotide under control of the Gall promoter, whose activity is induced by galactose as a carbon source in the media.

[0635] To create guide RNA for mitochondrial genome editing, the tRNA^{Lys} (tRK1 and modified tRK2 forms) that can be imported into mitochondria is used. Several versions of fusion RNA between tRNA and guide RNA are made. One approach is to fuse guide RNA to the 5' end of the tRNA (SEQ ID NO: 3 and 4). To suppress 5' end cleavage by RNase P, the first base of tRNA is modified in an alternative construct to prevent the pairing with the corresponding base on the acceptor stem of the tRNA (SEQ ID NO: 5 and 6). The second approach is to replace the intron of tRK2 with efficient mitochondrial import in the backbone of tRK2-2 and tRK1 (SEQ ID NO: 7 and 8, respectively). The third approach is to use the fact that tRK1 (tRNA^{Lys}) can be split into two molecules that together retain the property of mitochondrial import. In this case, guide RNA is fused to the

5' end of second half of the tRK1 in the region called variable loop in tRNA structure in a manner that retain the secondary structure of the tRNA splicing site (SEQ ID NO: 9). The guide RNA fused with B form (SEQ ID NO: 10) is co-expressed with A form to facilitate co-import into mitochondria.

[0636] A variation of creating synthetic guide RNA with RNA that serves as the efficient vehicle for mitochondrial import is to use the combination of F-hairpin and D-arm structures of tRK1. These structures are shown to facilitate import into mitochondria. In this approach, guide RNA is placed between two structures (SEQ ID NO: 11) or fused with one of them at the 5' or 3' ends (e.g. SEQ ID NO: 12 and 13).

[0637] For the site-specific cleavage sites, the following mitochondrial sequences were identified as target sites for guide RNA; the guide RNA variable targeting domain is shown below:

1. (cytochrome b gene) (SEQ ID NO: 14)
ACTGATAGAAGTGTAGTAAG_
2. (COX1 gene) (SEQ ID NO: 15)
ATGATTATTGCAATTCCAAC
3. (COX1 gene) (SEQ ID NO: 16)
ATTCCACGATACTTACTACG
4. (COX2 gene) (SEQ ID NO: 17)
TCAGCAACACCAAAATCAAGA

[0638] Each of the above variable targeting domains precedes a PAM sequence. SEQ ID NO: 14-17 precede the following PAM sequences: AGG, AGG, TGG and AGG, respectively.

[0639] Eleven nucleotides from the 3' end of each underlined sequence (adjacent to the PAM sequence), which are considered critical for Cas9 target site recognition, are unique to the yeast mitochondrial genome based on blast analyses. Each of the above variable targeting domains is fused at the 3' end with a tracrRNA sequence for Cas9 recognition (SEQ ID NO: 18). Polynucleotides encoding each engineered guide RNA are expressed in the nucleus under control of the SNR52 promoter and the SUP4 termination element (SEQ ID NO: 19 and 20, respectively). In this experiment, a yeast shuttle vector for transformation is used. For example, SNR52 expression cassettes are cloned into a yeast expression vector such as p 416-Gall (URA3+, multicopy plasmid purchased from ATCC). Expression cassettes encoding mitochondrial targeted Cas9 ("mCas9") are cloned into the Sall-XhoI sites of centromeric p 415-gall vector (LEU+) with expression under control of the Gall promoter whose activity is induced by galactose in the media as sole carbon source. Vectors are transformed into a yeast strain allowing auxotrophy selection such as BY4733 (mat a) line, and selected for Leu and Ura independent growth.

[0640] The transformants of each and/or the combination of mCas9 and guide RNA constructs are selected on media selective for corresponding auxotrophy as single colony lines. The expression of mCas9 endonuclease is induced by shifting media to the one containing galactose as sole carbon source. Cells derived from single colonies are grown in the

inducing media for several generations. These lines are analyzed for genome editing efficacies at the molecular level. Cells from multiple lines of each construct and each construct combination are combined together and their DNA are isolated by using standard DNA isolation protocols such as by using Yeast DNA Extraction Kit from ThermoFisher (cat#78870). Using PCR primer sets specific to corresponding genome editing sites, DNA at each editing site is amplified by PCR reaction. PCR products are subjected to high-throughput sequencing such as by using Illumina HiSeq protocols provided by the manufacture. The frequency of site-specific mutations at each target site is evaluated in comparison with corresponding control constructs. The efficacy of genome editing is also analyzed at the functional level. After obtaining single colony lines, each line is further grown for additional generations in non-selective glucose media to promote a homoplasmic state of the mitochondrial genome. Yeast cells are plated on glucose media such as YD medium. Single colonies are transferred to the glycerol media such as YG medium by replica plating. The efficacy of genome editing is evaluated by the output frequency of colonies incapable of growth on glycerol media, i.e. deficient in respiration due to the mutations in *cob*, *cox1* and *cox2* genes, respectively.

[0641] The next step of organellar genome editing is to create a dominant and sustainable state of the edited DNA in mitochondria, which initially contains a pool of multiple, if not hundreds of, unedited DNA. This is achieved by extending the period of enzymatic reactions of site-specific modifications in organelles. Depending on several factors such as the import efficiency of mCas9 and guide RNA into mitochondria, and the affinity between guide RNA, imported Cas9 and target sites, the length of the extended period suitable for each modification of an organelle varies. To assess the effect of extended periods, yeast lines transformed with appropriate mCas9 and guide RNA pairs are grown in selective media for corresponding constructs over a time course of hours, days and weeks. Then, each culture is subjected to evaluation at the molecular as well as functional levels as described above. The period of enzymatic states sufficient for the maintenance and phenotypic expression of edited mitochondrial genomes over generations is determined from the time course experiments.

Example 2

Targeting Cas9, Guide RNA and Donor DNA into Yeast Mitochondria

[0642] In order to edit organellar genomes precisely at the nucleotide level, donor DNA (comprising a polynucleotide modification template) is added to the site-specific endonuclease system. In one approach, donor DNA is introduced into mitochondria in combination with Cas9 and guide RNA; Cas9 and guide RNA are introduced into mitochondria as described in Example 1. In this example, the donor DNA is designed to create a specific mutation in the 15S rRNA gene in the mitochondrial genome to confer paromomycin resistance. The nucleotide substitution (C-to-G) at position 1514 can confer paromomycin resistance. To create the donor DNA with the resistance allele, one primer pair is designed to carry the corresponding substitution (SEQ ID NO: 21). PCR amplification is performed by using the primer set (SEQ ID NO: 21 and 22) and yeast total DNA as substrate following standard PCR protocols. The resulted

template DNA is transformed into mitochondria via DNA transformation procedures, such as biolistic methods. For transformation with donor DNA, the cells expressing Cas9 and guide RNA as described in Example 1 are used with the exception that the guide RNA is designed to cleave the vicinity of the paromomycin-resistance site of mitochondrial DNA as exemplified in SEQ ID NO: 23. The guide RNA is so designed that the cleavage site is covered by the donor DNA with overlapping sequences sufficient for homologous recombination at the both ends but the donor DNA is not recognized as the substrate for site-specific endonuclease activities. For instance, the donor DNA is modified to not include the PAM sequence that is targeted by the corresponding guide RNA. The variable targeting domain of the guide RNA is fused at the 3' end with tracrRNA sequence for association with Cas endonuclease; guide RNA expression constructs are made by using tRNA^{Lys} derived methods described in Example 1.

[0643] After transformation with donor DNA, cells are pooled together and grown in galactose media to induce Cas9 protein for several generations, following the favorable amplification of the engineered DNA by adding gradually increasing amount of paromomycin in the media over additional generations. Cells are plated to make single colonies. Single colonies are replica plated on media with glycine as the sole carbon source in the presence and absence of paromomycin to identify paromomycin resistant colonies. The efficiency of genome editing by this method is shown by an increased rate of producing paromomycin resistant cells with template DNA in comparison to control cells not transformed with donor DNA. Gene editing is confirmed by sequencing of the engineered site.

[0644] A subsequent genome editing step is performed to eliminate organellar DNA that does not carry the designed modification. This is achieved by any of several approaches. One approach is to expose cells under positive selection pressure as described above. Another approach is to eliminate or reduce the replication rate of unmodified organellar DNA. This can be achieved by cleaving unmodified DNA by use of site-specific endonucleases such as zinc finger proteins, TALEN and Cas9 systems. In the Cas9 approach, expression of specific guide RNAs is used to cleave unmodified organellar DNA and thereby increase the population of modified DNA.

Example 3

Replacement of Endogenous Organellar DNA

[0645] This is an alternative method for modification of an organellar genome. In this approach, the first step is to reduce or eliminate the endogenous organellar DNA by using site-specific endonucleases such as Cas9 systems. At the same time or subsequently, a replacement organellar DNA is introduced. The replacement DNA can be fragments of organellar DNA or complete organellar DNA that convey a new genotype and corresponding trait(s) when transformed into the organelle. In the case of organellar DNA fragments, they can be integrated into the remaining organellar DNA by homologous recombination. In the case of complete organellar DNA replacement, the replacement DNA can be isolated from cultivars, lines, sub species and other species which possess DNA compositions distinct from the endogenous organellar DNA of recipient cells. One requirement of the replacement DNA can be to contain a DNA element

functioning as a DNA replication origin in the recipient organelles. The replacement DNA can also be synthesized partially and/or completely. When replacement DNA is created in vitro, it can be a linear DNA with the inverted repeat sequence at the ends. The ends can facilitate homologous recombination in vitro or in vivo to create circular DNA for replication of organellar DNA in cells. The DNA created in vitro can also include exogenous DNA elements such as ones to allow selected amplification in bacterial cells.

[0646] To reduce or eliminate mitochondrial DNA, yeast cells are exposed to prolonged expression of guide RNA and Cas9 protein that are designed to be imported into mitochondria as described in Example 1 or to be synthesized directly in organelles as described in Example 4. The target sites are chosen to be unique to the endogenous mitochondrial DNA and not present in nuclear genome to reduce the chance of any damage occurring on nuclear genomes when taking the method described in Example 1. The target sites are also chosen to not be present in the replacement DNA.

[0647] Multiple cleavage sites enhance the rate of displacing endogenous organellar DNA. This can be attained by expressing multiple guide RNAs targeting different unique sequences in the endogenous mitochondrial DNA (e.g., see target sites of Example 1). After Cas9/guide RNA treatment, yeast cells that have lost mitochondrial DNA are identified by lack of respiration, inability to grow on media with glycerin as sole carbon source and the lack of mitochondrial DNA. The resulting rho⁰ condition can also be confirmed by absence of the mitochondrial DNA band in a CsCl gradient through the method described in Example 1. Once mitochondrial DNA is deleted, cells are then transformed with replacement DNA created in vitro or in vivo; e.g., mitochondrial DNA derived from different lines or species with traits distinct from the recipient cells. In this example, mitochondrial DNA from antibiotic resistant lines (e.g. IL8-8C/R53) is isolated and transformed into recipient cells that lack the resistant trait by using the transformation methods described in Example 2. Mitochondrial DNA for use in transformation can also be created by PCR amplification of organellar DNA by use of a primer set whose 3' ends are complementary with each other, sufficient for annealing in vivo. The resulted linear DNA molecules are transformed into mitochondria. Homologous recombination activity present in the organelle creates circular organellar DNA upon transformation. Alternatively, DNA for transformation can be created synthetically in a linear as well as a circular form.

Example 4

Introduction into Yeast Mitochondria of Donor DNA and Expression Cassettes for Cas9 and Guide RNA

[0648] In this example, a DNA plasmid ("Edit Plasmid") that can replicate in an organelle and encodes components of a site-specific endonuclease system such as Cas9, guide RNA and donor DNA is directly introduced into an organelle. The delivery of nucleic acids and proteins can be accomplished by utilizing methods such as bombardment ("biolistics"), electroporation and other suitable methods.

[0649] In yeast, DNA in a circular form with bacterial vector sequence (pBR322) can be transformed into mitochondria by utilizing a biolistic method. The resulted cells

were crossed with a line carrying a point mutation in mitochondrial DNA. They showed that the point mutation was recovered by recombination between the plasmid DNA and mitochondrial DNA. For efficient genome editing, a plasmid DNA to be transformed into yeast mitochondria is created with expression cassettes for Cas9 and guide RNA that are customized for expression in mitochondria. The plasmid DNA also contains donor DNA to facilitate site-specific genome editing. The Cas9 gene is optimized for mitochondrial expression (SEQ ID NO: 24) and is operably linked to a COX2 promoter and a terminator (SEQ ID NO: 25 and 26, respectively). The optimization is performed by changing CTN codons to TTA, GGG/GGC to GGT, GCG/GCC to GCT, CGG/CGC to CGT, CCG/CCC to CCT, AGC to AGT, AGG to AGA, ACG/ACC to ACT, TCG/TCC to TCT and GAG to GAA as well as TGA stop codon to TAA. The polynucleotide encoding a guide RNA that contains a variable targeting domain designed for the mitochondrial 21S rRNA gene (SEQ ID NO: 27) is operably linked to a promoter and terminator for the expression of the mitochondrial 15S rRNA gene (SEQ ID NO: 28 and 29, respectively). The donor DNA fragment carries the 21S rRNA gene with the chloramphenicol resistance allele, C^R321. The C^R321 mutation in the mitochondrial 21S rRNA gene can confer chloramphenicol resistance in yeast. For the selection of the plasmid in mitochondria, the plasmid can also carry a positive selectable marker such as active 15S rRNA gene with the paromomycin resistance mutation described above. This plasmid is transformed into mitochondria of yeast lines such as MCC123 [rho⁰] together with the other plasmid for nuclear transformation to select events of co-transformation of both plasmids in yeast. Transformed yeast cells are first colonized on media to allow the selection of nuclear transformants. By replica plating the colonized cells on the plates spread with a yeast line carrying the opposite mating type and wild-type mitochondrial genome, the colonies that are resistant to chloramphenicol are identified through subsequent replica-plating of mated cells on non-fermentable media such as YPGE with chloramphenicol (4 mg/ml). The increased frequency of chloramphenicol resistance colonies is confirmed by comparison with the frequency of chloramphenicol resistance colonies produced by the plasmid without Cas9 and guide RNA. Successful genome editing is further confirmed by sequencing of the edited site in mitochondrial DNA.

Example 5

Insertion of an Exogenous Gene into Mitochondrial DNA and Elimination of Unmodified Mitochondrial DNA

[0650] In this example, similar to Example 4, mitochondria are transformed with an Edit Plasmid. The Edit Plasmid contains an element that allows replication in mitochondria, and additional components of a site-specific endonuclease system such as Cas9, guide RNA and donor DNA. The donor DNA is designed to be bounded by two regions homologous to the mitochondrial genome for homologous recombination, which is facilitated by site-specific DNA cleavages. Between the two homologous regions, the insertion of an expression unit is demonstrated, consisting of a COXII promoter, a polynucleotide encoding GFP fluorescence protein and a terminator. The donor DNA can have multiple expression units with or without polycistronic

expression; i.e., where multiple coding regions are expressed under one promoter.

[0651] Two separate sites are targeted by Cas9-gRNA complexes in one demonstration. One Cas9 cleavage site in the COB gene is designed (variable targeting domain of: TGTCCCATTAAGACATAAGGTACTTCTACA SEQ ID NO:30; which precedes a TGG PAM sequence), and another cleavage site in the ATP9 gene (variable targeting domain of: TGGAGCAGGTATCTCAACAATTGGTTTATTAGGAGC SEQ ID NO:31; which precedes a AGG PAM sequence). One end of the donor DNA comprising polynucleotide covers the COB cleavage site and the other end covers the ATP9 gene to facilitate homologous recombination between the donor DNA and mitochondrial DNA. The donor DNA carries mutations in the sequence near the Cas9-gRNA cleavage sites to eliminate subsequent DNA cleavage after homologous recombination events. These mutations are designed to be “silent”; i.e., the mutated sequence has the same functionality as the wild type, such as replacement of one codon with a synonymous codon encoding the same amino acid. In addition to the modification at the cleavage sites, we also design Cas9-gRNA complexes that cleave additional sites between the two primary target sites in the wild-type mitochondrial DNA but not the donor DNA and the mitochondrial DNA produced by homologous recombination of donor DNA. Additional cleavage sites facilitate the “Genome Sweep” action; i.e., elimination of wild-type mitochondrial DNA without eliminating engineered mitochondrial DNA.

[0652] In a separate demonstration, the donor DNA contains a polynucleotide encoding lactoferrin in the place of GFP.

Example 6

Genome Editing of Mammalian Mitochondrial DNA

[0653] For Cas9 import into mammalian mitochondria, Cas9 protein without nuclear localization signal element is fused with a mitochondrial targeting peptide. One such peptide is NDUFV2 MTS which has 32 amino acid residues, NH2-MFFSAALRARAAGLTAHWGRHVRNLHKT-VMQN—COOH (SEQ ID NO:32). In this case, the NDUFV2 signal sequence is fused with the amino terminus of Cas9 to give a modified Cas9 (SEQ ID NO: 33). Alternatively, another signal peptide such as the one from citrate synthase (NH2-MALLTAAARLLGTKNASCLVLAARH—COOH; SEQ ID NO:34) that can function in human cells can be used to create a modified Cas9 (SEQ ID NO: 35). A polynucleotide encoding a modified Cas9 gene (with a mitochondrial target sequence) is operably linked to a promoter element such as CMV by utilizing the human transfection vector, pSF-CMV-Amp, purchased from Sigma Aldrich or is operably linked to a inducible promoter such as the TET-inducible promoter of pTRE2hyg vector, which can be purchased from Clontech.

[0654] Similar to other examples, guide RNA is fused to a mitochondrial targeting RNA; i.e., a sequence that allows import of RNA into mitochondria. In this experiment, RNAs that can be imported into human mitochondria are used. One of them is the yeast tRNA^{Lys}. The yeast tRNA^{Lys} and its variants can be imported into human mitochondria. The other RNA used is 5S rRNA, which can be imported into human mitochondria. In the latter case, the guide RNA is

cloned into Loop C that can be dispensable for mitochondrial import (SEQ ID NO: 36).

[0655] In this experiment, the guide RNA is designed to target the COX3 gene (SEQ ID NO: 37). In the guide RNA, the variable targeting domain is fused with the tracrRNA sequence as well as with a mitochondrial targeting RNA. The gRNA expression cassette consists of the polynucleotide encoding the guide RNA operably linked to a promoter and terminator that are functional in human cells. In this example, the U6 promoter for constitutive expression is used. For the 5S rRNA fusion, the promoter and terminator of the 5S rRNA gene (SEQ ID NO: 38) are also used. Guide RNA expression cassette is cloned into the plasmids carrying the Cas9 expression cassettes or cloned into distinct transfection vectors. Constructed plasmids are transfected into human cell lines such as HeLa and HEK293 as well as HeLa and HepG2 Tet-Off cells for Cas9 inducible expression from pTRE2hyg based constructs. Transfected cells undergo selection in the presence of hygromycin. Preparation of cell culture and transfection are performed for inducible expression.

[0656] Cells are harvested three days after transfection and total DNA of approximately 106 cells is extracted using a DNA extraction kit. PCR is conducted to amplify the regions encompassing the target sites and amplified DNA is deep sequenced by use of a high-throughput sequencer (e.g., MiSeq Illumina sequencer). The sequence data are analyzed to confirm modification at the target site.

Example 7

Genome Editing of Mammalian Mitochondrial DNA to Confer Resistance to Chloramphenicol

[0657] In this example with mammalian cells, mitochondrial DNA is edited to confer chloramphenicol resistance by a nucleotide substitution in the 16S rRNA gene. For the purpose, three components, Cas9 protein, guide RNA and donor DNA, are targeted to mitochondria.

[0658] The chloramphenicol resistance in a mouse cell line can be mapped to a single nucleotide change (CAP^R) in the mitochondrial 16S rRNA gene. The guide RNA is designed to include the CAP^R mutation site of the wild-type 16S rRNA gene. It is also designed in a manner that it will recognize the wild-type sequence but not the donor DNA with the CAP^R mutation (SEQ ID NO: 39). The donor DNA is produced by PCR amplification of the 16S rRNA region of the mouse CAP^R cells or is synthesized artificially (SEQ ID NO: 40).

[0659] Cas9 and guide RNA are targeted to mitochondria as described in Example 5. Plasmids with Cas9 and guide RNA expression cassettes are transfected into mouse cell lines such as NIH 3H3 as described above. The donor DNA is transformed into mitochondria. Transfected cells are cultured on media containing chloramphenicol (CAP). After the selection on CAP, the occurrence of resistant cells through genome editing is confirmed in comparison with controls. Finally, 16S rRNA of the CAP^R cells is sequenced to confirm genome editing at the molecular level.

Example 8

Introduction into Mammalian Mitochondria of Donor DNA and Expression Cassettes for Cas9 and Guide RNA

[0660] In this example, all components of genome editing including donor DNA are cloned in a plasmid DNA that is

introduced into mammalian mitochondria. The plasmid DNA is introduced into mitochondria either in a circular form or in a linear form that has the ability to circularize in mitochondria. The plasmid DNA contains sequence that allows for autonomous replication in mitochondria. It can also encode at least one selectable marker to allow for selection after transformation into mitochondria. Such a selectable marker can be the active 16S rRNA gene with CAP^R mutation. The rep/ori and other elements for gene expression in mitochondria present on the plasmid DNA may be derived from species different from the target species for mitochondrial DNA editing. Additional DNA cleavage sites can be designed for the wild-type sequences that differ from the donor DNA as described in previous examples.

Example 9

Introduction of Cas Endonuclease and Guide RNA into Plastids

[0661] To edit a chloroplast genome, Cas9 is modified to have a chloroplast targeting amino acid sequence (also known as transit peptide, TP) at the N-terminus of the protein and to remove any nuclear localization signal(s). In addition, the nucleotide sequence of Cas9 is codon-optimized for the plant species for optimum expression (SEQ ID NO: 41 & 42; for nucleic acid and amino acid sequences, respectively). The transit peptides from chloroplast-targeted proteins such as ribulose biphosphate carboxylase/oxygenase small subunit (rbcS), chlorophyll a/b binding protein (Cab) and DnaJ8 are used in the experiments. Each modified Cas9 is engineered to have a transit peptide fused translationally to the amino terminus of the Cas9 to create a TP-Cas9 (SEQ ID NO: 46). Expression of a polynucleotide encoding such a fusion protein is under control of a promoter functional in a plant, such as a CaMV 35S promoter. Cas9 without a transit peptide is used as a control (SEQ ID NO: 41 & 42).

[0662] For transport of a guide RNA into the chloroplast, RNA sequences are used that can import into the chloroplast. These plastid targeting RNAs (also referred to herein as “transit RNAs”), which can mediate import of attached heterologous RNA, include vd-5'UTR (SEQ ID NO: 48) and eIF4E1 mRNA (SEQ ID NO: 49). Transcription of polynucleotides encoding these fusion transcripts is under the control of a nuclear promoter functional in a plant, such as the 35S CaMV promoter (e.g., 1.3-kb 35S promoter of pBC-Yellow) or the U6 promoter; Chromosome 8 maize U6 polymerase III promoter). Guide RNA without a plastid targeting RNA serves as a control (SEQ ID NO: 50).

[0663] As an alternative method of creating gRNAs, a sequence-specific endoribonuclease is used, such as Csy4 which is responsible for processing CRISPR transcript from *Pseudomonas aeruginosa* (SEQ ID NO: 51-52, for nucleic acid and amino acid sequences, respectively). The Csy4 recognition sequence is: 5'-GTTCACTGCCGTATAG-GCAG-3' (SEQ ID NO: 53). Within the primary transcript, the gRNA sequence is flanked with Csy4 recognition sequences (SEQ ID NO: 54). A polynucleotide encoding this sequence fused with a 5' plastid targeting RNA is transcribed from either a 35S CaMV promoter or a U6 promoter in the nucleus and targeted into the chloroplast. For targeting Csy4 protein into the chloroplast, one of chloroplast transit peptides listed in SEQ ID NO: 43-45 is used, as an N-terminal translational fusion to Csy4.

Example 10

Introduction into Plastids of RNA Encoding Both Cas Endonuclease and Guide RNA

[0664] Plastid targeting RNA can transport heterologous RNAs into the plastid, which then are translated by the chloroplast translation machinery. This characteristic is utilized to transport all the genome editing components as RNA molecules into the chloroplast; transported mRNA is subsequently translated and the resulting proteins participate in the editing process. In this method, an expression cassette is made comprising a promoter operably linked to a polynucleotide encoding an RNA comprising the following: plastid targeting RNA, rbs (ribosome binding site), Cas9 coding sequence, rbs, Csy4 coding sequence, Csy4 recognition sequence, gRNA, and Csy4 recognition sequence. This expression cassette is integrated into the nuclear genome by transformation. The promoter in the above recombinant DNA construct is a promoter functional in a plant, such as a CaMV 35S promoter. The resulting RNA molecule is transported into chloroplast. Once it enters chloroplast, Cas9 and Csy4 proteins are produced by the chloroplast translation machinery. A complex of Cas9 and gRNA, which is processed from the transported RNA molecule by Csy4, finds and edits the target site in the chloroplast genome.

Example 11

Guide RNA Target Site Selection

[0665] Guide RNA target sites are selected from intergenic regions as well as genic regions of the chloroplast genome. The latter examples include rpoB, psbA, rps15, and rp133. Deletion of the rpoB gene can show a photosynthesis-defective phenotype. Deletion of the psbA gene can yield a photosystem II deficiency. Double deletion of rps15 and rp133 can result in synthetic lethality under autotrophic conditions. Use of web-based Bioinformatics program, APE (<http://biologylabs.utah.edu/jorgensen/wayned/ape/>), facilitates the selection process for gRNA target sites.

[0666] To select gRNAs target sites for *N. tabacum*, the *N. tabacum* chloroplast genome sequences are used. For gRNAs target sites for *N. benthamiana*, either public sequence deposition or direct sequencing of target regions in *N. benthamiana* chloroplast genome is used, as the total chloroplast genome sequence of *N. benthamiana* is not available. In addition, *N. tabacum* chloroplast DNA sequence is also used for the design of gRNA target sites for *N. benthamiana* since closely related plant species can have highly conserved chloroplast DNA sequences. Similarly, chloroplast *Glycine max* (strain: William 82) genomic sequence from Organelle Genome Resources at NCBI is used as a reference genome for designing tentative gRNA target sites in soybean chloroplast DNA, pending sequencing of the specific line that is transformed.

[0667] For editing of the indicated genic sequence regions, the following sequences are selected for variable targeting domains. The term “Nt” corresponds to “*Nicotiana tabacum*”, the term “Cp” corresponds to “Chloroplast” and the term “Glm” corresponds to “*Glycine max*”. When the variable targeting domain is on the reverse complement of the genic sequence, the term “reverse” is indicated.

For NtCp_rpoB (RNA polymerase beta chain) (SEQ ID NO: 55)

1. (SEQ ID NO: 56)
TTAGAGGAAGAGCCAAACAG

2. (SEQ ID NO: 57)
CTTGCTATAGCCGAACGCGA

For NtCp_psbA (photosystem II protein D1) (SEQ ID NO: 58)

1. (SEQ ID NO: 59)
GTTGATGAATGGTTATACAA

2. (SEQ ID NO: 60)
GATGATCCCTACCTTATTGA

For NtCp_rps15 (ribosomal protein S15) (SEQ ID NO: 61)

1. (SEQ ID NO: 62)
ATTTCTCAAGAAGAAAAGAG

2. (SEQ ID NO: 63)
TCAATTTACCAATAAGATA

For NtCp_rpl33 (50S ribosomal protein L33) (SEQ ID NO: 64)

1. (SEQ ID NO: 65)
GATATATTACTCAAAGAAG

2. (SEQ ID NO: 66)
AGTGTGATAAGGTATCAAG

For GlmaCp_rpoB (RNA polymerase beta chain) (SEQ ID NO: 67)

1. (SEQ ID NO: 68)
TGTCTAAACTACCTACAGG

2. (SEQ ID NO: 69)
AGCGGAATTTTCGGTCTATAC
(reverse)

For GlmaCp_psbA (photosystem II protein D1) (SEQ ID NO: 70)

1. (SEQ ID NO: 71)
GGTGTAGCTGGTGTATTCGG

2. (SEQ ID NO: 72)
TCTAGATCTAGCTGCGATCG
(reverse)

For GlmaCp_rps15 (ribosomal protein S15) (SEQ ID NO: 73)

1. (SEQ ID NO: 74)
ATAGAATACGAAGACTTACT
(reverse)

2. (SEQ ID NO: 75)
TGTCAAAGAAAGATAGAATA

For GlmaCp_rpl33 (50S ribosomal protein L33) (SEQ ID NO: 76)

-continued

1. (SEQ ID NO: 77)
CGTTGTTGCAAACATACAAT
(reverse)

2. (SEQ ID NO: 78)
ACAGAATACGCCTAGTCGAT

For *Nicotiana benthamiana* rps16 (ribosomal protein S16) (SEQ ID NO: 79)

1. (SEQ ID NO: 80)
TTGTGGATTTGTACATCCAC
(reverse)

2. (SEQ ID NO: 81)
TTGAACTGTTTGAAGTTAT
(reverse)

For *Nicotiana benthamiana* matK (maturase K) (SEQ ID NO: 82)

1. (SEQ ID NO: 83)
CTTGTGCTAGAACTTTAGCT

2. (SEQ ID NO: 84)
CGTTCATCTGGAAATCTTGG
(reverse)

[0668] For editing of the intergenic regions, the following sequences are selected for variable targeting domains.

Nicotiana tabacum:

1. (NtChrC;57408..57389) (SEQ ID NO: 85)
AAGAACTTCCCCCTTGACAG

2. (NtChrC;59412..59393) (SEQ ID NO: 86)
TATACAGGATGGGTAGAAAG

3. (NtChrC;59622..59603) (SEQ ID NO: 87)
ATATAATTTTAAATAAAGGG

4. (NtChrC;65704..65723) (SEQ ID NO: 88)
CTAGTCTTCGACACAAGAAA

Glycine max:

1. (GlmaCp_NC_007942.159039-59058) (SEQ ID NO: 89)
ATAACAGAAGTTAAAGAAGA

2. (GlmaCp_NC_007942.159100-59119) (SEQ ID NO: 90)
ATCTGGAACCATAGAACAG

3. (GlmaCp_NC_007942.162057-62038) (SEQ ID NO: 91)
CTATTTCGACACAAACAAGA

4. (GlmaCp_NC_007942.162361-62380) (SEQ ID NO: 92)
CTTTCTTTGACGAATTCGAG

Example 12

Transformation with Polynucleotides Encoding Cas Endonuclease and Guide RNAs

[0669] Gene cassettes encoding (a) Cas9 fused to a transit peptide; and (b) gRNA fused with vd-5'UTR or eIF4E1 mRNA as described above are subcloned into a binary

vector, such as pPZP and introduced into plants either for transient or for stable expression. DNA encoding Csy4 fused to a transit peptide is also transformed into plants in some experiments. Any of several methods may be used to transform plants with DNA sequences. These include agroinfiltration, biolistic bombardment, and floral dip method.

[0670] Similar approaches are also applicable for other plant species including dicots such as canola and monocots such as rice, wheat and corn.

Example 13

Introduction of Donor DNA into the Plastid Via Reverse Transcriptase

[0671] A donor DNA is introduced into the plastid genome to edit the genome in at least one way selected from the group consisting of: (1) creation of a point mutation in a target gene; (2) replacement of an endogenous coding region or regulatory sequence with a heterologous DNA sequence; and (3) insertion of a heterologous DNA sequence (e.g., for expression of a heterologous protein or RNA; for regulation of an endogenous gene).

[0672] In above examples several methods are presented for delivery of Cas9 and gRNAs into a chloroplast. In the current example, a donor DNA is also delivered into a chloroplast. In one method, a donor DNA for homologous recombination in a chloroplast is generated through reverse transcription of an RNA donor molecule which is transported into a chloroplast by transit RNA-guided transport. The RNA donor molecule, which is transcribed from transformed nuclear genome, contains the following: (1) a transit RNA, (2) sequences for homologous recombination; (3) a polynucleotide modification template sequence having at least one of the following: an endogenous sequence with an intended mutation (e.g., a site-specific mutation in the 16S rRNA) and a heterologous sequence (e.g., a heterologous protein coding sequence); and (4) a sequence that serve as a priming site for reverse transcriptase. In the homologous DNA regions, additional mutations, e.g., silent point mutations, are introduced into the sequence to distinguish these regions from additional gRNA target sites on the chloroplast DNA. The additional gRNA target sites are used to cleave non-transformed copies of chloroplast DNA. Reverse transcriptase protein is targeted into the chloroplast through a translational fusion with any of plastid targeting peptides described in SEQ ID NO: 43-45. Alternatively, an mRNA molecule (with a plastid rbs) encoding a reverse transcriptase is transported into the chloroplast as a fusion molecule with any one of plastid targeting RNAs described in SEQ ID NO: 48-49 and translated in chloroplast by the endogenous translation machinery.

Example 14

Introduction of Donor DNA into the Plastid Via Co-Bombardment with Two Polynucleotides

[0673] Another method to deliver donor DNA in conjunction with Cas9 and gRNAs is achieved through co-bombardment of two DNA molecules. In this approach, a first DNA molecule encoding Cas9 and gRNAs (employing chloroplast transport methods as described in previous examples) is targeted for transformation into the nuclear genome. A second DNA molecule, having a donor DNA sequence and homologous recombination sequences, is targeted for transformation into the chloroplast genome. The

second DNA molecule also can contain a chloroplast origin of replication. For transformation both DNA molecules are delivered to plant cells by biolistic bombardment. Biolistic particles are prepared as follows: (1) particles are coated with both DNA molecules either simultaneously or sequentially; or (2) particles are separately coated with each DNA molecule and then combined with the same molar ratio. For selection of nuclear transformation, commonly used antibiotic markers, such as nptII and bar, and/or fluorescent protein markers can be employed. For selection of chloroplast transformation, antibiotic markers such as aadA and/or fluorescent protein markers are used. The expression cassette for the chloroplast transformation selectable marker is either part of the donor DNA carrying polynucleotide that is integrated into the plastid genome or is placed outside of the donor DNA region, but remains on the delivered DNA molecule without being integrated into the chloroplast genome.

[0674] In a variation of above example of polynucleotide modification template delivery into the chloroplast, polynucleotides encoding Cas9 and gRNA (with or without Csy4) are transformed into the nuclear genome first. Gene expression of these components are under the control of inducible promoters. With the aid of selection markers (antibiotic markers and/or fluorescent marker proteins) stably transformed plants are selected. A second transformation is performed to transform chloroplast DNA with a DNA molecule containing a polynucleotide modification template DNA, homologous recombination sequences and a selectable marker such as aadA and/or a fluorescent marker protein. Selection of transformants is performed in the presence of selection agents for both nuclear and chloroplast transgenes and under conditions where the inducible promoter on the nuclear transgenes is active to transcribe Cas9 and gRNAs, which are subsequently transported into the chloroplast via the mechanism described in the previous examples.

Example 15

Introduction of Donor DNA into the Plastid Via *Agrobacterium*-Mediated Transformation

[0675] Donor DNA transport into the chloroplast is also performed via *Agrobacterium*-mediated transformation. A stable transgenic line which contains polynucleotides encoding Cas9 and gRNAs with an inducible promoter is created, as described above. This line is then transformed with a modified *Agrobacterium* strain, wherein the modification comprises the following: (1) addition of a chloroplast transit peptide fused to VirD2; (2) deletion of VirE2; and (3) removal of nuclear localization signals from VirD2. A binary vector is constructed having a polynucleotide modification template, homologous recombination sequences and a selection marker such as aadA and/or a fluorescent marker protein in between right and left T-DNA borders and transformed into *Agrobacteria*. For transformation, stable transgenic lines with polynucleotides encoding Cas9 and gRNAs are incubated with *Agrobacteria*. VirD2 protein which is covalently linked to single-stranded T-DNA enters into plant cells and is transported into the chloroplast via the N-terminal transit peptide. Transgenic selection is imposed by dual selection with nuclear (nptII) and chloroplast (aadA) markers and under conditions where the inducible promoter is active to transcribe polynucleotides encoding Cas9 and

gRNAs, which are subsequently transported into chloroplast by the mechanism described in the previous examples.

Example 16

Introduction into Plastids of Donor DNA and Expression Cassettes for Cas9 and Guide RNA

[0676] In this example, a DNA plasmid (“Edit Plasmid”) that can replicate in plastids and encodes components of a site-specific endonuclease system such as Cas9, guide RNA and donor DNA, is directly introduced into the plastid. The delivery of nucleic acids and proteins can be accomplished by use of methods such as bombardment (biolistics), electroporation and other available methods. Here an example in tobacco chloroplasts is shown.

[0677] The Edit Plasmid for tobacco chloroplasts is constructed as follows. Polynucleotides encoding Cas9 and guide RNA are cloned into the vector and are operably linked to appropriate promoters and terminators to allow for expression in tobacco chloroplasts. Alternatively, these two coding regions may be linked and transcribed polycistronically under one promoter. The polycistronic RNA may be processed to give rise to separate functional RNA molecules for genome editing, one for Cas9 translation and the other for guide RNA. A polynucleotide encoding a selectable marker that enables selection of the plasmid in chloroplasts, such as the *aadA* gene conferring spectinomycin resistance, is also present on the plasmid DNA, operably linked with an appropriate promoter and a terminator active in chloroplasts. An expression cassette encoding a negative selectable marker gene is also present on the plasmid to allow for counter selection, i.e., selection of chloroplasts without Edit Plasmid after editing and subsequent elimination of wild-type chloroplast DNA has been achieved. The *dao* gene is one such negative selectable marker gene. Furthermore, an element that allows for replication of the Edit Plasmid is also present in the vector. Such an element can be derived from the chloroplast DNA of the target species or alternatively from chloroplast DNA of another species, as well as from completely synthetic sources. In addition, donor DNA is present on the vector to allow for precise DNA editing and/or the precise insertion of heterologous DNA elements at specific sites in the chloroplast DNA.

[0678] As one example, the wild-type *psbA* gene in tobacco chloroplast DNA is replaced with an allele carrying a single nucleotide substitution that confers resistance to the herbicide triazine. Such a mutation can be present in herbicide tolerant plants in nature. For DNA cleavage in the vicinity of the mutation site, guide RNA to target the following DNA sequence is designed.

(SEQ ID NO: 93)

ACGAGAGTTGTTGAACTAGCATAT**TGGA**AGATCAA

[0679] The PAM sequence (TGG) is in bold font.

[0680] The donor DNA contains the following sequence with five mutations shown in bold font.

(SEQ ID NO: 94)

ACGAGAGTT**ATT**TGAAT**GT**AGCATACT**GAA**AGATCAA

[0681] The atrazine resistance mutation (G) is underlined. The four additional changes that do not alter protein

sequence are present to eliminate the donor DNA as being a target for the guide RNA designed for the endogenous wild-type *psbA* sequence. In particular, one change eliminates the PAM sequence critical for guide RNA pairing to the target polynucleic acid (e.g., target DNA) sites.

[0682] To facilitate homologous recombination, the donor DNA is bounded by longer homologous sequences upstream and downstream of the above sequence.

[0683] The Edit Plasmid is transformed into tobacco chloroplast by the biolistic approach as described in Chloroplast Biotechnology Methods and Protocols, Pal Maliga (Editor), Methods in Molecular Biology, Springer, New York (2014) (Cells with transformed chloroplasts are selected on the media containing spectinomycin. After the cultivation of callus cells on the selective media, calli are transferred to the media containing atrazine to assess the frequency of site-specific genome editing with the donor DNA. Sequencing of callus cells resistant to the herbicide confirms the successful genome editing at the molecular level.

[0684] To increase the rate of obtaining homoplasmic chloroplasts with engineered DNA, additional target sites are designed in the wild-type sequence covered by the corresponding homologous regions adjacent to the donor DNA. To protect the donor DNA and edited DNA in the chloroplast, donor DNA harbors silent mutations that avoid cleavage by Cas9 endonuclease; e.g., replacing codons with synonymous codons coding for the same amino acids. Expression cassettes encoding the gRNA(s) corresponding to those additional target sites are cloned into the Edit Plasmid vector for expression in chloroplasts. The donor DNA with the additional gRNA target sites mutated (for protection from Cas9 endonuclease) is also present in the Edit Plasmid.

[0685] The above Edit Plasmid with increased Genome Sweep activity is transformed into tobacco chloroplast as described above. Cells with transformed chloroplasts are selected on the media containing spectinomycin. After the cultivation of callus cells on the selective media, calli are transferred to the media containing atrazine to assess the frequency of site-specific genome editing with the template DNA. Sequencing of callus cells resistant to the herbicide confirms the successful genome editing at the molecular level.

[0686] When stable inheritance of edited organellar DNA is achieved, the Edit Plasmid can be segregated out in progeny plants under non-selective conditions for the Edit Plasmid. The segregation process can be facilitated by utilizing the negative selectable marker encoded in the Edit Plasmid, e.g., D-valine selection for the *dao* gene.

Example 17

Regulatory Elements for Plastid Gene Expression

[0687] Expression cassettes may be constructed that have a promoter functional in a plastid operably linked to either: (a) a donor polynucleotide; or (b) a plurality of donor polynucleotide arranged as a polycistronic unit. A desired 5'-UTR can also be present in the expression cassette, operably linked to the 3'-end of the promoter.

[0688] In one expression cassette, the polynucleotide (or polynucleotides) to be transcribed can be operably linked to the following promoter::5'-UTR regulatory elements:

[0689] (a) the maize *clpP* promoter in combination with the maize *clpP* 5'-UTR;

- [0690] (b) the maize clpP promoter in combination with the 5' UTR from gene 10 of bacteriophage T7;
- [0691] (c) the tomato psbA promoter in combination with the T7g10 5'-UTR; and
- [0692] (d) the tomato rml6 promoter in combination with the accD-mod 5'-UTR.
- [0693] The above regulatory elements can be obtained by PCR amplification.

Example 18

Pest Resistance Genes for Expression in Organelles

[0694] An expression cassette for use in organelle transformation is constructed using the wild-type nucleic acid sequence from *Bacillus thuringiensis* serovar kurstaki (U89872; SEQ ID NO:108) encoding the full-length native HD73 Cry1Ac delta-endotoxin (SEQ ID NO:109). Alternatively, a truncated native nucleic acid sequence (SEQ ID NO:110) is used, which encodes the active truncated Cry1Ac fragment. Additionally, in some cases, the nucleic acid sequence encoding the full-length or truncated Cry1Ac protein is codon-optimized for the organelle of interest.

[0695] In some cases, additional polynucleotides that encode proteins useful in conferring insect resistance to a plant are included in the above expression cassette as a polycistronic unit, or are expressed from separate expression cassettes. These polynucleotides encode the following: (a) the Cyt1Aa protein from *Bacillus thuringiensis* serovar israelensis (Gene ID: 5759908; SEQ ID NO:111); (b) the 20 kDa accessory protein from *Bacillus thuringiensis* serovar israelensis (pBt024; SEQ ID NO:112); and (c) the 19 kDa accessory protein from *Bacillus thuringiensis* serovar israelensis, (pBt022; SEQ ID NO:113).

Example 19

Engineered Plant with Increased Pest Resistance

[0696] In this example, a plant (e.g., soybean plant) is engineered with increased resistance to pests. Optionally, the plant also has increased resistance to herbicides.

[0697] The site-specific endonuclease system (e.g., Cas9, guide RNA, and donor DNA) of the disclosure is used to introduce one or more pesticidal proteins into the organellar (e.g., plastid) genome of a plant cell (e.g., soybean cell). The one or more pesticidal proteins or their fragments are selected from the group consisting of: Cry1Ac, Cyt1Aa (e.g., SEQ ID NO:109 or SEQ ID NO:110), Cry1Ab, Cry2Aa, CryII, Cry1C, Cry1D, Cry1E, Cry1Be, Cry1Fa and Vip3A.

[0698] In some cases, one or more accessory proteins are also introduced into the organellar (e.g., plastid) genome of the plant cell. The one or more accessory proteins can bind to a pesticidal protein and are selected from the group consisting of: a 20 kDa accessory protein and a 19 kDa accessory protein.

[0699] Additionally or independently, in some cases, the site-specific endonuclease system (e.g., Cas9, guide RNA, and donor DNA) is used to introduce one or more heterologous donor polynucleotides encoding a dsRNA, a siRNA, and/or a miRNA, wherein the dsRNA, the siRNA and the miRNA can suppress at least one target gene present in a plant pest, into the organellar (e.g., plastid) genome of the plant cell (e.g., the soybean cell). The dsRNA, the siRNA and the miRNA can suppress at least one target gene selected

from the group consisting of: proteasome A-type subunit peptide (Pas-4), ACT, SHR, EPIC2B and PnPMAl. The RNA interference-based mechanism can be used to protect the engineered plants from pests.

[0700] Optionally, in some cases, one or more herbicide tolerance proteins is also introduced into the organellar (e.g., plastid) genome of the plant cell using the site-specific endonuclease system (e.g., Cas9, guide RNA, and donor DNA) of the disclosure. The herbicide tolerance protein can be at least one selected from the group consisting of: a 4-hydroxyphenylpyruvate dioxygenase (HPPD), a sulfonylurea-tolerant acetolactate synthase (ALS), an imidazolinone-tolerant acetolactate synthase (ALS), a glyphosate-tolerant 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), a glyphosate-tolerant glyphosate oxidoreductase (GOX), a glyphosate N-acetyltransferase (GAT), a phosphinothricin acetyl transferase (PAT), a protoporphyrinogen oxidase (PROTOX), an auxin enzyme or receptor, a P450 polypeptide and an acetyl coenzyme A carboxylase (ACCase).

Example 20

Genetic Modification of Yeast Mitochondrial DNA by the Edit Plasmid Approach

[0701] To show mitochondrial genome editing with our methodology in yeast, *Saccharomyces cerevisiae*, various Edit Plasmid constructs were designed. The reference sequence used was a complete mitochondrial genome sequence from the *Saccharomyces* Genome Database (SGD), <https://www.yeastgenome.org/>. The targeted gene was the COX1 gene (also called oxi3 gene). Mutants of this gene previously have been shown to have a respiration-defective phenotype (<https://www.yeastgenome.org/locus/S000007260>). The following four guide RNA target sites in the COX1 gene were used (when the targeting sequence was on the reverse complement of the genic sequence, the term "reverse" is indicated):

- 1) (SEQ ID NO: 116)
TTCTTTGAAGTATCAGGAGGTGG;
- 2) (SEQ ID NO: 117)
ATGATTATTGCAATTCCAACAGG;
- 3) (SEQ ID NO: 118)
GCTATTTTGTAGTGGTATGGCAGG;
and
- 4) (SEQ ID NO: 119)
ACCATGTAAATATTGTGAACCCAGG.
(reverse)

[0702] The last three nucleotides in each sequence correspond to the PAM sequence. The first target site resided in exon 5, the second in exon 4, the third one in exon 1 and the fourth one at the junction of 3' end of exon 1 of the mitochondrial COX1 gene. Each Edit Plasmid contained a guide RNA expression cassette encoding guide RNA(s) directed to either one or two of the four COX1 target sites. The variable targeting domain of each guide RNA did not contain the 3-nucleotide PAM sequence listed above.

[0703] Yeast mitochondrial transformation was performed by following the protocol developed by the Fox lab (Fox et

al. 1988 Proc Natl Acad Sci USA 85:7288-7292; Bonnefoy and Fox 2001 Methods Enzymol 350:97-111). It previously has been shown that plasmids derived from pBR322 were capable of replicating in yeast mitochondria (Fox et al. 1988). One of the plasmids derived from pBR322, pHD6 was used, and the plasmids had been successfully transformed into yeast mitochondria in the past (Green-Willms et al. 2001 J Biol Chem 276: 6392-6397). All cloned fragments of pHD6 by digesting with PstI and HindIII are deleted except the genomic fragment of COX3 gene to leave the pBR322 backbone for creating our constructs. The COX3 fragment (0.75 kb PacI-MboI) was used as a screenable marker for mitochondrial transformants with its capability to rescue the *cox3* deletion mutant *cox3-10* as described in Fox et al., 1988. The Edit Plasmid constructs contained the following elements in the pBR322 backbone: Cas9 expression cassette, guide RNA expression cassette and donor DNA in the case of DNA replacement experiments. The Cas9 expression cassette had a Cas9 coding sequence that was optimized for the expression in yeast mitochondria (SEQ ID NO: 120). As part of codon optimization, the Cas9 codons that were not used at all or were rarely used in yeast mitochondria were replaced with codons that were used frequently. Also, a number of tryptophan codons were replaced with TGA, which is a stop codon in the universal codon table but is translated into tryptophan in yeast mitochondria (Fox 1979 Proc Natl Acad Sci USA 76: 6534-6538). This was designed to prevent expression of Cas9 in the nucleus after microprojectile DNA transformation. The expression cassette with the optimized Cas9 ORF was synthesized with the minimal promoter with 5' UTR and terminator of the COX2 gene; these regulatory elements were flanked with PstI and HindIII sites, respectively (SEQ ID NO: 121 and SEQ ID NO: 122). The minimal promoter and terminator, which had the length of 71 and 119 bp, respectively (Mireau et al. 2003 Mol Gen Genomics 270:1-8), were chosen with the purpose of suppressing homologous recombination at the sites and avoiding integration into the mitochondrial genome. Several unique restriction sites (XbaI, NotI and NcoI sites) were included at the HindIII end to facilitate cloning of additional elements. One such element was the guide RNA expression cassette. Guide RNAs targeting the COX1 sequences described above were created by fusion of each targeting sequence with the tracrRNA sequence (SEQ ID NO: 123). Each guide RNA expression cassette encoded either one or two guide RNAs, which were directed to the corresponding one or two of the four COX1 target sites.

[0704] The guide RNA expression cassette contained the following elements in 5' to 3' orientation: a minimal COX3 promoter (SEQ ID NO: 124); a tRNA gene, tF(GAA) (SEQ ID NO: 125); a single guide RNA directed to a COX1 site; a second tRNA gene, tW(UCA) (SEQ ID NO: 126); and a minimal COX3 terminator element (SEQ ID NO: 127). The constructs with two guide RNAs were created by combining

guide RNAs directed to COX1 sites 1 and 2, as well as to sites 3 and 4. When two guide RNA encoding sequences were present, the second one was fused directly after the tW(UCA) sequence and was flanked by a third tRNA gene, tM(CAU) (SEQ ID NO: 128) at the 3' end and before the COX3 terminator. The guide RNA expression cassettes with promoter and terminator elements were synthesized with a NotI site at the 5' end and a NcoI site at the 3' end to allow directional cloning into the pBR322 backbone that carries the Cas9 expression cassette.

[0705] For the DNA replacement experiments, the donor DNA carrying the GFP gene was synthesized and cloned into the NcoI site of constructs that encoded two guide RNAs. The nucleotide sequence (SEQ ID NO: 129) encoding GFP was codon optimized for expression in yeast mitochondria as done for Cas9 (see above). Several codons for tryptophan were changed to TGA, assuring GFP expression only in mitochondria. Also, the GFP coding region was designed to be in frame with the COX1 gene after DNA replacement. Both ends of the GFP ORF were fused with the COX1 genomic sequences at the external junction of the Cas9 cleavage sites. HR1-HR4 correspond to four short homology regions used in construction of the Edit Plasmids; they were each immediately adjacent to the corresponding guide RNA target site. The length of the homologous region at each end was chosen to be relatively short to minimize endogenous homologous recombination without Cas9 cleavages, i.e. 144 bp adjacent to the #1 guide RNA site (HR1; SEQ ID NO: 130), 115 bp adjacent to the #2 guide RNA site (HR2; SEQ ID NO: 131), 64 bp adjacent to the #3 guide RNA site (HR3; SEQ ID NO: 132) and 93 bp adjacent to the #4 guide RNA site (HR4; SEQ ID NO: 133). This design should facilitate DNA replacement induced by Cas9 activity and not by general homologous recombination. Additionally, the Edit Plasmids should remain autonomous without integrating into the genome. Furthermore, sequence variations were included at the guide RNA recognition sites within the donor DNA, so that the mitochondrial DNA after replacement would no longer be recognized by the guide RNA/Cas9 complex. This was done to prevent the deletion of the replaced DNA from the gene-edited mitochondrial genome. The variant of the first target site is listed under SEQ ID NO: 134, where 7 of the 20 nucleotides in the guide RNA recognition site have been changed. The variant of the second site was created by deleting 16 nucleotides at the 5' end of the recognition site (SEQ ID NO: 135). The third target site was modified by deleting the last five nucleotides (SEQ ID NO: 136). The fourth target site was modified by deleting 14 nucleotides at the 5' end (SEQ ID NO: 137).

[0706] The constructs made for this experiment are presented in Table 1.

TABLE 1

Components of Edit Plasmids for Yeast Mitochondria			
Construct	Expr Cassette 1*	Expr Cassette 2**	Donor DNA
HS1	Cas9m	tF:sgRNA-3:tW	N/A
HS2	Cas9m	tF:sgRNA-4:tW	N/A
HS3	Cas9m	tF:sgRNA-3:tW:sgRNA-4:tM	N/A

TABLE 1-continued

Components of Edit Plasmids for Yeast Mitochondria			
Construct	Expr Cassette 1*	Expr Cassette 2**	Donor DNA
HS4	Cas9m	tF:sgRNA-3:tW:sgRNA-4:tM	HR3:GFPm:HR4
HS5	Cas9m	tF:sgRNA-3:tW:sgRNA-4:tM	HR3:GFPm:HR4***
HS6	N/A	tF:sgRNA-2:tW:sgRNA-1:tM	HR1:GFPm:HR2
HS7	N/A	tF:sgRNA-3:tW:sgRNA-4:tM	HR3:GFPm:HR4
HS8	Cas9m	tF:sgRNA-2:tW:sgRNA-1:tM	HR1:GFPm:HR2
HS9	Cas9m	tF:sgRNA-2:tW:sgRNA-1:tM	N/A
HS10	Cas9m	tF:sgRNA-1:tW	N/A
HS11	Cas9m	tF:sgRNA-2:tW	N/A

*Each Expression Cassette 1 had the COX2 promoter with 5' UTR and the COX2 terminator.

**Each Expression Cassette 2 had the COX3 promoter and the COX3 terminator.

***The Donor DNA is in reverse orientation with respect to the construct HS4.

[0707] The constructs created were transformed into yeast lines that lacked mitochondrial DNA (ρ^0), MCC109 ρ^0 (MAT α ade2 ura3 kar 1), using the biolistic microprojectile method as described in Bonnefoy and Fox, 2001. The transformation was performed together with pYES2 as a carrier plasmid with URA3 selectable marker, so that URA⁺ nuclear transformants could be selected first on minimal medium lacking uracil in supplements. To identify mitochondrial transformants, URA⁺ colonies were assayed for the ability of rescuing a cox3 deletion mutant through a cross with MCC125 (MAT α lys2 ρ^+ cox3-10). The assay was repeated at least twice to obtain clean colonies with Edit Plasmids in the mitochondria. Isolated lines containing Edit Plasmids were then crossed with lines containing the wild-type mitochondrial genome, CUY563 (MAT α ura3 ade2 leu2 ade3 ρ^+) and NB80 (MAT α lys2 arg8 ura3 leu2 ρ^+), to analyze the genome editing effect by Cas9 at the target sites. In nuclear chromosomes subjected to double-strand breaks, one might expect a high frequency of mutations such as small deletions or insertions at the target sites. They are the results of Non-Homologous End-Joining (NHEJ) repair at the site of DNA cleavage triggered by the guide RNA dependent Cas9 activity. In yeast, 90% of the repair of double-strand breaks in chromosomes occurs by homologous recombination (Ricchetti et al. 1999 Nature 402:96-100). In mitochondria, where multiple copies of mitochondrial DNA are present in one organelle, the repair of dsDNA breaks through homologous recombination is expected to be significantly more frequent than in the nucleus. Under this circumstance, the frequency of indel mutations caused by re-ligation of DNA ends is expected to be extremely low in mitochondria. Due to this consideration, we focused on the detection of events caused by repair through homologous recombination, i.e., replacement with artificial donor DNA.

[0708] To assay for DNA replacement through Cas9 induced cleavages, the construct HS8 and its control construct HS6 were each transformed into a strain lacking mitochondrial DNA as described above. Each construct carried the donor DNA with GFP as well as two corresponding guide RNA genes (#1 and #2) but HS6 lacked the Cas9 expression cassette. Lines that contained each construct were identified by subsequent screening for their capability of rescuing the cox3 deletion mutant. The isolated mitochondrial transformants then were crossed with strains carrying the wild-type mitochondrial genome, CUY563 and NB80, to observe the effect of Edit Plasmids on the mitochondrial genomic DNA. The DNA replacement events at the cleavage sites then were assayed by PCR amplification

of pooled cells two days after the crosses. Primer sets were used wherein one primer was from the mitochondrial genomic region in the vicinity of the cleavage sites and the other primer was from the donor DNA region, selected so that the desired PCR product could only be amplified from a correctly replaced DNA in the mitochondrial genome but not from the wild-type mitochondrial DNA nor from the Edit Plasmid. The following four primer pairs were used: primers C and 12 for the 5' end junction; and for the 3' end junction, primers D and 11, E and 11, and F and 11. Primers C, D, E and F were specific to the genomic region of the COX1 gene (SEQ ID NO: 138, 139, 140 and 141, respectively). Primers 11 and 12 were specific to the GFP gene (SEQ ID NO: 142 and 143, respectively). The PCR amplification was performed as follows: Step 1: 94° C. for 7 min, step 2: 94° C. for 30 sec, step 3: 52° C. for 30 sec, step 4: 60° C. for 1 min 30 sec, step 5: go to step 2 for 39 times, step 6: 60° C. for 10 min. The low temperature for the extension reaction was chosen to accommodate AT-rich genomic sequences. After PCR amplification, we observed the expected size of the DNA fragments from each end of the replaced DNA by using the above four distinct pairs of primers. No corresponding DNA fragments were amplified in the cell samples that were crossed with the line carrying the control construct.

[0709] The amplified DNA fragments were sequenced directly. FIG. 1 presents the sequence (SEQ ID NO: 144) obtained from PCR amplification of the replaced DNA locus in transformed yeast mitochondrial DNA modified by the Edit Plasmid approach. Underlined sequences at the 5' and 3' ends indicate wild-type mitochondrial genomic sequences that are not present on the Edit Plasmid. Sequences in bold font indicate the short homologous regions present in the donor DNA and adjacent to the corresponding guide RNA target sites. Sequences that have double underlining indicate the modified guide RNA target sites present in the donor DNA; altered nucleotides are shown in bold font. The guide RNA target sites in the replaced DNA have been modified to prevent nuclease activity after integration into the mitochondrial genome. The codon-optimized GFP coding region is presented in italics. Sequences presented in lower case correspond to primers C and F that were used for amplification of the replaced DNA locus. Homologous recombination occurred as expected; i.e., there were no sequence changes either in the replaced DNA or in the surrounding wild-type mitochondrial DNA.

[0710] The sequence (SEQ ID NO: 144; FIG. 1) covering the replaced region matched with the construct completely.

Also shown in FIG. 1 are sequences at the 5' and 3' ends (shown with underlining) that are wild-type mitochondrial genomic sequences not present on the Edit Plasmid, which are contiguous to the HR regions (shown in bold font) present in the Edit Plasmid. In summary, DNA replacement was observed in yeast mitochondria by use of an Edit Plasmid that encodes a Cas9 expression cassette, a multiple guide RNA expression cassette and a donor DNA template.

[0711] Furthermore, single colonies were isolated from the cross between the HS8 line carrying the Edit Plasmid and wild-type strain, NB80. GFP signal was confirmed from a fraction of colonies when viewed through a fluorescence microscope.

[0712] In order to show the autonomously replicating nature of the Edit Plasmids in mitochondria, we attempted the rescue of plasmids from the cells after the crosses described above. 1 ml of overnight cell culture after each cross was sampled and subjected to the total DNA isolation. 200 ng of total DNA obtained by use of the Quick-DNA Miniprep Plus Kit (Zymo Research) were digested with *ApaI* and *SphI* to cleave pYES2 plasmid DNA in the total DNA fraction; the HS8 plasmid should remain intact as it doesn't possess these restriction sites. After inactivating the restriction enzymes at 65° C. for 20 min, the DNA was used to transform *E. coli* cells. Multiple colonies that grew on LB medium containing carbenicillin were identified. DNA was isolated, subjected to digestion with several restriction enzymes, and the digestion products were separated by gel electrophoresis. A number of plasmids were identified from two independent crosses that showed a digestion pattern identical to the original HS8 construct, demonstrating that rescue of the original Edit Plasmid HS8 was successful. This showed that Edit Plasmids remained as autonomously replicating DNA in the presence of wild-type mitochondrial DNA, not integrated into the organelle genome.

Example 21

Genetic Modification of *Chlamydomonas reinhardtii* Chloroplast DNA by the Edit Plasmid Approach

[0713] Guide RNA target sites were selected from genic regions of the *Chlamydomonas reinhardtii* chloroplast genome. The reference sequence used was a complete chloroplast genome sequence from NCBI (Accession number: NC_005353 and Version number: NC_005353.1). The targeted gene was *psaA*. Mutants of this gene previously have been shown to have a photosynthesis-defective phenotype (Redding et al. 1999, J Biol. Chem. 274: 10466-10473). To help design and select guide RNA target sites, a web-based Bioinformatics program was employed—CRISPOR (<http://crispor.tefor.net/>, Haeussler et al. 2016 Genome Biology 17:148-159). The following sequences were selected as guide RNA targeting sites for editing of exon 3 in the *psaA* gene. When the targeting sequence was on the reverse complement of the genic sequence, the term “reverse” is indicated. For each 23 nucleotide target site listed below, the first 20 nucleotides are the targeting sequence present in each corresponding guide RNA and the last 3 nucleotides are the PAM sequence.

1. (SEQ ID NO: 145)
GGTTTAAACCTGTTACTGGTGG
2. (SEQ ID NO: 146)
CTTCACCTGTAAATGGACACGG
(reverse)
3. (SEQ ID NO: 147)
TTTACAGGTGAAGGTACGTTGG
4. (SEQ ID NO: 148)
GTAGCTAAATAAGGGTATGGAGG
(reverse)

[0714] FIG. 2 presents the sequence (SEQ ID NO: 171) obtained from PCR amplification of the replaced DNA locus in transformed *Chlamydomonas* plastid DNA modified by the Edit Plasmid approach. Underlined sequences at the 5' and 3' ends indicate wild-type chloroplast genomic sequence that is not present on the Edit Plasmid. Sequences in bold font indicate the short homologous regions present in the donor DNA on the Edit Plasmid. Sequences that are both in bold font and underlined indicate guide RNA target sites present in the replaced DNA. The guide RNA target sites in the donor DNA have been modified to prevent nuclease activity after integration into the plastid genome. Sequences that have double underlining indicate silent mutations at the 3' side of guide RNA sites to preclude re-cleavage by Cas9/sgRNA. The codon-optimized GFP coding region is presented in italics. Homologous recombination occurred as expected; i.e., there were no sequence changes either in the replaced DNA or in the surrounding wild-type plastid DNA.

[0715] The Edit Plasmids for *Chlamydomonas* chloroplasts were constructed as follows. Polynucleotides encoding Cas9 and guide RNA were cloned into the vector and were operably linked to appropriate promoters and terminators to allow for expression in chloroplasts. The vector was either pBR322 or pUC19, each of which contained the replication origin of pMB1 which previously was shown to replicate in chloroplasts (Boynton et al. 1988 Science 240: 1534-1538).

[0716] The nucleic acid sequence (SEQ ID NO: 149) encoding SpCas9 (SEQ ID NO: 150) was codon-optimized for *Chlamydomonas* chloroplast expression. The optimization was performed using a web-based Codon Usage Database (Nakamura et al. 2000 Nucleic Acids Res. 28: 292). The optimized gene was synthesized by GenScript (Piscataway, N.J.). The promoter used for Cas9 gene expression was either the *Chlamydomonas* *psaA*-exon 1 promoter with its 5' UTR or the *Chlamydomonas* *psbD* promoter with its 5' UTR (SEQ ID NO: 151 & SEQ ID NO: 152, respectively). The terminator used for Cas9 gene expression was the *rbcL* 3' UTR (SEQ ID NO: 153).

[0717] For expression of sgRNA, a tRNA promoter and its corresponding 3' UTR (SEQ ID NO: 154 and SEQ ID NO: 155, respectively) were derived from the *Chlamydomonas* plastid *trnW* gene locus. For the proper processing of sgRNA after transcription, the endogenous chloroplast tRNA processing system was utilized as described in Xie et al. 2015 (Proc Natl Acad Sci USA 112: 3570-3575). For example, for expression of one guide RNA, a sgRNA sequence was placed between two tRNAs. The configuration was “tRNA-1-sgRNA-tRNA-2”. For expression of two sgRNAs, the configuration was “tRNA-1-sgRNA-1-tRNA-2-sgRNA-2-

tRNA-3". The following tRNA sequences from *Chlamydomonas* plastid DNA: trnW (SEQ ID NO: 156), trnK (SEQ ID NO: 157), and trnL (SEQ ID NO: 158) were employed.

[0718] A selectable marker expression cassette for the aadA coding region (SEQ ID NO: 159), to provide spectinomycin resistance, was also present on all the Edit Plasmid constructs. The promoter and terminator for the selectable marker expression cassette were the *Chlamydomonas* rbcL promoter with its 5' UTR (SEQ ID NO: 160) and the *Chlamydomonas* psbA 3' UTR (SEQ ID NO: 161), respectively. Plasmids that carried only a Cas9 expression cassette and selectable marker expression cassette were constructed for use as controls.

[0719] For DNA replacement experiments, donor DNA was designed which consisted of a GFP coding region surrounded by homologous recombination regions. The GFP coding sequence (SEQ ID NO: 162) was designed to be codon-optimized for *Chlamydomonas* chloroplast gene expression according to the method of Franklin et al. 2002 (Plant J 30: 733-744). For homologous recombination of the donor DNA after double-strand breaks by Cas9/double sgRNAs, we selected homologous regions of 74 or 76 bp each (HR1-HR4; SEQ ID NO: 163-SEQ ID NO: 166) from gRNA target sites in the *Chlamydomonas* chloroplast gene, psaA-Exon 3. The short length (74 or 76 bp) of each homologous sequence was chosen to minimize the occurrence of endogenous homologous recombination without double-strand breaks mediated by Cas9/guide RNA (Dauvillee et al. 2004 Photosynthesis Research 79: 219-224). The configuration of the donor DNA with its components is "1st HR-GFP-2nd HR". The GFP sequence was derived from Franklin et al. 2002 (Plant J. 30:733-744). To protect the donor DNA from further cleavage by Cas9 and to facilitate the Genome Sweep process, homologous recombination sequences also contained silent mutations at the target sites that precluded cleavage by Cas9 and guide RNAs. Homologous recombination was designed to give an in-frame fusion of GFP with the psaA gene product. Components in the Edit Plasmids for DNA replacement experiments included donor DNA as well as the Cas9, double sgRNAs and selectable marker expression cassettes described in the previous section. The same vector backbone was used as in the previous section, as well. As negative controls, plasmids lacking the Cas9 expression cassette were used.

[0720] Tables 2 and 3 list the components of the constructs described in this section.

TABLE 2

Components of Edit Plasmids for <i>Chlamydomonas</i> Chloroplasts			
Construct	Expr Cassette 1*	Expr Cassette 2**	Donor DNA
YP5	P _{psaA} :Cas9co	N/A	N/A
YP7	P _{psaA} :Cas9co	1X-sgRNA-1	N/A
YP8	P _{psaA} :Cas9co	1X-sgRNA-2	N/A
YP9	P _{psaA} :Cas9co	1X-sgRNA-3	N/A
YP10	P _{psaA} :Cas9co	1X-sgRNA-4	N/A
YP11	P _{psaA} :Cas9co	2X-sgRNA-1	N/A
YP12	P _{psaA} :Cas9co	2X-sgRNA-2	N/A
YP13	P _{psaA} :Cas9co	2X-sgRNA-1	HR1:GFPco:HR2
YP14	P _{psaA} :Cas9co	2X-sgRNA-2	HR3:GFPco:HR4
YP6	P _{psbD} :Cas9co	N/A	N/A
YP15	P _{psbD} :Cas9co	1X-sgRNA-1	N/A
YP16	P _{psbD} :Cas9co	1X-sgRNA-2	N/A
YP17	P _{psbD} :Cas9co	1X-sgRNA-3	N/A

TABLE 2-continued

Components of Edit Plasmids for <i>Chlamydomonas</i> Chloroplasts			
Construct	Expr Cassette 1*	Expr Cassette 2**	Donor DNA
YP18	P _{psbD} :Cas9co	1X-sgRNA-4	N/A
YP19	P _{psbD} :Cas9co	2X-sgRNA-1	N/A
YP20	P _{psbD} :Cas9co	2X-sgRNA-2	N/A
YP21	P _{psbD} :Cas9co	2X-sgRNA-1	HR1:GFPco:HR2
YP22	P _{psbD} :Cas9co	2X-sgRNA-2	HR3:GFPco:HR4
YP23	N/A	2X-sgRNA-1	HR1:GFPco:HR2
YP24	N/A	2X-sgRNA-2	HR3:GFPco:HR4
YP25	P _{psaA} :Cas9co	2X-sgRNA-1	HR1:GFPco:HR2
YP26	P _{psaA} :Cas9co	2X-sgRNA-2	HR3:GFPco:HR4
YP27	P _{psbD} :Cas9co	2X-sgRNA-1	HR1:GFPco:HR2
YP28	P _{psbD} :Cas9co	2X-sgRNA-2	HR3:GFPco:HR4
YP29	N/A	2X-sgRNA-1	HR1:GFPco:HR2
YP30	N/A	2X-sgRNA-2	HR3:GFPco:HR4
YP31	P _{psaA} :Cas9co	2X-sgRNA-1	N/A
YP32	P _{psaA} :Cas9co	2X-sgRNA-2	N/A
YP33	P _{psbD} :Cas9co	2X-sgRNA-1	N/A
YP34	P _{psbD} :Cas9co	2X-sgRNA-2	N/A

*Each Expression Cassette 1 used the rbcL terminator.

**Each Expression Cassette 2 encoded either one (1X) or two (2X) guide RNAs.

TABLE 3

Components of Expression Cassette 2 Encoding One or Two Guide RNAs	
Name	Component Detail*
1X-sgRNA-1	trnW-sgRNA591-trnK
1X-sgRNA-2	trnW-sgRNA717-trnK
1X-sgRNA-3	trnW-sgRNA747-trnK
1X-sgRNA-4	trnW-sgRNA843-trnK
2X-sgRNA-1	trnW-sgRNA591-trnK-sgRNA717-trnL
2X-sgRNA-2	trnW-sgRNA747-trnK-sgRNA843-trnL

*Each Expression Cassette 2 used both the trnW promoter and trnW terminator.

[0721] Edit Plasmids were transformed into wild-type *Chlamydomonas* (CC-125) according to the methods of Barrera et al. 2014 (Methods Mol. Biol. 1132: 391-399) and Ramesh et al. 2011 (Methods Mol. Biol. 684: 313-320). Chloroplast transformants were selected using Tris-Acetate-Phosphate (TAP) media supplemented with 100 µg/ml of Spectinomycin.

[0722] To assess DNA replacement events, we transformed Edit Plasmid YP13 containing donor DNA into CC-125 (wild-type *Chlamydomonas reinhardtii*) and randomly selected spectinomycin-resistant colonies. Control construct was YP23. Pooled transformed cell lines were used to prepare chloroplast DNAs according to Barrera et al. 2014 (Methods Mol. Biol. 1132: 391-399). Pool size for YP13 was 20 independent colonies and the pool size for YP23 was 16 independent colonies. For PCR amplification of the targeted recombination region, we used primer sets which consisted of a chloroplast genomic region-specific primer and a GFP gene-specific primer. Primer Set 1 (PS1) was designed to amplify the 5' end of GFP integration region while Primer Set 2 (PS2) was designed to amplify the 3' end.

1. PS1 FWD Primer
(SEQ ID NO: 167)
GCTGTTGGTTCCACTACCAC

2. PS1 REV Primer
(SEQ ID NO: 168)
CACCTTCAAATTTTACTTCAGCAGTG

-continued

3. PS2 FWD Primer (SEQ ID NO: 169)
 CATACGGTGTACAATGTTTCAGTCG

4. PS2 REV Primer (SEQ ID NO: 170)
 GTGAGAAATAATAGCATCACGGTGAC

[0723] The primer sets were designed to avoid amplification of wild-type chloroplast genome or of the Edit Plasmid. Using the above primer sets, the expected size of each amplicon is the following: 852 bp for Primer Set 1 and 712 bp for Primer Set 2. After PCR amplification, we successfully obtained amplicons of the expected sizes from two independent pools of *Chlamydomonas* cell lines transformed with YP13. The corresponding DNA fragments were not amplified from YP23, the control construct without the Cas9 expression cassette.

[0724] We sequenced the amplified DNA fragments to confirm successful DNA replacement through Cas9 activity. We obtained the sequence encompassing the donor DNA locus in the transformed *Chlamydomonas* chloroplast DNA (see FIG. 2) (SEQ ID NO: 171). The genomic sequence corresponded to the expected sequence from insertion of the donor DNA at the two Cas9 cleavage sites. As seen in FIG. 2, the replaced DNA contained the two modified guide RNA target sites in the *psaA* gene that were encoded in the donor DNA. Additionally, the 3-nt PAM sequence is no longer present adjacent to each target sequence, corresponding to the exact sequence of the donor DNA. Also shown in FIG. 2 are sequences at the 5' and 3' ends (shown with underlining) that are wild-type chloroplast genomic sequences not present on the Edit Plasmid, which are contiguous to the HR regions (shown in bold font) present in the Edit Plasmid. In summary, DNA replacement was observed in *Chlamydomonas* chloroplasts exactly as designed by use of an Edit Plasmid that encoded a Cas9 expression cassette, a multiple guide RNA expression cassette and a donor DNA template.

[0725] Once a chloroplast DNA site is cleaved by Cas9, DNA repair should be recognizable by the presence of any of the following: nucleotide substitution, small insertion or small deletion. We analyzed spectinomycin-resistant colonies transformed with YP11 and YP31 Edit Plasmid constructs for evidence of such DNA repair. We included YP29, the construct without the Cas9 expression cassette, as a control. To enrich for edited events, we utilized the presence

of the Avail recognition sequence (GGWCC where W is either A or T) at one of the Cas9/gRNA cleavage sites (SEQ ID NO: 146, CTTACCTGTAAATGGACCACGG). First, we extracted DNA from randomly selected *Chlamydomonas* colonies (15 colonies from YP11 transformants, 10 colonies from YP31 transformants, and five colonies from YP29 transformants). We then pooled extracted DNA for Q5® high-fidelity polymerase-based PCR amplification (New England BioLabs) of the genomic region containing the target site (one pool contained DNA from five colonies). We used the following primers: PS1 FWD Primer (SEQ ID NO: 167) and PS2 REV Primer (SEQ ID NO: 170). Amplified DNA products were purified and subjected to *Ava*II digestion overnight. After gel-electrophoresis, the region corresponding to 700-900 bp of each pool, containing undigested DNA of 795 bp, was cut out of an agarose gel and the DNA was extracted. Extracted DNA was then directly cloned into pMiniT2.0 vector according to a manufacturer's protocol (New England BioLabs, Ipswich, Mass.). We randomly selected twelve *E. coli* colonies from each pool of YP11 and YP31 transformants and eight colonies from the control YP29 pool and performed PCR amplification using the same primer pair, PS1 FWD Primer and PS2 REV Primer. Aliquots of PCR reactions were digested again with *Ava*II to further select candidates for DNA repair events. One each from two pools of YP11 constructs, one from one pool of YP31 transformants, four from the other pool of YP31 transformants and three from the YP29 transformants were identified and subjected to Sanger-sequencing to deduce the nucleotide composition of each candidate clone. In addition, we included PCR amplicons of 15 randomly selected colonies from the YP29 control pool for sequencing. Analysis of sequencing results showed that two transformants of YP11 and two of YP31, each from a different pool, had a single nucleotide substitution at the target sites. We observed the following two types of substitution: G to A, resulting in GAACC; and A to G, resulting in GGGCC; relative to the wild-type sequence, GGACC. Each of these two changes were detected in transformants from each construct, YP11 and YP31; however, none of the sequenced clones from the control YP29 transformants showed any change at the target site (i.e., each control transformant retained the Avail site). In summary, we have shown that four independent nucleotide substitution events have occurred at a guide RNA target site, consistent with cleavage by Cas9 and subsequent DNA repair in the chloroplast.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 172

<210> SEQ ID NO 1

<211> LENGTH: 4163

<212> TYPE: DNA

<213> ORGANISM: Artificial sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 1

aaaaaagaat ggttctacca agactatata cagctacaag tcgtgctget ctgtcgaccg 60

acaagaagta ctccattggg ctcgatatcg gcacaaacag cgtcggetgg gccgtcatta 120

cggacgagta caaggtgccg agcaaaaaat tcaaagtctt gggcaatacc gatcgccaca 180

-continued

gcataaagaa gaacctcatt ggcgcctcc tgttcgactc cggggagacg gccgaagcca	240
cgcggctcaa aagaacagca cggcgcagat ataccgcag aaagaatcgg atctgctacc	300
tgcaggagat ctttagtaat gagatggcta aggtggatga ctctttcttc cataggtcgg	360
aggagtctt tttggtggag gaggataaaa agcacgagcg ccaccaatc tttggcaata	420
tcgtggacga ggtggcgtag catgaaaagt acccaacat atatcatctg aggaagaagc	480
ttgtagacag tactgataag gctgacttgc ggttgatcta tctcgcgctg gcgcatatga	540
tcaaatttcg gggacacttc ctcatcgagg gggacctgaa ccagacaac agcgatgtcg	600
acaaactctt tatccaactg gttcagactt acaatcagct tttcgaagag aaccgatca	660
acgcattccg agttgacgcc aaagcaatcc tgagcgctag gctgtccaaa tcccgccggc	720
tcgaaaacct catcgcacag ctccctgggg agaagaagaa cggcctgttt ggtaattctta	780
tcgccctgtc actcgggctg accccaact ttaaatctaa cttcgacctg gccgaagatg	840
ccaagcttca actgagcaaa gacacctacg atgatgatct cgacaatctg ctggcccaga	900
tcggcgacca gtacgcagac ctttttttgg cggcaaagaa cctgtcagac gccattctgc	960
tgagtgatat tctcgcagtg aacacggaga tcaccaaacg tccgctgagc gctagtatga	1020
tcaagcgcta tgatgagcac caccaagact tgactttgct gaaggccctt gtcagacagc	1080
aactgcctga gaagtacaag gaaattttct tcgatcagtc taaaaatggc tacgcggat	1140
acattgacgg cggagcaagc caggaggaat tttaaaatt tattaagccc atcttggaag	1200
aatggacgg caccgaggag ctgctggtaa agcttaacag agaagatctg ttgcgcaaac	1260
agcgcacttt cgacaatgga agcatcccc accagattca cctgggcgaa ctgcacgcta	1320
tcctcaggcg gcaagaggat ttctaccct ttttgaaaga taacagggaa aagattgaga	1380
aatcctcac atttcgata cctactatg taggccccct cgcccgggga aattccagat	1440
tcgcgtggat gactcgcaaa tcagaagaga ccatcactcc ctggaacttc gaggaagtcg	1500
tggataaggg ggcctctgcc cagtccttca tcgaaaggat gactaacttt gataaaaatc	1560
tgcctaacga aaagggtgctt cctaaacact ctctgctgta cgagtacttc acagtttata	1620
acgagctcac caaggctaaa tacgtcacag aagggatgag aaagccagca ttctgtctg	1680
gagagcagaa gaaagctatc gtggacctcc tcttcaagac gaaccggaaa gttaccgtga	1740
aacagctcaa agaagactat ttcaaaaaga ttgaatgttt cgactctgtt gaaatcagcg	1800
gagtgaggga tcgcttcaac gcatccctgg gaacgtatca cgatctcctg aaaatcatta	1860
aagacaagga cttcctggac aatgaggaga acgaggacat tcttgaggac attgtcctca	1920
cccttacggt gtttgaagat agggagatga ttgaagaacg cttgaaaact tacgctcatc	1980
tcttcgacga caaagtcatg aaacagctca agaggcgccg atatacagga tggggggcggc	2040
tgtcaagaaa actgatcaat gggatccgag acaagcagag tggaaagaca atcctggatt	2100
ttcttaagtc cgatggatgt gccaacggga acttcatgca gttgatccat gatgactctc	2160
tcacctttaa ggaggacatc cagaaagcac aagtttctgg ccagggggac agtcttcacg	2220
agcacatcgc taatcttgca ggtagcccag ctatcaaaaa gggaatactg cagaccgtta	2280
aggctcgtga tgaactcgtc aaagtaatgg gaaggcataa gcccagaaat atcgttatcg	2340
agatggcccg agagaaccaa actaccaga agggacagaa gaacagtagg gaaaggatga	2400
agaggattga agagggtata aaagaactgg ggtcccaaat ccttaaggaa caccagttg	2460

-continued

```

aaaacaccca gtttcagaat gagaagctct acctgtacta cctgcagaac ggcagggaca 2520
tgtacgtgga tcaggaactg gacatcaatc ggctctccga ctacgacgtg gatcatatcg 2580
tgccccagtc ttttctcaaa gatgattcta ttgataataa agtggttgaca agatccgata 2640
aaaatagagg gaagagtgat aacgtcccct cagaagaagt tgtcaagaaa atgaaaaatt 2700
attggcggca gctgctgaac gccaaactga tcacacaacg gaagttcgat aatctgacta 2760
aggctgaacg aggtggcctg tctgagttgg ataaagccgg cttcatcaaa aggcagcttg 2820
ttgagacacg ccagatcacc aagcacgtgg cccaaattct cgattcacgc atgaacacca 2880
agtacgatga aaatgacaaa ctgattcgag aggtgaaagt tattactctg aagtctaagc 2940
tgggtctcaga tttcagaaag gactttcagt ttataaagg gagagagatc aacaattacc 3000
accatgcgca tgatgcctac ctgaatgcag tggtaggcac tgcacttacc aaaaaatacc 3060
ccaagcttga atctgaattt gtttacggag actataaagt gtacgatgtt aggaaaaatga 3120
tcgcaaaagtc tgagcaggaa ataggcaagg ccaccgctaa gtacttcttt tacagcaata 3180
ttatgaattt tttcaagacc gagattacac tggccaatgg agagattcgg aagcgaccac 3240
ttatcgaaac aaacggagaa acaggagaaa tcgtgtggga caagggtagg gatttcgcga 3300
cagtcgggaa ggtcctgtcc atgccgcagg tgaacatcgt taaaagacc gaagtacaga 3360
cgggaggctt ctccaaggaa agtatcctcc cgaaaaggaa cagcgacaag ctgatcgac 3420
gcaaaaaaga ttgggacccc aagaaatcgc gcggattcga ttctctaca gtcgcttaca 3480
gtgtactggt tgtggccaaa gtggagaaaag ggaagtctaa aaaactcaaa agcgtcaagg 3540
aactgctggg catcacatc atggagcgat caagcttcga aaaaaacccc atcgactttc 3600
tcgaggcgaa aggatataaa gaggtcaaaa aagacctcat cattaagctt cccaagtact 3660
ctctcttga gcttgaaaac ggcgggaaac gaatgctcgc tagtgcgggc gagctgcaga 3720
aaggtaacga gctggcactg cctctaaat acgttaattt cttgtatctg gccagccact 3780
atgaaaagct caaagggtct cccgaagata atgagcagaa gcagctgttc gtggaacaac 3840
acaaacacta ccttgatgag atcatcgac aaataagcga attctccaaa agagtgatcc 3900
tcgccgacgc taacctcgat aagggtcttt ctgcttaca taagcacagg gataagccca 3960
tcagggagca ggcagaaaac attatccact tgtttactct gaccaacttg ggcgcgctg 4020
cagccttcaa gtacttcgac accaccatag acagaaagcg gtacacctct acaaaggagg 4080
tcctggacgc cacttgatt catcagtaa ttacggggct ctatgaaaca agaatcgacc 4140
tctctcagct cggtgagac tga 4163

```

```

<210> SEQ ID NO 2
<211> LENGTH: 4145
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

```

```

<400> SEQUENCE: 2

```

```

gatccatgaa aagcttcatt acaaggaaca agacagccat tgacaagaag tactccattg 60
ggctcgatat cggcacaaac agcgtcggct gggcgcgtcat tacggacgag tacaagggtgc 120
cgagcaaaaa attcaaagtt ctgggcaata ccgatcgcca cagcataaag aagaacctca 180
ttggcgcctt cctgttcgac tccggggaga cggccgaagc cagcgggctc aaaagaacag 240

```

-continued

cacggcgag atataccgc agaaagaatc ggatctgcta cctgcaggag atctttagta	300
atgagatggc taaggtggat gactctttct tccataggct ggaggagtcc tttttggtgg	360
aggaggataa aaagcacgag cgccacccaa tctttggcaa tatcgtggac gaggtggcgt	420
accatgaaaa gtaccaaac atatatacgc tgaggaagaa gcttgtagac agtactgata	480
aggctgactt gcggttgatc tatctcgcgc tggcgcatac gatcaaattt cggggacact	540
tcctcatcga gggggacctg aaccagaca acagcgatgt cgacaaactc tttatccaac	600
tggttcagac ttacaatcag cttttcgaag agaaccgat caacgcaccc ggagttgacg	660
ccaaagcaat cctgagcgc aggtgtcca aatcccggcg gctcgaaaac ctcatcgac	720
agctccctgg ggagaagaag aacggcctgt ttggtaatct tatcgccctg tctactgggc	780
tgaccccaaa ctttaaatct aacttcgacc tggcgaaga tgccaagctt caactgagca	840
aagacaccta cgatgatgat ctgcacaatc tgctggccca gatcggcgac cagtacgcag	900
accttttttt ggcggcaaa aacctgtcag acgccattct gctgagtgat attctgcgag	960
tgaacacgga gatcaccaaa gctccgctga gcgctagtat gatcaagcgc tatgatgagc	1020
accaccaaga cttgactttg ctgaaggccc ttgtcagaca gcaactgctt gagaagtaca	1080
aggaaatfff cttcgatcag tctaaaaatg gctacgccgg atacattgac ggcggagcaa	1140
gccaggagga attttacaaa tttattaagc ccatcttgga aaaaatggac ggcaccgagg	1200
agctgctggt aaagcttaac agagaagatc tgttgcgcaa acagcgact ttcgacaatg	1260
gaagcatccc ccaccagatt cacctggcg aactgcacgc tatcctcagg cggcaagagg	1320
atttttcccc ctttttgaaa gataacaggg aaaagattga gaaaatcctc acatttcgga	1380
taccctacta tgtaggcccc ctgcgccggg gaaattccag attcgcgtgg atgactcgca	1440
aatcagaaga gaccatcact ccttggaact tcgaggaagt cgtggataag ggggcctctg	1500
cccagtcctt catcgaaagg atgactaact ttgataaaaa tctgcctaac gaaaagggtc	1560
ttcctaatac ctctctgctg tacgagtact tcacagtta taacgagctc accaaggta	1620
aatacgtcac agaagggatg agaaagccag cattcctgtc tggagagcag aagaaagcta	1680
tcgtggacct cctcttcaag acgaaccgga aagttaccgt gaaacagctc aaagaagact	1740
atttcaaaaa gattgaatgt ttcgactctg ttgaaatcag cggagtggag gatcgcttca	1800
acgcatacct gggaaacgtat cacgatctcc tgaaaatcat taaagacaag gacttcctgg	1860
acaatgagga gaacgaggac attcttgagg acattgtcct cacccttacg ttgtttgaag	1920
ataggagatg gattgaagaa cgcttgaaaa cttacgctca tctcttcgac gacaaagtca	1980
tgaaacagct caagaggcgc cgatatacag gatggggcgc gctgtcaaga aaactgatca	2040
atgggatccg agacaagcag agtggaaga caatcctgga ttttcttaag tccgatggat	2100
ttgccaacgc gaacttcacg cagttgatcc atgatgactc tctcaccttt aaggaggaca	2160
tccagaaagc acaagtttct ggcacggggg acagtcttca cgagcacatc gctaattctg	2220
caggtagccc agctatcaaa aagggaatac tgcagaccgt taaggctcgtg gatgaactcg	2280
tcaaagtaat ggggaaggcat aagcccagaa atatcgatat cgagatggcc cgagagaacc	2340
aaactaccca gaaggagacg aagaacagta gggaaaggat gaagaggatt gaagagggtg	2400
taaaagaact ggggtcccaa atccttaagg aacaccagc tgaaaacacc cagcttcaga	2460
atgagaagct ctacctgtac tacctgcaga acggcaggga catgtacgtg gatcaggaac	2520

-continued

```

tggacatcaa tcggtctccc gactacgacg tggatcatat cgtgccccag tcttttctca 2580
aagatgattc tattgataat aaagtgttga caagatccga taaaaataga gggaagagtg 2640
ataacgtccc ctcagaagaa gttgtcaaga aaatgaaaaa ttattggcgg cagctgctga 2700
acgccaaact gatcacacaa cggaggttcg ataatctgac taaggctgaa cgaggtggcc 2760
tgtctgagtt ggataaagcc ggcttcacaa aaaggcagct tgttgagaca cgccagatca 2820
ccaagcacgt ggcccaaatt ctgcattcac gcatgaacac caagtacgat gaaaatgaca 2880
aactgattcg agaggtgaaa gttattactc tgaagtctaa gctggtctca gatttcagaa 2940
aggactttca gttttataag gtgagagaga tcaacaatta ccaccatgcg catgatgcct 3000
acctgaatgc agtggtaggc actgcactta tcaaaaaata tcccaagctt gaatctgaat 3060
ttgtttacgg agactataaa gtgtacgatg ttaggaaaat gatcgcaaag tctgagcagg 3120
aaataggcaa ggccaccgct aagtacttct ttacagcaa tattatgaat tttttcaaga 3180
ccgagattac actggccaat ggagagattc ggaagcgacc acttatcgaa acaaacggag 3240
aaacaggaga aatcgtgtgg gacaagggtg gggatttcgc gacagtcggg aaggctcctgt 3300
ccatgccgca ggtgaacatc gttaaaaaga ccgaagtaca gaccggaggc ttctccaagg 3360
aaagtatcct cccgaaaagg aacagcgaca agctgatcgc acgcaaaaaa gattggggacc 3420
ccaagaaata cggcggattc gattctccta cagtcgctta cagtgtactg gttgtggcca 3480
aagtggagaa agggaaagtct aaaaaactca aaagcgtcaa ggaactgctg ggcatcaca 3540
tcatggagcg atcaagcttc gaaaaaaacc ccatcgactt tctcgaggcg aaaggatata 3600
aagaggtcaa aaaagacctc atcataaagc ttcccaagta ctctctcttt gagcttgaaa 3660
acggccggaa acgaatgctc gctagtgcgg gcgagctgca gaaaggtaac gagctggcac 3720
tgccctctaa atacgttaat ttcttgtatc tggccagcca ctatgaaaag ctcaaagggt 3780
ctccgaaga taatgagcag aagcagctgt tcgtggaaca acacaaacac taccttgatg 3840
agatcatcga gcaataaagc gaattctcca aaagagtgat cctcgccgac gctaacctcg 3900
ataagggtct ttctgcttac aataagcaca gggataagcc catcaggagg caggcagaaa 3960
acattatcca cttgtttact ctgaccaact tgggcgcgcc tgcagccttc aagtacttcg 4020
acaccacat agacagaaag cgttacacct ctacaaagga ggtcctggac gccacactga 4080
ttcatcagtc aattacgggg ctctatgaaa caagaatcga cctctctcag ctcggtggag 4140
actga 4145

```

```

<210> SEQ ID NO 3
<211> LENGTH: 342
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 3

```

```

nnnnnnnnnn nnnnnnnnnn gttttagagc tagaaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cgggtggtgag ccttggtggc gcaatcggtg 120
gcgcgtatga ctcttaataa taaggttagg ggttcgagcc cccatcaggc ctccattctt 180

```


-continued

ttttttttta aaacacgatg acataaattt cctttgtatg aaccgtaccc ttaataataa	240
aaggaaaaat catgcttttag gtataagatt ttctgttata ttaaaattta gtattttattt	300
ttattatgct attattttttt tcggtctcaa atgttactta gt	342

<210> SEQ ID NO 4
<211> LENGTH: 343
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 4

nnnnnnnnnn nnnnnnnnnn gtttttagagc tagaaatagc aagttaaaat aaggctagtc	60
cgttatcaac ttgaaaaagt ggcaccgagt cgggtggtgct ccttggttagc tcagttggta	120
gagcggttcgg ctcttaacgc aaatgtcagg ggttcgagcc ccctatgagg cgccatttct	180
tttttttttt aaaacacgat gacataaatt tcctttgtat gaaccgtacc cttaataata	240
aaaggaaaaa tcattgcttta ggtataagat ttctgttat attaaaaatt agtattttatt	300
tttattatgc tattatttttt ttccggtctca aatgttactt agt	343

<210> SEQ ID NO 5
<211> LENGTH: 342
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 5

nnnnnnnnnn nnnnnnnnnn gtttttagagc tagaaatagc aagttaaaat aaggctagtc	60
cgttatcaac ttgaaaaagt ggcaccgagt cgggtggtgct ccttggttagc gcaatcggtta	120
gcgcgtatga ctcttaatca taaggtttagg ggttcgagcc cccatcaggc ctccatttctt	180
ttttttttta aaacacgatg acataaattt cctttgtatg aaccgtaccc ttaataataa	240
aaggaaaaat catgcttttag gtataagatt ttctgttata ttaaaattta gtattttattt	300
ttattatgct attattttttt tcggtctcaa atgttactta gt	342

<210> SEQ ID NO 6
<211> LENGTH: 343
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 6

nnnnnnnnnn nnnnnnnnnn gtttttagagc tagaaatagc aagttaaaat aaggctagtc	60
cgttatcaac ttgaaaaagt ggcaccgagt cgggtggtgct ccttggttagc tcagttggta	120

-continued

gagcggttcgg ctcttaaccg aaatgtcagg ggctcgagcc ccctatgagg cgccatttct 180
tttttttttt aaaacacgat gacataaatt tcctttgtat gaaccgtacc cttaataata 240
aaaggaaaaa tcattgcttta ggtataagat tttctgttat attaaaaatt agtattttatt 300
tttattatgc tattattttt ttcgggtctca aatgttactt agt 343

<210> SEQ ID NO 7
<211> LENGTH: 358
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (39)..(58)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (138)..(140)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 7

gccttgtag ctcagttggt agagcggtcg gctcttaann nnnnnnnnnn nnnnnnnngt 60
ttagagcta gaaatagcaa gttaaaataa ggctagtcgg ttatcaactt gaaaaagtgg 120
caccgagtcg gtggtgcnnn ttaagcaagg ataccgaaat gtcagggggt cgagccccc 180
atgaggatcc attctttttt tttttaaaac acgatgacat aaatttcctt tgtatgaacc 240
gtacccttaa taataaaagg aaaaatcatg ctttaggtat aagattttct gttatattaa 300
aatttagtat ttatttttat tatgctatta tttttttcgg tctcaaatgt tacttagt 358

<210> SEQ ID NO 8
<211> LENGTH: 358
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (37)..(56)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (136)..(138)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 8

gccttgtag cgcaatcggt agcgcgtagt actcttnnnn nnnnnnnnnn nnnnnngttt 60
tagagctaga aatagcaagt taaaataagg ctagtccggt atcaactga aaaagtggca 120
ccgagtcggt ggtgcnnnntt aagcaaggat aaatcataag gttagggggt cgagccccc 180
tcagggtcc attctttttt tttttaaaac acgatgacat aaatttcctt tgtatgaacc 240
gtacccttaa taataaaagg aaaaatcatg ctttaggtat aagattttct gttatattaa 300
aatttagtat ttatttttat tatgctatta tttttttcgg tctcaaatgt tacttagt 358

<210> SEQ ID NO 9
<211> LENGTH: 293
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

-continued

<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 9

nnnnnnnnnn nnnnnnnnnn gtttttagagc tagaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cgggtggtgcg gggttcgagc ccccatcagg 120
gtccattctt tttttttttt aaaacacgat gacataaatt tcctttgtat gaaccgtacc 180
cttaataata aaaggaaaaa tcatgcttta ggtataagat tttctgttat attaaaaattt 240
agtatttatt tttattatgc tattattttt ttcggtctca aatgttactt agt 293

<210> SEQ ID NO 10
<211> LENGTH: 76
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 10

gccttggtgg cgcaatcggg agcgcgtagt actcttaatc ataattcttt ttttttttaa 60
aacacgatga cataaa 76

<210> SEQ ID NO 11
<211> LENGTH: 136
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (18)..(37)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 11

gcgcaatcgg tagcgcannn nnnnnnnnnn nnnnnnngtt ttagagctag aaatagcaag 60
ttaaataaag gctagtcggt tatcaacttg aaaaagtggc accgagtcgg tgggtgcgagc 120
cccctacagg gctctt 136

<210> SEQ ID NO 12
<211> LENGTH: 116
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (18)..(37)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 12

gcgcaatcgg tagcgcannn nnnnnnnnnn nnnnnnngtt ttagagctag aaatagcaag 60
ttaaataaag gctagtcggt tatcaacttg aaaaagtggc accgagtcgg tgggtgc 116

<210> SEQ ID NO 13
<211> LENGTH: 118
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

-continued

<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 13

nnnnnnnnnn nnnnnnnnnn gtttttagagc tagaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cgggtggtgcg agccccctac agggctct 118

<210> SEQ ID NO 14
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 14

actgatagaa gtgtagt 17

<210> SEQ ID NO 15
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 15

atgattattg caattccaac 20

<210> SEQ ID NO 16
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 16

attccacgat acttactacg 20

<210> SEQ ID NO 17
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 17

tcagcaacac caaatcaaga 20

<210> SEQ ID NO 18
<211> LENGTH: 79
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 18

gttttagagc tagaaatagc aagttaaaat aaggctagtc cgttatcaac ttgaaaaagt 60
ggcaccgagt cgggtggtgc 79

<210> SEQ ID NO 19
<211> LENGTH: 269
<212> TYPE: DNA
<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 19

tccttgaaaa gataatgtat gattatgctt tcaactcatat ttatacagaa acttgatgtt 60
ttctttcgag tataatacaag gtgattacat gtacgtttga agtacaactc tagattttgt 120

-continued

agtgccctct tgggctagcg gtaaagggtgc gcattttttc acaccctaca atgttctgtt 180
caaaagattt tgggtcaaacg ctgtagaagt gaaagttggg gcgcattgtt cggcgttcga 240
aacttctccg cagtgaaga taaatgatc 269

<210> SEQ ID NO 20
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 20

tttttttggt ttttatgtct 20

<210> SEQ ID NO 21
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 21

actaatcact catcaggcgt tga 23

<210> SEQ ID NO 22
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 22

caatggcatc cccttggaag c 21

<210> SEQ ID NO 23
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 23

agttaccgta ggggaacctg 20

<210> SEQ ID NO 24
<211> LENGTH: 4107
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 24

atggacaaga agtactctat tggtttagat atcgggtacaa acagtgtcgg ttgggctgtc 60
attactgacg aatacaagggt gcctagtaaa aaattcaaag ttttaggtaa tactgatcgt 120
cacagtataa agaagaactt aattggtgct ttattattcg actctgggtga aactgctgaa 180
gctactcgtt taaaaagaac agcacgtcgt agatatactc gtagaaagaa tcgtatctgc 240
tacttacagg aaatctttag taatgaaatg gctaagggtg atgactcttt cttccataga 300
ttagaagaat cttttttggt ggaagaagat aaaaagcacg aacgtcaccc aatctttggt 360
aatatcgtgg acgaagtggc ttaccatgaa agtaccctaa ctatatatca ttttaagaaag 420
aagtttagtag acagtactga taaggctgac ttgcgtttga tctattttage ttttagctcat 480
atgatcaaat ttcgtggaca cttcttaatc gaaggtgact taaaccacaga caacagtgat 540
gtcgacaaat tatttatcca attagttcag acttacaatc agttattcga agaaaacctt 600

-continued

atcaacgcat	ctggagttga	cgctaaagca	atcttaagtg	ctagattatc	taaatctcgt	660
cgtttagaaa	acttaatcgc	acagttacct	ggtgaaaaga	agaacggttt	atttggtaat	720
ttaatcgctt	tatcattagg	tttaactcct	aactttaaat	ctaacttcga	cttagctgaa	780
gatgctaagt	tacaattaag	taaagacact	tacgatgatg	atttagacaa	tttattagct	840
cagatcggtg	accagtaacg	agacttattt	ttggctgcaa	agaacttatc	agacgctatt	900
ttattaagtg	atattttacg	agtgaacact	gaaatcacta	aagctccttt	aagtgcctagt	960
atgatcaagc	gttatgatga	acaccaccaa	gacttgactt	tggtaaaggc	tttagtcaga	1020
cagcaattac	ctgaaaagta	caaggaaatt	ttcttcgatc	agtctaaaaa	tggttacgct	1080
ggatacattg	acggtggagc	aagtcaggaa	gaattttaca	aattttattaa	gcctatcttg	1140
gaaaaaatgg	acggtactga	agaattatta	gtaaagttaa	acagagaaga	tttattgcgt	1200
aaacagcgta	ctttcgacaa	tggaagtatc	cctcaccaga	ttcacttagg	tgaattacac	1260
gctatcttaa	gacgtcaaga	agattttctac	ccttttttga	aagataacag	agaaaagatt	1320
gaaaaaatct	taacatttcg	tatacettac	tatgtaggtc	ctttagctcg	tggaattctt	1380
agattcgctt	ggatgactcg	taaatacaga	gaaactatca	ctccttgga	cttcgaagaa	1440
gtcgtggata	agggtgcttc	tgctcagctc	ttcatcgaaa	gaatgactaa	ctttgataaa	1500
aatttaccta	acgaaaaggt	gttacctaaa	cactctttat	tatacgaata	cttcacagtt	1560
tataacgaat	taactaaggt	caaatacgtc	acagaaggta	tgagaaagcc	agcattctta	1620
tctggagaac	agaagaaagc	tatcgtggac	ttattattca	agactaaccg	taaagttact	1680
gtgaaacagt	taaaaaga	ctattttcaa	aagattgaat	gtttcgactc	tggtgaaatc	1740
agtggagtgg	aagatcgctt	caacgcatct	ttaggaactt	atcacgattt	attaaaaatc	1800
attaaagaca	aggacttcct	agacaatgaa	gaaaacgaag	acattttaga	agacattgtc	1860
ttaactttaa	ctttgtttga	agatagagaa	atgattgaag	aacgtttgaa	aacttacgct	1920
cattttattcg	acgacaaagt	catgaaacag	ttaaagagac	gtcgatatac	aggatggggt	1980
cgtttatcaa	gaaaattaat	caatggtatc	cgagacaagc	agagtggaaa	gacaatctta	2040
gattttttaa	agtctgatgg	atttgctaac	cgtaacttca	tgaggttgat	ccatgatgac	2100
tctttaactt	ttaaggaaga	catccagaaa	gcacaagttt	ctggtcaggg	tgacagttta	2160
cacgaacaca	tcgctaattt	agcaggtagt	ccagctatca	aaaagggaat	attacagact	2220
gttaaggctg	tggatgaatt	agtcaaagta	atgggaagac	ataagcctga	aaatatcggt	2280
atcgaaatgg	ctcgagaaaa	ccaaactact	cagaagggac	agaagaacag	tagagaaaga	2340
atgaagagaa	ttgaagaagg	tataaaagaa	ttaggttctc	aaatcttaaa	ggaacaccca	2400
gttgaaaaca	ctcagttaca	gaatgaaaag	ttatacttat	actacttaca	gaacggtaga	2460
gacatgtaacg	tggatcagga	attagacatc	aatcgtttat	ctgactacga	cgtggatcat	2520
atcgtgcctc	agtccttttt	aaaagatgat	tctattgata	ataaagtgtt	gacaagatct	2580
gataaaaata	gaggtaagag	tgataacgtc	ccttcagaag	aagttgtcaa	gaaaatgaaa	2640
aattattggc	gtcagttatt	aaacgctaaa	ttaatcacac	aacgtaagtt	cgataattta	2700
actaaggctg	aacgaggtgg	tttatctgaa	ttggataaag	ctggtttcat	caaaagacag	2760
ttagttgaaa	cacgtcagat	cactaagcac	gtggctcaaa	ttttagattc	acgtatgaac	2820
actaagtacg	atgaaaatga	caaattaatt	cgagaagtga	aagttattac	tttaaagtct	2880

-continued

aagttagtct cagatttcag aaaggacttt cagttttata aggtgagaga aatcaacaat	2940
taccaccatg ctcatgatgc ttacttaaat gcagtggtag gtactgcatt aatcaaaaaa	3000
tatcctaagt tagaatctga atttgtttac ggagactata aagtgtacga tgttagaaaa	3060
atgatcgcaa agtctgaaca ggaaataggt aaggctactg ctaagtactt cttttacagt	3120
aatattatga attttttcaa gactgaaatt acattagcta atggagaaat tcgtaagcga	3180
ccattaatcg aaacaaacgg agaaacagga gaaatcgtgt gggacaaggg tagagatttc	3240
gtacacgtcc gtaaggctct atctatgcct cagggtgaaca tcgttaaaaa gactgaagta	3300
cagactggag gtttctctaa ggaaagtatc ttacctaaaa gaaacagtga caagttaatc	3360
gcacgtaaaa aagattggga ccctaagaaa tacggtggat tcgattctcc tacagtcgct	3420
tacagtgtat tagttgtggc taaagtggaa aaaggttaagt ctaaaaaatt aaaaagtgtc	3480
aaggaattat taggtatcac aatcatggaa cgatcaagtt tcgaaaaaaa ccctatcgac	3540
tttttagaag ctaaggata taaagaagtc aaaaagact taatcattaa gttacctaag	3600
tactctttat ttgaattaga aaacggctgt aaacgaatgt tagctagtgc tggagaatta	3660
cagaaaggtg acgaattagc attaccttct aaatacgtta atttcttgta tttagctagt	3720
cactatgaaa agttaaagg ttctcttgaa gataatgaac agaagcagtt attcgtggaa	3780
caacacaaac actacttaga tgaatcctc gaacaaataa gtgaattctc taaaagagt	3840
atcttagctg acgctaact agataagggt ttatctgctt acaataagca cagagataag	3900
cctatcagag aacaggcaga aaacattatc cacttggtta ctttaactaa cttgggtgct	3960
cctgcagctt tcaagtact cgacactact atagacagaa agcgttacac ttctacaaag	4020
gaagtcttag acgctacatt aattcatcag tcaattactg gtttatatga aacaagaatc	4080
gacttatctc agttaggtgg agactaa	4107

<210> SEQ ID NO 25

<211> LENGTH: 1037

<212> TYPE: DNA

<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 25

tttatatata ttaaaataat attaataaat aattactcct cctagcagga ttcacatctc	60
cttcggccgg actccttcgg ggtccgcccc gcggggcgcg gccggactat tttattatta	120
ttaaatagat gttcattaaa taattataaa tataatttat cttttaaata tatatatata	180
atataatatt taaatatata ttataataa ataaataaat aattaattaa taaaaacata	240
taatgtatat ttatctataa aaaatattaa ttaaattaat atattattac agttccgggg	300
gccggccacg ggagccggaa ccccgaggga gataaataaa taaataaata taaataattc	360
ttcttcttta aaattaaata aaataaaata aaaagggggg cggactcctt cgggggcccg	420
ccccctccg cggggcgagc tatttttatt ttaaatatat attatattaa taatataaat	480
ataagtcccc gccccggcgg ggaccccgaa ggagtataaa taaaaattaa taatatatta	540
tatatatatt atattaataa taataataat aataataata ataaataata actccttgct	600
tcataccttt ataaataagg taatcactaa tatattataa taataaaaaa tatatatatt	660
atatataatc taaatattat atattttaat aaatattaat atatatgata tgaatattat	720
tagtttttgg gaagcgggaa tcccgtaagg agtgagggac ccctccctaa cgggaggagg	780

-continued

accgaaggag ttttagtatt tttttttttt taataaaata tatatttata tgattaataa	840
tattatatat attatttata aaaataatat ataattttta ttatttttaa taaaaaaagg	900
tgggggtgat aatataatat aatatttttt attttaattt ataatatata ataataaatt	960
ataaataaat ttttaattaaa agtagtatta acatattata aatagacaaa agagtctaaa	1020
ggttaagatt tattaaa	1037

<210> SEQ ID NO 26
 <211> LENGTH: 619
 <212> TYPE: DNA
 <213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 26

ttaatatatta cttattatta atatttttaa ttattaaaa taataataat aataataatt	60
ataataatat tcttaaatat aataaagata tagatttata ttctattcaa tcaccttata	120
ttaaaaaat aaatattatt aaaagagggt atcatacttc tttaaataat aaattaatta	180
ttgttcaaaa agataataaa aataataata agaataattt agaaatagat aatttttata	240
aatgattagt aggatttaca gatggagatg gtagttttta tattaaatta aatgataaaa	300
aatattttaag atttttttat ggtttttagaa tacatattga tgataaagca tgttttagaaa	360
agatttagaaa tatattaaat ataccttcta attttgaaga actacttaaa acaattatat	420
tagtaaattc acaaaagaaa tggttatatt ctaatattgt aactattttt gataagtatc	480
cttgtttaac aattaaatat tatagttatt ataaatgaaa aatagctata attaataatt	540
taaatgggat atcttataat aataaagatt tattaaatat taaaaatata attaataatt	600
atgaagtata atatccata	619

<210> SEQ ID NO 27
 <211> LENGTH: 23
 <212> TYPE: DNA
 <213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 27

gaggaaatgt tgagtcgaca tcg	23
---------------------------	----

<210> SEQ ID NO 28
 <211> LENGTH: 1000
 <212> TYPE: DNA
 <213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 28

taataaatat ttataaaaag aataatttat atttataata tataatttat atattttatt	60
tttattatac aattaatata aaatataaaa tattaaatat taaatattaa atattaaata	120
ttaaatatta atttttatag gggttatata ataattatat ttataattat ataattataa	180
aaagggtatt tttataatta ttacattttt attttattta taaaaatatt aattttaata	240
agtattgaat actttatata atataaatat taattacata attaataatt aaataatatt	300
taataaatatt atttaaatat attatttata attatttatt tataaaattc tatttttatt	360
attattattt ttattttatt attaaagatt aatataataa ttattaatat attaaaaatc	420
ttttattata ttaatattta taaaaaagta tttaataaaa agatgtata aatttataaa	480
ttatataata ttattaattt atataaatat aatattataa ctttgtgatt gtcaatttag	540

-continued

```

ttaatcattg ttattaataa aggaaagata taaaaaatat tctccttctt aaaaaggggt    600
tcgggtcccc cccgtaagg gggggtcct cactccttg gtcggaactc ttcgggggtcc    660
gccccgcggg ggcggggcgg actaatTTAA cttttaatat taatattaat attatttata    720
ttttaatat ataaaaataa ataattttat ttttattaat agtatattat ataaacaata    780
aaatagtatt aattatataa aatttatata aaatatatat aaatttatta tatatatata    840
tattaatatt ttaataaagt ttttattata aattttatta tttatttatt ataataataa    900
taattttatt attattatat aagtaataaa taatagtttt atataataat aataatatat    960
atatatatat attattatat tagttatata ataaggaaaa    1000

```

```

<210> SEQ ID NO 29
<211> LENGTH: 531
<212> TYPE: DNA
<213> ORGANISM: Saccharomyces cerevisiae

```

```

<400> SEQUENCE: 29
taaataattaa tctaaatatt aatataaata ttaataattaa tagttccggg gcccgccac    60
gggagccgga accccgaaag gagaaatatt aatataaata taaataattaa tataaatata    120
aatataaata taaatatatt ttaataataat ataataaat atataaatata ttatataaat    180
ataatatata aataatataa taaaatattt taatatatat ataataaat ataattatta    240
ttataattta atataaatta ttattataat ttaataataat aaataaataa ataattataa    300
ttataattat aattataatc tcaatatata aatgataaat tattataaat acaaaggaaa    360
taattgattt ttaaaatata ttaataaaaa tatataatat aaattatact tttttgtta    420
ttatataata attatattaa tatatttaat agaattaaac tccttcgggc ggactattat    480
tcattttata tattaatgat aaatcattaa ttattattaa taaatttatt t    531

```

```

<210> SEQ ID NO 30
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Saccharomyces cerevisiae

```

```

<400> SEQUENCE: 30
tgtcccatta agacataagg tacttctaca    30

```

```

<210> SEQ ID NO 31
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Saccharomyces cerevisiae

```

```

<400> SEQUENCE: 31
tgagagcagg atctcaacaa ttgggtttatt aggagc    36

```

```

<210> SEQ ID NO 32
<211> LENGTH: 32
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

```

```

<400> SEQUENCE: 32
Met Phe Phe Ser Ala Ala Leu Arg Ala Arg Ala Ala Gly Leu Thr Ala
1           5           10          15
His Trp Gly Arg His Val Arg Asn Leu His Lys Thr Val Met Gln Asn
          20          25          30

```

-continued

<210> SEQ ID NO 33
<211> LENGTH: 4200
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 33

atgttcttct cgcggcgct cggggcccg gcggtggcc tcaccgcca ctgggaaga	60
catgtaagga attgcataa gacagttatg caaatgaca agaagtact cattgggctc	120
gatatcgga caaacagcgt cggtgggccc gtcattacg acgagtaca ggtgccgagc	180
aaaaaattca aagttctggg caataccgat cgccacagca taaagaaga cctcattggc	240
gccctcctgt tcgactccg ggagacggcc gaagccacgc ggctcaaaag aacagcacgg	300
cgcagatata cccgcagaaa gaatcggatc tgctacctgc aggagatctt tagtaatgag	360
atggctaagg tggatgactc tttcttccat aggctggagg agtcctttt ggtggaggag	420
gataaaaagc acgagcgcca ccaatcttt ggcaatctc tggacgaggt ggcgtaccat	480
gaaaagtacc caaccatata tcacttgagg aagaagcttg tagacagtac tgataaggct	540
gacttgcggt tgatctatct cgcgctggcg catatgatca aatttcgggg acacttctc	600
atcgaggggg acctgaaccc agacaacagc gatgtcgaca aactctttat ccaactggtt	660
cagacttaca atcagctttt cgaagagAAC ccgatcaacg catccggagt tgacgcaaaa	720
gcaatcctga gcgctaggct gtccaaatcc cggcggtcgc aaaacctcat cgcacagctc	780
cctggggaga agaagaacgg cctgtttggt aatcttatcg cctgtcact cgggctgacc	840
cccaacttta aatctaactt cgacctggcc gaagatgcca agcttcaact gagcaaaagc	900
acctacgatg atgatctoga caatctgctg gccagatcg gcgaccagta cgcagacctt	960
tttttggcgg caaagaacct gtcagacgcc attctgctga gtgatattct gcgagtgaac	1020
acggagatca ccaaagctcc gctgagcgtc agtatgatca agcgtatga tgagcaccac	1080
caagacttga ctttgcgtga ggccttgtc agacagcaac tgcttgagaa gtacaaggaa	1140
attttcttcg atcagtctaa aaatggctac gccggataca ttgacggcgg agcaagccag	1200
gaggaaattht acaaattht taagccatc ttggaaaaaa tggacggcac caggagctg	1260
ctggtaaagc ttaacagaga agatctgttg gcgaacagc gcactttcga caatggaagc	1320
atccccacc agattcacct ggcggaactg cacgctatcc tcaggcgga agaggatttc	1380
taccctttt tgaaagataa cagggaag attgagaaaa tctcactt tcggatcccc	1440
tactatgtag gccccctgc cgggggaaat tccagattcg cgtggatgac tcgcaaatca	1500
gaagagacca tcaactccctg gaacttcgag gaagtcgttg ataagggggc ctctgccag	1560
tccttcactg aaaggatgac taacttgat aaaaatctgc ctaacgaaaa ggtgcttct	1620
aaacactctc tgctgtacga gtacttcaca gtttataacg agctaccaa ggtcaaatc	1680
gtcacagaag ggatgagaaa gccagcttc ctgtctggag agcagaaga agctatcgtg	1740
gacctctctc tcaagacgaa ccggaaagtt accgtgaac agctcaaga agactatttc	1800
aaaaagattg aatgtttcga ctctgttgaa atcagcgag tggaggatcg cttcaacgca	1860
tccctgggaa cgtatcacga tctcctgaaa atcattaaag acaaggactt cctggacaat	1920
gaggagaacg aggacattct tgaggacatt gtcctcacc ttacgttggt tgaagatagg	1980

-continued

gagatgattg aagaacgctt gaaaacttac gctcatctct tcgacgacaa agtcatgaaa	2040
cagctcaaga ggcgccgata tacaggatgg gggcggctgt caagaaaact gatcaatggg	2100
atccgagaca agcagagtgg aaagacaatc ctggattttc ttaagtccga tggatttgcc	2160
aaccggaact tcatgcagtt gatccatgat gactctctca cctttaagga ggacatccag	2220
aaagcacaag tttctggcca gggggacagt cttcacgagc acatcgctaa tcttgacagt	2280
agcccagcta tcaaaaaggg aatactgcag accgttaagg tcgtggatga actcgctaaa	2340
gtaatgggaa ggcataagcc cgagaatata gttatcgaga tggcccgaga gaaccaaact	2400
accagaagg gacagaagaa cagtagggaa aggatgaaga ggattgaaga gggataaaaa	2460
gaactggggg cccaaatcct taaggaaacac ccagttgaaa acaccagct tcagaatgag	2520
aagctctacc tgtactacct gcagaacggc agggacatgt acgtggatca ggaactggac	2580
atcaatcggc tctccgacta cgacgtggat catatcgtgc ccagctcttt tctcaaagat	2640
gattctattg ataataaagt gttgacaaga tccgataaaa atagaggga gagtgataac	2700
gtccctcag aagaagttgt caagaaaatg aaaaattatt ggcggcagct gctgaacgcc	2760
aaactgatca cacaacggaa gttcgataat ctgactaagg ctgaacgagg tggcctgtct	2820
gagttggata aagccgctt catcaaaagg cagcttgttg agacacgcca gatcaccaag	2880
cacgtggccc aaattctcga ttcacgcctg aacaccaagt acgatgaaaa tgacaaactg	2940
attcgagagg tgaaagtat tactctgaag tctaagctgg tctcagattt cagaaaggac	3000
tttcagtttt ataaggtgag agagatcaac aattaccacc atgcgcctga tgcctacctg	3060
aatgcagtgg taggcactgc acttatcaaa aaatatccca agcttgaatc tgaatttgtt	3120
tacggagact ataaagtga cgatgttagg aaaatgatcg caaagtctga gcaggaaata	3180
ggcaaggcca ccgctaagta cttcttttac agcaatatta tgaatttttt caagaccgag	3240
attacactgg ccaatggaga gattcgggaag cgaccactta tcgaaacaaa cggagaaaca	3300
ggagaaatcg tgtgggacaa gggtagggat ttcgcgacag tccggaaggc cctgtccatg	3360
ccgcaggtga acatcgctaa aaagaccgaa gtacagacgg gaggcctctc caaggaaagt	3420
atcctccgga aaaggaacag cgacaagctg atcgcacgca aaaaagattg ggaccccaag	3480
aaatacggcg gattcgattc tctacagtc gcttacagtg tactggttgt ggccaaagtg	3540
gagaaaggga agtctaaaaa actcaaaagc gtcaagggaac tgctgggcat cacaatcatg	3600
gagcgatcaa gcttcgaaaa aaaccccatc gactttctcg aggcgaaagg atataaagag	3660
gtcaaaaaag acctcatcat taagcttccc aagtactctc tctttgagct tgaaaaaggc	3720
cggaaacgaa tgctcgctag tgcgggagag ctgcagaaag gtaacgagct ggcaatgccc	3780
tctaaatacg ttaatttctt gtatctggcc agccactatg aaaagctcaa aggggtctccc	3840
gaagataatg agcagaagca gctgttcgtg gaacaacaca aacactacct tgatgagatc	3900
atcgagcaaa taagcgaatt ctccaaaaga gtgatcctcg ccgacgctaa cctcgataag	3960
gtgctttctg cttacaataa gcacagggat aagcccatca gggagcaggc agaaaacatt	4020
atccacttgt ttactctgac caacttgggc gcgcctgcag ccttcaagta cttcgacacc	4080
accatagaca gaaagcggtg cacctctaca aaggaggtcc tggacgccac actgattcat	4140
cagtcaatta cggggctcta tgaacaaga atcgacctct ctcagctcgg tggagactga	4200

-continued

<210> SEQ ID NO 34
 <211> LENGTH: 25
 <212> TYPE: PRT
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 34

Met Ala Leu Leu Thr Ala Ala Ala Arg Leu Leu Gly Thr Lys Asn Ala
 1 5 10 15

Ser Cys Leu Val Leu Ala Ala Arg His
 20 25

<210> SEQ ID NO 35
 <211> LENGTH: 4227
 <212> TYPE: DNA
 <213> ORGANISM: Artificial sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 35

atggctttac ttactgcggc cgccccgctc ttgggaacca agaatgcac tttgtctgtt	60
cttgagagccc ggcatatggc ttacttact gcggccgccc ggctcttggg aaccaagaat	120
gcagacaaga agtactccat tgggctcgat atcggcacaa acagcgctcg ctgggcccgc	180
attacggacg agtacaaggt gccgagcaaa aaattcaaa ttctgggcaa taccgatcgc	240
cacagcataa agaagaacct cattggcgcc ctctgttcg actccgggga gacggccgaa	300
gccacgcggc tcaaaagaac agcacggcgc agatataccc gcagaaagaa tcggatctgc	360
tacctgcagg agatctttag taatgagatg gctaagggtg atgactcttt cttccatagg	420
ctggaggagt ccttttttgt ggaggaggat aaaaagcacg agcggccccc aatctttggc	480
aatatcgtgg acgaggtggc gtaccatgaa aagtacccaa ccatatatca tctgaggaag	540
aagctttag acagtactga taaggctgac ttgcggttga tctatctcgc gctggcgcat	600
atgatcaaat ttccggggaca ctccctcacc gagggggacc tgaaccaga caacagcgat	660
gtcgacaaac tctttatcca actggttcag acttacaatc agcttttcga agagaacccg	720
atcaacgcat ccggagtga cgccaaagca atcctgagcg ctaggctgtc caaatcccgg	780
cggtctgaaa acctcatcgc acagctccct ggggagaaga agaacggcct gtttggtaat	840
cttatcgccc tgtcactcgg gctgaccccc aactttaaat ctaacttcga cctggccgaa	900
gatgccaagc ttcaactgag caaagacacc tacgatgatg atctcgacaa tctgctggcc	960
cagatcggcg accagtaacg agacctttt ttggcggaaga agaacctgtc agacgccatt	1020
ctgctgagtg atattctgcg agtgaacacg gagatcacca aagctccgct gagcgctagt	1080
atgatcaagc gctatgatga gcaccaccaa gacttgactt tgctgaaggc ccttgtcaga	1140
cagcaactgc ctgagaagta caaggaaatt ttcttcgacg agtctaaaaa tggtctacgc	1200
ggatacattg acggcggagc aagccaggag gaattttaca aatttattaa gcccattctg	1260
gaaaaaatgg acggcaccga ggagctgctg gtaaagctta acagagaaga tctgttgccg	1320
aaacagcgca ctttcgacaa tggaagcacc cccaccaga ttcacctggg cgaactgcac	1380
gctatcctca ggcggcaaga ggattttctac ccttttttga aagataacag ggaaaagatt	1440
gagaaaaatc tcacatttcg gataccctac tatgtaggcc ccctcgcccg gggaaattcc	1500
agattcgcgt ggatgactcg caaatcagaa gagaccatca ctccctggaa cttcgaggaa	1560
gtcgtggata agggggcctc tgcccagtc ttcacgaaa ggatgactaa ctttgataaa	1620

-continued

aatctgccta	acgaaaaggt	gcttctctaaa	cactctctgc	tgtacgagta	cttcacagtt	1680
tataacgagc	tcaccaaggt	caaatacgtc	acagaaggga	tgagaaagcc	agcatctctg	1740
tctggagagc	agaagaaagc	tatcgtggac	ctcctcttca	agacgaaccg	gaaagttacc	1800
gtgaaacagc	tcaaagaaga	ctatttcaaa	aagattgaat	gtttcgactc	tgttgaaatc	1860
agcggagtgg	aggatcgett	caacgcctcc	ctgggaacgt	atcacgatct	cctgaaaatc	1920
attaagaca	aggacttctc	ggacaatgag	gagaacgagg	acattcttga	ggacattgtc	1980
ctcaccctta	cgttggttga	agatagggag	atgattgaag	aacgcttgaa	aacttacgct	2040
catctctctg	acgacaaagt	catgaaacag	ctcaagaggc	gccgatatac	aggatggggg	2100
cggctgtcaa	gaaaactgat	caatgggac	cagacacaagc	agagtggaaa	gacaatcctg	2160
gattttctta	agtcgcatgg	atttgccaac	cggaaactca	tgcagttgat	ccatgatgac	2220
tctctcacct	ttaaggagga	catccagaaa	gcacaagttt	ctggccaggg	ggacagtctt	2280
cacgagcaca	tcgctaactc	tgcaggtagc	ccagctatca	aaaagggaat	actgcagacc	2340
gttaaggctg	tggatgaact	cgtcaaagta	atgggaaggc	ataagccga	gaatatcggt	2400
atcgagatgg	cccagagaga	ccaaactacc	cagaagggac	agaagaacag	tagggaaagg	2460
atgaagagga	ttgaagaggg	tataaaagaa	ctgggggtccc	aaatccttaa	ggaacaccca	2520
gttgaaaaca	cccagcttca	gaatgagaag	ctctacctgt	actacctgca	gaacggcagg	2580
gacatgtacg	tggatcagga	actggacatc	aatcggctct	ccgactacga	cgtggatcat	2640
atcgtgcccc	agtcctttct	caaagatgat	tctattgata	ataaagtgtt	gacaagatcc	2700
gataaaaata	gaggggaagag	tgataacgtc	ccctcagaag	aagttgtcaa	gaaaatgaaa	2760
aattattggc	ggcagctgct	gaacgccaaa	ctgatcacac	aacggaagtt	cgataatctg	2820
actaaggctg	aacgaggtgg	cctgtctgag	ttggataaag	ccggcttcat	caaaaggcag	2880
cttggtgaga	cacgccagat	caccaagcac	gtggcccaaa	ttctcgattc	acgcatgaac	2940
accaagtacg	atgaaaatga	caaactgatt	cgagaggtga	aagttattac	tctgaagtct	3000
aagctggctc	cagatttcag	aaaggacttt	cagttttata	aggtgagaga	gatcaacaat	3060
taccaccatg	cgcgatgatc	ctacctgaat	gcagtggtag	gcactgcact	tatcaaaaaa	3120
tatcccaagc	ttgaatctga	atttgtttac	ggagactata	aagtgtacga	tgtaggaaa	3180
atgatcgcaa	agtcctgagc	ggaaataggc	aaggccaccg	ctaagtactt	cttttacagc	3240
aatattatga	attttttcaa	gaccgagatt	acactggcca	atggagagat	tcggaagcga	3300
ccacttatcg	aaacaaacgg	agaaacagga	gaaatcgtgt	gggacaaggg	tagggatttc	3360
gcgacagtcc	ggaaggctct	gtccatgccg	caggtgaaca	tcgttaaaaa	gaccgaagta	3420
cagaccggag	gcttctccaa	ggaaagtatc	ctcccgaaaa	ggaacagcga	caagctgatc	3480
gcacgcaaaa	aagattggga	ccccaaagaa	tacggcggat	tcgattctcc	tacagtcgct	3540
tacagtgtac	tggttggtggc	caaagtggag	aaagggaagt	ctaaaaaact	caaaagcgtc	3600
aaggaaactgc	tgggcatcac	aatcatggag	cgatcaagct	tcgaaaaaaa	ccccatcgac	3660
tttctcgagg	cgaaggata	taaagaggtc	aaaaagacc	tcattcattaa	gcttcccaag	3720
tactctctct	ttgagcttga	aaacggccgg	aaacgaatgc	tcgctagtgc	ggcgagctg	3780
cagaaaggtg	acgagctggc	actgccctct	aaatacgtta	atttcttgta	tctggccagc	3840
cactatgaaa	agctcaaaag	gtctcccga	gataatgagc	agaagcagct	gttcgtggaa	3900

-continued

caacacaaac actaccttga tgagatcatc gagcaaataa gcgaattctc caaaagagtg 3960
atcctcgccg acgctaacct cgataagggtg ctttctgctt acaataagca cagggtataag 4020
cccatcaggg agcaggcaga aaacattatc cacttggtta ctctgaccaa cttgggagcg 4080
cctgcagcct tcaagtaact cgacaccacc atagacagaa agcggtagac ctctacaaaag 4140
gagggtcctgg acgccacact gattcatcag tcaattacgg ggctctatga aacaagaatc 4200
gacctctctc agctcgggtg agactga 4227

<210> SEQ ID NO 36
<211> LENGTH: 120
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 36

gctacggcc ataccacct gaacgcgcc gatctcgtct gatctcgga gctaagcagg 60
gtcgggctg gttagtactt ggatgggaga ccacctgga ataccgggtg ctgtaggctt 120

<210> SEQ ID NO 37
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 37

gtctggtgag tagtgcattg ct 22

<210> SEQ ID NO 38
<211> LENGTH: 460
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 38

agcccccgcg ccccggtctg gcggtgtcgg ctgcaatccg gcgggcacgg ccgggcccgg 60
ctgggctctt ggggcagcca ggcgcctcct tcagcgcta cgcccatacc accctgaacg 120
cgcccgatct cgtctgatct cggaagctaa gcagggtcgg gcctggttag tacttgatg 180
ggagaccacc tgggaatacc ggggtgtgta ggctttttct ttggcttttt gctgtttctt 240
tcctttttct ccagacggag tctcgccctc tcgcccaggc tggagtgcgg tggcgccatc 300
tcggctcact gcaagctcgg cctcccggtt ccacgccatt ccccgccctc agcctcccg 360
gtagctgggc ctacaggcgc ccgccaccac gcccgccac tttgttctat ttttcctaga 420
gacgggcttt caccctgtta gccgggatgg tctggagctc 460

<210> SEQ ID NO 39
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 39

gatgtcctga tccaacatcg 20

<210> SEQ ID NO 40
<211> LENGTH: 432
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 40

-continued

tttcggttgg ggtgacctcg gagaataaaa aatcctccga atgattataa cctagactta	60
caagtcaaag taaaatcaac atatcttatt gaccagata tattttgac aacggaccaa	120
gttaccctag ggataacagc gcaatcctat ttaagagttc atatcgacaa ttagggttta	180
cgacctcgac gttggatcag gacatcccaa tgggtgtaga gctattaatg gttcgtttgt	240
tcaacgatta aagtcctacg tgatctgagt tcagaccgga gcaatccagg tcggtttcta	300
tctatttacg atttctccca gtacgaaagg acaagagaaa tagagccacc ttacaaataa	360
gcgctctcaa cttaatttat gaataaaatc taaataaaat atatacgtac accctctaac	420
ctagagaagg tt	432

<210> SEQ ID NO 41

<211> LENGTH: 4296

<212> TYPE: DNA

<213> ORGANISM: Artificial sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 41

atggataaga agtactctat cggacttgac atcggaacca actctgttgg atgggctgtt	60
atcaccgatg agtacaaggt tccatctaag aagttcaagg ttcttggaac caccgataga	120
cactctatca agaagaacct tatcggtgct cttcttttctg attctggaga gaccgctgag	180
gctaccgatg tgaagagaac cgctagaaga agatacacca gaagaaagaa cagaatctgc	240
taccttcagg aaatctcttc taacgagatg gctaagggtg atgattcttt cttccacaga	300
cttgaggagt ctttccttgt tgaggaggat aagaagcacg agagacaccc aatcttcgga	360
aacatcggtg atgaggttgc ttaccacgag aagtacccaa ccatctacca ccttagaaag	420
aagttggttg attctaccga taaggctgat cttagactta tctacctgc tcttgctcac	480
atgatcaagt tcagaggaca cttccttacc gagggagacc ttaaccaga taactctgat	540
gttgataagt tgttcacca gcttggtcag acctacaacc agcttttcga ggagaacca	600
atcaacgctt ctggagttga tgctaaggct atcctttctg ctagactttc taagtctcgt	660
agacttgaga accttatcgc tcagcttcca ggagagaaga agaacggact ttctggaaac	720
cttatcgctc tttctcttgg acttaccca aacttcaagt ctaacttcga tcttgctgag	780
gatgctaagt tgcagcttcc taaggatacc tacgatgatg atcttgataa ccttcttgct	840
cagatcggag atcagtaacg tgatcttttc cttgctgcta agaaccttcc tgatgctatc	900
cttctttctg acatccttag agttaacacc gagatcacca aggctccact ttctgcttct	960
atgatcaaga gatacgatga gcaccaccag gatcttacct ttttgaaggc tcttggttaga	1020
cagcagcttc cagagaagta caaggaaatc ttcttcgac agtctaagaa cggatacgct	1080
ggatacatcg atggaggagc ttctcaggag gagttctaca agttcatcaa gccaatcctt	1140
gagaagatgg atggaaccga ggagcttctt gttaagttga acagagagga tcttcttaga	1200
aagcagagaa ccttcgataa cggatctatc ccacaccaga tccaccttgg agagcttcac	1260
gctatccttc gtacagagga ggattcttac ccattcttga aggataacag agagaagatc	1320
gagaagatcc ttaccttcag aatcccatc tacgttgac cacttgctag aggaaactct	1380
cgtttcgctt ggatgaccag aaagtctgag gagaccatca ccccttggaa cttcgaggag	1440
gtaagtttct gcttctacct ttgatatata tataataatt atcattaatt agtagtaata	1500

-continued

taatatattca aatatTTTT tcaaaataaa agaattgtagt atatagcaat tgcttttctg	1560
tagtttataa gtgtgtatat ttaatttat aacttttcta atatatgacc aaaatttgtt	1620
gatgtgcagg ttgttgataa gggagcttct gctcagtcctt tcatcgagag aatgaccaac	1680
ttcgataaga accttccaaa cgagaagggtt ctccaaaagc actctcttct ttacgagtac	1740
ttcacggtt acaacgagct taccaagggtt aagtacgtta ccgagggaaat gagaaagcca	1800
gctttccttt ctggagagca gaagaagggtt atcgttgatc ttcttttcaa gaccaacaga	1860
aagggtaccg ttaagcagtt gaaggaggat tacttcaaga agatcgagtg cttegattct	1920
gttgaaatct ctggagttga ggatagattc aacgcttctc ttggaacctc ccacgatctt	1980
ttgaagatca tcaaggataa ggatttctct gataacgagg agaacgagga catccttgag	2040
gacatcgctc ttacccttac ctttttcgag gatagagaga tgatcgagga gagactcaag	2100
acctacgctc accttttoga tgataagggtt atgaagcagt tgaagagaag aagatacacc	2160
ggatggggta gactttctcg taagttgatc aacggaatca gagataagca gtctggaaaag	2220
accatccttg atttcttgaa gtctgatgga ttcgctaaca gaaacttcat gcagcttctc	2280
cacgatgatt ctcttacctt caaggaggac atccagaagg ctacggttctc tggacaggga	2340
gattctcttc acgagccat cgctaacctt gctggatctc cagctatcaa gaagggaatc	2400
cttcagaccg ttaaggttgt tgatgagctt gttaagggtta tgggtagaca caagccagag	2460
aacatcgcta tcgagatggc tagagagaac cagaccacc cagaagggaca gaagaactct	2520
cgtgagagaa tgaagagaat cgaggaggga atcaaggagc ttgatctca aatcttgaag	2580
gagcaccag ttgagaacac ccagcttcag aacgagaagt tgtacctta ctaccttcag	2640
aacggaagag atatgtacgt tgatcaggag cttgacatca acagacttctc tgattacgat	2700
gttgatcaca tcgttccaca gtctttcttg aaggatgatt ctatcgataa caagggtctt	2760
acctgttctg ataagaacag aggaaagtct gataacgttc catctgagga ggttgtaaag	2820
aagatgaaga actactggag acagcttctt aacgctaagt tgatcaccac gagaaagtctc	2880
gataacctta ccaaggctga gagaggagga ctttctgagc ttgataaggc tggattccatc	2940
aagagacagc ttgttgagac cagacagatc accaagcacg ttgctcagat ccttgattct	3000
cgtatgaaca ccaagtacga tgagaacgat aagttgatca gagagggttaa ggttatcacc	3060
ttgaagtcta agttgggttc tgatttcaga aaggatttcc agttctacaa ggtagagag	3120
atcaacaact accaccagc tcacgatgct taccttaacg ctgttggttg aaccgctctt	3180
atcaagaagt acccaaagt ggagctgag ttcgtttacg gagattacaa ggtttacgat	3240
gttagaaaga tgatcgctaa gtctgagcag gagatcggaagggtaccgc taagtacttc	3300
ttctactcta acatcatgaa cttcttcaag accgagatca cccttgctaa cggagagatc	3360
agaaagagac cacttatcga gaccaacgga gagaccgag agatcgtttg ggataaggga	3420
agagatttgc ctaccgttag aaagggtctt tctatgccac aggttaacat cgtaagaaa	3480
accgaggttc agaccggagg attctctaag gagtctatcc ttccaaagag aaactctgat	3540
aagttgatcg ctgaaagaa ggattgggac ccaaagaagt acggaggatt cgattctcca	3600
accgttgctt actctgttct tgtgttgct aaggttgaga agggaaagtc taagaagttg	3660
aagctgtta aggagcttct tggaatcacc atcatggagc gttcttctt cgagaagaac	3720
ccaatcgatt tccttgaggc taagggatac aaggaggtta agaaggatct tatcatcaag	3780

-continued

```

ttgccaaagt actctctttt cgagcttgag aacggaagaa agagaatgct tgcttctgct 3840
ggagagcttc agaagggaaa cgagcttgct cttccatcta agtacgttaa cttectttac 3900
cttgcttctc actacgagaa gttgaaggga tctccagagg ataacgagca gaagcagctt 3960
ttcgttgagc agcacaagca ctaccttgat gagatcatcg agcaaatctc tgagttctct 4020
aagagagtta tccttgctga tgctaacctt gataagggtc tttctgctta caacaagcac 4080
agagataagc caatcagaga gcaggctgag aacatcatcc accttttcac ccttaccaac 4140
cttggtgctc cagctgcttt caagtacttc gataaccacca tcgatagaaa aagatacacc 4200
tctaccaagg aggttcttga tgctaccctt atccaccagt ctatcaccgg actttacgag 4260
accagaatcg atctttctca gcttgaggga gattga 4296

```

```

<210> SEQ ID NO 42
<211> LENGTH: 1368
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

```

```

<400> SEQUENCE: 42

```

```

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1      5      10     15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20     25     30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35     40     45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50     55     60
Lys Arg Thr Ala Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65     70     75     80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85     90     95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100    105    110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115    120    125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130    135    140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145    150    155    160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165    170    175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180    185    190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195    200    205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210    215    220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225    230    235    240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245    250    255

```

Asp	Leu	Ala	Glu	Asp	Ala	Lys	Leu	Gln	Leu	Ser	Lys	Asp	Thr	Tyr	Asp
			260					265					270		
Asp	Asp	Leu	Asp	Asn	Leu	Leu	Ala	Gln	Ile	Gly	Asp	Gln	Tyr	Ala	Asp
		275					280					285			
Leu	Phe	Leu	Ala	Ala	Lys	Asn	Leu	Ser	Asp	Ala	Ile	Leu	Leu	Ser	Asp
	290					295					300				
Ile	Leu	Arg	Val	Asn	Thr	Glu	Ile	Thr	Lys	Ala	Pro	Leu	Ser	Ala	Ser
305					310					315					320
Met	Ile	Lys	Arg	Tyr	Asp	Glu	His	His	Gln	Asp	Leu	Thr	Leu	Leu	Lys
				325					330					335	
Ala	Leu	Val	Arg	Gln	Gln	Leu	Pro	Glu	Lys	Tyr	Lys	Glu	Ile	Phe	Phe
			340					345					350		
Asp	Gln	Ser	Lys	Asn	Gly	Tyr	Ala	Gly	Tyr	Ile	Asp	Gly	Gly	Ala	Ser
		355					360					365			
Gln	Glu	Glu	Phe	Tyr	Lys	Phe	Ile	Lys	Pro	Ile	Leu	Glu	Lys	Met	Asp
	370					375					380				
Gly	Thr	Glu	Glu	Leu	Leu	Val	Lys	Leu	Asn	Arg	Glu	Asp	Leu	Leu	Arg
385					390					395					400
Lys	Gln	Arg	Thr	Phe	Asp	Asn	Gly	Ser	Ile	Pro	His	Gln	Ile	His	Leu
				405					410					415	
Gly	Glu	Leu	His	Ala	Ile	Leu	Arg	Arg	Gln	Glu	Asp	Phe	Tyr	Pro	Phe
			420					425					430		
Leu	Lys	Asp	Asn	Arg	Glu	Lys	Ile	Glu	Lys	Ile	Leu	Thr	Phe	Arg	Ile
		435					440					445			
Pro	Tyr	Tyr	Val	Gly	Pro	Leu	Ala	Arg	Gly	Asn	Ser	Arg	Phe	Ala	Trp
	450					455					460				
Met	Thr	Arg	Lys	Ser	Glu	Glu	Thr	Ile	Thr	Pro	Trp	Asn	Phe	Glu	Glu
465					470					475					480
Val	Val	Asp	Lys	Gly	Ala	Ser	Ala	Gln	Ser	Phe	Ile	Glu	Arg	Met	Thr
				485					490					495	
Asn	Phe	Asp	Lys	Asn	Leu	Pro	Asn	Glu	Lys	Val	Leu	Pro	Lys	His	Ser
			500					505					510		
Leu	Leu	Tyr	Glu	Tyr	Phe	Thr	Val	Tyr	Asn	Glu	Leu	Thr	Lys	Val	Lys
			515				520					525			
Tyr	Val	Thr	Glu	Gly	Met	Arg	Lys	Pro	Ala	Phe	Leu	Ser	Gly	Glu	Gln
	530					535					540				
Lys	Lys	Ala	Ile	Val	Asp	Leu	Leu	Phe	Lys	Thr	Asn	Arg	Lys	Val	Thr
545				550						555					560
Val	Lys	Gln	Leu	Lys	Glu	Asp	Tyr	Phe	Lys	Lys	Ile	Glu	Cys	Phe	Asp
				565					570					575	
Ser	Val	Glu	Ile	Ser	Gly	Val	Glu	Asp	Arg	Phe	Asn	Ala	Ser	Leu	Gly
			580					585					590		
Thr	Tyr	His	Asp	Leu	Leu	Lys	Ile	Ile	Lys	Asp	Lys	Asp	Phe	Leu	Asp
		595				600						605			
Asn															

-continued

Thr	Gly	Trp	Gly	Arg	Leu	Ser	Arg	Lys	Leu	Ile	Asn	Gly	Ile	Arg	Asp
			660					665					670		
Lys	Gln	Ser	Gly	Lys	Thr	Ile	Leu	Asp	Phe	Leu	Lys	Ser	Asp	Gly	Phe
		675					680					685			
Ala	Asn	Arg	Asn	Phe	Met	Gln	Leu	Ile	His	Asp	Asp	Ser	Leu	Thr	Phe
	690					695					700				
Lys	Glu	Asp	Ile	Gln	Lys	Ala	Gln	Val	Ser	Gly	Gln	Gly	Asp	Ser	Leu
705					710					715					720
His	Glu	His	Ile	Ala	Asn	Leu	Ala	Gly	Ser	Pro	Ala	Ile	Lys	Lys	Gly
			725						730					735	
Ile	Leu	Gln	Thr	Val	Lys	Val	Val	Asp	Glu	Leu	Val	Lys	Val	Met	Gly
			740					745					750		
Arg	His	Lys	Pro	Glu	Asn	Ile	Val	Ile	Glu	Met	Ala	Arg	Glu	Asn	Gln
		755					760					765			
Thr	Thr	Gln	Lys	Gly	Gln	Lys	Asn	Ser	Arg	Glu	Arg	Met	Lys	Arg	Ile
	770					775						780			
Glu	Glu	Gly	Ile	Lys	Glu	Leu	Gly	Ser	Gln	Ile	Leu	Lys	Glu	His	Pro
785					790					795					800
Val	Glu	Asn	Thr	Gln	Leu	Gln	Asn	Glu	Lys	Leu	Tyr	Leu	Tyr	Tyr	Leu
				805					810					815	
Gln	Asn	Gly	Arg	Asp	Met	Tyr	Val	Asp	Gln	Glu	Leu	Asp	Ile	Asn	Arg
			820					825					830		
Leu	Ser	Asp	Tyr	Asp	Val	Asp	His	Ile	Val	Pro	Gln	Ser	Phe	Leu	Lys
		835					840					845			
Asp	Asp	Ser	Ile	Asp	Asn	Lys	Val	Leu	Thr	Arg	Ser	Asp	Lys	Asn	Arg
	850					855					860				
Gly	Lys	Ser	Asp	Asn	Val	Pro	Ser	Glu	Glu	Val	Val	Lys	Lys	Met	Lys
865					870					875					880
Asn	Tyr	Trp	Arg	Gln	Leu	Leu	Asn	Ala	Lys	Leu	Ile	Thr	Gln	Arg	Lys
				885					890					895	
Phe	Asp	Asn	Leu	Thr	Lys	Ala	Glu	Arg	Gly	Gly	Leu	Ser	Glu	Leu	Asp
			900					905					910		
Lys	Ala	Gly	Phe	Ile	Lys	Arg	Gln	Leu	Val	Glu	Thr	Arg	Gln	Ile	Thr
		915					920					925			
Lys	His	Val	Ala	Gln	Ile	Leu	Asp	Ser	Arg	Met	Asn	Thr	Lys	Tyr	Asp
	930					935					940				
Glu	Asn	Asp	Lys	Leu	Ile	Arg	Glu	Val	Lys	Val	Ile	Thr	Leu	Lys	Ser
945					950					955					960
Lys	Leu	Val	Ser	Asp	Phe	Arg	Lys	Asp	Phe	Gln	Phe	Tyr	Lys	Val	Arg
				965					970					975	
Glu	Ile	Asn	Asn	Tyr	His	His	Ala	His	Asp	Ala	Tyr	Leu	Asn	Ala	Val
			980					985					990		
Val	Gly	Thr	Ala	Leu	Ile	Lys	Lys	Tyr	Pro	Lys	Leu	Glu	Ser	Glu	Phe
			995				1000					1005			
Val	Tyr	Gly	Asp	Tyr	Lys	Val	Tyr	Asp	Val	Arg	Lys	Met	Ile	Ala	
	1010					1015						1020			
Lys	Ser	Glu	Gln	Glu	Ile	Gly	Lys	Ala	Thr	Ala	Lys	Tyr	Phe	Phe	
	1025					1030						1035			
Tyr	Ser	Asn	Ile	Met	Asn	Phe	Phe	Lys	Thr	Glu	Ile	Thr	Leu	Ala	
	1040					1045					1050				
Asn	Gly	Glu	Ile	Arg	Lys	Arg	Pro	Leu	Ile	Glu	Thr	Asn	Gly	Glu	

```

<210> SEQ ID NO 43
<211> LENGTH: 80
<212> TYPE: PRT
<213> ORGANISM: Arabidopsis thaliana

<400> SEQUENCE: 43

Met Ala Ser Ser Met Leu Ser Ser Ala Thr Met Val Ala Ser Pro Ala
1             5             10             15

Gln Ala Thr Met Val Ala Pro Phe Asn Gly Leu Lys Ser Ser Ala Ala
                20             25             30

```

-continued

```

Phe Pro Ala Thr Arg Lys Ala Asn Asn Asp Ile Thr Ser Ile Thr Ser
      35              40              45

Asn Gly Gly Arg Val Asn Cys Met Gln Val Trp Pro Pro Ile Gly Lys
      50              55              60

Lys Lys Phe Glu Thr Leu Ser Tyr Leu Pro Asp Leu Thr Asp Ser Glu
      65              70              75              80

```

```

<210> SEQ ID NO 44
<211> LENGTH: 80
<212> TYPE: PRT
<213> ORGANISM: Arabidopsis thaliana

```

```

<400> SEQUENCE: 44

```

```

Met Ala Ser Asn Ser Leu Met Ser Cys Gly Ile Ala Ala Val Tyr Pro
1      5              10              15

Ser Leu Leu Ser Ser Ser Lys Ser Lys Phe Val Ser Ala Gly Val Pro
      20              25              30

Leu Pro Asn Ala Gly Asn Val Gly Arg Ile Arg Met Ala Ala His Trp
      35              40              45

Met Pro Gly Glu Pro Arg Pro Ala Tyr Leu Asp Gly Ser Ala Pro Gly
      50              55              60

Asp Phe Gly Phe Asp Pro Leu Gly Leu Gly Glu Val Pro Ala Asn Leu
      65              70              75              80

```

```

<210> SEQ ID NO 45
<211> LENGTH: 80
<212> TYPE: PRT
<213> ORGANISM: Arabidopsis thaliana

```

```

<400> SEQUENCE: 45

```

```

Met Thr Ile Ala Leu Thr Ile Gly Gly Asn Gly Phe Ser Gly Leu Pro
1      5              10              15

Gly Ser Ser Phe Ser Ser Ser Ser Ser Ser Phe Arg Leu Lys Asn Ser
      20              25              30

Arg Arg Lys Asn Thr Lys Met Leu Asn Arg Ser Lys Val Val Cys Ser
      35              40              45

Ser Ser Ser Ser Val Met Asp Pro Tyr Lys Thr Leu Lys Ile Arg Pro
      50              55              60

Asp Ser Ser Glu Tyr Glu Val Lys Lys Ala Phe Arg Gln Leu Ala Lys
      65              70              75              80

```

```

<210> SEQ ID NO 46
<211> LENGTH: 4467
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

```

```

<400> SEQUENCE: 46

```

```

atggcttctct ctatgctctc ttccgctact atggttgccct ctccggetca ggccactatg      60
gtcgtctcctt tcaacggact taagtctctc gctgccttcc cagccacccg caaggctaac      120
aacgacatta cttccatcac aagcaacggc ggaagagtta actgcatgca ggtggataag      180
aagtactcta tcggacttga catcggaacc aactctgttg gatgggctgt taccaccgat      240
gagtacaagg ttccatctaa gaagttcaag gttcttggaa acaccgatag acactctatc      300
aagaagaacc ttatcggtgc tcttcttttc gattctggag agaccgtga ggctaccaga      360

```

-continued

ttgaagagaa	ccgctagaag	aagatacacc	agaagaaaga	acagaatctg	ctaccttcag	420
gaaatcttct	ctaacgagat	ggctaagggt	gatgattctt	tcttcacag	acttgaggag	480
tctttccttg	ttgaggagga	taagaagcac	gagagacacc	caatcttcgg	aaacatcggt	540
gatgagggtg	cttaccacga	gaagtacca	accatctacc	accttagaaa	gaagtgggtt	600
gattctacog	ataaggctga	tcttagactt	atctaccttg	ctcttgctca	catgatcaag	660
ttcagaggac	acttccttat	cgagggagac	cttaaccag	ataactctga	tgttgataag	720
ttgttcatcc	agcttggtca	gacctacaac	cagcttttcg	aggagaaccc	aatcaacgct	780
tctggagttg	atgctaaggc	tatcctttct	gctagacttt	ctaagtctcg	tagacttgag	840
aaccttatcg	ctcagcttcc	aggagagaag	aagaacggac	tttcggaaa	ccttatcgct	900
ctttctcttg	gacttacccc	aaacttcaag	tctaacttcg	atcttgctga	ggatgctaag	960
ttgcagcttt	ctaaggatac	ctacgatgat	gatcttgata	accttcttgc	tcagatcgga	1020
gatcagtacg	ctgatctttt	ccttgctgct	aagaaccttt	ctgatgctat	ccttctttct	1080
gacatcctta	gagttaacac	cgagatcacc	aaggctccac	tttctgcttc	tatgatcaag	1140
agatacgatg	agcaccacca	ggatcttacc	cttttgaagg	ctcttgtag	acagcagctt	1200
ccagagaagt	acaaggaaat	cttcttcgat	cagtctaaga	acggatacgc	tggatacatc	1260
gatggaggag	cttctcagga	ggagttctac	aagttcatca	agccaatcct	tgagaagatg	1320
gatggaaccg	aggagcttct	tgtaagttg	aacagagagg	atcttcttag	aaagcagaga	1380
accttcgata	acggatctat	cccacaccag	atccaccttg	gagagcttca	cgctatcctt	1440
cgtagacagg	aggatttcta	ccattctctg	aaggataaca	gagagaagat	cgagaagatc	1500
cttaccttca	gaatcccata	ctacgttgga	ccacttgcta	gaggaaactc	tcgtttcgtc	1560
tggatgacca	gaaagtctga	ggagaccatc	acctcttgga	acttcgagga	ggtaagtttc	1620
tgcttctacc	tttgatatat	atataataat	tatcattaat	tagtagtaat	ataatatctc	1680
aaatatTTTT	ttcaaaataa	aagaatgtag	tatatagcaa	tgcttttct	gtagtttata	1740
agtgtgtata	ttttaattta	taacttttct	aatatatgac	caaaatttgt	tgatgtgcag	1800
gttggtgata	agggagcttc	tgtcagctct	ttcatcgaga	gaatgaccaa	cttcgataag	1860
aaccttccaa	acgagaaggt	tcttccaaag	cactctcttc	tttacgagta	cttcaccggt	1920
tacaacgagc	ttaccaaggt	taagtacgtt	accgagggaa	tgagaaagcc	agctttcctt	1980
tctggagagc	agaagaaggc	tatcgttgat	cttcttttca	agaccaacag	aaaggttacc	2040
gttaagcagt	tgaaggagga	ttacttcaag	aagatcgagt	gcttcgattc	tgttgaaatc	2100
tctggagttg	aggatagatt	caacgcttct	cttggaaacct	accacgatct	tttgaagatc	2160
atcaaggata	aggatttctc	tgataacgag	gagaacgagg	acatccttga	ggacatcggt	2220
cttaccttca	cccttttcga	ggatagagag	atgatcgagg	agagactcaa	gacctacgct	2280
caccttttcg	atgataaggt	tatgaagcag	ttgaagagaa	gaagatacac	cggatggggt	2340
agactttctc	gtaagttgat	caacggaatc	agagataagc	agtcctggaa	gaccatcctt	2400
gatttcttga	agtcctgatg	attcgctaac	agaaacttca	tgcagcttat	ccacgatgat	2460
tctcttacct	tcaaggagga	catccagaag	gtcagggttt	ctggacaggg	agattctctt	2520
cacgagcaca	tcgctaacct	tgtcggatct	ccagctatca	agaagggaat	ccttcagacc	2580
gttaaggttg	ttgatgagct	tgtaaggtt	atgggtagac	acaagccaga	gaacatcggt	2640

-continued

atcgagatgg cttagagagaa ccagaccacc cagaaggagc agaagaactc tcgtgagaga	2700
atgaagagaa tcgaggaggg aatcaaggag cttggatctc aaatcttgaa ggagcacca	2760
gttgagaaca cccagcttca gaacgagaag ttgtacctt actaccttca gaacggaaga	2820
gatatgtacg ttgatcagga gcttgacatc aacagacttt ctgattacga tgttgatcac	2880
atcgttccac agtctttctt gaaggatgat tctatcgata acaaggttct taccggttct	2940
gataagaaca gaggaaagtc tgataacgtt ccatctgagg aggttggtta gaagatgaag	3000
aactactgga gacagcttct taacgctaag ttgatcacc agagaaagtt cgataacctt	3060
accaaggctg agagaggagg actttctgag cttgataagg ctggattcat caagagacag	3120
cttggtgaga ccagacagat caccaagcac gttgctcaga tccttgattc tcgtatgaac	3180
accaagtacg atgagaacga taagttgatc agagaggtta aggttatcac cttgaagtct	3240
aagttggttt ctgatttcag aaaggatttc cagttctaca aggttagaga gatcaacaac	3300
taccaccacg ctcacgatgc ttaccttaac gctgtgtgtg gaaccgctct tatcaagaag	3360
taccaaaagt tggagtctga gttcgtttac ggagattaca aggtttacga tgttagaaag	3420
atgatcgcta agtctgagca ggagatcgga aaggctaccg ctaagtactt cttctactct	3480
aacatcatga acttcttcaa gaccgagatc acccttgcta acggagagat cagaaagaga	3540
ccacttatcg agaccaacgg agagaccgga gagatcgttt gggataaggg aagagatttc	3600
gctaccgtta gaaaggttct ttctatgcca caggttaaca tcgttaagaa aaccgaggtt	3660
cagaccggag gattctctaa ggagtctatc cttccaaaga gaaactctga taagttgatc	3720
gctagaaaga aggattggga cccaaagaag tacggaggat tcgattctcc aaccgttgct	3780
tactctgttc ttgttgttgc taaggttgag aagggaagt ctaagaagtt gaagtctgtt	3840
aaggagcttc ttggaatcac catcatggag cgttcttctt tcgagaagaa cccaatcgat	3900
ttccttgagg ctaagggata caaggagggtt aagaaggatc ttatcatcaa gttgccaaag	3960
tactctcttt tcgagcttga gaacggaaga aagagaatgc ttgcttctgc tggagagctt	4020
cagaagggaa acgagcttgc tcttccatct aagtacgtta acttccctta ccttgcttct	4080
cactacgaga agttgaaggg atctccagag gataacgagc agaagcagct ttctgttgag	4140
cagcacaagc actaccttga tgagatcacc gagcaaactc ctgagttctc taagagagtt	4200
atccttgctg atgctaacct tgataagggt ctttctgctt acaacaagca cagagataag	4260
ccaatcagag agcagggtga gaacatcacc caccctttca cccttaccaa ccttggtgct	4320
ccagctgctt tcaagtactt cgataccacc atcgatagaa aaagatacac ctctaccaag	4380
gaggttcttg atgctaccct tatccaccag tctatcaccg gactttacga gaccagaatc	4440
gatctttctc agcttgagg agattga	4467

<210> SEQ ID NO 47

<211> LENGTH: 1424

<212> TYPE: PRT

<213> ORGANISM: Artificial sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 47

Met Ala Ser Ser Met Leu Ser Ser Ala Thr Met Val Ala Ser Pro Ala

1

5

10

15

-continued

Gln	Ala	Thr	Met	Val	Ala	Pro	Phe	Asn	Gly	Leu	Lys	Ser	Ser	Ala	Ala
			20					25					30		
Phe	Pro	Ala	Thr	Arg	Lys	Ala	Asn	Asn	Asp	Ile	Thr	Ser	Ile	Thr	Ser
		35					40					45			
Asn	Gly	Gly	Arg	Val	Cys	Met	Gln	Val	Asp	Lys	Lys	Tyr	Ser	Ile	Gly
	50					55					60				
Leu	Asp	Ile	Gly	Thr	Asn	Ser	Val	Gly	Trp	Ala	Val	Ile	Thr	Asp	Glu
65					70				75						80
Tyr	Lys	Val	Pro	Ser	Lys	Lys	Phe	Lys	Val	Leu	Gly	Asn	Thr	Asp	Arg
				85					90					95	
His	Ser	Ile	Lys	Lys	Asn	Leu	Ile	Gly	Ala	Leu	Leu	Phe	Asp	Ser	Gly
			100					105					110		
Glu	Thr	Ala	Glu	Ala	Thr	Arg	Leu	Lys	Arg	Thr	Ala	Arg	Arg	Arg	Tyr
		115					120					125			
Thr	Arg	Arg	Lys	Asn	Arg	Ile	Cys	Tyr	Leu	Gln	Glu	Ile	Phe	Ser	Asn
	130					135					140				
Glu	Met	Ala	Lys	Val	Asp	Asp	Ser	Phe	Phe	His	Arg	Leu	Glu	Glu	Ser
145					150					155					160
Phe	Leu	Val	Glu	Glu	Asp	Lys	Lys	His	Glu	Arg	His	Pro	Ile	Phe	Gly
			165						170					175	
Asn	Ile	Val	Asp	Glu	Val	Ala	Tyr	His	Glu	Lys	Tyr	Pro	Thr	Ile	Tyr
		180						185					190		
His	Leu	Arg	Lys	Lys	Leu	Val	Asp	Ser	Thr	Asp	Lys	Ala	Asp	Leu	Arg
		195					200					205			
Leu	Ile	Tyr	Leu	Ala	Leu	Ala	His	Met	Ile	Lys	Phe	Arg	Gly	His	Phe
	210					215					220				
Leu	Ile	Glu	Gly	Asp	Leu	Asn	Pro	Asp	Asn	Ser	Asp	Val	Asp	Lys	Leu
225					230					235					240
Phe	Ile	Gln	Leu	Val	Gln	Thr	Tyr	Asn	Gln	Leu	Phe	Glu	Glu	Asn	Pro
			245						250					255	
Ile	Asn	Ala	Ser	Gly	Val	Asp	Ala	Lys	Ala	Ile	Leu	Ser	Ala	Arg	Leu
			260					265					270		
Ser	Lys	Ser	Arg	Arg	Leu	Glu	Asn	Leu	Ile	Ala	Gln	Leu	Pro	Gly	Glu
		275					280					285			
Lys	Lys	Asn	Gly	Leu	Phe	Gly	Asn	Leu	Ile	Ala	Leu	Ser	Leu	Gly	Leu
	290					295					300				
Thr	Pro	Asn	Phe	Lys	Ser	Asn	Phe	Asp	Leu	Ala	Glu	Asp	Ala	Lys	Leu
305					310					315					320
Gln	Leu	Ser	Lys	Asp	Thr	Tyr	Asp	Asp	Asp	Leu	Asp	Asn	Leu	Leu	Ala
			325						330					335	
Gln	Ile	Gly	Asp	Gln	Tyr	Ala	Asp	Leu	Phe	Leu	Ala	Ala	Lys	Asn	Leu
		340						345					350		
Ser	Asp	Ala	Ile	Leu	Leu	Ser	Asp	Ile	Leu	Arg	Val	Asn	Thr	Glu	Ile
		355					360					365			
Thr	Lys	Ala	Pro	Leu	Ser	Ala	Ser	Met	Ile	Lys	Arg	Tyr	Asp	Glu	His
	370					375					380				
His	Gln	Asp	Leu	Thr	Leu	Leu	Lys	Ala	Leu	Val	Arg	Gln	Gln	Leu	Pro
385					390					395					400
Glu	Lys	Tyr	Lys	Glu	Ile	Phe	Phe	Asp	Gln	Ser	Lys	Asn	Gly	Tyr	Ala
			405						410					415	
Gly	Tyr	Ile	Asp	Gly	Gly	Ala	Ser	Gln	Glu	Glu	Phe	Tyr	Lys	Phe	Ile

-continued

420							425					430				
Lys	Pro	Ile	Leu	Glu	Lys	Met	Asp	Gly	Thr	Glu	Glu	Leu	Leu	Val	Lys	
435							440					445				
Leu	Asn	Arg	Glu	Asp	Leu	Leu	Arg	Lys	Gln	Arg	Thr	Phe	Asp	Asn	Gly	
450							455					460				
Ser	Ile	Pro	His	Gln	Ile	His	Leu	Gly	Glu	Leu	His	Ala	Ile	Leu	Arg	
465							470					475				
Arg	Gln	Glu	Asp	Phe	Tyr	Pro	Phe	Leu	Lys	Asp	Asn	Arg	Glu	Lys	Ile	
485							490					495				
Glu	Lys	Ile	Leu	Thr	Phe	Arg	Ile	Pro	Tyr	Tyr	Val	Gly	Pro	Leu	Ala	
500							505					510				
Arg	Gly	Asn	Ser	Arg	Phe	Ala	Trp	Met	Thr	Arg	Lys	Ser	Glu	Glu	Thr	
515							520					525				
Ile	Thr	Pro	Trp	Asn	Phe	Glu	Glu	Val	Val	Asp	Lys	Gly	Ala	Ser	Ala	
530							535					540				
Gln	Ser	Phe	Ile	Glu	Arg	Met	Thr	Asn	Phe	Asp	Lys	Asn	Leu	Pro	Asn	
545							550					555				
Glu	Lys	Val	Leu	Pro	Lys	His	Ser	Leu	Leu	Tyr	Glu	Tyr	Phe	Thr	Val	
565							570					575				
Tyr	Asn	Glu	Leu	Thr	Lys	Val	Lys	Tyr	Val	Thr	Glu	Gly	Met	Arg	Lys	
580							585					590				
Pro	Ala	Phe	Leu	Ser	Gly	Glu	Gln	Lys	Lys	Ala	Ile	Val	Asp	Leu	Leu	
595							600					605				
Phe	Lys	Thr	Asn	Arg	Lys	Val	Thr	Val	Lys	Gln	Leu	Lys	Glu	Asp	Tyr	
610							615					620				
Phe	Lys	Lys	Ile	Glu	Cys	Phe	Asp	Ser	Val	Glu	Ile	Ser	Gly	Val	Glu	
625							630					635				
Asp	Arg	Phe	Asn	Ala	Ser	Leu	Gly	Thr	Tyr	His	Asp	Leu	Leu	Lys	Ile	
645							650					655				
Ile	Lys	Asp	Lys	Asp	Phe	Leu	Asp	Asn	Glu	Glu	Asn	Glu	Asp	Ile	Leu	
660							665					670				
Glu	Asp	Ile	Val	Leu	Thr	Leu	Thr	Leu	Phe	Glu	Asp	Arg	Glu	Met	Ile	
675							680					685				
Glu	Glu	Arg	Leu	Lys	Thr	Tyr	Ala	His	Leu	Phe	Asp	Asp	Lys	Val	Met	
690							695					700				
Lys	Gln	Leu	Lys	Arg	Arg	Arg	Tyr	Thr	Gly	Trp	Gly	Arg	Leu	Ser	Arg	
705							710					715				
Lys	Leu	Ile	Asn	Gly	Ile	Arg	Asp	Lys	Gln	Ser	Gly	Lys	Thr	Ile	Leu	
725							730					735				
Asp	Phe	Leu	Lys	Ser	Asp	Gly	Phe	Ala	Asn	Arg	Asn	Phe	Met	Gln	Leu	
740							745					750				
Ile	His	Asp	Asp	Ser	Leu	Thr	Phe	Lys	Glu	Asp	Ile	Gln	Lys	Ala	Gln	
755							760					765				
Val	Ser	Gly	Gln	Gly	Asp	Ser	Leu	His	Glu	His	Ile	Ala	Asn	Leu	Ala	
770							775					780				
Gly	Ser	Pro	Ala	Ile	Lys	Lys	Gly	Ile	Leu	Gln	Thr	Val	Lys	Val	Val	
785							790					795				
Asp	Glu	Leu	Val	Lys	Val	Met	Gly	Arg	His	Lys	Pro	Glu	Asn	Ile	Val	
805							810					815				
Ile	Glu	Met	Ala	Arg	Glu	Asn	Gln	Thr	Thr	Gln	Lys	Gly	Gln	Lys	Asn	
820							825					830				

Ser	Arg	Glu	Arg	Met	Lys	Arg	Ile	Glu	Gly	Ile	Lys	Glu	Leu	Gly	
835						840			845						
Ser	Gln	Ile	Leu	Lys	Glu	His	Pro	Val	Glu	Asn	Thr	Gln	Leu	Gln	Asn
850						855			860						
Glu	Lys	Leu	Tyr	Leu	Tyr	Tyr	Leu	Gln	Asn	Gly	Arg	Asp	Met	Tyr	Val
865						870			875			880			
Asp	Gln	Glu	Leu	Asp	Ile	Asn	Arg	Leu	Ser	Asp	Tyr	Asp	Val	Asp	His
			885						890			895			
Ile	Val	Pro	Gln	Ser	Phe	Leu	Lys	Asp	Asp	Ser	Ile	Asp	Asn	Lys	Val
			900						905			910			
Leu	Thr	Arg	Ser	Asp	Lys	Asn	Arg	Gly	Lys	Ser	Asp	Asn	Val	Pro	Ser
915						920						925			
Glu	Glu	Val	Val	Lys	Lys	Met	Lys	Asn	Tyr	Trp	Arg	Gln	Leu	Leu	Asn
930						935						940			
Ala	Lys	Leu	Ile	Thr	Gln	Arg	Lys	Phe	Asp	Asn	Leu	Thr	Lys	Ala	Glu
945						950			955			960			
Arg	Gly	Gly	Leu	Ser	Glu	Leu	Asp	Lys	Ala	Gly	Phe	Ile	Lys	Arg	Gln
			965						970			975			
Leu	Val	Glu	Thr	Arg	Gln	Ile	Thr	Lys	His	Val	Ala	Gln	Ile	Leu	Asp
			980						985			990			
Ser	Arg	Met	Asn	Thr	Lys	Tyr	Asp	Glu	Asn	Asp	Lys	Leu	Ile	Arg	Glu
995						1000						1005			
Val	Lys	Val	Ile	Thr	Leu	Lys	Ser	Lys	Leu	Val	Ser	Asp	Phe	Arg	
1010						1015						1020			
Lys	Asp	Phe	Gln	Phe	Tyr	Lys	Val	Arg	Glu	Ile	Asn	Asn	Tyr	His	
1025						1030						1035			
His	Ala	His	Asp	Ala	Tyr	Leu	Asn	Ala	Val	Val	Gly	Thr	Ala	Leu	
1040						1045						1050			
Ile	Lys	Lys	Tyr	Pro	Lys	Leu	Glu	Ser	Glu	Phe	Val	Tyr	Gly	Asp	
1055						1060						1065			
Tyr	Lys	Val	Tyr	Asp	Val	Arg	Lys	Met	Ile	Ala	Lys	Ser	Glu	Gln	
1070						1075						1080			
Glu	Ile	Gly	Lys	Ala	Thr	Ala	Lys	Tyr	Phe	Phe	Tyr	Ser	Asn	Ile	
1085						1090						1095			
Met	Asn	Phe	Phe	Lys	Thr	Glu	Ile	Thr	Leu	Ala	Asn	Gly	Glu	Ile	
1100						1105						1110			
Arg	Lys	Arg	Pro	Leu	Ile	Glu	Thr	Asn	Gly	Glu	Thr	Gly	Glu	Ile	
1115						1120						1125			
Val	Trp	Asp	Lys	Gly	Arg	Asp	Phe	Ala	Thr	Val	Arg	Lys	Val	Leu	
1130						1135						1140			
Ser	Met	Pro	Gln	Val	Asn	Ile	Val	Lys	Lys	Thr	Glu	Val	Gln	Thr	
1145						1150						1155			
Gly	Gly	Phe	Ser	Lys	Glu	Ser	Ile	Leu	Pro	Lys	Arg	Asn	Ser	Asp	
1160						1165						1170			
Lys	Leu	Ile	Ala	Arg	Lys	Lys	Asp	Trp	Asp	Pro	Lys	Lys	Tyr	Gly	
1175						1180						1185			
Gly	Phe	Asp	Ser	Pro	Thr	Val	Ala	Tyr	Ser	Val	Leu	Val	Val	Ala	
1190						1195						1200			
Lys	Val	Glu	Lys	Gly	Lys	Ser	Lys	Lys	Leu	Lys	Ser	Val	Lys	Glu	
1205						1210						1215			

-continued

Leu	Leu	Gly	Ile	Thr	Ile	Met	Glu	Arg	Ser	Ser	Phe	Glu	Lys	Asn
1220						1225					1230			
Pro	Ile	Asp	Phe	Leu	Glu	Ala	Lys	Gly	Tyr	Lys	Glu	Val	Lys	Lys
1235						1240					1245			
Asp	Leu	Ile	Ile	Lys	Leu	Pro	Lys	Tyr	Ser	Leu	Phe	Glu	Leu	Glu
1250						1255					1260			
Asn	Gly	Arg	Lys	Arg	Met	Leu	Ala	Ser	Ala	Gly	Glu	Leu	Gln	Lys
1265						1270					1275			
Gly	Asn	Glu	Leu	Ala	Leu	Pro	Ser	Lys	Tyr	Val	Asn	Phe	Leu	Tyr
1280						1285					1290			
Leu	Ala	Ser	His	Tyr	Glu	Lys	Leu	Lys	Gly	Ser	Pro	Glu	Asp	Asn
1295						1300					1305			
Glu	Gln	Lys	Gln	Leu	Phe	Val	Glu	Gln	His	Lys	His	Tyr	Leu	Asp
1310						1315					1320			
Glu	Ile	Ile	Glu	Gln	Ile	Ser	Glu	Phe	Ser	Lys	Arg	Val	Ile	Leu
1325						1330					1335			
Ala	Asp	Ala	Asn	Leu	Asp	Lys	Val	Leu	Ser	Ala	Tyr	Asn	Lys	His
1340						1345					1350			
Arg	Asp	Lys	Pro	Ile	Arg	Glu	Gln	Ala	Glu	Asn	Ile	Ile	His	Leu
1355						1360					1365			
Phe	Thr	Leu	Thr	Asn	Leu	Gly	Ala	Pro	Ala	Ala	Phe	Lys	Tyr	Phe
1370						1375					1380			
Asp	Thr	Thr	Ile	Asp	Arg	Lys	Arg	Tyr	Thr	Ser	Thr	Lys	Glu	Val
1385						1390					1395			
Leu	Asp	Ala	Thr	Leu	Ile	His	Gln	Ser	Ile	Thr	Gly	Leu	Tyr	Glu
1400						1405					1410			
Thr	Arg	Ile	Asp	Leu	Ser	Gln	Leu	Gly	Gly	Asp				
1415						1420								

<210> SEQ ID NO 48
 <211> LENGTH: 330
 <212> TYPE: DNA
 <213> ORGANISM: Artificial sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 48

ttggcgaaac cccatttcga cctttcggtc tcatcagggg tggcacacac caccctatgg	60
ggagaggtcg tcctctatct ctctcggaag gccggagcaa tccaaaagag gtacaccac	120
ccatgggtcg ggacttttaa ttcgaggat tcgtccttta aacgttcctc caagagtcce	180
ttccccaaac ccttactttg taagtgtggt tcggcgaatg taccgtttcg tcctttcgga	240
ctcatcaggg aaagtacaca ctttccgacg gtgggttcgt cgacacctct cccctccca	300
ggtactatcc cctttccagg atttgttccc	330

<210> SEQ ID NO 49
 <211> LENGTH: 942
 <212> TYPE: DNA
 <213> ORGANISM: Arabidopsis thaliana

<400> SEQUENCE: 49

acgagaggaa gtacattagt ttggagaaga gtaatagaca gagagataga gagaaagaga	60
agcagttcgg agaaacaatg gcggtagaag aactcccaa atctgttgta acggaagaag	120

-continued

ctaagcctaa ttcaatagag aatccgattg atcgatacca tgaggaaggt gatgatgccg	180
aagaaggaga gatcgccgga ggagaaggag acggaacgt tgacgaatcg agcaaatccg	240
gtgttcctga atcgatcct ctggaacatt catggacttt ctggttcgat aatcctgctg	300
tgaaatcgaa acaaacctct tggggaagtt ccttgcgacc cgtgtttacg ttttcaactg	360
ttgaggaatt ttggagtgtg tacaacaaca tgaagcatcc gagcaagtta gctcacggag	420
ctgacttcta ctgtttcaaa cacatcattg aacctaagtg ggaggatcct atttgtgcta	480
atggaggaaa atggactatg actttcccta aggagaagtc tgataagagc tggtcttaca	540
ctttgcttgc attgattgga gagcagtttg atcatggaga tgaaatatgt ggagcagttg	600
tcaacattag aggaagcaa gaaagatat ctatttgac taaaaatgct tcaaacgaag	660
ctgctcaggt gagcattgga aaacaatgga aggagtttct cgattacaac aacagcatag	720
gtttcatcat ccatgaggat gcgaagaagc tcgacaggaa tgcaagaac gcttacaccg	780
cttgaaacct ctcaaactct tgcatgtgtt caattacagt tttgtatgtg agagatctct	840
atttatctaa acatgacttg acagtctgtc ttgctagtgt ttgattgttc acgaagctct	900
aacatttcat ttagtaatat attagtatgg ttcttcataa ta	942

<210> SEQ ID NO 50
 <211> LENGTH: 96
 <212> TYPE: DNA
 <213> ORGANISM: Artificial sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthetic Construct
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (1)..(20)
 <223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 50

nnnnnnnnnn nnnnnnnnnn gttttagagc tagaaatagc aagttaaaat aaggctagtc	60
cgttatcaac ttgaaaaagt ggcaccgagt cgggtgc	96

<210> SEQ ID NO 51
 <211> LENGTH: 567
 <212> TYPE: DNA
 <213> ORGANISM: Pseudomonas aeruginosa

<400> SEQUENCE: 51

atgggtgatc attatctgga tattcggtg aggctgatc cagagtccc acctgcgcag	60
ctgatgtctg tccttttttg caaacttcat caggccctgg ttgccaggg cggagatcgg	120
ataggggtaa gctttccaga cctcgacgaa agccggagcc gcctgggaga acgctgcgg	180
atccacgctt ctgccgacga tctgagagcc ttgctggcaa ggccatggct tgaggggctc	240
cgggatcacc tgcagtgttg cgaaccgcc gttgttcccc acccaacccc ttatcggcag	300
gtgtctagag tgcaggccaa atctaacca gaacggctgc gacggcgact catgcggcga	360
catgatctta gcgaggaaga ggcccgaata agaataccctg ataccgtggc ccgcgcccct	420
gacttgctt ttgtcacact gcgggtcccag agtacggggc agcatttcag acttttcatt	480
cgacacgggc cactgcaagt taccgcccga gaaggaggt ttacttgtaa tggactctcc	540
aaggaggtt tcgtgccctg gttttga	567

<210> SEQ ID NO 52

-continued

<211> LENGTH: 188
<212> TYPE: PRT
<213> ORGANISM: *Pseudomonas aeruginosa*

<400> SEQUENCE: 52

Met Gly Asp His Tyr Leu Asp Ile Arg Leu Arg Pro Asp Pro Glu Phe
1 5 10 15

Pro Pro Ala Gln Leu Met Ser Val Leu Phe Gly Lys Leu His Gln Ala
20 25 30

Leu Val Ala Gln Gly Gly Asp Arg Ile Gly Val Ser Phe Pro Asp Leu
35 40 45

Asp Glu Ser Arg Ser Arg Leu Gly Glu Arg Leu Arg Ile His Ala Ser
50 55 60

Ala Asp Asp Leu Arg Ala Leu Leu Ala Arg Pro Trp Leu Glu Gly Leu
65 70 75 80

Arg Asp His Leu Gln Phe Gly Glu Pro Ala Val Val Pro His Pro Thr
85 90 95

Pro Tyr Arg Gln Val Ser Arg Val Gln Ala Lys Ser Asn Pro Glu Arg
100 105 110

Leu Arg Arg Arg Leu Met Arg Arg His Asp Leu Ser Glu Glu Glu Ala
115 120 125

Arg Lys Arg Ile Pro Asp Thr Val Ala Arg Ala Leu Asp Leu Pro Phe
130 135 140

Val Thr Leu Arg Ser Gln Ser Thr Gly Gln His Phe Arg Leu Phe Ile
145 150 155 160

Arg His Gly Pro Leu Gln Val Thr Ala Glu Glu Gly Gly Phe Thr Cys
165 170 175

Tyr Gly Leu Ser Lys Gly Gly Phe Val Pro Trp Phe
180 185

<210> SEQ ID NO 53
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: *Pseudomonas aeruginosa*

<400> SEQUENCE: 53

gttcactgcc gtataggcag

20

<210> SEQ ID NO 54
<211> LENGTH: 272
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (21)..(40)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (137)..(156)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 54

gttcactgcc gtataggcag nnnnnnnnnn nnnnnnnnnn gtttagagc tagaaatagc 60

aagttaaaat aaggctagtc cggtatcaac ttgaaaaagt ggcaccgagt cggtagcgttc 120

actgccgtat aggaggnnnn nnnnnnnnnn nnnnnngttt tagagctaga aatagcaagt 180

-continued

taaaaataagg ctagtccggt atcaacttga aaaagtggca ccgagtcggt gcgttcactg	240
ccgtataggc aggttcactg ccgtataggc ag	272
 <210> SEQ ID NO 55	
<211> LENGTH: 3213	
<212> TYPE: DNA	
<213> ORGANISM: Nicotiana tabacum	
 <400> SEQUENCE: 55	
atgctcgggg atggaaatga gggaaatatct acaatacctg gatttaatca gatacaattt	60
gaaggatttt gtaggttcat tgatcaaggt ttgacggaag aactttataa gtttccaaaa	120
attgaagata cagatcaaga aattgaattt caattatttg tggaacata tcaattggtc	180
gaacccttga taaaggaaag agatgctgtg tatgaatcac tcacatattc ttctgaatta	240
tatgtatccg cgggattaat ttggaaaaac agtagggata tgcaagaaca aacaattttt	300
atcggaaca ttctctaat gaattccctg ggaacttcta tagtcaatgg aatatataga	360
attgtgatca atcaaatatt gcaaagtccc ggtatttatt accgatcaga attggaccat	420
aacggaattt cgggtctatc cggcaccata atatcagatt ggggaggaag atcagaatta	480
gaaattgata gaaaagcaag gatatgggct cgtgtaagta ggaacaaaa aatatctatt	540
ctagtcttat catcagctat gggtttgaat ctaagagaaa ttctagagaa tgtttgetat	600
cctgaaattt ttttgtcttt tctgagtgat aaggagagaa aaaaaattgg gtcaaaagaa	660
aatgccattt tggagtttta tcaacaattt gcttgtgtag gtggcgatcc ggtattttct	720
gaatccttat gtaaggaatt acaaaagaaa ttctttcaac aaagatgtga attaggaagg	780
attggtcgac gaaatatgaa ccgaagactg aaccttgata taccocagaa caatacattt	840
ttgttaccac gagatatatt ggcagccgcc gatcatttga ttgggctgaa atttggatg	900
ggtgcacttg acgatatgaa tcatttgaaa aataaacgta ttcgttctgt agcagatctt	960
ttacaagatc aattcggatt ggcctctggt cgttttagaaa atgtggttcg ggggactata	1020
tgtggagcaa ttccgcataa attgataccg acacctcaga atttggtaac ctcaactcca	1080
ttaacaacta cttatgaatc ctttttcggt ttacacccat tatctcaagt ttggatcga	1140
actaatccat tgacacaaat agttcatggg agaaaattaa gttatttggg ccctggagga	1200
ctgacagggc gcaactgctag ttttcggata cgagatatcc atcctagtca ctatggacgt	1260
atttgcccaa ttgacacatc tgaaggaatc aatgttggac ttattggatc cttagcaatt	1320
catgcgagga ttggtcattg gggatctcta gaaagccctt tttatgaaat ttctgagagg	1380
tcaaccgggg tacggatgct ttatttatca ccaggtagag atgaatacta tatggtagcg	1440
gcaggaaatt ctttagcctt aaatcaggat attcaggaag aacagggtgt tccagctcga	1500
taccgtcaag aattcttgac tattgcatgg gaacagggtc atcttcgaag tatttttcct	1560
tttcaatatt tttctatttg agcttccctc attcctttta tcgaacataa tgatgcgaat	1620
cgagctttta tgagtcttaa tatgcaacgt caagcagttc ctctttctcg ctccgagaaa	1680
tgcatgttg gaactggggt ggaacgacaa gcagctctag attcgggggc tcttgctata	1740
gccgaacgag agggaagggt cgtttatacc aatactgaca agattctttt agcaggtaat	1800
ggagatatcc taagcattcc attagttata tatcaacggt ccaataaaaa tacttgatatg	1860
catcaaaaaa tccagggttc tcggggtaaa tgcattaaaa agggacaaa ttttagcggt	1920

-continued

```

ggtgctgcta cgggtggtgg cgaacttgct ttggggaaaa acgtattagt agcttatatg 1980
ccgtgggagg gttacaattc tgaagatgca gtacttatta gcgagcgttt ggtatatgaa 2040
gatatttata cttcttttca catacggaaa tatgaaattc agactcatgt gacaagccaa 2100
ggccctgaaa aagtaactaa tgaataaccg catttagaag cccatttact ccgcaattta 2160
gataaaaaatg gaattgtgat gctgggatct tgggtagaga caggtgatat tttagtaggt 2220
aaattaacac ccaggtcgt gaaagaatcg tcgtatgcc cggaagatag attgttacga 2280
gctatacttg gtattcaggt atctacttca aaagaaactt gtctaaaact acctataggt 2340
ggcaggggtc ggggtattga tgtgaggtgg atccagaaga ggggtggttc tagttataat 2400
cccgaaacga ttcgtgtata tattttacag aaacgtgaaa tcaaagtagg cgataaagta 2460
gctggaagac acggaaataa aggtatcatt tccaaaattt tgcctagaca agatatgcct 2520
tatttacaag atggaagatc cgttgatatg gtctttaacc cattaggagt accttcacga 2580
atgaatgtag gacagatatt tgaatgttca ctagggttag cagggagtct gctagacaga 2640
cattatcgaa tagcaccttt tgatgagaga tatgaacaag aagcttcgag aaaacttgtg 2700
ttttctgaat tatatgaagc cagtaagcaa acagcgaatc catgggtatt tgaaccggaa 2760
tatccaggaa aaagcagaat atttgatgga aggacgggga atccttttga acaaccggtt 2820
ataataggaa agccttatat cttgaaatta attcatcaag ttgatgataa aatccatggg 2880
cgctccagt gacattatgc gcttggttaca caacaacccc ttagaggaag agccaaacag 2940
gggggacagc gggtaggaga aatggagggt tgggctctag aaggggttgg ggttgctcat 3000
attttacaag agatgcttac ttataaatcg gatcatatta gagctcgcca ggaagtactt 3060
ggtactacga tcattggggg aacaatacct aatcccgaag atgctccaga atcttttcga 3120
ttgctcgttc gagaactacg atcttttagct ctggaactga atcatttctt tgtatctgag 3180
aagaacttcc agattaatag gaaggaagct taa 3213

```

```

<210> SEQ ID NO 56
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum

```

```

<400> SEQUENCE: 56

```

```

ttagaggaag agccaaacag 20

```

```

<210> SEQ ID NO 57
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum

```

```

<400> SEQUENCE: 57

```

```

cttgctatag ccgaacgcga 20

```

```

<210> SEQ ID NO 58
<211> LENGTH: 1062
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum

```

```

<400> SEQUENCE: 58

```

```

atgactgcaa ttttagagag acgcgaaagc gaaagcctat ggggtcgtt ctgtaactgg 60

```

```

ataactagca ctgaaaaccg tctttacatt ggatggtttg gtgttttgat gatccctacc 120

```

-continued

ttattgacgg caacttctgt atttattatt gccttcattg ctgctcctcc agtagacatt	180
gatggtattc gtgaacctgt ttcagggtct ctactttacg gaaacaatat ttttccggt	240
gccattattc ctacttctgc agctataggt ttacattttt acccaatctg ggaagcggca	300
tccgttgatg aatggttata caacggtggt ccttatgaac taattgttct acacttctta	360
cttggcgtag cttgtttacat gggtcgtgag tgggagctta gtttcctctt gggtagcgca	420
ccttggtattg ctgttgcata ttcagctcct gttgcagctg ctaccgcagt tttcttgatc	480
taccaaatg gtcaaggaag tttttctgat ggtatgctc taggaatctc tggtaacttc	540
aatttcatga ttgtattcca ggtcagcac aacatcctta tgcacccatt tcacatgtta	600
ggcgtagctg gtgtattcgg cggtcccta ttcagtgcata tgcattggtc cttggtaact	660
tctagtttga tcagggaaac cacagaaaat gaactctgta atgaaggta cagattcgggt	720
caagaggaag aaacttataa catcgtagcc gctcatggtt attttggcgg attgatcttc	780
caatatgcta gtttcaacaa ctctcgttcg ttacacttct tcctagctgc ttggcctgta	840
gtaggtatct ggtttaccgc tttaggtatc agcactatgg ctttcaacct aaatggtttc	900
aatttcaacc aatctgtagt tgacagtcaa ggcgtgtaa ttaatacttg ggctgatatc	960
attaaccgtg ctaaccttgg tatggaagtt atgcatgaac gtaatgctca caacttcctt	1020
ctagacctag ctgctatcga agctccatct acaaatggat aa	1062

<210> SEQ ID NO 59
 <211> LENGTH: 20
 <212> TYPE: DNA
 <213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 59

gttgatgaat ggttatacaa	20
-----------------------	----

<210> SEQ ID NO 60
 <211> LENGTH: 20
 <212> TYPE: DNA
 <213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 60

gatgatccct accttattga	20
-----------------------	----

<210> SEQ ID NO 61
 <211> LENGTH: 264
 <212> TYPE: DNA
 <213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 61

atggtaaaaa attctgtcat ttcagttatt tctcaagaag aaaagagagg atctgttgaa	60
tttcaagtat tcaatttcac caataagata cggagactta cttcacattt agaattgcac	120
aaaaaagact atttatctca gagaggtttg aagaaaaattt tgggaaaacg tcaacgactc	180
ctagcttatt tgtcaaaaaa aaatagagta cgttataaag aattaattaa tcagttggac	240
attcgagaga caaaaactcg ttaa	264

<210> SEQ ID NO 62
 <211> LENGTH: 20
 <212> TYPE: DNA
 <213> ORGANISM: Nicotiana tabacum

-continued

<400> SEQUENCE: 62

atttctcaag aagaaaagag 20

<210> SEQ ID NO 63

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 63

tcaatttcac caataagata 20

<210> SEQ ID NO 64

<211> LENGTH: 201

<212> TYPE: DNA

<213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 64

atggccaagg ggaagatgt ccgagtaacg gtgattttgg aatgtactag ttgtgtccga 60

aacagtgttg ataaggtatc aagaggtatt tccagatata ttactcaaaa gaaccggcac 120

aatacgccta atcgattaga attgaaaaaa ttctgtccct attgttacia acatacgatt 180

catggggaga taaagaaata g 201

<210> SEQ ID NO 65

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 65

gatatattac tcaaaagaac 20

<210> SEQ ID NO 66

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 66

agtgttgata aggtatcaag 20

<210> SEQ ID NO 67

<211> LENGTH: 3213

<212> TYPE: DNA

<213> ORGANISM: Glycine max

<400> SEQUENCE: 67

atgcttgggg atggaaatga aggaatgtct acactacctg gattgaatca gatacaattt 60

gaagggtttt gtaggttcat tgatcggggc ttaccagaag ggctttttta gtttcacaaa 120

attgaggata cagatcaaga aattgaattt caattatttg tagaaacata tcaattatta 180

gaacccttga taaacgaaaa agatgctgta tatgaatcgc ttacatatc tgctgaatta 240

tatgtatctg cgggattaat ttggaaaagt agtagggaca tacaagaaca aactattttt 300

gttggaacaa ttcctttaat gaattctctg ggaacttcta tagtaaatgg aatatacaga 360

attgtaatca atcaaatatt gcaaagccct ggtatttatt accgttcaga attggaccct 420

agcgggaattt cgggtctatc tggcaccata atatcagact gggggggtag attagaatta 480

gagattgata gaaaagcaag gatatgggct cgtgtgagta ggaaacagaa aatatctatt 540

-continued

ctagttttat	catcagctat	gggttcgaat	ttaagcgaaa	ttctagagaa	tgtttggtat	600
cctgaaattt	tcgtttcttt	cctaaatgat	aaggataaaa	aaaaaatagg	gtcaaaagaa	660
aatgccattt	tggagtttta	tcgacaattt	gcttgtgttg	gtggagatcc	agtattttct	720
gaatctttat	gtaaagaatt	acaaaaaaaa	ttttttcaac	aaagatgtga	attaggaagg	780
attggtcgac	gaaatatgaa	ccaaaagctt	aatcttgata	tacctcagaa	caatacattt	840
ttgttaccac	gagatatatt	gacagctgcg	gatcatttga	tggaatgaa	atttggaaatg	900
ggatatactg	acgatataaa	tcatttgaaa	aataaacgta	ttcgttcggt	agcagatcta	960
ttacaagatc	aatttggatt	ggccctgggt	cgtttagaaa	atatggttag	aggaactata	1020
tgtggagcaa	ttagacataa	attgataccg	actcctcaga	atttggtagc	ttcaactcca	1080
ttacaacta	cttatgaatc	tttttttga	ttacatccat	tatctcaagt	tttggatcaa	1140
actaatccat	tgacccaat	agttcatggg	agaaaattga	gttatttggg	ccctggagga	1200
ttgacggggc	gaactgctag	ttttcggata	cgagatatcc	accctagtca	ctatggacgc	1260
atttgtccaa	ttgacacgtc	ggaaggaatc	aatgttggac	ttattggatc	tctagcaatt	1320
catgcgagga	ttggtagtgt	gggggtccata	gaaagtccat	tttatgaaat	atctgagaga	1380
tcaaaaagaa	tacgcatgct	ttatttatca	ccaagtagag	atgaatacta	tatggtagca	1440
acaggaaatt	ctttggcact	taatcgagat	attcaggagg	aacagactgt	tccagcccga	1500
taccgtcaag	aatttcttac	gattgcatgg	gaacagggtc	atcttcgaag	tatttttccc	1560
ttccaatatt	tttctatttg	agcttctctg	attcctttta	ttgaacataa	tgatgccaat	1620
cgagctttaa	tgagtcttaa	tatgcaacgt	caagcagttc	cgctttctca	gtccgaaaaa	1680
tgcattgttg	gaactggatt	ggaacgccaa	gtagctttag	attcaggggt	ttccgctata	1740
gccgaacacg	agggaaacat	cattttatacc	aatactgaca	ggatattttt	atttggtaat	1800
ggagatactc	taagcattcc	attaactata	tatcaacgtt	ccaacaaaaa	tacttgatg	1860
catcaaaaac	cccagggtcg	ccgaggtaaa	tgtataaaaa	agggacaaat	tttagcggat	1920
gggtgtgcta	cagttgaacg	cgaactcgct	ttgggaaaaa	acgtcttagt	agcttatatg	1980
ccatgggaag	gttacaattc	tgaagatgct	gtactcatta	atgagcgtct	ggtctatgaa	2040
gatattttata	cttcttttca	catacggaaa	tatgaaatcc	agactcatat	gacaagctat	2100
ggttctgaaa	gaatcactaa	taaaattcca	catctagaag	ccattttact	cagaaattta	2160
gacaaaaatg	gaatttgtgat	cctcgggtcg	tgggtagaaa	cgggtgatat	tttagtgggt	2220
aaattaacac	ctcaaatggc	aaaagaatcc	tcgtattccc	ccgaagatag	attattacga	2280
gctatacttg	gcatttcaggt	atccacctca	aaggaaaactt	gtctaaaact	acctacaggc	2340
ggtaggggta	gagttattga	tgtgagatgg	atccaaaaaa	aggggggttc	cagttataat	2400
ccagaaacga	ttcgtatata	tatttttacag	aaacgtgaaa	ttaaagtagg	agataaagtg	2460
gctgggagac	atggaaataa	aggtatcggt	tcaaaaattt	tgtctagaca	ggatatgcct	2520
tatttgcaag	atggaagacc	cgttgatatg	gtcttcaatc	cactaggggt	accttcacga	2580
atgaatgtag	gacaaatatt	tgaatgctcg	ctcgggtagg	caggaggtat	gctagaaaga	2640
cattatcgaa	taacaccttt	tgatgagaga	tatgaacaag	aagcttcgag	aaaactagtg	2700
ttttctgaat	tatatgaagc	cagtaaacaa	acatctaata	catggatatt	tgaacccgag	2760
tatccaggaa	aaagcaaaat	ctttgatgga	agaacaggga	attcttttaa	acagcctgct	2820

-continued

ataatgggaa aaccttatat ttgaaatta attcatcaag ttgatgataa aatacatgga	2880
cgttcacgtg gacattatgc acttggtaca caacaaccac ttagaggaag ggccaagcag	2940
ggaggacaac gggtaggcga aatggagggt tgggccttgg aaggatttgg tgttgctcat	3000
attttacaag agatgcttac ttataaatct gatcatatta aaactcgcca agaagtactc	3060
gggactacga tcattggagg aacaatacct aaacctacag atgctccaga atctttttaga	3120
ttgctagtgc gagaattacg atcttttagct atggaactga atcatttctt tgtatccgag	3180
aagaacttcc ggattcatag gaaggaagct taa	3213

<210> SEQ ID NO 68
 <211> LENGTH: 20
 <212> TYPE: DNA
 <213> ORGANISM: Glycine max

<400> SEQUENCE: 68

tgtctaaaac tacctacagg	20
-----------------------	----

<210> SEQ ID NO 69
 <211> LENGTH: 20
 <212> TYPE: DNA
 <213> ORGANISM: Glycine max

<400> SEQUENCE: 69

agcgggaattt cggctctatac	20
-------------------------	----

<210> SEQ ID NO 70
 <211> LENGTH: 1062
 <212> TYPE: DNA
 <213> ORGANISM: Glycine max

<400> SEQUENCE: 70

atgactgcaa ttttagagag acgcgagagc gaaagcctat ggggtcgctt ctgtaactgg	60
ataaccagca ccgaaaatcg tctttacatt ggatgggttg gtgttttgat gattcctact	120
ttattgacgc caacttctgt atttattatc gcttttattg ctgcccctcc agtagatatt	180
gatggtattc gtgagcctgt tcttgatct ctactttatg gaaacaatat cttttctggt	240
gccattatcc ctacttctgc ggcctataggt ttgcactttt atcctatttg ggaagcggca	300
tctgttgatg aatgggtata caacggcggg ccttatgaac taattgttct acacttctta	360
cttggtgtag cttgctacat ggggcgtgag tgggaactta gtttctgttt gggtatgcgt	420
ccttgggattg ctgttgcata ttcagctcct gttgcagccg ctactgctgt tttcttgatc	480
tatcctattg gtcaggaag cttttcagat ggtatgcctc taggaatttc aggtactttc	540
aattttatga ttgtatttca ggctgagcat aatattctta tgcattccatt tcacatgtta	600
gggtgtagctg gtgtattcgg cggctcccta ttcagtgcata tgcattggctt cttggtaact	660
tctagtttga tcagggaaac cacagaaaat gaatctgcta atgaaggtta cagatttggt	720
caagaggaag aaacctataa tattgtagct gctcatgggt attttggccg attgatcttc	780
caatatgcaa gtttcaacaa ttctcgttct ttacatttct tcttagctgc ttggcctgta	840
gtaggtattt gggttaccgc tttaggtatc agcactatgg ctttcaactt aaatgggttc	900
aatttcaacc aatccgtagt ttagtagtaa ggtcgtgtaa ttaatacctg ggctgatatt	960
attaaccgag ctaaccttgg tatggaagta atgcatgaac gtaatgctca taatttccct	1020

-continued

ctagatctag ctgcatcg cgtccatct attaatgat aa 1062

<210> SEQ ID NO 71
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Glycine max

<400> SEQUENCE: 71

gggtgtagctg gtgtattcgg 20

<210> SEQ ID NO 72
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Glycine max

<400> SEQUENCE: 72

tctagatcta gctgcatcg 20

<210> SEQ ID NO 73
<211> LENGTH: 273
<212> TYPE: DNA
<213> ORGANISM: Glycine max

<400> SEQUENCE: 73

atggtaaaaa attcaattat acctgttatt tcacaagaaa aaaaagaaaa aaaccagga 60

tcgggtgaat ttcaaatatt caaatctacc gatagaatac gaagacttac ttcacatttt 120

gaattgcacc gaaaagacta tttatctcaa agaggtttac gtaaaatttt gggaaaaacga 180

caaagattgc tgtcttattt gtcaaagaaa gatagaatac ggtataaaaa attaataaat 240

cagtttgata ttcgagagtc acaaatcgt taa 273

<210> SEQ ID NO 74
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Glycine max

<400> SEQUENCE: 74

atagaatacg aagacttact 20

<210> SEQ ID NO 75
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Glycine max

<400> SEQUENCE: 75

tgtcaaagaa agatagaata 20

<210> SEQ ID NO 76
<211> LENGTH: 201
<212> TYPE: DNA
<213> ORGANISM: Glycine max

<400> SEQUENCE: 76

atggccaaag gtaaagatat ccgagtaatt gttattttgg aatgtaccgg ttgtgataaa 60

aagagtgtta ataaggaatc aacgggtatt tctagatata taactaaaaa gaatcgacag 120

aatacgcta gtcgattgga attgagaaaa ttttggtccc gttgttgcaa acatacaatt 180

cacgcagaaa taaagaaata g 201

-continued

<210> SEQ ID NO 77
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Glycine max

<400> SEQUENCE: 77

cgttgttgca aacatacaat 20

<210> SEQ ID NO 78
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Glycine max

<400> SEQUENCE: 78

acagaatacg cctagtcgat 20

<210> SEQ ID NO 79
<211> LENGTH: 864
<212> TYPE: DNA
<213> ORGANISM: Nicotiana benthamiana

<400> SEQUENCE: 79

gtgcgacttg aagcacagga tccgtgtggg attgtgacat ccaccatttt atgtaggaat 60
gaaggtgctc ttggtcgcac atcattgggt ctgtttcatt agattagaac ccctcttttt 120
tgttgtcttg gaatgtaaat agtccatgat ggagctcgag tagaaagtat taatttattt 180
ctcggggcaa gagtctaggg ttaatgcaa tcaataaaaa aattggaaca acttcgtaaa 240
tgtattttcg gtatggaaat cgaaagaatc caattcgagc aagtttccaa ttcaaaaatt 300
tcttggaatt gatcaaacct ttctgatcca aagtgtttca cgcggaatc catcgctctgt 360
aggattcttt catagaaatc gcaaaagggg tatgttgctg ccattttgaa aggattaaaa 420
agcaccgaag taatgtctaa acccaatgat ttaaaataaa acaagataa aggatcccag 480
aacaaggaaa cacctttttt attgtcttaa taactggatc gaactgaaga atccaaatcc 540
attttaaacg agacaaacat aaaaggagga aagaccgctc aataaatgaa attgccgaaa 600
gattttcctt tgaactgttt gaaagttatc caacttgagt tatgagagta cgaatggttt 660
ctttttcatt ttcaggaaga aagaagaaaa aaaagactta catctttaat tgatttgatc 720
attttatgga ccagttgtc atttcttaga tagaattcca tacagagata aaacctcgaa 780
tcaatcattt ttctcgagcc gtacgaggag aaagcttcct atacgtttct agggggggtg 840
ttgttcattt acatctatcc caat 864

<210> SEQ ID NO 80
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Nicotiana benthamiana

<400> SEQUENCE: 80

ttgtggattt gtacatccac 20

<210> SEQ ID NO 81
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Nicotiana benthamiana

<400> SEQUENCE: 81

-continued

ttgaactggt tgaaagtatt	20
-----------------------	----

<210> SEQ ID NO 82
<211> LENGTH: 1578
<212> TYPE: DNA
<213> ORGANISM: Nicotiana benthamiana

<400> SEQUENCE: 82

tttcaaatgg aagaaatcca aagatatatta cagccagata gatcgcaaca acacaacttc	60
ctatatccac ttatctttca ggagtatatt tatgcacttg ctcatgatca tggtttaa	120
agaaacaagt cgattttggt ggaatcca ggttataaca ataaatttag tttoctaatt	180
gtgaaacggt taattaccg aatgtatcaa cagaatcatt ttcttatttc tactaatgat	240
tctaacaaaa attcattttt ggggtgcaac aagagtttgt attctcaat gatatcagag	300
ggatttgcgt ttattgtgga aattccgttt tctctacgat taatatcttc tttatcttct	360
ttcgaaggca aaaaggtttt taaatctcat aatttacgat caattcattc aacatttcct	420
tttttagagg acaatttttc acatctaaat tatgtattag atatactaata accctacccc	480
gttcatctgg aaatcttgggt tcaaatctt cgctattggg taaaagatgc ctcttcttta	540
catttattac gattctttct ccatgaatat tggaatttga atagtcttat tacttcaaa	600
aagcccggtt actccttttc aaaaaaaat caaagattct tcttcttctt atataattct	660
tatgtatatg aatgcgaatc cactttcgtc tttctacgga accaatcttc tcatttacga	720
tcaacatctt ttggagccct tcttgaacga atatatctt atggaaaaat agaagctctt	780
gtagaagtct ttgctaagga ttttcagggt accctatgggt tattcaagga tcctttcatg	840
cattatgtta ggtatcaagg aaaatccatt ctggcttcaa aagggaaggt tcttttgatg	900
aataaatgga aattttacct tgtcaatttt tggcaatgtc atttttctct gtgctttcac	960
acaggaagga tccatataaa ccaattatcc aatcattccc gtaactttat gggtatctt	1020
tcaagtgtgc gactaaatcc ttcaatggta cgtagtcaaa tgtagaaaaa ttcatttcta	1080
atcaataatg caattaagaa gtctgatacc cttgttccaa ttattccttt gattggatca	1140
ttagctaaag caaacttttg taccgtatta gggcatccca ttagtaaac ggtttggtcc	1200
gatttatcag attctgatat tattgacgga tttgggcgta tatgcagaaa tctttttcat	1260
tattatagcg gatcttccaa aaaaaagact ttatatcgaa taaagtatat acttcgactt	1320
tcttgtgcta gaacttttagc tcggaacac aaaagtactg tacgcacttt ttgaaaaga	1380
tcgggctcgg aattattgga agaattttta acgtcggaag aacaagttct tcttttgacc	1440
ttcccacgag cttcttctag tttgtgggga gtatatagaa gtcggatttg gtatttgat	1500
attttttgta tcaatgatct ggcgaattat caatgattca ttcttagatt ttctaaatat	1560
aaatttggtt ctaaatga	1578

<210> SEQ ID NO 83
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Nicotiana benthamiana

<400> SEQUENCE: 83

cttgtgctag aacttttagt	20
-----------------------	----

<210> SEQ ID NO 84

-continued

<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: *Nicotiana benthamiana*

<400> SEQUENCE: 84

cgttcatctg gaaatcttg 20

<210> SEQ ID NO 85
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: *Nicotiana tabacum*

<400> SEQUENCE: 85

aagaacttcc cccttgacag 20

<210> SEQ ID NO 86
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: *Nicotiana tabacum*

<400> SEQUENCE: 86

tatacaggat gggtagaaag 20

<210> SEQ ID NO 87
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: *Nicotiana tabacum*

<400> SEQUENCE: 87

atataatttt taataaaggg 20

<210> SEQ ID NO 88
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: *Nicotiana tabacum*

<400> SEQUENCE: 88

ctagtcttcg acacaagaaa 20

<210> SEQ ID NO 89
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: *Glycine max*

<400> SEQUENCE: 89

ataacagaag ttaaagaaga 20

<210> SEQ ID NO 90
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: *Glycine max*

<400> SEQUENCE: 90

atctggaaac catagaacag 20

<210> SEQ ID NO 91
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: *Glycine max*

<400> SEQUENCE: 91

-continued

ctatttcgac acaaacaaga 20

<210> SEQ ID NO 92
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Glycine max

<400> SEQUENCE: 92

ctttctttga cgaattcgag 20

<210> SEQ ID NO 93
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 93

acgagagttg ttgaaactag catattggaa gatcaa 36

<210> SEQ ID NO 94
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 94

acgagagtta ttgaatgtag catactgaaa gatcaa 36

<210> SEQ ID NO 95
<211> LENGTH: 14
<212> TYPE: PRT
<213> ORGANISM: Saccharomyces cerevisiae

<400> SEQUENCE: 95

Met Val Leu Pro Arg Leu Tyr Thr Ala Thr Ser Arg Ala Ala
1 5 10

<210> SEQ ID NO 96
<211> LENGTH: 1384
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 96

Met Val Leu Pro Arg Leu Tyr Thr Ala Thr Ser Arg Ala Ala Leu Ser
1 5 10 15

Thr Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
20 25 30

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
35 40 45

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
50 55 60

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
65 70 75 80

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
85 90 95

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
100 105 110

Phe 115	Phe 130	His 145	Arg 165	Leu 180	Glu 150	Glu 135	Ser 120	Phe 170	Leu 185	Val 155	Glu 125	Asp 140	Lys 175	Lys 160
His 145	Glu 130	Arg 165	His 180	Pro 150	Ile 135	Phe 120	Gly 170	Asn 185	Ile 155	Val 140	Asp 125	Glu 175	Val 190	Tyr 160
His 145	Glu 130	Lys 165	Tyr 180	Pro 150	Thr 135	Ile 120	Tyr 170	His 185	Leu 155	Arg 140	Lys 125	Lys 175	Leu 190	Val 160
Ser 145	Thr 130	Asp 165	Lys 180	Ala 150	Asp 135	Leu 120	Arg 170	Leu 185	Ile 155	Tyr 140	Leu 125	Ala 175	Leu 190	Ala 160
Met 145	Ile 130	Lys 165	Phe 180	Arg 150	Gly 135	His 120	Phe 170	Leu 185	Ile 155	Glu 140	Gly 125	Asp 175	Leu 190	Asn 160
Asp 145	Asn 130	Ser 165	Asp 180	Val 150	Asp 135	Lys 120	Leu 170	Phe 185	Ile 155	Gln 140	Leu 125	Val 175	Gln 190	Thr 160
Asn 145	Gln 130	Leu 165	Phe 180	Glu 150	Glu 135	Asn 120	Pro 170	Ile 185	Asn 155	Ala 140	Ser 125	Gly 175	Val 190	Asp 160
Lys 145	Ala 130	Ile 165	Leu 180	Ser 150	Ala 135	Arg 120	Leu 170	Ser 185	Lys 155	Ser 140	Arg 125	Arg 175	Leu 190	Glu 160
Leu 145	Ile 130	Ala 165	Gln 180	Leu 150	Pro 135	Gly 120	Glu 170	Lys 185	Lys 155	Asn 140	Gly 125	Leu 175	Phe 190	Gly 160
Leu 145	Ile 130	Ala 165	Leu 180	Ser 150	Leu 135	Gly 120	Leu 170	Thr 185	Pro 155	Asn 140	Phe 125	Lys 175	Ser 190	Asn 160
Asp 145	Leu 130	Ala 165	Glu 180	Asp 150	Ala 135	Lys 120	Leu 170	Gln 185	Leu 155	Ser 140	Lys 125	Asp 175	Thr 190	Tyr 160
Asp 145	Asp 130	Leu 165	Asp 180	Asn 150	Leu 135	Leu 120	Ala 170	Gln 185	Ile 155	Gly 140	Asp 125	Gln 175	Tyr 190	Ala 160
Leu 145	Phe 130	Leu 165	Ala 180	Ala 150	Lys 135	Asn 120	Leu 170	Ser 185	Asp 155	Ala 140	Ile 125	Leu 175	Leu 190	Ser 160
Ile 145	Leu 130	Arg 165	Val 180	Asn 150	Thr 135	Glu 120	Ile 170	Thr 185	Lys 155	Ala 140	Pro 125	Leu 175	Ser 190	Ala 160
Met 145	Ile 130	Lys 165	Arg 180	Tyr 150	Asp 135	Glu 120	His 170	His 185	Gln 155	Asp 140	Leu 125	Thr 175	Leu 190	Leu 160
Ala 145	Leu 130	Val 165	Arg 180	Gln 150	Gln 135	Leu 120	Pro 170	Glu 185	Lys 155	Tyr 140	Lys 125	Glu 175	Ile 190	Phe 160
Asp 145	Gln 130	Ser 165	Lys 180	Asn 150	Gly 135	Tyr 120	Ala 170	Gly 185	Tyr 155	Ile 140	Asp 125	Gly 175	Gly 190	Ala 160
Gln 145	Glu 130	Glu 165	Phe 180	Tyr 150	Lys 135	Phe 120	Ile 170	Lys 185	Pro 155	Ile 140	Leu 125	Glu 175	Lys 190	Met 160
Gly 145	Thr 130	Glu 165	Glu 180	Leu 150	Leu 135	Val 120	Lys 170	Leu 185	Asn 155	Arg 140	Glu 125	Asp 175	Leu 190	Leu 160
Lys 145	Gln 130	Arg 165	Thr 180	Phe 150	Asp 135	Asn 120	Gly 170	Ser 185	Ile 155	Pro 140	His 125	Gln 175	Ile 190	His 160
Gly 145	Glu 130	Leu 165	His 180	Ala 150	Ile 135	Leu 120	Arg 170	Arg 185	Gln 155	Glu 140	Asp 125	Phe 175	Tyr 190	Pro 160
Leu 145	Lys 130	Asp 165	Asn 180	Arg 150	Glu 135	Lys 120	Ile 170	Glu 185	Lys 155	Ile 140	Leu 125	Thr 175	Phe 190	Arg 160
Pro 145	Tyr 130	Tyr 165	Val 180	Gly 150	Pro 135	Leu 120	Ala 170	Arg 185	Gly 155	Asn 140	Ser 125	Arg 175	Phe 190	Ala 160
Met 145	Thr 130	Arg 165	Lys 180	Ser 150	Glu 135	Glu 120	Thr 170	Ile 185	Thr 155	Pro 140	Trp 125	Asn 175	Phe 190	Glu 160
Val 145	Val 130	Asp 165	Lys 180	Gly 150	Ala 135	Ser 120	Ala 170	Gln 185	Ser 155	Phe 140	Ile 125	Glu 175	Arg 190	Met 160
Asn 145	Phe 130	Asp 165	Lys 180	Asn 150	Leu 135	Pro 120	Asn 170	Glu 185	Lys 155	Val 140	Leu			

-continued

515	520	525
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys		
530	535	540
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln		
545	550	555
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr		
565	570	575
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp		
580	585	590
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly		
595	600	605
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp		
610	615	620
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr		
625	630	635
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala		
645	650	655
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr		
660	665	670
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp		
675	680	685
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe		
690	695	700
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe		
705	710	715
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu		
725	730	735
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly		
740	745	750
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly		
755	760	765
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln		
770	775	780
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile		
785	790	795
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro		
805	810	815
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu		
820	825	830
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg		
835	840	845
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys		
850	855	860
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg		
865	870	875
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys		
885	890	895
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys		
900	905	910
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp		
915	920	925

-continued

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
 930 935 940
 Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
 945 950 955 960
 Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
 965 970 975
 Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
 980 985 990
 Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
 995 1000 1005
 Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu
 1010 1015 1020
 Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile
 1025 1030 1035
 Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe
 1040 1045 1050
 Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu
 1055 1060 1065
 Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly
 1070 1075 1080
 Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr
 1085 1090 1095
 Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys
 1100 1105 1110
 Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro
 1115 1120 1125
 Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp
 1130 1135 1140
 Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser
 1145 1150 1155
 Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu
 1160 1165 1170
 Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser
 1175 1180 1185
 Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr
 1190 1195 1200
 Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser
 1205 1210 1215
 Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala
 1220 1225 1230
 Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr
 1235 1240 1245
 Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly
 1250 1255 1260
 Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His
 1265 1270 1275
 Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser
 1280 1285 1290
 Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser
 1295 1300 1305

-continued

Ala Tyr 1310	Asn Lys His Arg 1315	Lys Pro Ile Arg 1320	Gln Ala Glu
Asn Ile 1325	Ile His Leu Phe Thr 1330	Leu Thr Asn Leu Gly 1335	Ala Pro Ala
Ala Phe 1340	Lys Tyr Phe Asp Thr 1345	Thr Ile Asp Arg Lys 1350	Arg Tyr Thr
Ser Thr 1355	Lys Glu Val Leu Asp 1360	Ala Thr Leu Ile His 1365	Gln Ser Ile
Thr Gly 1370	Leu Tyr Glu Thr Arg 1375	Ile Asp Leu Ser Gln 1380	Leu Gly Gly

Asp

```
<210> SEQ ID NO 97
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Saccharomyces cerevisiae
```

<400> SEQUENCE: 97

Met Lys Ser Phe Ile Thr Arg Asn Lys Thr Ala Ile
1 5 10

```
<210> SEQ ID NO 98
<211> LENGTH: 1379
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct
```

<400> SEQUENCE: 98

Met 1	Lys	Ser	Phe 5	Ile	Thr	Arg	Asn	Lys 10	Thr	Ala	Ile	Asp	Lys 15	Lys	Tyr
Ser	Ile	Gly	Leu 20	Asp	Ile	Gly	Thr	Asn 25	Ser	Val	Gly	Trp	Ala 30	Val	Ile
Thr	Asp	Glu	Tyr 35	Lys	Val	Pro	Ser 40	Lys	Lys	Phe	Lys 45	Val	Leu	Gly	Asn
Thr	Asp	Arg	His 50	Ser	Ile	Lys 55	Lys	Asn	Leu	Ile	Gly 60	Ala	Leu	Leu	Phe
Asp 65	Ser	Gly	Glu	Thr	Ala 70	Glu	Ala	Thr	Arg	Leu 75	Lys	Arg	Thr	Ala	Arg 80
Arg	Arg	Tyr	Thr 85	Arg	Arg	Lys	Asn	Arg	Ile 90	Cys	Tyr	Leu	Gln 95	Glu	Ile
Phe	Ser	Asn	Glu 100	Met	Ala	Lys	Val	Asp 105	Asp	Ser	Phe	Phe	His 110	Arg	Leu
Glu	Glu	Ser	Phe 115	Leu	Val	Glu	Glu	Asp 120	Lys	Lys	His	Glu 125	Arg	His	Pro
Ile 130	Phe	Gly	Asn	Ile	Val	Asp 135	Glu	Val	Ala	Tyr	His 140	Glu	Lys	Tyr	Pro
Thr 145	Ile	Tyr	His	Leu 150	Arg	Lys	Lys	Leu	Val	Asp 155	Ser	Thr	Asp	Lys	Ala 160
Asp	Leu	Arg	Leu 165	Ile	Tyr	Leu	Ala	Leu	Ala 170	His	Met	Ile	Lys	Phe	Arg 175
Gly	His	Phe	Leu 180	Ile	Glu	Gly	Asp	Leu 185	Asn	Pro	Asp	Asn 190	Ser	Asp	Val
Asp	Lys	Leu	Phe 195	Ile	Gln	Leu	Val	Gln 200	Thr	Tyr	Asn	Gln 205	Leu	Phe	Glu

Glu 210	Asn	Pro	Ile	Asn	Ala	Ser	Gly	Val	Asp	Ala	Lys	Ala	Ile	Leu	Ser
Ala 225	Arg	Leu	Ser	Lys	Ser	Arg	Arg	Leu	Glu	Asn	Leu	Ile	Ala	Gln	Leu
Pro	Gly	Glu	Lys	Lys	Asn	Gly	Leu	Phe	Gly	Asn	Leu	Ile	Ala	Leu	Ser
Leu	Gly	Leu	Thr	Pro	Asn	Phe	Lys	Ser	Asn	Phe	Asp	Leu	Ala	Glu	Asp
Ala	Lys	Leu	Gln	Leu	Ser	Lys	Asp	Thr	Tyr	Asp	Asp	Asp	Leu	Asp	Asn
Leu	Leu	Ala	Gln	Ile	Gly	Asp	Gln	Tyr	Ala	Asp	Leu	Phe	Leu	Ala	Ala
Lys 305	Asn	Leu	Ser	Asp	Ala	Ile	Leu	Leu	Ser	Asp	Ile	Leu	Arg	Val	Asn
Thr	Glu	Ile	Thr	Lys	Ala	Pro	Leu	Ser	Ala	Ser	Met	Ile	Lys	Arg	Tyr
Asp	Glu	His	His	Gln	Asp	Leu	Thr	Leu	Leu	Lys	Ala	Leu	Val	Arg	Gln
Gln	Leu	Pro	Glu	Lys	Tyr	Lys	Glu	Ile	Phe	Phe	Asp	Gln	Ser	Lys	Asn
Gly	Tyr	Ala	Gly	Tyr	Ile	Asp	Gly	Gly	Ala	Ser	Gln	Glu	Glu	Phe	Tyr
Lys 385	Phe	Ile	Lys	Pro	Ile	Leu	Glu	Lys	Met	Asp	Gly	Thr	Glu	Glu	Leu
Leu	Val	Lys	Leu	Asn	Arg	Glu	Asp	Leu	Leu	Arg	Lys	Gln	Arg	Thr	Phe
Asp	Asn	Gly	Ser	Ile	Pro	His	Gln	Ile	His	Leu	Gly	Glu	Leu	His	Ala
Ile	Leu	Arg	Arg	Gln	Glu	Asp	Phe	Tyr	Pro	Phe	Leu	Lys	Asp	Asn	Arg
Glu	Lys	Ile	Glu	Lys	Ile	Leu	Thr	Phe	Arg	Ile	Pro	Tyr	Tyr	Val	Gly
Pro 465	Leu	Ala	Arg	Gly	Asn	Ser	Arg	Phe	Ala	Trp	Met	Thr	Arg	Lys	Ser
Glu	Glu	Thr	Ile	Thr	Pro	Trp	Asn	Phe	Glu	Glu	Val	Val	Asp	Lys	Gly
Ala	Ser	Ala	Gln	Ser	Phe	Ile	Glu	Arg	Met	Thr	Asn	Phe	Asp	Lys	Asn
Leu	Pro	Asn	Glu	Lys	Val	Leu	Pro	Lys	His	Ser	Leu	Leu	Tyr	Glu	Tyr
Phe	Thr	Val	Tyr	Asn	Glu	Leu	Thr	Lys	Val	Lys	Tyr	Val	Thr	Glu	Gly
Met 545	Arg	Lys	Pro	Ala	Phe	Leu	Ser	Gly	Glu	Gln	Lys	Lys	Ala	Ile	Val
Asp	Leu	Leu	Phe	Lys	Thr	Asn	Arg	Lys	Val	Thr	Val	Lys	Gln	Leu	Lys
Glu	Asp	Tyr	Phe	Lys	Lys	Ile	Glu	Cys	Phe	Asp	Ser	Val	Glu	Ile	Ser
Gly	Val	Glu	Asp	Arg	Phe	Asn	Ala	Ser	Leu	Gly	Thr	Tyr	His	Asp	Leu

Leu 610	Lys	Ile	Ile	Lys	Asp	Lys 615	Asp	Phe	Leu	Asp	Asn 620	Glu	Glu	Asn	Glu
Asp 625	Ile	Leu	Glu	Asp	Ile 630	Val	Leu	Thr	Leu	Thr 635	Leu	Phe	Glu	Asp	Arg 640
Glu	Met	Ile	Glu	Glu 645	Arg	Leu	Lys	Thr	Tyr 650	Ala	His	Leu	Phe	Asp 655	Asp
Lys	Val	Met	Lys 660	Gln	Leu	Lys	Arg	Arg 665	Arg	Tyr	Thr	Gly	Trp 670	Gly	Arg
Leu	Ser	Arg	Lys	Leu	Ile	Asn 680	Gly	Ile	Arg	Asp	Lys 685	Gln	Ser	Gly	Lys
Thr	Ile 690	Leu	Asp	Phe	Leu	Lys 695	Ser	Asp	Gly	Phe	Ala 700	Asn	Arg	Asn	Phe
Met 705	Gln	Leu	Ile	His	Asp 710	Asp	Ser	Leu	Thr	Phe 715	Lys	Glu	Asp	Ile	Gln 720
Lys	Ala	Gln	Val	Ser 725	Gly	Gln	Gly	Asp	Ser 730	Leu	His	Glu	His	Ile 735	Ala
Asn	Leu	Ala	Gly 740	Ser	Pro	Ala	Ile	Lys 745	Lys	Gly	Ile	Leu	Gln 750	Thr	Val
Lys	Val 755	Val	Asp	Glu	Leu	Val	Lys 760	Val	Met	Gly	Arg 765	His	Lys	Pro	Glu
Asn 770	Ile	Val	Ile	Glu	Met	Ala 775	Arg	Glu	Asn	Gln	Thr 780	Thr	Gln	Lys	Gly
Gln 785	Lys	Asn	Ser	Arg	Glu 790	Arg	Met	Lys	Arg	Ile 795	Glu	Glu	Gly	Ile	Lys 800
Glu	Leu	Gly	Ser	Gln 805	Ile	Leu	Lys	Glu	His 810	Pro	Val	Glu	Asn	Thr 815	Gln
Leu	Gln	Asn	Glu 820	Lys	Leu	Tyr	Leu	Tyr 825	Tyr	Leu	Gln	Asn	Gly 830	Arg	Asp
Met	Tyr 835	Val	Asp	Gln	Glu	Leu	Asp 840	Ile	Asn	Arg	Leu 845	Ser	Asp	Tyr	Asp
Val 850	Asp	His	Ile	Val	Pro	Gln 855	Ser	Phe	Leu	Lys	Asp 860	Asp	Ser	Ile	Asp
Asn 865	Lys	Val	Leu	Thr	Arg 870	Ser	Asp	Lys	Asn	Arg 875	Gly	Lys	Ser	Asp	Asn 880
Val	Pro	Ser	Glu 885	Glu	Val	Val	Lys	Lys	Met 890	Lys	Asn	Tyr	Trp	Arg 895	Gln
Leu	Leu	Asn	Ala 900	Lys	Leu	Ile	Thr	Gln 905	Arg	Lys	Phe	Asp	Asn	Leu	Thr
Lys	Ala 915	Glu	Arg	Gly	Gly	Leu	Ser 920	Glu	Leu	Asp	Lys 925	Ala	Gly	Phe	Ile
Lys 930	Arg	Gln	Leu	Val	Glu	Thr 935	Arg	Gln	Ile	Thr	Lys 940	His	Val	Ala	Gln
Ile 945	Leu	Asp	Ser	Arg	Met 950	Asn	Thr	Lys	Tyr	Asp 955	Glu	Asn	Asp	Lys	Leu 960
Ile	Arg	Glu	Val 965	Lys	Val	Ile	Thr	Leu	Lys 970	Ser	Lys	Leu	Val	Ser	Asp 975
Phe	Arg	Lys	Asp 980	Phe	Gln	Phe	Tyr	Lys 985	Val	Arg	Glu	Ile	Asn 990	Asn	Tyr
His	His 995	Ala	His	Asp	Ala	Tyr	Leu	Asn 1000	Ala	Val	Val	Gly	Thr 1005	Ala	Leu
Ile	Lys	Lys	Tyr	Pro	Lys	Leu	Glu	Ser	Glu	Phe	Val	Tyr	Gly	Asp	

-continued

1010	1015	1020
Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln		
1025	1030	1035
Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile		
1040	1045	1050
Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile		
1055	1060	1065
Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile		
1070	1075	1080
Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu		
1085	1090	1095
Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr		
1100	1105	1110
Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp		
1115	1120	1125
Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly		
1130	1135	1140
Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala		
1145	1150	1155
Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu		
1160	1165	1170
Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn		
1175	1180	1185
Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys		
1190	1195	1200
Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu		
1205	1210	1215
Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys		
1220	1225	1230
Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr		
1235	1240	1245
Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn		
1250	1255	1260
Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp		
1265	1270	1275
Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu		
1280	1285	1290
Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His		
1295	1300	1305
Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu		
1310	1315	1320
Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe		
1325	1330	1335
Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val		
1340	1345	1350
Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu		
1355	1360	1365
Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp		
1370	1375	

-continued

<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 99

aaagagctcg gatcctctat gtattaatag aatctatag 39

<210> SEQ ID NO 100
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 100

aaaccatggt agattaatat tattaaattt aag 33

<210> SEQ ID NO 101
<211> LENGTH: 48
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 101

aaaccatggg tataatctct tctttaatt taagtaaaaa aactacac 48

<210> SEQ ID NO 102
<211> LENGTH: 80
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 102

aaaagagctc acatatacct tgggtgacac gagtatataa gtcattgtat actgttgaat 60
gggagaccac aacgggttcc 80

<210> SEQ ID NO 103
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 103

ttttatgcat atgtatatct ccttcttaaa gttaaac 37

<210> SEQ ID NO 104
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 104

aaaagagctc gctccccgc cgctggt 27

<210> SEQ ID NO 105
<211> LENGTH: 43
<212> TYPE: DNA

-continued

```

<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 105

ctgaatctgg gggaacactt tcccagaaat atagtcatcc ctg          43

<210> SEQ ID NO 106
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 106

gggatgacta ttttctggg aaagtgttcc cccagattca ga          42

<210> SEQ ID NO 107
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 107

ttttatgcat agagttttct tgccccctat ttg          33

<210> SEQ ID NO 108
<211> LENGTH: 3537
<212> TYPE: DNA
<213> ORGANISM: Bacillus thuringiensis

<400> SEQUENCE: 108

atggataaca atccgaacat caatgaatgc attccttata attgtttaag taaccctgaa      60
gtagaagtat taggtggaga aagaatagaa actggttaca cccaatcga ttttccttg      120
tcgctaacgc aatttccttt gagtgaattt gttcccggtg ctggatttgt gttaggacta      180
gttgatataa tatggggaat ttttgggtccc tctcaatggg acgcatttct tgtacaaatt      240
gaacagttaa ttaaccaaag aatagaagaa ttcgctagga accaagccat ttctagatta      300
gaaggactaa gcaatcttta tcaaatctac gcagaatctt ttagagagtg ggaagcagat      360
cctactaatc cagcattaag agaagagatg cgtattcaat tcaatgacat gaacagtgcc      420
cttacaacgc ctattcctct ttttgcagtt caaaattatc aagttcctct tttatcagta      480
tatgttcaag ctgcaaatat acatttatca gttttgagag atgtttcagt gtttgacaa      540
aggtggggat ttgatgccgc gactatcaat agtcgttata atgatttaac taggcttatt      600
ggcaactata cagattatgc tgtacgctgg tacaatacgg gattagaacg tgtatgggga      660
ccggattcta gagattgggt aaggtataat caatttagaa gagaattaac actaactgta      720
ttagatatcg ttgctctgtt cccgaattat gatagtagaa gatatccaat tcgaacagtt      780
tcccaattaa caagagaaat ttatacaaac ccagtattag aaaattttga tggtagtttt      840
cgaggctcgg ctcagggcat agaaagaagt attaggagtc cacatttgat ggatatactt      900
aacagtataa ccacttatac ggatgctcat aggggttatt attattggtc agggcatcaa      960
ataatggctt ctcctgtcgg ttttccgggg ccagaattca cgtttccgct atatggaacc     1020
atgggaaatg cagctccaca acaacgtatt gttgctcaac taggtcaggg cgtgtataga     1080

```

-continued

acattatcgt ccactttata tagaagacct tttaatatag ggataaataa tcaacaacta	1140
tctgttcttg acgggacaga atttgcttat ggaacctcct caaatttgcc atccgctgta	1200
tacagaaaaa gcggaacggt agattcgctg gatgaaatac cgccacagaa taacaacgtg	1260
ccacctaggc aaggatttag tcatcgatta agccatgttt caatgtttcg ttcaggcttt	1320
agtaatagta gtgtaagtat aataagagct cctatgttct cttggatata tcgtagtgct	1380
gaatttaata atataattgc atcggatagt attactcaaa tccctgcagt gaagggaac	1440
tttcttttta atggttctgt aatttcagga ccaggattta ctggtgggga cttagttaga	1500
ttaaatagta gtggaaataa cattcagaat agaggggata ttgaagtcc aattcacttc	1560
ccatcgacat ctaccagata tcgagttcgt gtacggtatg cttctgtaac cccgattcac	1620
ctcaacgtta attggggtaa ttcattccatt ttttccaata cagtaccagc tacagctacg	1680
tcattagata atctacaatc aagtgtttt ggttattttt aaagtgccaa tgcttttaca	1740
tcttcattag gtaatatagt agtggttaga aatttttagt ggactgcagg agtgataata	1800
gacagatttg aatttattcc agttactgca acactcgagg ctgaatataa tctggaaaga	1860
gcgcagaagg cgggtgaatgc gctgtttacg tctacaaacc aactagggct aaaaacaaat	1920
gtaacggatt atcatattga tcaagtgtcc aatttagtta cgtatttate ggatgaattt	1980
tgtctggatg aaaagcgaga attgtccgag aaagtcaaac atgcgaagcg actcagtgat	2040
gaacgcaatt tactccaaga ttcaaatttc aaagacatta ataggcaacc agaacgtggg	2100
tggggcggaa gtacagggat taccatccaa ggaggggatg acgtatttaa agaaaattac	2160
gtcacactat caggtacctt tgatgagtgc tatccaacat atttgatca aaaaatcgat	2220
gaatcaaaat taaaagcctt tacccgttat caattaagag ggatatatga agatagtcaa	2280
gacttagaaa tctatttaat tcgctacaat gcaaaacatg aaacagtaaa tgtgccaggt	2340
acgggttcct tatggccgct ttcagcccaa agtccaatcg gaaagtgtgg agagccgaat	2400
cgatgcgcgc cacacctga atggaatcct gacttagatt gttcgtgtag ggatggagaa	2460
aagtgtgccc atcattcgca tcatctctcc ttagacattg atgtaggatg tacagactta	2520
aatgaggacc taggtgtatg ggtgatcttt aagattaaga cgcaagatgg gcacgcaaga	2580
ctagggaatc tagagtttct cgaagagaaa ccattagtag gagaagcgct agctcgtgtg	2640
aaaagagcgg agaaaaaatg gagagacaaa cgtgaaaaat tggaatggga aacaaatctc	2700
gtttataaag aggcaaaaga atctgtagat gctttatttg taaactctca atatgatcaa	2760
ttacaagcgg atacgaatat tgccatgatt catgcggcag ataaacgtgt tcatagcatt	2820
cgagaagctt atctgcctga gctgtctgtg attccgggtg tcaatgcggc tatttttgaa	2880
gaattagaag ggcgtatttt cactgcattc tccctatatg atgcgagaaa tgtcattaaa	2940
aatggtgatt ttaataatgg cttatcctgc tggaaacgtga aagggcattg agatgtagaa	3000
gaacaaaaca accaacgttc ggtccttggt gttccggaat gggaagcaga agtgtcacaa	3060
gaagtctctg tctgtccggg tcgtggctat atccttcctg tcacagcgta caaggaggga	3120
tatggagaag gttgcgtaac cattcatgag atcgagaaca atacagacga actgaagttt	3180
agcaactcgc tagaagagga aatctatcca aataacacgg taacgtgtaa tgattatact	3240
gtaaatacaag aagaatacgg aggtgcgtac acttctcgtg atcgaggata taacgaagct	3300
ccttcctgac cagctgatta tgcgtcagtc tatgaagaaa aatcgtatac agatggacga	3360

-continued

```

agagagaatc cttgtgaatt taacagaggg tatagggatt acacgccact accagttggt 3420
tatgtgacaa aagaattaga atacttocca gaaaccgata aggtatggat tgagattgga 3480
gaaacggaag gaacatttat cgtggacagc gtggaattac tccttatgga ggaatag 3537

```

```

<210> SEQ ID NO 109
<211> LENGTH: 1178
<212> TYPE: PRT
<213> ORGANISM: Bacillus thuringiensis

```

```

<400> SEQUENCE: 109

```

```

Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu
1           5           10           15
Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly
20           25           30
Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser
35           40           45
Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile
50           55           60
Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile
65           70           75           80
Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala
85           90           95
Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu
100          105          110
Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu
115          120          125
Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala
130          135          140
Ile Pro Leu Phe Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val
145          150          155          160
Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser
165          170          175
Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg
180          185          190
Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp Tyr Ala Val
195          200          205
Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg
210          215          220
Asp Trp Val Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val
225          230          235          240
Leu Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ser Arg Arg Tyr Pro
245          250          255
Ile Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val
260          265          270
Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Ser Ala Gln Gly Ile Glu
275          280          285
Arg Ser Ile Arg Ser Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr
290          295          300
Ile Tyr Thr Asp Ala His Arg Gly Tyr Tyr Tyr Trp Ser Gly His Gln
305          310          315          320
Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro
325          330          335

```

-continued

Leu	Tyr	Gly	Thr	Met	Gly	Asn	Ala	Ala	Pro	Gln	Gln	Arg	Ile	Val	Ala	
			340					345					350			
Gln	Leu	Gly	Gln	Gly	Val	Tyr	Arg	Thr	Leu	Ser	Ser	Thr	Leu	Tyr	Arg	
		355					360					365				
Arg	Pro	Phe	Asn	Ile	Gly	Ile	Asn	Asn	Gln	Gln	Leu	Ser	Val	Leu	Asp	
		370				375						380				
Gly	Thr	Glu	Phe	Ala	Tyr	Gly	Thr	Ser	Ser	Asn	Leu	Pro	Ser	Ala	Val	
385					390					395					400	
Tyr	Arg	Lys	Ser	Gly	Thr	Val	Asp	Ser	Leu	Asp	Glu	Ile	Pro	Pro	Gln	
			405						410					415		
Asn	Asn	Asn	Val	Pro	Pro	Arg	Gln	Gly	Phe	Ser	His	Arg	Leu	Ser	His	
			420					425					430			
Val	Ser	Met	Phe	Arg	Ser	Gly	Phe	Ser	Asn	Ser	Ser	Val	Ser	Ile	Ile	
		435					440					445				
Arg	Ala	Pro	Met	Phe	Ser	Trp	Ile	His	Arg	Ser	Ala	Glu	Phe	Asn	Asn	
		450				455					460					
Ile	Ile	Ala	Ser	Asp	Ser	Ile	Thr	Gln	Ile	Pro	Ala	Val	Lys	Gly	Asn	
465					470					475					480	
Phe	Leu	Phe	Asn	Gly	Ser	Val	Ile	Ser	Gly	Pro	Gly	Phe	Thr	Gly	Gly	
			485						490					495		
Asp	Leu	Val	Arg	Leu	Asn	Ser	Ser	Gly	Asn	Asn	Ile	Gln	Asn	Arg	Gly	
			500					505					510			
Tyr	Ile	Glu	Val	Pro	Ile	His	Phe	Pro	Ser	Thr	Ser	Thr	Arg	Tyr	Arg	
		515					520					525				
Val	Arg	Val	Arg	Tyr	Ala	Ser	Val	Thr	Pro	Ile	His	Leu	Asn	Val	Asn	
		530				535					540					
Trp	Gly	Asn	Ser	Ser	Ile	Phe	Ser	Asn	Thr	Val	Pro	Ala	Thr	Ala	Thr	
545					550					555					560	
Ser	Leu	Asp	Asn	Leu	Gln	Ser	Ser	Asp	Phe	Gly	Tyr	Phe	Glu	Ser	Ala	
			565						570					575		
Asn	Ala	Phe	Thr	Ser	Ser	Leu	Gly	Asn	Ile	Val	Gly	Val	Arg	Asn	Phe	
			580					585					590			
Ser	Gly	Thr	Ala	Gly	Val	Ile	Ile	Asp	Arg	Phe	Glu	Phe	Ile	Pro	Val	
		595				600					605					
Thr	Ala	Thr	Leu	Glu	Ala	Glu	Tyr	Asn	Leu	Glu	Arg	Ala	Gln	Lys	Ala	
		610				615					620					
Val	Asn	Ala	Leu	Phe	Thr	Ser	Thr	Asn	Gln	Leu	Gly	Leu	Lys	Thr	Asn	
625					630					635					640	
Val	Thr	Asp	Tyr	His	Ile	Asp	Gln	Val	Ser	Asn	Leu	Val	Thr	Tyr	Leu	
			645						650					655		
Ser	Asp	Glu	Phe	Cys	Leu	Asp	Glu	Lys	Arg	Glu	Leu	Ser	Glu	Lys	Val	
		660						665					670			
Lys	His	Ala	Lys	Arg	Leu	Ser	Asp	Glu	Arg	Asn	Leu	Leu	Gln	Asp	Ser	
		675					680						685			
Asn	Phe	Lys	Asp	Ile	Asn	Arg	Gln	Pro	Glu	Arg	Gly	Trp	Gly	Gly	Ser	
		690				695					700					
Thr	Gly	Ile	Thr	Ile	Gln	Gly	Gly	Asp	Asp	Val	Phe	Lys	Glu	Asn	Tyr	
705					710					715					720	
Val	Thr	Leu	Ser	Gly	Thr	Phe	Asp	Glu	Cys	Tyr	Pro	Thr	Tyr	Leu	Tyr	
				725					730						735	

-continued

Gln	Lys	Ile	Asp	Glu	Ser	Lys	Leu	Lys	Ala	Phe	Thr	Arg	Tyr	Gln	Leu	740	745	750
Arg	Gly	Tyr	Ile	Glu	Asp	Ser	Gln	Asp	Leu	Glu	Ile	Tyr	Leu	Ile	Arg	755	760	765
Tyr	Asn	Ala	Lys	His	Glu	Thr	Val	Asn	Val	Pro	Gly	Thr	Gly	Ser	Leu	770	775	780
Trp	Pro	Leu	Ser	Ala	Gln	Ser	Pro	Ile	Gly	Lys	Cys	Gly	Glu	Pro	Asn	785	790	800
Arg	Cys	Ala	Pro	His	Leu	Glu	Trp	Asn	Pro	Asp	Leu	Asp	Cys	Ser	Cys	805	810	815
Arg	Asp	Gly	Glu	Lys	Cys	Ala	His	His	Ser	His	His	Phe	Ser	Leu	Asp	820	825	830
Ile	Asp	Val	Gly	Cys	Thr	Asp	Leu	Asn	Glu	Asp	Leu	Gly	Val	Trp	Val	835	840	845
Ile	Phe	Lys	Ile	Lys	Thr	Gln	Asp	Gly	His	Ala	Arg	Leu	Gly	Asn	Leu	850	855	860
Glu	Phe	Leu	Glu	Glu	Lys	Pro	Leu	Val	Gly	Glu	Ala	Leu	Ala	Arg	Val	865	870	875
Lys	Arg	Ala	Glu	Lys	Lys	Trp	Arg	Asp	Lys	Arg	Glu	Lys	Leu	Glu	Trp	885	890	895
Glu	Thr	Asn	Ile	Val	Tyr	Lys	Glu	Ala	Lys	Glu	Ser	Val	Asp	Ala	Leu	900	905	910
Phe	Val	Asn	Ser	Gln	Tyr	Asp	Gln	Leu	Gln	Ala	Asp	Thr	Asn	Ile	Ala	915	920	925
Met	Ile	His	Ala	Ala	Asp	Lys	Arg	Val	His	Ser	Ile	Arg	Glu	Ala	Tyr	930	935	940
Leu	Pro	Glu	Leu	Ser	Val	Ile	Pro	Gly	Val	Asn	Ala	Ala	Ile	Phe	Glu	945	950	955
Glu	Leu	Glu	Gly	Arg	Ile	Phe	Thr	Ala	Phe	Ser	Leu	Tyr	Asp	Ala	Arg	965	970	975
Asn	Val	Ile	Lys	Asn	Gly	Asp	Phe	Asn	Asn	Gly	Leu	Ser	Cys	Trp	Asn	980	985	990
Val	Lys	Gly	His	Val	Asp	Val	Glu	Glu	Gln	Asn	Asn	Gln	Arg	Ser	Val	995	1000	1005
Leu	Val	Val	Pro	Glu	Trp	Glu	Ala	Glu	Val	Ser	Gln	Glu	Val	Arg		1010	1015	1020
Val	Cys	Pro	Gly	Arg	Gly	Tyr	Ile	Leu	Arg	Val	Thr	Ala	Tyr	Lys		1025	1030	1035
Glu	Gly	Tyr	Gly	Glu	Gly	Cys	Val	Thr	Ile	His	Glu	Ile	Glu	Asn		1040	1045	1050
Asn	Thr	Asp	Glu	Leu	Lys	Phe	Ser	Asn	Cys	Val	Glu	Glu	Glu	Ile		1055	1060	1065
Tyr	Pro	Asn	Asn	Thr	Val	Thr	Cys	Asn	Asp	Tyr	Thr	Val	Asn	Gln		1070	1075	1080
Glu	Glu	Tyr	Gly	Gly	Ala	Tyr	Thr	Ser	Arg	Asn	Arg	Gly	Tyr	Asn		1085	1090	1095
Glu	Ala	Pro	Ser	Val	Pro	Ala	Asp	Tyr	Ala	Ser	Val	Tyr	Glu	Glu		1100	1105	1110
Lys	Ser	Tyr	Thr	Asp	Gly	Arg	Arg	Glu	Asn	Pro	Cys	Glu	Phe	Asn		1115	1120	1125
Arg	Gly	Tyr	Arg	Asp	Tyr	Thr	Pro	Leu	Pro	Val	Gly	Tyr	Val	Thr				

-continued

1130	1135	1140
Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp Lys Val Trp Ile Glu		
1145	1150	1155
Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp Ser Val Glu Leu		
1160	1165	1170
Leu Leu Met Glu Glu		
1175		
<210> SEQ ID NO 110 <211> LENGTH: 1848 <212> TYPE: DNA <213> ORGANISM: Bacillus thuringiensis <400> SEQUENCE: 110		
atggataaca atccgaacat caatgaatgc attccttata attgtttaag taaccctgaa	60	
gtagaagtat taggtggaga aagaatagaa actggttaca cccaatcga tatttccttg	120	
tcgctaacgc aatttcctttt gagtgaattt gttcccggtg ctggatttgt gttaggacta	180	
gttgatataa tatggggaat ttttggtccc tctcaatggg acgcatttct tgtacaaatt	240	
gaacagttaa ttaaccaaag aatagaagaa ttcgctagga accaagccat ttctagatta	300	
gaaggactaa gcaatcttta tcaaatctac gcagaatctt ttagagagtg ggaagcagat	360	
cctactaatc cagcattaag agaagagatg cgtattcaat tcaatgacat gaacagtgcc	420	
cttacaacgc ctattcctct ttttgagttt caaaattatc aagttcctct tttatcagta	480	
tatgttcaag ctgcaaatct acatttatca gttttgagag atgtttcagt gtttggacaa	540	
aggtggggat ttgatgcgcg gactatcaat agtcgttata atgatttaac taggcttatt	600	
ggcaactata cagattatgc tgtacgctgg tacaatacgg gattagaacg tgtatgggga	660	
cgggattcta gagattgggt aagggtataat caatttagaa gagaattaac actaactgta	720	
ttagatatcg ttgctctggt ccgaattat gatagtagaa gatatccaat tcgaacagtt	780	
tcccaattaa caagagaaat ttatacaaac ccagtattag aaaattttga tggtagtttt	840	
cgaggctcgg ctcagggcct agaagaagt attaggagtc cacatttgat ggatatactt	900	
aacagtataa ccactctatac ggatgctcat aggggttatt attattggtc agggcatcaa	960	
ataatggctt ctcctgtcgg tttttcgggg ccagaattca cgtttccgct atatggaacc	1020	
atgggaaatg cagctccaca acaacgtatt gttgctcaac taggtcaggg cgtgtataga	1080	
acattatcgt ccactttata tagaagacct tttaatatag ggataaataa tcaacaacta	1140	
tctgttcttg acgggacaga atttgcttat ggaacctcct caaatttgcc atccgctgta	1200	
tacagaaaaa gcggaaacgt agattcgtg gatgaaatac cgccacagaa taacaacgtg	1260	
ccacctaggg aaggatttag tcatcgatta agccatgttt caatgtttcg ttcaggcttt	1320	
agtaatagta gtgtaagtat aataagagct cctatgttct cttggataga tcgtagtgtc	1380	
gaatttaata atataattgc atcggatagt attactcaaa tccctgcagt gaagggaaac	1440	
tttcttttta atgggttctgt aatttcagga ccaggattta ctggtgggga cttagttaga	1500	
ttaaatagta gtggaaataa cattcagaat agaggggata ttgaagttcc aattcacttc	1560	
ccatcgacat ctaccagata tcgagttcgt gtacggtatg cttctgtaac cccgattcac	1620	
ctcaacgtta attggggtaa ttcattccatt ttttccaata cagtaccagc tacagctacg	1680	
tcattagata atctacaatc aagtgatattt gggtattttg aaagtgccaa tgcttttaca	1740	

-continued

tcttcattag gtaatatagt aggtgttaga aatttttagtg ggactgcagg agtgataata 1800

gacagatttg aatttattcc agttactgca acactcgagg ctgaatag 1848

<210> SEQ ID NO 111

<211> LENGTH: 750

<212> TYPE: DNA

<213> ORGANISM: Bacillus thuringiensis

<400> SEQUENCE: 111

atggaaaatt taaatcattg tccattagaa gatataaagg taaatccatg gaaaaccctt 60

caatcaacag caagggttat tacattacgt gttgaggatc caaatgaaat caataatctt 120

ctttctatta acgaaattga taatccgaat tatatattgc aagcaattat gtttagcaaat 180

gcatttcaaa atgcattagt tcccacttct acagattttg gtgatgccct acgcttttagt 240

atgccaaaag gtttagaaat cgcaaacaca attacaccga tgggtgctgt agtgagtatt 300

gttgatcaaa atgtaactca aacgaataac caagtaagtg ttatgattaa taaagtctta 360

gaagtgttaa aaactgtatt aggagttgca ttaagtggat ctgtaataga tcaattaact 420

gcagcagtta caaatacgtt taaaaattta aatactcaaa aaaatgaagc atggattttc 480

tggggcaagg aaactgctaa tcaaacaaat tacacataca atgtcctggt tgcaatccaa 540

aatgccccaa ctggtggcgt tatgtattgt gtaccagttg gttttgaaat taaagtatca 600

gcagtaaagg aacaagtttt atttttcaca attcaagatt ctgcgagcta caatgttaac 660

atccaatctt tgaaatttgc acaaccatta gttagctcaa gtcagtatcc aattgcagat 720

cttactagcg ctattaatgg aaccctctaa 750

<210> SEQ ID NO 112

<211> LENGTH: 549

<212> TYPE: DNA

<213> ORGANISM: Bacillus thuringiensis

<400> SEQUENCE: 112

atgacagaaa atggagtggt ttataaaata ttcacaacag aaaataataa tttttgtata 60

aatcctactt tgttagaag ggttttttaa aataatttag atgaatttga tttttcgcta 120

gtaaaaaaaa acttagaaca tgagaagaat tgtgtgatta cttctacaat gaatcaaaca 180

atttctttcg agaatatgaa tagtacagaa atggggcata agacatatc ttttttaaat 240

caaacagtat taaataataa ggggaattct tcttttagagg aacaagtctc taatattttt 300

tatagatgtg tatatatgga agttggaaaa tcaagttcat atattaaacc tcttgagcag 360

gattctaata aaataaggta tgtttgtagt ttgctcttta tagtgccta taagaataac 420

ataacatcaa ttattccagt aaatttacia ctaacattat tatcgaaaaa tgtaaaacaa 480

tcctcttcta caaatatatt ttcaggagat atacatttta atatggtaac aatgacttat 540

ttaacttaa 549

<210> SEQ ID NO 113

<211> LENGTH: 540

<212> TYPE: DNA

<213> ORGANISM: Bacillus thuringiensis

<400> SEQUENCE: 113

atgaatatga attttgattt cgaggatcat gaaaataaga atttatctgt gcaggaggaa 60

-continued

catcaccatt gtagtgaagg aggggaacat aaaatagcat tttgttgtgt agtctcaatt	120
ccaaaaggtt ttaaataatgt tgccattgt gatccgaaat ttgtatataa ccttgattgt	180
ctatccgttt caaaagaaaa atgccgaag gttgttccta tagaaggatg tggatgtgca	240
gaggtagatt tacatgtatt aaaggtaaag ggatgcacat catttgtatc gaatatagaa	300
atagaaccta ttcatgaatg catgacctgc tcagcaaacc cacataaaga aaacattgct	360
gtgagttgcc aagatactgt ctgcgtagat caagttttgt attgcagtgt agattgtttg	420
ccagattgtg atattaattg tgataatgta aaaatttgcg atgtgagcat tgaaccaatt	480
ggagattgtg attgtcacgc ggtgaaaatt aaagggaat tttcacttca ctataaataa	540

<210> SEQ ID NO 114

<211> LENGTH: 2155

<212> TYPE: DNA

<213> ORGANISM: *Heterodera glycines*

<400> SEQUENCE: 114

tattcttggg tctgcaacta acaaatccca aagaattttt ccggtagaaa catgttttgg	60
acacgttggg gaggaatcaa aatatgttgc tgtgatccaa ataagaaaac aaatttatat	120
caaacattta gatccaaatt caatcatttt ctaaacaaca gctgacaaat aattgaattc	180
atcaagaagt ttccatcggg tttctgttgt cgataggccc aatttgactc aacagcgctc	240
caatcgcttc gacatttaaa atttgcgtgc cacaataccc aggctgaata gttacaacag	300
ttctgttagt atttattagc tttctaaaaa tgaacttgcc tccccattgg tgccatccac	360
ccatttgagc accttaaaaa tgaattgcta gtaaattaac gtgtattttt gtattcaccg	420
caatttcgat attctgcgcc cccgattgga ccactggggc taaagacttg agcatcagtt	480
taagcgtaga ctctcatca gctgtgttgt cgttgccgta ttctttctcc agatactcct	540
tcacaatctt ttcatttcgc cctattgaac cggccaacaa ttcttagtaa actccggacg	600
gttccgtctt gaaaagatga ggggtcccat cagaatcgaa gcctccgaca agcattgaaa	660
ttccaaaagg ccgacggcca gtgggttgag tgtatctcta acaaaacaaa aagttactca	720
gatgttgtaa gtcaattgac ctgttttata tcagctatga tgcgagagat atgcatgaca	780
gatacgcggt cctcaagcgt caatttgtaa ttttcgcatt caactcgagc acggtcgata	840
aggacgcgtg catcggcgct gagtcggcg aatgcgacca taacatgcta aacataggcg	900
tgcattgaag aggaataagg aatttaccga atccaatgca tgtattttac gaatggtacg	960
ttcgtcttgc agagtcggga tagatttctg caaaataaaa gttcaccaga acagtttaag	1020
caatttttat gtctgtgaat acactagcta aaaataattt actttttcga ctccaattac	1080
aacacaattt tttcctttca cggcaacctt aagcaatgca cattaatatt tttaaaaagc	1140
aataaaacgc accgctgttg agcccttctt cactgcttct tgcgcatagt caacttgaaa	1200
aagcttgcca tccggagaaa aaatcgtaat tgcacgatca taacgctcca tagtttattt	1260
gctaaaaatc agcttacaaa aagcgggttag gcatttaaaa tttaaagctcc gtaaaaattc	1320
aattaaaaat catcacattt atttttttaa tttttcaatt tttaaatttt tcttttttgg	1380
cgaactgtct acccttgtaa cttctaaaaa aggagttgag actgaaatcc cgcgccagat	1440
cccgtccgga cctgtctctg ttttccatag tgaacaaata attattgttg tatttttact	1500
ctttcgctgc tccacacaca tctctttcat gaactttaga caaaaagtat ttaaatgcgt	1560

-continued

```

cttgagaagt gttgggtttt ttcatacaaca atttatccgg gccacggaat tcaattcgta 1620
cgtaacgaag caacggtgaa aacaatttat tgttatagta cataataaat taaaattttt 1680
gttttaggttt tcaagttttg taggtcaaaa tgcaacaaat tatttaaaaa gaagaagaac 1740
ccgcgcgaat tgaatggac gaaggcatcg cggcgaattc ggggaaaagt tagtgtttga 1800
tttttgtttt tactctttta catttatgtt aaatttaaat ttcttttact cttaggaat 1860
tgggtcaacga tgttactcaa gcgatggaaa ttcgcagaaa cgaaccgaca aaatatgata 1920
gaaacctttg ggaactgca ggtaaatcgt ccatatatac caacaaaccg taacgacgaa 1980
aaaaagtacc ggaagggaga atccgcaaaa tctttgctct cggacactta aacatttttc 2040
ctgtttttaa tttttcatgg acgaaaaaac atatacagcg gttttcgcca aaaaaaaat 2100
aaccaatttg ggtagacaag tatgtctaataaatcttcca tttgaatttt gattt 2155

```

```

<210> SEQ ID NO 115
<211> LENGTH: 148
<212> TYPE: DNA
<213> ORGANISM: Heterodera glycines

```

```

<400> SEQUENCE: 115

```

```

tcatttcgcc ctattgaacc ggccaacaat tcgtagtaaa ctccggacgg ttccgtcttg 60
aaaagatgag gggccccatc agaatcgaag cctccgacaa gcattgaaat tccaaaaggc 120
cgacggccag tggtttgagt gtatctct 148

```

```

<210> SEQ ID NO 116
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Saccharomyces cerevisiae

```

```

<400> SEQUENCE: 116

```

```

ttctttgaag tatcaggagg tgg 23

```

```

<210> SEQ ID NO 117
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Saccharomyces cerevisiae

```

```

<400> SEQUENCE: 117

```

```

atgattattg caattccaac agg 23

```

```

<210> SEQ ID NO 118
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Saccharomyces cerevisiae

```

```

<400> SEQUENCE: 118

```

```

gctattttta gtggtatggc agg 23

```

```

<210> SEQ ID NO 119
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Saccharomyces cerevisiae

```

```

<400> SEQUENCE: 119

```

```

accatgtaaa tattgtgaac cagg 24

```

-continued

```

<210> SEQ ID NO 120
<211> LENGTH: 4107
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 120

atggataaaa aatattcaat cggtttagat atcgggtacaa attcagtagg ttgagctgta    60
atcacagatg aatataaagt accttcaaaa aaatttaaag tattaggtaa tacagataga    120
cattcaatca aaaaaaatTT aatcgggtgct ttattatttg attcagggtga aacagctgaa    180
gctacaagat taaaaagaac agctagaaga agatatacaa gaagaaaaaa tagaatctgt    240
tatttacaag aaatcttttc aaatgaaatg gctaaagtag atgattcatt tttocataga    300
ttagaagaat catttttagt tgaagaagat aaaaaacatg aaagacatcc tatctttggt    360
aatatcgtag atgaagtagc ttatcatgaa aaatatccta caatctatca ttaagaaaa    420
aaattagtag attcaactga taaagctgat ttaagattaa tctatttagc tttagctcat    480
atgatcaaat ttgaggtgca tttcttaatc gaaggtgatt taaatcctga taattcagat    540
gtagataaat tattcatcca attagtacaa acatataatc aattatttga agaaaatcct    600
atcaatgctt cagggtgtaga tgctaaagca atcttatcag ctagattatc aaaatcaaga    660
agattagaaa atttaatgcg tcaattacct ggagaaaaaa aaaatggttt atttggtaat    720
ttaatcgcot tatcattagg ttaactcct aatttcaaat caaatttcga tttagctgaa    780
gatgcaaaat tacaattatc taaagataca tatgatgatg atttagataa tttattagct    840
caaatcggtg atcaatatgc tgattttatc tttagctgcta aaaatttatc agatgctatc    900
ttattatcag atatcttaag agtaaataca gaaatcacia aagcaccttt atcagcttca    960
atgatcaaaa gatatgatga acatcatcaa gatttaacat tattaaaagc tttagtaaga   1020
caacaattac cagaaaaata taaagaaatc ttctttgatc aatcaaaaaa tggttatgct   1080
ggttatatcg atggtggtgc ttctcaagaa gaattctata aattcatcaa acctatctta   1140
gaaaaaatgg atggtacaga agaattatta gtaaaattaa atagagaaga tttattaaga   1200
aaacaagaaa catttgataa tggttcaatc cctcatcaaa tccatttagg tgaattacat   1260
gcaatcttaa gaagacaaga agatTTTTat cctttcttaa aagataatag agaaaaaatc   1320
gaaaaaatct taacatttag aatcccttat tatgtaggtc ctttagctag aggtaattca   1380
agatttgctt gaatgacaag aaaatcagaa gaaacaatca caccttggaa ttttgaagaa   1440
gtagtagata aaggagcttc agcacaatca tttatcgaaa gaatgacaaa ttttgataaa   1500
aatttaccta atgaaaaagt ttacctaata cattcattat tatatgaata tttcacagta   1560
tataatgaat taacaaaagt aaaatatgta acagaaggta tgagaaaaac tgctttttta   1620
tcagggtgaac aaaaaaaagc aatcgtagat ttattattta aaacaaatag aaaagtaaca   1680
gtaaaacaat taaaagaaga ttatttcaaa aaaatcgaat gttttgattc agtagaaatc   1740
tctggtgtag aagatagatt taatgcttct ttaggtacat atcatgattt attaaaaatc   1800
atcaaagata aagatttctt agataatgaa gaaaatgaag atatcttaga agatattcgt   1860
ttaacattaa ctttattoga agatagagaa atgatcgaag aaagattaaa aacatatgct   1920
catttatttg atgataaagt aatgaaacaa ttaaaaagaa gaagatatac tggttgaggt   1980
agattatcaa gaaaattaat caatggtatc agagataaac aatctggtaa aacaatctta   2040

```

-continued

gatttcttaa aatcagatgg ttttgctaag agaaatttca tgcaattaat ccatgatgat	2100
agtttaactt ttaaagaaga tatccaaaaa gctcaagtat caggtcaagg tgattcatta	2160
catgaacata tcgctaattt agctgggtct cctgctatca aaaaaggat cttacaaact	2220
gtaaaagttg tagatgaatt agttaagtt atgggtagac ataaacctga aaatatcgta	2280
atcgaaatgg caagagaaaa tcaacaaca caaaaaggac aaaaaattc aagagaaaga	2340
atgaaaagaa tcgaagaagg tatcaagaa ttaggttcac aaatcttaa agaaccatcct	2400
gtagaaaata cacaattaca aaatgaaaa ttatatattt attatttaca aaatggtaga	2460
gatatgtatg tagatcaaga attagatatt aatagattat ctgattatga ttagatcat	2520
atcgtagctc aatcattctt aaaagatgat tcaatcgata ataaagtatt aacaagatca	2580
gataaaaata gaggtaaaag tgataatgta ccttctgaag aagttgtaaa aaaaatgaaa	2640
aattattgaa gacaattatt aaatgctaaa ttaatcacac aaagaaaatt cgataattta	2700
acaaaagctg aaagaggtgg tttatcagaa ttagataaag ctggtttcat caaaagacaa	2760
ttagttgaaa caagacaaat cactaaacat gttgctcaaa tcttagatag tagaatgaat	2820
acaaaatatg atgaaaatga taaattaatc agagaagtaa aagtaatcac attaaaatct	2880
aaattagtat cagatttttag aaaagatttt caattctata aagtaagaga aatcaataat	2940
tatcatcatg ctcatgatgc ttattttaa gctgtagtag gtacagcttt aatcaaaaaa	3000
tatccaaaat tagaatcaga atttgtatat ggagattata aagtatatga tggtagaaaa	3060
atgatcgcta aatcagaaca agaaatcggg aaagctactg ctaaatattt cttttattca	3120
aatatcatga attttttcaa aactgaaatc acttttagct atggtgaaat cagaaaaaga	3180
cctttaatcg aaacaaatgg tgaaactggg gaaatcgtat gagataaagg tagagatttt	3240
gctacagtaa gaaaagtatt atcaatgcct caagtaaata tcgttaaaaa aactgaagta	3300
caaatcgggt gttttttctaa agaataatc ttacaaaaa gaaattcaga taaattaatc	3360
gctagaaaaa aagattgaga tccaaaaaaa tatggtggtt tcgattcacc tacagtagca	3420
tattcagtat tagtagtagc aaaagtagaa aaaggtaaat ctaaaaaatt aaaatcagta	3480
aaagaattat taggtatcac aatcatggaa agatcatcat tcgaaaaaa tccaatcgat	3540
tttttagaag ctaagggtta taaagaagtt aaaaagatt taatcatcaa attacctaaa	3600
tatagtttat ttgaattaga aaatggaaga aaaagaatgt tagcatcagc tggatgaatta	3660
caaaaaggta atgaattagc attaccatct aaatatgtta atttcttata ttagcatca	3720
cattatgaaa aattaaaagg ttctcctgaa gataatgaac aaaaacaatt attttagaaa	3780
caacataaac attatttaga tgaaatcatc gaacaaatct cagaattttc aaaaagagta	3840
atcttagcag atgcaattt agataaagtt ttatctgctt ataataaaca tagagataaa	3900
cctatcagag aacaagcaga aaatatcatc catttattca cattaacaaa ttaggtgct	3960
cctgctgctt tcaaatattt cgatacaaca atcgatagaa aaagatatat tcaacaaaa	4020
gaagtattag atgcaacatt aatccatcaa tcaatcacag gtttatatga aactagaatc	4080
gatttatctc aattaggtgg tgattaa	4107

<210> SEQ ID NO 121

<211> LENGTH: 88

<212> TYPE: DNA

<213> ORGANISM: *Saccharomyces cerevisiae*

-continued

<400> SEQUENCE: 121

ctgcaggact agtaaataaa ttttaattaa aagtagtatt aacatattat aaatagacaa 60
aagagtctaa aggttaagat ttattaaa 88

<210> SEQ ID NO 122

<211> LENGTH: 148

<212> TYPE: DNA

<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 122

ttaatatatta cttattatta atatttttaa ttattaaaaa taataataat aataataatt 60
ataataatat tcttaaatat aataaagata tagatttata ttctattcaa tcaccttatt 120
ctagaagcgg cgcaccatg gaaagctt 148

<210> SEQ ID NO 123

<211> LENGTH: 97

<212> TYPE: DNA

<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 123

ttctttgaag tatcaggagg gtttttagagc tagaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cgggtgct 97

<210> SEQ ID NO 124

<211> LENGTH: 75

<212> TYPE: DNA

<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 124

tatatattat gtattattat ataaatatat atatattatta tattataagt aataataagt 60
attatattat atata 75

<210> SEQ ID NO 125

<211> LENGTH: 75

<212> TYPE: DNA

<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 125

gcttttatag cttagtggta aagcgataaa ttgaagattt atttacctgt agttcgattc 60
tcattaaggg caata 75

<210> SEQ ID NO 126

<211> LENGTH: 76

<212> TYPE: DNA

<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 126

aggagattag cttaattggg atagcattcg ttttacacac gaaagattat aggttcgaac 60
cctatatttc ctaaat 76

<210> SEQ ID NO 127

<211> LENGTH: 118

<212> TYPE: DNA

<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 127

-continued

 ttattaataa ttaacaataa ttaatatatt ataatttata tatatatatt ttatattatt 60

ataataatat tcttacaaat ataattatta tatattattc cttcaaaact cctaacgg 118

<210> SEQ ID NO 128

<211> LENGTH: 76

<212> TYPE: DNA

<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 128

gagcttgat agtttaattg gttaaaacat ttgttcata aataaataat gtaaggttca 60

attccttcta caagta 76

<210> SEQ ID NO 129

<211> LENGTH: 744

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 129

atgacacatt tagaagaag tagacaaatg tcaaaagggtg aagaattatt cactggagta 60

gtacctatct tagtagaatt agatgggat gtaaatggtc ataaattctc agtatcaggt 120

gaagtggaag gtgatgctac ataggtgaaa ttaacattaa aattcatctg tacaacaggt 180

aaattacctg taccttgacc tacattagta acaacattcg gatatggagt acaatgtttc 240

gcaagatata ctgatcatat gaaacaacat gatttcttca aatcagcaat gcctgaaggt 300

tacgtacaag aaagaacaat cttcttcaaa gatgatggta attataaaac aagagctgaa 360

gtaaaattcg aaggtgatac attagtaaat agaatcgagt taaaagggtat cgatttcaaa 420

gaagatggta atactctagg tcataaatta gaatataatt ataattcaca taatgtatat 480

atcatggctg ataaacaaaa aaatgggtatc aaagtaaatt tcaaaatcag acataatatc 540

gaagacgggt cagtacaatt agcagatcat tatcaacaaa atacacctat cggatgatgg 600

cctgtattat tacctgataa tcattactta agtacacaat cagctttatc aaaagatcct 660

aatgaaaaaa gagatcatat ggtattatta gaatttgtaa cagctgctgg tatcacacat 720

ggtatggatg aattatataa ataa 744

<210> SEQ ID NO 130

<211> LENGTH: 144

<212> TYPE: DNA

<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 130

atgagaacaa atggtatgac aatgcataaa ttaccattat ttgtatgac aattttcatt 60

acagcgttct tattattatt atcattacct gtattatctg ctggtattac aatgttatta 120

ttagatagaa acttcaatac ttca 144

<210> SEQ ID NO 131

<211> LENGTH: 115

<212> TYPE: DNA

<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 131

aattaaaatt ttctcatgat taataaatcc ctttagcaag gataaaaaata aaaataaaaa 60

-continued

taaaaagtgtg atcagaaatt atcaaaaaat aaataataat aatataataa aaaca 115

<210> SEQ ID NO 132
<211> LENGTH: 64
<212> TYPE: DNA
<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 132

aatggtacaa agatgattat attcaacaaa tgcaaaagat attgcagtat tatatattat 60

gtta 64

<210> SEQ ID NO 133
<211> LENGTH: 93
<212> TYPE: DNA
<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 133

aattcacaat tatttaatgg tgcgcctctc agtgcgtata ttctgttgat gcgtctagca 60

ttagtattat gaatcatcaa tagatactta aaa 93

<210> SEQ ID NO 134
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 134

tttttcggag ttctcgttgg agg 23

<210> SEQ ID NO 135
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(16)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 135

nnnnnnnnnn nnnnnncaac agg 23

<210> SEQ ID NO 136
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (19)..(23)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 136

gctattttta gtggtatgnn nnn 23

<210> SEQ ID NO 137
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (8)..(24)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 137

accatgtnnn nnnnnnnnnn nnnn 24

<210> SEQ ID NO 138
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 138

ctattcaggc acattcagga cc 22

<210> SEQ ID NO 139
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 139

ttttatcctt gctaaagggga 20

<210> SEQ ID NO 140
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 140

tttgataatt tctgatcaac 20

<210> SEQ ID NO 141
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 141

agaggtatac caacacaaga ttc 23

<210> SEQ ID NO 142
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 142

caggtgaagg tgaaggtgat gc 22

<210> SEQ ID NO 143
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 143

gatctgctaa ttgtactgaa ccg 23

-continued

<210> SEQ ID NO 144
<211> LENGTH: 1308
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 144

ctattcaggc acattcagga cctagtgtag atttagcaat ttttgcatta catttaacat	60
caatttcacg attattaggt gctattaatt tcattgtaac aacattaaat atgagaacaa	120
atggtatgac aatgcataaa ttaccattat ttgtatgac aattttcatt acagcgttct	180
tattattatt atcattacct gtattatctg ctggtattac aatgttatta ttagatagaa	240
acttcaatag ttcatttttc ggagtttctg gtggagggtg tggaatgaca catttagaaa	300
gaagtagaca aatgtcaaaa ggtgaagaat tattcactgg agtagtacct atcttagtag	360
aattagatgg tgatgtaaat ggtcataaat tctcagtac aggtgaaggt gaaggatg	420
ctacatatgg taaattaaca ttaaaattca tctgtacaac aggtaaatta cctgtacctt	480
gacctacatt agtaacaaca ttccgatag gagtacaatg ttccgcaaga taccctgac	540
atatgaaaca acatgatttc ttcaaatcag caatgcctga aggttacgta caagaaagaa	600
caatcttctt caaagatgat ggtaattata aaacaagagc tgaagtaaaa ttcgaagggtg	660
atacattagt aaatagaatc gagttaaag gtatcgattt caaagaagat ggtaatatct	720
taggtcataa attagaatat aattataatt cacataatgt atatatcatg gctgataaac	780
aaaaaaatgg tatcaaagta aatttcaaaa tcagacataa tatcgaagac gggtcagtac	840
aattagcaga tcattatcaa caaaatacac ctatcggtga tggctcgtga ttattacctg	900
ataatcatta cttaagtaca caatcagctt tatcaaaaga tcctaataaa aaaagagatc	960
atatgggtatt attagaattt gtaacagctg ctggtatcac acatgggtatg gatgaattat	1020
ataaataaca acaggaatta aaattttctc atgattaata aatcccttta gcaaggataa	1080
aaataaaaat aaaaataaaa agttgatcag aaattatcaa aaaaataata ataataatat	1140
aataaaaaca tatttaataa ataataatat aattataata aatatatata aaggtaattt	1200
atatgatatt tatccaagat caaatagaaa ttatatcaa ccaataata ttaataaaga	1260
attagtagta tatggttata atttagaatc ttgtgttggt atacctct	1308

<210> SEQ ID NO 145
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 145

ggtttaaaacc ctgttactgg tgg	23
----------------------------	----

<210> SEQ ID NO 146
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 146

cttcacctgt aaatggacca cgg	23
---------------------------	----

<210> SEQ ID NO 147
<211> LENGTH: 23

-continued

```

<212> TYPE: DNA
<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 147

tttacagggtg aaggtcacgt tgg                23

<210> SEQ ID NO 148
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 148

gtagctaaat aagggtatgg agg                23

<210> SEQ ID NO 149
<211> LENGTH: 4107
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 149

atggacaaaa aatactcaat tggtttagat attggtacaa attcagttgg ttgggctggt    60
attacagatg aatataaagt tccaagtaaa aaatttaaa ttttaggtaa tacagatcgt    120
cactcaatta agaaaaactt aattggtgct ttattatttg attcaggtga aacagctgaa    180
gctacacggt taaaacgtac agctcgtcgt cgttatacac gtcgtaaaaa tcgtatttgt    240
tatttacaag aaattttctc aaatgaaatg gctaaagtgt atgattcatt ttttcaccgt    300
ttagaagaat catttttagt tgaagaagat aaaaaacacg aacgtcaccc aatttttggt    360
aatattgttg atgaagttgc ttatcacgaa aaatatccaa caatttatca cttacgtaaa    420
aaattagttg attcaactga taaagctgat ttacgtttaa tttatttagc tttagctcac    480
atgattaaat tccgtggtca cttcttaatt gaaggtgatt taaaccaga taattcagat    540
gttgacaaat tattcattca attagttcaa acatataatc aattatttga agaaaatcca    600
attaatgctt cagggtgttg tgctaaagca attttatcag ctcgtttattc aaaatcacgt    660
cgtttagaaa acttaattgc tcaattacca ggtgaaaaga aaaatggttt attcggtaac    720
ttaattgcat tatcattagg tttaacacca aatttcaaat caaacttcga tttagctgaa    780
gatgctaaat tacaattatc aaaagataca tacgatgatg atttagataa cttattagca    840
caaattggtg atcaatatgc tgattttatc ttagctgcta aaaacttatc agatgctatt    900
ttattatcag atattttacg tgtaataaca gaaattacaa aagctccatt atcagcttca    960
atgattaaac gttatgatga acaccaccaa gatttaacat tattaaaagc tttagttcgt   1020
caacaattac ctgaaaaata caaagaaatt ttcttcgata aatctaaaaa tggttatgct   1080
ggttatattg atggtggtgc ttcacaagaa gaattctata aattcattaa acctatttta   1140
gaaaaaatgg atggtacaga agaattatta gttaaattaa atcgtgaaga tttattacgt   1200
aaacaacgta catttgataa tggttcaatt cctcaccaaa ttcatttagg tgaattacac   1260
gcaattttac gtcgtcaaga agatttttat ccattcttaa aagataatcg tgaaaaaatt   1320
gaaaaaatat taacatttcg tattccatat tatgtaggtc cattagctcg tggtaattca   1380
cgtttcgctt ggatgacacg taaatctgaa gaaacaatta caccttgga ttttgaagaa   1440
gttggtgata aagggtgctag tgctcaatca tttattgaac gtatgacaaa tttcgacaaa   1500

```

-continued

aacttaccaa atgaaaaagt ttaccacaaa cactcattat tatatgaata tttcacagtt	1560
tataatgaat taacaaaagt taaatatgtt acagaaggtg tgcgtaaacc tgcattttta	1620
agtgggtgaac aaaagaagc tattgttgac ttattattca aaacaaatcg taaagttaca	1680
gttaaacat taaaagaaga ttactttaag aaaattgaat gttttgattc agtagaaatt	1740
tcaggtgtag aagatcgttt caatgcttca ttaggtacat accacgattt attaaaaatt	1800
attaaagaca aagacttttt agataatgaa gaaaatgaag atatttttaga agatattgtt	1860
ttacattaa cattattoga agatcgtgaa atgattgaag aacgtttaaa aacatatgct	1920
cacttatttg atgataaagt tatgaaacaa ttaaaacgtc gtcgttacac aggttggggt	1980
cgtttatctc gtaaatat taacggtatt cgtgacaaac aatcaggtaa aacaatttta	2040
gatttcttaa aatcagatgg ttttgctaat cgtaacttta tgcaattaat tcacgatgat	2100
tctttaacat tcaagaaga tattcaaaaa gctcaagttt cagggtcaagg tgattcatta	2160
cacgaacaca ttgctaactt agctggttct ccagctatta aaaaagggtat ttacacaa	2220
gttaaagttg tagatgaatt agtaaaagta atgggtcgtc acaaacacaga aaacattgtt	2280
attgaaatgg cacgtgaaaa tcaacaaca caaaaaggtc aaaagaactc acgtgaacgt	2340
atgaaacgta ttgaagaagg tattaagaa ttaggttcac aaattttaaa agaacaccca	2400
gttgaaaata cacaattaca aaacgaaaaa ttatatattt actatttaca aaatggtcgt	2460
gatatgtatg tagatcaaga attagatatt aaccgtttat cagattatga tgttgatcac	2520
attgttcac aatctttctt aaaagacgat tcaattgata acaaagtttt aacacgttca	2580
gataaaaaac gtggtaaatc agataatgta ccatcagaag aagtagttaa gaaaatgaaa	2640
aactattggc gtcaattatt aaatgcaaaa ttaattacac aacgtaaatt cgataactta	2700
acaaaagctg aacgtgggtg ttatcagaa ttagacaaag ctggtttcat taaacgtcaa	2760
ttagtagaaa cacgtcaaat tactaaacac gttgctcaaa ttttagactc tcgtatgaat	2820
acaaaatatg atgaaaatga taaattaatt cgtgaagtta aagttattac attaaaatca	2880
aaattagtat cagatttcog taaagatttc caattctaca aagttcgtga aattaacaac	2940
tatcaccacg ctcacgatgc ttacttaaat gctgtgttg gtactgcatt aattaaaaaa	3000
tacccaaaat tagaatctga attcgtttat ggtgactata aagtttatga tgtacgtaaa	3060
atgattgcta aatcagaaca agaaattggt aaagctactg ctaaatactt tttctattca	3120
aacattatga atttctttaa aactgaaatt acattagcta acggtgaaat tcgtaaacgt	3180
ccattaattg aaactaatgg tgaaactggt gaaattgtat gggataaagg tcgtgatttc	3240
gctacagttc gtaaaagtatt atcaatgcca caagttaata ttgttaaaaa aactgaagtt	3300
caaacaggtg gtttttcaaa agaattctatt ttacctaaac gtaactcaga caaattaatt	3360
gctcgtaaaa aagattggga tcctaaaaaa tatggtggtt tcgattcacc aacagtagct	3420
tattcagtat tagttgtagc taaagtagaa aaaggtaaat ctaaaaaatt aaaatcagta	3480
aaagaattat taggtattac aattatggaa cgttcacatc tcgagaaaaa cccaattgat	3540
ttcttagaag ctaagggtta taaagaagtt aaaaagatt taattattaa attaccaaaa	3600
tactctttat ttgaattaga aaacggtcgt aaacgtatgt tagcttctgc tggatgaatta	3660
caaaaaggta atgaattagc attaccatca aaatatgtaa atttcttata cttagcttca	3720
cactacgaaa aattaaaagg ttcaccagaa gataacgaac aaaaacaatt attcgttgaa	3780

-continued

```

caacataaac actatttaga tgaattatt gaacaaattt cagaattttc aaaacgtgtt 3840
atatttagctg atgctaattt agataaagtt ttatctgctt ataacaaaca ccgtgataaa 3900
cctattcgtg aacaagctga aaacattatt cacttattta cattaacaaa tttaggtgct 3960
ccagctgctt tcaaatattt cgatacaaca attgaccgta aacgttacac atcaacaaaa 4020
gaagtttttag acgctacatt aattcatcaa tcaattacag gtttatatga aacacgtatt 4080
gatttaagtc aattaggtgg tgattaa 4107

```

<210> SEQ ID NO 150

<211> LENGTH: 1368

<212> TYPE: PRT

<213> ORGANISM: Streptococcus pyogenes

<400> SEQUENCE: 150

```

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1      5      10      15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20     25     30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35     40     45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50     55     60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65     70     75     80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85     90     95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100    105    110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115    120    125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130    135    140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145    150    155    160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165    170    175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180    185    190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195    200    205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210    215    220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225    230    235    240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245    250    255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260    265    270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275    280    285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

```

-continued

290	295	300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser		
305	310	315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys		
	325	330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe		
	340	345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser		
	355	360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp		
	370	375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg		
	385	390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu		
	405	410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe		
	420	425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile		
	435	440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp		
	450	455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu		
	465	470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr		
	485	490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser		
	500	505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys		
	515	520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln		
	530	535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr		
	545	550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp		
	565	570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly		
	580	585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp		
	595	600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr		
	610	615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala		
	625	630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr		
	645	650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp		
	660	665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe		
	675	680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe		
	690	695 700

-continued

Lys	Glu	Asp	Ile	Gln	Lys	Ala	Gln	Val	Ser	Gly	Gln	Gly	Asp	Ser	Leu	705	710	715	720
His	Glu	His	Ile	Ala	Asn	Leu	Ala	Gly	Ser	Pro	Ala	Ile	Lys	Lys	Gly	725	730	735	
Ile	Leu	Gln	Thr	Val	Lys	Val	Val	Asp	Glu	Leu	Val	Lys	Val	Met	Gly	740	745	750	
Arg	His	Lys	Pro	Glu	Asn	Ile	Val	Ile	Glu	Met	Ala	Arg	Glu	Asn	Gln	755	760	765	
Thr	Thr	Gln	Lys	Gly	Gln	Lys	Asn	Ser	Arg	Glu	Arg	Met	Lys	Arg	Ile	770	775	780	
Glu	Glu	Gly	Ile	Lys	Glu	Leu	Gly	Ser	Gln	Ile	Leu	Lys	Glu	His	Pro	785	790	795	800
Val	Glu	Asn	Thr	Gln	Leu	Gln	Asn	Glu	Lys	Leu	Tyr	Leu	Tyr	Tyr	Leu	805	810	815	
Gln	Asn	Gly	Arg	Asp	Met	Tyr	Val	Asp	Gln	Glu	Leu	Asp	Ile	Asn	Arg	820	825	830	
Leu	Ser	Asp	Tyr	Asp	Val	Asp	His	Ile	Val	Pro	Gln	Ser	Phe	Leu	Lys	835	840	845	
Asp	Asp	Ser	Ile	Asp	Asn	Lys	Val	Leu	Thr	Arg	Ser	Asp	Lys	Asn	Arg	850	855	860	
Gly	Lys	Ser	Asp	Asn	Val	Pro	Ser	Glu	Glu	Val	Val	Lys	Lys	Met	Lys	865	870	875	880
Asn	Tyr	Trp	Arg	Gln	Leu	Leu	Asn	Ala	Lys	Leu	Ile	Thr	Gln	Arg	Lys	885	890	895	
Phe	Asp	Asn	Leu	Thr	Lys	Ala	Glu	Arg	Gly	Gly	Leu	Ser	Glu	Leu	Asp	900	905	910	
Lys	Ala	Gly	Phe	Ile	Lys	Arg	Gln	Leu	Val	Glu	Thr	Arg	Gln	Ile	Thr	915	920	925	
Lys	His	Val	Ala	Gln	Ile	Leu	Asp	Ser	Arg	Met	Asn	Thr	Lys	Tyr	Asp	930	935	940	
Glu	Asn	Asp	Lys	Leu	Ile	Arg	Glu	Val	Lys	Val	Ile	Thr	Leu	Lys	Ser	945	950	955	960
Lys	Leu	Val	Ser	Asp	Phe	Arg	Lys	Asp	Phe	Gln	Phe	Tyr	Lys	Val	Arg	965	970	975	
Glu	Ile	Asn	Asn	Tyr	His	His	Ala	His	Asp	Ala	Tyr	Leu	Asn	Ala	Val	980	985	990	
Val	Gly	Thr	Ala	Leu	Ile	Lys	Lys	Tyr	Pro	Lys	Leu	Glu	Ser	Glu	Phe	995	1000	1005	
Val	Tyr	Gly	Asp	Tyr	Lys	Val	Tyr	Asp	Val	Arg	Lys	Met	Ile	Ala		1010	1015	1020	
Lys	Ser	Glu	Gln	Glu	Ile	Gly	Lys	Ala	Thr	Ala	Lys	Tyr	Phe	Phe		1025	1030	1035	
Tyr	Ser	Asn	Ile	Met	Asn	Phe	Phe	Lys	Thr	Glu	Ile	Thr	Leu	Ala		1040	1045	1050	
Asn	Gly	Glu	Ile	Arg	Lys	Arg	Pro	Leu	Ile	Glu	Thr	Asn	Gly	Glu		1055	1060	1065	
Thr	Gly	Glu	Ile	Val	Trp	Asp	Lys	Gly	Arg	Asp	Phe	Ala	Thr	Val		1070	1075	1080	
Arg	Lys	Val	Leu	Ser	Met	Pro	Gln	Val	Asn	Ile	Val	Lys	Lys	Thr		1085	1090	1095	

-continued

Glu Val	Gln Thr	Gly Gly	Phe	Ser Lys	Glu Ser	Ile	Leu Pro	Lys	
1100			1105			1110			
Arg Asn	Ser Asp	Lys Leu	Ile	Ala Arg	Lys Lys	Asp	Trp Asp	Pro	
1115			1120			1125			
Lys Lys	Tyr Gly	Gly Phe	Asp	Ser Pro	Thr Val	Ala	Tyr Ser	Val	
1130			1135			1140			
Leu Val	Val Ala	Lys Val	Glu	Lys Gly	Lys Ser	Lys	Lys Leu	Lys	
1145			1150			1155			
Ser Val	Lys Glu	Leu Leu	Gly	Ile Thr	Ile Met	Glu	Arg Ser	Ser	
1160			1165			1170			
Phe Glu	Lys Asn	Pro Ile	Asp	Phe Leu	Glu Ala	Lys	Gly Tyr	Lys	
1175			1180			1185			
Glu Val	Lys Lys	Asp Leu	Ile	Ile Lys	Leu Pro	Lys	Tyr Ser	Leu	
1190			1195			1200			
Phe Glu	Leu Glu	Asn Gly	Arg	Lys Arg	Met Leu	Ala	Ser Ala	Gly	
1205			1210			1215			
Glu Leu	Gln Lys	Gly Asn	Glu	Leu Ala	Leu Pro	Ser	Lys Tyr	Val	
1220			1225			1230			
Asn Phe	Leu Tyr	Leu Ala	Ser	His Tyr	Glu Lys	Leu	Lys Gly	Ser	
1235			1240			1245			
Pro Glu	Asp Asn	Glu Gln	Lys	Gln Leu	Phe Val	Glu	Gln His	Lys	
1250			1255			1260			
His Tyr	Leu Asp	Glu Ile	Ile	Glu Gln	Ile Ser	Glu	Phe Ser	Lys	
1265			1270			1275			
Arg Val	Ile Leu	Ala Asp	Ala	Asn Leu	Asp Lys	Val	Leu Ser	Ala	
1280			1285			1290			
Tyr Asn	Lys His	Arg Asp	Lys	Pro Ile	Arg Glu	Gln	Ala Glu	Asn	
1295			1300			1305			
Ile Ile	His Leu	Phe Thr	Leu	Thr Asn	Leu Gly	Ala	Pro Ala	Ala	
1310			1315			1320			
Phe Lys	Tyr Phe	Asp Thr	Thr	Ile Asp	Arg Lys	Arg	Tyr Thr	Ser	
1325			1330			1335			
Thr Lys	Glu Val	Leu Asp	Ala	Thr Leu	Ile His	Gln	Ser Ile	Thr	
1340			1345			1350			
Gly Leu	Tyr Glu	Thr Arg	Ile	Asp Leu	Ser Gln	Leu	Gly Gly	Asp	
1355			1360			1365			

<210> SEQ ID NO 151

<211> LENGTH: 279

<212> TYPE: DNA

<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 151

tcttaattca acatttttaa gtaaaactg tttaatgtta tacttttacg aatacacata	60
tggtaaaaaa taaacaata tctttaaaat aagtaaaaat aatttgtaaa ccaataaaaa	120
atatatttat ggtataatat aacatatgat gtaaaaaaaa ctatttgtct aatttaataa	180
ccatgcattt tttatgaaca cataataatt aaaagcgttg ctaatggtgt aaataatgta	240
tttattaaat taaataattg ttattataag gagaaatcc	279

<210> SEQ ID NO 152

<211> LENGTH: 414

<212> TYPE: DNA

-continued

<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 152

```
aatggatat ttgttacatt taattccaca aaaatgtcca atacttaaaa tacaaaatta      60
aaagtattag ttgtaaacctt gactaacatt ttaaatttta aattttttcc taattatata    120
ttttacttgc aaaatttata aaaattttat gcatttttat atcataataa taaaaccttt    180
attcatggtt tataatataa taattgtgat gactatgcac aaagcagttc tagtcccata    240
tatataacta tatataaccg gtttaaagat ttatttaaaa atatgtgtgt aaaaaatgct    300
tatttttaat tttattttat ataagttata atattaaata cacaatgatt aaaattaaat    360
aataataaat ttaacgtaac gatgagttgt ttttttattt tggagataca cgca          414
```

<210> SEQ ID NO 153

<211> LENGTH: 258

<212> TYPE: DNA

<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 153

```
tttttatttt tcatgatggt tatgtgaata gcataaacat cgtttttatt tttatggtgt      60
ttaggttaaa tacctaaaca tcattttaca tttttaaaat taagttctaa agttatcttt    120
tgtttaaatt tgcctgtctt tataaattac gatgtgccag aaaaataaaa tcttagcttt    180
ttattataga atttatcttt atgtattata ttttataagt tataataaaa gaaatagtaa    240
cataactaaag cggatgta                                258
```

<210> SEQ ID NO 154

<211> LENGTH: 102

<212> TYPE: DNA

<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 154

```
ttaacccatg attaacaact atatcaataa aatcaatttg tagtgaaata ctctgattga      60
cattaaaata ataccatgat aaaaattata ataacaaatt tt                          102
```

<210> SEQ ID NO 155

<211> LENGTH: 101

<212> TYPE: DNA

<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 155

```
tttttcctaa tgtactttgt tgtaaaagtg gctggtttaa cttttttagg ttctggattg      60
aacaataatg gcagttaaga gtcactaaag ctgctgtata g                          101
```

<210> SEQ ID NO 156

<211> LENGTH: 73

<212> TYPE: DNA

<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 156

```
acgtcccttag ttcagtcggt agaacgcagg tttccaaaac ctgatgtcgt gggttcaatt      60
cctacagggc gtg                                73
```

<210> SEQ ID NO 157

<211> LENGTH: 72

<212> TYPE: DNA

-continued

<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 157

gggttgctaa ctcaatggta gactactcgg ctcttaacgg ataagttctg ggttcgagtc 60

ccaggttaacc ca 72

<210> SEQ ID NO 158

<211> LENGTH: 82

<212> TYPE: DNA

<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 158

gccttcgtga tggaaactggt agacatcctg gttttaggaa ccagtgtga aaggcgtgcc 60

ggttcaaatac cgccgaagga ca 82

<210> SEQ ID NO 159

<211> LENGTH: 795

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 159

atggctcgtg aagcgggttat cgccgaagta tcaactcaac tatcagaggt agttggcgctc 60

atcgagcgcc atctcgaacc gacgttgctg gccgtacatt tgtacggctc cgagtggtat 120

ggcggcctga agccacacag tgatattgat ttgctgggta cggtgaccgt aaggcttgat 180

gaaacaacgc ggcgagcttt gatcaacgac cttttggaaa cttcggcttc ccctggagag 240

agcgagattc tccgcgctgt agaagtcacc attgttgctg acgacgacat cattccgtgg 300

cgttatccag ctaagcgoga actgcaattt ggagaatggc agcgcaatga cattcttgca 360

ggtatcttgc agccagccac gatcgacatt gatctggcta tcttgctgac aaaagcaaga 420

gaacatagcg ttgccttggg aggtccagcg gcggaggaaac tctttgatcc ggttcctgaa 480

caggatctat ttgaggcgct aaatgaaacc ttaacgctat ggaactcgcc gcccgactgg 540

gctggcgatg agcgaaatgt agtgcttacg ttgtcccgca tttggtacag cgcagtaacc 600

ggcaaaatcg cgccgaagga tgctcgtgcc gactgggcaa tggagcgccct gccggcccag 660

tatcagcccg tcatacttga agctagacag gcttatcttg gacaagaaga agatcgcttg 720

gcctcgcgcg cagatcagtt ggaagaattt gtccactacg tgaaaggcga gatcactaag 780

gtagttggca aataa 795

<210> SEQ ID NO 160

<211> LENGTH: 189

<212> TYPE: DNA

<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 160

catataccta aaggcccttt ctatgctcga ctgataagac aagtacataa atttgctagt 60

ttacattatt ttttatttct aaatatataa tatatttaaa tgtatttaaa atttttcaac 120

aatttttaaa ttatatttcc ggacagatta ttttaggata gtcaaaagaa gttacattta 180

tttatataa 189

<210> SEQ ID NO 161

-continued

```

<211> LENGTH: 400
<212> TYPE: DNA
<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 161
ttttttttta aactaaaata aatctgggta accatacctg gtttatttta gtttatacac    60
acttttcata tatatatact taatagctac cataggcagt tggcaggacg tccccttacg    120
ggacaaatgt attttatggt gcctgccaac tgcctaatat aaatattagt ggacgtcccc    180
ttccccttac gggcaagtaa acttagggat tttaatgctc cgtaggaggg caaataaatt    240
ttagtggcag ttgcctcgcc tatcggtcaa caagttcctt cggagtatat aaatatacctg    300
ccaaatgcg atatttatat actaggcagt ggcggtacca ctgactaat atttatattc    360
cgtaagacgt cctccttcgg agtatgtaaa catgctaagt    400

```

```

<210> SEQ ID NO 162
<211> LENGTH: 717
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 162
atggctaag gtgaagaatt attcacaggt gttgtaccta ttttagtaga attagacggt    60
gatgtaaacg gtcacaaatt ttcagtttct ggtgaagggt aaggtagcgc aacttatggt    120
aaattaacac ttaaatatcat ttgtactaca ggtaaaattac cagtaccttg gccatcatta    180
gttacaactt ttacatacgg tgtacaatgt ttcagtcgtt accctgatca catgaaacaa    240
catgactttt tcaaatctgc tatgccagaa ggttatgttc aagaacgtac tatttttttc    300
aaagatgaag gtaattataa aacacgtgct gaagtaaaat ttgaagggtga tacttttagtt    360
aaccgtattg aattaaaagg tattgacttc aaagaagatg gtaatatattt aggtcacaaa    420
cttgaatata actacaattc acataacgta tatattatgg cagacaaaac aaaaaatggt    480
attaaagtaa actttaaaat tcgtcataat atcgaggatg gttctgtaca attagctgac    540
cactatcaac aaaacacacc aattgggtgat ggtcctgttt tacttccaga caatcattat    600
ttaagtactc aatctgcttt atcaaaagat cctaacgaaa aacgtgacca catggtatta    660
cttgaatttg ttacagcagc tgggtattact cacggtatgg atgaattata caaataa    717

```

```

<210> SEQ ID NO 163
<211> LENGTH: 74
<212> TYPE: DNA
<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 163
gtccttttct ttactttaaa ctggagtgaa tacagtgatt tcttaacatt taaagggtgt    60
ttaaaccttg ttac    74

```

```

<210> SEQ ID NO 164
<211> LENGTH: 76
<212> TYPE: DNA
<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 164
tccatttaca ggtgaagggt acgttggttt atatgaaatt ttaacaactt cttggcatgc    60

```

-continued

acaattagct attaac 76

<210> SEQ ID NO 165
<211> LENGTH: 76
<212> TYPE: DNA
<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 165

gtactaactg gggatttggt cacagtatga aagaaatttt agaagctcac cgtgggtccat 60

ttacaggtga aggtca 76

<210> SEQ ID NO 166
<211> LENGTH: 76
<212> TYPE: DNA
<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 166

tacccttatt tagctactga ttacgggtaca caattatcat tatttacaca ccacacatgg 60

attggtgggt tctgta 76

<210> SEQ ID NO 167
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 167

gctggttggt tccactacca c 21

<210> SEQ ID NO 168
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 168

caccttcaaa ttttacttca gcacgtg 27

<210> SEQ ID NO 169
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 169

catacgggtg acaatgtttc agtcg 25

<210> SEQ ID NO 170
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 170

gtgagaaata atagcatcac ggtgac 26

<210> SEQ ID NO 171
<211> LENGTH: 1408
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: Synthetic Construct

<400> SEQUENCE: 171

```

gctggttggt tccactacca caaagctgct ccaaaactag aatggttcca aaacgttgaa    60
tcaatgttaa accaccactt aggtggtctt ctggttttag gtagtttagc ttgggctggt    120
caccaaattc acgtttcttt accagtaaac aaattattag atgctgggtg agatccaaaa    180
gaaattccac ttctcatga tttattatta aatcgtgcta ttatggctga cttataccca    240
agttttgcta aaggtattgc tcctttcttt actttaaac ggagtgaata cagtgatctc    300
ttaacattta aaggtggttt aaaccctggt acattatcag gttctgctgg ttcagcagct    360
ggtagtgcta aaggtgaaga attattcaca ggtgtgttac ctattttagt agaattagac    420
ggtagtgtaa acggtcacaa attttcagtt tctggtgaag gtgaagggtga cgcaacttat    480
ggtaaattaa cacttaaat cttttgtact acaggtaaat taccagtacc ttggccatca    540
ttagttacaa cttttacata cgggtgtaca tgtttcagtc gttaccctga tcacatgaaa    600
caacatgact ttttcaaadc tgctatgcca gaaggttatg ttcaagaacg tactattttt    660
ttcaaagatg acggttaatta taaaacacgt gctgaagtaa aatttgaagg tgatacttta    720
gttaaccgta ttgaattaaa aggtattgac ttcaaagaag atggtaatat tttaggtcac    780
aaacttgaat ataactacaa ttcacataac gtatatatta tggcagacaa acaaaaaaat    840
ggatttaaag taaactttaa aattcgtcat aatatcgagg atggttctgt acaattagct    900
gaccactatc aacaaaacac accaattggt gatggtcctg ttttacttcc agacaatcat    960
tatttaagta ctcaatctgc tttatcaaaa gatcctaacg aaaaacgtga ccacatggta   1020
ttacttgaat ttgttacagc agctggtatt actcacggtg tggatgaatt atacaaataa   1080
tccatttaca ggtgaaggtc acgttggttt atatgaaatt ttaacaactt cttggcatgc   1140
acaattagct attaacttag ctttatttgg ttcggtatca attattgtag ctcaccacat   1200
gtacgcaatg cctccatacc cttatttagc tactgattac ggtacacaat tatcattatt   1260
tacacaccac acatggattg gtggtttctg tattgttggt gctgggtgctc acgcagctat   1320
tttcatgggt cgtgactacg atcctactaa taactacaac aacttattag accgtgtaat   1380
tcgtcacctg gatgctatta tttctcac                                     1408

```

<210> SEQ ID NO 172

<211> LENGTH: 16

<212> TYPE: PRT

<213> ORGANISM: Drosophila melanogaster

<400> SEQUENCE: 172

```

Arg Gln Ile Lys Ile Trp Phe Gln Asn Arg Arg Met Lys Trp Lys Lys
1           5           10          15

```

That which is claimed:

1. A method for altering the genome of an organelle, the method comprising:

a. introducing into an organelle a recombinant DNA construct comprising the following:

i. a first polynucleotide encoding at least one guide RNA, wherein the at least one guide RNA directs a polynucleotide guided polypeptide to cleave at least one target sequence present in an organelle genome;

ii. a second polynucleotide encoding a polynucleotide guided polypeptide, wherein the polynucleotide guided polypeptide, when associated with the guide RNA, cleaves the at least one target sequence;

iii. optionally, a third polynucleotide encoding at least one homologous organelle DNA sequence, wherein the at least one homologous organelle DNA is of sufficient size for homologous recombination, wherein integration of the at least one homologous organelle DNA sequence into the organelle genome results in removal of the at least one target sequence;

- iv. optionally, a fourth polynucleotide encoding at least one selectable marker or at least one screenable marker, or both; wherein the fourth polynucleotide is operably linked to a promoter that is functional in the organelle; and
 - v. optionally, a fifth polynucleotide encoding an origin of replication that is functional in the organelle; and
 - b. growing a cell comprising the organelle of (a) under conditions in which the first polynucleotide of (i) and the second polynucleotide of (ii) are each expressed.
2. The method of claim 1, wherein the method further comprises: (c) selecting a cell having an organelle that comprises an altered genome.
3. The method of claim 2, wherein the method further comprises: (d) selecting a cell that is homoplasmic for the altered genome of the organelle.
4. The method of claim 3, wherein the organelle is a plastid.
5. The method of claim 3, wherein the organelle is a mitochondrion.
6. The method of claim 1, comprising the third polynucleotide of (iii), wherein the third polynucleotide of (iii) comprises a sixth and a seventh polynucleotide, wherein the sixth and the seventh polynucleotides correspond to two adjacent regions of homology in the organelle genome, wherein the sixth and seventh polynucleotides are separated by a sequence that is heterologous to the organelle DNA.
7. The method of claim 6, wherein the sequence that is heterologous to the organelle DNA comprises at least one selected from the group consisting of: the first polynucleotide, the second polynucleotide, the fourth polynucleotide, an eighth polynucleotide, and any combination thereof, wherein the eighth polynucleotide encodes an RNA that is heterologous to the organelle.
8. The method of claim 1, wherein the at least one guide RNA is present on a polycistronic transcription unit.
9. The method of claim 8, wherein the at least one guide RNA is processed from a polycistronic RNA after transcription of the polycistronic transcription unit by use of at least one selected from the group consisting of: an RNA cleavage site, a Csy4 cleavage site, a ribozyme cleavage site, a polynucleotide guided polypeptide cleavage site, the presence of a tRNA sequence, and any combination thereof.
10. The method of claim 9, wherein the polycistronic RNA comprises a first tRNA sequence 5' to the at least one guide RNA and a second tRNA sequence 3' to the at least one guide RNA.
11. The method of claim 6, wherein at least one selected from the group consisting of: the first polynucleotide, the second polynucleotide, the fourth polynucleotide, the fifth polynucleotide, and any combination thereof, is located outside the region bounded by the sixth and the seventh polynucleotide.
12. The method of claim 6, comprising the fourth and the fifth polynucleotides, wherein both the fourth and the fifth polynucleotides are located outside the region bounded by the sixth and the seventh polynucleotides.
13. The method of claim 6, wherein said polynucleotide-guided polypeptide is selected from the group consisting of: a Cas9 protein, a MAD2 protein, a MAD7 protein, a CRISPR nuclease, a nuclease domain of a Cas protein, a Cpf1 protein, an Argonaute, modified versions thereof, and any combination thereof.
14. The method of claim 6, wherein the method further comprises introducing into the organelle a polynucleotide encoding at least one selectable marker selected from the group consisting of: a positive selectable marker, a negative selectable marker, and any combination thereof.
15. The method of claim 14, wherein the method further involves growing the cell in the presence of a positive selection agent and selecting a cell that is homoplasmic for the altered genome of the organelle.
16. The method of claim 15, wherein the method further involves growing the cell in the absence of the positive selection agent, followed by selecting a cell that lacks a non-integrated recombinant DNA construct.
17. The method of claim 15, wherein the method further involves growing the cell in the absence of the positive selection agent, followed by growing the cell in the presence of a negative selection agent, followed by selecting a cell that lacks a non-integrated recombinant DNA construct.
18. The method of claim 6, wherein the cell is a plant cell, wherein the organelle is a plastid or a mitochondrion, and wherein the method further comprises regenerating a plant from the plant cell comprising an altered organelle genome.
19. The method of claim 6, wherein the cell is a yeast cell or an algal cell.
20. A recombinant DNA construct comprising the following:
- i. a first polynucleotide encoding at least one guide RNA, wherein the at least one guide RNA directs a polynucleotide guided polypeptide to cleave at least one target sequence present in an organelle genome;
 - ii. a second polynucleotide encoding a polynucleotide guided polypeptide, wherein the polynucleotide guided polypeptide, when associated with the guide RNA, cleaves the at least one target sequence;
 - iii. a third polynucleotide comprising a sixth and a seventh polynucleotide, wherein the sixth and the seventh polynucleotides correspond to two adjacent regions of homology in the organelle genome, wherein the sixth and seventh polynucleotides are separated by a sequence that is heterologous to the organelle DNA, wherein the sequence that is heterologous to the organelle DNA comprises at least one selected from the group consisting of: the first polynucleotide, the second polynucleotide, the fourth polynucleotide, an eighth polynucleotide, and any combination thereof, wherein the eighth polynucleotide encodes an RNA that is heterologous to the organelle;
 - iv. optionally, a fourth polynucleotide encoding at least one selectable marker or at least one screenable marker, or both; wherein the fourth polynucleotide is operably linked to a promoter that is functional in the organelle; and
 - v. optionally, a fifth polynucleotide encoding an origin of replication that is functional in the organelle.
21. A yeast cell, algal cell, plant cell, plant, seed, root, stem, leaf, flower, fruit, or bean comprising the recombinant DNA construct of claim 20.
22. A method for altering the genome of an organelle, the method comprising:
- a. introducing into a cell:
 - i. a polynucleotide encoding an RNA sequence comprising an organelle targeting RNA operably linked to a guide polynucleic acid, wherein the guide polynucleic acid directs a polynucleotide guided poly-

peptide to cleave a target sequence present in an organelle genome, wherein the polynucleotide is operably linked to at least one regulatory element; and either

- ii. a second polynucleotide encoding a modified polynucleotide guided polypeptide, wherein the second polynucleotide is operably linked to at least one regulatory element, and wherein the modified polynucleotide guided polypeptide comprises a polynucleotide guided polypeptide operably linked to an organelle targeting peptide; wherein the organelle targeting RNA of (i) and the organelle targeting peptide of (ii) each target the same organelle; or
 - iii. a third polynucleotide, wherein the third polynucleotide is operably linked to at least one regulatory element, wherein the third polynucleotide encodes an RNA molecule comprising an organelle targeting RNA operably linked to an RNA sequence encoding a polynucleotide guided polypeptide; wherein the organelle targeting RNA of (i) and the organelle targeting RNA of (iii) each target the same organelle; and
- b. growing the cell under conditions in which the polynucleotide of (i) and the second polynucleotide of (ii) or the third polynucleotide of (iii) are both expressed.

23. The method of claim **22**, further comprising introducing a polynucleotide comprising at least one donor polynucleotide into the organelle, wherein the at least one donor polynucleotide comprises at least one homologous sequence

with respect to the organelle genome, wherein integration of all or part of the at least one donor polynucleotide into the organelle genome results in removal of the target site of the guide polynucleic acid.

24. A method for altering a genome of an organelle, the method comprising:

- (a) introducing into an organelle of a cell:
 - (i) at least one guide RNA, wherein the at least one guide RNA directs a polynucleotide guided polypeptide to cleave at least one target sequence present in the genome of the organelle;
 - (ii) a polynucleotide guided polypeptide, wherein the polynucleotide guided polypeptide, when associated with the at least one guide RNA, cleaves the at least one target sequence; and
 - (iii) a replacement DNA; and
- (b) selecting a cell comprising an organelle comprising the replacement DNA.

25. The method of claim **24**, wherein the replacement DNA of step (a) part (iii) comprises fragments of organellar DNA or a complete organellar DNA from a cultivar, line, sub-species and other species and is distinct from the genome of the organelle of step (a).

26. The method of claim **24**, wherein the at least one target sequence is not present in the replacement DNA.

27. The method of claim **24**, wherein after step (a) part (ii) and prior to step (a) part (iii), a cell is selected in which the genome of the organelle has been eliminated.

* * * * *