(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2018/0320164 A1**

(43) **Pub. Date:** **Nov. 8, 2018**

Hay et al.

(54) **DNA SEQUENCE MODIFICATION-BASED GENE DRIVE**

(71) Applicant: **California Institute of Technology,** Pasadena, CA (US)

(72) Inventors: **Bruce A Hay**, Pasadena, CA (US); **Georg Oberhofer**, Pasadena, CA (US); **Tobin William Ivy**, Pasadena, CA (US)

(21) Appl. No.: **15/970,728**

(22) Filed: **May 3, 2018**

Related U.S. Application Data

(60) Provisional application No. 62/502,338, filed on May 5, 2017.

Publication Classification

(51) **Int. Cl.**
C12N 15/10	(2006.01)
C12N 15/90	(2006.01)
C12N 5/14	(2006.01)
C12N 5/16	(2006.01)

(52) **U.S. Cl.**
CPC *C12N 15/102* (2013.01); *C12N 15/905* (2013.01); *C12N 15/80* (2013.01); *C12N 5/14* (2013.01); *C12N 5/16* (2013.01); *C12N 15/907* (2013.01)

(57) **ABSTRACT**
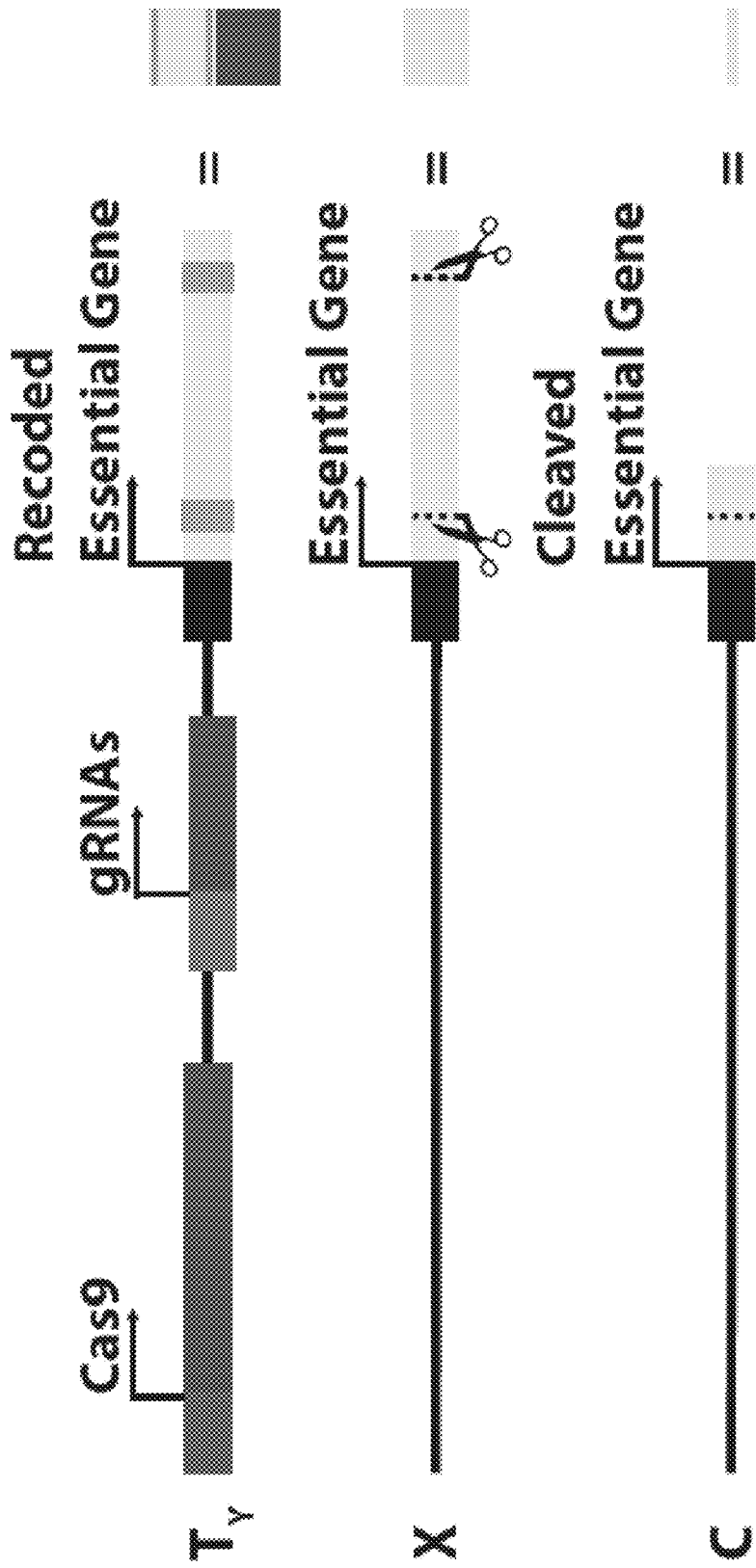
Described herein are embodiments relating to manipulation of populations and sex ratio in populations through DNA sequence modifications.
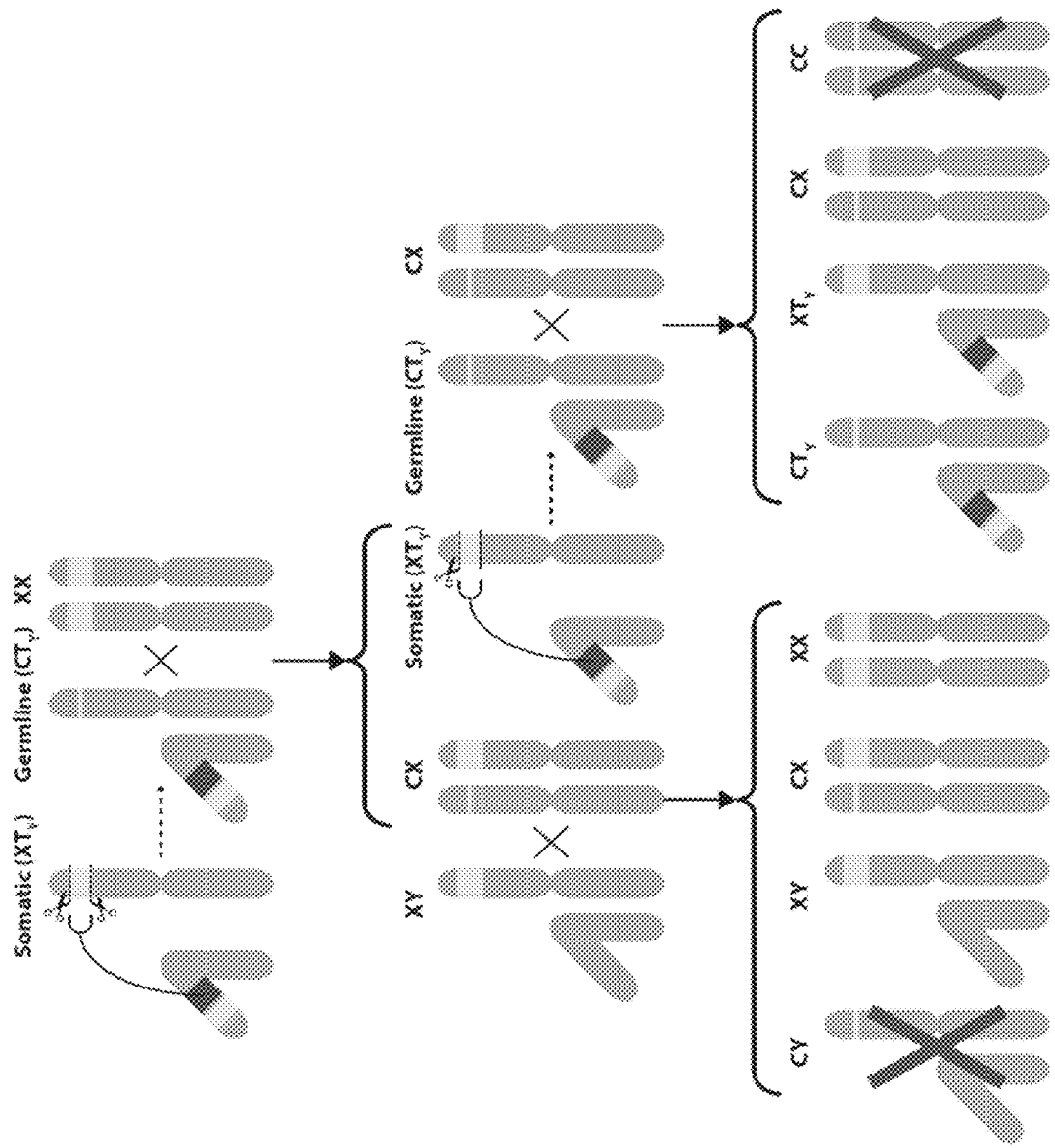
Specification includes a Sequence Listing.

FIG. 1A

FIG. 1B

FIG. 1C
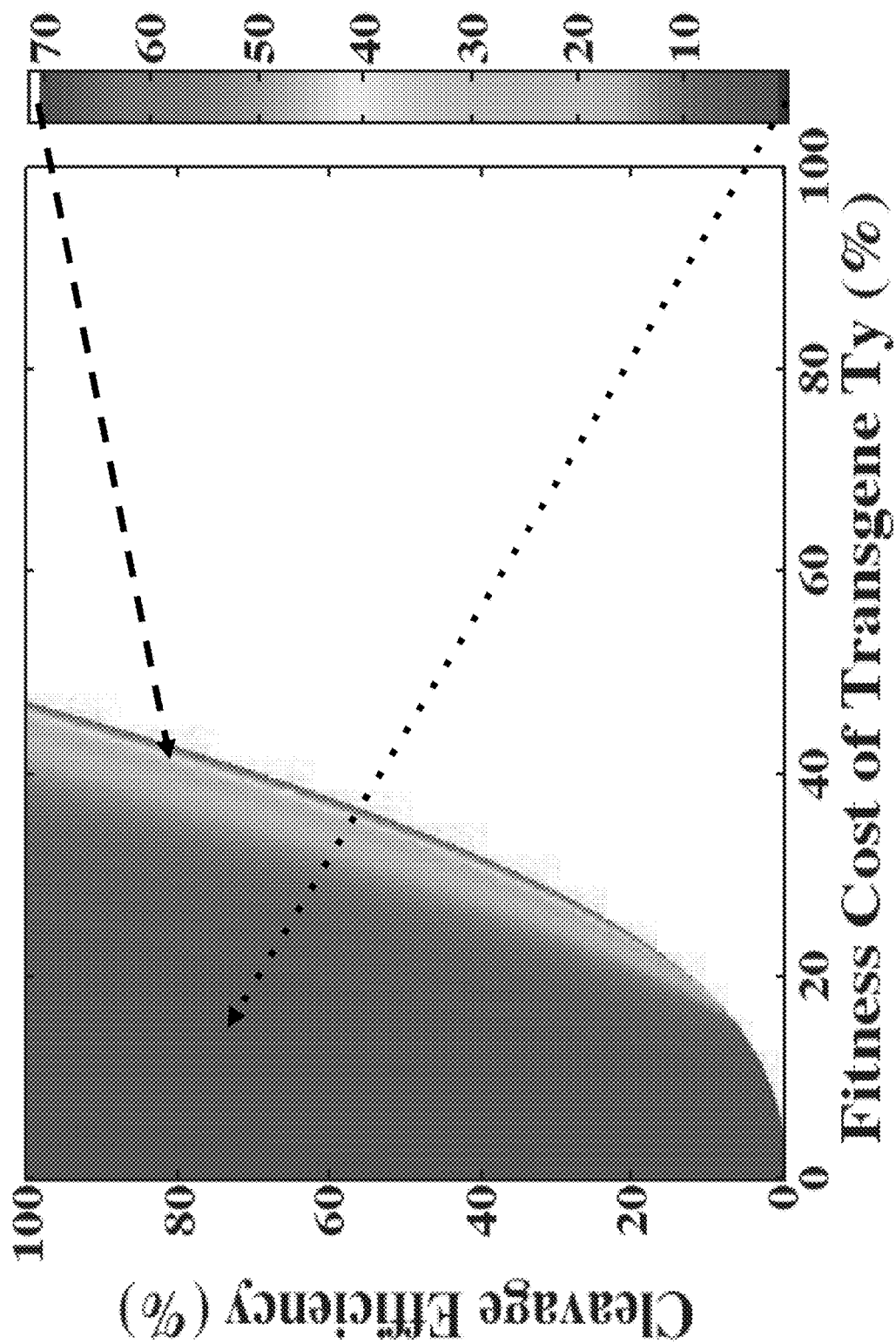
FIG. 2A

FIG. 2B

FIG. 2C
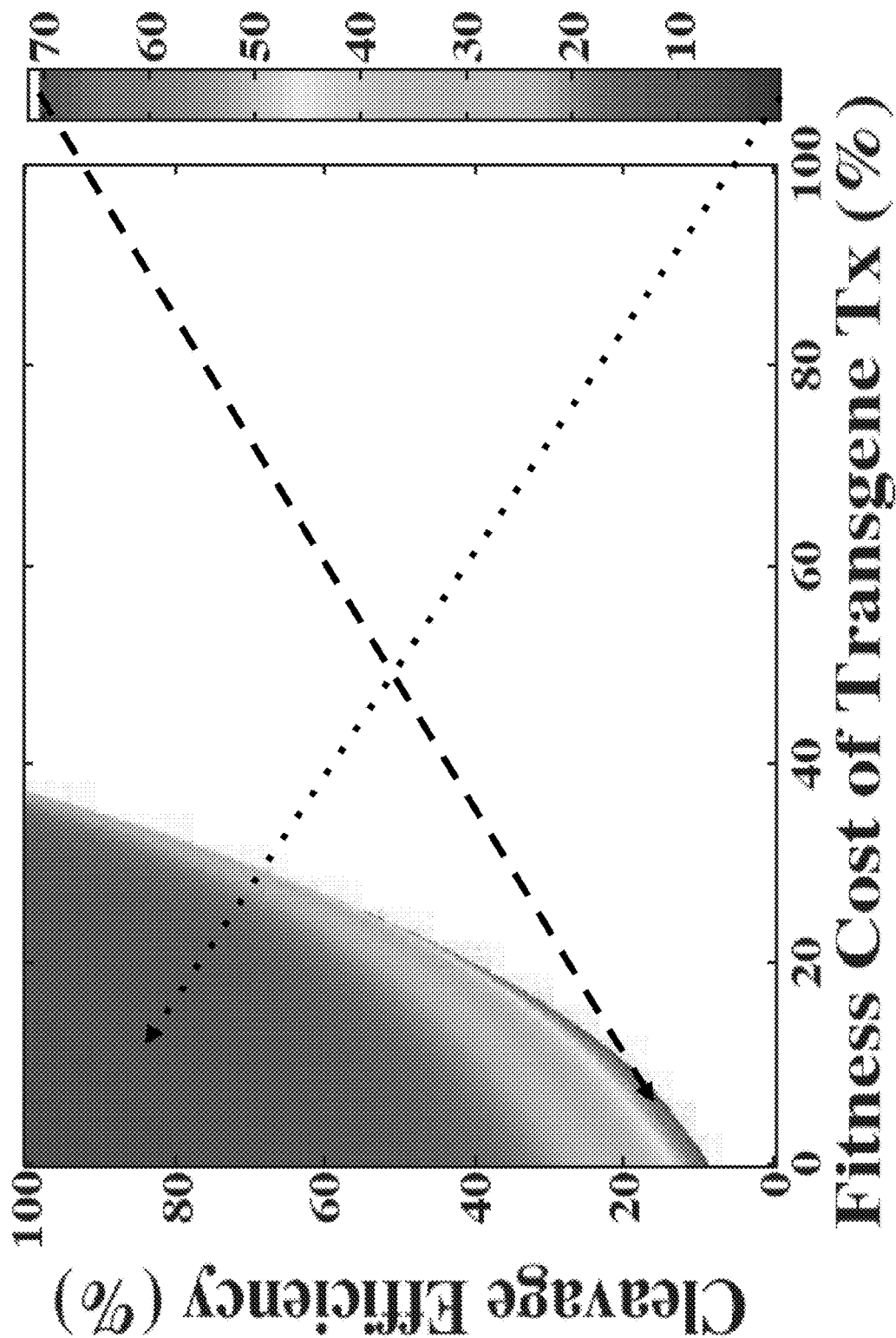
FIG. 3A

FIG. 3B

FIG. 3C
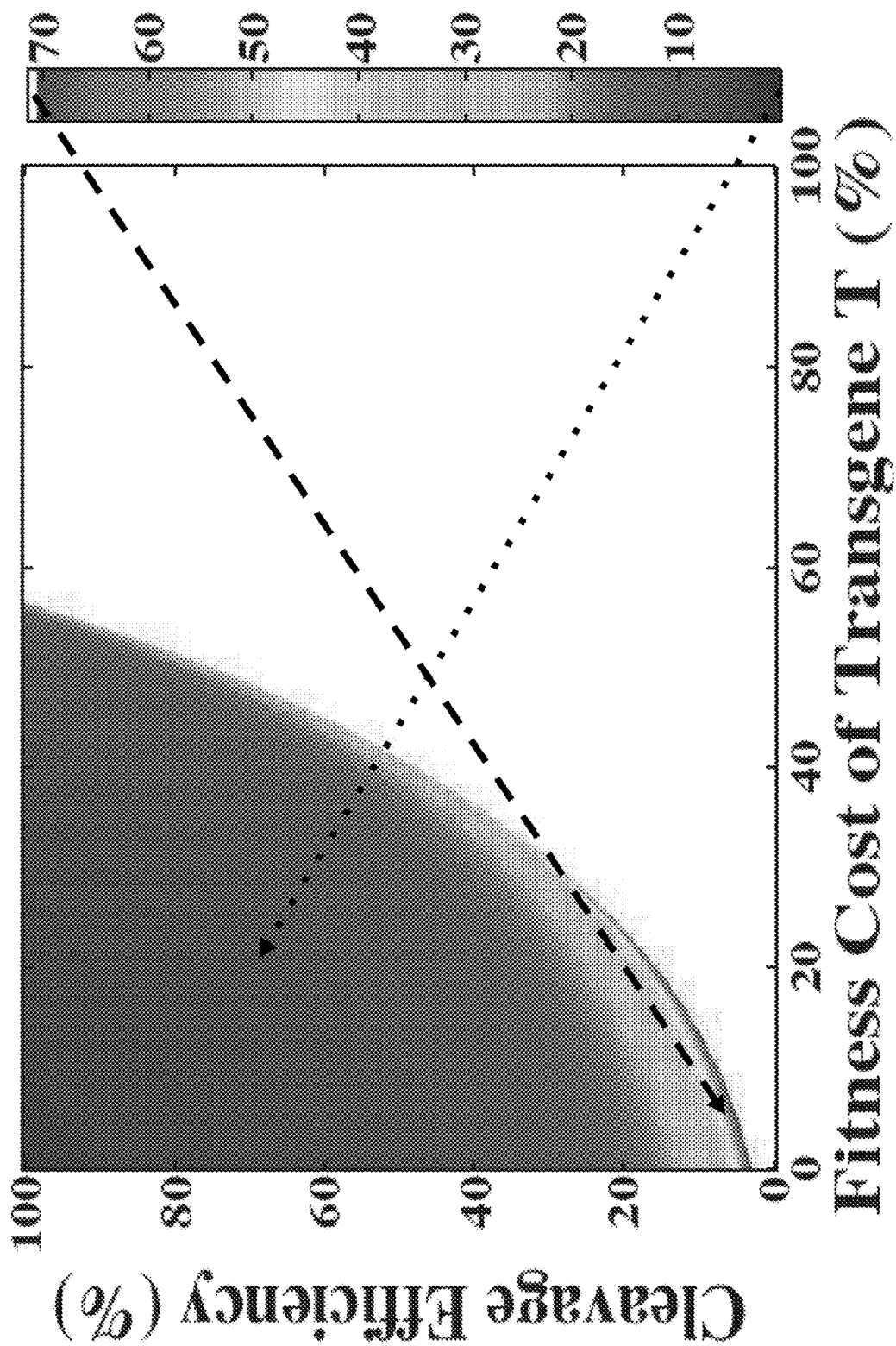
FIG. 4A
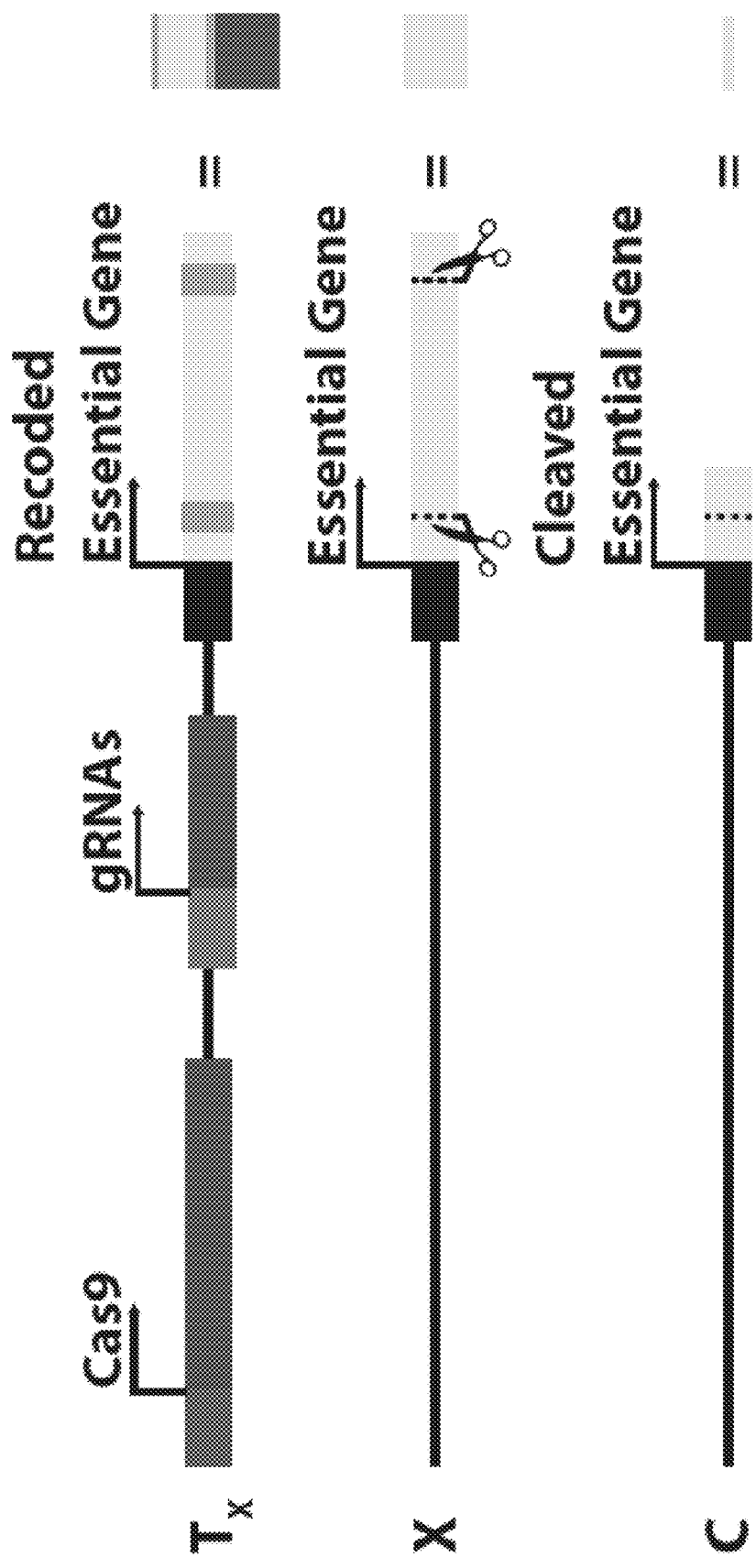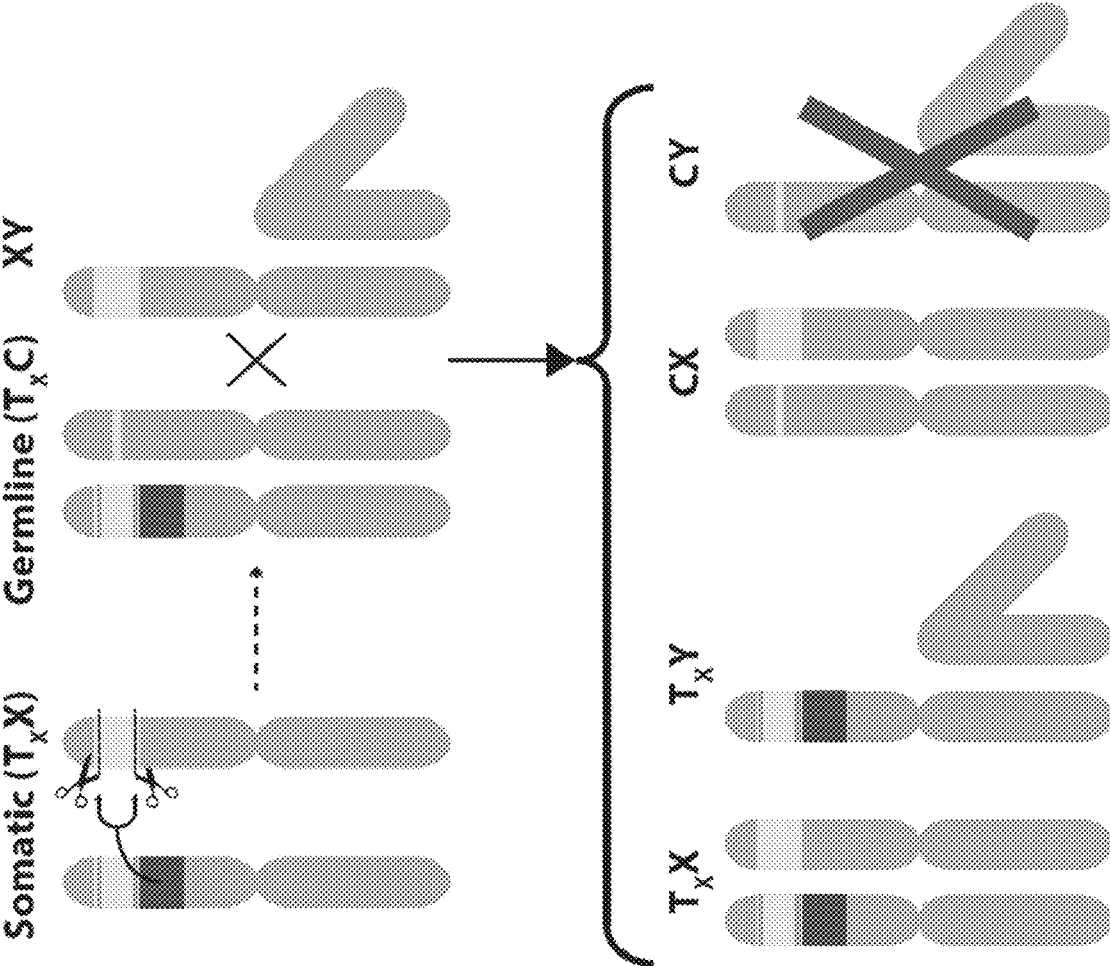
FIG. 4B

FIG. 4C

FIG. 5A

FIG. 5B

FIG. 5C

FIG. 6A – FIG. 6F

FIG. 6A

FIG. 6B

FIG. 6C

FIG. 6D

FIG. 6E

FIG. 6F

FIG. 7

FIG. 8A
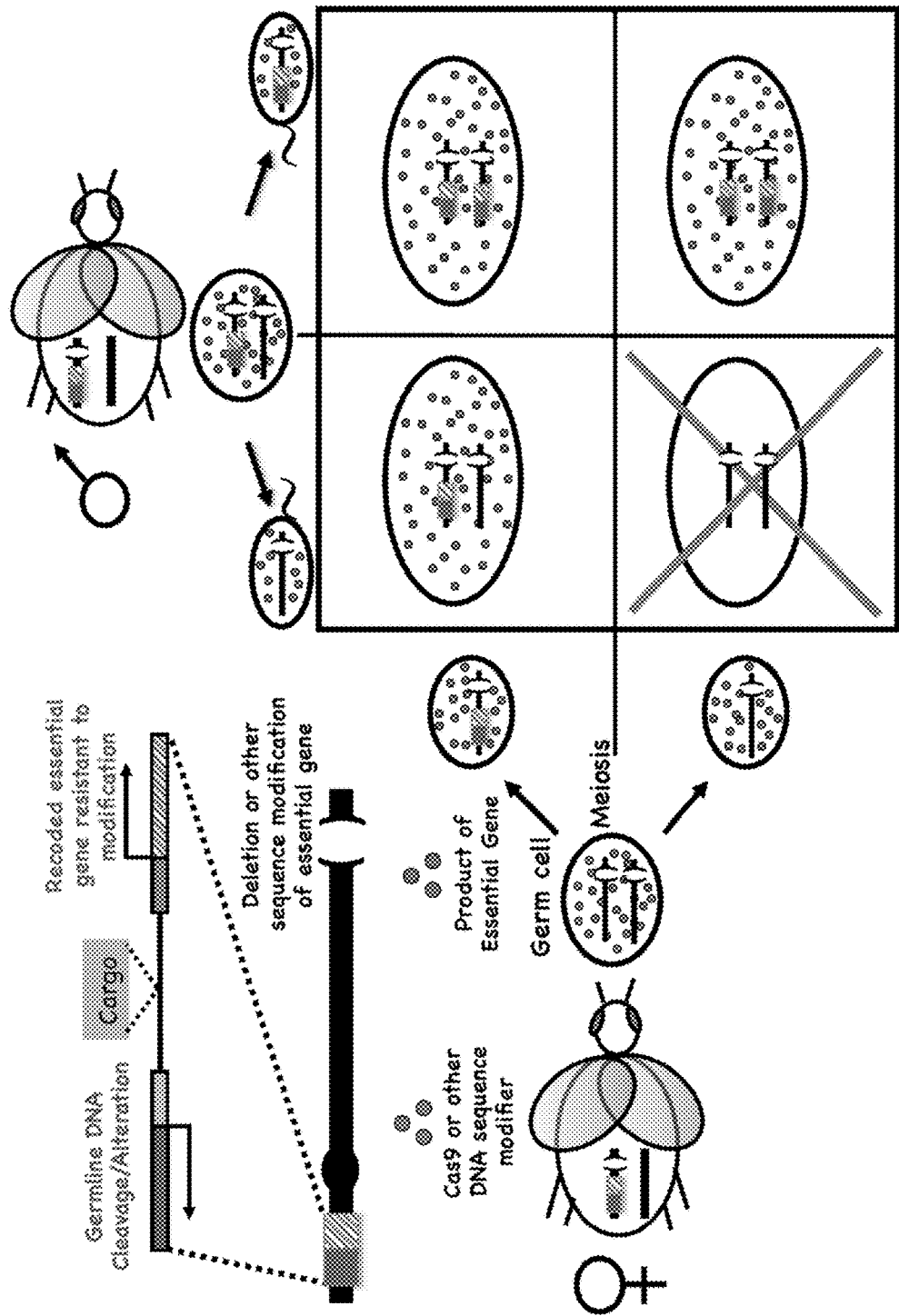
FIG. 8B
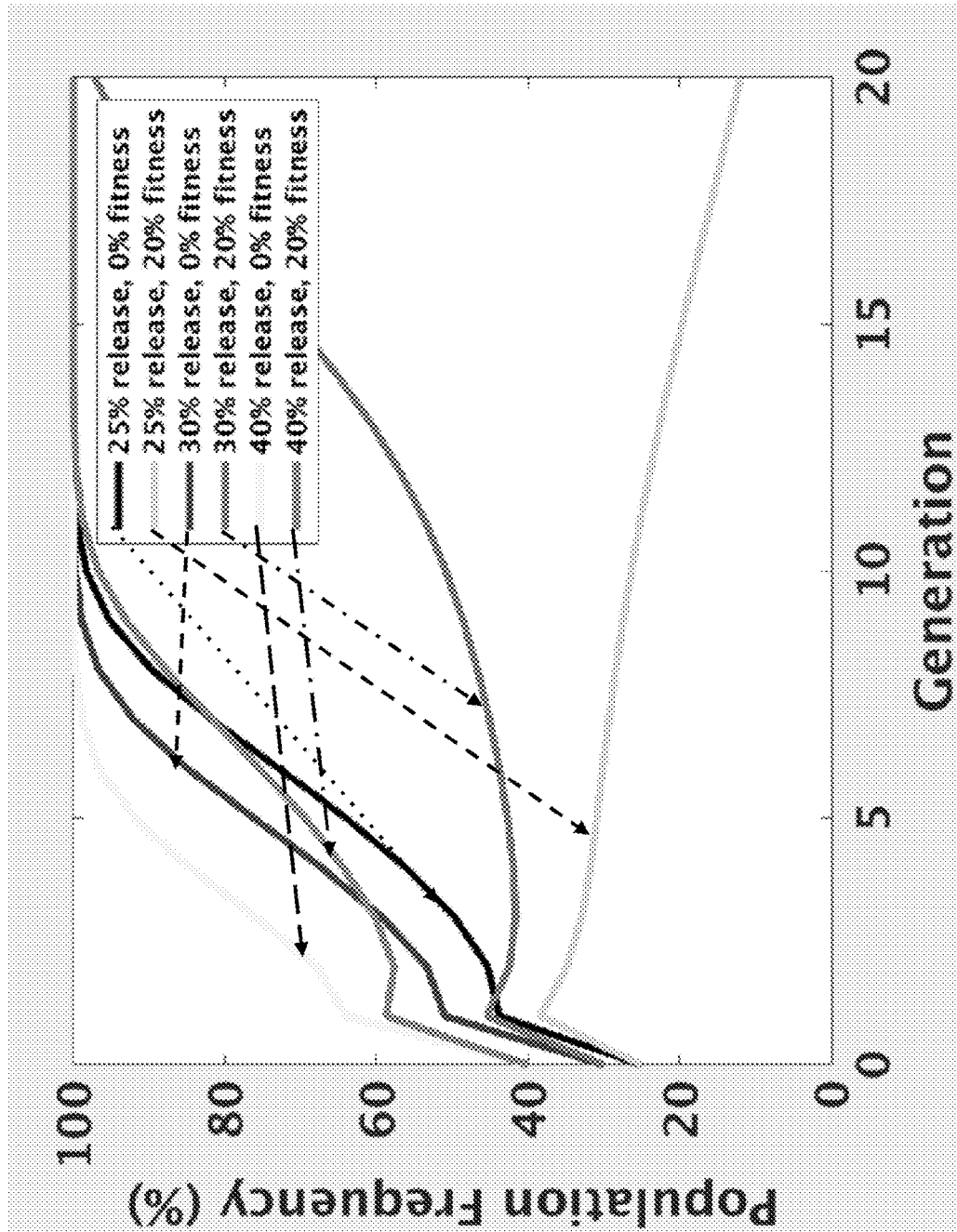
FIG. 9A
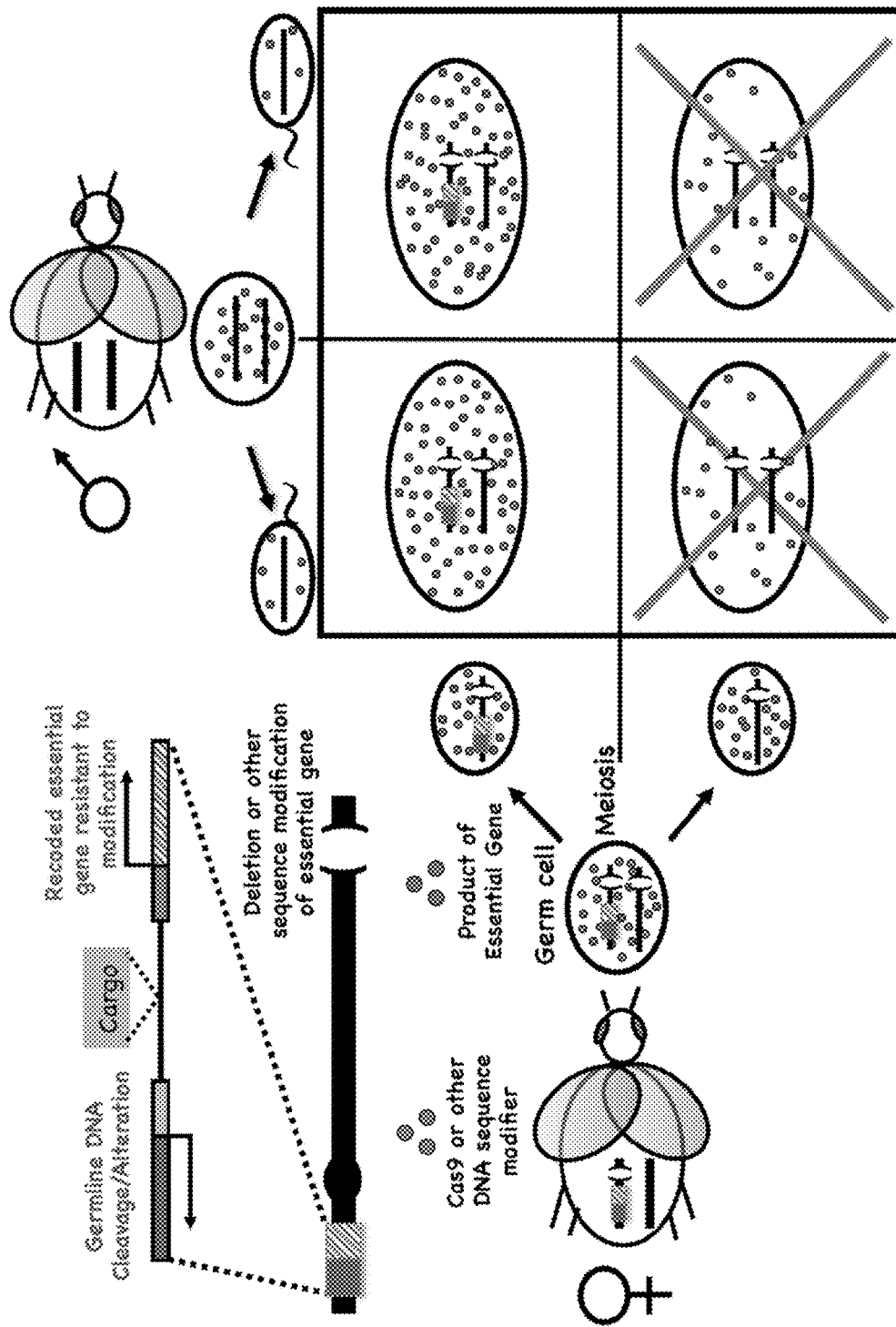
FIG. 9B
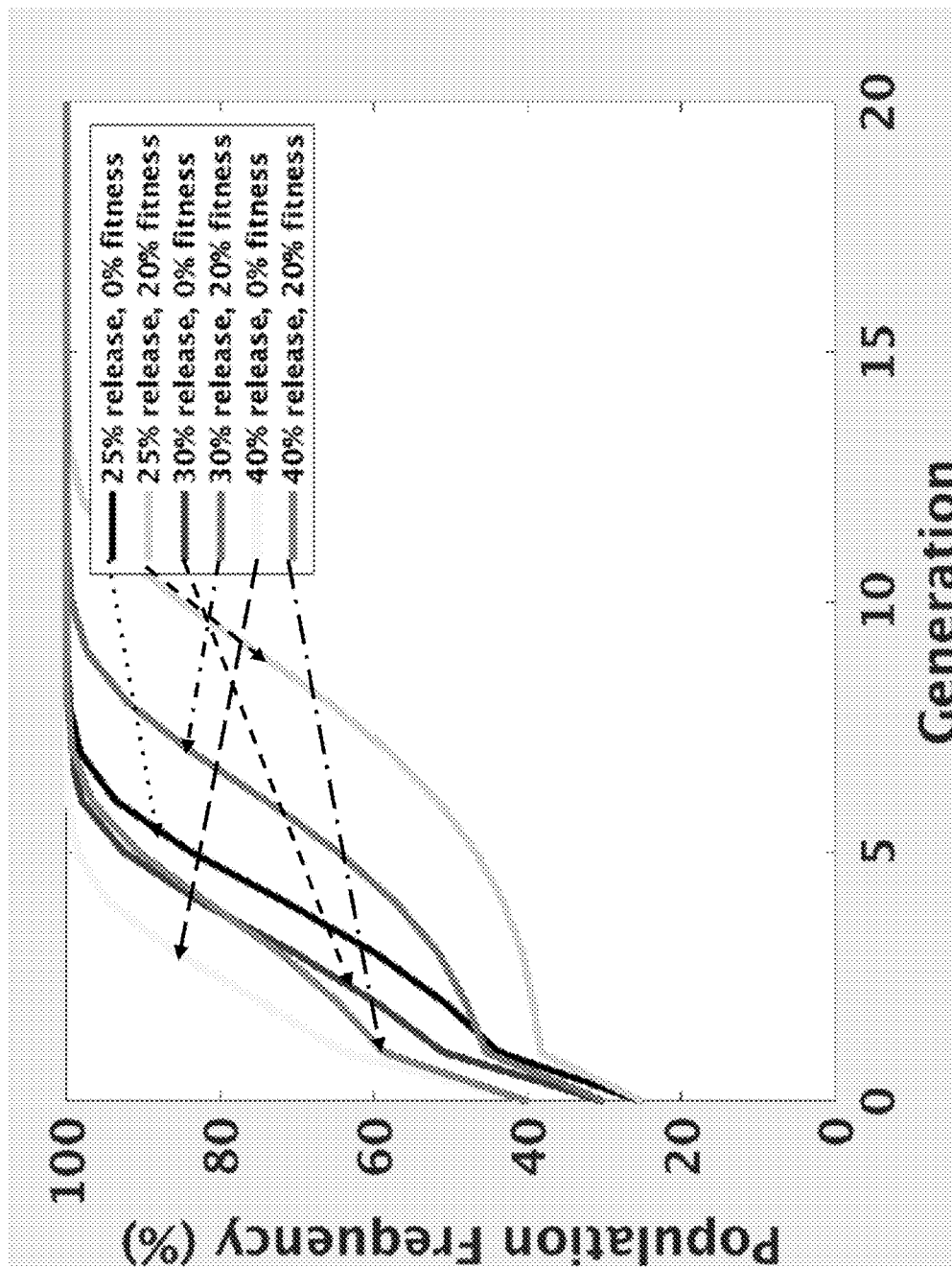
FIG. 10
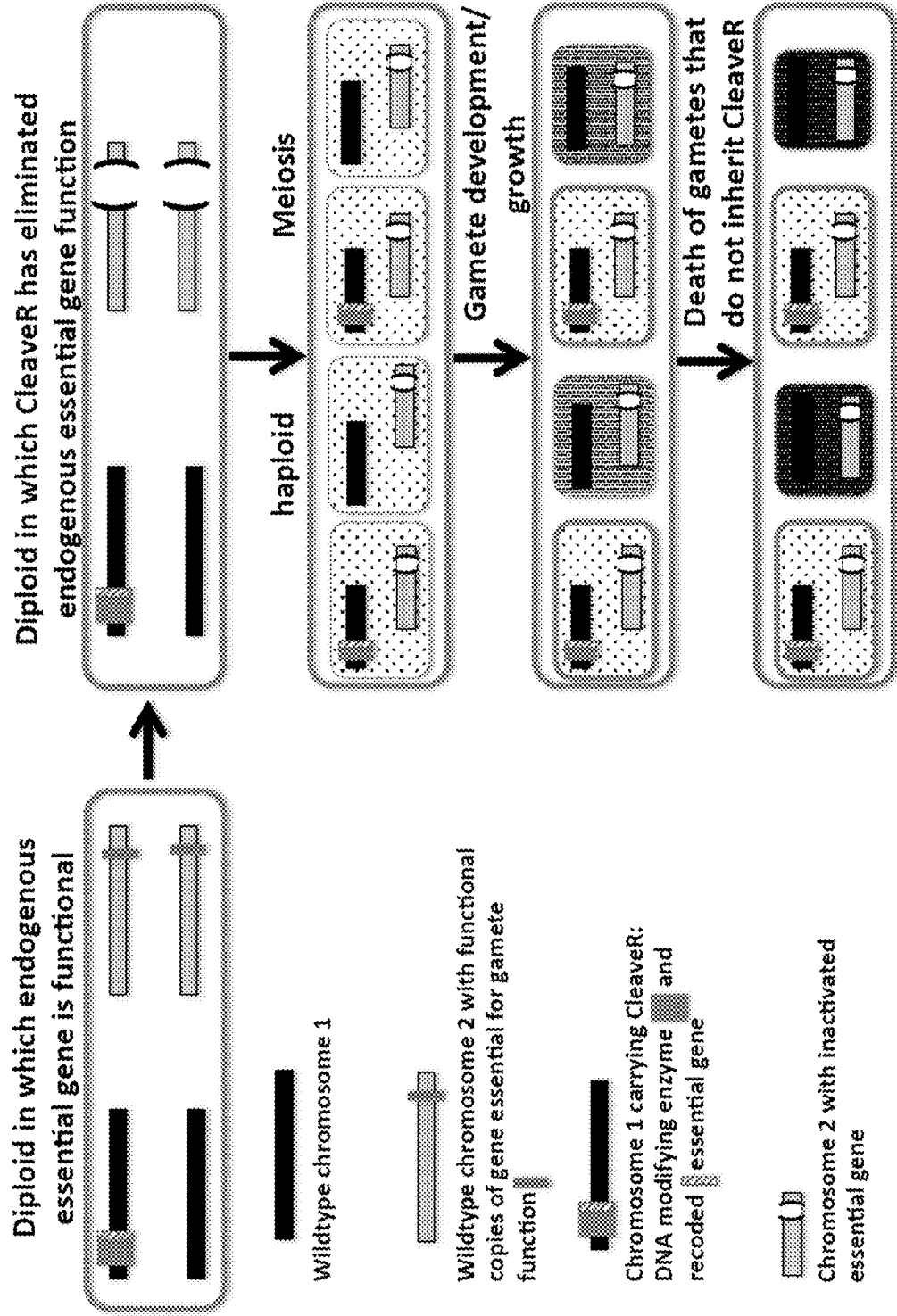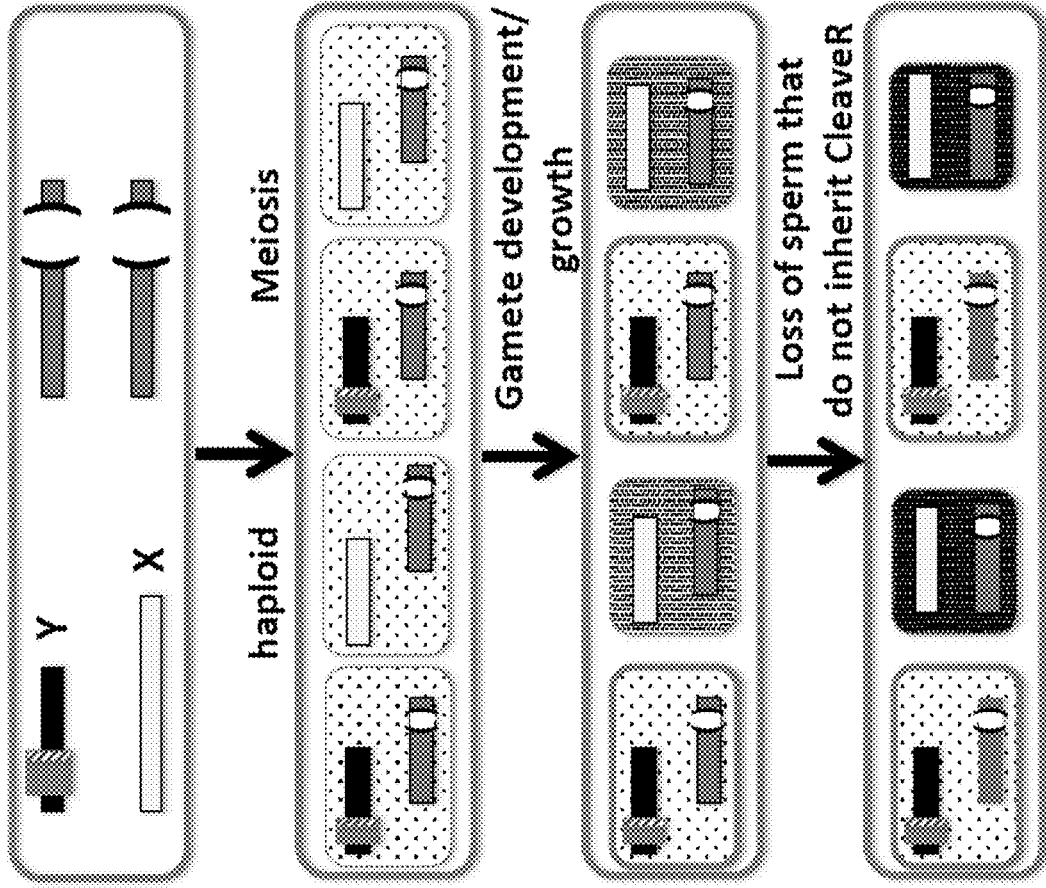
FIG. 11

FIG. 12
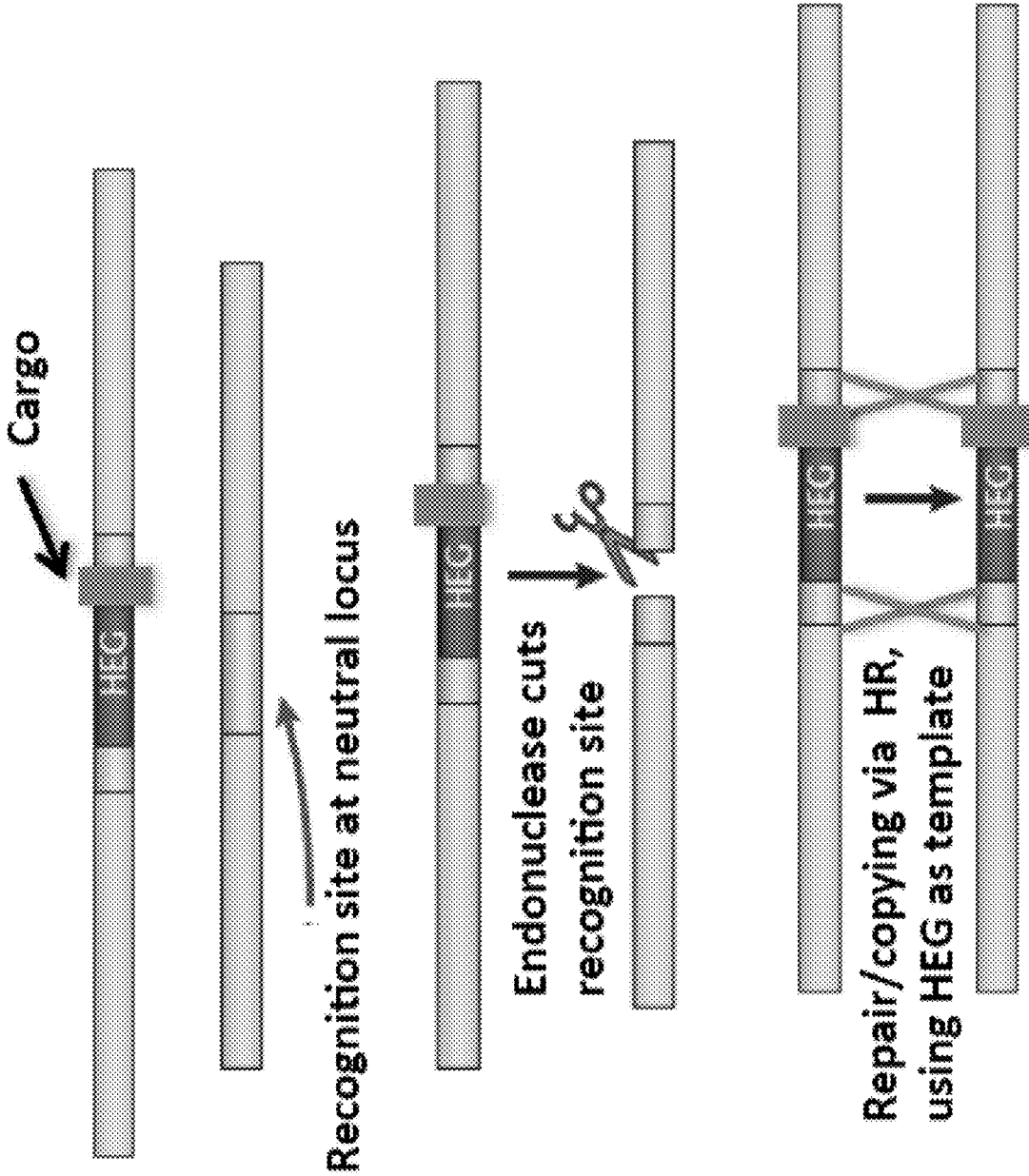
FIG. 13
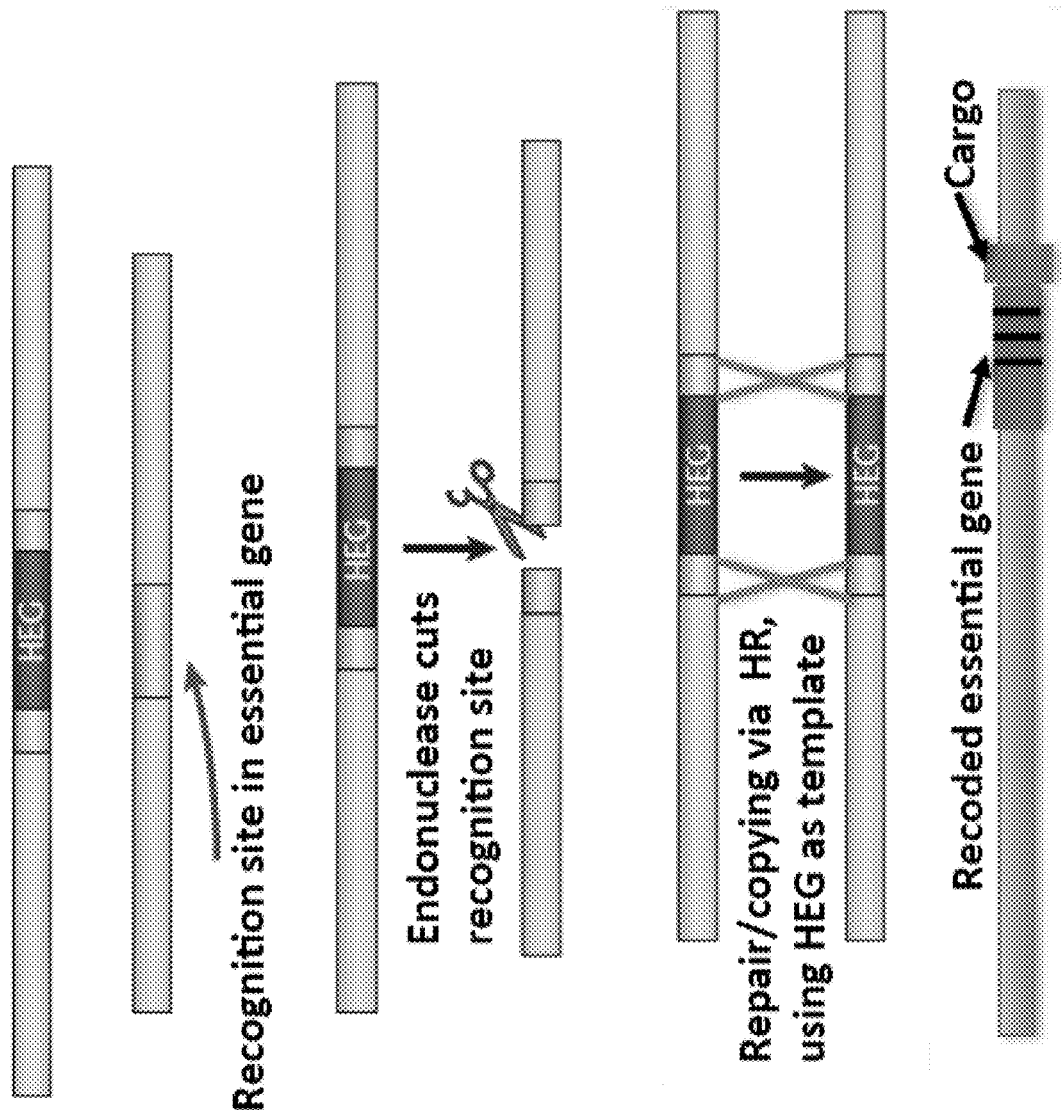
FIG. 14
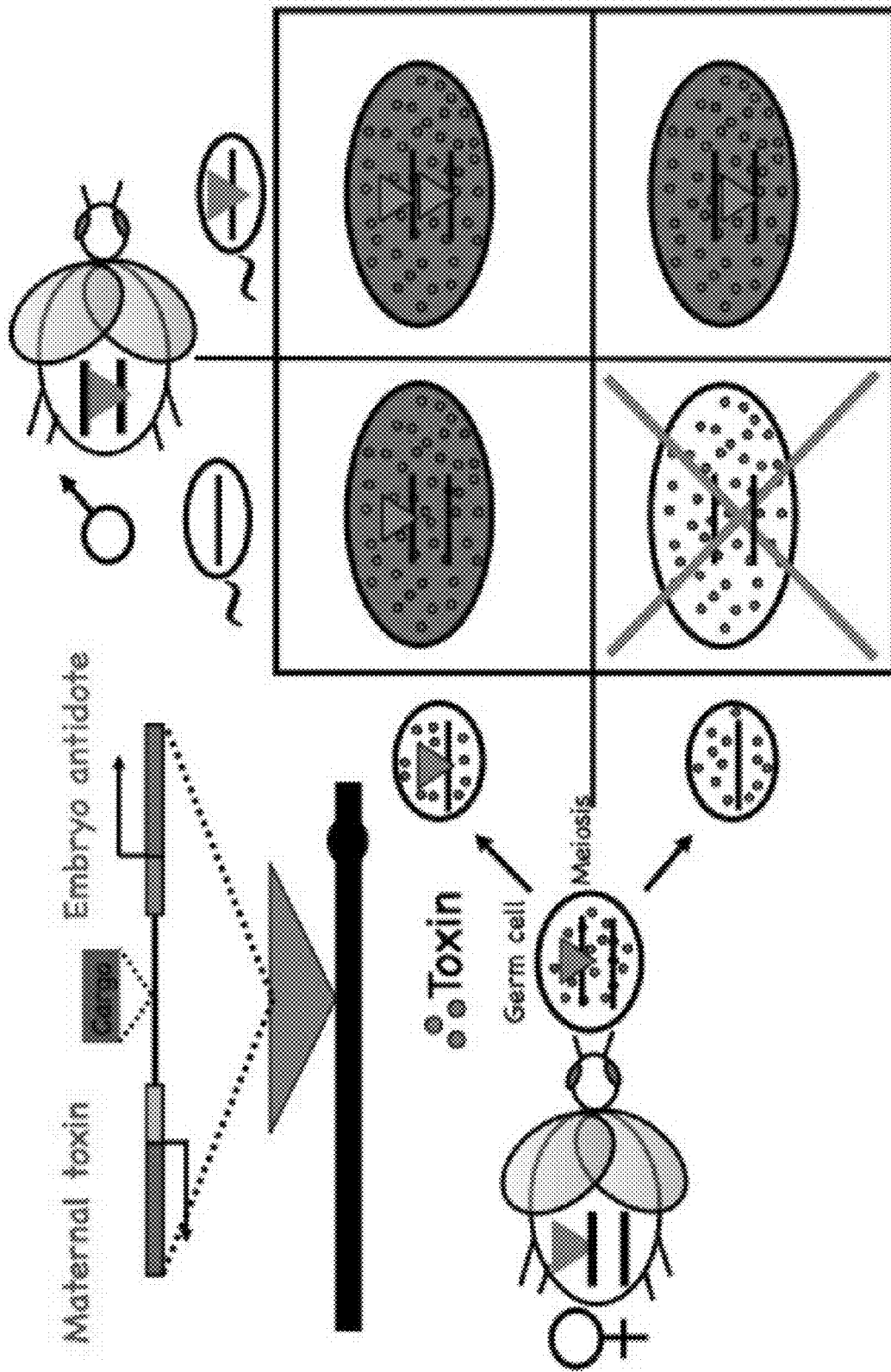
FIG. 15A – FIG. 15C
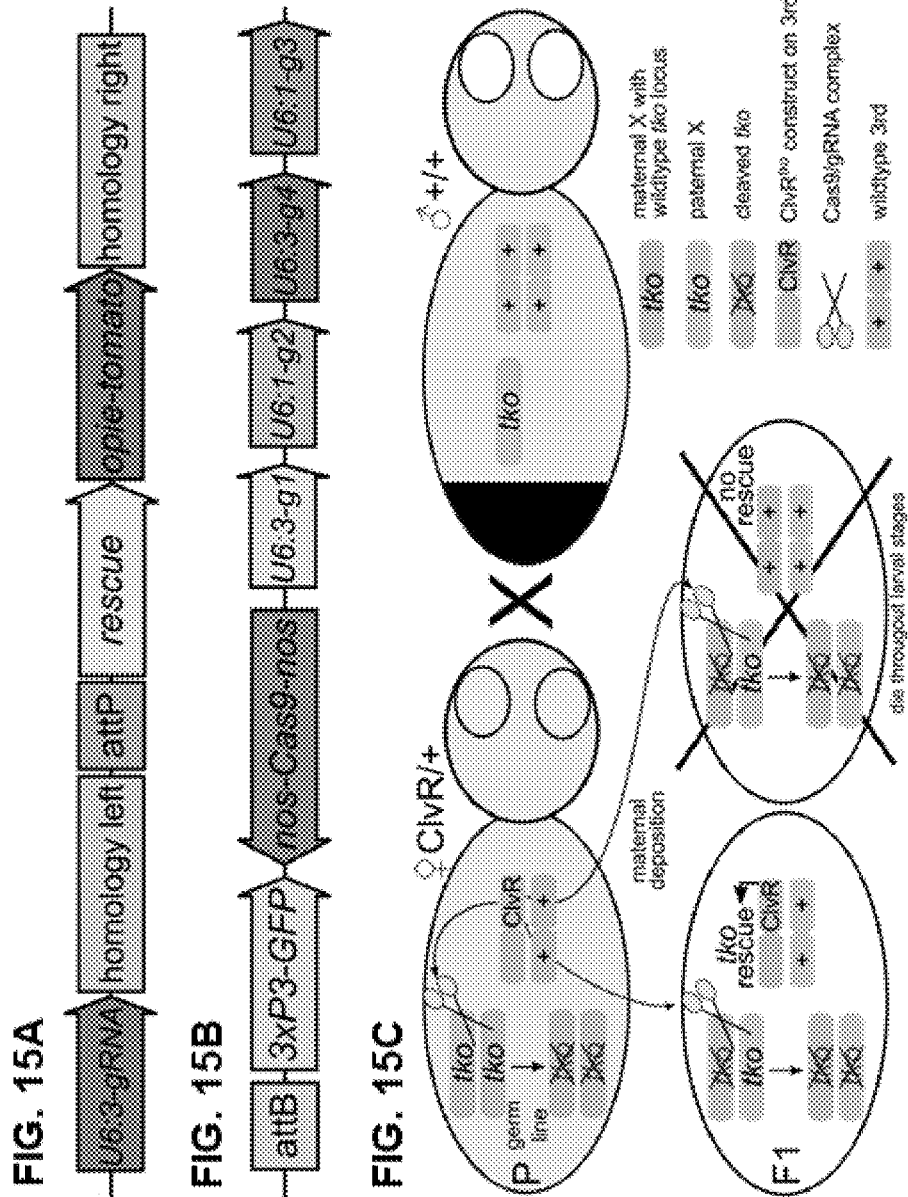
FIG. 15A

FIG. 15B

FIG. 15C

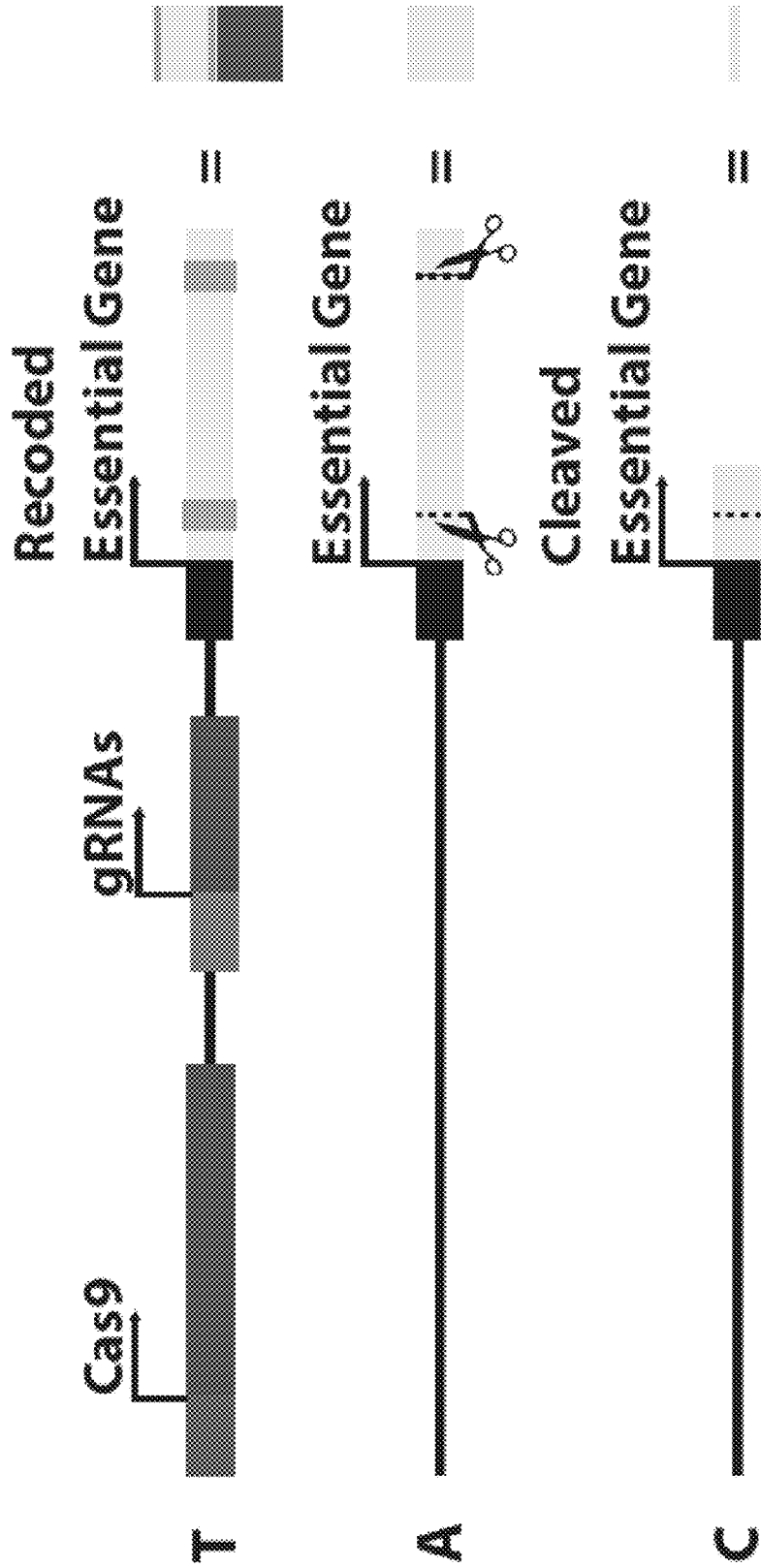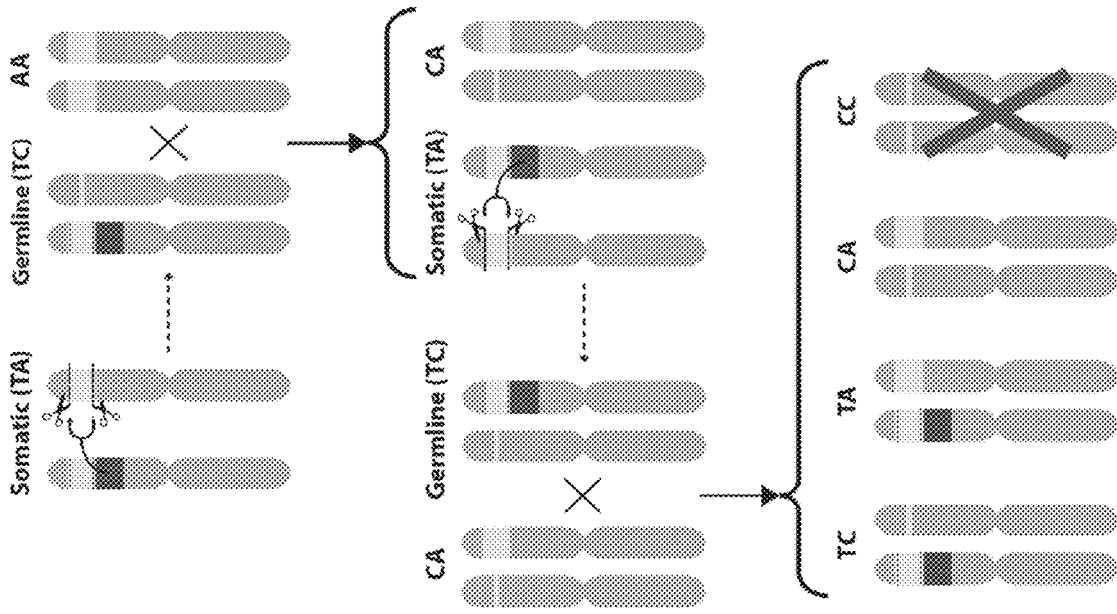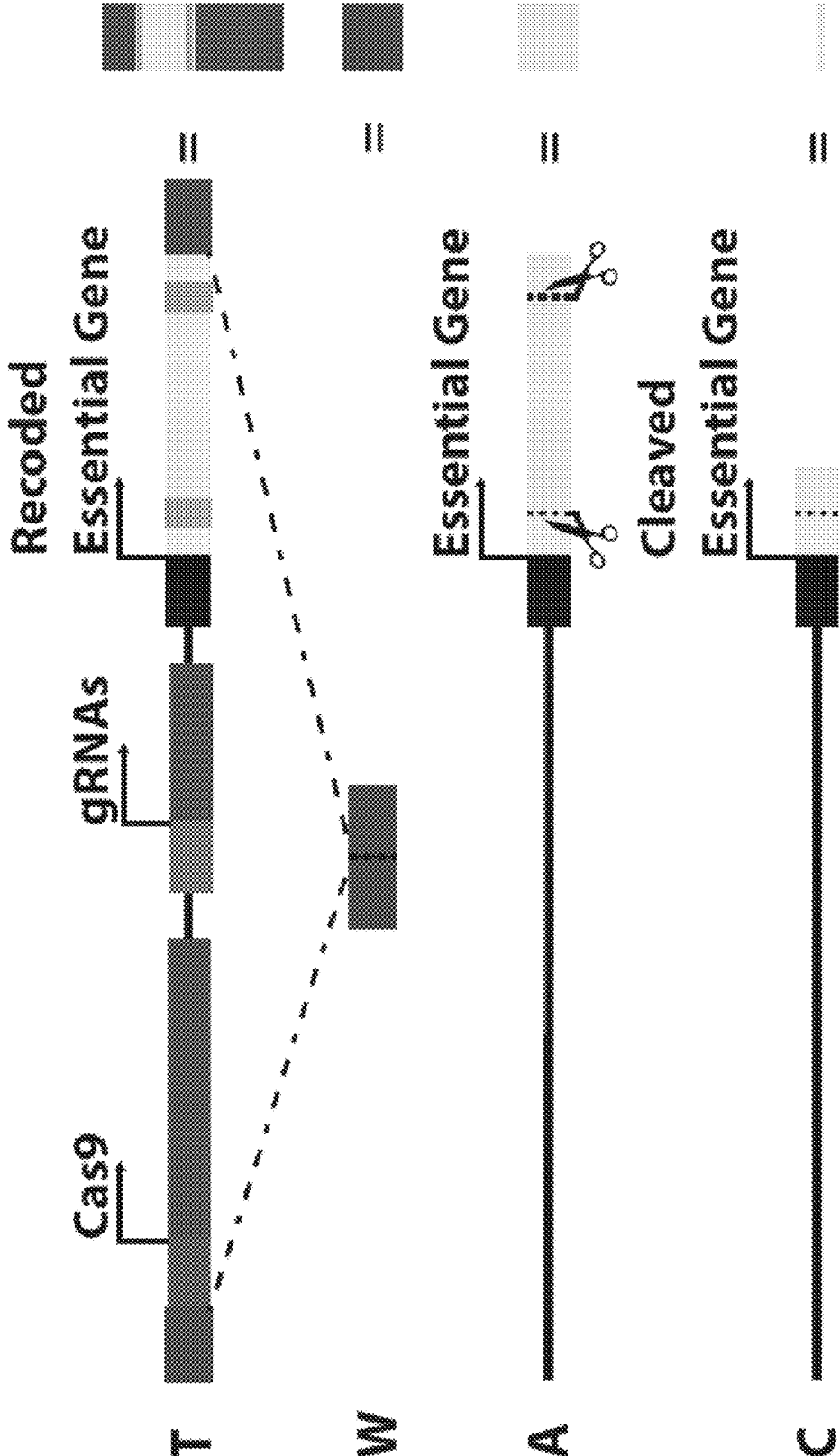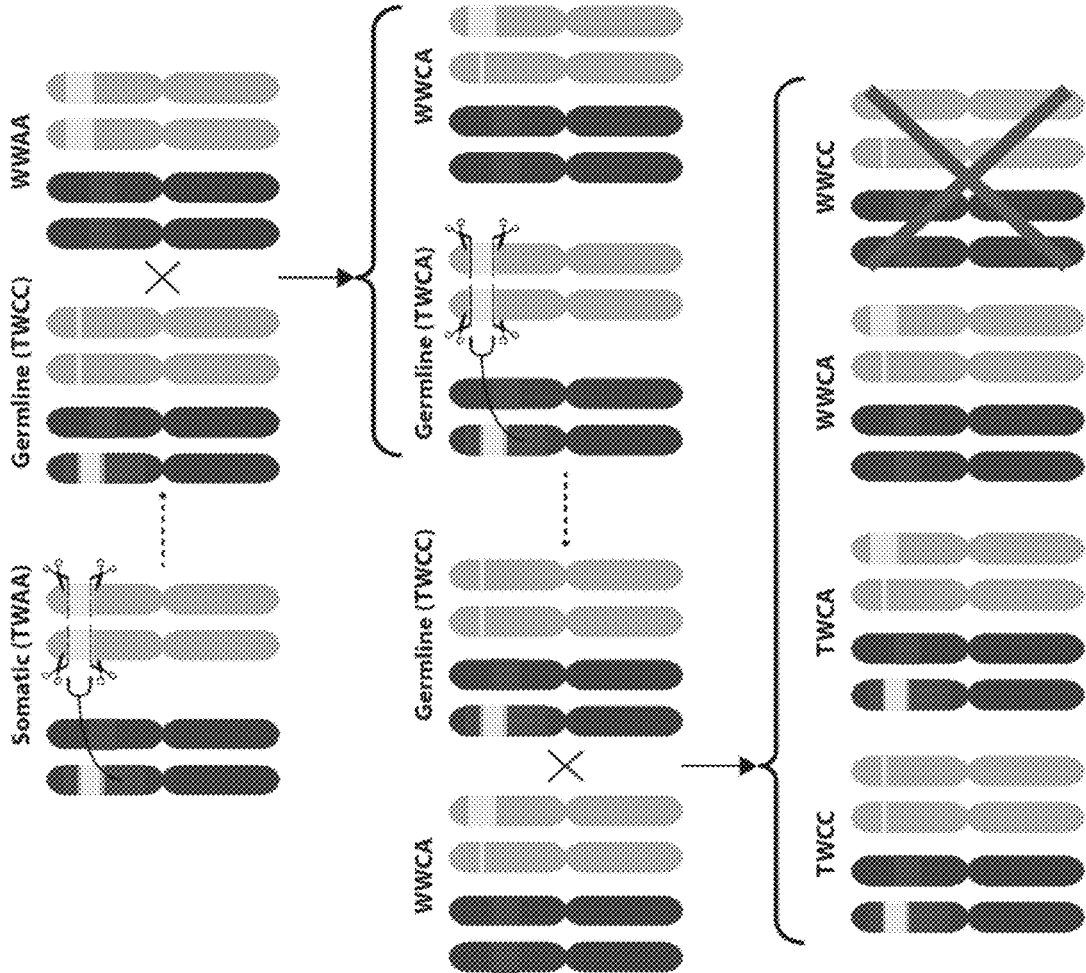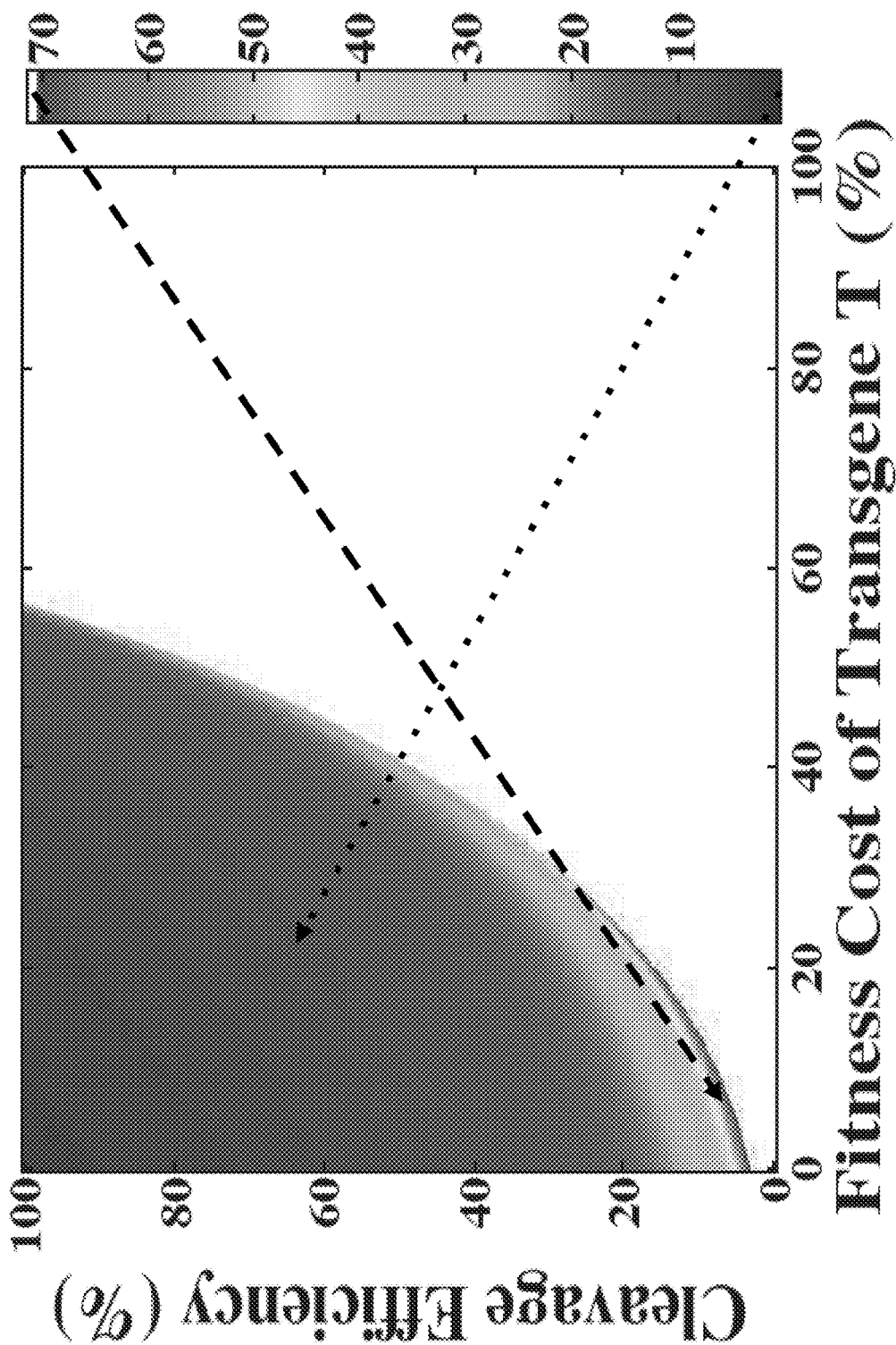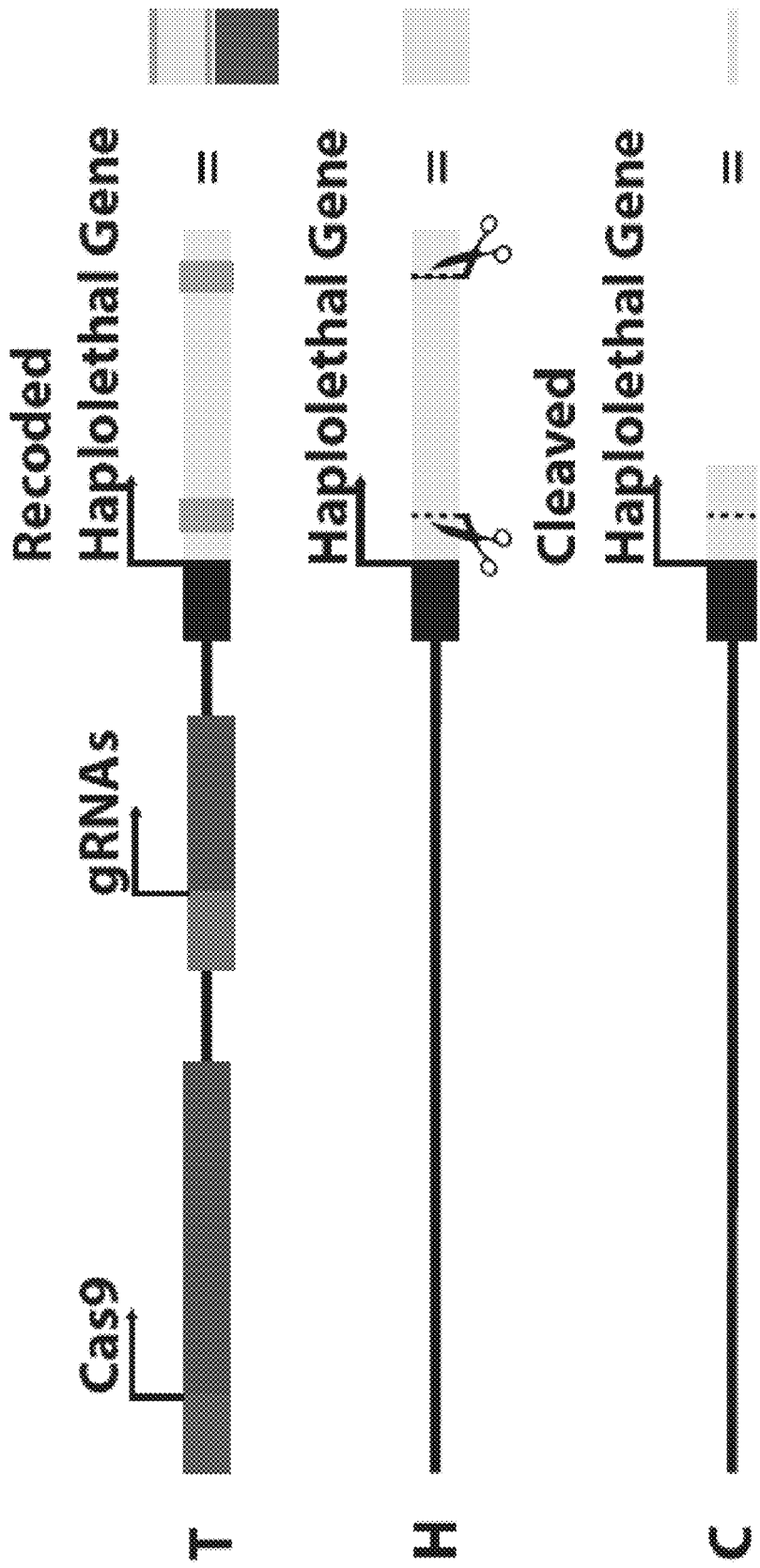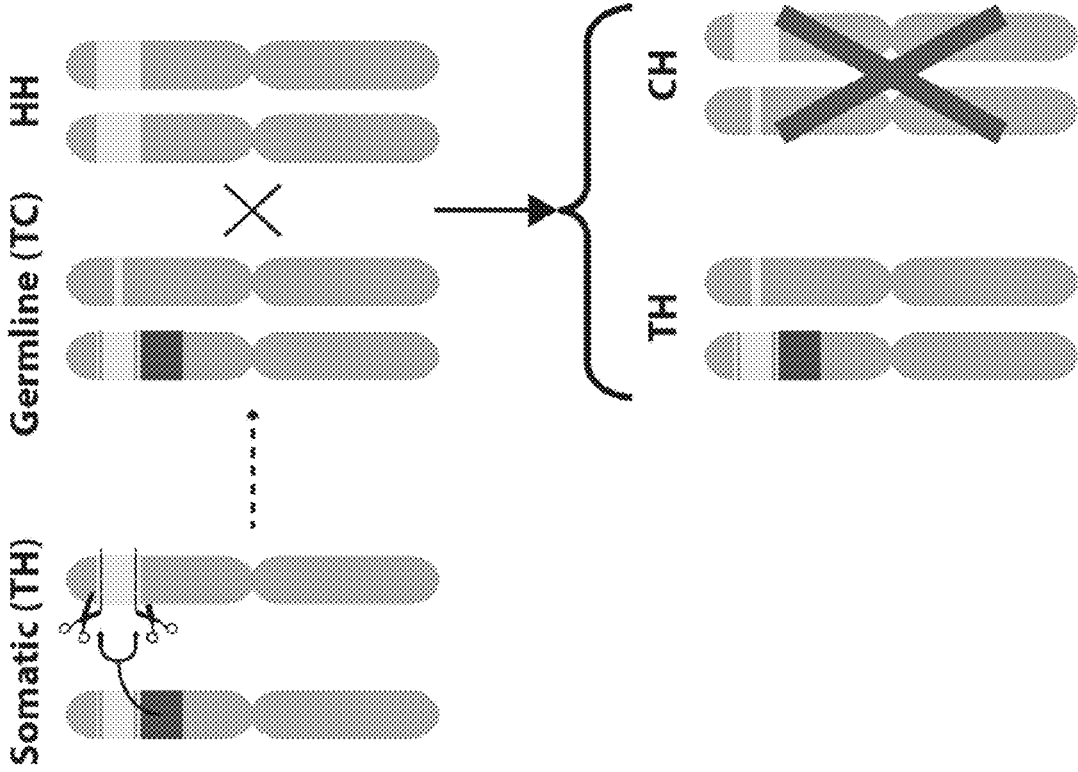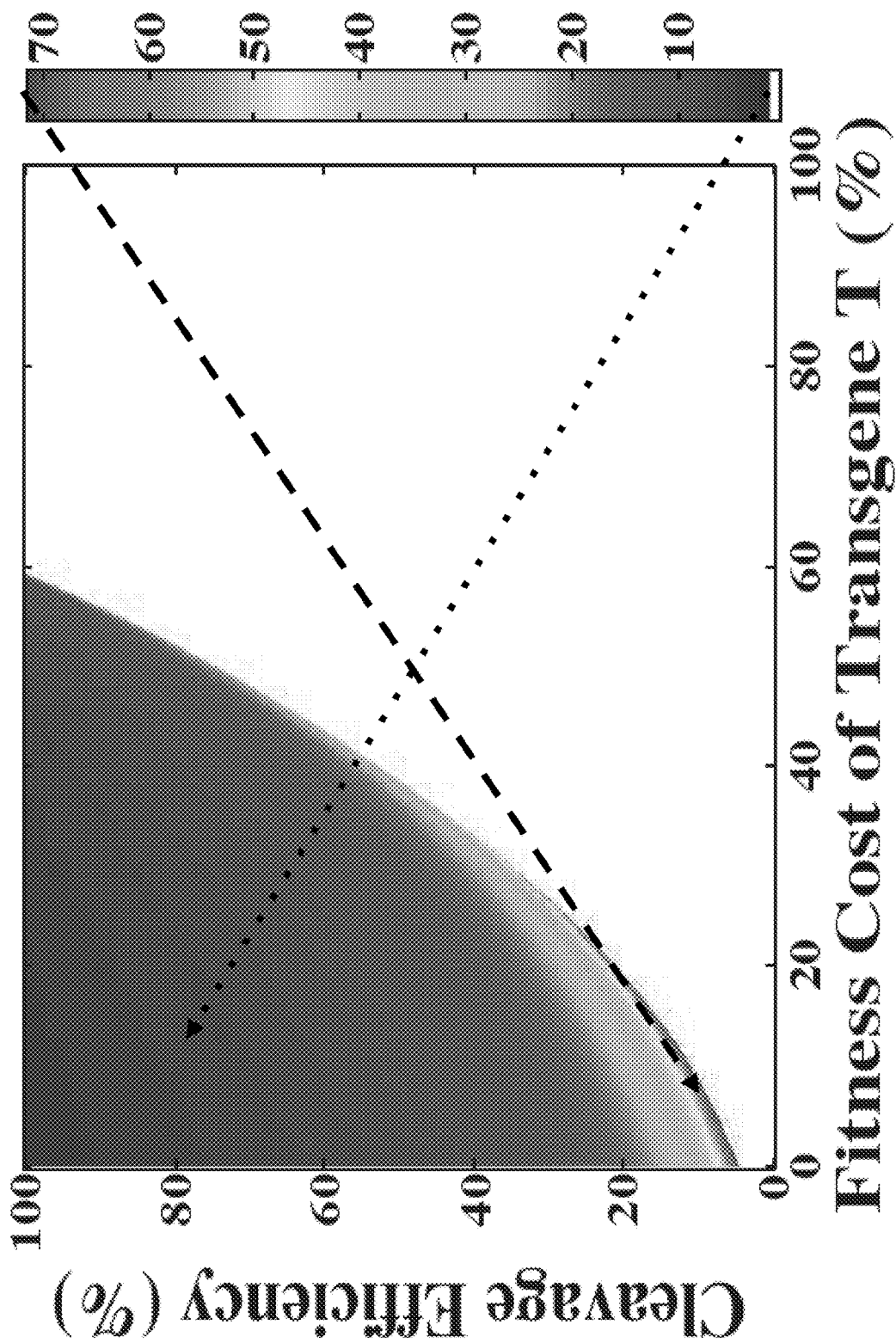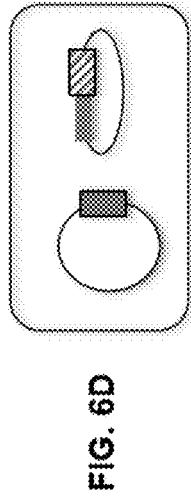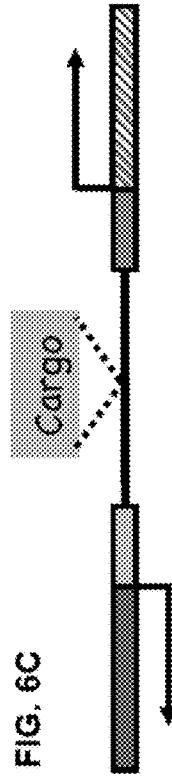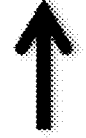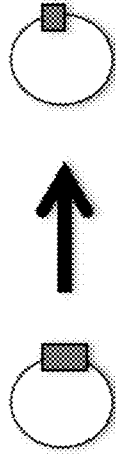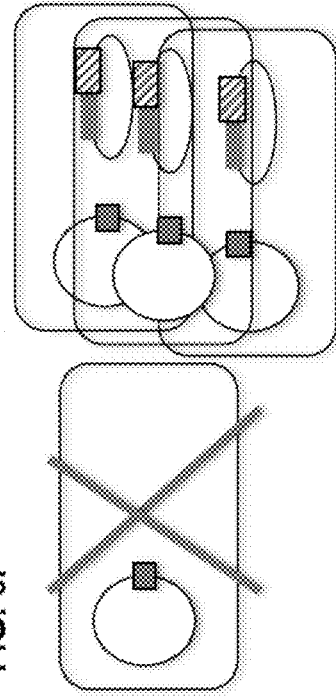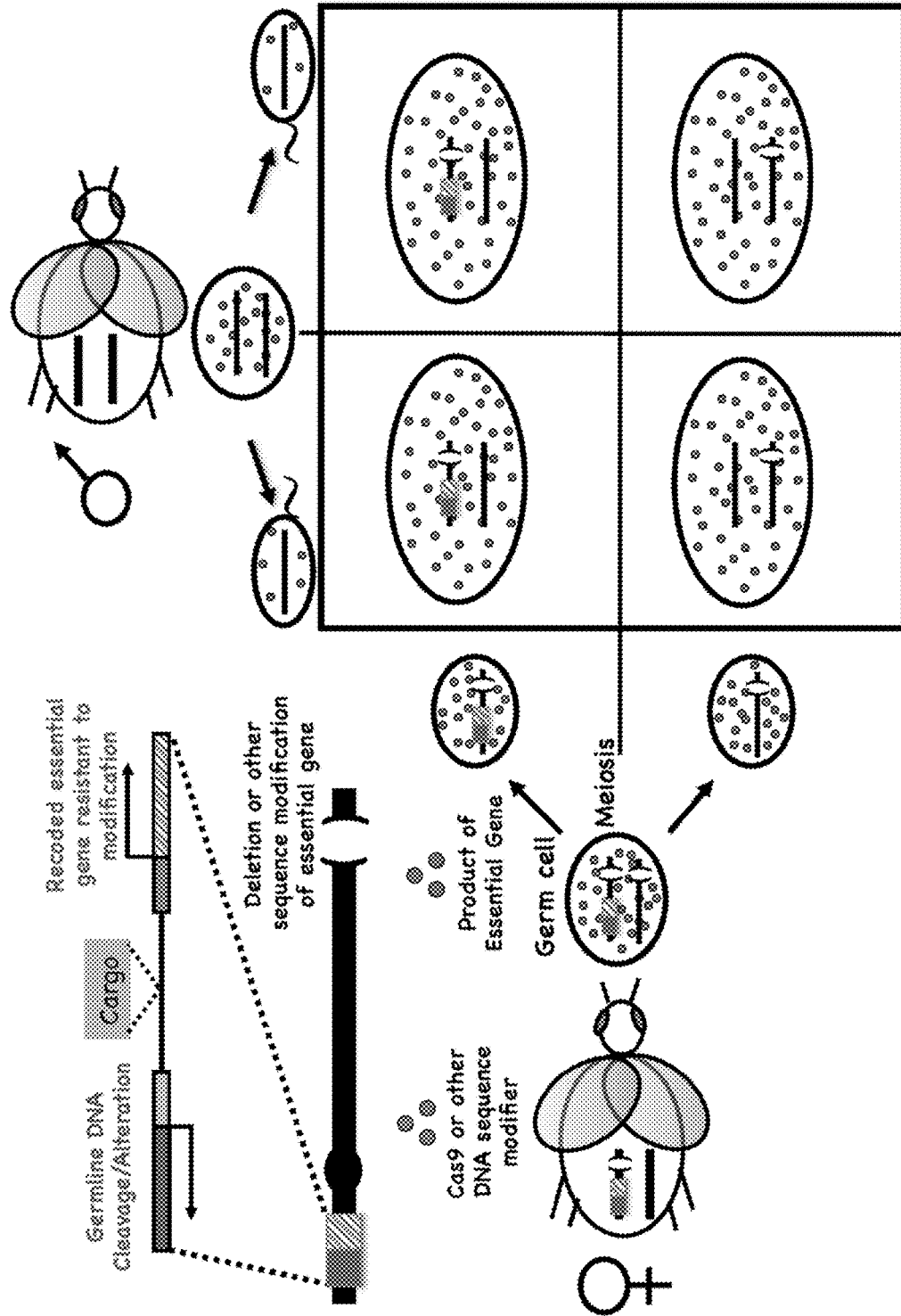FIG. 16A – FIG. 16B
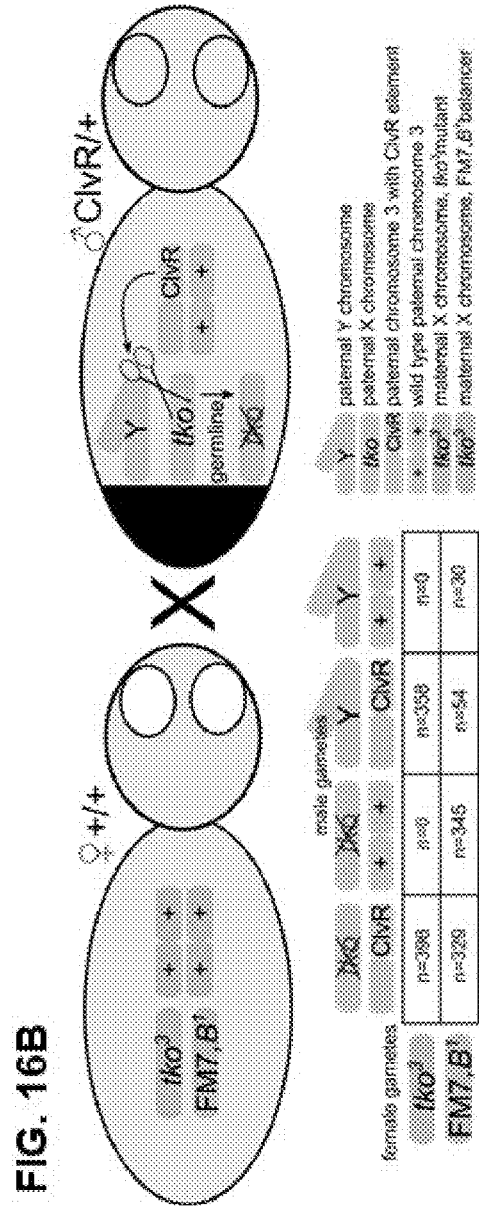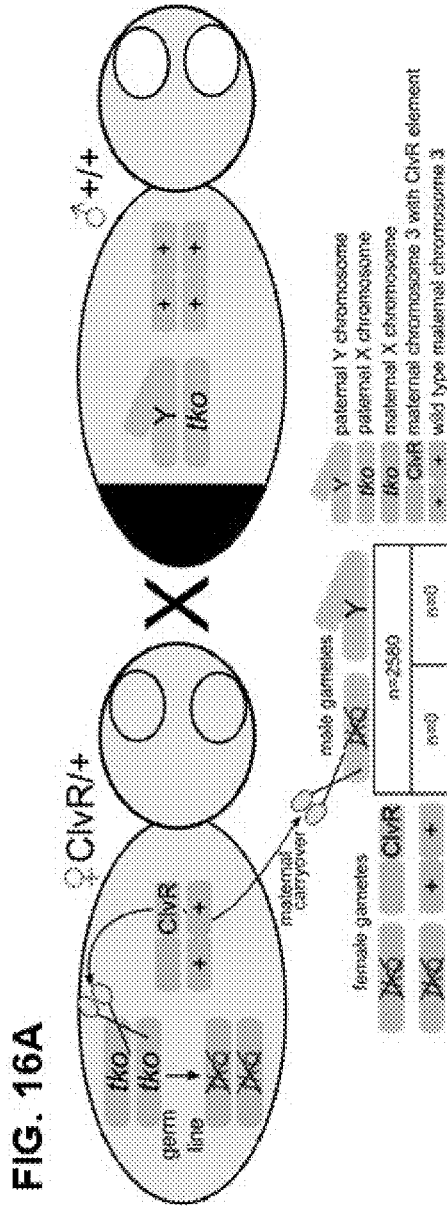
FIG. 16A
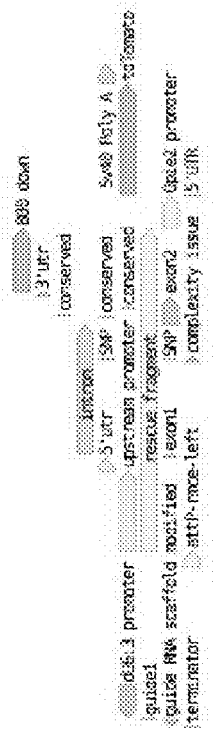
FIG. 16B

FIG. 17

FIG. 18A

FIG. 18B

FIG. 18C

FIG. 19
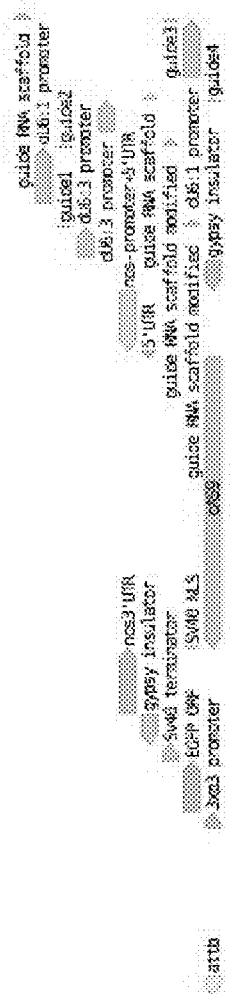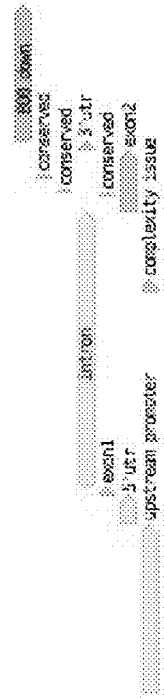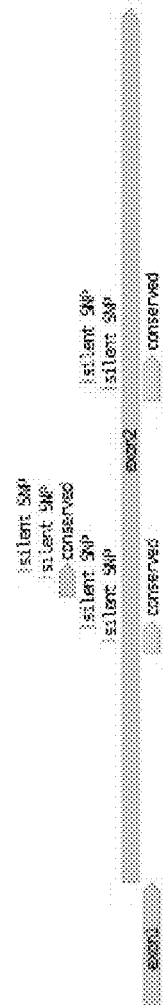
DNA SEQUENCE MODIFICATION-BASED GENE DRIVE

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application 62/502,338 filed on May 5, 2017, which is hereby incorporated by reference in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED R&D

[0002] This invention was made with government support from the US Department of Agriculture, National Institute of Food and Agriculture specialty crop initiative under USDA NIFA Award No. 2012-51181-20086. The government has certain rights in the invention.

REFERENCE TO ELECTRONIC SEQUENCE LISTING

[0003] The present application is being filed along with an Electronic Sequence Listing. The Electronic Sequence Listing is provided as a file entitled CALTE130ASEQLIST.txt which is 83,106 bytes in size, created on May 3, 2018. The information in the Electronic Sequence Listing is incorporated herein by reference in its entirety.

BACKGROUND

Field

[0004] The disclosure is generally related to DNA sequence modification-based modification of a population.

Description of the Related Art

[0005] Gene drive occurs when genetic elements—including genes, gene complexes, entire chromosomes and endosymbiotic bacteria—are transmitted to viable, fertile progeny at rates greater than those due to Mendelian transmission, resulting in an increase in their frequency in the population over time, even if their presence results in a fitness cost to carriers.

SUMMARY

[0006] In some embodiments, a vector is provided. In some embodiments, the vector comprises a first gene encoding a DNA sequence modifying enzyme, wherein the DNA modifying enzyme modifies an endogenous copy of an essential gene, a first promoter operably linked to the first gene encoding the DNA sequence modifying enzyme, a second gene encoding a rescue transgene, a second promoter operably linked to the rescue transgene, and optionally, one or more cargo sequences, wherein the vector is configured to be positioned in a chromosome or an extra-chromosomal element.

[0007] In some embodiments of the vector, the DNA sequence modifying enzyme is a nuclease or a base editor. In some embodiments of the vector, the nuclease cleaves and generates one or more double strand breaks in the endogenous copy of the essential gene. In some embodiments of the vector, the one or more double strand breaks are repaired to create an altered sequence of the essential gene. In some embodiments of the vector, the base editor creates one or more base changes in the endogenous copy of the essential gene to create an altered sequence of the essential gene. In some embodiments of the vector, the one or more base changes comprise one or more point mutations in the endogenous copy of the essential gene.

[0008] In some embodiments of the vector, the rescue transgene is either a recoded copy of the essential gene or is a gene of unrelated sequence, wherein the rescue transgene encodes a protein that is functionally equivalent to a protein encoded by the essential gene, and wherein the DNA sequence modifying enzyme does not modify the rescue transgene.

[0009] In some embodiments of the vector, the chromosome is an autosome, X chromosome, Y chromosome, or supernumerary chromosome. In some embodiments of the vector, the extra-chromosomal element is a plasmid or a virus.

[0010] In some embodiments of the vector, the one or more cargo sequences comprise one or more foreign gene sequences, or one or more alleles of an endogenous chromosomal or extra-chromosomal gene to which the vector has been linked through nearby insertion on the chromosome or extra-chromosomal element that carries the endogenous allele of interest.

[0011] In some embodiments of the vector, the DNA sequence modifying enzyme is selected from the group consisting of Cas9, Cas-9-related RNA-guided nucleases, ZFN, TALEN, homing endonuclease, restriction enzymes, natural site-specific nucleases, engineered site-specific nucleases, base editing enzymes, cytidine deaminase, and adenine deaminase.

[0012] In some embodiments, the vector further comprises one or more additional sequences, wherein the one or more additional sequences allow the vector to be positioned in the chromosome or the extra-chromosomal element. In some embodiments of the vector, the one or more additional sequences is selected from the group consisting of transposase binding site, LTRs, recombinase binding site, a sequence with homology to a desired location on the chromosome or the extra-chromosomal element.

[0013] In some embodiments of the vector, the first promoter is selected from the group consisting of a germline promoter, a male specific germline promoter, a female specific germline promoter, a cell-type specific promoter, a tissue-specific promoter, a ubiquitous promoter, a promoter activated at a specific stage of mitosis, and a promoter activated at a specific stage of meiosis.

[0014] In some embodiments of the vector, the double strand break is repaired by a mechanism selected from the group consisting of non-homologous end joining, microhomology-mediated end joining, and incomplete homologous recombination.

[0015] In some embodiments of the vector, the size of the one or more cargo sequences ranges from about 0.5 kb to about 500 kb.

[0016] In some embodiments of the vector, the nuclease comprises at least one nuclease domain and one or more DNA binding domains. In some embodiments of the vector, when the nuclease is Cas9 or a Cas9-related enzyme, the vector further comprise one or more genes encoding a guide RNA, wherein the guide RNA enables the nuclease to target specific sequences within the essential gene through Watson-Crick base pairing. In some embodiments of the vector, when the nuclease is Cas9, the nuclease domain of Cas9 is inactivated through one or more mutations, and the vector

2

comprises a different nuclease domain. In some embodiments of the vector, the different nuclease domains is single chain variant of FokI. In some embodiments of the vector, when the DNA binding domain is a TALE, the nuclease domain is provided as a single active nuclease domain, such as single chain variants of FokI.

[0017] In some embodiments, a method of modifying a population by a vector is provided. In some embodiments, the method comprises obtaining an organism of the population, positioning the vector, configured to be positioned in at least one chromosome or extra-chromosomal element in the organism, comprising a first gene encoding a DNA sequence modifying enzyme, wherein the DNA modifying enzyme modifies an endogenous copy of an essential gene, a first promoter operably linked to the first gene encoding the DNA sequence modifying enzyme, a second gene encoding a rescue transgene, a second promoter operably linked to the rescue transgene, and optionally, one or more cargo sequences, expressing the DNA sequence modifying enzyme in the organism, inducing one or more sequence modifications in the essential gene in one or more cells in the organism, such that the one or more sequence modifications result in the essential gene being rendered partially or wholly non-functional and result in a defect in survival, growth control, fertility, or differentiation of the one or more cells if the one or more cells lack the rescue transgene, rescuing the defects in survival, growth control, or differentiation of the one or more cells in which the essential gene has been rendered partially or wholly non-functional, by the rescue transgene, generating an altered organism, wherein the altered organism carries one or more copies of the vector, and wherein the defects in survival, growth control, or differentiation of the one or more cells in which the essential gene has been rendered partially or wholly non-functional have been rescued by the rescue transgene, introducing the altered organism in an environment wherein an increase in a frequency of the altered organism is desired relative to a frequency of a wild type organism in the population; replacing the wild type organism with the altered organism in the population in the environment wherein the altered organism is introduced, thereby modifying the population.

[0018] In some embodiments of the method, an organism with the defect in survival, growth control, fertility, or differentiation of the one or more cells is eliminated if the one or more cells of the organism lack the rescue transgene.

[0019] In some embodiments of the method, the DNA sequence modifying enzyme does not modify the rescue transgene.

[0020] In some embodiments of the method, rescuing the defects in survival, growth control, or differentiation is achieved by restoring normal survival, growth control, fertility, or differentiation of the one or more cells by the rescue transgene.

[0021] In some embodiments of the method, the one or more cells comprise somatic cells, germline cells, gametes, or a combination thereof.

[0022] In some embodiments of the method, the DNA sequence modifying enzyme is a nuclease or a base editor. In some embodiments of the method, the nuclease cleaves and generates one or more double strand breaks in the endogenous copy of the essential gene.

[0023] In some embodiments of the method, the one or more double strand breaks are repaired to create an altered sequence comprising insertions, deletions, base alterations, or a combination thereof.

[0024] In some embodiments of the method, the base editor creates one or more base changes or small insertions/deletions in the endogenous copy of the essential gene. In some embodiments of the method, the one or more base changes comprise one or more point mutations, or deamidated bases that are replaced with nucleotides of a different sequence.

[0025] In some embodiments of the method, the altered organism is heterozygous or homozygous for the vector. In some embodiments of the method, the organism is haploid, diploid, or polyploid. In some embodiments of the method, the organism is selected from the group consisting of prokaryotes, fungi, plants, and animals.

[0026] In some embodiments of the method, the environment comprises an open environment, a bioreactor, a multicellular body, or a colony of individual cells.

[0027] In some embodiments of the method, the wild type organism is replaced at a high frequency with the altered organism in the environment wherein the wild type organism is present. In some embodiments of the method, the high frequency is defined as replacement of at least 90% of the wild type organism with the altered organism after 100 generations in the population. In some embodiments of the method, the wild type organism is replaced at a rapid rate with the altered organism in the environment wherein the wild type organism is present. In some embodiments of the method, the rapid rate is defined as replacement of at least 90% of the wild type organisms by organisms carrying the vector in the population after at most 100 generations.

[0028] In some embodiments of the method, the one or more sequence modifications in the one or more cells is a result of the one or more cells carrying the first gene encoding the DNA sequence modifying enzyme or is a result of the DNA sequence modifying enzyme being transmitted to the one or more cells from one or more cells expressing the DNA sequence modifying enzyme through diffusion, active transport, or movement of the DNA sequence modifying enzyme from a cell that expresses the DNA sequence modifying enzyme to a cell that does not express the DNA sequence modifying enzyme.

[0029] In some embodiments of the method, the vector is positioned on the chromosome or the extra-chromosomal element by a homologous recombination-dependent integration. In some embodiments of the method, the vector is positioned on the chromosome or extra-chromosomal element by random integration, integration using transposition, integration using a recombinase, or a combination thereof.

[0030] In some embodiments of the method, the one or more cargo sequences comprise one or more foreign gene sequences, or one or more alleles of an endogenous chromosomal or extra-chromosomal gene to which the vector has been linked through nearby insertion on the chromosome or extra-chromosomal element that carries the endogenous allele of interest.

[0031] In some embodiments of the method, the vector is positioned on the chromosome or the extra-chromosomal element, the first gene operably linked to the first promoter, the second gene operably linked to the second promoter, and the one or more cargo transgenes are genetically linked.

[0032] In some embodiments of the method, the nuclease cleaves and generates one or more double strand breaks in the endogenous copy of the essential gene with a high cleavage efficiency. In some embodiments of the method, the high cleavage frequency is defined as the nuclease cleaving the endogenous copy of the essential gene in at least 30% of the organisms carrying the vector and the endogenous copy of the essential gene in each generation. In some embodiments of the method, the base editor creates one or more base changes in the endogenous copy of the essential gene with a high base editing frequency. In some embodiments of the method, the high base editing frequency is defined as the base editor modifying the endogenous copy of the essential gene in at least 20% of the organisms carrying the vector and the endogenous copy of the essential gene in each generation.

BRIEF DESCRIPTION OF THE DRAWINGS

[0033] FIG. 1A-FIG. 1C show an embodiment of an X cleavage mediated Y drive. The vector is present on the Y chromosome. Cleavage of an essential gene located on the X chromosome is brought about by Cas9 and associated gRNAs. The Y chromosome also carries a recoded version of the essential gene that is resistant to cleavage by Cas9.

[0034] FIG. 1A shows a schematic of the mechanism of action an embodiment of a vector (transgenic construct) that brings about drive of a vector-bearing Y chromosome through cleavage of an essential gene on the X chromosome.

[0035] FIG. 1B shows a schematic of an embodiment of inheritance, and viable or non-viable progeny, of an X chromosome cleavage mediated Y chromosome drive process. X (linear) and Y (kinked) chromosomes are indicated.

[0036] FIG. 1C shows a graph of an embodiment of a population frequency modeling of X cleavage mediated Y drive for different fitness costs and introduction frequencies. The heat map to the right indicates the number of generations required for the vector to reach a population frequency of >99%.

[0037] FIG. 2A-FIG. 2C show an embodiment of a cleavage mediated X drive with the vector also located on the X.

[0038] FIG. 2A shows a schematic of the mechanism of action an embodiment of a vector (transgenic construct) for cleavage mediated X drive with the vector located on the X.

[0039] FIG. 2B shows a schematic of inheritance, and viable or non-viable progeny, of a cleavage mediated X drive process with the vector located on the X.

[0040] FIG. 2C shows a graph of an embodiment of a population frequency modeling of cleavage mediated X drive with the vector located on the X.

[0041] FIG. 3A-FIG. 3C show an embodiment of a cleavage mediated autosomal drive.

[0042] FIG. 3A shows a schematic of the mechanism of action an embodiment of a vector (transgenic construct) for cleavage mediated autosomal drive.

[0043] FIG. 3B shows a schematic of inheritance and viable or non-viable progeny of a cleavage mediated autosomal drive process.

[0044] FIG. 3C shows a graph of an embodiment of a population frequency modeling of cleavage mediated autosomal drive.

[0045] FIG. 4A-FIG. 4C show an embodiment of a cleavage mediated 2-locus autosomal drive.

[0046] FIG. 4A shows a schematic of the mechanism of action an embodiment of a vector (transgenic construct) for cleavage mediated 2-locus autosomal drive.

[0047] FIG. 4B shows a schematic of inheritance, and viable or non-viable progeny, of a cleavage mediated 2-locus autosomal drive process.

[0048] FIG. 4C shows a graph of an embodiment of a population frequency modeling of cleavage mediated 2-locus autosomal drive.

[0049] FIG. 5A-FIG. 5C show an embodiment of a cleavage mediated haplolethal drive.

[0050] FIG. 5A shows a schematic of the mechanism of action an embodiment of a vector (transgenic construct) for cleavage mediated haplolethal drive.

[0051] FIG. 5B shows a schematic of inheritance and viable or non-viable progeny of a cleavage mediated haplolethal drive process.

[0052] FIG. 5C shows a graph of an embodiment of a population frequency modeling of cleavage mediated haplolethal drive.

[0053] FIG. 6A-FIG. 6F show a schematic of an embodiment of maintenance of extrachromosomal element.

[0054] FIG. 7 shows a schematic of an embodiment the results of a cross between a female insect heterozygous for the vector with germline expression of the DNA sequence modifying enzyme and a wild type male when there is no carryover of DNA cleavage/alteration activity from germline into embryo.

[0055] FIG. 8A shows a schematic of an embodiment the results of a cross between an insect heterozygous for the vector with germline expression of the DNA sequence modifying enzyme an a second insect heterozygous for the vector when there is no maternal transfer of DNA cleavage/alteration activity from germline into embryo. Individuals that inherit no functional copies of the essential gene die, while those that inherit at least one copy of the vector and its associated rescue transgene survive.

[0056] FIG. 8B shows a graph of an embodiment of vector-mediated gene drive/population replacement for an autosomal two locus scenario, with different fitness costs and introduction frequencies, and without maternal transfer of DNA cleavage/alteration activity.

[0057] FIG. 9A shows a schematic of an embodiment the results of a cross when there is maternal transfer of DNA cleavage/alteration activity from germline into embryo.

[0058] FIG. 9B shows a graph of an embodiment of vector-mediated gene drive/population replacement for different fitness costs and introduction frequencies with maternal transfer of DNA cleavage/alteration activity.

[0059] FIG. 10 shows a schematic of an embodiment of a meiotic gene drive. Spores that fail to inherit a functional copy of the essential gene die.

[0060] FIG. 11 shows a schematic of an embodiment of vector-mediated sex ratio distortion.

[0061] FIG. 12 shows a schematic of an embodiment of homing endonuclease gene (HEG)-based population replacement in which the cargo gene is included as a component of the HEG.

[0062] FIG. 13 shows a schematic of an embodiment of HEG-based population replacement in which the cargo is located at a different site in the genome.

[0063] FIG. 14 shows a schematic of an embodiment of a Medea-based gene drive.

[0064] FIG. **15**A-FIG. **15**C show an embodiment of DNA sequence modification based gene drive (herein referred to as ClvR when a nuclease is used for DNA sequence modification) construct design and principle according to the present disclosure.

[0065] FIG. **16**A-FIG. **16**B show an embodiment of a determination of the effects of a CleaveR drive when transmitted through the female (FIG. **16**A) or male (FIG. **16**B) germline.

[0066] FIG. **17** shows an embodiment of an alignment of the target gene (*Drosophila melanogaster* tko—Examples 15 and 16) with the recoded rescue based on *Drosophila virilis* tko. Target exon in red, recoded rescue exon in green, gRNA target sites in pink, PAM in bold letters, additional silent point mutations introduced into the rescue copy to reduce homology in blue.

[0067] FIG. **18**A show a schematic of an embodiment of the components of the DNA sequence modification-based gene drive (Example 17).

[0068] FIG. **18**B shows a schematic of an embodiment of the components of the step 1 of FIG. **18**A (Example 17).

[0069] FIG. **18**C shows a schematic of an embodiment of the components of the step 2 of FIG. **18**A (Example 17).

[0070] FIG. **19** shows an embodiment of the results of Sanger sequencing from Example 17.

DETAILED DESCRIPTION

[0071] In nature gene drive is brought about by a number of mechanisms, in a number of contexts (Ben-David et al. 2017; Burt and Trivers 1998; Seidel et al. 2011; Nuckolls et al. 2017; Hu et al. 2017). A number of novel methods of engineering gene drive have also been proposed, and in several cases implemented.

[0072] There are two general contexts in which gene drive is considered as a technological tool. In one, the goal is population replacement: to spread a trait throughout an extant population. For organisms such as beneficial insects such traits include insecticide, natural pathogen resistance or resistance to other stresses. For a pest/disease vector traits of interest include insecticide sensitivity, the inability to carry or transmit specific pathogens, or a change in life history that preclude pathogen transmission. Genes that confer conditional lethality in response to an environmental cue, so as to ultimately bring about population suppression, are also of interest. A second goal is population suppression or elimination. Targets of interest include invasive species of plants and animals, pests that cause damage directly to plants or animals, and vectors of plant or animal disease. Finally, gene drive is also of interest as a tool for maintaining the presence of a trait in a population in which the genetic element (plasmid, chromosome, virus) in which the gene drive element and any associated cargo genes are sometimes lost, for example during cell division. This is related to population replacement.

[0073] A number of methods have been considered for bringing about self-sustaining population replacement. Many of these take as their starting point naturally occurring selfish genetic elements to which cargo genes could be linked (Braig and Yan 2001; Burt and Trivers 1998; Chen et al. 2007). Others involve the use of novel, engineered systems, many of which utilize, in one way or another, the phenomenon of underdominance (heterozygote disadvantage) (Gould and Schliekelman 2004; Marshall and Hay 2011; Marshall and Hay 2012; Marshall et al. 2011; Akbari

et al. 2013; Altrock et al. 2010; Altrock et al. 2011; Davis et al. 2001; Gokhale et al. 2014; Reeves et al. 2014). An important characteristic of any gene drive mechanism is its level of invasiveness: its ability to increase in frequency both at the point of release and in surrounding areas linked to the release site by various levels of migration, when introduced at various population frequencies. Here we divide gene drive mechanisms somewhat arbitrarily into low and high threshold variants, with the understanding that these distinctions lie along a continuum. Low threshold gene drive mechanisms require that only a small fraction of individuals in the population carry the drive element in order for spread to occur locally (Marshall 2009; Marshall and Hay 2012). Examples include transposons, engineered Medea chromosomal elements (Chen et al. 2007; Wade and Beeman 1994; Ward et al. 2011), several other possible single locus chromosomal elements (Marshall and Hay 2012), site-specific nucleases that home into their target site (Burt 2003; Gantz and Bier 2015; Gantz et al. 2015; Hammond et al. 2016; Simoni et al. 2014; Windbichler et al. 2011), and site-specific nucleases located on the Y chromosome that cleave and thereby (somehow) block development of X-bearing sperm, resulting in sex ratio distortion (Galizi et al. 2014). These mechanisms are predicted to be invasive because low levels of migration of drive element-bearing individuals into areas outside the release area may, depending on the threshold and the migration rate (Beaghton et al. 2016; Beaghton et al. 2017; Godfray et al. 2017; Marshall 2009; Marshall and Hay 2012), result in these areas being seeded with enough transgene-bearing individuals that drive is likely to occur. Low threshold, invasive gene drive mechanisms are attractive when the goal is to spread transgenes over a large area, and migration rates between the release site and surrounding areas of interest are low. However, for these same reasons, it is likely to be challenging to restore the population to the pre-transgenic state if desired. High (or higher) threshold gene drive mechanisms require, as their name implies, that transgenes make up a much larger fraction of the total insect population (important examples range from ~15-70%) before gene drive occurs. Below this frequency transgenes are instead actively eliminated from the population. These drive mechanisms thus behave as frequency-dependent bistable switches. High transgene frequencies are needed to initiate drive at the release site, limiting the possibility that unintended release of a few individuals could initiate replacement (Marshall 2009). Furthermore, once replacement has occurred at the release site, spread to high frequency in areas connected to the release site by low levels of migration is prevented because the transgene never reaches the threshold frequency needed for drive (Altrock et al. 2010; Altrock et al. 2011; Marshall and Hay 2012). Finally, transgenes can be eliminated from the population if the release of wildtypes results in the frequency of transgenics being driven below the threshold required for drive. A number of gene drive mechanisms that could in principal bring about high threshold gene drive have been proposed. Examples include a number of single locus toxin-antidote gene drive mechanisms (Marshall and Hay 2011; Marshall and Hay 2012; Marshall et al. 2011), reciprocal chromosome translocations, inversions and compound chromosomes (Gould and Schliekelman 2004), and several forms of engineered underdominance (Akbari et al. 2013; Altrock et al. 2010; Altrock et al. 2011; Davis et al. 2001; Gokhale et al. 2014; Marshall and Hay 2012; Reeves

5

et al. 2014). Two of these, UD^{MEL}(double Medea), and engineered reciprocal translocations, have recently been shown to drive reversible population replacement into populations of wildtype *Drosophila* (Akbari et al. 2013; Buchman et al. 2018). A third system has been shown to drive high threshold population replacement in *Drosophila* in a split configuration (Reeves et al. 2014). In each of these systems gene drive occurs when transgene-bearing chromosomes experience frequency-dependent changes in fitness with respect to non-transgene-bearing counterparts, with the former having high fitness at high frequency and lower fitness at low frequency. These systems all rely, in one way or another, on the phenomena of underdominance, in which transgene-bearing heterozygotes (or some fraction of them or their progeny) have a lower fitness than either homozygous wildtypes or homozygous transgenics (or transgene-bearing trans-heterozygote in some three allele cases). If the frequency of one allele or pair of alleles or chromosome type is above a critical threshold it spreads to genotype, and in some cases allele fixation. Conversely, if it falls below the critical threshold it is lost in favor of the other allele or chromosome type, usually wildtype. In broad outline, this behavior occurs because when transgene-bearing individuals are common they mate mostly with each other, producing transgene-bearing offspring of high fitness (high survival and/or fecundity), while wildtypes mate mostly with transgene-bearing individuals, producing a preponderance of heterozygous offspring of low fitness (inviable and/or with reduced fecundity). However, when the frequency of wildtypes is high the tables are turned, with transgene-bearing individuals producing high frequencies of unfit heterozygous progeny, and wildtypes producing a high frequency of fit homozygous progeny.

[0074] The only gene drive mechanisms shown to drive population replacement in otherwise wildtype organisms are Medea (Akbari et al. 2012; Buchman et al. 2018; Chen et al. 2007), UDMEL (Akbari et al. 2013), and reciprocal chromosome translocations (Buchman et al. 2018), all in *Drosophila melanogaster* or *Drosophila suzukii*. Several other methods, including engineered underdominance (Reeves et al. 2014) and homing endonucleases (Windbichler et al. 2011; Windbichler et al. 2007; Simoni et al. 2014; Gantz and Bier 2015; Gantz et al. 2015; Hammond et al. 2016; Champer et al. 2017; Chan et al. 2011; Chan et al. 2013), have seen important progress, though population replacement has not been demonstrated.

[0075] There is a need for robust mechanisms of gene drive that can easily be developed for diverse species, and that are robust to mechanisms that can cause failure of gene drive to occur. Thus, while Medea elements have been generated in Drosophila, it has not yet been possible to develop them in other insects. In addition, Medea is inherently challenging because it requires that early zygotic promoters be available, along with antidotes, which together are capable of rescuing maternal lethality. These reagents, as well as specific mechanisms for bringing about toxicity in embryos but not oocytes, are challenging to identify and create, and their implementation requires that one have detailed biological knowledge of the species under consideration (Hay et al. 2010). UDMEL represents a more complicated version of Medea, and therefore suffers from the same problems (Akbari et al. 2013). Homing-based population replacement is challenging for several reasons. First, it requires that DNA cleavage be followed by DNA repair

using homologous recombination, and that homologous recombination proceed through the entire gene drive element that must be copied. Since the cell utilizes multiple repair pathways, and HR is inefficient, complete copying through HR often does not happen. Second, because homing requires the targeting and cleavage of a specific sequence, its efficacy is sensitive to genomic sequence variation. Variation can occur as pre existing sequence polymorphisms in a population. It can also arise from mutation, and as a result of break repair through non-homologous end joining, which is error prone (Preston et al. 2006; Windbichler et al. 2011). Regardless of the mechanism, sequence variants that are not cleaved are resistant to homing, and may retain some or complete wildtype gene function. The presence of such resistant alleles can block HEG spread and thereby prevent population replacement. Thus, the question of how to bring about high frequency homing that is gene specific, but insensitive to some level of sequence variation within the gene, is central to the development of HEG-based population replacement technologies, and remains to be solved. Translocations can only provide high threshold population replacement. They also require a significant amount of chromosomal engineering, in that two large chromosome fragments must become linked to each other, while maintaining high levels of organism fitness (Buchman et al. 2018; Marshall and Hay 2012). Finally, shredding of the X chromosome through the use of a P-linked transgene that thereby causes the loss of X-bearing sperm has also been proposed (Burt 2003), and significant progress has been made (Galizi et al. 2014; Galizi et al. 2016; Windbichler et al. 2008). However, this approach is limited to population suppression and species that have clear X and Y chromosomes in which males are Y. Many species of interest lack this configuration. In summary, gene drive for population replacement is an important technological goal, but methods for easily engineering it in diverse species are lacking.

[0076] As a specific example of the need for population replacement gene drive, despite a myriad of approaches to controlling mosquito-borne infections, ranging from insecticide treated bed nets, new anti-malarial drugs such as artemisinin, and suppression attempts using sterile males, there are still over 600,000 deaths from malaria each year [WHO World Malaria Report 2014]. This stems from a combination of lack of human compliance, emerging drug resistance, and selection for mosquitoes preferring to bite outdoors. These failures show the need for novel molecular approaches to combating insect-borne disease [Alphey, 2014].

[0077] However, the approaches proposed face substantial barriers to their development. In toxin-antidote systems, the toxin has to be strong enough to suppress one or both copies of the target gene and the recoded 'antidote' version of this gene has to have strong enough and timely zygotic expression to compensate for the loss of the maternal product Chen et al 2007, [Akbari, 2013; Akbari, 2014]. These are already difficult requirements for the development of the original gene drive, let alone successive drives in case the original mutates to inactivity. Additionally, what works in one species, such as the Medea^{myd88} in *drosophila melanogaster*, does not necessarily work in other species, such as *Aedes aegypti*, despite sharing the molecular components involved in the drive.

[0078] HEG approaches are elegant in that they increase their frequency not through the destruction of competing

alleles as in toxin-antidote drives but by copying themselves onto non HEG containing homologs, thus forcing heterozygotes for the HEG to become homozygous. However, they suffer from the being limited in what they can target due to their inherent base specificity and from potential replication errors every time they are copied.

[0079] HEG based approaches to gene drive are predicted to be very powerful, driving from low frequency and in relatively few generations. The emergence of TALENs and ZFNs have vastly expanded the number of possible target sites while maintaining specificity, but their multiple repeats make them prone to mutation due to recombination [Simoni, 2014; Esvelt, 2014]. An alternative now being very actively explored utilizes the CRISPR nuclease Cas9 and gRNAs that target Cas9 to specific sequences for cleavage based on Watson-Crick base pairing interactions. While HEGs based on Cas9 can target virtually any sequence, a Cas9 drive construct is likely to be quite large, making homing more difficult and the construct much more prone to copying errors.

[0080] While drives like Medea can incorporate new toxins in addition to old ones to perform additional stages of replacement, adding additional gRNAs will buffer a Cas9 HEG against NHEJ resistant alleles but will only make the construct even larger and thus more prone to other problems, such as abortive gap repair.

[0081] Cas9 can be used at the heart of any of the gene drives previously proposed for use as HEGs, with a substantially larger pool of potential targets while maintaining specificity. However, these strategies have the major drawback of susceptibility to DNA loss or drive dysfunction due to the imperfect copying of Cas9 during homology directed repair.

[0082] Described herein is a novel mechanism for gene drive that is very simple, yet powerful, and utilizes only two simple components that can be readily engineered in any organism for which genetic engineering can be achieved. The first component is a gene expressing an enzyme that bring about DNA sequence modification, and thus inactivation, of an essential gene. The second component is a transgene (the rescue transgene) that is able to rescue the loss of function phenotype due to inactivation of the endogenous copies of the essential gene, and is insensitive to enzyme-mediated DNA sequence modification. This method requires only two components: a site-specific DNA modifying enzyme that targets a gene required for viability or fertility in any way (an essential gene), and a second, functional version of the essential gene that includes sequences that are resistant to modification by the site-specific DNA modifying enzyme (the rescue transgene). When these two elements are linked together, for example, in a vector (e.g., plasmid), organisms that carry the vector always survive because they always carry the rescue transgene. In contrast, organisms that do not carry the rescue transgene will die or be sterile if they only carry inactive copies of the essential gene that are inherited from vector-bearing parents or created de novo through site-specific DNA modifying enzyme activity that is brought into these cells through diffusion, transport, or cell-cell movement.

[0083] In some embodiments, the gene drive disclosed herein is an alternative form of gene drive that utilizes Cas9 or other nucleases to bring about cleavage and repair of an essential gene that does not involve or require homing. This form of gene drive can also make use of base editing enzymes such as adenosine or cytosine deaminase to modify specific bases to create non-functional versions of an essential gene. Without being limited by any particular theory, the mechanism simply involves DNA sequence modifying enzyme such as Cas9, a set of gRNAs targeting an essential gene for cleavage, (or a sequence targeted base editor) and a recoded version of the target that is immune to modification linked as a single construct. In some embodiments, individuals carrying one or more copies of this construct bring about modification of the sequence of one or more copies of the endogenous version of the essential gene such that it is no longer functional. Individuals who end up inheriting only non-functional versions of the essential gene die or are sterile, while those that carry one or more copies of the construct, which includes a rescue transgene, will survive and/or be fertile. Over multiple generations this behavior is predicted to result in the spread of the construct/vector into the population at the expense of the wildtype version of the same chromosome.

[0084] In some embodiments, characterized and disclosed herein are multiple forms of this DNA sequence modification mediated drive. A discrete generation, deterministic population frequency model is used to demonstrate that there are a variety of conditions, that include various fitness costs, DNA sequence modification frequencies, and introduction frequencies, under which population replacement is predicted to occur.

Definitions

[0085] As used herein, the section headings are for organizational purposes only and are not to be construed as limiting the described subject matter in any way. All literature and similar materials cited in this application, including but not limited to, patents, patent applications, articles, books, treatises, and internet web pages are expressly incorporated by reference in their entirety for any purpose. When definitions of terms in incorporated references appear to differ from the definitions provided in the present teachings, the definition provided in the present teachings shall control. It will be appreciated that there is an implied "about" prior to the temperatures, concentrations, times, etc discussed in the present teachings, such that slight and insubstantial deviations are within the scope of the present teachings herein.

[0086] In this application, the use of the singular includes the plural unless specifically stated otherwise. Also, the use of "comprise", "comprises", "comprising", "contain", "contains", "containing", "include", "includes", and "including" are not intended to be limiting. It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive. Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. See, for example Singleton et al., Dictionary of Microbiology and Molecular Biology 2nd ed., J. Wiley & Sons (New York, N.Y. 1994); Sambrook et al., Molecular Cloning, A Laboratory Manual, Cold Springs Harbor Press (Cold Springs Harbor, N.Y. 1989). For purposes of the present invention, the following terms are defined below. It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the invention as claimed. In this application, the use of the

singular includes the plural unless specifically stated otherwise. In this application, the use of "or" means "and/or" unless stated otherwise. Furthermore, the use of the term "including", as well as other forms, such as "includes" and "included", is not limiting.

[0087] As used in this specification and claims, the singular forms "a," "an" and "the" include plural references unless the content clearly dictates otherwise.

[0088] As used herein, "about" means a quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length that varies by as much as 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1% to a reference quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length.

[0089] As used herein, "regulatory element" refers to nucleic acid elements that can influence the expression of a coding sequence (for example, a gene) in a particular host organism. These terms are used broadly and encompass all elements that promote or regulate transcription, including promoters, core elements required for basic interaction of RNA polymerase and transcription factors, upstream elements, enhancers, and response elements (see, for example, Lewin, "Genes V" (Oxford University Press, Oxford) pages 847-873).

[0090] As used herein, the term "insertion site" refers a nucleic acid sequence that allows for insertion of the constructs as provided herein into a genome of a multicellular organism (for example, an insect genome). In some embodiments, a construct as provided herein can comprise a "insertion sequence" that allows for insertion of the construct into a genome of the host organism. Some embodiments that can be employed include the piggybac transposable element, mariner type transposable elements, and the P-element. Also, plasmids can be site specifically integrated into the genome using attb/attp or even by using CRISPR/Cas9, TALEN, MegaTAL and homologous recombination.

[0091] As used herein, a "vector," interchangeably referred to as a transgenic construct, a targeting construct, or simply a construct, is a nucleic acid. As used herein, "nucleic acid" refers to deoxyribonucleic acid (DNA). In some embodiments, nucleic acid may refer to ribonucleic acid (RNA). In some embodiments, the construct as provided herein comprise one or more regulatory elements. Exemplary regulatory elements in prokaryotes include promoters, operators and ribosome binding sites. Regulatory elements that are used in eukaryotic cells can include, without limitation, transcriptional and translational control sequences, such as promoters, terminators, enhancers, insulators, splicing signals, polyadenylation signals, terminators, protein degradation signals, internal ribosome-entry element (IRES), 2A sequences, and the like, that provide for and/or regulate expression of a coding sequence and/or production of an encoded polypeptide in a host cell. For example, a promoter is a nucleotide sequence that permits binding of RNA polymerase and directs the transcription of a gene. Typically, a promoter is located in the 5' non-coding region of a gene, proximal to the transcriptional start site of the gene. Sequence elements within promoters that function in the initiation of transcription are often characterized by consensus nucleotide sequences. Examples of promoters include, but are not limited to, promoters from bacteria, yeast, plants, viruses, and mammals (including humans). A promoter can be inducible, repressible, and/or constitutive. Inducible promoters initiate increased levels of transcription

from DNA under their control in response to some change in culture conditions (for example, a change in temperature).

[0092] As used herein, "homologous recombination" refers to exchange of nucleotide sequences between two identical nucleic acid sequences. Homologous recombination also refers to exchange of nucleotide sequences between two similar nucleic acid sequences. In some embodiments, when the two nucleic acid sequences are similar, a similarity between the two nucleic acid sequences can be about 90% to about 99.9%. In some embodiments, the similarity between the two nucleic acid sequences can be about 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8 or 99.9%.

[0093] As used herein, "gene drive" refers to a situation in which genetic elements—including alleles of specific genes, gene complexes, entire chromosomes or endosymbiotic bacteria—are transmitted to viable, fertile progeny at rates greater than those due to Mendelian transmission, resulting in an increase in their frequency in the population over time, even if their presence results in a fitness cost to carriers. Without being limited by any particular theory, gene drive can occur by a number of mechanisms. In some embodiments gene drive has evolved in wild populations of various organisms, through a variety of mechanisms that are still under study (Burt and Trivers, 2006). In some embodiments, the gene drive is engineered. In some embodiments, the gene drive represents a naturally occurring mechanism or is engineered depending on the context and environment in which it occurs. A number of novel methods of engineering gene drive have also been proposed, and in several cases implemented.

[0094] In some embodiments, the present disclosure is related to vectors and methods for DNA sequence modification-based modification of populations, and beneficial and commercial applications of the vectors and methods.

[0095] In one implementation of this system, detailed below in the examples and sometimes referred to as CleaveR (also referred to herein as ClvR), the nuclease includes a member of the RNA-guided nucleases, such as Cas9. In this implementation Cas9 is expressed in the germline of male, females, or both sexes. Multiple gRNAs are also expressed, preferably three or four of them. They are designed to engage in Watson-Crick base pairing with, and therefore target for cleavage, distinct sequences within a target gene, so as to bring about its cleavage at multiple sites. These multiple breakages are expected to result in the creation of repair products—deletions, base changes, small additions—that create a non-functional version of the targeted gene. In summary, the purpose of the nuclease is to bring about loss-of-function mutants of the targeted gene. There are two important characteristics of the system described thus far. First, the cassette encoding the nuclease can sit at any position in the genome. Second, the gene being targeted for inactivation is in some sense an essential gene: required for organism survival or fertility, broadly defined as fitness.

[0096] The second component of the CleaveR gene drive system is the existence of a version of the targeted essential gene that can rescue the lethality or infertility of those individuals in which both copies (for a diploid) of the essential gene have been inactivated, but that is itself resistant to cleavage by the RNA-guided Cas9 component of the construct. Resistance to cleavage is brought about by recoding the transgene so that it no longer productively interacts with the guide RNA Cas9 complex, according to rules that

8

are well known in the field. Further recoding of the rescue transgene, in both the coding region and non-coding and regulatory regions, is also carried out. This recoding is done so as to minimize homology between the wildtype, endogenous version of the gene and the rescue version of the gene. This recoding is also done so as to minimize/eliminate the possibility that the cleaved version of the wildtype endogenous essential gene can be repaired and restored to functionality through ectopic homologous recombination, using the rescue transgene as a template for repair based on existing homology at the broken ends of the former. The literature provides guidance on the level of homology needed to prevent or promote homologous recombination. Without being limited by any particular theory, recoding can successfully achieved even when the rescue transgene has essentially no nucleotide homology to the endogenous copy of the gene. Demonstration that this can be achieved comes from multiple reports showing that bacterial and/or human versions of a large number of essential genes can successfully replace their yeast counterparts, resulting in yeast with high fitness.

[0097] In the CleaveR construct, also often referred to as the vector or the construct, when these two genes are located near each other (tightly linked), they behave, as illustrated below, as a novel selfish genetic element, able to spread itself into a population and/or maintain itself in a population (bring about population replacement) under a variety of conditions that include varying levels of fitness cost associated with carrying the vector and any associated cargo genes, and introduction frequencies. The details of these characters are described in more below.

Overview of CleaveR-Based Gene Drive

[0098] Without being limited by any particular theory, when the CleaveR construct is present in an organism, wildtype copies of the essential are a risk for cleavage and inactivation. The individuals carrying CleaveR themselves do not experience any cost from this cleavage, which happens in the germline and also in some cases in somatic cells, because they also carry a tightly linked copy of the rescue transgene. However, the gametes they pass on will in many cases not carry a functional copy of the endogenous essential gene, and they may also lack the CleaveR construct. In some cases the Cas9/gRNA complexes will also be deposited into oocytes/eggs, resulting in cleavage of the endogenous copy of the essential gene in early embryos that do not carry the CleaveR construct. In all of these cases, which arise through normal Mendelian segregation of chromosomes during meiosis in males and females, and in some cases diffusion or transport of Cas9/gRNA into daughter cells or products of cell-cell fusion (fertilization), progeny are often created that carry no functional copies of the essential gene. These individuals are of low fitness (dead, sterile or otherwise dysfunctional [flightless]) and do not contribute further to the population.

[0099] The above behavior results in some loss in each generation of chromosomes and individuals that do not carry the CleaveR. This results, over multiple generations, in a progressive increase in the frequency of CleaveR-bearing individuals. Modeling, discussed further below, shows that under a variety of conditions CleaveR is predicted to spread to high frequency such that most or all individuals in the population bear at least one copy of the CleaveR chromosome. The CleaveR chromosome is in some sense "held" in

the population because as it has been spread (and the mechanism by which it has been spreading) it has necessarily caused inactivation of most or all of the wildtype copies of the essential gene. Thus the population has become "locked" into a configuration in which it now depends on the presence of CleaveR in order to maintain viability or fertility.

[0100] A similar principle, cleavage associated with rescue of those who carry the CleaveR vector, allows CleaveR to act as a gamete killer (known as spore killers in yeast), and to force its inheritance in conditions in which it is episomal (as in a plasmid). In both cases the presence of the CleaveR element selects for those who carry it, and against those who fail to inherit it.

[0101] In some embodiments, the method of gene drive described herein is agnostic as to the mechanism by which sequence modification-dependent inactivation of the essential gene is brought about. It can involve cleavage and error-prone repair, as discussed above. It can also involve the use of base editing enzymes known from the literature. It can also utilize other DNA modifying enzymes such as sequence targeted transposases, recombinases, integrases, topoisomerases, or other enzymes that can be targeted to specific sequences in DNA to bring about sequence changes. Importantly, the exact nature of the sequence changes brought about is not critical since there are many ways of rendering nonfunctional any particular gene through sequence modification.

Vectors

[0102] FIG. 15A-FIG. 15C show an embodiment of ClvR construct design and principle according to the present disclosure (Example 15). In some embodiments, the disclosure is related to a vector. In some embodiments, the vector comprises a first gene encoding a DNA sequence modifying enzyme. In some embodiments, the DNA modifying enzyme modifies the sequence of an endogenous copy of an essential gene. As used herein, an "essential gene" is defined as a gene that is critical for survival, growth or fertility, and whose loss of function is either lethal, prevents growth or is sterilizing. Some essential genes are critical for survival under all circumstances. Some essential genes are critical for survival only under particular circumstances and/or particular environmental conditions (e.g., in the presence of toxic drugs, toxins, etc., or in the absence of nutrients, vitamins, etc.). In some embodiments, more than one or more endogenous copies of the essential gene are present. In some embodiments, when one or more endogenous copies of the essential gene are present they are alleles or allelic variants of the essential gene. As used herein, the "endogenous copy" refers to the wild type version of the essential gene.

[0103] In some embodiments, vector comprises first promoter operably linked to the first gene encoding the DNA sequence modifying enzyme. In some embodiments, the first gene is operably linked to one or more additional regulatory elements. In some embodiments, the vector further comprises a second gene encoding a rescue transgene. In some embodiments of the vector, a second promoter is operably linked to the rescue transgene. In some embodiments, the second gene is operably linked to one or more additional regulatory elements.

[0104] In some embodiments, the vector optionally comprises one or more cargo sequences. In some embodiments, a cargo sequence is a nucleic acid. In some embodiments,

the vector is configured to be positioned in a chromosome. In some embodiments, the vector is configured to be positioned in an extra-chromosomal element. Non-limiting examples of cargo genes include are sequences encoding antibodies against Plasmodium, the causal agent of malaria (Isaacs et. al. 2011, Hollingdale et. al. 1984, and Li et. al. 2005), or non-coding RNAs to bring about cleavage of the dengue virus RNA genome (Yen et. al. 2018, Franz et. al. 2006, Mathur et. al. 2010, Travanty et. al. 2004, and Castillo et. al. 2016). In some embodiments, the vector is configured to be positioned in a chromosome and an extra chromosomal element. In some embodiments, the vector is configured to be positioned in a chromosome but not in an extra chromosomal element. In some embodiments, the vector is configured to be positioned in an extra chromosomal element but not in a chromosome.

[0105] In some embodiments, the DNA sequence modifying enzyme is a nuclease. Non-limiting examples of nucleases include Flap endonucleases, restriction endonucleases (e.g., F-EcoT5I, F-EcoT5II, F-EcoT5IV, F-SceI, F-TevI, F-TevII, I-AchMI, I-AniI, I-BasI, I-BmoI, I-Bth0305I, I-BthII, I-BthORFAP, I-CeuI, I-ChuI, I-CpaI, I-CpaII, I-CreI, I-CsmI, I-CvuI, I-DdiI, I-DmoI, I-GpiI, I-GzeI, I-HjeMI, I-HmuI, I-HmuII, I-LlaI, I-LtrI, I-LtrWI, I-MpeMI, I-MsoI, I-NanI, I-NitI, I-NjaI, I-OnuI, I-PakI, I-PanMI, I-PnoMI, I-PogTE7I, I-PorI, I-PpoI, I-ScaI, I-SceI, I-SceII, I-SceIII, I-SceVI, I-SpomI, I-SscMI, I-Ssp6803I, I-TevI, I-TevII, I-TevIII, I-TslI, I-Ts1WI.AY76, I-Vdi141I, PI-AvaI, PI-BciPI, PI-HvoWI, PI-MleSI, PI-MtuI, PI-PkoI, PI-PkoII, PI-PspI, PI-SceI, PI-TfuI, PI-TfuII, PI-TliI, PI-TliII, PI-TmaI, PI-TmaKI), Cas9, and Cas9-like enzymes (including but not limited to CPf1, C2c1, C2c2, and C2c3 (Shmakov et. al. 2015, Shmakov et. al. 2017, Koonin et. al. 2017-1, Koonin et. al. 2017-2), ZFNs, MegaTALs, TALENS, HEGs, meganucleases, etc.

[0106] In some embodiments, DNA modifications are achieved through cleavage by site-specific nucleases. Without being limited by any particular theory, it should be understood that equivalent effects can be obtained through the use of any enzyme that brings about modification of a target DNA sequence. Non-limiting examples include cytosine and adenine base changes brought about through the targeted use of deaminases and site-specific integrases.

[0107] In some embodiments, the nuclease cleaves the endogenous copy of the essential gene. In some embodiments, the nuclease generates one or more double strand breaks in the endogenous copy of the essential gene. In some embodiments, the nuclease cleaves and generates one or more double strand breaks in the endogenous copy of the essential gene. In some embodiments, the one or more double strand breaks in the endogenous copy of the essential gene are staggered. In some embodiments, the one or more double strand breaks in the endogenous copy of the essential gene are not staggered. In some embodiments, the nuclease cleaves and generates one or more single strand breaks in the endogenous copy of the essential gene.

[0108] In some embodiments, the one or more double strand breaks (DSBs) are repaired. In some embodiments, the one or more DSBs are repaired to create an altered sequence of the essential gene. In some embodiments, the one or more DSBs are repaired by one or more of non-homologous end joining (NHEJ), microhomology-mediated end joining (MMEJ), homologous recombination (HR), complete HR, and incomplete HR. In some embodiments,

the altered sequence comprises substitutions, insertions, deletions, frame-shifts, or a combination thereof.

[0109] In some embodiments, the DNA sequence modifying enzyme is a base editor. Non-limiting examples of a base editor include cytosine deaminase, and adenine deaminases.

[0110] In some embodiments, the base editor creates one or more base changes in endogenous copy of the essential gene. In some embodiments, the one or more base changes comprise transitions, transversions, or both. In some embodiments, the one or more base changes occur due to tautomerism, depurination, deamidation, or a combination thereof. In some embodiments, the one or more base changes creates an altered sequence of the essential gene. In some embodiments, the one or more base changes comprise one or more point mutations in the endogenous copy of the essential gene. In some embodiments, the one or more point mutations comprise frameshift mutation, nonsense mutation, missense mutation, neutral mutation, silent mutation, or a combination thereof.

[0111] In some embodiments, the promoter of the first gene expresses within females such that the DNA-modifying enzyme produced by the first gene is deposited into eggs and can modify target sequences inherited from a father who lacks the vector. This activity, while unnecessary, for the majority of cases wherein this drive method successfully replaces a population, results in more rapid population replacement than without, for a given fitness cost and/or introduction frequency. Where the DNA-modifying enzyme is a version of Cas9 or a Cas9-related enzyme (guided to a target sequence by a guide RNA), both Cas9 and any and all associated gRNAs are deposited into the eggs of such females together to enable modification of alleles inherited from a non-vector bearing male.

[0112] In some embodiments, there is paternal carryover of the DNA modifying enzyme, allowing for modification of alleles inherited from the mother, even in those who have not inherited the vector.

[0113] In some embodiments, the rescue transgene is a recoded copy of the essential gene. In some embodiments, when the rescue transgene is a recoded copy of the essential gene, the protein encoded by the recoded copy of the essential gene (recoded protein) is about 90% to about 99.9% identical to protein encoded by the endogenous copy of the essential gene (endogenous protein). In some embodiments, the recoded protein is about 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, or 99.9% to the endogenous protein. In some embodiments, the rescue transgene is a gene of unrelated sequence. In some embodiments, when the rescue transgene is a gene of unrelated sequence, the protein encoded by the recoded copy of the essential gene (recoded protein) is functionally equivalent to the protein encoded by the endogenous copy of the essential gene (endogenous protein). In some embodiments, the DNA sequence modifying enzyme does not modify the rescue transgene.

[0114] In some embodiments, the chromosome in which the vector is positioned is one or more of an autosome, X chromosome, Y chromosome, or supernumerary chromosome. In some embodiments, the vector is positioned in one or more combinations of an autosome, X chromosome, Y chromosome, or supernumerary chromosome. For example, in some embodiments, the vector is positioned in an autosome and an X chromosome, in some embodiments, the

vector is positioned in an autosome and a Y chromosome, in some embodiments, the vector is positioned in an autosome and a supernumerary chromosome, in some embodiments, the vector is positioned in an X chromosome and a Y chromosome, in some embodiments, the vector is positioned in an X chromosome and a supernumerary chromosome, in some embodiments, the vector is positioned in an Y chromosome and a supernumerary chromosome, and in some embodiments, the vector is positioned in an autosome, X chromosome, Y chromosome, and supernumerary chromosome.

[0115] In some embodiments, the vector is positioned in an extra-chromosomal element. In some embodiments, the extra-chromosomal element is a plasmid. In some embodiments, the extra-chromosomal element is a virus. In some embodiments, the extra-chromosomal element is a plasmid and a virus. In some embodiments, the vector is positioned in combinations of one or more chromosomes and one or more extra-chromosomal elements.

[0116] In some embodiments, the vector optionally comprises one or more cargo sequences. In some embodiments, the one or more cargo comprise foreign gene sequences, or one or more alleles of an endogenous chromosomal or extra-chromosomal gene.

[0117] In some embodiments the cargo comprises one or more alleles of an endogenous chromosomal or extra-chromosomal gene to which the vector has been linked through nearby insertion on the chromosome or extra-chromosomal element that carries the endogenous allele of interest.

[0118] In some embodiments, the cargo can be physically part of the vector prior to its insertion in a chromosomal or an extra-chromosomal element. In some embodiments, the cargo can be a chromosomal/extra-chromosomal allele of a gene that becomes linked to the vector after the insertion of the vector near that allele. In some embodiments, a fraction of the cargo can be physically part of the vector prior to its insertion in a chromosomal or an extra-chromosomal element, and a remainder of the cargo can be a chromosomal/extra-chromosomal allele of a gene that becomes linked to the vector after the insertion of the vector near that allele. In some embodiments, the cargo does not have to be a part of the vector, i.e., in some embodiments, the cargo is optional and can be physically part of the vector prior to its insertion in a chromosomal or an extra chromosomal element. In some embodiments, the cargo does not have to be a part of the vector, i.e., in some embodiments, a fraction of the cargo can optionally be physically part of the vector prior to its insertion in a chromosomal or an extra chromosomal element, and a remainder of the cargo can be a chromosomal/extra-chromosomal allele of a gene that becomes linked to the vector after the insertion of the vector near that allele.

[0119] In some embodiments herein, the vector comprising the first gene encoding the DNA sequence modifying enzyme and the second gene encoding the rescue transgene is referred to as CLeaveR (e.g., FIG. 6C), which comprises and/or consists of two components: (1) a site-specific DNA modifying enzyme designed to alter the sequence of an endogenous gene required for survival, proliferation, fertility, or differentiation so as to render it non-functional (left); (2) a recoded version of the essential gene resistant to cleavage, and having reduced nucleotide identity with the endogenous gene (right). Optionally, one or more cargo sequences are present (center).

[0120] In some embodiments, DNA sequence modifying enzyme is, without limitation, Cas9, Cas-9-related RNA-guided nucleases, ZFNs, TALENs, homing endonucleases, restriction enzymes, natural site-specific nucleases, engineered site-specific nucleases, base editing enzymes, cytidine deaminase, and adenine deaminase.

[0121] In some embodiments, the vector further comprises one or more additional sequences. In some embodiments, the one or more additional sequences allow the vector to be positioned in the chromosome. In some embodiments, the one or more additional sequences allow the vector to be positioned in the extra-chromosomal element. In some embodiments, the one or more additional sequences allow the vector to be positioned in the chromosome and the extra-chromosomal element. In some embodiments, the one or more additional sequences allow the vector to be positioned in the chromosome but not the extra-chromosomal element. In some embodiments, the one or more additional sequences allow the vector to be positioned in the extra-chromosomal element but not the chromosome.

[0122] In some embodiments, the one or more additional sequences is, without limitations, transposase binding site, LTRs, recombinase binding site, a sequence with homology to a desired location on the chromosome or a sequence with homology to a desired location on the extra-chromosomal element, or combinations thereof.

[0123] In some embodiments, the vector further comprises one or more additional sequences, wherein the one or more additional sequences serve as dominant marker genes that allow individuals carrying the vector to be easily identified either visually, as with expression of a fluorescent protein, or by virtue of surviving a negative selection procedure, as with expression of a gene that encodes resistance to a toxin (such as an antibiotic, insecticide, herbicide), in the presence of the toxin. In some embodiments, the vector comprises one or more sequences that encode marker proteins that can be expressed under the control of suitable regulatory elements. Non-limiting examples of marker proteins include dsRed, GFP, EGFP, CFP, ECFP, BFP, EBFP, mHoneydew, mBanana, mOrange, tdTomato, mTangering, mStrawberry, mCherry, mGrape1, mGrape2, mRaspberry, mPlum, YFP or EYFP, and can be chosen by one of skilled in the art according to need. Fluorescent marker protein can be visualized by illuminating with a suitable excitatory wavelength and observing the fluorescence (e.g., by fluorescence microscopy). In some embodiments, a marker protein would allow for easy identification of organisms carrying the vector.

[0124] In some embodiments, the first promoter is, without limitations, a germline promoter, a male specific germline promoter, a female specific germline promoter, a cell-type specific promoter, a tissue-specific promoter, a ubiquitous promoter, a promoter activated at a specific stage of mitosis, a promoter activated at a specific stage of meiosis, or combinations thereof.

[0125] In some embodiments, the size of the one or more cargo sequences ranges from about is about 0.5 kb to about 500 kb. In some embodiments, the size ranges from about 1 kb to about 1000 kb. In some embodiments, the size ranges from about 5 kb to about 5000 kb. In some embodiments, the size ranges from about 10 kb to about 10000 kb. In some embodiments, the size is about 0.1, 0.5, 1, 5, 10, 25, 50, 75, 100, 250 500, 750 1000, 2500, 5000, 7500, or 10000 kb.

[0126] In some embodiments, the nuclease comprises at least one nuclease domain. In some embodiments, the nuclease comprises one or more DNA binding domains. In some embodiments, the nuclease comprises at least one nuclease domain and one or more DNA binding domains.

[0127] In some embodiments, when the nuclease is Cas9 or a Cas9-related enzyme, the vector further comprise one or more genes encoding a guide RNA. In some embodiments, the guide RNA enables the nuclease to target specific DNA sequences through Watson-Crick base pairing, thereby allowing targeting of very many positions in any genome. In some embodiments, the guide RNA enables the nuclease to target specific sequences within the endogenous copy of the essential gene. In some embodiments, the guide RNA enables the nuclease to target specific sequences within the protein coding region of endogenous copy of the essential gene. In some embodiments, the guide RNA allows the nuclease to target specific sequences within the non-coding region of endogenous copy of the essential gene. In some embodiments, the guide RNA allows the nuclease to target specific sequences outside the endogenous copy of the essential gene.

[0128] In some embodiments, when the nuclease is Cas9, the nuclease domain of Cas9 is inactivated through one or more mutations and the vector comprises a different nuclease domain. In some embodiments, the different nuclease domain is single chain variant of FokI. In some embodiments, when the DNA binding domain is a TALE, the nuclease domain is provided as a single active nuclease domain. In some embodiments, the single active nuclease domain is a single chain variants of FokI. In some embodiments of the vector, when the DNA binding domain is a TALE, the nuclease domain is provided as a single active nuclease domain, such as single chain variants of FokI (Sun and Zhao 2014).

Methods

[0129] One of ordinary skill in the art would appreciate that any of the methods disclosed herein can be performed by any of the vectors provided herein.

[0130] In some embodiments, a method of modifying a population by a vector is provided. In some embodiments, the method comprises obtaining an organism of the population. In some embodiments, the organism is, without limitations, bacteria, archaea, fungi, plants and animals, including rodents, amphibians, mammals, reptiles, insects, mosquitoes, fish, etc.

[0131] In some embodiments, the method comprises positioning the vector in at least one chromosome or extra-chromosomal element in the organism. In some embodiments, the vector is any of the embodiments of the vectors provided herein.

[0132] In some embodiments, the DNA sequence modifying enzyme is expressed in the organism. In some embodiments, the organism is unicellular or multicellular. In some embodiments, when the organism is multicellular, the DNA sequence modifying enzyme is expressed in all cells of the organism. In some embodiments, the DNA sequence modifying enzyme is not expressed in all cells of the multicellular organism. In some embodiments, the DNA sequence modifying enzyme is expressed in a fraction of cells of the multicellular organism. In some embodiments, the DNA sequence modifying enzyme is expressed only in the male or female germline, or in the germline of both sexes.

[0133] In some embodiments, the expression of the DNA sequence modifying enzyme induces one or more sequence modifications. In some embodiments, the expression of the DNA sequence modifying enzyme induces one or more sequence modifications in an essential gene. In some embodiments, the expression of the DNA sequence modifying enzyme induces one or more sequence modifications in an essential gene in one or more cells in the organism. In some embodiments, the one or more sequence modifications result in the essential gene being rendered partially non-functional. In some embodiments, the one or more sequence modifications result in the essential gene being rendered wholly non-functional. In some embodiments, the one or more sequence modifications result in the essential gene being rendered partially non-functional in some circumstances and wholly non-functional in other circumstances. In some embodiments, the result of the essential gene being rendered partially or wholly non-functional is in a defect in the organism. In some embodiments, the defect is, without limitations, a defect in survival, growth control, fertility, differentiation, or combinations thereof.

[0134] In some embodiments, the defect occurs when the one or more cells in which the essential gene being rendered partially or wholly non-functional lack a rescue transgene. In some embodiments, the rescue transgene expresses a recoded protein that rescues the defects in survival, growth control, differentiation, or combinations thereof.

[0135] In some embodiments, the expression of the recoded protein by the rescue transgene results in the generations of an altered organism. In some embodiments, the altered organism expresses the recoded protein in the one or more cells in which the essential gene has been rendered partially non-functional. In some embodiments, the altered organism expresses the recoded protein in the one or more cells in which the essential gene has been rendered wholly non-functional. In some embodiments, the altered organism expresses the recoded protein in the one or more cells in which the essential gene has been rendered partially non-functional in some circumstances and wholly non-functional in other circumstances.

[0136] In some embodiments, the altered organism carries one or more copies of the vector, and wherein the defects in survival, growth control, or differentiation of the one or more cells in which the essential gene has been rendered partially non-functional have been rescued the rescue transgene expressed from the one or more copies of the vector. In some embodiments, the altered organism carries one or more copies of the vector, and wherein the defects in survival, growth control, or differentiation of the one or more cells in which the essential gene has been rendered wholly non-functional have been rescued the rescue transgene expressed from the one or more copies of the vector. In some embodiments, the altered organism carries one or more copies of the vector, and wherein the defects in survival, growth control, or differentiation of the one or more cells in which the essential gene has been rendered partially non-functional in some circumstances and wholly non-functional in other circumstances have been rescued the rescue transgene expressed from the one or more copies of the vector.

[0137] In some embodiments, the altered organism is introduced in a population. In some embodiments, the altered organism is introduced in a population in which an increase in a frequency of the altered organism is desired relative to a frequency of a wild type organism. In some

embodiments, the altered organism is introduced in a population in a particular environment. In some embodiments, the altered organism is introduced in a population in a particular environment in which an increase in a frequency of the altered organism is desired relative to a frequency of a wild type organism in the population in the particular environment. In some embodiments, the altered organisms is introduced in the population such that the percent of the altered organism in the population ranges from about 0.0001% to about 50%. In some embodiments, the percent is about 0.00001, 0.0005, 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 5, 10, 20, 30, 40, or 50%. In some embodiments, the percent is greater than 100%, so as to achieve a more rapid change in the population.

[0138] In some embodiments, introducing the altered organism in the population results in replacement of the wild type organism with the altered organism in the population. In some embodiments, introducing the altered organism in the population results in replacement of the wild type organism with the altered organism in the population in the particular environment.

[0139] In some embodiments, the altered organism exhibits one or more altered traits. In some embodiments, the altered organism introduces and spreads one or more traits of interest in the population. In some embodiments, the one or more traits of interest is a novel trait not previously prevalent in the population, a trait that is a modified version of a trait previously present in the population (e.g., an enhance or a suppressed version of a trait previously present in the population) or a combination thereof. In some embodiments, the population is modified by the introduction of the altered organism in the population. In some embodiments, the population is modified by the introduction of the altered organism in the population in the particular environment. In Non-limiting examples of traits of interest include but are not limited to pathogen resistance, insecticide resistance, environmentally triggered death or sterility, herbicide resistance, fungicide resistance, phage resistance, resistance to viral infection, resistance to abiotic environmental factors, such as cold, heat, and stress.

[0140] In some embodiments, an organism with the defect in survival, growth control, fertility, or differentiation of the one or more cells is eliminated if the one or more cells of the organism lack the rescue transgene.

[0141] In some embodiments, the DNA sequence modifying enzyme does not modify the rescue transgene.

[0142] In some embodiments, rescuing the defects in one or more of survival, growth control, or differentiation is achieved by restoring one or more of normal survival, growth control, fertility, or differentiation of the one or more cells by the rescue transgene.

[0143] In some embodiments, the one or more cells comprise somatic cells, germline cells, gametes, or a combination thereof.

[0144] In some embodiments, the DNA sequence modifying enzyme is a nuclease or a base editor according to the embodiments herein.

[0145] In some embodiments, the nuclease cleaves and generates one or more double strand breaks in the endogenous copy of the essential gene as described herein.

[0146] In some embodiments, the one or more double strand breaks are repaired to create an altered sequence comprising insertions, deletions, base alterations, or a combination thereof.

[0147] In some embodiments, the base editor creates one or more base changes or small insertions/deletions in the endogenous copy of the essential gene.

[0148] In some embodiments, the one or more base changes comprise one or more point mutations, or deamidated bases that are replaced with nucleotides of a different sequence.

[0149] In some embodiments, the altered organism is heterozygous or homozygous for the vector.

[0150] In some embodiments, the organism is haploid. Non-limiting example of haploid organisms include prokaryotes. In some embodiments, the organism is diploid. Non-limiting example of diploid organisms include insects, fungi, many plants and animals. In some embodiments, the organism is polyploidy. Non-limiting example of polyploidy organisms include some fungi and animals and many plants.

[0151] In some embodiments, the organism is selected from the group consisting of prokaryotes, fungi, plants, and animals. In some embodiments, the organism is, without limitations, a prokaryote (bacteria, archaea), fungi, insect, mammal, rodent, fish, amphibian, reptile or plant. In some embodiments, any of the embodiments of the vectors and and/or methods can be one or more of the following: *Autographa gamma* Silver Y moth *Chilo suppressalis* Asiatic rice borer *Diabrotica speciosa* Cucurbit beetle *Harpophora maydis* Late wilt of corn *Helicoverpa armigera* Old world bollworm *Heteronychus arator* Black maize beetle *Peronosclerospora maydis* Java downy mildew *Peronosclerospora philippinensis* Philippine downy mildew *Punctodera chalcoensis* Mexican corn cyst nematode *Sclerophthora rayssiae* var. *zeae* Brown stripe downy mildew *Spodoptera littoralis* Egyptian cottonworm *Spodoptera litura* Cotton cutworm *Thaumatotibia leucotreta* False codling moth *Anthonomus grandis* Boll weevil *Autographa gamma* Silver Y moth *Eutetranychus orientalis* Citrus brown mite *Helicoverpa armigera* Old World bollworm *Oxycarenus hyalinipennis* Cotton seed bug *Pectinophora gossypiella* Pink bollworm *Spodoptera littoralis* Egyptian cottonworm *Spodoptera litura* Cotton cutworm *Thaumatotibia leucotreta* False codling moth *Adoxophyes orana* Summer fruit *tortrix* moth *Aeolesthes sarta* City longhorned beetle *Agrilus biguttatus* Oak splendour beetle *Archips xylosteanus* Variegated golden *tortrix Epiphyas postvittana* Light brown apple moth *Lymantria dispar asiatica* Asian gypsy moth *Lymantria mathura* Rosy moth *Massicus raddei* Mountain oak longhorned beetle *Phytophthora quercina* Oak decline *Platypus quercivorus* Oak ambrosia beetle *Raffaelea quercivora* Japanese oak wilt *Scolytus intricatus* European oak bark beetle *Spodoptera littoralis* Egyptian cottonworm *Thaumatotibia leucotreta* False codling moth *Thaumetopoea processionea* Oak processionary moth *Tortrix viridana* Green oak *tortrix Tremex fuscicornis* Tremex woodwasp Candidatus Phytoplasma pini Pine witches' broom *Cronartium flaccidum* Scots pine blister rust *Dendroctonus micans* Great spruce bark beetle *Dendrolimus pini* Pine-tree lappet *Dendrolimus punctatus* Masson pine moth *Dendrolimus sibiricus* Siberian silk moth *Diprion pini* Pine sawfly *Hylobius abietis* Large pine weevil *Lymantria mathura* Rosy moth *Monochamus saltuarius* Japanese pine sawyer *Monochamus sutor* Small white-marmorated longhorned beetle *Mycosphaerella gibsonii* Needle blight of pine *Panolis flammea* Pine beauty moth *Tomicus destruens* No common name, a pine shoot beetle *Autographa gamma* Silver Y moth *Cernuella virgate* Maritime garden snail

Cochlicella spp. Exotic species *Diabrotica speciosa* Cucurbit beetle *Helicoverpa armigera* Old world bollworm *Heterodera filipjevi* Cereal cyst nematode *Heterodera latipons* Mediterranean cereal cyst nematode *Heteronychus arator* Black maize beetle *Lobesia botrana* European grape vine moth *Meloidogyne artiellia* British root-knot nematode *Nysius huttoni* Wheat bug *Peronosclerospora philippinensis* Philippine downy mildew *Spodoptera littoralis* Egyptian cottonworm *Spodoptera litura* Cotton cutworm *Adoxophyes orana* Summer fruit *tortrix* moth *Alectra vogelii* Yellow witchweed *Autographa gamma* Silver Y moth *Cernuella virgata* Maritime garden snail *Chrysodeixis chalcites* Golden twin spot moth *Crocidosema aporema* Bud borer *Diabrotica speciosa* Cucurbit beetle *Eutetranychus orientalis* Citrus brown mite *Helicoverpa armigera* Old world bollworm *Spodoptera littoralis* Egyptian cottonworm *Adoxophyes orana* Summer fruit *tortrix* moth *Autographa gamma* Silver Y moth Candidatus Phytoplasma australiense Australian grapevine yellows *Cryptoblabes gnidiella Epiphyas* Honeydew moth *postvittana Eupoecilia ambiguella* Candidatus Phytoplasma vitis 1 Light brown apple moth *Heteronychus arator Lobesia botrana Pseudopezicula tracheiphila Spodoptera* European grape berry moth *littoralis Spodoptera litura Thaumatotibia leucotreta* Flavescence doree Black maize beetle European grape vine *Bursaphelenchus cocophilus* Red ring nematode Candidatus Phytoplasma palmae Palm lethal yellowing Cocadviroid Coconut cadang cadang Coconut cadang cadang viroid *Darna pallivitta* Nettle caterpillar *Haplaxius crudus* American palm cixiid *Metamasius hemipterus* West Indian cane weevil *Oryctes rhinoceros* Coconut *rhinoceros* beetle *Paysandisia archon* No common name, a palm borer *Raoiella indica* Red palm mite *Rhabdoscelus obscurus* New Guinea sugarcane weevil *Rhynchophorus ferrugineus* Red palm weevil *Rhynchophorus palmarum* South American palm weevil *Autographa gamma* Silver-Y moth Candidatus Phytoplasma australiense Australian grapevine yellows *Chrysodeixis chalcites* Golden twin spot moth *Globodera pallida* Pale cyst nematode *Globodera rostochiensis* Golden nematode *Helicoverpa armigera* Old world bollworm *Meloidogyne fallax* False Columbia root-knot nematode *Meloidogyne minor* Root-knot nematode *Neoleucinodes elegantalis* Tomato fruit borer *Ralstonia solanacearum* race 3 Bacterial wilt/Southern biovar 2 bacterial Wilt *Spodoptera littoralis* Egyptian cottonworm *Spodoptera litura* Cotton cutworm *Synchytrium endobioticum* Potato wart *Tecia solanivora* Guatemalan potato tuber moth *Thaumatotibia leucotreta* False codling moth *Tuta absoluta* Tomato leaf miner *Adoxophyes orana* Summer fruit *tortrix Argyresthia pruniella* Cherry blossom moth *Bactrocera zonata* Peach fruit fly Candidatus Phytoplasma prunorum European stone fruit yellows *Enarmonia formosana* Cherry bark *tortrix Epiphyas postvittana* Light brown apple moth *Grapholita funebrana* (Syn.: Plum fruit moth *Cydia funebrana*) *Leucoptera malifoliella* Pear leaf blister moth *Lobesia botrana* European grape vine moth *Monilia polystroma* Asiatic brown rot *Monilinia fructigena* Brown rot, Apple brown rot *Potyvirus* Plum Pox Virus Plum pox *Rhagoletis cerasi* European cherry fruit fly *Thaumatotibia leucotreta* False codling moth *Globodera pallida* Pale cyst nematode *Globodera rostochiensis* Golden nematode *Heterodera cajani* Pigeonpea cyst nematode *Heterodera ciceri* Chickpea cyst nematode *Heterodera filipjevi* Cereal cyst nematode *Heterodera latipons* Mediterranean cereal cyst nematode *Heterodera*

sacchari Sugarcane cyst nematode *Punctodera chalcoensis* Mexican corn cyst nematode *Agrilus auroguttatus* Goldspotted oak borer *Agrilus biguttatus* Oak splendour beetle *Agrilus planipennis* Emerald ash borer *Anoplophora chinensis* Citrus longhorned beetle *Anoplophora glabripennis* Asian longhorned beetle *Chlorophorus annularis* Bamboo borer *Chlorophorus strobilicola* Slender-banded pine cone longhorn beetle *Dendroctonus micans* Great spruce bark beetle *Ips sexdentatus* Six-toothed bark beetle *Ips typographus* European spruce bark beetle *Megaplatypus mutatus* No common name, an ambrosia beetle *Monochamus alternatus* Japanese pine sawyer *Monochamus saltuarius* Japanese pine sawyer *Monochamus sutor* Small white-marmorated longhorned beetle *Orthotomicus erosus* Mediterraneran pine engraver *Pityogenes chalcographus* Sixtoothed spruce bark beetle *Platypus quercivorus* Oak ambrosia beetle *Scolytus intricatus* European oak bark beetle *Tetropium castaneum* Black spruce beetle *Tetropium fuscum* Brown spruce longhorned beetle *Tomicus destruens* No common name, a pine shoot beetle *Tomicus minor* Lesser pine shoot beetle *Tomicus piniperda* Pine shoot beetle *Trichoferus campestris* Velvet longhorned beetle *Trypodendron* European hardwood ambrosia beetle domesticum Redbay ambrosia beetle *Belocaulus* spp. No common name, leatherleaf slugs *Cernuella* spp. No common name, hygromiid snails *Cochlicella* spp. No common name, cochlicellid snails *Colosius* spp. No common name, leatherleaf slugs *Laevicaulis* spp. No common name, leatherleaf slugs *Lissachatina fulica* Giant African snail *Meghimatium pictum* Chinese slug *Monacha* spp. No common name, hygromiid snails *Sarasinula* spp. No common name, leatherleaf slugs *Semperula* spp. No common name, leatherleaf slugs *Veronicella* spp. No common name, leatherleaf slugs *Dendrolimus pini* Pine-tree lappet *Dendrolimus punctatus* Masson pine moth *Dendrolimus sibiricus* Siberian silk moth *Lymantria albescens* Okinawa gypsy moth *Lymantria dispar asiatica* Asian gypsy moth *Lymantria dispar japonica* Japanese gypsy moth *Lymantria mathura* Rosy moth *Lymantria monacha* Nun moth *Lymantria postalba* White-winged gypsy moth *Lymantria umbrosa* Hokkaido gypsy moth *Lymantria xylina* Casuarina tussock moth.

[0152] In some embodiments, an insect can be a direct pest or indirect pest. A "direct pest" refers to insects that can cause damage at one or more stage of their life cycle by, for example, eating crops or damaging animals. The New World screw-worm fly *Cochliomyia hominivorax,* for example, is a direct pest of cattle, and the spotted wing *Drosophila, Drosophila suzukii* is pest of many fruit crops. An "indirect pest" refers to insects that transmit human diseases, for example, mosquitoes which carry malaria. Indirect pests of organisms other than humans, such as livestock or plants are also known.

[0153] In some embodiments, insect refers to, without limitations, one or more of *Drosophila,* mosquitoes, bumblebees, hoverflies, grasshoppers, dragonfly, dancefly, weevil, cricket, wasp, moth, beetles, honey bee, robberfly or butterfly. Additional examples of insects include, but are not limited to, Asian citrus psyllid (*diaphorini citrii,* Australian sheep blowfly (*Lucilia cuprina,* Asian tiger mosquito (*Aedes albopictus*); Japanese beetle (*Popilla japonica*), White-fringed beetle (*Graphognatus* spp.), Citrus blackfly (*Aleurocanthus woglumi*), Oriental fruit fly (*Dacus dorsalis*), Olive fruit fly (*Dacus oleae*), tropical fruit fly (*Dacus cucurbitae, Dacus zonatus*), Mediterranean fruit fly (*Cer-*

atitis capitata), Natal fruit fly (*Ceratitis rosa*), Chemy fruit fly (*Rhagoletis cerasi*), Queensland fruit fly (*Bactrocera tryoni*), Caribbean fruit fly (*Anastrepha suspensa*), imported fire ants (*Solenopis richteri, Solenopis invictai,* Gypsy moth (*Lyman tria dispar*), Codling moth (*Cydia pomonella*), Brown tail moth (*Euproctis chrysorrhoea*), yellow fever mosquito (*Aedes aegypti*), malaria mosquitoes (*Anopheles gambiae, Anopheles stephensi*), New world screwworm (*Cochliomyia hominivorax*), Old World Screwworm (*Chrysomya bezziana*), Tsetse fly (*Glossina* spp), Boll weevil (*Anthonomous grandis*), Damsel fly (*Enallagma hageni*), Dragonfly (*Libellula luctuosa*), and rice stem borer (*Tryporyza incertulas*). In some embodiments, the insect either transmits human disease or are agricultural pests. In some embodiments, the insects are wild insect populations.

[0154] In some embodiments, the insects are mosquitoes or flies (for example fruit flies, tsetse flies, sand flies). The mosquitoes can be, for example, *Aedes* sp. or *Anopheles* sp. In some embodiments, the mosquito is yellow fever mosquito (*Aedes aegypti*), malaria mosquito (*Anopheles gambiae, Anopheles stephensi*), Asian tiger mosquito (*Aedes albopictus*) or Culex mosquitoes. In some embodiments, the insect is one that transmits a disease of a mammal. The disease can be any disease, for example, malaria and/or yellow fever. In some embodiments, the insect is a Spotted wing *Drosophila* (*Drosophila Suzukii*).

[0155] In some embodiments, insect refers to an insect that spreads a disease of humans. In some embodiments, insect refers to an insect that spreads a disease of economically important animals. In some embodiments, insect refers to an insect that spreads a disease of companion animals. In some embodiments, insect refers to an insect that spreads a disease of plants.

[0156] In some embodiments, mosquitoes can be, without limitations, of *Aedes, Anopheles, Culex, Coquillettidia, Haemagogus, Mansonia, Ochlerotatus, Psorophora* or other genera that transmit diseases. In some embodiments, the diseases transmitted by mosquitoes can be one or more of Malaria, Chikungunya, Dog Heartworm, Dengue, Yellow Fever, Eastern Equine Encephalitis, St. Louis Encephalitis, LaCrosse Encephalitis, Western Equine Encephalitis, West Nile Virus, or Zika Virus, lymphatic filariasis.

[0157] In some embodiments, the population has about 10,000 to about 100,000,000,000 organisms. In some embodiments, the population has about 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500, 10,000, 100,000, 500,000, 1,000,000, 100,000,000, 1,000,000,000, 100,000, 000,000 or 1,000,000,000,000 organisms, or a number within a range defined by any two of the aforementioned values.

[0158] In some embodiments, the environment comprises an open environment, a bioreactor, or a multicellular body, a closed container, or combinations thereof. In some embodiments, the environment is a combination of an open environment, a bioreactor, or a multicellular body, a closed container, and the environment changes sequentially from one to the other.

[0159] In some embodiments, the wild type organism is replaced at a high frequency with the altered organism in the population wherein the wild type organism is present. In some embodiments, the wild type organism is replaced at a high frequency with the altered organism in the population in a particular environment wherein the wild type organism

is present. In some embodiments, high frequency is defined as replacement of at least 90% of the wild type organism with the altered organism after 50 generations. In some embodiments, high frequency is defined as replacement of at least 80% of the wild type organism with the altered organism after 50 generations. In some embodiments, high frequency is defined as replacement of at least 70% of the wild type organism with the altered organism after 50 generations. In some embodiments, high frequency is defined as replacement of at least 60% of the wild type organism with the altered organism after 50 generations. In some embodiments, high frequency is defined as replacement of at least 50% of the wild type organism with the altered organism after 50 generations.

[0160] In some embodiments, the wild type organism is replaced at a rapid rate with the altered organism in the population wherein the wild type organism is present. In some embodiments, the wild type organism is replaced at a rapid rate with the altered organism in the population in a particular environment wherein the wild type organism is present. In some embodiments, rapid rate is defined as replacement of at least 90% of the wild type organism with the altered organism after 50 generations. In some embodiments, rapid rate is defined as replacement of at least 80% of the wild type organism with the altered organism after 50 generations. In some embodiments, rapid rate is defined as replacement of at least 70% of the wild type organism with the altered organism after 50 generations. In some embodiments, rapid rate is defined as replacement of at least 60% of the wild type organism with the altered organism after 50 generations. In some embodiments, rapid rate is defined as replacement of at least 50% of the wild type organism with the altered organism after 50 generations.

[0161] In some embodiments, at least 90% of the wild type organism is replaced with the altered organism after 50 generations. In some embodiments, at least 90% of the wild type organism is replaced with the altered organism after 50 generations. In some embodiments, at least 80% of the wild type organism is replaced with the altered organism after 50 generations. In some embodiments, at least 70% of the wild type organism is replaced with the altered organism after 50 generations. In some embodiments, at least 60% of the wild type organism is replaced with the altered organism after 50 generations. In some embodiments, at least 50% of the wild type organism is replaced with the altered organism after 50 generations.

[0162] In some embodiments, the one or more sequence modifications in the one or more cells is a result of the one or more cells carrying the first gene encoding the DNA sequence modifying enzyme. In some embodiments, the one or more sequence modifications in the one or more cells is a result of the DNA sequence modifying enzyme being transmitted to the one or more cells from one or more cells expressing the DNA sequence modifying enzyme through diffusion. In some embodiments, the one or more sequence modifications in the one or more cells is a result of the DNA sequence modifying enzyme being transmitted to the one or more cells from one or more cells expressing the DNA sequence modifying enzyme through active transport. In some embodiments, the one or more sequence modifications in the one or more cells is a result of the one or more cells carrying the first gene encoding the DNA sequence modifying enzyme and is a result of the DNA sequence modifying enzyme being transmitted to the one or more cells from one

or more cells expressing the DNA sequence modifying enzyme through diffusion. In some embodiments, the one or more sequence modifications in the one or more cells is a result of the one or more cells carrying the first gene encoding the DNA sequence modifying enzyme and is a result of the DNA sequence modifying enzyme being transmitted to the one or more cells from one or more cells expressing the DNA sequence modifying enzyme through active transport. In some embodiments, the one or more sequence modifications in the one or more cells is a result of the DNA sequence modifying enzyme being transmitted to the one or more cells from one or more cells expressing the DNA sequence modifying enzyme through diffusion and is a result of the DNA sequence modifying enzyme being transmitted to the one or more cells from one or more cells expressing the DNA sequence modifying enzyme through active transport. In some embodiments, the one or more sequence modifications in the one or more cells is a result of the DNA sequence modifying enzyme being transmitted to the one or more cells from one or more cells expressing the DNA sequence modifying enzyme through intercellular movement.

[0163] In some embodiments, the vector is positioned in the chromosome or the extra-chromosomal element by a homologous recombination-dependent integration, random integration, integration using transposition, integration using a recombinase, or combinations thereof.

[0164] In some embodiments, the one or more cargo sequences comprise a one or more foreign gene sequences, or one or more alleles of an endogenous chromosomal or extra-chromosomal gene to which the vector has been linked through nearby insertion on the chromosome or extra-chromosomal element that carries the endogenous allele of interest.

[0165] In some embodiments, when the vector is positioned on the chromosome or the extra-chromosomal element, the first gene operably linked to the first promoter, the second gene operably linked to the second promoter, and the one or more cargo transgenes are genetically linked.

[0166] In some embodiments of the method, the vector and cargo are located in a small chromosomal inversion. In some embodiments of the method, the vector and cargo are located in a small chromosomal inversion further limits the possibility that the vector and cargo can separate from each other during any stage of DNA replication, mitosis, and/or or meiosis.

[0167] In some embodiments, the nuclease cleaves and generates one or more double strand breaks in the endogenous copy of the essential gene with a high cleavage efficiency. In some embodiments, the high cleavage frequency is defined as at least 30% of individuals carrying the nuclease cleave the endogenous copy of the essential gene in each generation. In some embodiments, the high cleavage frequency is defined as at least 40% of individuals carrying the nuclease cleave the endogenous copy of the essential gene being cleaved in each generation. In some embodiments, the high cleavage frequency is defined as at least 50% of individuals carrying the nuclease cleave the endogenous copy of the essential gene being cleaved in each generation. In some embodiments, the high cleavage frequency is defined as at least 60% of individuals carrying the nuclease cleave the endogenous copy of the essential gene being cleaved in each generation.

[0168] In some embodiments, the high cleavage frequency is defined as the nuclease cleaving one or more alleles of the endogenous copy of the essential gene in at least 30% of organisms carrying the vector and the one or more alleles of the endogenous copy of the essential gene in each generation. In some embodiments, the high cleavage frequency is defined as the nuclease cleaving one or more alleles of the endogenous copy of the essential gene in at least 40% of organisms carrying the vector and the one or more alleles of the endogenous copy of the essential gene in each generation. In some embodiments, the high cleavage frequency is defined as the nuclease cleaving one or more alleles of the endogenous copy of the essential gene in at least 50% of organisms carrying the vector and the one or more alleles of the endogenous copy of the essential gene in each generation. In some embodiments, the high cleavage frequency is defined as the nuclease cleaving one or more alleles of the endogenous copy of the essential gene in at least 60% of organisms carrying the vector and the one or more alleles of the endogenous copy of the essential gene in each generation.

[0169] In some embodiments, the base editor creates one or more base changes in endogenous copy of the essential gene with a high base editing frequency. In some embodiments, the high base editing frequency is defined as base editing in at least 20% of organisms that carry the vector in each generation. In some embodiments, the high base editing frequency is defined as base editing in at least 30% of organisms that carry the vector in each generation. In some embodiments, the high base editing frequency is defined as base editing in at least 40% of organisms that carry the vector in each generation. In some embodiments, the high base editing frequency is defined as base editing in at least 50% of organisms that carry the vector in each generation.

[0170] In some embodiments, the high base editing frequency is defined as the base editor modifying one or more alleles of the endogenous copy of the essential gene in at least 20% of the organisms carrying the vector and the one or more alleles of the endogenous copy of the essential gene in each generation. In some embodiments, the high base editing frequency is defined as the base editor modifying one or more alleles of the endogenous copy of the essential gene in at least 30% of the organisms carrying the vector and the one or more alleles of the endogenous copy of the essential gene in each generation. In some embodiments, the high base editing frequency is defined as the base editor modifying one or more alleles of the endogenous copy of the essential gene in at least 40% of the organisms carrying the vector and the one or more alleles of the endogenous copy of the essential gene in each generation. In some embodiments, the high base editing frequency is defined as the base editor modifying one or more alleles of the endogenous copy of the essential gene in at least 50% of the organisms carrying the vector and the one or more alleles of the endogenous copy of the essential gene in each generation.

[0171] In some embodiments of the method, the nuclease cleaves and generates one or more double strand breaks in the endogenous copy of the essential gene with a high cleavage efficiency. In some embodiments of the method, the high cleavage frequency is defined as the nuclease cleaving the endogenous copy of the essential gene in at least 30% of the organisms carrying the vector in each generation. In some embodiments of the method, the base editor creates one or more base changes in the endogenous

copy of the essential gene with a high base editing frequency. In some embodiments of the method, the high base editing frequency is defined as the base editor modifying the endogenous copy of the essential gene in at least 20% of the organisms carrying the vector in each generation.

[0172] In some embodiments, the promoter of the first gene is a female-specific promoter such that the first gene encoding the DNA sequence modifying enzyme is expressed within females only. In some embodiments, female-specific expression of the DNA sequence modifying enzyme results in the DNA-modifying enzyme being present in the eggs. In some embodiments, when an egg expressing the DNA sequence modifying enzyme is fertilized by a male gamete, the DNA sequence modifying enzyme from the egg can modify target sequence in the paternal copy provided by the father. In some embodiments, there is paternal carryover wherein sperm contribute DNA modifying activity to eggs, resulting in modification of the copy of the target sequence provided by the mother. In some embodiments, there is potential for carryover. In some embodiments, modification of an allele in a fertilized egg is achieved even when the allele is inherited from a parent that did not carry the vector.

[0173] In some embodiments, as used herein, "fitness cost" is defined as a relative reduction in the number of offspring produced by, or survival of, individuals carrying the transgenic construct, as compared with wild type individuals. In some embodiments fitness cost is defined as a relative reduction in the number of offspring produced by, or survival of, individuals not carrying the transgenic construct, as compared with those who do. In some embodiments, fitness benefit is defined as a relative increase in the number of offspring produced by, or survival of, individuals carrying the transgenic construct as compared with wild type individuals.

[0174] In some embodiments, the first gene expresses within females (the female germline or cells that contribute components to the female germline), such that the DNA-modifying enzyme and any associated cofactors such as guide RNAs are deposited into all oocytes/eggs, and modify target sequences in the version of the essential gene provided by the father. This represents maternal carryover of DNA sequence modifying activity.

[0175] In some embodiments, paternal carryover of the DNA modifying enzyme results in modification of the maternal copy of the essential locus in eggs, including those that do not inherit the vector.

Applications

[0176] In some embodiments, the methods provided herein can be applied for modification of populations for beneficial outcomes. For example, in some embodiments, to prevent mosquito-borne diseases (e.g., malaria, dengue, etc.), mosquitoes can be engineered based on the embodiments of the vectors and methods disclosed herein to resist infection. The engineered mosquito can be used to replace wild mosquito population in order to achieve less transmission and less disease. Such a trait (e.g., refractoriness of the engineered mosquito to disease transmission) is unlikely to spread into a population in the absence of gene drive because the trait results in a fitness cost to carriers. A gene drive solution to this problem described herein is to increase the fitness cost associated with not carrying the gene of interest through DNA sequence modification-based gene drive.

[0177] In comparison to other low threshold gene drive systems (Example,12-14), the Cas9 based gene drive mechanisms in Examples 1-5 do not require any homing to occur (homing is known to vary in its relative rate compared to other forms of DNA repair in different species), and they are predicted to rapidly take over wild type populations even when the associated cargo results in significant fitness costs. The presently proposed DNA sequence modification-based drives are all predicted to replace wild type populations quickly while bearing substantial fitness costs, and each of the five drives displays characteristics that qualify them for different scenarios.

[0178] While all five of these mechanisms have been considered in the context of Cas9, these drive results could apply to any endonuclease or base editor used to disrupt the function of an endogenous gene. For some embodiments, one of the biggest advantages of these drives is their adaptability to new species. This is because the primary requirements are the identification of an essential gene (thousands in each organism), a recoded or sequence unrelated version of the gene (including associated regulatory sequences) that has wildtype or close to wildtype function, and a promoter and DNA sequence-modifying enzyme capable of bringing about sequence alteration of the endogenous copy of the essential gene in the germline, or germline and early embryo cells exposed to the enzyme.

Additional Embodiments

[0179] Without being limited by any particular theory, one implementation of a DNA sequence-based modification-based gene drive is as follows: a cell expresses a DNA sequence modifying enzyme that alters the sequence of an essential gene, inactivating it. The DNA sequence modifying enzyme is transmitted through cytoplasm to offspring (either maternally, paternally, or both), where it modifies the target gene, regardless of whether the gene encoding the DNA sequence modifying enzyme is transmitted to these progeny. Progeny that inherit the DNA sequence modifying enzyme-encoding gene also inherit a rescue copy of the wildtype gene that has been cleaved. This rescue copy is both functional and uncleavable. In this way key features required for gene drive are brought about

[0180] In some embodiments, the above system is applicable to insects. A DNA sequence modifying enzyme is expressed under the control of a germline promoter. The promoter may be expressed in both the male and female germline. However, let us also consider a case in which the nuclease is expressed under the control of a late female germline specific promoter. In this case the DNA sequence modifying enzyme is transmitted from the oocyte, where its mRNA (and any associated co-factors such as gRNAs) is synthesized, to the mature oocyte/fertilized egg. In the zygote (fertilized egg) the DNA sequence modifying enzyme alters wildtype copies of the gene, resulting in their inactivation. This inactivation can occur in the nuclei that will ultimately give rise to cells of various somatic tissues of the animal. It can also occur in the cells that will give rise to the embryonic germline. Without being limited by any particular theory, provided that endogenous copies of the essential gene are altered in a sufficient number of nuclei, and are inactivated in both copies (for diploid organisms), which can happen early in embryogenesis (at the single diploid nucleus stage) or later, after some number of nuclei have been generated, then development will be disrupted,

17

resulting in animal death. However, if the zygote inherits along with the DNA sequence modifying enzyme-encoded gene a tightly linked copy of the rescue transgene that cannot be modified, the progeny will survive, even if both copies of the wildtype copy of the gene have been modified. This occurs because for most genes in diploids heterozygosity for one wildtype copy of the gene is sufficient to provide enough function to allow the organisms to survive and thrive. Good evidence for this conclusion comes from several sources: the many examples of phenotypically normal heterozygotes in many species; and the ability to create and maintain healthy stocks for deletions that eliminate, one at a time, one copy of most regions of the *drosophila* genome (flybase.org). If there is a modest fitness cost associated with heterozygosity this will be decreased over time as the construct spreads. This is because as spread occurs the frequency of homozygotes for the construct rises, in which case individuals now carry two copies of the rescue gene of interest and are therefore have increased fitness.

[0181] Without being limited by any particular theory, the model can be generalized and extrapolated to prokaryotes or other haploids carrying a plasmid that encodes a DNA sequence modifying enzyme and a recoded or sequence unrelated version of an essential gene. In this case progeny that fail to inherit the plasmid will still inherit the chromosomal mutation that results in loss of function of the wildtype copy of the gene. They may also inherit the DNA sequence modifying enzyme, in which case the sequence of any wildtype copies of the essential gene (incorporated through horizontal gene transfer, transduction, transformation, or conjugation) will be altered and the cell will die. However, those cells that inherit the plasmid inherit a functional copy of the gene, even though the chromosomal version of the gene has been altered.

[0182] Without being limited by any particular theory, the model can be generalized and extrapolated to organisms such as yeast that go from a haploid to diploid phase and back to haploid through sporulation. A chromosomal copy of the DNA sequence modifying enzyme and a recoded or sequence unrelated version of the rescue will be transmitted to only some progeny during sporulation. Those haploids that fail to inherit the rescue copy of the gene will die because the DNA sequence modifying enzyme, which is transmitted through cytoplasm, will cause alteration of the wildtype copy. The wildtype copy of the gene will likely also have been cleaved during the diploid stage in which case cytoplasmic inheritance of the nuclease is not essential. In any case, only haploids that inherit the tightly linked rescue construct will survive. This constitutes a kind of gamete killing. Most generally, the system described applies to any biological situation in which a DNA sequence modifying enzyme alters the sequence of an essential gene so as to disrupt essential functions in haploid, diploid or polyploid cells. This modification can occur in the parental cell, which can be haploid, diploid, tetraploid, or polyploid. Alternatively the DNA sequence modifying enzyme, the transcript and/or protein for which is produced in the parental cell, can alter the sequence of the essential gene in the progeny cells in which it becomes located through cytoplasmic diffusion or active transport. The operative principle in all cases is that in the relevant cell type, or in a multicellular organism, in some fraction of these cells, all or most copies of the endogenous copies of the essential gene are altered so as to produce non-functional copies of the gene. This results in

death of all cells that fail to inherit one or more copies of the rescue transgene. The DNA sequence modifying enzyme and the rescue transgene are tightly linked and behave as a single genetic unit.

[0183] In some embodiments, the model is extrapolated to a diploid animal such as a rodent, mosquito, fish, amphibian, or other organism in which spermatogenesis utilizes haploid-specific promoters to drive the expression of genes essential for spermatogenesis. In some embodiments, the DNA sequence modifying enzyme is expressed in the germline at some point. It is not critical when, or in which sex. What matters is that ultimately one will end up with post-meiotic cells that carry a copy of the rescue transgene, while their post-meiotic brothers do not. To the extent that it is true that the product of the rescue transgene, which will have all the endogenous regulatory sequences of the endogenous gene, does not diffuse into the meiotic brothers to which they are still connected by cytoplasmic bridges, those sperm will die, be resorbed, or be ejaculated in a state that is non-functional. This will result in nuclease and rescue transgene-bearing meiotic products being preferentially represented in the next generation, a form of population replacement.

[0184] In some embodiments, a rescue version of a post-meiotic expressed gene that is normally present on an autosome can be expressed from the Y chromosome along with the DNA sequence modifying enzyme. Provided the DNA sequence modifying enzyme alters the wildtype endogenous copy of the gene in the germline then only Y-bearing sperm will generate this product. This holds even if the haplo-expressed gene is on the autosome (FIG. 11). In early generations there may be some reduced sex ratio bias if some wildtype copies are not cleaved, and depending on when in germline development the nuclease acts (before or after this generations post-meiotic expression). However, the bottom line is the same. Eventually, wildtype copies of the haplo-expressed gene are lost and the only remaining functional copies are those on the Y chromosome. This can result in sex ratio distortion if the sperm in which the gene has been inactivated are unable to carry out fertilization.

[0185] In some embodiments, the model is applicable to chickens with ZW sex chromosomes. W is the sex chromosome. Males are ZZ. A W chromosome that carries a rescue cassette and a nuclease. It is inherited only into females. Males that inherit the Z chromosome inherit a cleaved copy of an essential Z gene, or cleaved copies of an autosomal essential gene. In any case, ultimately a population in which there are only females is obtained because males do not inherit a rescue construct. The male eggs simply do not develop. A big male egg is still obtained because the actual embryo with chickens is quite small. However, baby male chickens are not obtained. Ultimately female chickens carrying the rescue construct and no wildtype copies of the gene are mated with wildtype males. Female progeny survive. Male progeny do not survive if there is maternal carryover that causes killing of the wildtype loci inherited from the male. If W-bearing females are mated with to wildtype males, which is what is done in a breeding or hybrid generation situation, the males will all die if the gene that is essential is normal on the Z and there is enough maternal carryover of the DNA sequence modifying enzyme to cause the wildtype copy of whatever chromosome carries the wildtype copy of the gene from males to undergo sequence modification such that males inherit no good copies of the essential gene.

EXAMPLES

[0186] Outlined in Examples 1-5 are the designs of five proposed cleavage mediated gene drives. Discrete generation, deterministic population frequency models were developed for each of the five drive mechanisms that demonstrate the range of fitness costs and Cas9 cleavage efficiencies for which they will take over a wildtype population.

Example 1

X Chromosome Cleavage Mediated Y Chromosome Drive

[0187] X chromosome cleavage mediated Y chromosome Drive (also referred to herein as X cleavage mediated Y drive) consists of Cas9, gRNAs which target an essential (i.e. recessive lethal) gene on the X chromosome, and a recoded copy of this target X gene which is immune to gRNA targeting, which are situated together at the same locus on the Y chromosome (FIG. 1A). The transgenic construct (TY) is situated on the Y chromosome and consists of Cas9 (long green rectangle), gRNAs (short green rectangle) targeting an essential gene on the X chromosome, and a recoded version of the target gene (yellow rectangle with recoded gRNA target sites in orange) (FIG. 1A). Potential cleavage sites on the target essential gene (X) are indicated by dashed lines and scissors, and the cleaved locus (C) is a null form of the target gene made of what remains of the gene from the outer ends of the cleavage sites (FIG. 1A).

[0188] In males who carry this construct (TY) and a normal X chromosome (X), the target gene is cleaved multiple times during spermatogenesis, destroying the wild type copy of the gene on the X chromosome (C) and resulting in either TY or C bearing sperm (FIG. 1B). In transgenic males that bear wild type X chromosomes, Cas9 (stacked green square and yellow square with orange bars) can find and cleave a copy of the target gene (yellow square). The resulting cleaved locus (yellow bar) is passed on instead of the original target wildtype locus. When two individuals bearing a cleaved locus mate and the cleaved X loci are paired together (CC) or when the cleaved locus is passed on to a male (CY), the resulting offspring is unviable, removing wild type alleles are from the population (FIG. 1B) As TY males mate with wild type females, C's will begin to accumulate in heterozygotes (CX). All CY males and all CC females will die from the absence of a functional copy of the target essential X gene, leaving the viable genotypes CTY, XTY, XY, CX, and XX (FIG. 1B). Events proceed from left to right. The vector on the Y expresses a site-specific nuclease (dark square) and a rescue transgene (light square). The nuclease has the ability to cleave a wildtype version of the essential gene on the X at multiple positions (scissors). Cleavage does not necessarily happen in somatic cells. The left-most panel (X,Ty) simply indicates where cleavage occurs. Cleavage occurs in germline cells (CTy), resulting in the creation of an X chromosome that lacks a functional copy of the essential gene (thin light line). When a male carrying these chromosomes mates with a wildtype female new opportunities for cleavage of a wildtype X are created (second line). In the third generation matings are shown that result in the death of several genotypes.

[0189] The discrete generation, deterministic population frequency model for this drive mechanism demonstrates that if Cas9 cleaves the target gene with 100% efficiency, TY can drive to fixation amongst Y chromosomes with just a few moderate releases of CTY males while bearing a fitness cost of up to approximately 45% (FIG. 1C). TY can still drive male replacement when Cas9 is cleaving at non-optimal rates, but it can only tolerate correspondingly reduced fitness costs as a result (FIG. 1C). Discrete generation, deterministic population frequency modeling of X cleavage mediated Y drive is shown in FIG. 1C. Each data point uses a few moderate releases of transgenic mosquitoes (three releases of CTY males at 50% of the population) with the specified fitness cost and Cas9 cleavage efficiency. The color of each data point indicates the number of generations (as indicated by the colorbar) before TY bearing individuals make up >99% of all males. White indicates the inability of TY to take over under the specified conditions or failure to do so within 70 generations (FIG. 1C).

[0190] The X CM Y drive is capable of quickly driving a transgene to fixation on the Y chromosome while bearing ~40% fitness costs at high cleavage efficiency. As males are the only transgenics, it cannot be used as a replacement mechanism for attacking mosquitoes because only the females are vectors. However, it can still be useful in the context of suppression if the cargo is a lethal gene under an environmentally triggered promoter. In this way, the transgene can spread to fixation in males, killing all males once the environmental trigger activates and resulting in a population crash. Alternatively, this construct can be used in ZW species where the female is the heterogametic sex, such as the pink bollworm.

Example 2

Cleavage Mediated X Drive

[0191] Cleavage mediated X drive consists of Cas9, gRNAs which target an essential gene on the X chromosome, and a recoded or sequence unrelated copy of this target X gene which is immune to gRNA targeting, which are situated together at the same locus as the target gene (FIG. 2A). The transgenic construct (TX) is situated on the X chromosome and consists of Cas9 (long green rectangle), gRNAs (short green rectangle) targeting an essential gene on the X chromosome (at the same locus as TX), and a recoded version of the target gene (yellow rectangle with recoded gRNA target sites in orange) (FIG. 2A). Potential cleavage sites on the target essential gene (X) are indicated by dashed lines and scissors, and the cleaved locus (C) is a null form of the target gene made of what remains of the gene from the outer ends of the cleavage sites (FIG. 2A).

[0192] In females who carry this construct (TX) and a normal X chromosome (X), the target gene is cleaved multiple times during oogenesis, destroying the wild type copy of the gene on the X chromosome (C) and resulting in either TX or C bearing eggs (FIG. 2B). In transgenic females that bear wild type X chromosomes (TX X) Cas9 can find and cleave a copy of the target gene. The resulting cleaved locus is passed on instead of the original target wildtype locus. When the cleaved locus is passed on to a male (CY), the resulting offspring is unviable, removing a wild type allele are from the population (FIG. 2B). As transgenic individuals mate with wild types, cleaved copies of the essential X gene will begin to accumulate in females (CX). All males that receive a cleaved X chromosome (CY) will die from the absence of a functional copy of the target

essential X, leaving the viable genotypes TXY, XY, TXTX, TXC, TXX, CX, and XX (FIG. 2B).

[0193] The discrete generation, deterministic population frequency model for this drive mechanism demonstrates that if Cas9 cleaves the target gene with 100% efficiency, TX can drive to fixation with just a few moderate releases of TXY males while bearing a fitness cost of up to approximately 35% (FIG. 2C). TX can still drive population replacement when Cas9 is cleaving at non-optimal rates, but it can only tolerate correspondingly reduced fitness costs as a result (FIG. 2C). Discrete generation, deterministic population frequency modeling of cleavage mediated X drive is shown in FIG. 2C. Each data point uses a few moderate releases of transgenic mosquitoes (three releases of TXY males at 50% of the population) with the specified fitness cost and Cas9 cleavage efficiency. The color of each data point indicates the number of generations (as indicated by the colorbar) before TX bearing individuals make up >99% of the population. White indicates the inability of TX to take over under the specified conditions or failure to do so within 70 generations (FIG. 2C).

[0194] The X drive can tolerate ~35% fitness costs at high cleavage efficiency. This drive is well suited to replacement in XY species of mosquitoes such as *Anopheles gambiae.*

Example 3

Autosomal Cleavage Mediated Autosomal Drive

[0195] Cleavage mediated autosomal drive consists of Cas9, gRNAs which target an essential autosomal gene, and a recoded or sequence unrelated copy of this target gene which is immune to gRNA targeting, which are situated together at the same locus as the target gene (FIG. 3A). The transgenic construct (T) is situated on an autosome and consists of Cas9 (long green rectangle), gRNAs (short green rectangle) targeting an essential gene (at the same autosomal locus as T), and a recoded version of the target gene (yellow rectangle with recoded gRNA target sites in orange) (FIG. 3A). Potential cleavage sites on the target essential gene (A) are indicated by dashed lines and scissors, and the cleaved locus (C) is a null form of the target gene made of what remains of the gene from the outer ends of the cleavage sites (FIG. 3A).

[0196] In males and females who carry the construct (T) and a wild type copy of the its target (A), the target gene is cleaved multiple times during gametogenesis, destroying the wild type copy of the gene (C) and resulting in either T or C bearing gametes (FIG. 3B). As transgenic individuals mate with wild types, cleaved copies of the essential gene will begin to accumulate in heterozygotes (CA individuals). All individuals that receive two cleaved autosomes (CC) will die from the absence of a functional copy of the target essential autosomal gene, leaving the viable genotypes TT, TC, TA, CA, and AA (FIG. 3B). In heterozygotes (TA) Cas9 can find and cleave a copy of the target gene. The resulting cleaved locus is passed on instead of the original target wildtype locus. When two individuals bearing cleaved locus mate and the cleaved loci are paired together (CC), the resulting offspring is unviable, removing two wild type alleles are from the population (FIG. 3B).

[0197] The discrete generation, deterministic population frequency model for this drive mechanism demonstrates that if Cas9 cleaves the target gene with 100% efficiency, T can drive to fixation with just a few moderate releases of TT

males while bearing a fitness cost of up to approximately 55% (FIG. 3C). T can still drive population replacement when Cas9 is cleaving at non-optimal rates, but it can only tolerate correspondingly reduced fitness costs as a result (FIG. 3C). Discrete generation, deterministic population frequency modeling of cleavage mediated autosomal drive is shown in FIG. 3C. Each data point uses a few moderate releases of transgenic mosquitoes (three releases of TT males at 50% of the population) with the specified fitness cost and Cas9 cleavage efficiency. The color of each data point indicates the number of generations (as indicated by the colorbar) before T bearing individuals make up >99% of the population. White indicates the inability of T to take over under the specified conditions or failure to do so within 70 generations (FIG. 3C).

[0198] The autosomal drive is very potent, capable of driving even with ~55% fitness costs at high cleavage efficiency. Because the construct is autosomal, it can be used to drive replacement in any species, importantly covering both *Anopheles gambiae* and *Aedes aegypti.* It is also perhaps the easiest to implement, as the only knowledge it requires about the target species are an essential gene on an autosome and an appropriate promoter to drive expression of the DNA sequence modifying enzyme (either pre-meiotic or gametogenic).

Example 4

Cleavage Mediated 2-Locus Autosomal Drive

[0199] Cleavage mediated 2-locus autosomal drive consists of Cas9, gRNAs which target an essential autosomal gene, and a recoded or sequence unrelated copy of this target gene which is immune to gRNA targeting, which are situated together on a different autosome (wild type W) than the target gene (FIG. 4A). The transgenic construct (T) is situated on an autosome and consists of Cas9 (long green rectangle), gRNAs (short green rectangle) targeting an essential gene (at a different autosomal locus than T), and a recoded version of the target gene (yellow rectangle with recoded gRNA target sites in orange). The transgenic construct T is generated by targeted insertion at a wild type locus indicated by the blue rectangle (W). Potential cleavage sites on the target essential gene (A) are indicated by dashed lines and scissors, and the cleaved locus (C) is a null form of the target gene made of what remains of the gene from the outer ends of the cleavage sites (FIG. 4A).

[0200] In males and females who carry at least one copy of the construct (T) and at least one copy of the wild type target (A), the target gene is cleaved multiple times during gametogenesis, destroying the wild type copy of the gene (C) and resulting in C bearing gametes (FIG. 4B). As transgenic individuals mate with wild types, cleaved copies of the essential gene will begin to accumulate in heterozygotes (-CA individuals). Only individuals who do not bear a T and receive two cleaved genes (WWCC) will die from the absence of a functional copy of the target essential autosomal gene, leaving the viable genotypes TTCC, TTCA, TTAA, TWCC, TWCA, TWAA, WWCA, and WWAA (FIG. 4B). In individuals which possess at least one T and at least one A, Cas9 can find and cleave a copy of the target gene. The resulting cleaved locus is passed on instead of the original target wildtype locus. When two individuals bearing the cleaved locus mate and the cleaved loci are paired together in the absence of the transgene (WWCC), the

resulting offspring is unviable, removing two wild type alleles are from the population (FIG. **4**B).

[0201] The discrete generation, deterministic population frequency model for this drive mechanism demonstrates that if Cas9 cleaves the target gene with 100% efficiency, this drive mechanism is identical to the single locus cleavage based autosomal drive mechanism. However, if Cas9 cleavage efficiency is imperfect, then this 2-locus cleavage based drive can tolerate larger fitness costs than the single locus version (FIG. **4**C, as compared to FIG. **3**C). Discrete generation, deterministic population frequency modeling of cleavage mediated 2-locus autosomal drive is shown in FIG. **4**C. Each data point uses a few moderate releases of transgenic mosquitoes (three releases of TTCC males at 50% of the population) with the specified fitness cost and Cas9 cleavage efficiency. The color of each data point indicates the number of generations (as indicated by the colorbar) before T bearing individuals make up >99% of the population. White indicates the inability of T to take over under the specified conditions or failure to do so within 70 generations (FIG. **4**C).

[0202] The dynamics of the 2-locus autosomal drive makes it identical to the autosomal drive when the cleavage efficiency of Cas9 is perfect, but when that cleavage efficiency is reduced 2-locus drive becomes the stronger drive. As a result, it can maintain higher fitness costs at reduced cleavage efficiencies while sharing the same applicability to species and ease of creation as with single locus versions.

Example 5

Cleavage Mediated Haplolethal Drive

[0203] Cleavage mediated haplolethal drive is slightly different from the other four cleavage based mechanisms. It consists of Cas9, gRNAs which target an autosomal haplolethal gene (instead of a recessive lethal gene), and a recoded or sequence unrelated copy of this haplolethal target gene which is immune to gRNA targeting, which are situated together at the same locus as the target gene (FIG. **5**A). The transgenic construct (T) is situated on an autosome and consists of Cas9 (long green rectangle), gRNAs (short green rectangle) targeting a haplolethal gene (at the same autosomal locus as T), and a recoded version of the target gene (yellow rectangle with recoded gRNA target sites in orange). Potential cleavage sites on the target haplolethal gene (long yellow rectangle, H) are indicated by dashed lines and scissors, and the cleaved locus (short yellow rectangle, C) is a null form of the target gene made of what remains of the gene from the outer ends of the cleavage sites (FIG. **5**A).

[0204] Cleavage is male specific, so in males who carry the construct (T) and a wild type copy of its target (H), the target gene is cleaved multiple times during spermatogenesis, destroying the wild type copy of the gene (C) and resulting in either T or C bearing sperm (FIG. **5**B). As transgenic males mate, cleaved copies of the haplolethal gene will immediately result in the death of their carrier (both TC and CH genotypes), leaving the viable genotypes TT, TH, and HH (FIG. **5**B). In heterozygotes (TH) Cas9 can find and cleave a copy of the target essential gene. The resulting cleaved locus is passed on instead of the original wildtype locus, and any offspring that receives the cleaved locus is unviable, removing either a transgene and a cleaved locus (TC) or two wild type alleles (CH) from the population (FIG. **5**B). Related constructs can be implemented, as

described above for the two-locus autosomal situation, in which the construct is located at a position different from that of the gene being targeted.

[0205] The discrete generation, deterministic population frequency model for this drive mechanism demonstrates that if Cas9 cleaves the target gene with 100% efficiency, T can drive to fixation with just a few moderate releases of TT males while bearing a fitness cost of up to approximately 60% (FIG. **5**C). T can still drive population replacement when Cas9 is cleaving at non-optimal rates, but it can only tolerate correspondingly reduced fitness costs as a result (FIG. **5**C). Discrete generation, deterministic population frequency modeling of cleavage mediated haplolethal drive is shown in FIG. **5**C. Each data point uses a few moderate releases of transgenic mosquitoes (three releases of TT males at 50% of the population) with the specified fitness cost and Cas9 cleavage efficiency. The color of each data point indicates the number of generations (as indicated by the colorbar) before T bearing individuals make up >99% of the population. White indicates the inability of T to take over under the specified conditions or failure to do so within 70 generations.

[0206] The haplolethal drive is even stronger than the autosomal drive, capable of driving even with ~60% fitness costs at high cleavage efficiency. However, at reduced cleavage efficiency it withstands a smaller range of fitness costs than the 2-locus drive. Additionally, haplolethal drives rely on identifying a haploethal locus on which to base this drive mechanism as well as a pre-meiotic promoter to drive expression of either Cas9 and a post-meiotic promoter for the gRNAs, with one or both promoters also driving male specific expression. The latter two requirements are necessary for getting cleavage of the haplolethal locus in sperm without causing cleavage in the rest of the individual, thereby resulting in death of the construct-bearing individual. In some implementations Cas9 expression is limited to stages of spermatogenesis after those that require activity of the gene being targeted.

Example 6

Maintenance of Extrachromosomal Element

[0207] FIG. **6**A shows a chromosome (circle) carrying essential gene (green rectangle). In this example a prokaryotic chromosome carries a wildtype copy of an essential gene. FIG. **6**B shows an extra-chromosomal element such as a plasmid carrying the construct (red and green rectangles) and any other genes (e.g., one or more cargo sequences) to be maintained in the population. An extrachromosomal element carries the vector, which carries a recoded or sequence unrelated version of the essential gene (diagonal lines) and the DNA modifying enzyme driven by a promoter (solid rectangle). FIG. **6**C shows the construct, which consists of two components: (1) a site-specific DNA modifying enzyme designed to alter the sequence of an endogenous gene required for survival, proliferation, fertility, or differentiation so as to render it non-functional (left); (2) a recoded or sequence unrelated version of the essential gene resistant to cleavage, and having reduced nucleotide identity with the endogenous gene (right). Optionally, one or more cargo sequences are present (center). FIG. **6**D shows the chromosome (FIG. **6**A) and the extra extra-chromosomal element (FIG. **6**B) in a cell and forced inheritance of the extra-chromosomal element. The endogenous copy of essen-

tial gene is altered within the cell by CleaveR to render it non-functional (FIG. **6**E). However, cells that inherit CleaveR survive, proliferate, differentiate, or are fertile, whereas those that fail to inherit C fail are survive, proliferate, differentiate, or are sterile (FIG. **6**F). An expanded view of the vector shown in (FIG. **6**B). Recoded essential gene (or functional equivalent that lacks significant sequence homology) transcribes to the right. DNA sequence modifying enzyme transcribes to the left. A cargo gene is located in between the two in the figure, though the actual arrangement between cargo, rescue and DNA modifying enzyme can take a number of forms. FIG. **6**D shows a cell carrying the wildtype chromosome and the extrachromosomal element including the vector. FIG. **6**E shows DNA modifying activity of the element results in sequence changes to the wildtype copy of essential chromosomal gene (horizontal arrow leading to a chromosome carrying a smaller version of the essential gene). FIG. **6**F shows the extrachromosomal element is spontaneously lost from some cells (left). These cells die because they lack essential gene activity. Those on the right, that carry the vector and associated rescue transgene survive and proliferate.

[0208] Example 7

[0209] FIG. **7** shows a schematic of an embodiment the results of a cross between organisms (in this example insects) heterozygous for the construct and a wild type organism when there is no carryover of DNA cleavage/ alteration activity from germline into embryo. DNA sequence modified (parentheses) version of the essential gene is created in the female germline of heterozygotes. Both copies are cleaved, but the diploid germline cell survives because it carries one copy of the rescue transgene. Female haploid meiotic products (oocytes) survive because the essential product is provided to them from the rescue transgene. These products are inherited by progeny. All individuals inherit chromosomes carrying one sequence modified version of the essential gene. No progeny die. However, crosses between heterozygotes for the nonfunctional version of the essential gene in subsequent generations will create dead homozygotes (not shown). Note that in this example the essential locus is located on the same chromosome as the vector. This is simply for illustrative purposes as it decreases the number of genotypes that need to be shown to capture important aspects of vector behavior. As noted in the figures above, the vector can be located on any chromosome, and act to bring about sequence modifications of any essential gene, on any chromosome or extrachromosomal element. All progeny express one or both versions of the essential gene in the example provided. Therefore, all progeny survive.

Example 8

[0210] FIG. **8**A shows a schematic of an embodiment of the results of a cross between heterozygous organisms when there is no carryover of DNA cleavage/alteration activity from germline into embryo. Cleavage of the essential gene occurs in the parental cell resulting in survival of progeny that express the recoded protein survive and death of offspring that do no inherit CleaveR (FIG. **8**A). The outcome of a cross between heterozygotes is the same whether or not there is maternal carryover. Progeny that inherit the construct survive while those that do not die. FIG. **8**B shows an graph of an embodiment of CleaveR gene drive for different

fitness costs and introduction frequencies without maternal transfer of DNA cleavage/alteration activity.

Example 9

[0211] FIG. **9**A shows a schematic of an embodiment the results of a cross when there is maternal transfer of DNA cleavage/alteration activity from germline into embryo. Cleavage of the essential gene occurs in the parental cell and in products of cell fusion/fertilization into which the DNA cleavage/alteration activity (or the encoding RNA(s)) is introduced during oogenesis, resulting in death of offspring that do no inherit the construct (FIG. **9**A). Only progeny that express the recoded protein survive. FIG. **9**B shows a graph of an embodiment of gene drive for different fitness costs and introduction frequencies with maternal transfer of DNA cleavage/alteration activity.

Example 10

Meiotic Gene Drive

[0212] FIG. **10** shows a schematic of an embodiment of a meiotic gene drive. Cleavage of the essential gene occurs in the parental cell. As a result, gametes that fail to inherit CleaveR do not survive. In such a system chromosomes that carry the vector have a selective advantage and increase in frequency. Such a system can also be used to guarantee that gametes arising from a transgenic individual always carry the transgenes of interest (by virtue of tight genetic linkage to the construct). This ability has applications in agriculture, as it provides a method for regulating gene flow between populations of different genotypes.

Example 11

Sex Ratio Distortion

[0213] FIG. **11** shows a schematic of an embodiment of vector-mediated sex ratio distortion. A gene essential for post-meiotic sperm development is expressed on the Y chromosome as a part of the drive element. Only Y-bearing sperm, generated from diploids in which the drive element/ vector has eliminated a gene required in haploid stages for sperm function, will express the product of this essential gene and be able to complete spermatogenesis/carry out fertilization. This results in sex-ratio distortion if sperm in which the gene has been inactivated fail to develop/undergo fertilization. Such a technology has many uses when the goal is to bring about population reduction or elimination by biasing the sex ratio towards males. A related approach can also be used to bias sex ratios towards males in species in which males are the homogametic sex (ZZ) and females the heterogametic sex (ZW). It can also be used for similar ends in species in which maleness is determined by a dominant allele of a male-determining locus. The primary requirement is that it be possible to eliminate and replace the activity of a gene required in haploid stages of sperm function, and that this product not be able to rescue meiotic brothers to which they may be linked by cytoplasmic bridges until late in spermatogenesis.

Example 12

Comparison of DNA Sequence Modification-Based Gene Drive with Homing-Based Gene Drive—1

[0214] FIG. **12** shows a schematic of an embodiment of a homing endonuclease-based cleavage of target gene for gene

drive. The HEG cuts at a neutral locus in the wildtype chromosome, located at the same position in the genome as the HEG. The presence of the HEG disrupts the HEG cleavage site. In this example, the HEG carries a cargo gene located between the homology arms. In the middle panel, the HEG cleaves the wildtype allele. In the lower panel homologous recombination (HR) is used to repair the DNA break using the HEG-bearing chromosome as a template. Successful HR results in copying of the HEG into the cleaved chromosome. Cleavage of neutral locus by the homing nuclease results in the homing of gene drive and cargo genes into cleaved chromosome. This results in an increase in the population frequency of the HEG and its cargo transgene. However, homing to the neutral locus is required, which may be inefficient. Additionally, the cargo gene needs to be copied, which may not always occur, and development of resistance of neutral locus sequences to cleavage is very common. In contrast, with the DNA sequence modification-based drive method described herein (FIG. **8**A, FIG. **9**A), cleavage of the essential gene results in death of progeny that lack functional copies of the essential gene, i.e., both endogenous copies are cleaved and the recoded copy of the essential gene is not inherited, and results in survival of only those progeny that inherit cargo and recoded copy of the essential gene. Additionally, there is no need for the cargo to be copied as the cargo transmitted with the chromosome. Additionally, homing is not required or utilized, and occurrence of essential genes resistant to cleavage would be rare. Additionally, some species have low rate of germline HDR, greatly if not completely hindering homing based strategies.

Example 13

Comparison of DNA Sequence Modification-Based Gene Drive with Homing-Based Gene Drive—2

[0215] FIG. **13** shows a schematic of an embodiment of a homing endonuclease-based cleavage of target gene for gene drive. The HEG cleaves an essential gene. Homing occurs into the cleaved essential gene, resulting in loss of essential gene function, and an increase in the frequency of the HEG, but only under specific conditions, since loss of both copies of an essential gene results in death or infertility. The recoded essential gene and a cargo are located elsewhere in the genome. As the frequency of the HEG increases, versions of the other chromosome that carry the recoded rescue and cargo are selected for, resulting in their spread. It is important to note that homing is required for this version of population replacement to work. Cleavage alone is not sufficient as it only results in loss of essential gene function, but not an increase in HEG frequency. It is only with homing that the frequency of the HEG increases. Progeny that inherit the chromosome with recoded essential gene and cargo survive but may experience a fitness cost in an otherwise background, which would result in their loss. Only progeny that inherit two inactive copies of the essential gene die. In contrast, with the DNA sequence modification-based drive method described herein (CleaveR; FIG. **8**A, FIG. **9**A), only cleavage is required, and cleavage of the essential gene results in death of progeny that lack functional copies of the essential gene, i.e., both endogenous copies are cleaved and the recoded copy of the essential gene is not inherited, and results in survival of only those progeny that inherit cargo and recoded copy of the essential gene, which are tightly linked. The DNA sequence modification-based drive mecha-

nism described herein does not utilize or depend on homing, only DNA sequence modification and tight linkage to a rescuing transgene.

Example 14

Comparison of DNA Sequence Modification-Based Gene Drive with Medea

[0216] FIG. **14** shows a schematic of an embodiment of a Medea-based gene drive. In Medea-based gene drive a maternally deposited toxin (which may consist of maternally expressed miRNAs that result in a loss of an essential gene, as well as a protein-based toxin, (c.f. Chen et al., 2007)) has the potential to cause the death of all embryos. However, those that inherit a tightly linked antidote survive (which may include a version of the maternally expressed gene being targeted by the maternally expressed miRNAs (c.f. Chen et al., 2007)) because they turn on expression of the antidote just in time to prevent toxin action. In this drive mechanism there is no DNA sequence modification of an endogenous locus. The mechanism of action requires that a maternal (or paternal) toxin be deposited into the embryo. In the Medea-based system, a toxin is expressed in maternal germline resulting in the toxin being present in all oocytes/eggs. Embryos that inherit Medea survive because they express an antidote in the early embryo, while those that do not inherit Medea die. In the Medea-based system, maternal expression of a toxin which can kill embryos but not oocytes is required, and rescue is achieved through early embryo expression of an antidote. In contrast, with the DNA sequence modification-based drive method described herein (CleaveR; FIG. **8**A, FIG. **9**A), cleavage of the essential gene results in death of progeny that lack functional copies of the essential gene, i.e., both endogenous copies are cleaved and the recoded copy of the essential gene is not inherited, and results in survival of only those progeny that inherit cargo and recoded copy of the essential gene. The DNA sequence modification-based drive mechanism described herein only requires DNA sequence modification and does not require maternal or paternal deposition of a toxin. Additionally, germline expression of a DNA modifying enzyme that targets an essential gene occurs, and rescue achieved through inheritance of a recoded version of an essential gene.

Example 15

Cleavage Mediated Drive Targeting an Essential Gene on the X-Chromosome, Proof of Concept in *Drosophila melanogaster*

[0217] Example 15 is an embodiment of a CleaveR drive system showing reduction to practice. FIG. **15**A shows a schematic of an embodiment of a Construct A with a U6:3-gRNA, an attP site, the tko rescue copy from *Drosophila virilis* (Dv) and a ubiquitous opie2-td-tomato marker. Only elements between the homology arms were inserted into the germline via Cas9 mediated HR. FIG. **15**B shows an embodiment of a Construct B with an attB site, a 3xP3-GFP marker, Cas9 driven by nanos regulatory elements, and a set of four U6 driven gRNAs. Construct B was integrated into the attP landing site of construct A via phiC31 integrase. FIG. **15**C shows an embodiment of the principle of ClvR. Females heterozygous for the ClvR construct create cleaved tko alleles in the germline. Additionally, active Cas9/gRNA

23

complex is deposited maternally to all embryos. Offspring without the rescue copy will die.

[0218] The cleavage mediated 2-locus autosomal drive described herein (referred to as CleaveR) consists of Cas9, 4 gRNAs which target an essential gene on the X-chromosome, and a recoded copy of this target gene which is immune to gRNA targeting, which are situated together on a different autosome (chromosome 3) than the target gene (FIG. 15C). FIG. 17 shows an embodiment of an alignment of the target gene (Drosophila melanogaster tko [second line]—Examples 15 and 16) with the recoded rescue based on Drosophila virilis tko.

[0219] In males and females who carry at least one copy of the construct and at least one copy of the wild type target, the target gene is cleaved multiple times during gametogenesis, destroying the wild type copy of the gene and resulting in gametes bearing cleaved tko alleles (FIG. 15C). As transgenic individuals mate with wild types, cleaved copies of the essential gene will begin to accumulate in heterozygotes.

[0220] Additionally, if the CleaveR drive system is inherited through the female germline, all of the offspring will inherit Cas9 and gRNASs. Only offspring that carries the rescue encoded by CleaveR will survive (FIG. 15C).

Target Gene Selection and gRNA Design

[0221] Two versions of the ClvR constructs were constructed using tko (technical knockout) as the target. The tko gene encodes a mitochondrial ribosome protein (Royden, Pirrotta, and Jan 1987). It is a recessive lethal. Benchling software suite was used to design gRNAs targeting the exonic regions of the genes at 4 sites. gRNAs were used based on on-target activity ranking (Doench et al. 2016). In addition gRNAs were selected so as to not cut in the rescue constructs (see below).

Cloning of ClvR Constructs and Fly Germline Transformation

[0222] All plasmids were assembled with standard molecular cloning techniques and Gibson assembly (Gibson et al. 2009). All restriction enzymes, enzymes for Gibson Assembly mastermix and Q5 polymerase used in PCRs were from NEB. Gel extraction kits and JM109 cells for cloning from Zymo Research. The gRNA cassette and Cas9 were based on pCFD3(4)-dU6:3gRNA and pnos-Cas9-nos which were a gift from Simon Bullock (Port et al. 2014) (Addgene #49410 and #62208) and modified as described previously (Oberhofer, Ivy, and Hay 2018). Construct A (FIG. 15A) was inserted into the fly germline via Cas9 mediated homologous recombination. Construct B (FIG. 15B) was integrated into an attP landing site within construct A using the phiC31 system.

[0223] The experiment was started with a plasmid having a dU6:3 promoter and a modified guide scaffold (Dang et al. 2015) separated by BsmBI cutsites from our previous work (Oberhofer, Ivy, and Hay 2018), which was based on pCFD3-dU6:3gRNA, a gift from Simon Bullock (Addgene plasmid #49410) (Port et al. 2014). Restriction digestion was performed with BsmBI and ligated annealed oligos (P0-68E FWD+P0-68E REV) as described on flycrispr.molbio.wisc.edu. This gRNA targets a region on the third chromosome (68E) which was chosen based on the location of an attP

landing site in a widely used fly strain, zh-68E (Bischof et al. 2007). Next, the plasmid was cut with HindIII and SpeI and the following 4 fragments were assembled in a Gibson reaction (Gibson et al. 2009) to yield plasmid p68-tko-step1 (see FIG. 15A):

[0224] Two homology arms, approximately 1 kb in length up and downstream of the above gRNA target site were amplified from genomic DNA with primers P9+P10 and P15+P16; an attP site with primers P11+P12; a 4.2 kb rescue fragment with primers P13+P14. The rescue fragment was based on the tko genomic region of Drosophila virilis, a distant Drosophila species (Drosophila 12 Genomes Consortium et al. 2007). Additionally, 6 silent point mutations were introduced in the ORF of Dv-tko in order to avoid homology stretches >14bp. The rescue was gene synthesized by IDT as two gBlock fragments with an additional 2 point mutations introduced in the intron to work around a synthesis complexity issue. Finally, a td-tomato marker (Shaner et al. 2004) driven by the ubiquitous opie2 promoter (Theilmann and Stewart 1992) with primers P15+P16 was used as the dominant marker.

[0225] Construct p68-tko-step1 (see FIG. 15A) was injected into a fly strain expressing Cas9 in the germline under nanos regulatory regions (Bloomington stock #54591) (Port et al. 2014). All injections were carried out by Rainbow Transgenic Flies.

[0226] Male injected G0 flies were outcrossed to w– and the progeny was scored for ubiquitous td-tomato expression. Male transformants were crossed to a TM3,Sb/TM6b,Tb balancer stock. Flies carrying the marker over TM3,Sb, were pooled and used as the injection strain for the 2nd construct following below.

[0227] For construct tko-step2 (FIG. 15B), two constructs having two gRNAs each were subcloned. Construct pU6:3-U6:1-tandem (Oberhofer, Ivy, and Hay 2018) (based on (Port et al. 2014)) was digested with BsmBI and ligated back in two gRNAs encoded in the primer overhangs: P21+P22 and P23+P24.

[0228] A plasmid that had a 3xP3-GFP marker gene, an attB site as well as parts of nos-Cas9-nos flanked by gypsy insulators was digested with EcoRV and BglII. In a three fragment Gibson reaction full length nos-Cas9-nos, as well as the two gRNA cassettes from above were assembled to yield the final construct ptko-B. Cas9 was amplified with primers P25-nosCas9 FWD+P26-nos-Cas9 REV, guide cassette A with P27-guidesA FWD+P28-guidesA REV, and guide cassette B with P29-guidesB FWD+P30-guidesB REV.

[0229] Construct B was injected along with a phiC31 helper plasmid (Rainbow Transgenic Flies). Injected G0 flies were outcrossed to w– and the progeny was screened for 3xP3-GFP expression. Transgenic males were used to cross to the balancer stock TM3,Sb/TM6b,Tb as well as w[1118]. Flies carrying the GFP marker over TM3,Sb were pooled to generate the balanced stock and flies homozygous for the ClvR construct were collected in the next generation. All primers are shown in TABLE 1, and vector sequences are provided in SEQ ID NO: 39 (p68-tko-step1; FIG. 15A), SEQ ID NO: 40 (tko-step2; FIG. 15B), and SEQ ID NO: 41 (Dvir-rescue-modified; "rescue" in FIG. 15A and FIG. 17).

TABLE 1

	PRIMERS	
Primer	Sequence	SEQ ID NO:
P0-68E FWD	gtcgTGCACAACCAGAGACTGGAG	1
P0-68E REV	aaacCTCCAGTCTCTGGTTGTGCA	2
P9-68E-hr-left FWD	cttattacgtggccaactaggtgcccaaaatgtgtgtgga	3
P10-68E-hr-left REV	GCTTCGGTGTGTCCGTCAGTgagaggttttgccgcgattt	4
P11-attP FWD	aaatcgcggcaaaacctctcACTGACGGACACACCGAAGCC	5
P12-attP REV	ccttgctgcccgcctgcagcAGTCGCGCTCGCGCGACTGA	6
P13-dv-tko FWD	TCAGTCGCGCGAGCGCGACTgctgcaggcgggcagcaagg	7
P14-dv-tko REV	gcagtgcaaaaaagttggtggggtcggacctcaagttgcat atgg	8
P15-68E-hr-right FWD	tgcaacttgaggtccgaccccaccaacttttttgcactgc	9
P16-68E-hr-right REV	gggcgaattgggtacaagctaggatgatgggatgctggaa	10
P21-tko-guidesA FWD	ctattttcaatttaacgtcgctgcagcgatgccattccaGT TTCagagctaTGCTGgaaa	11
P22-tko-guidesA REV	ttcCAGCAtagctctGAAACtcgccaagggcgagtcctgCg aagttcacccggatatct	12
P23-tko-guidesB FWD	ctattttcaatttaacgtcgcaacattgtactgtgccgcgG TTTCagagctaTGCTGgaa	13
P24-tko-guidesB REV	ttcCAGCAtagctctGAAACatcgaaagtgcgtgctggtgC gaagttcacccggatatct	14
P25-nosCas9 FWD	GTTGTCTATACTATAAGATCTATAGGCACGGGAT AACGCT	15
P26-nos-Cas9 REV	GCAATCACAGGTGAGCAAAAAAGCTTGGATTTC ACTGGAACT	16
P27-guidesA FWD	AGTTCCAGTGAAATCCAAGCttttttgctcacctgtgattgc	17
P28-guidesA REV	aatcacaggtgagcaaaaaaaattaaccctcactaaaggga	18
P29-guidesB FWD	cccttttagtgagggttaattttttttgctcacctgtgatt	19
P30-guidesB REV	gcagcctcgagatcgatgattgccgagcacaattgtctag	20
tko-seq1	aagcgttccaagctgcacag	21
tko-seq2	cgcacatccatttccaattg	22
tko-seq3	cacacacacaggtgcgttc	23
tko-seq4	acaactagacgttggcaatcTCACACCTTCCTCTTCTTCTT	24
tko-seq5	tcagcgggattagtgtaagt	25
tko-seq6	catatgcaacttgaggtccg	26
s2-attB-rev	ttcgagaccgtgacctacat	27
s2-u631-seq	AGTTCCAGTGAAATCCAAGC	28
T3-seq REV	gttccctttagtgagggttaatt	29
T3-seq FWD	ATTAACCCTCACTAAAGGGA	30
CAS91F	ATGGACAAGAAGTACTCCATTG	31
CAS91R	GATCGGTATTGCCCAGAACT	32
CAS92F2	AGCGCTAGGCTGTCCAAATC	33

25

TABLE 1 -continued

PRIMERS		
Primer	Sequence	SEQ ID NO:
CAS93F	GAGAAAATCCTCACATTTCGG	34
CAS94F2	AGAGTGGAAAGACAATCCTGG	35
CAS95F	CTGAACGCCAAACTGATCAC	36
CAS96F	TGGACGCCACACTGATTCAT	37
CAS96R	TCACACCTTCCTCTTCTTCTT	38

Example 16

ClvR Effect in Females and Males

[0230] To determine the rate of germline cleavage and carryover effect in females carrying the ClvR element, heterozygous females were crossed to w[1118] males and scored the progeny for the dominant opie2-td-tomato marker of the ClvR construct. Under normal mendelian rules only half of the progeny should carry this marker. Among the 2580 progeny from these crosses all carried the opie2-td-tomato dominant marker, showing that the system works efficiently when transmitted through females (see FIG. **9A** and FIG. **16**A), data in the Punnett square below each cross figure.

[0231] To determine the cleavage rate in the male germline, crosses were set up between males heterozygous for the ClvR element and females carrying a mutant copy of tko over the FM7a X-chromosome balancer (tko³/FM7a/Dp(1; 2;Y)w+, BDSC_4283). Female offspring of this cross will inherit one X-chromosome from the father and one from the mother. Female offspring inheriting the mutant tko allele from the mother and not carrying the ClvR element with the rescue copy of tko will be dead, if tko was cleaved in the male germline (see FIG. **16**B).

[0232] For FIG. **16**A, B all female chromosomes in red, male chromosomes in blue, Cas9/gRNA complex indicated as green scissors. Top row in Panel A and B indicates the cross, lower row shows a punnett square with gametes indicated and numbers of scored progeny in the corresponding fields. Numbers showing the effect of CleaveR are indicated in red (FIG. **16**A) Females heterozygous for the CleaveR system were crossed to wildtype males. The Cas9/gRNA complex encoded by the CleaveR element, cleaves all wildtype copies of tko in the female germline. In addition active complexes get deposited maternally into all embryos, leading to subsequent cleavage of the paternal tko allele in the zygote. Only offspring that inherited the rescue copy from the CleaveR construct were viable, showing that the CleaveR system works efficiently in the female germline and also brings about maternal carryover-dependent cleavage. In FIG. **16**B, males heterozygous for the CleaveR element were crossed to a tko mutant. The only copy of wildtype tko on the single male X-chromosome was cleaved in the male germline by the CleaveR system. When the cleaved tko allele was paired with the maternal mutant X-chromosome (tko3), only those animals that also inherit the rescue encoded by the CleaveR element survived, all others died. Actual data is shown in the Punnet squares below each cross.

Results showed successful implementation of the DNA sequence modification-based gene drive according to the embodiments disclosed herein.

Example 17

ClvR Effect in Females and Males

[0233] FIG. **18**A show a schematic of an embodiment of the components of the DNA sequence modification-based gene drive implemented in the example below, targeting the X-linked locus tko in *Drosophila*, using a third chromosome-based gene drive element. FIG. **18**B (SEQ ID NO: 42) shows a schematic of an embodiment of the components of the step 1 transgenic created for the DNA sequence modification-based gene drive implemented for targeting the X-linked locus tko in *Drosophila*, using a third chromosome-based gene drive element. This construct was inserted into the *Drosophila* genome using homologous recombination, based on the left and right homology arms. FIG. **18**C (SEQ ID NO: 43) shows a schematic of an embodiment of the components of the step 2 construct created for the DNA sequence modification-based gene drive implemented for targeting the X-linked locus tko in *Drosophila*, using a third chromosome-based gene drive element. This construct was inserted into the step 1 genomic region using the attb site-specific integrase target site. FIG. **19** shows an embodiment Sanger sequencing results of the gRNA1 target region of the *Drosophila* wildtype version of the tko gene and offspring: ♂ClvRt^{tko}/+ offspring from ♀ClvR^{tko}/+XX ♂w[1118] parents. The wildtype sequence is shown as well as products of ClvR action, which contain indels. Two flies were sequenced from 9 different single fly crosses each. All 18 analyzed flies showed indels of varying sizes at the gRNA1 target site. Results showed successful implementation of the DNA sequence modification-based gene drive according to the embodiments disclosed herein.

REFERENCES

[0234] WHO World Malaria Report 2014. WHO at <who.int/malaria/publications/world_malaria_report_2014/en/>

[0235] Alphey, L. Genetic Control of Mosquitoes. Annu. Rev. Entomol. 59, 205-224 (2014).

[0236] Resnik, D. B. Ethical Issues in Field Trials of Genetically Modified Disease-Resistant Mosquitoes. Dev. World Bioeth. 14, 37-46 (2014).

[0237] Malavasi, A. Project *Aedes* transgenic population control in Juazeiro and Jacobina Bahia, Brazil. BMC Proc. 8, O11 (2014).

[0238] Popovici, J. et al. Assessing key safety concerns of a *Wolbachia*-based strategy to control dengue transmission by *Aedes* mosquitoes. Mem. Inst. Oswaldo Cruz 105, 957-964 (2010).

[0239] Walker, T. et al. The wMel *Wolbachia* strain blocks dengue and invades caged *Aedes aegypti* populations. Nature 476, 450-453 (2011).

[0240] Hoffmann, A. A. et al. Successful establishment of *Wolbachia* in *Aedes* populations to suppress dengue transmission. Nature 476, 454-457 (2011).

[0241] Sebrovskii, A. S. A New Possible Method of Pest Control. Zool Zh 19, 618-630 (1940).

[0242] Curtis, C. F. Possible Use of Translocations to fix Desirable Genes in Insect Pest Populations. Nature 218, 368-369 (1968).

[0243] Gould, F. & Schliekelman, P. POPULATION GENETICS OF AUTOCIDAL CONTROL AND STRAIN REPLACEMENT. Annu. Rev. Entomol. 49, 193-217 (2004).

[0244] Sinkins, S. P. & Gould, F. Gene drive systems for insect disease vectors. Nat. Rev. Genet. 7, 427-435 (2006).

[0245] Chen, C.-H. et al. A Synthetic Maternal-Effect Selfish Genetic Element Drives Population Replacement in *Drosophila*. Science 316, 597-600 (2007).

[0246] Gould, F., Huang, Y., Legros, M. & Lloyd, A. L. A Killer-Rescue system for self-limiting gene drive of anti-pathogen constructs. Proc. R. Soc. B Biol. Sci. 275, 2823-2829 (2008).

[0247] Marshall, J. M. & Hay, B. A. GENERAL PRINCIPLES OF SINGLE-CONSTRUCT CHROMOSOMAL GENE DRIVE: SINGLE-CONSTRUCT GENE DRIVE. Evolution 66, 2150-2166 (2012).

[0248] Marshall, J. M. The Impact of Dissociation on Transposon-Mediated Disease Control Strategies. Genetics 178, 1673-1682 (2008).

[0249] Davis, S., Bax, N. & Grewe, P. Engineered Underdominance Allows Efficient and Economical Introgression of Traits into Pest Populations. j. Theor. Biol. 212, 83-98 (2001).

[0250] Magori, K. Genetically Engineered Underdominance for Manipulation of Pest Populations: A Deterministic Model. Genetics 172, 2613-2620 (2005).

[0251] Brelsfoard, C. L. & Dobson, S. L. *Wolbachia*-based strategies to control insect pests and disease vectors. Asia Pac j Mol Biol Biotechnol 17, 55-63 (2009).

[0252] Huang, Y., Magori, K., Lloyd, A. L. & Gould, F. INTRODUCING DESIRABLE TRANSGENES INTO INSECT POPULATIONS USING Y-LINKED MEIOTIC DRIVE?A THEORETICAL ASSESSMENT. Evolution 61, 717-726 (2007).

[0253] Lyttle, T. W. Experimental population genetics of meiotic drive systems I. Pseudo-Y chromosomal drive as a means of eliminating cage populations of *Drosophila melanogaster*. Genetics 86, 413-445 (1977).

[0254] Magnusson, K. et al. Demasculinization of the *Anopheles gambiae* X chromosome. BMC Evol. Biol. 12, 69 (2012).

[0255] Simoni, A. et al. Development of synthetic selfish elements based on modular nucleases in *Drosophila melanogaster*. Nucleic Acids Res. 42, 7461-7472 (2014).

[0256] Burt, A. & Koufopanou, V. Homing endonuclease genes: the rise and fall and rise again of a selfish element. Curr. Opin. Genet. Dev. 14, 609-615 (2004).

[0257] Gimble, F. S. Invasion of a multitude of genetic niches by mobile endonuclease genes. FEMS Microbiol. Lett. 185, 99-107 (2000).

[0258] Galizi, R. et al. A synthetic sex ratio distortion system for the control of the human malaria mosquito. Nat. Commun. 5, (2014).

[0259] Akbari, O. S. et al. A Synthetic Gene Drive System for Local, Reversible Modification and Suppression of Insect Populations. Curr. Biol. 23, 671-677 (2013).

[0260] Akbari, O. S. et al. Novel Synthetic Medea Selfish Genetic Elements Drive Population Replacement in *Drosophila;* a Theoretical Exploration of Medea-Dependent Population Suppression. ACS Synth. Biol. 3, 915-928 (2014).

[0261] Esvelt, K. M., Smidler, A. L., Catteruccia, F. & Church, G. M. Concerning RNA-guided gene drives for the alteration of wild populations. eLife e03401 (2014). doi:10. 7554/eLife.03401

[0262] Isaacs, A. T. et al. Engineered Resistance to *Plasmodium falciparum* Development in Transgenic *Anopheles stephensi*. PLOS Pathog. 7, e1002017 (2011).

[0263] Hollingdale, M. R., Nardin, E. H., Tharavanij, S., Schwartz, A. L. & Nussenzweig, R. S. Inhibition of entry of *Plasmodium falciparum* and *P. vivax* sporozoites into cultured cells; an in vitro assay of protective antibodies. J. Immunol. 132, 909-913 (1984).

[0264] Li, F., Patra, K. P. & Vinetz, J. M. An Anti-Chitinase Malaria Transmission-Blocking Single-Chain Antibody as an Effector Molecule for Creating a *Plasmodium falciparum*-Refractory Mosquito. J. Infect. Dis. 192, 878-887 (2005).

[0265] Yen, P.-S., James, A., Li, J.-C., Chen, C.-H. & Failloux, A.-B. Synthetic miRNAs induce dual arboviral-resistance phenotypes in the vector mosquito *Aedes aegypti*. Commun. Biol. 1, 11 (2018).

[0266] Franz, A. W. E. et al. Engineering RNA interference-based resistance to dengue virus type 2 in genetically modified *Aedes aegypti*. Proc. Natl. Acad. Sci. 103, 4198-4203 (2006).

[0267] Mathur, G. et al. Transgene-mediated suppression of dengue viruses in the salivary glands of the yellow fever mosquito, *Aedes aegypti*. Insect Mol. Biol. 19, 753-763 (2010).

[0268] Travanty, E. A. et al. Using RNA interference to develop dengue virus resistance in genetically modified *Aedes aegypti*. Insect Biochem. Mol. Biol. 34, 607-613 (2004).

[0269] Castillo, J. A. et al. Complex interaction between dengue virus replication and expression of miRNA-133a. BMC Infect. Dis. 16, (2016).

[0270] Shmakov, S. et al. Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. Mol. Cell 60, 385-397 (2015).

[0271] Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. Curr. Opin. Microbiol. 37, 67-78 (2017).

[0272] Shmakov, S. et al. Diversity and evolution of class 2 CRISPR-Cas systems. Nat. Rev. Microbiol. 15, 169-182 (2017).

[0273] Koonin, E. V., Makarova, K. S. & Wolf, Y. I. Evolutionary Genomics of Defense Systems in Archaea and Bacteria. Annu. Rev. Microbiol. 71, 233-261 (2017).

[0274] Bischof, Johannes, Robert K. Maeda, Monika Hediger, Francois Karch, and Konrad Basler. 2007. "An Optimized Transgenesis System for *Drosophila* Using Germ-Line-Specific phiC31 Integrases." Proceedings of the National Academy of Sciences of the United States of America 104 (9): 3312-17.

[0275] Dang, Ying, Gengxiang Jia, Jennie Choi, Hongming Ma, Edgar Anaya, Chunting Ye, Premlata Shankar, and Haoquan Wu. 2015. "Optimizing sgRNA Structure to Improve CRISPR-Cas9 Knockout Efficiency." Genome Biology 16 (December): 280.

[0276] Doench, John G., Nicolo Fusi, Meagan Sullender, Mudra Hegde, Emma W. Vaimberg, Katherine F. Donovan, Ian Smith, et al. 2016. "Optimized sgRNA Design to Maximize Activity and Minimize off-Target Effects of CRISPR-Cas9." Nature Biotechnology 34 (2): 184-91.

[0277] *Drosophila* 12 Genomes Consortium, Andrew G. Clark, Michael B. Eisen, Douglas R. Smith, Casey M. Bergman, Brian Oliver, Therese A. Markow, et al. 2007. "Evolution of Genes and Genomes on the *Drosophila* Phylogeny." Nature 450 (7167): 203-18.

[0278] Gibson, Daniel G., Lei Young, Ray-Yuan Chuang, J. Craig Venter, Clyde A. Hutchison, and Hamilton O. Smith. 2009. "Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases." Nature Methods 6 (5): 343-45.

[0279] Oberhofer, Georg, Tobin Ivy, and Bruce A. Hay. 2018. "Behavior of Homing Endonuclease Gene Drives Targeting Genes Required for Viability or Female Fertility with Multiplexed Guide RNAs." bioRxiv. https://doi.org/10. 1101/289546.

[0280] Port, Fillip, Hui-Min Chen, Tzumin Lee, and Simon L. Bullock. 2014. "Optimized CRISPR/Cas Tools for Efficient Germline and Somatic Genome Engineering in *Drosophila.*" Proceedings of the National Academy of Sciences of the United States of America 111 (29): E2967-76.

[0281] Royden, C. S., V. Pirrotta, and L. Y. Jan. 1987. "The Tko Locus, Site of a Behavioral Mutation in *D. Melanogaster,* Codes for a Protein Homologous to Prokaryotic Ribosomal Protein S12." Cell 51 (2): 165-73.

[0282] Shaner, Nathan C., Robert E. Campbell, Paul A. Steinbach, Ben N. G. Giepmans, Amy E. Palmer, and Roger Y. Tsien. 2004. "Improved Monomeric Red, Orange and Yellow Fluorescent Proteins Derived from *Discosoma* Sp. Red Fluorescent Protein." Nature Biotechnology 22 (12): 1567-72.

[0283] Theilmann, D. A., and S. Stewart. 1992. "Molecular Analysis of the Trans-Activating IE-2 Gene of *Orgyia Pseudotsugata* Multicapsid Nuclear Polyhedrosis Virus." Virology 187 (1): 84-96.

[0284] Akbari, Omar S., Chun-Hong Chen, John M. Marshall, Haixia Huang, Igor Antoshechkin, and Bruce A. Hay. 2012. "Novel Synthetic Medea Selfish Genetic Elements Drive Population Replacement in *Drosophila;* a Theoretical Exploration of Medea-Dependent Population Suppression." *ACS Synthetic Biology,* December. https://doi.org/10.1021/ sb300079h.

[0285] Akbari, Omar S., Kelly D. Matzen, John M. Marshall, Haixia Huang, Catherine M. Ward, and Bruce A. Hay. 2013. "A Synthetic Gene Drive System for Local, Reversible Modification and Suppression of Insect Populations." Current Biology: CB 23 (8): 671-77.

[0286] Altrock, P. M., A. Traulsen, and F. A. Reed. 2011. "Stability Properties of Underdominance in Finite Subdivided Populations." *PLoS Computational Biology* 7 (11): e1002260.

[0287] Altrock, P. M., A. Traulsen, R. G. Reeves, and F. A. Reed. 2010. "Using Underdominance to Bi-Stably Transform Local Populations." *Journal of Theoretical Biology* 267 (1): 62-75.

[0288] Beaghton, A., P. J. Beaghton, and A. Burt. 2016. "Gene Drive through a Landscape: Reaction-Diffusion Models of Population Suppression and Elimination by a Sex Ratio Distorter." Theoretical Population Biology 108 (April): 51-69.

[0289] Beaghton, A., A. Hammond, T. Nolan, A. Crisanti, H. C. Godfray, and A. Burt. 2017. "Requirements for Driving Antipathogen Effector Genes into Populations of Disease Vectors by Homing." Genetics 205 (4): 1587-96.

[0290] Ben-David, E., A. Burga, and L. Kruglyak. 2017. "A Maternal-Effect Selfish Genetic Element in *Caenorhabditis Elegans.*" Science 356 (6342): 1051-55.

[0291] Braig, H. R., and G. Yan. 2001. "The Spread of Genetic Constructs in Natural Insect Populations." In *Genetically Engineered Organisms: Assessing Environmental and Human Health Effects,* edited by D. K. Letourneau and B. E. Burrows, 251-314. CRC Press.

[0292] Buchman, A. B., T. Ivy, J. M. Marshall, O. S. Akbari, and B. A. Hay. 2018. "Engineered Reciprocal Chromosome Translocations Drive High Threshold, Reversible Population Replacement in *Drosophila.*" *ACS Synthetic Biology.* https://doi.org/10.1021/acssynbio.7b00451.

[0293] Burt, A. 2003. "Site-Specific Selfish Genes as Tools for the Control and Genetic Engineering of Natural Populations." *Proceedings. Biological Sciences/The Royal Society* 270 (1518): 921-28.

[0294] Burt, A., and R. Trivers. 1998. "Genetic Conflicts in Genomic Imprinting." *Proceedings. Biological Sciences/ The Royal Society* 265 (1413): 2393-97.

[0295] Champer, J., R. Reeves, S. Y. Oh, C. Liu, J. Liu, A. G. Clark, and P. W. Messer. 2017. "Novel CRISPR/Cas9 Gene Drive Constructs Reveal Insights into Mechanisms of Resistance Allele Formation and Drive Efficiency in Genetically Diverse Populations." *PLoS Genetics* 13 (7): e1006796.

[0296] Chan, Yuk-Sang, David S. Huen, Ruth Glauert, Eleanor Whiteway, and Steven Russell. 2013. "Optimising Homing Endonuclease Gene Drive Performance in a Semi-Refractory Species: The *Drosophila Melanogaster* Experience." *PLoS One* 8 (1): e54130.Chan, Yuk-Sang, Daniel A. Naujoks, David S. Huen, and Steven Russell. 2011. "Insect Population Control by Homing Endonuclease-Based Gene Drive: An Evaluation in *Drosophila Melanogaster.*" Genetics 188 (1): 33-44.

[0297] Chen, C. H., H. Huang, C. M. Ward, J. T. Su, L. V. Schaeffer, M. Guo, and B. A. Hay. 2007. "A Synthetic Maternal-Effect Selfish Genetic Element Drives Population Replacement in *Drosophila.*" Science 316 (5824): 597-600.

[0298] Davis, Stephen, Nicholas Bax, and Peter Grewe. 2001. "Engineered Underdominance Allows Efficient and Economical Introgression of Traits into Pest Populations." *Journal of Theoretical Biology* 212 (1): 83-98.

[0299] Galizi, R., L. A. Doyle, M. Menichelli, F. Bernardini, A. Deredec, A. Burt, B. L. Stoddard, N. Windbichler, and A. Crisanti. 2014. "A Synthetic Sex Ratio Distortion

System for the Control of the Human Malaria Mosquito." *Nature Communications* 5: 3977.

[0300] Galizi, R., A. Hammond, K. Kyrou, C. Taxiarchi, F. Bernardini, S. M. O'Loughlin, P. A. Papathanos, T. Nolan, N. Windbichler, and A. Crisanti. 2016. "A CRISPR-Cas9 Sex-Ratio Distortion System for Genetic Control." *Scientific Reports* 6: 31139.Gantz, V. M., and E. Bier. 2015. "The Mutagenic Chain Reaction: A Method for Converting Heterozygous to Homozygous Mutations." *Science* 348 (6233): 442-44.

[0301] Gantz, V. M., N. Jasinskiene, O. Tatarenkova, A. Fazekas, V. M. Macias, E. Bier, and A. A. James. 2015. "Highly Efficient Cas9-Mediated Gene Drive for Population Modification of the Malaria Vector Mosquito *Anopheles Stephensi."* Proceedings of the National Academy of Sciences of the United States of America* 112 (49): E6736-43.

[0302] Godfray, H. C. J., A. North, and A. Burt. 2017. "How Driving Endonuclease Genes Can Be Used to Combat Pests and Disease Vectors." *BMC Biology* 15 (1): 81.

[0303] Gokhale, Chaitanya S., Richard Guy Reeves, and Floyd A. Reed. 2014. "Dynamics of a Combined Medea-Underdominant Population Transformation System." *BMC Evolutionary Biology* 14: 98.

[0304] Gould, Fred, and Paul Schliekelman. 2004. "Population Genetics of Autocidal Control and Strain Replacement." *Annual Review of Entomology* 49: 193-217.

[0305] Hammond, A., R. Galizi, K. Kyrou, A. Simoni, C. Siniscalchi, D. Katsanos, M. Gribble, et al. 2016. "A CRISPR-Cas9 Gene Drive System Targeting Female Reproduction in the Malaria Mosquito Vector *Anopheles Gambiae."* Nature Biotechnology* 34 (1): 78-83.

[0306] Hay, Bruce A., Chun-Hong Chen, Catherine M. Ward, Haixia Huang, Jessica T. Su, and Ming Guo. 2010. "Engineering the Genomes of Wild Insect Populations: Challenges, and Opportunities Provided by Synthetic Medea Selfish Genetic Elements." *Journal of Insect Physiology* 56 (10): 1402-13.

[0307] Hu, W., Z. D. Jiang, F. Suo, J. X. Zheng, W. Z. He, and L. L. Du. 2017. "A Large Gene Family in Fission Yeast Encodes Spore Killers That Subvert Mendel's Law." *eLife* 6. https://doi.org/10.7554/eLife.26057.

[0308] Marshall, J. M. 2009. "The Effect of Gene Drive on Containment of Transgenic Mosquitoes." *Journal of Theoretical Biology* 258 (2): 250-65.

[0309] Marshall, J. M., and B. A. Hay. 2011. "Inverse Medea as a Novel Gene Drive System for Local Population Replacement: A Theoretical Analysis." *The Journal of Heredity* 102 (3): 336-41.

[0310] _____. 2012. "Confinement of Gene Drive Systems to Local Populations: A Comparative Analysis." *Journal of Theoretical Biology* 294: 153-71.

[0311] Marshall, John M., and Bruce A. Hay. 2012a. "Confinement of Gene Drive Systems to Local Popu . . . [J Theor Biol. 2012]—PubMed—NCBI." *Journal of Theoretical Biology* 294 (February): 153-71.

[0312] _____. 2012b. "General Principles of Single-Construct Chromosomal . . . [Evolution. 2012]—PubMed—NCBI." *Evolution; International Journal of Organic Evolution* 66 (7):2150-66.

[0313] Marshall, John M., Geoffrey W. Pittman, Anna B. Buchman, and Bruce A. Hay. 2011. "Semele: A Killer-Male, Rescue-Female System for Suppression and Replacement of Insect Disease Vector Populations." *Genetics* 187 (2): 535-51.

[0314] Nuckolls, N. L., M. A. Bravo Nunez, M. T. Eickbush, J. M. Young, J. J. Lange, J. S. Yu, G. R. Smith, S. L. Jaspersen, H. S. Malik, and S. E. Zanders. 2017. "Wtf Genes Are Prolific Dual Poison-Antidote Meiotic Drivers." *eLife* 6. https://doi.org/10.7554/eLife.26033.

[0315] Preston, Christine R., Carlos C. Flores, and William R. Engels. 2006. "Differential Usage of Alternative Pathways of Double-Strand Break Repair in *Drosophila."* Genetics* 172 (2): 1055-68.

[0316] Reeves, R. G., J. Bryk, P. M. Altrock, J. A. Denton, and F. A. Reed. 2014. "First Steps towards Underdominant Genetic Transformation of Insect Populations." *PLoS One* 9 (5): e97557.

[0317] Seidel, H. S., M. Ailion, J. Li, A. van Oudenaarden, M. V. Rockman, and L. Kruglyak. 2011. "A Novel Sperm-Delivered Toxin Causes Late-Stage Embryo Lethality and Transmission Ratio Distortion in *C. Elegans."* PLoS Biology* 9 (7): e1001115.

[0318] Simoni, A., C. Siniscalchi, Y. S. Chan, D. S. Huen, S. Russell, N. Windbichler, and A. Crisanti. 2014. "Development of Synthetic Selfish Elements Based on Modular Nucleases in *Drosophila Melanogaster."* Nucleic Acids Research* 42 (11): 7461-72.

[0319] Wade, M. J., and R. W. Beeman. 1994. "The Population Dynamics of Maternal-Effect Selfish Genes." *Genetics* 138 (4): 1309-14.

[0320] Ward, Catherine M., Jessica T. Su, Yunxin Huang, Alun L. Lloyd, Fred Gould, and Bruce A. Hay. 2011. "Medea Selfish Genetic Elements as Tools for Altering Traits of Wild Populations: A Theoretical Analysis." *Evolution; International Journal of Organic Evolution* 65 (4): 1149-62.

[0321] Windbichler, Nikolai, Philippos Aris Papathanos, Flaminia Catteruccia, Hilary Ranson, Austin Burt, and Andrea Crisanti. 2007. "Homing Endonuclease Mediated Gene Targeting in *Anopheles Gambiae* Cells and Embryos." *Nucleic Acids Research* 35 (17): 5922-33.

[0322] Windbichler, N., M. Menichelli, P. A. Papathanos, S. B. Thyme, H. Li, U. Y. Ulge, B. T. Hovde, et al. 2011. "A Synthetic Homing Endonuclease-Based Gene Drive System in the Human Malaria Mosquito." *Nature* 473 (7346): 212-15.

[0323] Windbichler, N., P. A. Papathanos, and A. Crisanti. 2008. "Targeting the X Chromosome during Spermatogenesis Induces Y Chromosome Transmission Ratio Distortion and Early Dominant Embryo Lethality in *Anopheles Gambiae."* PLoS Genetics* 4 (12): e1000291. Sun, N., and H. Zhao. 2014. "A Single-Chain TALEN Architecture for Genome Engineering." *Molecular bioSystems* 10 (3): 446-53.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 43

<210> SEQ ID NO 1

-continued

```
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P0-68E FWD

<400> SEQUENCE: 1

gtcgtgcaca accagagact ggag                                         24


<210> SEQ ID NO 2
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P0-68E REV

<400> SEQUENCE: 2

aaacctccag tctctggttg tgca                                         24


<210> SEQ ID NO 3
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P9-68E-hr-left FWD

<400> SEQUENCE: 3

cttattacgt ggccaactag gtgcccaaaa tgtgtgtgga                        40


<210> SEQ ID NO 4
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P10-68E-hr-left REV

<400> SEQUENCE: 4

gcttcggtgt gtccgtcagt gagaggtttt gccgcgattt                        40


<210> SEQ ID NO 5
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P11-attP FWD

<400> SEQUENCE: 5

aaatcgcggc aaaacctctc actgacggac acaccgaagc c                      41


<210> SEQ ID NO 6
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P12-attP REV

<400> SEQUENCE: 6

ccttgctgcc cgcctgcagc agtcgcgctc gcgcgactga                        40


<210> SEQ ID NO 7
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
```

-continued

<400> SEQUENCE: 7

tcagtcgcgc gagcgcgact gctgcaggcg ggcagcaagg   40

<210> SEQ ID NO 8
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P14-dv-tko REV

<400> SEQUENCE: 8

gcagtgcaaa aaagttggtg gggtcggacc tcaagttgca tatgg   45

<210> SEQ ID NO 9
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P15-68E-hr-right FWD

<400> SEQUENCE: 9

tgcaacttga ggtccgaccc caccaacttt tttgcactgc   40

<210> SEQ ID NO 10
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P16-68E-hr-right REV

<400> SEQUENCE: 10

gggcgaattg ggtacaagct aggatgatgg gatgctggaa   40

<210> SEQ ID NO 11
<211> LENGTH: 60
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P21-tko-guidesA FWD

<400> SEQUENCE: 11

ctattttcaa tttaacgtcg ctgcagcgat gccattccag tttcagagct atgctggaaa   60

<210> SEQ ID NO 12
<211> LENGTH: 60
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P22-tko-guidesA REV

<400> SEQUENCE: 12

ttccagcata gctctgaaac tcgccaaggg cgttgtcctg cgaagttcac ccggatatct   60

<210> SEQ ID NO 13
<211> LENGTH: 60
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P23-tko-guidesB FWD

<400> SEQUENCE: 13

-continued

```
ctattttcaa tttaacgtcg caacattgta ctgtgccgcg gtttcagagc tatgctggaa          60


<210> SEQ ID NO 14
<211> LENGTH: 60
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P24-tko-guidesB REV

<400> SEQUENCE: 14

ttccagcata gctctgaaac atcgaaagtg cgtgctggtg cgaagttcac ccggatatct          60


<210> SEQ ID NO 15
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P25-nosCas9 FWD

<400> SEQUENCE: 15

gttgtctata ctataagatc tataggcacg ggataacgct                                40


<210> SEQ ID NO 16
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P26-nos-Cas9 REV

<400> SEQUENCE: 16

gcaatcacag gtgagcaaaa aagcttggat ttcactggaa ct                            42


<210> SEQ ID NO 17
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P27-guidesA FWD

<400> SEQUENCE: 17

agttccagtg aaatccaagc ttttttgctc acctgtgatt gc                            42


<210> SEQ ID NO 18
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P28-guidesA REV

<400> SEQUENCE: 18

aatcacaggt gagcaaaaaa aattaaccct cactaaaggg a                             41


<210> SEQ ID NO 19
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P29-guidesB FWD

<400> SEQUENCE: 19

ccctttagtg agggttaatt ttttttgctc acctgtgatt                               40


<210> SEQ ID NO 20
```

-continued

```
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P30-guidesB REV

<400> SEQUENCE: 20

gcagcctcga gatcgatgat tgccgagcac aattgtctag                            40


<210> SEQ ID NO 21
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: tko-seq1

<400> SEQUENCE: 21

aagcgttcca agctgcacag                                                 20


<210> SEQ ID NO 22
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: tko-seq2

<400> SEQUENCE: 22

cgcacatcca tttccaattg                                                 20


<210> SEQ ID NO 23
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: tko-seq3

<400> SEQUENCE: 23

cacacacaca ggtgcgttc                                                  19


<210> SEQ ID NO 24
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: tko-seq4

<400> SEQUENCE: 24

acaactagac gttggcaatc tcacaccttc ctcttcttct t                         41


<210> SEQ ID NO 25
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: tko-seq5

<400> SEQUENCE: 25

tcagcgggat tagtgtaagt                                                 20


<210> SEQ ID NO 26
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
```

-continued

```
<223> OTHER INFORMATION: tko-seq6

<400> SEQUENCE: 26

catatgcaac ttgaggtccg                                               20


<210> SEQ ID NO 27
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: s2-attB-rev

<400> SEQUENCE: 27

ttcgagaccg tgacctacat                                               20


<210> SEQ ID NO 28
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: s2-u631-seq

<400> SEQUENCE: 28

agttccagtg aaatccaagc                                               20


<210> SEQ ID NO 29
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: T3-seq REV

<400> SEQUENCE: 29

gttcccttta gtgagggtta att                                          23


<210> SEQ ID NO 30
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: T3-seq FWD

<400> SEQUENCE: 30

attaaccctc actaaaggga                                               20


<210> SEQ ID NO 31
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: CAS91F

<400> SEQUENCE: 31

atggacaaga agtactccat tg                                            22


<210> SEQ ID NO 32
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: CAS91R

<400> SEQUENCE: 32
```

-continued

```
gatcggtatt gcccagaact                                        20


<210> SEQ ID NO 33
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: CAS92F2

<400> SEQUENCE: 33

agcgctaggc tgtccaaatc                                        20


<210> SEQ ID NO 34
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: CAS93F

<400> SEQUENCE: 34

gagaaaatcc tcacatttcg g                                      21


<210> SEQ ID NO 35
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: CAS94F2

<400> SEQUENCE: 35

agagtggaaa gacaatcctg g                                      21


<210> SEQ ID NO 36
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: CAS95F

<400> SEQUENCE: 36

ctgaacgcca aactgatcac                                        20


<210> SEQ ID NO 37
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: CAS96F

<400> SEQUENCE: 37

tggacgccac actgattcat                                        20


<210> SEQ ID NO 38
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: CAS96R

<400> SEQUENCE: 38

tcacaccttc ctcttcttct t                                      21


<210> SEQ ID NO 39
```

-continued

```
<211> LENGTH: 12817
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: p68-tko-step1

<400> SEQUENCE: 39

agcttgtacc caattcgccc tatagtgagt cgtattacaa ttcactggcc gtcgttttac      60

aacgtcgtga ctgggaaaac cctggcgtta cccaacttaa tcgccttgca gcacatcccc     120

ctttcgccag ctggcgtaat agcgaagagg cccgcaccga tcgcccttcc caacagttgc     180

gcagcctgaa tggcgaatgg gacgcgccct gtagcggcgc attaagcgcg gcgggtgtgg     240

tggttacgcg cagcgtgacc gctacacttg ccagcgccct agcgcccgct cctttcgctt     300

tcttcccttc ctttctcgcc acgttcgccg gctttccccg tcaagctcta atcggggggc     360

tccctttagg gttccgattt agtgctttac ggcacctcga ccccaaaaaa cttgattagg     420

gtgatggttc acgtagtggg ccatcgccct gatagacggt ttttcgccct ttgacgttgg     480

agtccacgtt ctttaatagt ggactcttgt tccaaactgg aacaacactc aaccctatct     540

cggtctattc ttttgattta taggggattt tgccgatttc ggcctattgg ttaaaaaatg     600

agctgattta acaaaaattt aacgcgaatt ttaacaaaat attaacgctt acaatttagg     660

tggcacttt cggggaaatg tgcgcggaac ccctatttgt ttatttttct aaatacattc     720

aaatatgtat ccgctcatga gacaataacc ctgataaatg cttcaataat attgaaaaag     780

gaagagtatg agtattcaac atttccgtgt cgcccttatt ccctttttg cggcattttg     840

ccttcctgtt tttgctcacc cagaaacgct ggtgaaagta aaagatgctg aagatcagtt     900

gggtgcacga gtgggttaca tcgaactgga tctcaacagc ggtaagatcc ttgagagttt     960

tcgccccgaa gaacgttttc caatgatgag cacttttaaa gttctgctat gtggcgcggt    1020

attatcccgt attgacgccg ggcaagagca actcggtcgc cgcatacact attctcagaa    1080

tgacttggtt gagtactcac cagtcacaga aaagcatctt acggatggca tgacagtaag    1140

agaattatgc agtgctgcca taaccatgag tgataacact gcggccaact tacttctgac    1200

aacgatcgga ggaccgaagg agctaaccgc ttttttgcac aacatggggg atcatgtaac    1260

tcgccttgat cgttgggaac cggagctgaa tgaagccata ccaaacgacg agcgtgacac    1320

cacgatgcct gtagcaatgg caacaacgtt gcgcaaacta ttaactggcg aactacttac    1380

tctagcttcc cggcaacaat taatagactg gatggaggcg gataaagttg caggaccact    1440

tctgcgctcg gcccttccgg ctggctggtt tattgctgat aaatctggag ccggtgagcg    1500

tgggtctcgc ggtatcattg cagcactggg gccagatggt aagccctccc gtatcgtagt    1560

tatctacacg acggggagtc aggcaactat ggatgaacga aatagacaga tcgctgagat    1620

aggtgcctca ctgattaagc attggtaact gtcagaccaa gtttactcat atatacttta    1680

gattgattta aaacttcatt tttaatttaa aaggatctag gtgaagatcc ttttttgataa    1740

tctcatgacc aaaatccctt aacgtgagtt ttcgttccac tgagcgtcag accccgtaga    1800

aaagatcaaa ggatcttctt gagatccttt ttttctgcgc gtaatctgct gcttgcaaac    1860

aaaaaaacca ccgctaccag cggtggtttg tttgccggat caagagctac caactctttt    1920

tccgaaggta actggcttca gcagagcgca gataccaaat actgtccttc tagtgtagcc    1980

gtagttaggc caccacttca agaactctgt agcaccgcct acatacctcg ctctgctaat    2040
```

-continued

```
cctgttacca gtggctgctg ccagtggcga taagtcgtgt cttaccgggt tggactcaag   2100

acgatagtta ccggataagg cgcagcggtc gggctgaacg gggggttcgt gcacacagcc   2160

cagcttggag cgaacgacct acaccgaact gagataccta cagcgtgagc tatgagaaag   2220

cgccacgctt cccgaaggga gaaaggcgga caggtatccg gtaagcggca gggtcggaac   2280

aggagagcgc acgagggagc ttccaggggg aaacgcctgg tatctttata gtcctgtcgg   2340

gtttcgccac ctctgacttg agcgtcgatt tttgtgatgc tcgtcagggg ggcggagcct   2400

atggaaaaac gccagcaacg cggcctttt acggttcctg ccttttgct ggccttttgc   2460

tcacatgttc tttcctgcgt tatcccctga ttctgtggat aaccgtatta ccgcctttga   2520

gtgagctgat accgctcgcc gcagccgaac gaccgagcgc agcgagtcag tgagcgagga   2580

agcggaagag cgcccaatac gcaaaccgcc tctccccgcg cgttggccga ttcattaatg   2640

cagctggcac gacaggtttc ccgactggaa agcgggcagt gagcgcaacg caattaatgt   2700

gagttagctc actcattagg caccccaggc tttacacttt atgcttccgg ctcgtatgtt   2760

gtgtggaatt gtgagcggat aacaatttca cacaggaaac agctatgacc atgattacgc   2820

caagcgcgca attaaccctc actaaaggga acaaaagctg gagctcctgc aggttgttgg   2880

ttggcacacc acaaatatac tgttgccgag cacaattgtc tagaatgcat acgcattaag   2940

cgaacattaa aaagatgtga aaacataact attatgtcta aataaacaca cgtcagatgt   3000

atgtacgtca acgaaaaacc attgtctata tattacaatt actaaataca taccaaattg   3060

aatacatatt gatgaaaaat aataaatact ggcgaaagca aaaaaacgaa acatttttat   3120

tttattgaac aactctcagg ctccaggtag gcaaaaaagc accgactcgg tgccactttt   3180

tcaagttgat aacggactag ccttatttca acttgctatg ctgtttccag catagctctg   3240

aaacctccag tctctggttg tgcacgacgt taaattgaaa ataggtctat atatacgaac   3300

tgagtctgga aaaagaagtt gagaattata aaaagtagtg agcactggcg ccctctctgc   3360

ttggcgagct aacctttcg cctcttggct gagtaggtgg cgtttcattc tactctgtaa   3420

aattaatgta gaattgaaac actatggtca aaaaatactt aggggcataa gttataaaac   3480

gtatgaaatg aatttttatc aacctgggct attcaaaaat tttcgaatta ttttatgtat   3540

tttttttaat cgttttttcat tataggttaa aatacacttt aaaaggaatt ctttcctgta   3600

aaataaatat aaataaatat gctttattga cagaaaattt gatgttttg tatttgagta   3660

ggagcaatca caggtgagca aaaaagaatt catcaattga tcggctaaat ggtatggcaa   3720

gaaaaggtat gcaatataat aatctttat tgggtatgca acgaaaattt gtttcgtcaa   3780

cgtatgcaat attctttatt aaaagagggt atgcaatgta ttttattaaa aacgggtatg   3840

caatataata atcttttatt gggtatgcaa cgaaaatttg tttcgtcaaa gtatgcaata   3900

ttttttatta aaagagggta tgcaatgtat tttattaaaa acgggtatgc aataaaaaat   3960

tatttggttt ctctaaaaag tatgcagcac ttatttttg ataaggtatg caacaaaatt   4020

ttactttgcc gaaatatgc aatgtttttg cgaataaatt caacgcacac ttattacgtg   4080

gccaactagg tgcccaaaat gtgtgtggac tacgaaattc tccaaattta agatgctatc   4140

tttaaaccaa tgaaatatgg ttcgtatact atgaattttc aattaggcga acatcaatga   4200

ttcccccccc agaaaaccga catagcagag cacacgagca ggcgcaaatt gagaaaccca   4260

tccgcgtgaa gtcggttaat ttgcccatct tcttctggac gcgttcgtgc accgctgct   4320
```

-continued

```
catccggcgg agtattgtac cattgtgtac ggccgtagtc cgtgtgcctt cgttttggcg    4380

ttcatgcatg agcagcccaa ttccttgctg ccccattcgg ttacattgca cagtggacac    4440

aaaagctagt tttgtagtca aagtacagaa ttcacaaatt atataaactg atatagttca    4500

tagatagtat aaactgatac caagtaacag atacacattt aaataggtaa actgtgtctg    4560

tgatcaaact gtttcttttc gtgtcgaaga atcaattaaa aatgattgaa tcattatatt    4620

tatttccgtt aaaagctgtg caggctgttc aaaatgtttt aatgaaaaaa tacgaatttt    4680

tagactgtct gaatcacagt gtgctcgctt acatttccgc tttcctcttt tggcaactcg    4740

atgtcgcctt tggggctctt ttggagaccg gaaaaaggca acattttcta ttcgattctt    4800

tttgccaatt gcccgagact gtgtcctgtc ggcatatgac gaatacgtaa cgtacgtgac    4860

ggcgacgtta ctcatacgca ccgtgggtag ctgcagacat ctcagcaccc acgttcgcga    4920

attattttga attcgactcc ctgggcgata tttgtttttc gcttttgcat attttgcggg    4980

caatttgggt aaaaggattt ccgcactctg cgacgccgtc ttcagtttgc ggctttcgtt    5040

tttttcctag tagttcggca cacatttttcc tcgccgcttc ggcaaatcgc tcacgtaaaa    5100

tatgcatgcg tttccttggc ggttttgcgc tctcaagtgc ctgcaattca attacatttc    5160

gattgatttt catgtttggc cccaaatcgc ggcaaaacct ctcactgacg gacacaccga    5220

agccccggcg gcaaccctca gcggatgccc cggggcttca cgttttccca ggtcagaagc    5280

ggttttcggg agtagtgccc caactggggt aacctttgag ttctctcagt tggggggcgta    5340

gggtcgccga catgacacaa ggggttgtga ccggggtgga cacgtacgcg ggtgcttacg    5400

accgtcagtc gcgcgagcgc gactgctgca ggcgggcagc aaggcgtccc atccgcatta    5460

cgtgcccagc tatttgccag ctatgcccga tcctcatgcc tatattcgaa cgcccacgca    5520

caagcagccc gtaaccgaat acgaggcaat aagggaaaag gcagccagtc agaagcgtga    5580

cgttgagaag gcgctgacca aatttctgtg caaaacaaca gaaacaaaca atctctttcc    5640

caccgaggac aacatgtttc cgtgtaagta agcgctgcga ttaatggttc ttggttcttt    5700

attcaaatgt ttcgacttct ttttctgaat gcaacagtaa tcgcctgtaa gcccgccttt    5760

ccggcgtatg cagctgcctt gaatcccaca gatcaggtat ttgacttcga ggagctggag    5820

taccactact tggtggccaa tcgtacggaa gatgtgccca gtaaaggtag gtccaaattg    5880

tacacaatag atattccaat gaacacaggc tctactttca tttgcagagg agggcgagga    5940

gggtgacagt gagaatgagg aactggatgg cgacaagtcc aaggaggaga gcccgagct    6000

ggagatcaag cccaattcaa caacaaataa agctatttta gagaatccca atatagacaa    6060

tccctacttg cgtgccgcta cactgccaaa gcgttccaag ctgcacagtg agtgcactac    6120

accacgcatg gtgccctcac gaagtataca ctcggcttca cccacgacac cgacgccctc    6180

aactctagag ataaccaaaa gtagtgctta gttataatta taaatagatg cattgtaatt    6240

gtgtatagtt ttttaaaaaa aaaatattgg ataaacaaac tcttttcttc ttatcgatag    6300

ttcgtgcttt tgcttaaaat ggtgtgcgat ggcagcgctg cggcaacaaa cagctgtttc    6360

gatataaaag tacattttac ttatcgatag ctcgtgctat tgcataaaat gatgttaggt    6420

ggcaacgctg cggcaacaat cagctgttta ccaggccgca gcaacgttac agtgcatttt    6480

acattttacc aagttgaatt aataaaattg ctttttaaaa gtgtttacta aattaaaaag    6540

ccaacaaatt gttgttgttt tcgttgctta caagcggctg ctgtacataa attatacata    6600
```

-continued

```
ttagcgctaa acgtgctcaa catgaatttc ctgcgccaaa cattcaacgt tacgaaacaa    6660

ttgacggcac aaggtaagtt ttaacaaaaa tccctattta aaacattgcg ttgcggctat    6720

ttattcaact tcgagtcccg tgttctatat acatacgcgc ccacgcgcct aattgccaac    6780

catgtgaggc agccggtagc cgcttgcgca catccatttc caattggtga ctgtgcgcat    6840

tttgtgttta tccaaggatc ctgcgttcca ttgtgtgcac acaatgattt gtattgtctg    6900

ttgtttgcct gcgatctcaa ctcttttaca tgggcgcgtg gccggcttgc gagcctgtcg    6960

cccgtctgcc agttctctag ttgtcgtcgt accccccttc ccctgccca gcccttatc      7020

gtgtgtctag tctgtgaata ttttttataag cattttctca tgtgtgtttc ctgtttgtgt    7080

gttttaatgt gtcctcaaaa ctgttcacgg agcctacaaa gtgtgtattg agaatatata    7140

tatatatata tagtccatct gtccatcttg gatatttgtc attggaacgg gcgagcgaaa    7200

aaagggtttg tcaatgaaaa acttatcatt ttcattatgt gcaacattta ctaaccaaat    7260

ctattcaata cataggttgg acaaacttgc cttctgttct tcgagataac ttcagcaaag    7320

tctgccaatg cgatctgaag tccattcaat ttttggccta gcaaaaaacg cattcgtttt    7380

tctgcttgtt ttaattaaaa ttcacaacaa aatccgcata acatgaggcc caccctcaac    7440

aataggaatt tgcatgacat gcacacaata aggaaaaaac aacacaagaa aaaaaaatta    7500

tgagaaaagg acacacacac acacacaggt gcgttcggat cgcggcagac aatgcacgga    7560

gctgtgattg gcatagttct tgctgtgcgc ctgctcccat tgtaagcgat tgtccagcgt    7620

tatggtaatt attacctgtg tgtacgtgtg tgtgtgtgtg tgtggcattt aattaaaaat    7680

tgttgtcgtt tgcgatttttg gctgcagtac agtcgagtcc agtcgggagt ccagctgaac    7740

agaaatctga gcatcagaca gtcaaccccc gtgcatggct aaaggttctc aatgcttaaa    7800

aggcttgaga actgcagttg ccgctgaccc acagccgcgt catttggctg caattatttg    7860

tgaaaataac cttatatatg catgatatgt ggatggatat ggatggatat atggatgtgc    7920

gcagcataac aattattttg cgattttcac agagattagc cacaacaaaa ggcgaatggc    7980

cattgttgct tgggcatttg gaactggcca actgtttctg accctttgt catgttgtgt     8040

ccgttctctc gtttgtgtca aatgttttta gccgctcgct gcggctgcgc tcacacatgc    8100

ggcagcagct accatataca atttatatac caatatatgt acacatattt aattggtaca    8160

gttgtgtcca cttgcattgt atgtgtacac ttaacgcact cttgcaattc cggacaagtc    8220

aagaggagac aactagacgt tggcaatcgg aaattggaag ccttacagaa acactgcgtt    8280

tataacttgt tctcagctgt ttctctctct catcttgatt acattgcagc gctgcagagc    8340

aattatttgt gtgccgcatt gcgcggcatg gcatcgttga atcaaatgca tcgcactggg    8400

ccgcatataa agaagcgtcc gccacgtcag cccctggacg gtaaaccgtt tgccaaggga    8460

gtggtgctca agacactgat caagaagcca aagaaaccaa actcggcgaa tcgtaaatgc    8520

gcgctggtgc gcttatccac gggaaaggag atggtcgcct atataccgg cattggacat     8580

aatctgcagg agcataatat tgtactgtgt cgcgtcggac gactgcagga tgtgcccggc    8640

gtcaagctga aggcggtgcg cggtgtctac gatctggcgc acgttatcaa gaagggccaa    8700

tgacaaccaa ctaccatgta attctcttcc ataaaaaaac aaaaaaaaaa taagaaaaga    8760

aaacaagcca aatctttgag tactctgcta ttcttgtgca gcatatatta ttatgatttt    8820

tttaatggaa aattatgcag ctcagcggga ttagtgtaag tagccaacac acaacaagtg    8880
```

-continued

agctctggcc	tcgcctcatc	ccaacttgtc	ttgccgtaat	cttaagtcaa	caggccaaat	8940
tgcgagccaa	acaattggcc	agtgttgcca	acgacgctgc	cgaaaaagga	gctaaatccc	9000
attggaaaat	agctaaaaaa	tagccagagc	atgaattgga	cgactgaaga	cagctgaaat	9060
tggccagaat	ttggccagaa	tatagctgat	atcgcaacac	tgcacatttg	ttgcccactc	9120
gaaatatgat	tttaacggca	tttttacgct	ttagcaggca	aatccctttt	tgaagaaacg	9180
gcccgtcttt	acttttttaac	agaatttgct	tgcacaaatt	ttatgccaaa	taatcgttaa	9240
gcgaaatggg	cgtcgacagg	ccacgccaca	ttttcacagt	acgcagagcc	tcgccacgcc	9300
tttgcaacag	gatacaacaa	attttgaata	ggcaccgacc	aggttgtcgt	gcatctgggc	9360
tgaacgataa	catctgtatt	aaatcaatcc	catattcaag	cttccaaagg	atttcggcaa	9420
catgccaatt	ccgctattaa	tctttcgggt	tatcctgaat	atctgggtag	tcctaggtgt	9480
ggagctggtg	tcgcttatgg	cagagcagct	gaatgctaac	atatacgagc	ataaaaagtt	9540
tcatcaggaa	tccatatgca	acttgaggtc	cgaccccacc	aacttttttg	cactgcaaaa	9600
aaacacgctt	ttgcacgcgg	gcccatacat	agtacaaact	ctacgtttcg	tagactattt	9660
tacataaata	gtctacaccg	ttgtatacgc	tccaaataca	ctaccacaca	ttgaaccttt	9720
ttgcagtgca	aaaagtacg	tgtcggcagt	cacgtaggcc	ggccttatcg	ggtcgcgtcc	9780
tgtcacgtac	gaatcacatt	atcggaccgg	acgagtgttg	tcttatcgtg	acaggacgcc	9840
agcttcctgt	gttgctaacc	gcagccgac	gcaactcctt	atcggaacag	gacgcgcctc	9900
catatcagcc	gcgcgttatc	tcatgcgcgt	gaccggacac	gaggcgcccg	tcccgcttat	9960
cgcgcctata	aatacagccc	gcaacgatct	ggtaaacaca	gttgaacaga	tggtgagcaa	10020
gggcgaggag	gtcatcaaag	agttcatgcg	cttcaaggtg	cgcatggagg	gctccatgaa	10080
cggccacgag	ttcgagatcg	agggcgaggg	cgagggccgc	ccctacgagg	gcacccagac	10140
cgccaagctg	aaggtgacca	agggcggccc	cctgcccttc	gcctgggaca	tcctgtcccc	10200
ccagttcatg	tacggctcca	aggcgtacgt	gaagcacccc	gccgacatcc	ccgattacaa	10260
gaagctgtcc	ttccccgagg	gcttcaagtg	ggagcgcgtg	atgaacttcg	aggacggcgg	10320
tctggtgacc	gtgacccagg	actcctccct	gcaggacggc	acgctgatct	acaaggtgaa	10380
gatgcgcggc	accaacttcc	cccccgacgg	ccccgtaatg	cagaagaaga	ccatgggctg	10440
ggaggcctcc	accgagcgcc	tgtaccccg	cgacggcgtg	ctgaagggcg	agatccacca	10500
ggccctgaag	ctgaaggacg	gcggccacta	cctggtggac	ttcaagacca	tctacatggc	10560
caagaagccc	gtgcaactgc	ccggctacta	ctacgtggac	accaagctgg	acatcacctc	10620
ccacaacgag	gactacacca	tcgtggaaca	gtacgagcgc	tccgagggcc	gccaccacct	10680
gttcctgggg	catggcaccg	gcagcaccgg	cagcggcagc	tccggcaccg	cctcctccga	10740
ggacaacaac	atggccgtta	tcaaggaatt	tatgcgcttc	aaagttagga	tggagggatc	10800
catgaacgga	catgagttcg	agatcgaggg	agagggcgag	ggacgccgt	atgaaggcac	10860
acaaacagcc	aaactcaagg	tcaccaaggg	cggaccactg	cccttcgcct	gggatatcct	10920
gagtccccag	tttatgtacg	gcagcaaggc	ctacgttaag	caccccgctg	acataccgga	10980
ctacaaaaag	ctgtcctttc	cggaaggctt	caagtgggag	cgcgtgatga	atttcgaaga	11040
cggaggactg	gtcactgtga	cccaagatag	cagtttgcag	gacggtacac	tgatctataa	11100
ggttaaaatg	cgcggcacta	actttccgcc	agatggccca	gtgatgcaga	agaagaccat	11160

-continued

```
gggttgggag gcatccaccg aacgtctgta ccctcgagac ggagtgctca agggcgagat    11220

ccatcaggcc ctcaaactga aagatggtgg tcactacctg gtcgaattta agaccattta    11280

catggccaag aagccggttc agctgcccgg atattattat gtggatacga aactggatat    11340

aacttcgcat aacgaagact acaccattgt cgagcagtat gagcgcagcg aaggccgaca    11400

tcacctgttc ctctacggca tggacgagct gtacaagtag gcggccgcga ctctagatca    11460

taatcagcca taccacattt gtagaggttt tacttgcttt aaaaaacctc ccacacctcc    11520

ccctgaacct gaaacataaa atgaatgcaa ttgttgttgt taacttgttt attgcagctt    11580

ataatggtta caaataaagc aatagcatca caaatttcac aaataaagca ttttttttcac   11640

tgcattctag ttgtggtttg tccaaactca tcaatgtatc ttaaagcttc agtctctggt    11700

tgtgcatgga gcgtgttttt ggggcgattt acgagcgtga accgatattt tccatacgtc    11760

atgtttgcct tgctttacac tccaggaaac ctcatccgtt taatttagcc ggaagtgttg    11820

caatagatgc cacatcacaa tcgacttaat aattttttta ggagcaagtt ttaatggaaa    11880

cagtttctga taaataaata tacatatcaa cttagtacaa gaatatccag ctgaaagaat    11940

ggtatatata tatatatata tacttgtgtt tgtttctatg acagtttctt cacagctttc    12000

gattttctta tggcacatcg cgcgacagtt gaaatgaaaa actgaaatca gtgaaacccc    12060

gaaaaaactg aaaaagccac ggaaattgta cagatacaca gatacgcaga tacaccagca    12120

aatgtacaaa aaggtctata tacatatata tgtagctact cacttatgtt gtccttcgca    12180

gattgctccc tttaagcaaa taaaaaaaag ttggctccac gccgaaaaga aataaaatta    12240

aatggagaat cgcaaatcca tagtgagcca aaggcaaatc tataaaagaa atgaaacgaa    12300

attcattcat tttcgttttc gagttcgaat atttaagtta tatatataaa cgcagtattt    12360

atccatgtaa tcgaaccaca aaagcccaat gagaaaaccc tacattttat gctgagcatc    12420

acaaaatgcc tttcctttca catgaattta tgtattttaa tcaatttccc tcgctgtggc    12480

agttaaatat cctaaatttg tccaacgaaa ttgatgcttc aattattcga atgacgacgt    12540

ttaatgggct ttcgaggaat aaaagcaaaa attcacaaga aaaacgcctc tgcatccatg    12600

ctcattatcg gaatcaatta aaatttcaca tgtatcgtta gcatggccat gtcagcaaat    12660

ccacgggatt cggctagagt cctccaaaat acgcccacgg gacccataca ccttcgaaat    12720

gatccaacat caatccctat ccaaatgtat acttagatat gtacatacct tgtcttttct    12780

tggtcggcga atgggggttc cagcatccca tcatcct                             12817
```

```
<210> SEQ ID NO 40
<211> LENGTH: 14148
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: tko-step2

<400> SEQUENCE: 40

atcatcgatc tcgaggctgc atccaacgcg cgcgttggga gctctccgga tcaattcggc     60

ttcaggtacc gtcgacgatg taggtcacgg tctcgaagcc gcggtgcggg tgccagggcg    120

tgcccttggg ctccccgggc gcgtactcca cctcacccat ctggtccatc atgatgaacg    180

ggtcgaggtg gcggtagttg atcccggcga acgcgcggcg caccgggaag ccctcgccct    240

cgaaaccgct gggcgcggtg gtcacggtga gcacgggacg tgcgacggcg tcggctgggt    300
```

-continued

```
gcggatacgc ggggcagcgt cagcgggttc tcgacggtca cggcgggcat gtcgacgaca   360

tgttcgcctc atttgtgttc gtttatgtat tcgatgttat gtgtatgctc atgtgatgtt   420

tagcttgtaa gcgcgagatg tgggtagcag gagatgcagt gcagccaaca gcagtgacca   480

gatgatatat gctatgctac tactactact tatatgctat gatttgtggc gcggaggcgt   540

gtctgcgaca cataatcccg cccatttagc tttaagattc aggcactaag aagcaattcg   600

atcaataaat tattgtaacc actctgcatg tgagcaaaag gccagcaaaa ggccaggaac   660

cgtaaaaagg ccgcgttgct ggcgtttttc cataggctcc gcccccctga cgagcatcac   720

aaaaatcgac gctcaagtca gaggtggcga acccgacag gactataaag ataccaggcg   780

tttccccctg gaagctccct cgtgcgctct cctgttccga ccctgccgct taccggatac   840

ctgtccgcct ttctcccttc gggaagcgtg gcgctttctc atagctcacg ctgtaggtat   900

ctcagttcgg tgtaggtcgt tcgctccaag ctgggctgtg tgcacgaacc ccccgttcag   960

cccgaccgct gcgccttatc cggtaactat cgtcttgagt ccaacccggt aagacacgac  1020

ttatcgccac tggcagcagc cactggtaac aggattagca gagcgaggta tgtaggcggt  1080

gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaagaac agtatttggt  1140

atctgcgctc tgctgaagcc agttaccttc ggaaaaagag ttggtagctc ttgatccggc  1200

aaacaaacca ccgctggtag cggtggtttt tttgtttgca agcagcagat tacgcgcaga  1260

aaaaaaggat ctcaagaaga tcctttgatc ttttctacgg ggtctgacgc tcagtggaac  1320

gaaaactcac gttaagggat tttggtcatg agattatcaa aaaggatctt cacctagatc  1380

ctttttaaatt aaaaatgaag ttttaaatca atctaaagta tatatgagta aacttggtct  1440

gacagttacc aatgcttaat cagtgaggca cctatctcag cgatctgtct atttcgttca  1500

tccatagttg cctgactccc cgtcgtgtag ataactacga tacgggaggg cttaccatct  1560

ggccccagtg ctgcaatgat accgcgagac ccacgctcac cggctccaga tttatcagca  1620

ataaaccagc cagccggaag ggccgagcgc agaagtggtc ctgcaacttt atccgcctcc  1680

atccagtcta ttaattgttg ccgggaagct agagtaagta gttcgccagt taatagtttg  1740

cgcaacgttg ttgccattgc tacaggcatc gtggtgtcac gctcgtcgtt tggtatggct  1800

tcattcagct ccggttccca acgatcaagg cgagttacat gatcccccat gttgtgcaaa  1860

aaagcggtta gctccttcgg tcctccgatc gttgtcagaa gtaagttggc cgcagtgtta  1920

tcactcatgg ttatggcagc actgcataat tctcttactg tcatgccatc cgtaagatgc  1980

ttttctgtga ctggtgagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccg  2040

agttgctctt gcccggcgtc aatacggat aataccgcgc cacatagcag aactttaaaa  2100

gtgctcatca ttggaaaacg ttcttcgggg cgaaaactct caaggatctt accgctgttg  2160

agatccagtt cgatgtaacc cactcgtgca cccaactgat cttcagcatc ttttactttc  2220

accagcgttt ctgggtgagc aaaaacagga aggcaaaatg ccgcaaaaaa gggaataagg  2280

gcgacacgga aatgttgaat actcatactc ttccttttc aatattattg aagcatttat  2340

cagggttatt gtctcatgag cggatacata tttgaatgta tttagaaaaa taaacaaata  2400

ggggttccgc gcacatttcc ccgaaaagtg ccacctgacg tcgacggatc gggagatcgg  2460

cgcgggatct aattcaatta gagactaatt caattagagc taattcaatt aggatccaag  2520

cttatcgatt tcgaaccctc gaccgccgga gtataaatag aggcgcttcg tctacggagc  2580
```

-continued

```
gacaattcaa ttcaaacaag caaagtgaac acgtcgctaa gcgaaagcta agcaaataaa    2640

caagcgcagc tgaacaagct aaacaatcgg ctcgagaccg gtcgccacca tggtgagcaa    2700

gggcgaggag ctgttcaccg gggtggtgcc catcctggtc gagctggacg gcgacgtaaa    2760

cggccacaag ttcagcgtgt ccggcgaggg cgagggcgat gccacctacg gcaagctgac    2820

cctgaagttc atctgcacca ccggcaagct gcccgtgccc tggcccaccc tcgtgaccac    2880

cctgacctac ggcgtgcagt gcttcagccg ctaccccgac cacatgaagc agcacgactt    2940

cttcaagtcc gccatgcccg aaggctacgt ccaggagcgc accatcttct tcaaggacga    3000

cggcaactac aagacccgcg ccgaggtgaa gttcgagggc gacaccctgg tgaaccgcat    3060

cgagctgaag ggcatcgact tcaaggagga cggcaacatc ctggggcaca gctggagta    3120

caactacaac agccacaacg tctatatcat ggccgacaag cagaagaacg gcatcaaggt    3180

gaacttcaag atccgccaca acatcgagga cggcagcgtg cagctcgccg accactacca    3240

gcagaacacc cccatcggcg acggccccgt gctgctgccc gacaaccact acctgagcac    3300

ccagtccgcc ctgagcaaag accccaacga gaagcgcgat cacatggtcc tgctggagtt    3360

cgtgaccgcc gccgggatca ctctcggcat ggacgagctg tacaagtaaa gcggccgcga    3420

ctctagatca taatcagcca taccacattt gtagaggttt tacttgcttt aaaaaacctc    3480

ccacacctcc ccctgaacct gaaacataaa atgaatgcaa ttgttgttgt taacttgttt    3540

attgcagctt ataatggtta caaataaagc aatagcatca caaatttcac aaataaagca    3600

tttttttcac tgcattctag ttgtggtttg tccaaactca tcaatgtatc ttagttgttg    3660

gttggcacac cacaaatata ctgttgccga gcacaattga tcggctaaat ggtatggcaa    3720

gaaaaggtat gcaatataat aatctttat tgggtatgca acgaaaattt gtttcgtcaa    3780

cgtatgcaat attctttatt aaaagagggt atgcaatgta ttttattaaa aacgggtatg    3840

caatataata atctttttatt gggtatgcaa cgaaaatttg tttcgtcaaa gtatgcaata    3900

ttttttatta aaagagggta tgcaatgtat tttattaaaa acgggtatgc aataaaaaat    3960

tatttggttt ctctaaaaag tatgcagcac ttattttttg ataaggtatg caacaaaatt    4020

ttactttgcc gaaaatatgc aatgtttttg cgaataaatt caacgcacac ttattacgtg    4080

gccaacgcgc ctagtggatc cttcctggcc cttttcgaga aacgccgcga gggcgaaaag    4140

gattagttgt ttcaaacgca agaaggacat ttgtttcctt aaattgtaac catttcttta    4200

tttggcactc gagccattga attttttcatt ttcagaatat gtgtacacat ttttttaaaaa    4260

aataaaaaaa ttatataatg ctggcggttg tttcatgtgt gaaaaattga tcaatggtaa    4320

acaaaattga ataaatatat aacatatata tatagatatg tgtgttgaaa tgaatacttg    4380

cgatacatgt aataaaaata ctcttcgctt atctatcaaa aagtgcggaa tgtcaaaatt    4440

taaaattttta caatgaatgc gtagccgacg acgaaagtgt ccttgctat ttcctttagc    4500

aagatttaaa tttagattaa attctaatga tacgattgac agttcgaaat tcaaagtgtt    4560

ccttttttcaa aatttagtaa agattgtata tcaattgtag atatatcgaa atttttcggc    4620

cgcaagcgaa cattttacaa aatgaaggcg accagttgca gaccaattcc attcatcaac    4680

tttcggattg taagatattt ctatcggcca cgacgattga acaagtatta cgatattgta    4740

agtcttcttt aacaaaatta gtttcccttt cacagaaaca gacataaatt cttgaattat    4800

tgacttggat ttgagtgatc gttcgttgtc tatactataa gatctatagg cacgggataa    4860
```

-continued

```
cgctctaaat ctctttaaaa tcgaacgcgc caggcgctag ttaaacgtta ctatctatct    4920

ggttaaccca gctttgatcg gaatgcgtat atatatttca tgttatataa acgctgcaaa    4980

agctgccaga gcctctgctc cagagctgga ttcgctcaca ccttcctctt cttcttgggg    5040

tcagccctgc tgtctccacc gagctgagag aggtcgattc ttgtttcata gagccccgta    5100

attgactgat gaatcagtgt ggcgtccagg acctcctttg tagaggtgta ccgctttctg    5160

tctatggtgg tgtcgaagta cttgaaggct gcaggcgcgc ccaagttggt cagagtaaac    5220

aagtggataa tgttttctgc ctgctccctg atgggcttat ccctgtgctt attgtaagca    5280

gaaagcacct tatcgaggtt agcgtcggcg aggatcactc ttttggagaa ttcgcttatt    5340

tgctcgatga tctcatcaag gtagtgtttg tgttgttcca cgaacagctg cttctgctca    5400

ttatcttcgg gagacccttt gagctttca tagtggctgg ccagatacaa gaaattaacg    5460

tatttagagg gcagtgccag ctcgttacct ttctgcagct cgcccgcact agcgagcatt    5520

cgtttccggc cgttttcaag ctcaaagaga gagtacttgg gaagcttaat gatgaggtct    5580

tttttgacct ctttatatcc tttcgcctcg agaaagtcga tggggttttt ttcgaagctt    5640

gatcgctcca tgattgtgat gcccagcagt tccttgacgc ttttgagttt tttagacttc    5700

cctttctcca ctttggccac aaccagtaca ctgtaagcga ctgtaggaga atcgaatccg    5760

ccgtatttct tggggtccca atcttttttg cgtgcgatca gcttgtcgct gttccttttc    5820

gggaggatac tttccttgga gaagcctccg gtctgtactt cggtcttttt aacgatgttc    5880

acctgcggca tggacaggac cttccggact gtcgcgaaat ccctacccct gtcccacacg    5940

atttctcctg tttctccgtt tgtttcgata agtggtcgct tccgaatctc tccattggcc    6000

agtgtaatct cggtcttgaa aaaattcata atattgctgt aaaagaagta cttagcggtg    6060

gccttgccta tttcctgctc agactttgcg atcattttcc taacatcgta cactttatag    6120

tctccgtaaa caaattcaga ttcaagcttg ggatatttt tgataagtgc agtgcctacc    6180

actgcattca ggtaggcatc atgcgcatgg tggtaattgt tgatctctct caccttataa    6240

aactgaaagt cctttctgaa atctgagacc agcttagact tcagagtaat aactttcacc    6300

tctcgaatca gtttgtcatt ttcatcgtac ttggtgttca tgcgtgaatc gagaatttgg    6360

gccacgtgct tggtgatctg gcgtgtctca acaagctgcc ttttgatgaa gccggcttta    6420

tccaactcag acaggccacc tcgttcagcc ttagtcagat tatcgaactt ccgttgtgtg    6480

atcagtttgg cgttcagcag ctgccgccaa taatttttca ttttcttgac aacttcttct    6540

gaggggacgt tatcactctt ccctctattt ttatcggatc ttgtcaacac tttattatca    6600

atagaatcat ctttgagaaa agactggggc acgatatgat ccacgtcgta gtcggagagc    6660

cgattgatgt ccagttcctg atccacgtac atgtccctgc cgttctgcag gtagtacagg    6720

tagagcttct cattctgaag ctgggtgttt tcaactgggt gttccttaag gatttgggac    6780

cccagttctt ttataccctc ttcaatcctc ttcatccttt ccctactgtt cttctgtccc    6840

ttctgggtag tttggttctc tcgggccatc tcgataacga tattctcggg cttatgcctt    6900

cccattactt tgacgagttc atccacgacc ttaacggtct gcagtattcc cttttttgata    6960

gctgggctac ctgcaagatt agcgatgtgc tcgtgaagac tgtcccccctg gccagaaact    7020

tgtgctttct ggatgtcctc cttaaaggtg agagagtcat catggatcaa ctgcatgaag    7080

ttccggttgg caaatccatc ggacttaaga aaatccagga ttgtctttcc actctgcttg    7140
```

```
tctcggatcc cattgatcag ttttcttgac agccgccccc atcctgtata tcggcgcctc    7200

ttgagctgtt tcatgacttt gtcgtcgaag agatgagcgt aagtttcaa gcgttcttca     7260

atcatctccc tatcttcaaa caacgtaagg gtgaggacaa tgtcctcaag aatgtcctcg     7320

ttctcctcat tgtccaggaa gtccttgtct ttaatgattt tcaggagatc gtgatacgtt     7380

cccaggggatg cgttgaagcg atcctccact ccgctgattt caacagagtc gaaacattca    7440

atctttttga aatagtcttc tttgagctgt ttcacggtaa cttttccggtt cgtcttgaag    7500

aggaggtcca cgatagcttt cttctgctct ccagacagga atgctggctt tctcatccct     7560

tctgtgacgt atttgacctt ggtgagctcg ttataaactg tgaagtactc gtacagcaga     7620

gagtgtttag gaagcacctt ttcgttaggc agatttttat caaagttagt catcctttcg     7680

atgaaggact gggcagaggc cccctttatcc acgacttcct cgaagttcca gggagtgatg    7740

gtctcttctg atttgcgagt catccacgcg aatctggaat ttccccgggc gagggggcct     7800

acatagtagg gtatccgaaa tgtgaggatt ttctcaatct tttccctgtt atctttcaaa     7860

aaggggtaga aatcctcttg ccgcctgagg atagcgtgca gttcgcccag gtgaatctgg     7920

tggggatgc ttccattgtc gaaagtgcgc tgtttgcgca acagatcttc tctgttaagc      7980

tttaccagca gctcctcggt gccgtccatt ttttccaaga tgggcttaat aaatttgtaa     8040

aattcctcct ggcttgctcc gccgtcaatg tatccggcgt agccattttt agactgatcg     8100

aagaaaattt ccttgtactt ctcaggcagt tgctgtctga caagggccctt cagcaaagtc    8160

aagtcttggt ggtgctcatc atagcgcttg atcatactag cgctcagcgg agctttggtg     8220

atctccgtgt tcactcgcag aatatcactc agcagaatgg cgtctgacag gttctttgcc     8280

gccaaaaaaa ggtctgcgta ctggtcgccg atctgggcca gcagattgtc gagatcatca     8340

tcgtaggtgt ctttgctcag ttgaagcttg gcatcttcgg ccaggtcgaa gttagattta    8400

aagttgggggg tcagcccgag tgacagggcg ataagattac caaacaggcc gttcttcttc   8460

tccccaggga gctgtgcgat gaggttttcg agccgccggg atttggacag cctagcgctc     8520

aggattgctt tggcgtcaac tccggatgcg ttgatcgggt tctcttcgaa aagctgattg     8580

taagtctgaa ccagttggat aaagagtttg tcgacatcgc tgttgtctgg gttcaggtcc     8640

ccctcgatga ggaagtgtcc ccgaaatttg atcatatgcg ccagcgcgag atagatcaac     8700

cgcaagtcag ccttatcagt actgtctaca agcttcttcc tcagatgata tatggttggg     8760

tactttttcat ggtacgccac ctcgtccacg atattgccaa agattgggtg gcgctcgtgc   8820

tttttatcct cctccaccaa aaaggactcc tccagcctat ggaagaaaga gtcatccacc    8880

ttagccatct cattactaaa gatctcctgc aggtagcaga tccgattctt tctgcgggta    8940

tatctgcgcc gtgctgttct tttgagccgc gtggcttcgg ccgtctcccc ggagtcgaac    9000

aggagggcgc caatgaggtt cttctttatg ctgtggcgat cggtattgcc cagaactttg    9060

aattttttgc tcggcacctt gtactcgtcc gtaatgacgg cccagccgac gctgtttgtg    9120

ccgatatcga gcccaatgga gtacttcttg tccatggcga aaatccgggt cgaaagttac     9180

ggttatcgcg cactctactt tccacaaatc ctcacccaaa aaccaagcac agtttattca     9240

actgaagtat tcgcgatact tctttatcta ataataatgt acatgtaact aaaactcgctt    9300

ttgggttaaa atcgtgacgc agaggcaaaa aaaatcgtat gtcccttaga caacttgaaa    9360

caactgcgaa gcgtacggca attccaggaa ttttgtggta aagctacgcg ccaactaacg     9420
```

45

```
gttcttgctt agaggtggaa taatgtagtt ttccagcgat aataaatata tcgatatttt      9480

tagtaaaatt gaaaaggtaa acttaatttt agaaaataat ttataagaaa tttaatagta      9540

tgcaaaataa tttttacttg ctaagaatat gtgccactaa ttaaaagctg gacaccgcgc      9600

aatggaaaat agtactacaa cacagcaaca aagcctgagt tatcaacaaa aaaatacgaa      9660

aacatctccc aaaactaagc acccacacgc gccactcgcc gtcacaacac aatcactgca      9720

caccaccatt cgaatttcgc gcactgtgac aacatcacat gatatcggcg cggcaacatc      9780

ggattaccga caaaacgaac tatcgcacga gccaccgccg gcgaagagcg ctcgttttgc      9840

aacaccggcg cgcgctgaac gaagagaaca gctgactgct tgatacgtgc gtgtttcgcg      9900

gcaggaatta cataaagttt agagcctctg acgccagacc ccccgaacat tcgctccgat      9960

caaactacct gcgaacggtc acctaatccc caccatgcat ggtaggttac ctctgatccc     10020

ggtcatcact ggcgttcgct cacatccgtc cttacatgtg catatttcga ggttaaaacg     10080

gtcgaagctt ggatccgcta gcgttgttgg ttggcacacc acaaatatac tgttgccgag     10140

cacaattgat cggctaaatg gtatggcaag aaaaggtatg caatataata atctttttatt     10200

gggtatgcaa cgaaaatttg tttcgtcaac gtatgcaata ttctttatta aaagagggta     10260

tgcaatgtat tttattaaaa acgggtatgc aatataataa tcttttattg ggtatgcaac     10320

gaaaatttgt ttcgtcaaag tatgcaatat ttttattaa aagagggtat gcaatgtatt     10380

ttattaaaaa cgggtatgca ataaaaaatt atttggtttc tctaaaaagt atgcagcact     10440

tatttttga taaggtatgc aacaaaattt tactttgccg aaaatatgca atgttttttgc     10500

gaataaattc aacgcacact tattacgtgg ccaactagcc tagttccagt gaaatccaag     10560

cacttgaggt ccgacccgat gaattctttt ttgctcacct gtgattgctc ctactcaaat     10620

acaaaaacat caaattttct gtcaataaag catatttatt tatatttatt ttacaggaaa     10680

gaattccttt taaagtgtat tttaacctat aatgaaaaac gattaaaaaa aatacataaa     10740

ataattcgaa aattttttgaa tagcccaggt tgataaaaat tcatttcata cgtttttataa     10800

cttatgcccc taagtatttt ttgaccatag tgtttcaatt ctacattaat tttacagagt     10860

agaatgaaac gccacctact cagccaagag gcgaaaaggt tagctcgcca agcagagagg     10920

gcgccagtgc tcactacttt ttataattct caacttcttt ttccagactc agttcgtata     10980

tatagaccta tttttcaattt aacgtcgctg cagcgatgcc attccagttt cagagctatg     11040

ctggaaacag catagcaagt tgaaataagg ctagtccgtt atcaacttga aaaagtggca     11100

ccgagtcggt gcttttttgc ctacctggag cctgagagtt gttcaataaa ataaaaatgt     11160

ttcgtttttt tgctttcgcc agtatttatt attttttcatc aatatgtatt caatttggta     11220

tgtatttagt aattgtaata tatagacaat ggttttccgt tgacgtacat acatctgacg     11280

tgtgtttatt tagacataat agttatgttt tcacatcttt ttaatgttcg cttaatgcgt     11340

atgcattcta gattttcaac gtcctcgata gtatagtggt tagtatcccc gcctgtcacg     11400

cgggagaccg gggttcaatt ccccgtcggg gagaatctgt gattcttttt tttttttcttt     11460

tactttgtta tataaacaat ttttgtttta attgaatcta atttgccatt gcttttttagga     11520

atctcaggca tccagcaagc gtttgtccgc cgaatcgccc atcagtgaag aagatcctgt     11580

ggcggctacg aaaatctccc cggccatgtc ggcctccacc tccagcgaaa aacccatcag     11640

cgagctggcc acctctgtgc tgacccaccg ctttccagac tccacctcct cacccggcga     11700
```

-continued

```
acatggcctt ggacgaatgc agttgtcgat ccgctacagc gcccagcgtc aaaaactaga   11760

cgtgaccata cacaaaatcc agaagatacc acttcgcgat cccagcaata tccccgatcc   11820

gtatgttaag ctgtatctgt tgcctggacg caccaaggag tcgaaacgca agacgagcgt   11880

gatcaaggac aactgcaacc ccgtctacga tgcatccttt gagtacctga tttccattgc   11940

cgaactcagg cagacggaac tggaggtgac ggtgtgcacc caaaagggat tcctatccgg   12000

cggtagtccc atcattggca tggtaggtac ccgaaagcaa cccccttagtt acagacacag   12060

cgcgtacgtc cttcgcatcc ttatgattcc caagtacata ttctgcaaga gtacagtata   12120

tataggaaag atatccgggt gaacttcgca ggacaacgcc cttggcgagt ttcagagcta   12180

tgctggaaac agcatagcaa gttgaaataa ggctagtccg ttatcaactt gaaaaagtgg   12240

caccgagtcg gtgctttttt gcctacctgg agcctgagag ttgttcaatc tagacaattg   12300

tgctcggcaa cagtatattt gtggtgtgcc aaccaacaac ctgcaggagc tccagctttt   12360

gttcccttta gtgagggtta attttttttg ctcacctgtg attgctccta ctcaaataca   12420

aaaacatcaa attttctgtc aataaagcat atttattat atttattta caggaaagaa   12480

ttccttttaa agtgtatttt aacctataat gaaaaacgat taaaaaaaat acataaaata   12540

attcgaaaat ttttgaatag cccaggttga taaaaaattca tttcatacgt tttataactt   12600

atgcccctaa gtatttttttg accatagtgt ttcaattcta cattaatttt acagagtaga   12660

atgaaacgcc acctactcag ccaagaggcg aaaaggttag ctcgccaagc agagagggcg   12720

ccagtgctca ctactttttta taattctcaa cttcttttttc cagactcagt tcgtatatat   12780

agacctattt tcaatttaac gtcgcaacat tgtactgtgc cgcggtttca gagctatgct   12840

ggaaacagca tagcaagttg aaataaggct agtccgttat caacttgaaa aagtggcacc   12900

gagtcggtgc ttttttgcct acctggagcc tgagagttgt tcaataaaat aaaaatgttt   12960

cgtttttttg ctttcgccag tatttattat ttttcatcaa tatgtattca atttggtatg   13020

tatttagtaa ttgtaatata tagacaatgg ttttccgttg acgtacatac atctgacgtg   13080

tgtttatttta gacataatag ttatgttttc acatctttt aatgttcgct taatgcgtat   13140

gcattctaga ttttcaacgt cctcgatagt atagtggtta gtatccccgc ctgtcacgcg   13200

ggagaccggg gttcaattcc ccgtcgggga gaatctgtga ttctttttttt ttttctttta   13260

ctttgttata taaacaattt ttgtttttaat tgaatctaat ttgccattgc ttttaggaat   13320

ctcaggcatc cagcaagcgt ttgtccgccg aatcgcccat cagtgaagaa gatcctgtgg   13380

cggctacgaa aatctccccg gccatgtcgg cctccacctc cagcgaaaaa cccatcagcg   13440

agctggccac ctctgtgctg acccaccgct ttccagactc cacctcctca cccggcgaac   13500

atggccttgg acgaatgcag ttgtcgatcc gctacagcgc ccagcgtcaa aactagacg   13560

tgaccataca caaaatccag aagataccac ttcgcgatcc cagcaatatc cccgatccgt   13620

atgttaagct gtatctgttg cctggacgca ccaaggagtc gaaacgcaag acgagcgtga   13680

tcaaggacaa ctgcaacccc gtctacgatg catcctttga gtacctgatt tccattgccg   13740

aactcaggca gacggaactg gaggtgacgg tgtgcaccca aaagggattc ctatccggcg   13800

gtagtcccat cattggcatg gtaggtaccc gaaagcaacc ccttagttac agacacagcg   13860

cgtacgtcct tcgcatcctt atgattccca agtacatatt ctgcaagagt acagtatata   13920

taggaaagat atccgggtga acttcgcacc agcacgcact ttcgatgttt cagagctatg   13980
```

-continued

```
ctggaaacag catagcaagt tgaaataagg ctagtccgtt atcaacttga aaaagtggca    14040

ccgagtcggt gcttttttgc ctacctggag cctgagagtt gttcaatcta gacaattgtg    14100

ctcggcaaca gtatatttgt ggtgtgccgt accgggccaa ttcgagct                 14148


<210> SEQ ID NO 41
<211> LENGTH: 4152
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Dvir-rescue-modified

<400> SEQUENCE: 41

gctgcaggcg ggcagcaagg cgtcccatcc gcattacgtg cccagctatt tgccagctat      60

gcccgatcct catgcctata ttcgaacgcc cacgcacaag cagcccgtaa ccgaatacga     120

ggcaataagg gaaaaggcag ccagtcagaa gcgtgacgtt gagaaggcgc tgaccaaatt     180

tctgtgcaaa acaacagaaa caaacaatct ctttcccacc gaggacaaca tgtttccgtg     240

taagtaagcg ctgcgattaa tggttcttgg ttctttattc aaatgtttcg acttcttttt     300

ctgaatgcaa cagtaatcgc ctgtaagccc gcctttccgg cgtatgcagc tgccttgaat     360

cccacagatc aggtatttga cttcgaggag ctggagtacc actacttggt ggccaatcgt     420

acggaagatg tgcccagtaa aggtaggtcc aaattgtaca caatagatat tccaatgaac     480

acaggctcta ctttcatttg cagaggaggg cgaggagggt gacagtgaga atgaggaact     540

ggatggcgac aagtccaagg aggagaagcc cgagctggag atcaagccca attcaacaac     600

aaataaagct attttagaga atcccaatat agacaatccc tacttgcgtg ccgctacact     660

gccaaagcgt tccaagctgc acagtgagtg cactacacca cgcatggtgc cctcacgaag     720

tatacactcg gcttcaccca cgacaccgac gccctcaact ctagagataa ccaaaagtag     780

tgcttagtta taattataaa tagatgcatt gtaattgtgt atagtttttt aaaaaaaaaa     840

tattggataa acaaactctt ttcttcttat cgatagttcg tgctttttgct taaaatggtg     900

tgcgatggca gcgctgcggc aacaaacagc tgtttcgata taaaagtaca ttttacttat     960

cgatagctcg tgctattgca taaaatgatg ttaggtggca acgctgcggc aacaatcagc    1020

tgtttaccag gccgcagcaa cgttacagtg cattttacat tttaccaagt tgaattaata    1080

aaattgcttt ttaaaagtgt ttactaaatt aaaaagccaa caaattgttg ttgttttcgt    1140

tgcttacaag cggctgctgt acataaatta tacatattag cgctaaacgt gctcaacatg    1200

aatttcctgc gccaaacatt caacgttacg aaacaattga cggcacaagg taagttttaa    1260

caaaaatccc tatttaaaac attgcgttgc ggctatttat tcaacttcga gtcccgtgtt    1320

ctatatacat acgcgcccac gcgcctaatt gccaaccatg tgaggcagcc ggtagccgct    1380

tgcgcacatc catttccaat tggtgactgt gcgcattttg tgtttatcca aggatcctgc    1440

gttccattgt gtgcacacaa tgatttgtat tgtctgttgt ttgcctgcga tctcaactct    1500

tttacatggg cgcgtggccg gcttgcgagc ctgtcgcccg tctgccagtt ctctagttgt    1560

cgtcgtaccc cccttccccc tgcccagccc cttatcgtgt gtctagtctg tgaatatttt    1620

tataagcatt ttctcatgtg tgtttcctgt ttgtgtgttt taatgtgtcc tcaaaactgt    1680

tcacggagcc tacaaagtgt gtattgagaa tatatatata tatatatagt ccatctgtcc    1740

atcttggata tttgtcattg gaacgggcga gcgaaaaaag ggtttgtcaa tgaaaaactt    1800
```

-continued

```
atcattttca ttatgtgcaa catttactaa ccaaatctat tcaatacata ggttggacaa    1860

acttgccttc tgttcttcga gataacttca gcaaagtctg ccaatgcgat ctgaagtcca    1920

ttcaattttt ggcctagcaa aaaacgcatt cgttttctg cttgttttaa ttaaaattca     1980

caacaaaatc cgcataacat gaggcccacc ctcaacaata ggaatttgca tgacatgcac    2040

acaataagga aaaacaaca caagaaaaaa aaattatgag aaaaggacac acacacacac     2100

acaggtgcgt tcggatcgcg gcagacaatg cacggagctg tgattggcat agttcttgct    2160

gtgcgcctgc tcccattgta agcgattgtc cagcgttatg gtaattatta cctgtgtgta    2220

cgtgtgtgtg tgtgtgtgtg gcatttaatt aaaaattgtt gtcgtttgcg attttggctg    2280

cagtacagtc gagtccagtc gggagtccag ctgaacagaa atctgagcat cagacagtca    2340

accccgtgc atggctaaag gttctcaatg cttaaaaggc ttgagaactg cagttgccgc    2400

tgacccacag ccgcgtcatt tggctgcaat tatttgtgaa aataaccta tatatgcatg     2460

atatgtggat ggatatggat ggatatatgg atgtgcgcag cataacaatt attttgcgat    2520

tttcacagag attagccaca acaaaaggcg aatggccatt gttgcttggg catttggaac    2580

tggccaactg tttctgaccc ttttgtcatg ttgtgtccgt tctctcgttt gtgtcaaatg    2640

tttttagccg ctcgctgcgg ctgcgctcac acatgcggca gcagctacca tatacaattt    2700

atataccaat atatgtacac atatttaatt ggtacagttg tgtccacttg cattgtatgt    2760

gtacacttaa cgcactcttg caattccgga caagtcaaga ggagacaact agacgttggc    2820

aatcggaaat tggaagcctt acagaaacac tgcgtttata acttgttctc agctgtttct    2880

ctctctcatc ttgattacat tgcagcgctg cagagcaatt atttgtgtgc cgcattgcgc    2940

ggcatggcat cgttgaatca aatgcatcgc actgggccgc atataaagaa gcgtccgcca    3000

cgtcagcccc tggacggtaa accgtttgcc aagggagtgg tgctcaagac actgatcaag    3060

aagccaaaga aaccaaactc ggcgaatcgt aaatgcgcgc tggtgcgctt atccacggga    3120

aaggagatgg tcgcctatat acccggcatt ggacataatc tgcaggagca taatattgta    3180

ctgtgtcgcg tcggacgact gcaggatgtg cccggcgtca agctgaaggc ggtgcgcggt    3240

gtctacgatc tggcgcacgt tatcaagaag ggccaatgac aaccaactac catgtaattc    3300

tcttccataa aaaacaaaa aaaaataag aaagaaaac aagccaaatc tttgagtact      3360

ctgctattct tgtgcagcat atattattat gattttttta atggaaaatt atgcagctca    3420

gcgggattag tgtaagtagc caacacacaa caagtgagct ctggcctcgc ctcatcccaa    3480

cttgtcttgc cgtaatctta agtcaacagg ccaaattgcg agccaacaa ttggccagtg     3540

ttgccaacga cgctgccgaa aaaggagcta aatcccattg gaaaatagct aaaaaatagc    3600

cagagcatga attggacgac tgaagacagc tgaaattggc cagaatttgg ccagaatata    3660

gctgatatcg caacactgca catttgttgc ccactcgaaa tatgatttta acggcatttt    3720

tacgctttag caggcaaatc ccttttgaa gaaacggccc gtctttactt tttaacagaa     3780

tttgcttgca caatttttat gccaaataat cgttaagcga aatgggcgtc gacaggccac    3840

gccacatttt cacagtacgc agagcctcgc cacgcctttg caacaggata caacaaattt    3900

tgaataggca ccgaccaggt tgtcgtgcat ctgggctgaa cgataacatc tgtattaaat    3960

caatcccata ttcaagcttc caaaggattt cggcaacatg ccaattccgc tattaatctt    4020

tcgggttatc ctgaatatct gggtagtcct aggtgtggag ctggtgtcgc ttatggcaga    4080
```

-continued

```
gcagctgaat gctaacatat acgagcataa aaagtttcat caggaatcca tatgcaactt    4140

gaggtccgac cc                                                        4152


<210> SEQ ID NO 42
<211> LENGTH: 12817
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: tko-step1-genbank

<400> SEQUENCE: 42

agcttgtacc caattcgccc tatagtgagt cgtattacaa ttcactggcc gtcgttttac      60

aacgtcgtga ctgggaaaac cctggcgtta cccaacttaa tcgccttgca gcacatcccc     120

ctttcgccag ctggcgtaat agcgaagagg cccgcaccga tcgcccttcc caacagttgc     180

gcagcctgaa tggcgaatgg gacgcgccct gtagcggcgc attaagcgcg gcgggtgtgg     240

tggttacgcg cagcgtgacc gctacacttg ccagcgccct agcgcccgct cctttcgctt     300

tcttcccttc ctttctcgcc acgttcgccg gctttccccg tcaagctcta atcggggggc     360

tccctttagg gttccgattt agtgctttac ggcacctcga ccccaaaaaa cttgattagg     420

gtgatggttc acgtagtggg ccatcgccct gatagacggt ttttcgccct ttgacgttgg     480

agtccacgtt ctttaatagt ggactcttgt tccaaactgg aacaacactc aaccctatct     540

cggtctattc ttttgattta taagggattt tgccgatttc ggcctattgg ttaaaaaatg     600

agctgattta acaaaaattt aacgcgaatt ttaacaaaat attaacgctt acaatttagg     660

tggcactttt cggggaaatg tgcgcggaac ccctatttgt ttatttttct aaatacattc     720

aaatatgtat ccgctcatga caataaccc tgataaatg cttcaataat attgaaaaag     780

gaagagtatg agtattcaac atttccgtgt cgcccttatt cccttttttg cggcattttg     840

ccttcctgtt tttgctcacc cagaaacgct ggtgaaagta aaagatgctg aagatcagtt     900

gggtgcacga gtgggttaca tcgaactgga tctcaacagc ggtaagatcc ttgagagttt     960

tcgccccgaa gaacgttttc caatgatgag cacttttaaa gttctgctat gtggcgcggt    1020

attatcccgt attgacgccg ggcaagagca actcggtcgc cgcatacact attctcagaa    1080

tgacttggtt gagtactcac cagtcacaga aaagcatctt acggatggca tgacagtaag    1140

agaattatgc agtgctgcca taaccatgag tgataacact gcggccaact tacttctgac    1200

aacgatcgga ggaccgaagg agctaaccgc tttttttgcac aacatggggg atcatgtaac    1260

tcgccttgat cgttgggaac cggagctgaa tgaagccata ccaaacgacg agcgtgacac    1320

cacgatgcct gtagcaatgg caacaacgtt gcgcaaacta ttaactggcg aactacttac    1380

tctagcttcc cggcaacaat aatagactg gatggaggcg ataaagttg caggaccact    1440

tctgcgctcg gcccttccgg ctggctggtt tattgctgat aaatctggag ccggtgagcg    1500

tgggtctcgc ggtatcattg cagcactggg gccagatggt aagccctccc gtatcgtagt    1560

tatctacacg acggggagtc aggcaactat ggatgaacga aatagacaga tcgctgagat    1620

aggtgcctca ctgattaagc attggtaact gtcagaccaa gtttactcat atatacttta    1680

gattgattta aaacttcatt tttaatttaa aaggatctag gtgaagatcc ttttttgataa    1740

tctcatgacc aaaatccctt aacgtgagtt ttcgttccac tgagcgtcag accccgtaga    1800

aaagatcaaa ggatcttctt gagatccttt ttttctgcgc gtaatctgct gcttgcaaac    1860
```

-continued

```
aaaaaaacca ccgctaccag cggtggtttg tttgccggat caagagctac caactcttt     1920

tccgaaggta actggcttca gcagagcgca gataccaaat actgtccttc tagtgtagcc     1980

gtagttaggc caccacttca agaactctgt agcaccgcct acatacctcg ctctgctaat     2040

cctgttacca gtggctgctg ccagtggcga taagtcgtgt cttaccgggt tggactcaag     2100

acgatagtta ccgataagg cgcagcggtc gggctgaacg gggggttcgt gcacacagcc     2160

cagcttggag cgaacgacct acaccgaact gagataccta cagcgtgagc tatgagaaag     2220

cgccacgctt cccgaaggga gaaaggcgga caggtatccg gtaagcggca gggtcggaac     2280

aggagagcgc acgagggagc ttccaggggg aaacgcctgg tatctttata gtcctgtcgg     2340

gtttcgccac ctctgacttg agcgtcgatt tttgtgatgc tcgtcagggg ggcggagcct     2400

atggaaaaac gccagcaacg cggcctttt acggttcctg gccttttgct ggccttttgc     2460

tcacatgttc tttcctgcgt tatcccctga ttctgtggat aaccgtatta ccgcctttga     2520

gtgagctgat accgctcgcc gcagccgaac gaccgagcgc agcgagtcag tgagcgagga     2580

agcggaagag cgcccaatac gcaaaccgcc tctccccgcg cgttggccga ttcattaatg     2640

cagctggcac gacaggtttc ccgactggaa agcgggcagt gagcgcaacg caattaatgt     2700

gagttagctc actcattagg cacccaggc tttacacttt atgcttccgg ctcgtatgtt     2760

gtgtggaatt gtgagcggat aacaatttca cacaggaaac agctatgacc atgattacgc     2820

caagcgcgca attaaccctc actaaaggga acaaaagctg gagctcctgc aggttgttgg     2880

ttggcacacc acaaatatac tgttgccgag cacaattgtc tagaatgcat acgcattaag     2940

cgaacattaa aaagatgtga aaacataact attatgtcta aataaacaca cgtcagatgt     3000

atgtacgtca acggaaaacc attgtctata tattacaatt actaaataca taccaaattg     3060

aatacatatt gatgaaaaat aataaatact ggcgaaagca aaaaacgaa acatttttat     3120

tttattgaac aactctcagg ctccaggtag gcaaaaaagc accgactcgg tgccactttt     3180

tcaagttgat aacggactag ccttatttca acttgctatg ctgtttccag catagctctg     3240

aaacctccag tctctggttg tgcacgacgt taaattgaaa ataggtctat atatacgaac     3300

tgagtctgga aaaagaagtt gagaattata aaaagtagtg agcactggcg ccctctctgc     3360

ttggcgagct aaccttttcg cctcttggct gagtaggtgg cgtttcattc tactctgtaa     3420

aattaatgta gaattgaaac actatggtca aaaaatactt aggggcataa gttataaaac     3480

gtatgaaatg aatttttatc aacctgggct attcaaaaat tttcgaatta ttttatgtat     3540

tttttttaat cgttttttcat tataggttaa aatacacttt aaaaggaatt ctttcctgta     3600

aaataaatat aaataaatat gctttattga cagaaaattt gatgtttttg tatttgagta     3660

ggagcaatca caggtgagca aaaaagaatt catcaattga tcggctaaat ggtatggcaa     3720

gaaaaggtat gcaatataat aatcttttat tgggtatgca acgaaaattt gtttcgtcaa     3780

cgtatgcaat attctttatt aaaagagggt atgcaatgta ttttattaaa aacgggtatg     3840

caatataata atcttttatt gggtatgcaa cgaaaatttg tttcgtcaaa gtatgcaata     3900

ttttttatta aaagagggta tgcaatgtat tttattaaaa acgggtatgc aataaaaaat     3960

tatttggttt ctctaaaaag tatgcagcac ttatttttg ataaggtatg caacaaaatt     4020

ttactttgcc gaaaatatgc aatgtttttg cgaataaatt caacgcacac ttattacgtg     4080

gccaactagg tgcccaaaat gtgtgtggac tacgaaattt ccaaatttta agatgctatc     4140
```

-continued

```
tttaaaccaa tgaaatatgg ttcgtatact atgaattttc aattaggcga acatcaatga    4200

ttcccccccc agaaaaccga catagcagag cacacgagca ggcgcaaatt gagaaaccca    4260

tccgcgtgaa gtcggttaat ttgcccatct tcttctggac gcgttcgtgc acccgctgct    4320

catccggcgg agtattgtac cattgtgtac ggccgtagtc cgtgtgcctt cgttttggcg    4380

ttcatgcatg agcagcccaa ttccttgctg ccccattcgg ttacattgca cagtggacac    4440

aaaagctagt tttgtagtca aagtacagaa ttcacaaatt atataaactg atatagttca    4500

tagatagtat aaactgatac caagtaacag atacacattt aaataggtaa actgtgtctg    4560

tgatcaaact gtttcttttc gtgtcgaaga atcaattaaa aatgattgaa tcattatatt    4620

tatttccgtt aaaagctgtg caggctgttc aaaatgtttt aatgaaaaaa tacgaatttt    4680

tagactgtct gaatcacagt gtgctcgctt acatttccgc tttcctcttt tggcaactcg    4740

atgtcgcctt tggggctctt ttggagaccg gaaaaaggca acattttcta ttcgattctt    4800

tttgccaatt gcccgagact gtgtcctgtc ggcatatgac gaatacgtaa cgtacgtgac    4860

ggcgacgtta ctcatacgca ccgtgggtag ctgcagacat ctcagcaccc acgttcgcga    4920

attattttga attcgactcc ctgggcgata tttgtttttc gcttttgcat attttgcggg    4980

caatttgggt aaaaggattt ccgcactctg cgacgccgtc ttcagtttgc ggctttcgtt    5040

ttttttcctag tagttcggca cacattttcc tcgccgcttc ggcaaatcgc tcacgtaaaa    5100

tatgcatgcg tttccttggc ggttttgcgc tctcaagtgc ctgcaattca attacatttc    5160

gattgatttt catgtttggc cccaaatcgc ggcaaaacct ctcactgacg gacacaccga    5220

agccccggcg gcaaccctca gcggatgccc cggggcttca cgttttccca ggtcagaagc    5280

ggttttcggg agtagtgccc caactggggt aacctttgag ttctctcagt tggggcgta    5340

gggtcgccga catgacacaa ggggttgtga ccggggtgga cacgtacgcg ggtgcttacg    5400

accgtcagtc gcgcgagcgc gactgctgca ggcgggcagc aaggcgtccc atccgcatta    5460

cgtgcccagc tatttgccag ctatgcccga tcctcatgcc tatattcgaa cgcccacgca    5520

caagcagccc gtaaccgaat acgaggcaat aagggaaaag gcagccagtc agaagcgtga    5580

cgttgagaag gcgctgacca aatttctgtg caaaacaaca gaaacaaaca atctctttcc    5640

caccgaggac aacatgtttc cgtgtaagta agcgctgcga ttaatggttc ttggttcttt    5700

attcaaatgt ttcgacttct ttttctgaat gcaacagtaa tcgcctgtaa gcccgccttt    5760

ccggcgtatg cagctgcctt gaatcccaca gatcaggtat ttgacttcga ggagctggag    5820

taccactact tggtggccaa tcgtacggaa gatgtgccca gtaaaggtag gtccaaattg    5880

tacacaatag atattccaat gaacacaggc tctactttca tttgcagagg agggcgagga    5940

gggtgacagt gagaatgagg aactggatgg cgacaagtcc aaggaggaga gcccgagct    6000

ggagatcaag cccaattcaa caacaaataa agctatttta gagaatccca atatagacaa    6060

tccctacttg cgtgccgcta cactgccaaa gcgttccaag ctgcacagtg agtgcactac    6120

accacgcatg gtgccctcac gaagtataca ctcggcttca cccacgacac cgacgccctc    6180

aactctagag ataaccaaaa gtagtgctta gttataatta taaatagatg cattgtaatt    6240

gtgtatagtt ttttaaaaaa aaaatattgg ataaacaaac tcttttcttc ttatcgatag    6300

ttcgtgcttt tgcttaaaat ggtgtgcgat ggcagcgctg cggcaacaaa cagctgtttc    6360

gatataaaag tacattttac ttatcgatag ctcgtgctat tgcataaaat gatgttaggt    6420
```

-continued

```
ggcaacgctg cggcaacaat cagctgttta ccaggccgca gcaacgttac agtgcatttt   6480

acattttacc aagttgaatt aataaaattg cttttaaaaa gtgtttacta aattaaaaag   6540

ccaacaaatt gttgttgttt tcgttgctta caagcggctg ctgtacataa attatacata   6600

ttagcgctaa acgtgctcaa catgaatttc ctgcgccaaa cattcaacgt tacgaaacaa   6660

ttgacggcac aaggtaagtt ttaacaaaaa tccctattta aaacattgcg ttgcggctat   6720

ttattcaact tcgagtcccg tgttctatat acatacgcgc ccacgcgcct aattgccaac   6780

catgtgaggc agccggtagc cgcttgcgca catccatttc caattggtga ctgtgcgcat   6840

tttgtgttta tccaaggatc ctgcgttcca ttgtgtgcac acaatgattt gtattgtctg   6900

ttgtttgcct gcgatctcaa ctcttttaca tgggcgcgtg gccggcttgc gagcctgtcg   6960

cccgtctgcc agttctctag ttgtcgtcgt acccccccttc ccctgccca gccccttatc   7020

gtgtgtctag tctgtgaata ttttttataag cattttctca tgtgtgtttc ctgtttgtgt   7080

gttttaatgt gtcctcaaaa ctgttcacgg agcctacaaa gtgtgtattg agaatatata   7140

tatatatata tagtccatct gtccatcttg gatatttgtc attggaacgg gcgagcgaaa   7200

aaagggtttg tcaatgaaaa acttatcatt ttcattatgt gcaacattta ctaaccaaat   7260

ctattcaata cataggttgg acaaacttgc cttctgttct tcgagataac ttcagcaaag   7320

tctgccaatg cgatctgaag tccattcaat ttttggccta gcaaaaaacg cattcgtttt   7380

tctgcttgtt ttaattaaaa ttcacaacaa aatccgcata acatgaggcc caccctcaac   7440

aataggaatt tgcatgacat gcacacaata aggaaaaaac aacacaagaa aaaaaaatta   7500

tgagaaaagg acacacacac acacacaggt gcgttcggat cgcggcagac aatgcacgga   7560

gctgtgattg gcatagttct tgctgtgcgc ctgctcccat tgtaagcgat tgtccagcgt   7620

tatggtaatt attacctgtg tgtacgtgtg tgtgtgtgtg tgtggcattt aattaaaaat   7680

tgttgtcgtt tgcgattttg gctgcagtac agtcgagtcc agtcgggagt ccagctgaac   7740

agaaatctga gcatcagaca gtcaaccccc gtgcatggct aaaggttctc aatgcttaaa   7800

aggcttgaga actgcagttg ccgctgaccc acagccgcgt catttggctg caattatttg   7860

tgaaaataac cttatatatg catgatatgt ggatggatat ggatggatat atggatgtgc   7920

gcagcataac aattattttg cgattttcac agagattagc cacaacaaaa ggcgaatggc   7980

cattgttgct tgggcatttg gaactggcca actgtttctg accctttttgt catgttgtgt   8040

ccgttctctc gtttgtgtca aatgtttta gccgctcgct gcggctgcgc tcacacatgc   8100

ggcagcagct accatataca atttatatac caatatatgt acacatattt aattggtaca   8160

gttgtgtcca cttgcattgt atgtgtacac ttaacgcact cttgcaattc cggacaagtc   8220

aagaggagac aactagacgt tggcaatcgg aaattggaag ccttacagaa acactgcgtt   8280

tataacttgt tctcagctgt ttctctctct catcttgatt acattgcagc gctgcagagc   8340

aattatttgt gtgccgcatt gcgcggcatg gcatcgttga atcaaatgca tcgcactggg   8400

ccgcatataa agaagcgtcc gccacgtcag cccctggacg gtaaaccgtt tgccaaggga   8460

gtggtgctca agacactgat caagaagcca aagaaaccaa actcggcgaa tcgtaaatgc   8520

gcgctggtgc gcttatccac gggaaaggag atggtcgcct atatacccgg cattggacat   8580

aatctgcagg agcataatat tgtactgtgt cgcgtcggac gactgcagga tgtgcccggc   8640

gtcaagctga aggcggtgcg cggtgtctac gatctggcgc acgttatcaa gaagggccaa   8700
```

-continued

```
tgacaaccaa ctaccatgta attctcttcc ataaaaaaac aaaaaaaaaa taagaaaaga   8760

aaacaagcca aatctttgag tactctgcta ttcttgtgca gcatatatta ttatgatttt   8820

tttaatggaa aattatgcag ctcagcggga ttagtgtaag tagccaacac acaacaagtg   8880

agctctggcc tcgcctcatc ccaacttgtc ttgccgtaat cttaagtcaa caggccaaat   8940

tgcgagccaa acaattggcc agtgttgcca acgacgctgc cgaaaaagga gctaaatccc   9000

attggaaaat agctaaaaaa tagccagagc atgaattgga cgactgaaga cagctgaaat   9060

tggccagaat ttggccagaa tatagctgat atcgcaacac tgcacatttg ttgcccactc   9120

gaaatatgat tttaacggca tttttacgct ttagcaggca aatccctttt tgaagaaacg   9180

gcccgtcttt acttttttaac agaatttgct tgcacaaatt ttatgccaaa taatcgttaa   9240

gcgaaatggg cgtcgacagg ccacgccaca tttttcacagt acgcagagcc tcgccacgcc   9300

tttgcaacag gatacaacaa attttgaata ggcaccgacc aggttgtcgt gcatctgggc   9360

tgaacgataa catctgtatt aaatcaatcc catattcaag cttccaaagg atttcggcaa   9420

catgccaatt ccgctattaa tctttcgggt tatcctgaat atctgggtag tcctaggtgt   9480

ggagctggtg tcgcttatgg cagagcagct gaatgctaac atatacgagc ataaaaagtt   9540

tcatcaggaa tccatatgca acttgaggtc cgaccccacc aacttttttg cactgcaaaa   9600

aaacacgctt ttgcacgcgg gcccatacat agtacaaact ctacgtttcg tagactattt   9660

tacataaata gtctacaccg ttgtatacgc tccaaataca ctaccacaca ttgaaccttt   9720

ttgcagtgca aaaaagtacg tgtcggcagt cacgtaggcc ggccttatcg ggtcgcgtcc   9780

tgtcacgtac gaatcacatt atcggaccgg acgagtgttg tcttatcgtg acaggacgcc   9840

agcttcctgt gttgctaacc gcagccggac gcaactcctt atcggaacag gacgcgcctc   9900

catatcagcc gcgcgttatc tcatgcgcgt gaccggacac gaggcgcccg tcccgcttat   9960

cgcgcctata aatacagccc gcaacgatct ggtaaacaca gttgaacaga tggtgagcaa  10020

gggcgaggag gtcatcaaag agttcatgcg cttcaaggtg cgcatggagg gctccatgaa  10080

cggccacgag ttcgagatcg agggcgaggg cgagggccgc ccctacgagg gcacccagac  10140

cgccaagctg aaggtgacca agggcggccc cctgcccttc gcctgggaca tcctgtcccc  10200

ccagttcatg tacggctcca aggcgtacgt gaagcacccc gccgacatcc ccgattacaa  10260

gaagctgtcc ttccccgagg gcttcaagtg ggagcgcgtg atgaacttcg aggacggcgg  10320

tctggtgacc gtgacccagg actcctccct gcaggacggc acgctgatct acaaggtgaa  10380

gatgcgcggc accaacttcc cccccgacgg ccccgtaatg cagaagaaga ccatgggctg  10440

ggaggcctcc accgagcgcc tgtacccccg cgacggcgtg ctgaagggcg agatccacca  10500

ggccctgaag ctgaaggacg gcggccacta cctggtggag ttcaagacca tctacatggc  10560

caagaagccc gtgcaactgc ccggctacta ctacgtggac accaagctgg acatcacctc  10620

ccacaacgag gactacacca tcgtggaaca gtacgagcgc tccgagggcc gccaccacct  10680

gttcctgggg catggcaccg gcagcaccgg cagcggcagc tccggcaccg cctcctccga  10740

ggacaacaac atggccgtta tcaaggaatt tatgcgcttc aaagttagga tggagggatc  10800

catgaacgga catgagttcg agatcgaggg agagggcgag ggacgcccgt atgaaggcac  10860

acaaacagcc aaactcaagg tcaccaaggg cggaccactg cccttcgcct gggatatcct  10920

gagtcccag tttatgtacg gcagcaaggc ctacgttaag caccccgctg acataccgga  10980
```

-continued

```
ctacaaaaag ctgtcctttc cggaaggctt caagtgggag cgcgtgatga atttcgaaga    11040

cggaggactg gtcactgtga cccaagatag cagtttgcag gacggtacac tgatctataa    11100

ggttaaaatg cgcggcacta actttccgcc agatggccca gtgatgcaga agaagaccat    11160

gggttgggag gcatccaccg aacgtctgta ccctcgagac ggagtgctca agggcgagat    11220

ccatcaggcc ctcaaactga aagatggtgg tcactacctg gtcgaattta agaccattta    11280

catggccaag aagccggttc agctgcccgg atattattat gtggatacga aactggatat    11340

aacttcgcat aacgaagact acaccattgt cgagcagtat gagcgcagcg aaggccgaca    11400

tcacctgttc ctctacggca tggacgagct gtacaagtag gcggccgcga ctctagatca    11460

taatcagcca taccacattt gtagaggttt tacttgcttt aaaaaacctc ccacacctcc    11520

ccctgaacct gaaacataaa atgaatgcaa ttgttgttgt taacttgttt attgcagctt    11580

ataatggtta caaataaagc aatagcatca caaatttcac aaataaagca ttttttttcac    11640

tgcattctag ttgtggtttg tccaaactca tcaatgtatc ttaaagcttc agtctctggt    11700

tgtgcatgga gcgtgttttt ggggcgattt acgagcgtga accgatattt ccatacgtc     11760

atgtttgcct tgctttacac tccaggaaac ctcatccgtt taatttagcc ggaagtgttg    11820

caatagatgc cacatcacaa tcgacttaat aattttttta ggagcaagtt ttaatggaaa    11880

cagtttctga taaataaata tacatatcaa cttagtacaa gaatatccag ctgaaagaat    11940

ggtatatata tatatatata tacttgtgtt tgtttctatg acagtttctt cacagctttc    12000

gatttcctta tggcacatcg cgcgacagtt gaaatgaaaa actgaaatca gtgaaacccc    12060

gaaaaaactg aaaagccac  ggaaattgta cagatacaca gatacgcaga tacaccagca    12120

aatgtacaaa aaggtctata tacatatata tgtagctact cacttatgtt gtccttcgca    12180

gattgctccc tttaagcaaa taaaaaaaag ttggctccac gccgaaaaga aataaaatta    12240

aatggagaat cgcaaatcca tagtgagcca aaggcaaatc tataaaagaa atgaaacgaa    12300

attcattcat tttcgttttc gagttcgaat atttaagtta tatatataaa cgcagtattt    12360

atccatgtaa tcgaaccaca aaagcccaat gagaaaaccc tacattttat gctgagcatc    12420

acaaaatgcc tttcctttca catgaattta tgtattttaa tcaatttccc tcgctgtggc    12480

agttaaatat cctaaatttg tccaacgaaa ttgatgcttc aattattcga atgacgacgt    12540

ttaatgggct ttcgaggaat aaaagcaaaa attcacaaga aaaacgcctc tgcatccatg    12600

ctcattatcg gaatcaatta aaatttcaca tgtatcgtta gcatggccat gtcagcaaat    12660

ccacgggatt cggctagagt cctccaaaat acgcccacgg gacccataca ccttcgaaat    12720

gatccaacat caatccctat ccaaatgtat acttagatat gtacatacct tgtcttttct    12780

tggtcggcga atggggggttc cagcatccca tcatcct                             12817
```

```
<210> SEQ ID NO 43
<211> LENGTH: 14148
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: tko-step2-genbank

<400> SEQUENCE: 43

atcatcgatc tcgaggctgc atccaacgcg cgcgttggga gctctccgga tcaattcggc         60

ttcaggtacc gtcgacgatg taggtcacgg tctcgaagcc gcggtgcggg tgccagggcg        120
```

-continued

```
tgcccttggg ctccccgggc gcgtactcca cctcacccat ctggtccatc atgatgaacg      180

ggtcgaggtg gcggtagttg atcccggcga acgcgcggcg caccgggaag ccctcgccct      240

cgaaaccgct gggcgcggtg gtcacggtga gcacgggacg tgcgacggcg tcggctgggt      300

gcggatacgc ggggcagcgt cagcgggttc tcgacggtca cggcgggcat gtcgacgaca      360

tgttcgcctc atttgtgttc gtttatgtat tcgatgttat gtgtatgctc atgtgatgtt      420

tagcttgtaa gcgcgagatg tgggtagcag gagatgcagt gcagccaaca gcagtgacca      480

gatgatatat gctatgctac tactactact tatatgctat gatttgtggc gcggaggcgt      540

gtctgcgaca cataatcccg cccatttagc tttaagattc aggcactaag aagcaattcg      600

atcaataaat tattgtaacc actctgcatg tgagcaaaag gccagcaaaa ggccaggaac      660

cgtaaaaagg ccgcgttgct ggcgtttttc cataggctcc gcccccctga cgagcatcac      720

aaaaatcgac gctcaagtca gaggtggcga acccgacag gactataaag ataccaggcg      780

tttccccctg gaagctccct cgtgcgctct cctgttccga ccctgccgct taccggatac      840

ctgtccgcct ttctcccttc gggaagcgtg gcgctttctc atagctcacg ctgtaggtat      900

ctcagttcgg tgtaggtcgt tcgctccaag ctgggctgtg tgcacgaacc ccccgttcag      960

cccgaccgct gcgccttatc cggtaactat cgtcttgagt ccaacccggt aagacacgac     1020

ttatcgccac tggcagcagc cactggtaac aggattagca gagcgaggta tgtaggcggt     1080

gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaagaac agtatttggt     1140

atctgcgctc tgctgaagcc agttaccttc ggaaaaagag ttggtagctc ttgatccggc     1200

aaacaaacca ccgctggtag cggtggtttt tttgtttgca agcagcagat tacgcgcaga     1260

aaaaaaggat ctcaagaaga tcctttgatc ttttctacgg ggtctgacgc tcagtggaac     1320

gaaaactcac gttaagggat tttggtcatg agattatcaa aaaggatctt cacctagatc     1380

ctttttaaatt aaaaatgaag ttttaaatca atctaaagta tatatgagta aacttggtct     1440

gacagttacc aatgcttaat cagtgaggca cctatctcag cgatctgtct atttcgttca     1500

tccatagttg cctgactccc cgtcgtgtag ataactacga tacgggaggg cttaccatct     1560

ggccccagtg ctgcaatgat accgcgagac ccacgctcac cggctccaga tttatcagca     1620

ataaaccagc cagccggaag ggccgagcgc agaagtggtc ctgcaacttt atccgcctcc     1680

atccagtcta ttaattgttg ccgggaagct agagtaagta gttcgccagt taatagtttg     1740

cgcaacgttg ttgccattgc tacaggcatc gtggtgtcac gctcgtcgtt tggtatggct     1800

tcattcagct ccggttccca acgatcaagg cgagttacat gatcccccat gttgtgcaaa     1860

aaagcggtta gctccttcgg tcctccgatc gttgtcagaa gtaagttggc cgcagtgtta     1920

tcactcatgg ttatggcagc actgcataat tctcttactg tcatgccatc cgtaagatgc     1980

ttttctgtga ctggtgagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccg     2040

agttgctctt gcccggcgtc aatacgggat aataccgcgc cacatagcag aactttaaaa     2100

gtgctcatca ttggaaaacg ttcttcgggg cgaaaactct caaggatctt accgctgttg     2160

agatccagtt cgatgtaacc cactcgtgca cccaactgat cttcagcatc ttttactttc     2220

accagcgttt ctgggtgagc aaaaacagga aggcaaaatg ccgcaaaaaa gggaataagg     2280

gcgacacgga aatgttgaat actcatactc ttcctttttc aatattattg aagcatttat     2340

cagggttatt gtctcatgag cggatacata tttgaatgta tttagaaaaa taaacaaata     2400
```

-continued

```
ggggttccgc gcacatttcc ccgaaaagtg ccacctgacg tcgacggatc gggagatcgg    2460

cgcgggatct aattcaatta gagactaatt caattagagc taattcaatt aggatccaag    2520

cttatcgatt tcgaaccctc gaccgccgga gtataaatag aggcgcttcg tctacggagc    2580

gacaattcaa ttcaaacaag caaagtgaac acgtcgctaa gcgaaagcta agcaaataaa    2640

caagcgcagc tgaacaagct aaacaatcgg ctcgagaccg tcgccacca tggtgagcaa    2700

gggcgaggag ctgttcaccg gggtggtgcc catcctggtc gagctggacg gcgacgtaaa    2760

cggccacaag ttcagcgtgt ccggcgaggg cgagggcgat gccacctacg gcaagctgac    2820

cctgaagttc atctgcacca ccggcaagct gcccgtgccc tggcccaccc tcgtgaccac    2880

cctgacctac ggcgtgcagt gcttcagccg ctaccccgac cacatgaagc agcacgactt    2940

cttcaagtcc gccatgcccg aaggctacgt ccaggagcgc accatcttct tcaaggacga    3000

cggcaactac aagacccgcg ccgaggtgaa gttcgagggc gacaccctgg tgaaccgcat    3060

cgagctgaag ggcatcgact tcaaggagga cggcaacatc ctggggcaca gctggagta    3120

caactacaac agccacaacg tctatatcat ggccgacaag cagaagaacg gcatcaaggt    3180

gaacttcaag atccgccaca acatcgagga cggcagcgtg cagctcgccg accactacca    3240

gcagaacacc cccatcggcg acggccccgt gctgctgccc gacaaccact acctgagcac    3300

ccagtccgcc ctgagcaaag accccaacga gaagcgcgat cacatggtcc tgctggagtt    3360

cgtgaccgcc gccgggatca ctctcggcat ggacgagctg tacaagtaaa gcggccgcga    3420

ctctagatca taatcagcca taccacattt gtagaggttt acttgctttt aaaaaacctc    3480

ccacacctcc ccctgaacct gaaacataaa atgaatgcaa ttgttgttgt taacttgttt    3540

attgcagctt ataatggtta caaataaagc aatagcatca caaatttcac aaataaagca    3600

tttttttcac tgcattctag ttgtggtttg tccaaactca tcaatgtatc ttagttgttg    3660

gttggcacac cacaaatata ctgttgccga gcacaattga tcggctaaat ggtatggcaa    3720

gaaaaggtat gcaatataat aatctttat tgggtatgca acgaaaattt gtttcgtcaa    3780

cgtatgcaat attctttatt aaaagagggt atgcaatgta ttttattaaa aacgggtatg    3840

caatataata tctttttatt gggtatgcaa cgaaaatttg tttcgtcaaa gtatgcaata    3900

ttttttatta aaagagggta tgcaatgtat tttattaaaa acgggtatgc aataaaaaat    3960

tatttggttt ctctaaaaag tatgcagcac ttatttttttg ataaggtatg caacaaaatt    4020

ttactttgcc gaaaatatgc aatgttttttg cgaataaatt caacgcacac ttattacgtg    4080

gccaacgcgc ctagtggatc cttcctggcc cttttcgaga aacgccgcga gggcgaaaag    4140

gattagttgt ttcaaacgca agaaggacat ttgtttcctt aaaattgtaac catttcttta    4200

tttggcactc gagccattga attttttcatt ttcagaatat gtgtacacat ttttttaaaaa    4260

aataaaaaaa ttatataatg ctggcggttg tttcatgtgt gaaaaattga tcaatggtaa    4320

acaaaattga ataaatatat aacatatata tatagatatg tgtgttgaaa tgaatacttg    4380

cgatacatgt aataaaaata ctcttcgctt atctatcaaa aagtgcggaa tgtcaaaatt    4440

taaaatttta caatgaatgc gtagccgacg acgaaagtgt tccttgctat ttcctttagc    4500

aagatttaaa tttagattaa attctaatga tacgattgac agttcgaaat tcaaagtgtt    4560

cctttttcaa aatttagtaa agattgtata tcaattgtag atatatcgaa attttttcggc    4620

cgcaagcgaa cattttacaa aatgaaggcg accagttgca gaccaattcc attcatcaac    4680
```

-continued

```
tttcggattg taagatattt ctatcggcca cgacgattga acaagtatta cgatattgta   4740

agtcttcttt aacaaaatta gtttcccttt cacagaaaca gacataaatt cttgaattat   4800

tgacttggat ttgagtgatc gttcgttgtc tatactataa gatctatagg cacgggataa   4860

cgctctaaat ctctttaaaa tcgaacgcgc caggcgctag ttaaacgtta ctatctatct   4920

ggttaaccca gctttgatcg gaatgcgtat atatatttca tgttatataa acgctgcaaa   4980

agctgccaga gcctctgctc cagagctgga ttcgctcaca ccttcctctt cttcttgggg   5040

tcagccctgc tgtctccacc gagctgagag aggtcgattc ttgtttcata gagccccgta   5100

attgactgat gaatcagtgt ggcgtccagg acctcctttg tagaggtgta ccgctttctg   5160

tctatggtgg tgtcgaagta cttgaaggct gcaggcgcgc ccaagttggt cagagtaaac   5220

aagtggataa tgttttctgc ctgctccctg atgggcttat ccctgtgctt attgtaagca   5280

gaaagcacct tatcgaggtt agcgtcggcg aggatcactc ttttggagaa ttcgcttatt   5340

tgctcgatga tctcatcaag gtagtgtttg tgttgttcca cgaacagctg cttctgctca   5400

ttatcttcgg gagacccttt gagcttttca tagtggctgg ccagatacaa gaaattaacg   5460

tatttagagg gcagtgccag ctcgttacct ttctgcagct cgcccgcact agcgagcatt   5520

cgttccggc cgtttt caag ctcaaagaga gagtacttgg gaagcttaat gatgaggtct   5580

tttttgacct cttttatatcc tttcgcctcg agaaagtcga tggggttttt ttcgaagctt   5640

gatcgctcca tgattgtgat gcccagcagt tccttgacgc ttttgagttt tttagacttc   5700

cctttctcca ctttggccac aaccagtaca ctgtaagcga ctgtaggaga atcgaatccg   5760

ccgtatttct tggggtccca atctttttttg cgtgcgatca gcttgtcgct gttccttttc   5820

gggaggatac tttccttgga gaagcctccg gtctgtactt cggtcttttt aacgatgttc   5880

acctgcggca tggacaggac cttccggact gtcgcgaaat ccctacccctt gtcccacacg   5940

atttctcctg tttctccgtt tgtttcgata agtggtcgct tccgaatctc tccattggcc   6000

agtgtaatct cggtcttgaa aaaattcata atattgctgt aaaagaagta cttagcggtg   6060

gccttgccta tttcctgctc agactttgcg atcattttcc taacatcgta cactttatag   6120

tctccgtaaa caaattcaga ttcaagcttg ggatattttt tgataagtgc agtgcctacc   6180

actgcattca ggtaggcatc atgcgcatgg tggtaattgt tgatctctct caccttataa   6240

aactgaaagt cctttctgaa atctgagacc agcttagact tcagagtaat aactttcacc   6300

tctcgaatca gtttgtcatt ttcatcgtac ttggtgttca tgcgtgaatc gagaatttgg   6360

gccacgtgct tggtgatctg gcgtgtctca acaagctgcc ttttgatgaa gccggcttta   6420

tccaactcag acaggccacc tcgttcagcc ttagtcagat tatcgaactt ccgttgtgtg   6480

atcagtttgg cgttcagcag ctgccgccaa taatttttca ttttcttgac aacttcttct   6540

gaggggacgt tatcactctt ccctctattt ttatcggatc ttgtcaacac tttattatca   6600

atagaatcat ctttgagaaa agactggggc acgatatgat ccacgtcgta gtcggagagc   6660

cgattgatgt ccagttcctg atccacgtac atgtccctgc cgttctgcag gtagtacagg   6720

tagagcttct cattctgaag ctgggtgttt tcaactgggt gttccttaag gatttgggac   6780

cccagttctt ttatacccctc ttcaatcctc ttcatccttt ccctactgtt cttctgtccc   6840

ttctgggtag tttggttctc tcgggccatc tcgataacga tattctcggg cttatgcctt   6900

cccattactt tgacgagttc atccacgacc ttaacggtct gcagtattcc cttttttgata   6960
```

-continued

```
gctgggctac ctgcaagatt agcgatgtgc tcgtgaagac tgtccccctg gccagaaact      7020

tgtgctttct ggatgtcctc cttaaaggtg agagagtcat catggatcaa ctgcatgaag      7080

ttccggttgg caaatccatc ggacttaaga aaatccagga ttgtctttcc actctgcttg      7140

tctcggatcc cattgatcag ttttcttgac agccgccccc atcctgtata tcggcgcctc      7200

ttgagctgtt tcatgacttt gtcgtcgaag agatgagcgt aagttttcaa gcgttcttca      7260

atcatctccc tatcttcaaa caacgtaagg gtgaggacaa tgtcctcaag aatgtcctcg      7320

ttctcctcat tgtccaggaa gtccttgtct ttaatgattt tcaggagatc gtgatacgtt      7380

cccagggatg cgttgaagcg atcctccact ccgctgattt caacagagtc gaaacattca      7440

atcttttttga aatagtcttc tttgagctgt ttcacggtaa ctttccggtt cgtcttgaag      7500

aggaggtcca cgatagcttt cttctgctct ccagacagga atgctggctt tctcatccct      7560

tctgtgacgt atttgacctt ggtgagctcg ttataaactg tgaagtactc gtacagcaga      7620

gagtgtttag gaagcacctt ttcgttaggc agattttat caaagttagt catcctttcg      7680

atgaaggact gggcagaggc cccccttatcc acgacttcct cgaagttcca gggagtgatg      7740

gtctcttctg atttgcgagt catccacgcg aatctggaat ttccccgggc gagggggcct      7800

acatagtagg gtatccgaaa tgtgaggatt ttctcaatct tttccctgtt atctttcaaa      7860

aaggggtaga aatcctcttg ccgcctgagg atagcgtgca gttcgcccag gtgaatctgg      7920

tgggggatgc ttccattgtc gaaagtgcgc tgtttgcgca acagatcttc tctgttaagc      7980

tttaccagca gctcctcggt gccgtccatt ttttccaaga tgggcttaat aaatttgtaa      8040

aattcctcct ggcttgctcc gccgtcaatg tatccggcgt agccattttt agactgatcg      8100

aagaaaattt ccttgtactt ctcaggcagt tgctgtctga caagggcctt cagcaaagtc      8160

aagtcttggt ggtgctcatc atagcgcttg atcatactag cgctcagcgg agctttggtg      8220

atctccgtgt tcactcgcag aatatcactc agcagaatgg cgtctgacag gttctttgcc      8280

gccaaaaaaa ggtctgcgta ctggtcgccg atctgggcca gcagattgtc gagatcatca      8340

tcgtaggtgt ctttgctcag ttgaagcttg gcatcttcgg ccaggtcgaa gttagattta      8400

aagttggggg tcagcccgag tgacagggcg ataagattac caaacaggcc gttcttcttc      8460

tccccaggga gctgtgcgat gaggttttcg agccgccggg atttggacag cctagcgctc      8520

aggattgctt tggcgtcaac tccggatgcg ttgatcgggt tctcttcgaa aagctgattg      8580

taagtctgaa ccagttggat aaagagtttg tcgacatcgc tgttgtctgg gttcaggtcc      8640

ccctcgatga ggaagtgtcc ccgaaatttg atcatatgcg ccagcgcgag atagatcaac      8700

cgcaagtcag cctatcagt actgtctaca agcttcttcc tcagatgata tatggttggg      8760

tacttttcat ggtacgccac ctcgtccacg atattgccaa agattgggtg gcgctcgtgc      8820

tttttatcct cctccaccaa aaaggactcc tccagcctat ggaagaaaga gtcatccacc      8880

ttagccatct cattactaaa gatctcctgc aggtagcaga tccgattctt tctgcgggta      8940

tatctgcgcc gtgctgttct tttgagccgc gtggcttcgg ccgtctcccc ggagtcgaac      9000

aggagggcgc caatgaggtt cttctttatg ctgtggcgat cggtattgcc cagaactttg      9060

aattttttgc tcggcacctt gtactcgtcc gtaatgacgg cccagccgac gctgtttgtg      9120

ccgatatcga gccaatgga gtacttcttg tccatggcga aaatccgggt cgaaagttac      9180

ggttatcgcg cactctactt tccacaaatc ctcacccaaa aaccaagcac agtttattca      9240
```

-continued

```
actgaagtat tcgcgatact tctttatcta ataataatgt acatgtaact aaactcgctt    9300

ttgggttaaa atcgtgacgc agaggcaaaa aaaatcgtat gtcccttaga caacttgaaa    9360

caactgcgaa gcgtacggca attccaggaa ttttgtggta aagctacgcg ccaactaacg    9420

gttcttgctt agaggtggaa taatgtagtt ttccagcgat aataaatata tcgatatttt    9480

tagtaaaatt gaaaaggtaa acttaatttt agaaatataat ttataagaaa tttaatagta    9540

tgcaaaataa ttttttacttg ctaagaatat gtgccactaa ttaaaagctg gacaccgcgc    9600

aatggaaaat agtactacaa cacagcaaca aagcctgagt tatcaacaaa aaaatacgaa    9660

aacatctccc aaaactaagc acccacacgc gccactcgcc gtcacaacac aatcactgca    9720

caccaccatt cgaatttcgc gcactgtgac aacatcacat gatatcggcg cggcaacatc    9780

ggattaccga caaaacgaac tatcgcacga gccaccgccg gcgaagagcg ctcgttttgc    9840

aacaccggcg cgcgctgaac gaagagaaca gctgactgct tgatacgtgc gtgtttcgcg    9900

gcaggaatta cataaagttt agagcctctg acgccagacc ccccgaacat tcgctccgat    9960

caaactacct gcgaacggtc acctaatccc caccatgcat ggtaggttac tctctgatccc   10020

ggtcatcact ggcgttcgct cacatccgtc cttacatgtg catatttcga ggttaaaacg   10080

gtcgaagctt ggatccgcta gcgttgttgg ttggcacacc acaaatatac tgttgccgag   10140

cacaattgat cggctaaatg gtatggcaag aaaaggtatg caatataata atcttttatt   10200

gggtatgcaa cgaaaatttg tttcgtcaac gtatgcaata ttctttatta aaagagggta   10260

tgcaatgtat tttattaaaa acgggtatgc aatataataa tcttttattg ggtatgcaac   10320

gaaaatttgt ttcgtcaaag tatgcaatat ttttttattaa aagagggtat gcaatgtatt   10380

ttattaaaaa cgggtatgca ataaaaaatt atttggtttc tctaaaaagt atgcagcact   10440

tattttttga taaggtatgc aacaaaattt tactttgccg aaaatatgca atgttttgc   10500

gaataaattc aacgcacact tattacgtgg ccaactagcc tagttccagt gaaatccaag   10560

cacttgaggt ccgacccgat gaattctttt ttgctcacct gtgattgctc ctactcaaat   10620

acaaaaacat caaattttct gtcaataaag catatttatt tatatttatt ttacaggaaa   10680

gaattccttt taaagtgtat tttaacctat aatgaaaaac gattaaaaaa aatacataaa   10740

ataattcgaa aatttttgaa tagcccaggt tgataaaaat tcattccata cgttttataa   10800

cttatgcccc taagtatttt ttgaccatag tgtttcaatt ctacattaat tttacagagt   10860

agaatgaaac gccacctact cagccaagag gcgaaaggt tagctcgcca agcagagagg   10920

gcgccagtgc tcactacttt ttataattct caacttcttt ttccagactc agttcgtata   10980

tatagaccta ttttcaattt aacgtcgctg cagcgatgcc attccagttt cagagctatg   11040

ctggaaacag catagcaagt tgaaataagg ctagtccgtt atcaacttga aaaagtggca   11100

ccgagtcggt gctttttttgc ctacctggag cctgagagtt gttcaataaa ataaaaatgt   11160

ttcgtttttt tgctttcgcc agtatttatt atttttcatc aatatgtatt caatttggta   11220

tgtatttagt aattgtaata tatagacaat ggttttccgt tgacgtacat acatctgacg   11280

tgtgtttatt tagacataat agttatgttt tcacatcttt ttaatgttcg cttaatgcgt   11340

atgcattcta gattttcaac gtcctcgata gtatagtggt tagtatcccc gcctgtcacg   11400

cgggagaccg gggttcaatt ccccgtcggg gagaatctgt gattcttttt tttttttcttt   11460

tactttgtta tataaacaat ttttgttta attgaatcta atttgccatt gcttttagga   11520
```

-continued

```
atctcaggca tccagcaagc gtttgtccgc cgaatcgccc atcagtgaag aagatcctgt   11580

ggcggctacg aaaatctccc cggccatgtc ggcctccacc tccagcgaaa aacccatcag   11640

cgagctggcc acctctgtgc tgacccaccg cttttccagac tccacctcct cacccggcga   11700

acatggcctt ggacgaatgc agttgtcgat ccgctacagc gcccagcgtc aaaaactaga   11760

cgtgaccata cacaaaatcc agaagatacc acttcgcgat cccagcaata tccccgatcc   11820

gtatgttaag ctgtatctgt tgcctggacg caccaaggag tcgaaacgca agacgagcgt   11880

gatcaaggac aactgcaacc ccgtctacga tgcatccttt gagtacctga tttccattgc   11940

cgaactcagg cagacggaac tggaggtgac ggtgtgcacc caaaagggat tcctatccgg   12000

cggtagtccc atcattggca tggtaggtac ccgaaagcaa ccccttagtt acagacacag   12060

cgcgtacgtc cttcgcatcc ttatgattcc caagtacata ttctgcaaga gtacagtata   12120

tataggaaag atatccgggt gaacttcgca ggacaacgcc cttggcgagt ttcagagcta   12180

tgctggaaac agcatagcaa gttgaaataa ggctagtccg ttatcaactt gaaaaagtgg   12240

caccgagtcg gtgctttttt gcctacctgg agcctgagag ttgttcaatc tagacaattg   12300

tgctcggcaa cagtatattt gtggtgtgcc aaccaacaac ctgcaggagc tccagctttt   12360

gttcccttta gtgagggtta attttttttg ctcacctgtg attgctccta ctcaaataca   12420

aaaacatcaa attttctgtc aataaagcat atttatttat atttatttta caggaaagaa   12480

ttccttttaa agtgtatttt aacctataat gaaaaacgat taaaaaaaat acataaaata   12540

attcgaaaat ttttgaatag cccaggttga taaaaattca tttcatacgt tttataactt   12600

atgcccctaa gtattttttg accatagtgt ttcaattcta cattaatttt acagagtaga   12660

atgaaacgcc acctactcag ccaagaggcg aaaaggttag ctcgccaagc agagagggcg   12720

ccagtgctca ctacttttta taattctcaa cttctttttc cagactcagt tcgtatatat   12780

agacctattt tcaatttaac gtcgcaacat tgtactgtgc cgcggtttca gagctatgct   12840

ggaaacagca tagcaagttg aaataaggct agtccgttat caacttgaaa aagtggcacc   12900

gagtcggtgc ttttttgcct acctggagcc tgagagttgt tcaataaaat aaaaatgttt   12960

cgttttttg ctttcgccag tatttattat ttttcatcaa tatgtattca atttggtatg   13020

tatttagtaa ttgtaatata tagacaatgg tttttccgttg acgtacatac atctgacgtg   13080

tgtttattta gacataatag ttatgttttc acatcttttt aatgttcgct taatgcgtat   13140

gcattctaga ttttcaacgt cctcgatagt atagtggtta gtatccccgc ctgtcacgcg   13200

ggagaccggg gttcaattcc ccgtcgggga gaatctgtga ttctttttttt ttttcttttta   13260

ctttgttata taaacaattt ttgtttttaat tgaatctaat ttgccattgc ttttaggaat   13320

ctcaggcatc cagcaagcgt ttgtccgccg aatcgcccat cagtgaagaa gatcctgtgg   13380

cggctacgaa aatctccccg gccatgtcgg cctccacctc cagcgaaaaa cccatcagcg   13440

agctggccac ctctgtgctg acccaccgct ttccagactc cacctcctca cccggcgaac   13500

atggccttgg acgaatgcag ttgtcgatcc gctacagcgc ccagcgtcaa aaactagacg   13560

tgaccataca caaaatccag aagataccac ttcgcgatcc cagcaatatc cccgatccgt   13620

atgttaagct gtatctgttg cctggacgca ccaaggagtc gaaacgcaag acgagcgtga   13680

tcaaggacaa ctgcaacccc gtctacgatg catcctttga gtacctgatt ccattgccg   13740

aactcaggca gacggaactg gaggtgacgg tgtgcaccca aaagggattc ctatccggcg   13800
```

```
gtagtcccat cattggcatg gtaggtaccc gaaagcaacc ccttagttac agacacagcg   13860

cgtacgtcct tcgcatcctt atgattccca agtacatatt ctgcaagagt acagtatata   13920

taggaaagat atccgggtga acttcgcacc agcacgcact ttcgatgttt cagagctatg   13980

ctggaaacag catagcaagt tgaaataagg ctagtccgtt atcaacttga aaaagtggca   14040

ccgagtcggt gcttttttgc ctacctggag cctgagagtt gttcaatcta gacaattgtg   14100

ctcggcaaca gtatatttgt ggtgtgccgt accgggccaa ttcgagct                14148
```

1. A vector, comprising:
a first gene encoding a DNA sequence modifying enzyme, wherein the DNA modifying enzyme modifies an endogenous copy of an essential gene;
a first promoter operably linked to the first gene encoding the DNA sequence modifying enzyme;
a second gene encoding a rescue transgene;
a second promoter operably linked to the rescue transgene; and
optionally, one or more cargo sequences;
wherein the vector is configured to be positioned in a chromosome or an extra-chromosomal element.

2. The vector of claim 1, wherein the DNA sequence modifying enzyme is a nuclease or a base editor.

3. The vector of claim 2, wherein the nuclease cleaves and generates one or more double strand breaks in the endogenous copy of the essential gene.

4. The vector of claim 3, wherein the one or more double strand breaks are repaired to create an altered sequence of the essential gene.

5. The vector of claim 2, wherein the base editor creates one or more base changes in endogenous copy of the essential gene to create an altered sequence of the essential gene.

6. The vector of claim 5, wherein the one or more base changes comprise one or more point mutations in the endogenous copy of the essential gene.

7. The vector of claim 1, wherein the rescue transgene is either a recoded copy of the essential gene or is a gene of unrelated sequence, wherein the rescue transgene encodes a protein that is functionally equivalent to a protein encoded by the essential gene, and wherein the DNA sequence modifying enzyme does not modify the rescue transgene.

8. The vector of claim 1, wherein the chromosome is an autosome, X chromosome, Y chromosome, or supernumerary chromosome.

9. The vector of claim 1, wherein the extra-chromosomal element is a plasmid or a virus.

10. The vector of claim 1, wherein the one or more cargo sequences comprise a one or more foreign gene sequences, or one or more alleles of an endogenous chromosomal or extra-chromosomal gene to which the vector has been linked through nearby insertion on the chromosome or extra-chromosomal element that carries the endogenous allele of interest.

11. The vector of claim 1, wherein the DNA sequence modifying enzyme is selected from the group consisting of Cas9, Cas-9-related RNA-guided nucleases, ZFN, TALEN, homing endonuclease, restriction enzymes, natural site-specific nucleases, engineered site-specific nucleases, base editing enzymes, cytidine deaminase, and adenine deaminase.

12. The vector of claim 1, wherein the vector further comprises one or more additional sequences, wherein the one or more additional sequences allow the vector to be positioned in the chromosome or the extra-chromosomal element.

13. The vector of claim 12, wherein the one or more additional sequences is selected from the group consisting of transposase binding site, LTRs, recombinase binding site, a sequence with homology to a desired location on the chromosome or the extra-chromosomal element.

14. The vector of claim 1, wherein the first promoter is selected from the group consisting of a germline promoter, a male specific germline promoter, a female specific germline promoter, a cell-type specific promoter, a tissue-specific promoter, a ubiquitous promoter, a promoter activated at a specific stage of mitosis, and a promoter activated at a specific stage of meiosis.

15. (canceled)

16. (canceled)

17. The vector of claim 1, wherein the nuclease comprises at least one nuclease domain and one or more DNA binding domains.

18. The vector of claim 11, wherein when the nuclease is Cas9 or a Cas9-related enzyme, the vector further comprise one or more genes encoding a guide RNA, wherein the guide RNA enables the nuclease to target specific sequences within the essential gene through Watson-Crick base pairing.

19. The vector of claim 11, wherein when the nuclease is Cas9, the nuclease domain of Cas9 is inactivated through one or more mutations, and the vector comprises a different nuclease domain.

20. The vector of claim 19, wherein the different nuclease domains is single chain variant of FokI.

21. The vector of claim 1, wherein when the DNA binding domain is a TALE, the nuclease domain is provided as a single active nuclease domain, such as single chain variants of FokI.

22. A method of modifying a population by a vector, the method comprising:
obtaining an organism of the population,
positioning the vector, configured to be positioned in at least one chromosome or extra-chromosomal element in the organism, comprising:
a first gene encoding a DNA sequence modifying enzyme, wherein the DNA modifying enzyme modifies an endogenous copy of an essential gene;
a first promoter operably linked to the first gene encoding the DNA sequence modifying enzyme;
a second gene encoding a rescue transgene;
a second promoter operably linked to the rescue transgene; and
optionally, one or more cargo sequences;

expressing the DNA sequence modifying enzyme in the organism,

inducing one or more sequence modifications in the essential gene in one or more cells in the organism, such that the one or more sequence modifications result in the essential gene being rendered partially or wholly non-functional and result in a defect in survival, growth control, fertility, or differentiation of the one or more cells if the one or more cells lack the rescue transgene,

rescuing the defects in survival, growth control, or differentiation of the one or more cells in which the essential gene has been rendered partially or wholly non-functional, by the rescue transgene,

generating an altered organism, wherein the altered organism carries one or more copies of the vector, and wherein the defects in survival, growth control, or differentiation of the one or more cells in which the essential gene has been rendered partially or wholly non-functional have been rescued the rescue transgene,

introducing the altered organism in an environment wherein an increase in a frequency of the altered organism is desired relative to a frequency of a wild type organism in the population;

replacing the wild type organism with the altered organism in the population in the environment wherein the altered organism is introduced,

thereby modifying the population.

23.-**48**. (canceled)

*   *   *   *   *