Described herein are embodiments relating to manipulation of populations and sex ratio in populations through DNA sequence modifications.

Specification includes a Sequence Listing.
Recognition site in essential gene

Endonuclease cuts recognition site

Repair/copying via HR, using HEG as template

Recoded essential gene

Cargo
FIG. 18A

68-tko-step1 (12817 bp)

tko-step2 (14148 bp)

Dvir-rescue frag (6000 bp)

Dvir-rescue-cds (426 bp)
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 a increase in the 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 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, Cas9-related RNA-guided nucleases, ZFN, TALEN, homing endonucleases, restriction enzymes, natural site-specific nucleases, engineered site-specific nucleases, base editing enzymes, cytidine deaminase, and adenosine 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 are selected from the group consisting of transposase binding site, LTRs, recombine 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
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.

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.

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.

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

In some embodiments of the method, the rescue transgene, 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.

In some embodiments of the method, the one or more cells comprise somatic cells, germine cells, gametes, or a combination thereof.

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.

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.

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.

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.

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

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.

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.

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.

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.

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.
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

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 recessed version of the essential gene that is resistant to cleavage by Cas9.

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.

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.

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%.

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

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.

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.

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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 germine into embryo.

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 germine 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.

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.

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

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.

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.

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

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.

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.

FIG. 14 shows a schematic of an embodiment of a Medea-based gene drive.
FIG. 15A–FIG. 15C show an embodiment of DNA sequence modification based gene drive (herein referred to as ClVir when a nuclease is used for DNA sequence modification) construct design and principle according to the present disclosure.

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

FIG. 17 shows an embodiment of an alignment of the target gene (Drosophila melanogaster tko—Examples 15 and 16) with the recorded rescue based on Drosophila virilis tko. Target exon in red, recorded 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.

FIG. 18 shows a schematic of an embodiment of the components of the DNA sequence modification-based gene drive (Example 17).

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

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

FIG. 19 shows an embodiment of the results of Sanger sequencing from Example 17.

DETAILED DESCRIPTION

In nature gene drive is brought about by a number of mechanisms, in a number of contexts (Ben-David et al. 2017; Bart 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.

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 generic 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.

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; Bart 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 (somewhat) 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 transgenes 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 et al. 2014).
et al. 2014). Two of these, UL39 (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 cleared 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, shedding 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 luck 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 system 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 nucleases 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”, “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 context 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 to 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 C1vR), 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 one 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
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.

In the CleveR 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 CleveR-Based Gene Drive

Without being limited by any particular theory, when the CleveR construct is present in an organism, wildtype copies of the essential are a risk for cleavage and inactivation. The individuals carrying CleveR 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 CleveR 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 CleveR 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.

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

A similar principle, cleavage associated with rescue of those who carry the CleveR vector, allows CleveR 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 CleveR element selects for those who carry it, and against those who fail to inherit it.

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

FIG. 15A-FIG. 15C show an embodiment of ClyR 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.

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.

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 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-EcoTSI, F-EcoTSII, F-EcoTSIV, F-Scel, F-TeVI, F-TeVII, I-AchMl, I-Anil, I-BasI, I-Bmol, I-Bthh0305I, I-BthII, I-BthORFAP, I-CluI, I-Chul, I-Cpal, I-CpalII, I-Crel, L-CsmI, L-Corn, L-DdII, L-Dmol, L-GpiI, L-GzeI, L-HjeMI, L-Hmnl, L-Hmol, L-IalI, L-ItrI, L-ItrWl, L-IpeMI, L-MsoI, L-Nanl, L-NfiI, L-Nqal, L-Omal, L-Pakl, L-PanMI, L-PnoMI, L-PogT71I, L-PorI, L-Ppol, L-Scl, L-SceI, L-SceII, L-SceIII, L-SceVIII, L-SpomI, L-SsmMI, L-Ssp6803I, L-TeVl, L-TeVII, L-TeVIII, L-Tsal, L-TsIWIAY76, L-Vdi1411, P-Aval, P-BeriP, P-HvoWI, P-MelEI, P-MruI, P-PkoI, P-PkoII, P-PspI, P-Scel, P-TfuI, P-TviI, P-TlII, P-TlIII, P-Tmlal, P-TmalaKl), 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 integrase.

[0107] In some embodiments, the nuclelease cleaves the endogenous copy of the essential gene. In some embodiments, the nuclelease generates one or more double strand breaks in the endogenous copy of the essential gene. In some embodiments, the nuclelease 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 nuclelease 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-hologous end joining (NHEJ), microhomology-mediated end joining (MMMEJ), 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 tautomersism, depurination, deamination, 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 recorded copy of the essential gene. In some embodiments, when the rescue transgene is a recorded copy of the essential gene, the protein encoded by the recorded copy of the essential gene (recorded 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 recorded 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 recorded copy of the essential gene (recorded 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 adenosine 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 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 dRed, GFP, EGFP, CFP, ECFP, BFP, EBFP, mHoneydew, m3Anna, mOrange, tdTomato, mTangerine, mStrawberry, mCherry, mGrape1, mGrape2, mRaspberry, mPlum, YFP or EYFP, and can be chosen by one of skill in the art according to need. Fluorescent marker protein can be visualized by illuminating with a suitable excitation 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 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.
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.

In some embodiments, when the nuclease is Cas9 or a Cas9-related enzyme, the vector further comprises 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.

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, 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

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.

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.

In some embodiments, the method comprises positioning the vector in at least one chromosome or extrachromosomal element in the organism. In some embodiments, the vector is any of the embodiments of the vectors provided herein.

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.

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.

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.

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.

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 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 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
in the particular environment. 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 the population in the particular environment. In some embodiments, the altered organism 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.

[0139] 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.

[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 deaminated 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 polyploid. Non-limiting example of polyploid 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, archaee), fungi, insect, mammal, rodent, fish, amphibian, reptile or plant. In some embodiments, any of the embodiments of the vectors and/or methods can be one or more of the following: Autographa gamma Silver Y moth Chilo suppressalis Asian 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 nematode Selarophthora rassiae var. zeae Brown stripe downy mildew Spodoptera liturata Egyptian cottonworm Spodoptera litura Cotton cutworm Thaumatotibia leucotreta False codling moth Anthonomus grandis Boll weevil Autographa gamma Silver Y moth Euxetranycys 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 Epiphas postvittana Light brown apple moth Lymandria dispar asiatica Asian gypsy moth Lymandria mathura Rosy moth Massicus raddei Mountain oak longhorned beetle Phytophthora quercina Oak decline Platypus quercivorus Oak ambrosia beetle Raffaelea quercivorae Japanese oak wilt Scytalidus intricatus European oak bark beetle Spodoptera littoralis Egyptian cottonworm Thaumatotibia leucotreta False codling moth Thaumetopoea processionea Oak processionary moth Tortrix viridana Green oak tortrix Trelmom xiescicornis Trelmom woodwasp Candidatus Phytoplasma pini Pine witches’ broom Cronartium flaccidum Scots pine blister rust Denstroconus micans Great spruce bark beetle Dendrolimus pini Pine-tree lappet Dendrolimus punctatus Masson pine moth Dendrolimus sibiricus Siberian silk moth Dipiron pini Pine sawfly Hylobius abietis Large pine weevil Lymandria mathura Rosy moth Monochamus saltuarius Japanese pine sawyer Monochamus sutor Small white-marmorated longhorned beetle Mecynorrhina girondii Needle blight of pine Panolis flammea Pine beauty moth Tomicus destruens No common name, a pine shoot beetle Autographa gamma Silver Y moth Cerneuellia virgate Maritime garden snail
Cochlicella spp. Exotic species Diabrotica speciosa Cucurbit beetle Helicoverpa armigera Old world bollworm Heteroderas filipjevi Cereal cyst nematode Heteroderas latipons Mediterranean cereal cyst nematode Heterorichia arator Black maize beetles Lobesia botrana European grapevine moth Meloidogyne arcellia British root-knot nematode Nysius huttoni Wheat bug Peronosclerospora philippinensis Philippine downy mildew Spodoptera littoralis Egyptian cotton worm Spodoptera litura Cotton cutworm Adoxophyes orana Summer fruit tortrix moth Aclecta vogelli 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 Eutetanychus orientalis Citrus brown mite Helicoverpa armigera Old world bollworm Spodoptera littoralis Egyptian cotton worm Adoxophyes orana Summer fruit tortrix moth Autographa gamma Silver Y moth Candidatus Phytoplasma australasianum Australian grapevine yellows Cryptoblas'm gniidiella Epiphera Honeydew moth postvittana Euopeudia ambigua Candidatus Phytoplasma vitis 1 Light brown apple moth Heterorichia arator Lobesia botrana Pseudopeziza trachiphila Spodoptera European grape berry moth Lithorilus Spodoptera litura Thaumatomatia leucotreta Flavescence doree Black maize beetle European grape vine Bursaphelenchus cocophilus Red ring nematode Candidatus Phytoplasma palmae Palm lethal yellowing Cocadivir.coconut cadang cadang Coconuts cadang cadang virus Darna palivitta Nettle caterpillar Hapliasipalus crudus American palm cixiid Metamastus hemipterus West Indian cane weevil Orcetes rhinoceros Coconuts rhinoceros beetle Paysandisia archon No common name, a palm borer Roberia indica Red palm mite Rhodacoccus obscurus New Guinea sugarcane weevil Rhyynchophorus ferrugineus Red palm weevil Rhyynchophorus palmarum South American palm weevil Autographa gamma Silver-Y moth Candidatus Phytoplasma australasiense 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 Neoleinlucodes eleganalis Tomato fruit borer Ralstonia solanacearum race 3 Bacterial wilt/Southern blight 2 Bacterial Wilt Spodoptera littoralis Egyptian cotton worm Spodoptera litura Cotton cutworm Synchysium endobioticum Potato wart Tectia solanivora Guatemalan potato tuber moth Thaumatomatia leucotreta False codling moth Tuta absoluta Tomato leaf miner Adoxophyes orana Summer fruit tortrix moth Amystria pinnatella Cherry blossom moth Actrocera zonata Peach fruit fly Candidatus Phytoplasma prunorum European stone fruit yellows Enormaria formosa Cherry bark tortrix Epiphas postvittana Light brown apple moth Grapholitha fuscana (Syn.: Plum fruit moth Cydia fuscana) Lecithocera mali folia Leaf blistet moth Lobesia botrana European grape vine moth Monilia polyspora Asiatie brown rot Monilia fructigena Brown rot, Apple brown rot Potyviruses Plum Pox Virus Plum pox Phytoplasma ecer a sandy European cherry fruit fly Thaumatomatia leucotreta False codling moth Globodera pallida Pale cyst nematode Globodera rostochiensis Golden nematode Heteroder a cajani Pigeonpea cyst nematode Heteroder a ciceri Chickpea cyst nematode Heteroder a filipjevi Cereal cyst nematode Heterodera latipons Mediterranean cereal cyst nematode Heterodera sacchari Sugarcane cyst nematode Punctodera chalcoensis Mexican corn cyst nematode Agriplus agrolatus Goldspot ocaked borer Agriplus biguttatus Oak splendid beetle Agriplus planipennis Emerald ash borer Anoplocnora chinensis Citrus longhorned beetle Anoplocnora glabripennis Asian longhorned beetle Chlorophorus annularis Bamboo borer Chlorophorus strobilicola Slender-bunched pine longhorn beetle Dendroctonus micans Great spruce bark beetle Ips sexdentatus Six-toothed bark beetle Ips typographus European spruce bark beetle Megaplas ma mutatus No common name, an ambrosia beetle Monochamus alternatus Japanese pine sawyer Monochamus saltuarius Japanese pine sawyer Monochasmus atertorn Small white-marmorated longhorned beetle Orthotomicus erosus Mediterranean pine engraver Pityogenes chalcographus Six-toothed spruce bark beetle Platypus quercivoros Oak ambrosia beetle Scolytus intricatus European oak bark beetle Tetroplum castaneum Black spruce beetle Tetroplum 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 Trichoderus canes Fris Desert longhorned beetle Trypodendron European hardwood ambrosia beetle domesticum Redbay ambrosia beetle Belo caulis spp. No common name, leafhopper slugs Cerrellia spp. No common name, hygromiid slugs Chouchicella spp. No common name, coccidei slugs Colossia spp. No common name, leafhopper slugs Laevicaudia spp. No common name, leafhopper slugs Lissachatina fulica Giant African snail Meghinatium pictum Chinese slug Monacha spp. No common name, hygromiid slugs Sarasinula spp. No common name, leafhopper slugs Semperula spp. No common name, leafhopper slugs Veronicella spp. No common name, leafhopper slugs Dendrolimus pini Pine-tree lappet Dendrolimus punctatus Masson pine moth Dendrolimus sibiricus Siberian silk moth Lymantria albescens Okinawan gypsy moth Lymantria dispar asiatica Asian gypsy moth Lymantria dispar japonica Japanese gypsy moth Lymantria malthar Rosy moth Lymantria monacha Nun moth Lyman tria posthila White-winged gypsy moth Lymantria umbrosa Hokkaido gypsy moth Lymantria sylina Casularina 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 Cochliomya hominivorax, for example, is a direct pest of cattle, and the spotted wing Drosophila, Drosophila suzuki, 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, dragonflies, dacefly, weevil, cricket, wasp, moth, beetles, honey bee, robberfly or butterfly. Additional examples of insects include, but are not limited to, Asian citrus psyllid (diaphorina citri), Australian sheep blowfly (Lucilia cuprina), Asian tiger mosquito (Aedes albopictus); Japanese beetle (Popillia japonica), White-fringed beetle (Graphimogaster spp.), Citrus blackfly (Aleucanthis voglioni), Oriental fruit fly (Dacus dorsalis), Olive fruit fly (Dacus oleae), tropical fruit fly (Dacus cirebitae), Dacus zonatus), Mediterranean fruit fly (Cer-
atitisa capitata), Natal fruit fly (Ceratitis rosa), Cherry fruit fly (Rhopalodera cerasi), Queensland fruit fly (Bactrocera tryoni), Caribbean fruit fly (Anastrepha suspensa), imported fire ants (Solenopsis richteri), Solenopsis invicta, Gypsy moth (Lymantria 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 hominivora), Old World Screwworm (Chrysonyza bezziana), Tsetse fly (Glossina spp.), Boll weevil (Anthonomus grandis), Damsel fly (Enallagma hageni), Dragonfly (Libellula luctuosa), and rice stem borer (Trypoziza incertulas). In some embodiments, the insects either transmit 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, insects refers to an insect that spreads a disease of economically important animals. In some embodiments, insects refers to an insect that spreads a disease of companion animals. In some embodiments, insects 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 germin 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 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.

[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. Pregenin 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 unreakable. 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,
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 whole 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 nucleate 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 nuclear change 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 autosomal (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 nucleate 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 nucleate. 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 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 viable, removing wild type alleles 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 wolfworm.

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 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 viable, removing a wild type allele 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 tolerate correspondingly reduced fitness costs as a result (FIG. 2C). 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 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 T 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 recorded 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 recorded version of the target gene (yellow rectangle with recombined 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 inviable, removing two wild type alleles 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, TX 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 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 recorded 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 recorded version of the target gene (yellow rectangle with recombined 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 TT, TT, TT, TWCC, TWCA, TWA, 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 inviable, removing two wild type alleles are from the population (FIG. 4B).

[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. 4C, as compared to FIG. 3C). Discrete generation, deterministic population frequency modeling of cleavage mediated 2-locus autosomal drive is shown in FIG. 4C. 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. 4C).

[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. 5A). 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. 5A).

[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. 5B). 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. 5B). 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 inviable, removing either a transgene and a cleaved locus (TC) or two wild type alleles (CH) from the population (FIG. 5B). 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. 5C). 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. 5C). Discrete generation, deterministic population frequency modeling of cleavage mediated haplolethal drive is shown in FIG. 5C. 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 haplolethal 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. 6A shows a chromosome (circle) carrying essential gene (green rectangle). In this example a prokaryotic chromosome carries a wildtype copy of an essential gene. FIG. 6B shows an extrachromosomal 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. 6C 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 as 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. 6D shows the chromosome (FIG. 6A) and the extra extrachromosomal element (FIG. 6B) in a cell and forced inheritance of the extrachromosomal element. The endogenous copy of essen-
tial gene is altered within the cell by CleaveR to render it non-functional (FIG. 6J). 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. 6F). An expanded view of the vector shown in (FIG. 6J). 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 between the two in the figure, though the actual arrangement between cargo, rescue, and DNA modifying enzyme can take a number of forms. FIG. 6J shows a cell carrying the wildtype chromosome and the extrachromosomal element including the vector. FIG. 6E 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. 6F 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.

Example 7

Example 9

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 (ooocytes) 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

FIG. 8A 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 not inherit CleaveR (FIG. 8A). 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. 8B shows a 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

FIG. 9A shows a schematic of an embodiment of a system to create 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 not inherit the construct (FIG. 9A). Only progeny that express the recoded protein survive. FIG. 9B 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

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 chromosomal that carry the vector have a selective advantage and increase in frequency. Such a system can also be used to guarantee that progeny 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

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

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. 8A, FIG. 9A), 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. 8A, FIG. 9A), 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 mechanism 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. 8A, FIG. 9A), 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. 15A 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. 15B shows an embodiment of a Construct B with an attP 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. 15C shows an embodiment of the principle of CleaveR. Females heterozygous for the CleaveR construct create cleaved tko alleles in the germline. Additionally, active Cas9/gRNA
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 germ line, all of the offspring will inherit Cas9 and gRNAs. Only offspring that carries the rescue encoded by CleaveR will survive (FIG. 15C).

Target Gene Selection and gRNA Design

[0221] Two versions of the CleaveR constructs were constructed using tko (technical knockout) as the target. The tko gene encodes a mitochondrial ribosome protein (Royden, Pirlotta, and Jan 1987). It is a recessive lethal. Benchmarking 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 CleaveR 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-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 cutesies 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 HindII and SpeI and the following 4 fragments were assembled in a Gibson reaction (Gibson et al. 2009) to yield plasmid p68-tdo-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 P94+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-tdo 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 synthetic complexity issue. Finally, a td-tomato marker (Shaner et al. 2004) driven by the ubiquitously opie2 promoter (Theilmann and Stewart 1992) with primers P15+P16 was used as the dominant marker.

[0225] Construct p68-tdo-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,Sh/TM6b,Tb balancer stock. Flies carrying the marker over TM3,Sh, 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 pU63-U61-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,Sh/TM6b,Tb as well as w[1118]. Flies carrying the GFP marker over TM3,Sh were pooled to generate the balanced stock and flies homozygous for the CiVR 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-tdo-step1; FIG. 15A), SEQ ID NO: 40 (tko-step2; FIG. 15B), and SEQ ID NO: 41 (Dvtd-rescue-modified; “rescue” in FIG. 15A and FIG. 17).
<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>SEQ ID NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>P0-68E FWD</td>
<td>gtcgTCCACAACCAGAAGACTGAG</td>
<td>1</td>
</tr>
<tr>
<td>P0-68E REV</td>
<td>aacctTCCACTCTCTGTTGTCGCA</td>
<td>2</td>
</tr>
<tr>
<td>P9-68E-hr-left FWD</td>
<td>cttattaagtcgacactagggctcaccctactgtggtgga</td>
<td>3</td>
</tr>
<tr>
<td>P10-68E-hr-left REV</td>
<td>GCTCGGTGTCGGCGCTGGTgagggtttgtgcgcctgattt</td>
<td>4</td>
</tr>
<tr>
<td>P11-attP FWD</td>
<td>aatcgcgcgaacacttctcactgcACCGACACACCCAGC</td>
<td>5</td>
</tr>
<tr>
<td>P12-attP REV</td>
<td>cttcgctggccgctgcagccAGTCGCGCTCCGCGCACTGCA</td>
<td>6</td>
</tr>
<tr>
<td>P13-dv-tko FWD</td>
<td>TCGTGGCGCGAGCGACGACTCgtcgcagggcgcagccag</td>
<td>7</td>
</tr>
<tr>
<td>P14-dv-tko REV</td>
<td>gcagtcggacaagttgtgttggtgggtcgagatctcagtttact</td>
<td>8</td>
</tr>
<tr>
<td>P15-68E-hr-right FWD</td>
<td>tggcaactctggcgccaccccacactctcttcttcagtcg</td>
<td>9</td>
</tr>
<tr>
<td>P16-68E-hr-right REV</td>
<td>ggggcgattcgytgccaggtcagggtggtggtggtgga</td>
<td>10</td>
</tr>
<tr>
<td>P21-tko-guidesA FWD</td>
<td>ctttttaacttaacgccgctgccgctgccgctgccgctgccgctccagt</td>
<td>11</td>
</tr>
<tr>
<td>P22-tko-guidesA REV</td>
<td>ttcACACCTAcgtctGAAAtccggcaagggcagctctcgCg</td>
<td>12</td>
</tr>
<tr>
<td>P23-tko-guidesB FWD</td>
<td>cttttttacttaacgccgctgccgctgccgctgccgctgccgctccagt</td>
<td>13</td>
</tr>
<tr>
<td>P24-tko-guidesB REV</td>
<td>ttcACACCTAcgtctGAAAtccggcaagggcagctctcgCg</td>
<td>14</td>
</tr>
<tr>
<td>P25-nosCas9 FWD</td>
<td>GGTGTCTATAACTATAAGTCTACTAGTACCGCACAGGAT</td>
<td>15</td>
</tr>
<tr>
<td>P26-nos-Cas9 REV</td>
<td>GCAATCGAAGTTTTCAGAAAAGATGGGTTATTCCGGAAATCT</td>
<td>16</td>
</tr>
<tr>
<td>P27-guidesA FWD</td>
<td>AGTTCAGTGAAAATCCAAGCtttttgtctcaacctgagtgc</td>
<td>17</td>
</tr>
<tr>
<td>P28-guidesA REV</td>
<td>aatcaccagtgcttcccaaggcagctcactacaaagggag</td>
<td>18</td>
</tr>
<tr>
<td>P29-guidesB FWD</td>
<td>cccctttgagtggttaaattttttttttgtcaacctgtgattt</td>
<td>19</td>
</tr>
<tr>
<td>P30-guidesB REV</td>
<td>gcagctgctgccgatcgatgtcctgcagcaccatattctcag</td>
<td>20</td>
</tr>
<tr>
<td>tko-seq1</td>
<td>aagcgcttccaaaggtgcacag</td>
<td>21</td>
</tr>
<tr>
<td>tko-seq2</td>
<td>cgccatccatctctacttg</td>
<td>22</td>
</tr>
<tr>
<td>tko-seq3</td>
<td>cccacacacagttggctcc</td>
<td>23</td>
</tr>
<tr>
<td>tko-seq4</td>
<td>acaactgacgttggccactTCACACCTTCTTCCTTCTTT</td>
<td>24</td>
</tr>
<tr>
<td>tko-seq5</td>
<td>tccgcppaggttagtgtgattg</td>
<td>25</td>
</tr>
<tr>
<td>tko-seq6</td>
<td>catcgcaactctggctgccccccctg</td>
<td>26</td>
</tr>
<tr>
<td>s2-attB-rev</td>
<td>ttggagaactgctgacatc</td>
<td>27</td>
</tr>
<tr>
<td>s2-u631-seq</td>
<td>AGTTCAGTGAAAATCCAAGGCG</td>
<td>28</td>
</tr>
<tr>
<td>T3-seq REV</td>
<td>gtccccctttagtgaggggttaatt</td>
<td>29</td>
</tr>
<tr>
<td>T3-seq FWD</td>
<td>ATTACCCCTACTAAAGGGAC</td>
<td>30</td>
</tr>
<tr>
<td>CAS91F</td>
<td>ATGGAGAAAGAGTACTTCATAGC</td>
<td>31</td>
</tr>
<tr>
<td>CAS91R</td>
<td>GATCCGTATTGCCGCAAGACT</td>
<td>32</td>
</tr>
<tr>
<td>CAS92F2</td>
<td>AGGCCTAGGGCTGCTGCTAATCG</td>
<td>33</td>
</tr>
</tbody>
</table>
TABLE 1 -continued

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>SEQ ID NO.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAS93F</td>
<td>GGAATAATCTCTACATTTTGG</td>
<td>34</td>
</tr>
<tr>
<td>CAS94F</td>
<td>AGGATGGAAGACAGACTCTTG</td>
<td>35</td>
</tr>
<tr>
<td>CAS95F</td>
<td>CTGACGCCAACAGTATCAC</td>
<td>36</td>
</tr>
<tr>
<td>CAS96F</td>
<td>TOGAACCCACACCTGATCAT</td>
<td>37</td>
</tr>
<tr>
<td>CAS96R</td>
<td>TCAACCTTCCCTCTTCTTCTT</td>
<td>38</td>
</tr>
</tbody>
</table>

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-tld-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-tld-tomato dominant marker, showing that the system works efficiently when transmitted through females (see FIG. 9A and FIG. 16A), 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 (tko5/FM7aDpt1; 2;Y)w+, BDSC_4233. 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. 16B).

[0232] For FIG. 16A, 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. 16A) 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. 16B, 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 (tko5), only those animals that also inherit the rescue encoded by the CleaveR element survived, all others died. Actual data is shown in the Punnett 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. 18A 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. 18B (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. 18C (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: δ ClvR<sup>tko5</sup> off+ offspring from δ ClvR<sup>tko5</sup>+XX δ 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


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


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

<400> SEQUENCE: 2

aadctccag tctctgtgtgc gcag
<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

cttattagtg ggcxaactag gtgcacaaaa tgtgtgtgga
<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

gctcggtggt gtcgtaagaagaggttttg gcggcgatttt
<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

aaatcgccggc aacacctcttc actgacggac aaccaagac c
<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

cctttggtgcc cgcttcgactc agtcgctc gcggactga
gtgagcactg
<210> SEQ ID NO 7
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
-continued

<223> OTHER INFORMATION: P13-dv-tko FWD
<400> SEQUENCE: 7

tcaacctgc gacgcgac ggtgcaggc ggcagcaagg
<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

gcaacctgaa aagtttggt gggcgagacc tcaagttgca tattg
<210> SEQ ID NO 9
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P15-60E-hr-right FWD

<400> SEQUENCE: 9

tgcaaacctga ggcggagacc cacaacttt ttgcaacctg
<210> SEQ ID NO 10
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P16-60E-hr-right REV

<400> SEQUENCE: 10

gggcsaattg ggtcaagcct aagatgatgg gatgttgga
<210> SEQ ID NO 11
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P21-tko-guide1A FWD

<400> SEQUENCE: 11

catttccag tttacgctg cgtcagcgat gccatccag ttccagcgt atgttgga
<210> SEQ ID NO 12
<211> LENGTH: 60
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P22-tko-guide1A REV

<400> SEQUENCE: 12

ttcgcgacg gcgcagcaag cgtggtcctg gsaagttcag cggatatct
<210> SEQ ID NO 13
<211> LENGTH: 60
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P23-tko-guide2B FWD

<400> SEQUENCE: 13
catttccaa ttaacgtcg caacattgta ctgtgcccg gtttcaagtc tatgctggaa

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

<400> SEQUENCE: 14
tttccgacata gctctgaaac atcgaagagtcttgctggtg cgaaagtccccaggtatct

<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 tatagcagcg ggataacgtc

<210> SEQ ID NO 16
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: P25-nos-Cas9 REV

<400> SEQUENCE: 16
gacaatccagc gtgagcagaagacgtggat ttcgctggaa ctggtgctttttgcctttttgctactgtgatt gc

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

<400> SEQUENCE: 17
agttccagtg aatcccaagc ttttttgctc acctggtatt gc

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

<400> SEQUENCE: 18
aattccaggt gacaaaaaa aatcaacctc cactaaagg g a

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

<400> SEQUENCE: 19
ccctttgtag agggttaatt ttttttgctc acctggtatt

<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
goagctcga gatgatgat tgccgagc ac 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 agtgcacag 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 tttcccaattg 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 ggtgcggttc 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
acacagagc gttgcaact tcacaccttc cttctctctct 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
tcagcgggt ttagttaagt 20

<210> SEQ ID NO 26
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
OTHER INFORMATION: tko-seq6
SEQUENCE: 26
catatgcacacctgaggtccg 20

SEQ ID NO: 27
LENGTH: 20
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: s2-at8-rev

SEQUENCE: 27
ttcgagaccctgacctacat 20

SEQ ID NO: 28
LENGTH: 20
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: s2-u63l-seq

SEQUENCE: 28
agttccagtgaaatccaagc 20

SEQ ID NO: 29
LENGTH: 23
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: T3-seq REV

SEQUENCE: 29
gtcccctttaaggggttaatt 23

SEQ ID NO: 30
LENGTH: 20
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: T3-seq FWD

SEQUENCE: 30
atacccttcactaagggas 20

SEQ ID NO: 31
LENGTH: 22
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: CAS91F

SEQUENCE: 31
atggacaagaagtaactcattg 22

SEQ ID NO: 32
LENGTH: 20
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: CAS91R

SEQUENCE: 32
-continued

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

<400> SEQUENCE: 33
agcgttaggc tgtccaaatc

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

<400> SEQUENCE: 34
gagaaaaatc tcacatttcg g

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

<400> SEQUENCE: 35
agactgagaa gacaatctg g

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

<400> SEQUENCE: 36
cgacgcgca aactgatcag

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

<400> SEQUENCE: 37	
tggacgcac acggatctcat

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

<400> SEQUENCE: 38
tccacacctt ccctctctct t

<210> SEQ ID NO 39
agctgtgacc caaatgcgcc tataaggtg cgtatcaaca ttcaactgccc gtctgttttac 60
aacgtcctga ctgggaaac ccctgccctta tgccttgcga gcacatacc 120
cctggccag ctggcgtaat agcagaagg gcccagcaga tgcgcctccc caacagttgc 180
gcagccgaa tggcgaatgg gaagccgcct gtacgggccc attaagcgcg aggggttgtg 240
tgttgaaggg ccagcgtacc cctacacttg cgacgccgct aagcgccgccct tttcctcgtt 300
tttttccttc ctgggctccc acgtgtgccc gttttccccgg tcagcaacta aatggggggc 360
ttcttgggt gttgctattt acgtgttac gggaacctga ccccaaaaaa cttgattaag 420
gtattggtgg acgtatgctgg ccatacgccc gctagacggt tttgcgcctgt ttagcgtg 480
agttcgcagtt cttaaagtgt gacatcttt gcaccacaccc tccaaactgg aacaacacotaactccattct 540
cggttcattc tttaaggtttta cagccagtta ccaagctttgc ggcctatttg gttaaaaaattg 600
agttgatta aacaataattt aacgagatt tttacaaata ttaacagcct aaaaattcagt 660
tggcaactttt cggggaagtg tggcgggaac cccattttgt tttttttcct attacaccc 720
aatatatgtg cggctctagaa gacaataact cttgataaag tctcaaatattt attaaag 780
gaacaggattc atgtttctggt cgccttatt tccccccctttt cggcattttggtt 840
cctgtaggt tttgtgaccc caggaacgtt cgtgaagta aagatgtcgg aagatcagtt 900
ggtgcaacga cttgggatta tcacactgga ttcgacaagc gtgtaagcctt ttagaagttt 960
tgcgcaacca gacgcttttca cactgtgag coctttttaaa gttctgttac tgggagcgtt 1020
aatcagcgtg atgaagcgcc ggcaagacca aactggtgccc gcatacaact atttctcagaa 1080
tgacttggtg gtaactcgac cagtcacaga aaggcatctt aacggatggca tgaagtaaag 1140
agatattcg acgttgcacca taacaccagt gtaaactag ggcgcacatc taacctgac 1200
aacagacgga ggacggcaggg agtcaggccc tttttgcac acaatgaggg gacatgtac 1260
tgctcttga cgtgagggac ccagacgaa gagaacgttg tgaagacata ccaaacacgc aacgtgaaac 1320
caccggtatg ctagaggtg ccacaaacgtt ggcgaactcata taacttgccg aacticatcc 1380
ttagctccc ccgacccaaat taataagctgt gatggagggc gataaagttct caggcaactt 1440
tgtcgctctgc ggccttccgg cttgggcgttct tttcacgtag aatactggag cgggtacgcg 1500
ttggtcgctcg ggtatcaactg cgcacactgg ggccagatgg aagccctccc gtcctgtag 1560
taatcagcag acgggaggct cagccacttag gtaagacgga aatagacaca ctagctcagat 1620
aggtgctctcaa ctttagaagc attggttaact gtcgacacaa gttcatcaca atatacctta 1680
gatgtgatta aacaactttt ttaatattaata aaggtttagtg gtgaagatcct tttttgatas 1740
tttcatgaccc aaatgcctttt aacgtgaggct ttcggcttcag tggcgcagga aacccgactga 1800
aaagatcaaa gatctttcttg gatctcctttt tttttgcgc gtaaatctgct gcgacgacac 1860
aaaaaaccacc ggcgacccgg cttgggtcttt gccgbggagt ggagcgctac ccaaatctttttt 1920
tccgaagttg atcgggtctc ggcagagcga gataccaaat atcgctccgg tagtgcagcgg 1980
gtaggttagg ccaccaactca agaactctgtg agccagcctc acetactcctg ctctgtcaat 2040
cctgttacca tggcctgctg ccagttgcga taaggctgtg cttaacccggt tggactcaag 2100
agatatgta cggataaag gcagagcggtc gggtcgaag gggggsttngt gcaacacagc 2160
cagctttgag gcagaagact aacccgaact gatataacca cagctgtgag tattgaaga 2220
cgccagcgtt ccggaagagga gaaagccgga caggtactcg gtaaagcgca ggtgcggca 2280
agagagcgcc acaggggagcg tccacggggg aaacccgcgtg tatctttatata gttcctgcttgg 2340
gttgccccag ctctgcctctg agcgtgcatt tttgtgtgtgc tcctcagggg gcgggagcct 2400
atggaaaaac gcaacgcaacg ccggcctttc agcgttccttg gctcccttttc ggccttttgc 2460
tccactgttc ctcctgctgat tatactctgg ttaagcctgat aaccgtatcctt cccgctttga 2520
gtgacgagtt cccgctgccc gcaccccgaac gacgcagcgcc agcagtcag tcagcggag 2580
agcgagaagcg cggcactatttac gcaacccgcc tttcccccgcc cgttgccgca ttccttgttac 2640
caggccgacag ccaggctttcc ccaactgggaa agcggccagt gacgcaaacg caatctttagt 2700
gattgaggca actcatgggg cccacccgagc cttaccccttt atgtccttctg cttgcgctgt 2760
gtgacagatt gtggagcctt acacaattcct acagggaaac acgtatcgtc acatgattcg 2820
cacgcggga atttacccctt actaaacgctt ccaaaaaacgattagcttgc agccttttgg 2880
ttgccacactt aacaataact tggccgagcc ccaatccagtatgatactagcacgagtaat 2940
cgaacacttta aaataagttta aaaaaataact ttatacgcttat aataacaaacgc gtcgagatgt 3000
atgtaagctca agcggaaaaac atttgtctata cattacaact actaaatccta tcccaaatgt 3060
aatacataact gtauacaggta aaataacaaactt ccacacacgtc aaaaaacgcagc acatcttttat 3120
ctttttccgt acltctccacct gcctagggca gcaaaaaaagcgacctttc atctccttctg 3180
tgaactgtgat acagcaagtact tctatattct tattgtcttaga aaaaacagagt 3240
aaacccctcag cctctctggttc ctgcaagctg taatgctttattatatc attacacacc 3300
tgatgcttgga aaaaagaggt gagaattaata aaaaagactg agcaatggcgc cctctctgctg 3360
ttgccagtct atcctctcggt cctctctggt ctgtaggtgc gctttcattct tacctgttta 3420
aataaactgta gaaatgaaac acatgtgtca aaaaaaactact ccagggctata ctataacaa 3480
gtatgaaaat aatattaatc acctggcctt ctctgttattt aatcttagtt 3540
ttttttttaaag ctgtttctat cttatttgtt aaatatccttta aaaaagactt cttctctgta 3600
aaaataaat aatataaat ctatatttga cgaagaaattt gatgtttattg tatttgaga 3660
gggaacacact caggtgacga aaaaagaaatc atcataattgt ctgctaaat atgtatggc 3720
gaaagagtatt gcataataaat aatcttcttct tgggtatcga acaaaaaattt tttccgtcag 3780
gcttccgat attctttctttaa aaaaagaggtt atgcaattata tttaatatat aacgggtgatt 3840
cattataa aactttttttt ctggtcaggt gcggaaatttct ttccgtcaga gtaacacca 3900
tttttttttaaagatggttga ctcagattgttat attatattaaagcgctgctgctgataa 3960
tatttgcagct ctataaacaag tataagcagc acattttttc ataagattagc aacacaaattt 4020
ttttagcttga gaaataatac cattttttttt ggaataaattat ccaacgcaac aatatctctg 4080
gcgaacttag ctgccccacagt gttttggtga ctaaaattttatc tccaaaaattttt ggattatc 4140
ttttacccaa gtaaatattt ccctgtatatct ataaaaatctta aattacggtc gataattta 4200	ttttttttttt gtaaatattctct gttcctattct aaaaatatgttta aaatatccgca 4260
tctctgtggta gtcctgtatt tttgccacgt tctcttgagg gcctcctgtgc acgcttctc 4320
catcggggtcg agtatgtagc cattgtgtac gcgcggtagtc cggtgcccttg ctgttggcgc 4380
ttcattgtgg agaagcccaag ttcctggcttg cccacgctgg taccattgca cagtgacg 4440
aaaaatgtgt tttgtagcga aagtacagaa ttcacaaatt atataacttg atataaggtca 4500
ttagatttag aactagtaac cagataatac atataacttg atataaggtca 4560
tgatcactac gttctttcttc gttgctgaaga atccattcag atagattgaa tccttatatt 4620	tatatatcat aataaggtgtg cagcggcttc aaaaagttttt aatggaaaaa tagaatatc 4680

tagatgttcat gatacagacgt gttgctgtgcct agatactccg tgtctctcttt ttggcagctcg 4740
agtgcgtcgt tggcgcttttt cttcagagcc gaaaagggca acaatattta ttcagattttt 4800
tttgccaaagtg ggcgagacgt gttgcgtttgc ggcgagctgc gatactgtag ccgctagcag 4860
gggtcgattc ctcagacagc cggtgtaggt ctcgagcact ctcagcagcgg aagttgctgg 4920
atttatattgt atccagctgcct tgtctctctg tttgtttgttg gttttgttac atttgctgg 4980
caatgccgtt ataaaggatttt cctccgctcct gcagccacggt cttcgatttt gcgttgctttt 5040
	ttttctctag tgtggtggca cacactttcc tggcgcottt ggcaggat aggctgtgaa 5100	tatccttgttg ttcctgtctg cttccagtitc ctgcgaaattt cattcaacatc 5160
gatttgattgg catggtgtgag cccaaactgcgg gggaaacact ctcagtagcgc gaggactcga 5220
agccggcggc gccacccctcag gcggctgccc cggggtcttca gtttttttccca gtttaagcag 5280
gttctgtggtttt cggcctagcgg cgcgggggttag acaatcggag tttcctccg gttttgggtg 5340
gggtgactgcag cagctgagcg ggtctatggg cggggttggtt ccacgtacgg ggtgtcttacg 5400
aacggctactgc gcggagcccc gcagctgacgg cgggctagtgg gaaagctttg caaggactgta 5460
cgggctgctgg ctggtggtggt cttcagcgcc cagctgactg cttcagcagc cgtctggtct 5520
cagaggggtt cgtctggtgc cttcaggcaat aacgttaga cgggagtactg cggggctaaat 5580
cgggtgaag gctggtgccac atactttgtgg gaaaaaacc cggcagcttc gagaagttgt 5640
caccggagcc acaatgttgg ccggtgatag aagcgtgctga ttaactgttt ttgtgttttt 5700
attcaacatg tgtgctttttct tgttgtataa gcgcagttga tagctgtttg gctggttttt 5760
cgggttagat cagctgtgctt gaccagcagg atcagctgtt ggagtcggag 5820
taccatcact tgggtgcttg ccgtagcgaa gatggtgcttg ctaaagttta gtagcaaatg 5880	taccaatattc atattctttcat gaaacgcgaggc tcctaaacgc tttgtccaggg aggccagag 5940
gggggagct ggaacagagga aactgtgtag cagcagcggc aagggagaga aggccagag 6000
ggagactag cccacacagg ccaacaatgg aagttaattt gagaacacca atataagcaaac 6060
tccoatcaggt ccgcgccagt gcggccacgg cgggccagag ctggcctgctg agttgcctga 6120
accagctagt ggccggccat gaaattactag atcggctgca ccacagcagc caggccccgg 6180
aacacttacg atatacagaaa atagttgtact gttataaatg tataaagctg ccattttt 6240
gttgttagt tttttttttttt aataattgg aaaaaaattttt cttgttctttc ctttgctatg 6300
tctggttttttt gtttgcggtc ggccaggctcg cggctacaaa cagctggtctt 6360
gatataagaat cttctttttttt tttgtcacag cgggtcatttt tttcttggttgg ttcatttgtgg 6420
ggcaacgtgctggcagcag cagctgttta ccagcgcgcc gcacgcgtttc agtgcatttt 6480
acatttttactgcacgtttt tttttttttttt gtttctttta ccacagacac aattaagaag 6540
cacacaaattt gtttggtttt ctgtgtgttt caagccgctgt ctgtacataa attatacaat 6600
<table>
<thead>
<tr>
<th>Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>ttacctcagc tattcagtc ggtgacttc</td>
<td>6660</td>
</tr>
<tr>
<td>ttgctctcagctgatgactcttctgctgtgtgt</td>
<td>6720</td>
</tr>
<tr>
<td>tttactcactctgactctgactcactgactcactgactcactg</td>
<td>6780</td>
</tr>
<tr>
<td>cagctcactctgactctgactcactgactcactgactcactg</td>
<td>6840</td>
</tr>
<tr>
<td>ggtgacttcgctcactgactcactgactcactgactcactg</td>
<td>6900</td>
</tr>
<tr>
<td>ggtgacttcgctcactgactcactgactcactgactcactg</td>
<td>6960</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>7020</td>
</tr>
<tr>
<td>ggtgacttcgctcactgactcactgactcactgactcactg</td>
<td>7080</td>
</tr>
<tr>
<td>ggtgacttcgctcactgactcactgactcactgactcactg</td>
<td>7140</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>7200</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>7260</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>7320</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>7380</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>7440</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>7500</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>7560</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>7620</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>7680</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>7740</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>7800</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>7860</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>7920</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>7980</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>8040</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>8100</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>8160</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>8220</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>8280</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>8340</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>8400</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>8460</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>8520</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>8580</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>8640</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>8700</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>8760</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>8820</td>
</tr>
<tr>
<td>cagctcactgactcactgactcactgactcactgactcactg</td>
<td>8880</td>
</tr>
<tr>
<td>Sequence</td>
<td>Start Position</td>
</tr>
<tr>
<td>----------</td>
<td>---------------</td>
</tr>
<tr>
<td>agctctgcct cgcctcata ccaacctgtc ttgccgtaat ctaatgctaa caggccaaat</td>
<td>9940</td>
</tr>
<tr>
<td>tgcgaacca acaattgagc agttggtgca acaagctggc cgaaaaagga gttaaatccc</td>
<td>9000</td>
</tr>
<tr>
<td>atgggaataa aagttaaaaa tagccagagc atgaatttapa cagactgaaga cagctgaat</td>
<td>9060</td>
</tr>
<tr>
<td>ttgcacagaa ttggcgagaa tatagctgtat atcgcaaaaa tgcacatcgg tggccaccc</td>
<td>9120</td>
</tr>
<tr>
<td>gaaatagtat tttaaagcggc tttaaacgctt ttgccggacct aatctcccgtt tgaagaacag</td>
<td>9180</td>
</tr>
<tr>
<td>gcctgtcttt tcttaatttc gcagatagtc attctaat ctatattcatt tgaagcctttt</td>
<td>9240</td>
</tr>
<tr>
<td>ggcaaatgagc ctggagacagc ccagcccaaa ttctacagtt aagcagagcc tgcgcacggc</td>
<td>9300</td>
</tr>
<tr>
<td>ttgcaacagc gatacaacaac attttgaata ggccacgacc aggttagtga gccttgggagc</td>
<td>9360</td>
</tr>
<tr>
<td>tgaacgataa cattctgtatt aaatacatac catatctcaag cttccaaaggt atttccggca</td>
<td>9420</td>
</tr>
<tr>
<td>ctgcacatt cgcctattac tctttctggtc tattctgtaat atctgggtag tctcagttct</td>
<td>9480</td>
</tr>
<tr>
<td>gggctgtcgtc tgcctttcag cagacgacag cagatgttaag aataacgagc ataaaggttc</td>
<td>9540</td>
</tr>
<tr>
<td>tcatccagaa tcaatactcga acttggagtc gcaccccaac acctttttgt acctgcaaaaa</td>
<td>9600</td>
</tr>
<tr>
<td>aaacgaagcct ttgcaacgagc gcccatacag agtacacaaat ctacaggtct cagctctttt</td>
<td>9660</td>
</tr>
<tr>
<td>tacatataa aagcctagag cgcctctagc tttctacaag tccaaaaata caccaacaca ttgaacctttt</td>
<td>9720</td>
</tr>
<tr>
<td>ttgcagtgca aaaaaagtacg tggcgacgcc aagctgtgctt ctattctggtg aagcagacgcc</td>
<td>9780</td>
</tr>
<tr>
<td>ttgcctagtc aagcatgacagc atgggacggc acagaggtgtt tttctagctg acagcagcc</td>
<td>9840</td>
</tr>
<tr>
<td>agttctctct ctccagagac gcacccctcg gccttgggtt atgggaaccag gcagacggcc</td>
<td>9900</td>
</tr>
<tr>
<td>catcagcaag gcgctttcag tctcctggtc gcagccggac ccagcgcagt tacgcgttat</td>
<td>9960</td>
</tr>
<tr>
<td>cogcagatata aacacgacgg ccaagctatc gttaaacaag tggtgaagcag ttcgagccaa</td>
<td>10020</td>
</tr>
<tr>
<td>ggccgagagc tgcataaaga agttctgacag ctctcattgg gcgcttggag cggctcctag</td>
<td>10080</td>
</tr>
<tr>
<td>cgcacagcag ccagctagcgg caggggcggc cgcctccacagt ggcacacagc gcagacggagt</td>
<td>10140</td>
</tr>
<tr>
<td>ggcaagatcg aggtggacca aagggggggc cctgctttc gcctgggacca cctgctcttc</td>
<td>10200</td>
</tr>
<tr>
<td>coatgcctag ttaacgcaga cggcgacagt aagccacccc gcgcaactcc cctgcttttc</td>
<td>10260</td>
</tr>
<tr>
<td>gaaggtgtcc ttcoccgagc gttcagagtg ggaagcggtgg atgaactcttg acaggccgagc</td>
<td>10320</td>
</tr>
<tr>
<td>ttctggtacg ttgcagagcc actacctctt gcagcagcagc acggctgatct cagaagatggag</td>
<td>10380</td>
</tr>
<tr>
<td>gatggcgagcc acaacacgcc cccgccaggg ccgggttaatg cggagagagtc tttccttcag</td>
<td>10440</td>
</tr>
<tr>
<td>ggaggctccc acggagcgcc tgtacccccc cgagggcggtg ctggagaggg gatcccaacs</td>
<td>10500</td>
</tr>
<tr>
<td>ggccctgtag ctaagagagc cggccacaact cctggtgag ttcagagcaaa ttcataacggc</td>
<td>10560</td>
</tr>
<tr>
<td>caaagaacccc tcgaactcgc cccggtactaa tctagggag accaatctggg aacaactcctcc</td>
<td>10620</td>
</tr>
<tr>
<td>ccacacagcag gcacccacac cttatgaagca tttggcggcgc gcacaccacag</td>
<td>10680</td>
</tr>
<tr>
<td>gttctcggag atggagcaag gcacagcagc cagggcgcag tccggccagc cctctcctcag</td>
<td>10740</td>
</tr>
<tr>
<td>ggcaacacag atgaggctgta tcaagagatg aagagttgag gaggccagg atgaagagcaac</td>
<td>10800</td>
</tr>
<tr>
<td>catgagagcgg ctatcgcttc gagaactggc aaggggctgg gagacgggcgt atgggagccac</td>
<td>10860</td>
</tr>
<tr>
<td>acaaaagcgc aaacttaaag ggcaacaggg cgaggcaactg cttcttcgct gcctgtagct</td>
<td>10920</td>
</tr>
<tr>
<td>gatgccccag tttataagcg gcagcagggc tacgttaag acctccggcgc gccatacggcgc</td>
<td>10980</td>
</tr>
<tr>
<td>ctcaaaagag cggtcccttt tggtggactg aagctgtaga atttggcagc</td>
<td>11040</td>
</tr>
<tr>
<td>cggaggctcgt gtactgtgtc cccaaagtag cagtttggag gacgcgtaga tggctgctaaa</td>
<td>11100</td>
</tr>
<tr>
<td>ggttaaatgc cgcggcacaact tttctcggcc agatgtggca gttgactcga agaagccact</td>
<td>11160</td>
</tr>
</tbody>
</table>
-continued

gcgatacgc ggccagcgt cagcgggttc tcgaagtcga cgaggggcac gcgcacgaca 360
tgttgccttc atttgggttc ggtagaatg attgaattg gcaggtgtcctaat 420
tagtagtaa ggcggcagag cgggaagcag gcatgccgct cgcggcgcac 480
ggattttgct cattgagatt gagcgggagt ggtgaacatc gagcgagctc 540
gcggcaacataaaccggcccccattagattctgaaggctaatagcccataacac 600
cggctgatc tattttgaa gwcaagggagt gccgcggaag ccacgaaacc 660

cgccgaggggagccggtttttcattgcttcccgcgccccctgtgcagcatc 720
aaaatatgacgtaatgctgcgaagggcagcaacgctatgataaaatcagggcag 780
	tcgagcctgggaggtgtgttctctcccgtcctgctgccgtacattcatt 840
tgctgcgtccttcgatttgcattgtttcgagttaatagtttgatctggtgcg 900
tcaggattggcctgattatcctgtgcggtgacatgctttgattttggccgctg 960
tccgactgctgagccgctgcttctgttcctcattcattagaaaaataagc 1020
ttttaaattcgttttttttttttttttttaagagattctattctttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
cagtctaatc cttctttaaa tccgaacgcc cagggcgctag tttaaggctta ctatctcttc 4920
gatatacct gagttgtagc gaatgctat atatattttca ttttatataa aagotgcaaa 4980
agtcgcaaga gctctgctgc cagactgtag ttgctgacaca cttctctctt cttctctgggg 5040
tcagcctcag tgcctcagcag acgtgagatc tttgcttcatagacgacga 5100
attgaaggtg tatacaggtgg ggcgctcagg acctccttgtagaggtgtaga cgcctctcttg 5160
tctttgctct gttgcaagat cttgcaaggt gctgggctgct ccaagttggc cagagtaaac 5220
aagtttattata tggtttctggc ctctgctctg atggcttat tccctgtgcctt atgtaaggca 5280
gaaagacgcct tatacaggtg acgatggcgg gagaataactc tttttgagga aatccctatt 5340
tgctgatagag tttcttatcag ctgtagtgta gttgttctaca cgcagacgtct tttttgctca 5400
tatcttctggg gacagcttcag tcgggtggg agcagttatat cgggttaaag 5460
tttgtgacatt gtttctttca ctcgggctct cttgagtttg acagtcgacag 5520
cytttctgccag ctcttctctcg ctggctcagc acgacgcact agcagatagt 5580
tttttctgag cccttggtgac ctcgctcagc ctttctctttt ctttctctttt 5640
gttgctgaact ttgatgctag gaccccgtctg ctttgctctctc tttgctctctc 5700
cotttccact tttgtggccat aaacctgctca ctgtaagctt ctgtgagttt acgatagt 5760
cgttattcag cttttttcactc atttttttttc aagttttttt gttttttttt 5820
ggcgtaatac tttcttttttg gccgaacagt cttttttctct ctgtagtgctg 5880
acactggctaga tttttttgtgc ctgtagtgctg gttttttttt 5940
attttctctg tttttttctct gcgtgtagag cagctgcagct atctgctctc atcagttttt 6000
gagtaactct ttttcgctatc aaaaatcacta atctgtagttt aaaaagaggtt ctcagtttt 6060
gcttctctct gttttttttc cattttttg aatctttttc tttttttttttt 6120
tcttgctgaa tccttttttg atgggtttttt aacacagtttt ctcagtttttt 6180
acacaggatac ccttttctttttt cttttttttttt 6240
aactgaacctt ctatcactatt ttttctgctgtt cttttttttttt 6300
tctcggcata gcgtgcgagt ttttctctgtc tttttttttttt 6360
gcacgcgggt cttcatgctag cttcagcagct gttcgcctgag ggtatttttt 6420
tttttttttttt tttttttttttt 6480
tttttttttttt tttttttttttt 6540
gaggggggtgt tttttttttttt tttttttttttt 6600
atcactgttct ctttcttttttt cttttttttttt 6660
cagtttaggtg atcactgttct cttttttttttt 6720
tagacgcttc cctcttttttt cttttttttttt 6780
cctttttttttt cttttttttttt 6840
tttttttttttt cttttttttttt 6900
cctttttttttt cttttttttttt 6960
gacgggtttc ccttcttttttt cttttttttttt 7020
tttttttttttt cttttttttttt 7080
tttttttttttt cttttttttttt 7140
-continued

ttcggtgtcc cattgatcag ttttcttgac agccccccac atccgtgata tcggtgcctc 7200
ttgagcaact ttatcaacct gctggaagaag agatgagcgt aagttttaaa gogttcttca 7260
atcactctccc tatcttccaa caagaatgag gtagagcaca acgtctccag aatgtcctcg 7320
attccctct ccgctcgagga gttcaagct ttaatagtt tcggagatc gtgatacgtt 7380
ccaggcttgg cgttgaaggg atcctccact ccgcttcttt caacagatgc gaaauatca 7440
acattttttga aataacccct tcggagttgc ttcccgggttc gtcacctggg 7500
agaggctcca cgatagcctt ctctctgtct cccacaggg aagctcgggttt tctctctctt 7560
ctttgatcgct atttgaacctt ggctgagctcg tttaaaacctg tcaagaatact ctacacgcaga 7620
gagttttttta gagaaaccttt ttcgcttcgg agatgtgttttat caagatgttg cattctttcgg 7680
atgaaggaact ggcggagcgc cccctcttccc aacgaacctt cgaagtcttcca ggaggtgttg 7740
gtttccctcttg attgctcagt ctactcaagaa aatctggaat ttcocccggg gagggggcct 7800
acatagtgggc atacgccgaa ctggtgagttt ttcocctcct tttgtctgtgt atctttccaa 7860
aaggggttggg aactctcttt ccgctggagg aatcgcgtgca gtcgctccag agtacattg 7920
tgggggtttc tttcgtgttc gaaagcgcgc tgcgtcgcag cagagtctttc tctgtttttgg 7980
tttccccgcg ccgctgtgctt gctcccagttt tttcccaaga taggctttatt aaatgtttga 8040
aatttcctct gcgtcttgcoc gctggttctat tctcagccgt agccatctttc agactgtacc 8100
aagaaaaattt ctggctcttt cctgcagggg tgcggctgca gtcgctccag caggaatgctc 8160
aatctcgttt gcgtctcgat atgcgctttg atcataactag ctgcgcagcgc agctttgtttg 8220
atcctcgtct ttcactcgcag aataactaag aagacagaggt ccgctgaaag cggtagttttc 8280
goaaacaaaaa gccgggctgtgtt ctgtgctgcgt ctggcggcag cacaggtagt gacagaca 8340
tgtgtaggtct ttttctctctg tgggctccgtt cccgctgggaa tcttttttttg 8400
aagtgggggg cccgcggccg ccgggctgct tccaagagtc aaagaaaatc caaatcagggc gttctctctt 8460
tccacgggag cctgtgctctt gagttttttg aagcgcgggg atttgagcag cctacgcgct 8520
agaggttgtg ggctgtttgct cccggttgggt ctctggctctt gacttttttg 8580
taagtcggtta caaggtgttg tgcgctttgg tggctgctgcgccc gttggcggcag gttaccagcag 8640
cocctgtgag aagaaattgg cggctgaggt cggctgaaag cagagcagtag atagacagcc 8700
cyccagtcag ccttcagcttt aatgctcttc gcagctgatt atagttttttg 8760
tactttttcg gataagccca ctctccataag attggcggaa agatggggtgg ggtgctgctg 8820
ttttatcct cctccaccaaa aaaaaatcct tccagcctat ggaaagagga gtcattccacc 8880
ttctccctg caatttcgct ctattccact cagttgctctg cccgtcgggta 8940
tacctgcag gcgtctgcctt ttttggcggc gcgtctctct tgggttgccgg 9000
agaggggggc caaatgaagtt cttccttattg ggtgctgggg ggattggtgg 9060
aattttttttg cgcggccattagctgcct gtaatgcagc ccacgcagag gctttttttg 9120
cgctatagca gccccaggt ttttcccttt tccgctgggg cagggcaggg acaatcagcttt 9180
gggtatgcg cactctttct ccccaacctt caacgcagcag acgttttacatc 9240
acttgaagtt ctgcgcattc tttttttact ataattaagt acatgtaaatt ccaatctcgg 9300
ttggtttaaa atctgtgagc agagcctaaa aaactcgtcg gtctccttga caacttggaa 9360
cacgctgcgag ccgtacgcgc atccagggaa ttcttggtta aagctcgcgc ccagctaatgg 9420
gtcttgctt agaggtcga taatgtagtt ttccagcgt atataaatata tcgatatatt 9480
tagtaaatgt gaaagaagat tatcaagattttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
ctggaacag cacagcaagt tgaacaaggg ctggtcctgt tttacatcgtga aaaaagctg 14040
cogagtcggt ggtttttttgc ttacgtgagag cctgaggatt gttcaatota gcaaatgtg 14100
cctgggcaca gtagatgttg ggtgtgcaag aacggggccaa ttggacctg 14148

<216> SEQ ID NO 41
<217> LENGTH: 4152
<218> ORGANISM: Artificial
<220> OTHER INFORMATION: Dvir-rescue-modified
<400> SEQUENCE: 41

gctgcaagcgg gcagcagcaag gcctcactsc gcattaagct gcactactatt tgcagctat 60
gcctcgatct cattccataa ttcgaaagcc caagcacaag gcagcctgaa cgaataac 120
ggaccaagaa gaaaagcgag caagtcagaa gcagtcagtt gagaaggcgg gcaccaaat 180
tttggtcata acacacagaa aacatcactct tcttccacc gggaccaacag tctttctg 240
taagaagctc ctgcgtttaaa tggcctttggt tttctatttc aatgttttctg aactccttttt 300
tgaatgcacca cgaagtttgcg tctgtaagcc gcctttccgg gttgtgcagc tgcctgtta 360
cacgggact gcattttctga cttcaggagc gttggagact actacctttgt ggcacacag 420
acgccagatg gcgccagtca aggtagtgcac aatagttcaacta atctacgatt a 480
acgcggccta tttcttattgc cagagggcgg cgagggagtg gacagttcag atgaggaatt 540
gatggcgagc acgttaccaag gcggagcagc atcaacgcaca attaccac 600
aatcataagactg atctctagag atcccaaatag aacaaatcctct cccctctgtg cgtcatac 660
accaacagtt ctaaacggtgc acaggtgtggt caactacaac agctcagttgc cctcccagag 720
tatccactgc gttctccagca gcgcacccgg gcctctatctct tgtacagatt ccacaatgag 780
tgtgttagtta taattataar tagtggtcatt gtaattgtgt atagttttttt aaaaaaaa 840
attgctataa ccaacactct tttctttttcat gctatggttc tgctttttatt taaaatttgt 900
tgctagcttc ggctgtgggg aaacacaagc ttgcttgataa taaaatctca aaaaaaatttt 960
cgtatcagtcgc tcagactgaa aacatggtgct ccagggcggcc aacccatcagc 1020
tgttttcacag ggcataagtt cattttcatt tttacacaagtg tgaattaat 1080
aaatggcttt ttcaaagttgt ttcataaat aaaaaagcca ccaaatgtgtg ttgcatttt 1140
tgtcttaaag gcagctgtgtg cataaaattt tagcatattgag cgttaaagtgt gtoaaactgt 1200
aatctctccgc gcacaacactt cactgttacgc aacaattttga ccacacgagaag tgaattttta 1260
ccacatactc ctattcnanaa atgctgttgc gcctattttat cccacttgga gttccggtgt 1320
ctatactactg acgcggccac ggcctacttc gcacaaaatcg acagggcggct ggtagcctt 1380
tgctgcacat cattctccatt cggagtgggtc agcttttttggt tcctatattaca aggatccctgc 1440
gttcaattgt tgggagcaac gttatatggt tgttttggtt tgtgtgtgaga tctoaactttt 1500
ttattacagg ggctgagccg gtctggcagc gcctgcagct cccagtgtg cttctagtct 1560
cgcctgaccc ctcctccccc tggcagccgc cttactcgtgt gcttttgctg tgaattttttt 1620
taagaacatt tttcataagtt tgttttttggt tgttttggtt tataacggtg ccacaaactg 1680
tccggcagcc tcaaaagttg gtaattgaaata tataaatagtg cccatctgccc 1740
attctgagata tttgctatag gcaaggggcca gcggaaaaga ggtttttcata tgaaaactt 1800
...-continued

```
acatttttca ttatgtgca acaaaatctaa ccaaatctat tcaatcatcata ggtggacaa 1860
aottcttc tcggcctgca tataaaccct gcaaatgcagtcgcaagctcgt 1920
tcsaatcttt tggctacaaa aaaaagcattc cgtttctttg ctttttattctttaaaattctaa 1980
cacaacaat cgcatcatac gggcccccggg ctcacccgctc gcaggtcttttgcgacagcttcggc 2040
acaatatgaag aaaaagcaca caaaataa aaaaatgag aaaaagcaca caaaataa 2100
acagttgctg tcgtacgctgc cgcagctctg cagcagctgc ttcctctcagcttttcgct 2160
ggctgcctcgc tcggcattgtc acggcatgtc cagcttttgacgtaaatttaa cctgtgtgta 2220
cctggtctgc tggctgctgc gcatttaattt aaaattttgtc gctttttgcag aatgtgctgct 2280
cagctacgct gcggccacg cggcagctgg cttgctccccgg gtcagctgctgc cagcagcttcggc 2340
acccccctgc tggctcaactgc tgggctgctgctg cgcgggaatgg gccagtgtgctg cggccggtgcag 2400
tgggctcag cgcggcaacg tgggctgctgctg cgcgggtgctg cgcgggtgctg cggccggtgcag 2460
atggtggcg gagaatccgg aatctggcgcttgagtgcggcgc cggccggtgcag aatgtgctgct 2520
cttttccacg cggcagctgc ttcctctcagcttttcgct 2580
tggcacaact ctttttttttc tgggctgctgctg cgcgggtgctg cggccggtgcag 2640
ttttttgcag cggcagctgc ttcctctcagcttttcgct 2700
atatacctc atatgtgatac atatattaattt ggagatattt ctttctccct tggcagctgc 2760
gtaaccaata ccgagcttgcc ccagacctgc agagcggcagc cggcagctgc 2820
acagggctct cggctgctgc gacagcggcagc cggcagctgc 2880
ttcctctcagcttttttgcag cgcgggtgctg cggccggtgcag 2940
gggttgctgc gtagctgcag cgcggtgctg cggccggtgcag 3000
cggcagctgc ttcctctcagcttttcgct 3060
acgacgacg cggcagctgc ttcctctcagcttttcgct 3120
acgacgacg cggcagctgc ttcctctcagcttttcgct 3180
ctgtctgctgc gacagcggcagc cggcagctgc 3240
acgacgacg cggcagctgc ttcctctcagcttttcgct 3300
acgacgacg cggcagctgc ttcctctcagcttttcgct 3360
ctgtctgctgc gacagcggcagc cggcagctgc 3420
acgacgacg cggcagctgc ttcctctcagcttttcgct 3480
acgacgacg cggcagctgc ttcctctcagcttttcgct 3540
acgacgacg cggcagctgc ttcctctcagcttttcgct 3600
acgacgacg cggcagctgc ttcctctcagcttttcgct 3660
acgacgacg cggcagctgc ttcctctcagcttttcgct 3720
acgacgacg cggcagctgc ttcctctcagcttttcgct 3780
acgacgacg cggcagctgc ttcctctcagcttttcgct 3840
acgacgacg cggcagctgc ttcctctcagcttttcgct 3900
acgacgacg cggcagctgc ttcctctcagcttttcgct 3960
acgacgacg cggcagctgc ttcctctcagcttttcgct 4020
acgacgacg cggcagctgc ttcctctcagcttttcgct 4080
```
-continued

gcagctgat gctaacctat cagagcataa aagttttctat caggaatccca tattgcaacctt 4140

gaggtcggac cc 4162

<210> SEQ ID NO: 42
<211> LENGTH: 12917
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE: OTHER INFORMATION: tko-step1-genbank

<400> SEQUENCE: 42

agcgttgcacc caattgcgcc tatagtgagt cgtaacctaa ttgctggcc gcaggttttac 60
aagcggctgaaa cctggcggtta cccaaactta ttgcccggc caaccaacccc 120
ccttgcctccc agcggctcag cccgcaacgga cccggccttc ccaccaagttgc 180
gacagctgaa tggcgaatgg gcagcgccct ttgagcgccg cttaaggggagc ggaggggttgg 240
tggttaagcg gcaggtggac gcaacctgttg cccgcgcctt ccacgcgctgtt cccgcgttct 300
tttcccttt ctttgcgcac acgtgttcgc gcctttcccg gcaagctctta aatgaggggc 360
tccttttgag gttgctgttt tggagctcgg gcagcctgac cccgtaaaag cccggaagttc 420
gtggagttgc acggctgggg ccaatgcgcct gataagccgg gattgctgctt ttgtaagtgg 480
agcgttaagct tttaatagtt gcacgttggc tccaaacttg aacacactgc aacactctct 540
ejggctcttt cttggtttta aaggtatgtc tggcagcttg ggcctatttg taaaaacttg 600
agcgttttta aacgcggaatt ttaacaaact attaagctgtt aacatattgg 660
tgaggctttc cggggaaaag ttggcgggaac cccatatttg tattttttct aaataactct 720
aatatgtag cggcgctcaga acaataacc tctgaaatag ctctcaatag attgaaaag 780
gagcgattag gcagttggcc attttgggtt cggcgttatt cccccttcgg tggcggtttg 840
cotcctgttt ttctgctgtc acggaacgct gcggtagtaa aagatgatgta cagcagagtt 900
gggtcgatga ctgggcttttc ctcgaatcga tttaaatcagcg gtaaggtctt ttggacgttt 960
tgcggccggag ccaacgttcc tcaagtagg ccctttttat ccggcgtgct gtgggcgggtt 1020
attacccgggt tattggccg gcaagacgcc actcgctggcg ccgataccact attttcagttac 1080
tggactgtgtt tggcgtaatc caagttgtga cgcataattc tgggcgttgg 1140
agaattatgc agtggctgca taccACCtag tggaaaaacct accgtcctagc tattttctgc 1200
aacgttttgg gggctagcagc gattacccgg ttttttttgca aaccaggtggg ttcgtcattaa 1260
tgcttttttt cgtggcgtgt gggatggata tggaaactctg gcaaagctc aacagtggcc 1320
cacgactct gtagtggaca cgaacctttt ccacagcttg gcaaatggg ataccttcac 1380
tccccccc gcgcaacacg aataagacgt gttagagggc gaaacctagt cagaccacact 1440
tctgcgcggt ggcctggctgg gttgcttggtt tattgctgatt actttcgtag gacggggagcc 1500
tggcgagctt ggtcagctgtt cggcacgagc gcagatgttt aagcctcaacct gtaatttagt 1560
taatctagcc cggcggcata cggagaactg ggtgacacg ggtgaacgctg aatacagtatt 1620
aggtgctcct gctgatagctt atcggtgttaat gtaacaaacg cggatttacg tactaatcct 1680
gattgatgaa aacacctatt ctttaattaa aaggaattgtag gtaagatgcc cttttttgtaa 1740	tctctgacc aaaaatccct cgttcggtt tggctggcc aacccgtaga 1800
aaagactaa gcagttttttt tttttgctgc gtaaatgcgct gcttgcaac 1860
-continued

```
aaaaaaacc ccgtacccag cggtgtggtttg ttggcggagat cagaagctac caactctttt
1920
tccaaagta aotggtottca gcagagcaca gataccaaat aotgctcttoo tagtgtaccc
1980
gttagttgagg caccacactca aagaactctctg aagcaacgacct aataacccctg cttcggtaat
2040
cctggttacca gttgtgttctc gcagttggca taagctgttg ttaaccgggt tgtaactcaag
2100
acgatagttg cccgattaaag gcgcacgagtc gggtgtgagac gggggtttctgt gcacacagcc
2160
cagctttggagg gcaagacgct ctacaagacttta gacagccact cagcgttgaag cagataaaag
2220
cgcagcgcctt cccgaaagga gaaagggcgca caggatacccg tgaagcggca gggtcgagac
2280
aggagaagc acagggagag tttcaggggag aacgcttacg tattctttata gttctgtgagc
2340
gttcgcagcac ctcgccactgtg acgctgatt ttttgtgatgc tgcgacgggg ggccccagcct
2400
atggaaaaacc gcagcaagacgc cggcccttttt acggttctcttg gcccttttgc gcgctttttc
2460
toacattgttc tttctcctgt tttcttggat aacgctttta cccgctttgta
2520
gtgcagctgc accgctggc gcacggcgc acgcagttcag tgggcgagaa
2580
agcgagaggg ccgcacacaac gcacacgcgc cttcccggcg ccgttgccoga ctgcattaatg
2640
cagctgggacc gcagcgttccgg cccgacccgg ggcggcaggg ggcagcagacg ccattaatgt
2700
gagggggtcgc atcactaggct ccccgaggc cttccaccttt atgccctcgg ccgcctgtgt
2760
gtgaaggatt tggaagtgag aaaaattctca cacaagggac gcgtctgacc atgcatcgc
2820
cacgcgagca attaacctcc aataaaggga aaaaaagcttg ggtccctgcg aggtgttttg
2880
tggcagaaacc acaaatatacc tgggccggag ccaatattgc tgcagttcagc acgaaattag
2940
cgaacattaa aaraagttga aaaaattcaacct atattgctaa ataaacacaca cctggataatg
3000
atgtaacgca agcggaaaaacc atttgtctata tttatacatttta ataaacctaa aacgaaattg
3060
aatsccattat gatgaaatatt ataaattctca gggcaagcga aaaaaagcttg aacttgttcc
3120
ttattgaac aacgactaggg ccctcagggg gacaaaaacgc cccagtttcgg gcccacatttt
3180
toaagtttgag tccacaggtacct ttttaattta aotgctctgg cttgctccag cattagccttg
3240
aaoaccocac cttctggttgc gcacagctg ctaattgaaaa atagttctct atataacgaa
3300
tgagctcggaa aaaaaaggtg ggggattata aaaaagttga agccatgccc gcttctctgg
3360
tggggagggtt atccttcggc gctgagttgg cgtcttctttcg tattgctgact
3420
aattactgta gaaattgaaacc ataatgytca aaaaaactttt aaggcccataaa gttaaaacc
3480
gtatgaatag aatatttatttt aatcgggttgt ctacaaaaat cttgccatttttt tttttatgtat
3540
ttttttttttt ctgctttccc tattagttta aataacattttta aaaaagatttt cttcctctga
3600
aataaatatc aataaatattt gcttttattg cggaaaaattt gttgatttattt tattttagtta
3660
gagcgaacct cgcgttgacga aaaaaagattt ccaattaggt gcgtcctaatt ggtataggcc
3720
gaaactgagtg caatactaaatt ctttcttttttg tgtgtgatgc gcacaaaat caaggggaattc
3780
genatgcaat aatcatttatt ctttctctgga atttaataatatg aaaaaggtttgttg aatatttttt
3840
caatatata aatctttattt ggtttcgtgaa caaggattttt cttgctcaatt gttgcataa
3900
tttttattaa aaaaaaggttg tgcgtatgtat ttattttttt caaggggtatgc aatattttttt
3960
ttttttttttt cttttttaaaa tttttttttt tattttaaaa aacggtgtatg gtaaaaatttt
4020
ctctttcgca gaaaaatgtcc gattttttttt ctaaaaattt caacgcctcag tataattgtg
4080
gcgaaccttg ccgctaaattt tttggttttttt ctaaaattttt aaggttctata
4140
```
tttaaccaag tgaatatggt ttcggattact atgaatttttcc aataggcgag accaattag 4200
atcaccctcag agaatatcag gcttcggctg gttcctggtgct aacgtcctgctg 4260
ttacctgtggat acgtcgctgc gctcctggtgctt ccggcctac ccggcctac 4320
ccttctgag tggctac gctgtgctgc gctgtgctgc gctgtgctgc 4380
tcttcgggttg gttccgggttg gttccgggttg gttccgggttg gttccgggttg 4440
tgccgaatg tgcgccgggt gtcctcgggttg gtcctcgggttg gtcctcgggttg 4500
atgtgatggtg tcaatcttatc tcttttcttg tcttttcttg tcttttcttg 4560
tccttcttcttg gttccgggttg gttccgggttg gttccgggttg gttccgggttg 4620
tgttggcag ctcctttcttg ctcctttcttg ctcctttcttg ctcctttcttg 4680
tcttttttttt ttgccttttttt ttgccttttttt ttgccttttttt ttgccttttttt 4740
tgcgtgtgtg gtcggtgtgtg gtcggtgtgtg gtcggtgtgtg gtcggtgtgtg 4800
tgccttttttt ttgccttttttt ttgccttttttt ttgccttttttt ttgccttttttt 4860
gagagggag gacgtcgtcgtc gacgtcgtcgtc gacgtcgtcgtc gacgtcgtcgtc 4920
tttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
-continued

tgacaacca ctacactgtg attctttccc ataaaaaac ccataaaaaa taagaaagaa 9760
aaacaagcct aatctttggc tactctgtota ttctgtgca gccattatata ttatgattttt 9820
tttaattggaa aatattcagc tctacoggggg ttaattgtaa tagcacaacac acaaaagaaggttg 8880
agtctgggc ctgctctata ccaactttgct tcggctgtaat cttaagtcaag cggccaaat 9940
tgcgacgcca aacattaacc gcgttggcaca agcaacgcttc gaaaaaagga gctaatccc 9000
atitgcacat gatcgaacacg tacaacatttgt atfagcacaac gcacaactttt gttgcccttc 9120
gaaatatgata ttataacgcca tttaatacgttt taagccccga aatccctttct tgaagaagctc 9180
gcccctctct atcctttttaa aagatatttct tgcaacaatttt ctgcaccata tatactgtaaa 9240
gcgaatattgc cgctgacaggg ccaacggcaca ttcttaacagt acogcagagct gcgcceggcc 9300
ctcccaacag atcattggtaa ggaacggcagc aggtttgtgtg gaaactggtc agatctggggc 9360
tgacacgtat cacgtgttttttcat aactcagcact cctcgaaggg gccttcgccct tcggcagcags 9420
catcgaacat cctcttagtt tatttcgcgtgt tatactgtgat atctgggttag toctaggttg 9480
gagagttgtgt gtctttatgc gcagcaagctag gaattgtaaac atatagcagc ataaaaagttt 9540
tacattggatt tacaatcctt caagctgggtt ccctctccat cactttttttc ctatctgtaaa 9600
aaacagctgg tgcggacgggc gcctcataagat atgcaaaacta ctacguttctg gtagatattttt 9660
ctacttaata gtcacacggt gtgggacaatc ctataatcctg caacacacagc ttcgaacatc 9720
tgtcagttcsa aaaaaaagcag ctggcagcgact caagcgtggtc ccgctttacgc ggtgtgcttc 9780
tgtcataagc caataacactt atccggcaggg acaggtgttct cttattctgtg agaagacgccc 9840
agcttctctg gttgatccac gcggcggggag gcaacactttc atggggacac gcagccctctc 9900
catctcagcc gcgcgcgtac tcaattgcgtg gcggcggcagc gacgctggcggg ctcggttattt 9960
cggggtatatg gtaataacgcc gcaacagctgt gtaacacaca gttgagccac 10020
nggcgcaggga gctccaaaaggc atggatcgggt tctctagggct cgcctagagc gttcctgaa 10080
cgccgacaggct ttcagctgtgc gcctggggcg cccgctgac gccttcaggg gcacagccagc 10140
cgccgacggtc aggctgacac gcggcgccctt cctctctctgc gcctgagcaga tttgctggccc 10200
cagttctgtg taagccgtagc gcggcgtgctg gagaacacccc gcgcgactcc agcttacacaa 10260
aggacttttcc ttcctgaggt gcctctagttc gcggcgcggag atgaagctgtg aggacggcccg 10320
tttcggtacgc gcttgagccg gctcgctttcc gcgagcaaggg acgcgtgtctt acaagatgagtt 10380
ggatgggagc accaaccttc ccctgggctcg ccctgtattgt cagaaagaca coctagggctg 10440
gggctccctcc accagacggcg tgcactccagcg cagcctgcgtg ttgaggggag gcctccacac 10500
ggctctgaag tgggtgagac gcggcgcactc cctggtggag tctaagacgtc ctatactggcc 10560
cgaagaaggg tggcagcttgc gcgctctacta ttcagttgag gcctagccgtgc acatcccctct 10620
cagcagcggcag atacccactg tgcgggaacag tgcggagccg gcctcctccct gcctcctctct 10680
gttcctttggc cattggcagc gtcgagaccc gacaagcaccgc ttcgagacgct ctccttcgcga 10740
gacacacacat aggctgtatt ctgcagccgtt ctaagcttatt gcgtggagcta cggggtacctc 10800
cctccagga cattggagttc gcgtgagggag gacgctggtg gacgctggcgc atggagcaga 10860
acacaagccg aacacttccag rcaacaaaaag ggagggcgctg ccctccgtct gcgggatactt 10920
gtgccccccct ttatgatcgag gcagcggccg ccctctctgtc gcatcagcagc acatgccgga 10980
ctacaaaaag ctgtccttct cggagaagct cagagtgga gacgcgtgatga attcgaaga
11040
cggagacgt gtacctgta gccaagaatag caagttgcaag gacggttaacg tgatactaaa
11100
ggttaaatg cgggagcact aaccttcgcg ccagtgccca gtatgcgcga agaagcacat
11160
ggggtggggc gatocacagcg aagctgtgta cctctggaac gaggagtcag aaggccgagat
11220
catccagccc tccaacgtaa aagaagtttgtc tctactctgtg tcaataaatttg agaccattta
11280
catggccagc aagccggttc aagtgcccgg atattattat ttggagacga aactgtgatat
11340
aacttagcgt aacggaagac acaccatggt gcagacgtat gacgcgacgc agaagccgaca
11400
tcaacctgct ccctacgcca cggacgagct gtcaaggtaag ggcgcgagcga cctctagata
11460
taatcagccca taccacatttt gtzagagttt tactgctttt aaaaaacctc aaacccactcc
11520
cctgaacct gaaacatata atgaatgcaaa tttgtttgttg taattgcttt atttgccgatt
11580
ataatggtcta caaaatagac aatagcaatca ccaatctccac aataagaacca ttttttcccac
11640
tgcctctcct tggaggtttt gggagcgttg tccagaggtt acggagagtg taccaggttctc
11700
tagttgtccct tggaggtttt gggagcgttg tccagaggtt acggagagtg taccaggttctc
11760
atgtttgtccct tggaggtttt gggagcgttg tccagaggtt acggagagtg taccaggttctc
11820
caatagatgcc cacactcaca ctagatcaatat aaaaaaaatg gagaagcgtta gaatggcga
11880
cagttttctgaa taattatat ctaatatca ctagtacca gaatatacag taggaggaat
11940
ggatatata tataatatata tatttgggttattt agtaggttattt accagcgctc
12000
gtttttctt tggaggtttt gggagcgttg tccagaggtt acggagagtg taccaggttctc
12060
gaaacatgct aaaaaaaaaa tggctttccaa gcggaaagaag aatataatata
12120
aatgtagttc caggtatgta cagatcagca ctagcagtagc tccacccca
12180
gtttcttcct tttacgaaat taaaatttttt tggctttccaa gcggaaagaag aatataatata
12240
aaggtgacta cagaaacgtc cagatttcgttta atcattattatatatat cagttttttt
12300
attcatttcct tttctttctt cagatcagca ctagcagtagc tccacccca
12360
atcatttcct tttctttctt cagatcagca ctagcagtagc tccacccca
12420
aacatatgcc ttcccccttc catataatta tgtaatttttc taattttttc tgcggtgtcg
12480
agttttaaat cotaatttttg ccaagaaaatttttgattgccatacagcaggtatggactagt
12540
atttaggcgt tcggagattt aaaaagccaaa atcacacagaa aaacgcttctg tcatcctgtg
12600
tcatctagc ataatcattata aattttcctg tcatgtgttt ccgtggtcctgt ctaggcacgt
12660
tcgacgggtatt cggctacagtt cccacccaat aacgccacagg gacccataca cctccgaaat
12720
gatttcttcct tttacttcttt cagatcagca ctagcagtagc tccacccca
12780
tgatgctggag cagctgtgtct cagatcagca ctagcagtagc tccacccca
12840
<210> SEQ ID NO 43
<211> LENGTH: 14149
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: tko-step2-genbank
<400> SEQUENCE: 43
atcatttgtc gcagagttgtc atcacaagcc gcgcgttgga gtcctccgga tcaattcggc
ttcagttac gcagacgtat tagttaagcct cttcgaagcc gcgggtcgagg tgtccaggcgc
60
120
tgcccttggg ctcccccgggc gctgactcca cctcaccctcct ctgccccacat agatgaagcg 180
gtctgaggtg gctgataggt atcccgccgca aacgagccgg gaacgcgggaac ccctgcctct 240
cggaaaccagt ggccgctgct gtaacaggtg gcacagggcct gcggagcctg gctgccggt 300
gggactagct ggccccaggt cagcaggtttc tggagagcct gggccaggct gcggagcaca 360
ttgtgctcc tattctgctc tttgggtgatt tagagttgatt gtgatgtgtg atgaggttgt 420
tagttgctca gcggagactg tggctagcag gatgacgctg gcacacacac gcaagtgcacc 480
gatgataat gctatgtctac tactactact ataatgctat gattttggtgc gcggagcagt 540
gtctgacaca ctaaactccg cccatttactg ttaaatgatcc agaacaactagg agaaccattg 600
tccaaatct tattgtaacc actgctagtt ctagcgaaag gagcgcacacc aagacagccag 660
cgttaaaagg cgcgctttgt gcggcttttc aataacgcct gcgtcctgg gaacgcactac 720
aaaatacgcc gctcaagtcag gacggctgga aacccgagac gacctaaag ataccaggcg 780
tttccctctg gaagacccct gcgtgctgcc gtaacgctgag ccctgcctgt taccacatc 840
cgctcctcgcttctctcct gggagcctgt gcggctctcc atggtctggt atagctagctg 900
tcagctcggg tttggtcagct tgcgcccaag ctgtgctgtc tgcgaacgca tcccgtctcg 960
cgcccagccg ggcctaatcg ctggctgagt tccaacgctt aagaaagcag 1020
ttatccgccg tgccccagcgg cacctgtaac aggattagcc gagccaggtactgtaggtgcc 1080
gtctgcaagt tttggtcagct ctgctgcaag tctgcgcagat ctagctgtgg 1140
atcgctcgct gcgtcagctag gtttatccctt gccaaacagg tiggttagctc tttgtgcagc 1200
aagcagctccttcttaatcag ggcgccgat cggcctgacgt gcacacacac gcaagtgcacc 1260
aaaaaggtgc ccaccaagag cctcttcagt tttatctgta tttatatcag atctatctattag 1320
gaaaccacct gcggctgact gttccgatat ggtttcatag aagacagcttt cccctcagg 1380
ctttataactt aaaaataagg tttttatcata atctaatgta tattatgta aacctgtgtct 1440
gactggtacat aagctgatca cagctgggca ccatcctcag cggatctttt atttgtcttg 1500
tccatgcttt gttgcttcgcc cggcttattt aataacgcct ccgaggcagag cgaagcgagtcttg 1560
ggccccagctg tgcgcgtacag aacgcagcgc cccgctcgcag ggcgcctctgt tttatgca 1620
ataacagcgagcggccagtcg aaggaagttgcc ctgcaacactt atgcggtcct 1680
atcagctctat ttaatggttc cgggaagct agatgaagcg tttgcgctag taatagagtgg 1740
cggcagcggtt tcggactgtcg ttcaggggtactgcgcgt tgggtgctgg tggtagttgctg 1800
tccattcacgt ccggctccca aacgtacaggg cggctctgtat ctgccccacat gctgtgcacaa 1860
aagcagcttc gttcggttcgct ttcggaggttc gttctctggtc ggcagctgtta 1920
tacaaacctc ttcgcggagtcc ccoutcattt cttcctgca acgtagagtc 1980
cccttcgtcc tcggctgtcata ctccttccttg atgctgccagc ctttaccttc 2040
agctgtcttt gcggcggctgcc attaggggt aatacggcag ccacagcagag cacccttaaa 2100
gtctcctcata ttgaaaaacg cttcctggg cggaaaaactc caggtacttc acgcagcttg 2160
agatctgagt ctggctgatcc ccaccccgtct cctccgctct ttatctcttc 2220
acgccgtct ggccgctgctg gaaacaggaa aggcaaatcg cggcacaaga aagcataagg 2280
gcgcacacac aatggtttcct atccatactc ttctttttct ataatatttg aacgatttat 2340
caggtcttaat gctctcagat gcggataagat ctgagtaatg caaatggatag 2400
<table>
<thead>
<tr>
<th>Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>ggggttcgc gcacatgcc ccgaaaaagt ccaacctgagc tgcaggggtc ggagcgcgg</td>
<td>2460</td>
</tr>
<tr>
<td>ecggggctct aatccataa gagaactta caattaagtc taatccaa attggccaaag</td>
<td>2520</td>
</tr>
<tr>
<td>ctatcgtatt tccaagcctt gaacgctgga gatataaatg agggctccct tctaagggcc</td>
<td>2580</td>
</tr>
<tr>
<td>gacaactcc tcccaacag caaagtgaac agctggctaa gcaaagctga agcaaaaaat</td>
<td>2640</td>
</tr>
<tr>
<td>caagcgcaag ttcagacgtg ctgcgagaag ctgctgcaataa gtttgagcaa</td>
<td>2700</td>
</tr>
<tr>
<td>ggacagggag cgggtgtgcct cgcccttota gcagctgcag gcgacgctga</td>
<td>2760</td>
</tr>
<tr>
<td>cgccacacag ttccagcgtg cccggcaggg cgcggggcgg tgcaccaagt gcacacctac</td>
<td>2820</td>
</tr>
<tr>
<td>ctctgaagtc atctggcaaca cccggcagct gcggctgcgc cggcggcaac tcgctgcaac</td>
<td>2880</td>
</tr>
<tr>
<td>ctctgacatt ttcgctgcag gttcagccct cctacctgacc acacacagt gcagcagctt</td>
<td>2940</td>
</tr>
<tr>
<td>ctctgaagtc gcacagccag aagggcagct cccggcagcgc acacattctct tcggagacag</td>
<td>3000</td>
</tr>
<tr>
<td>ccgcaatc aagacaggcc cccaggtgta gttcagggg gacacccctgg tgaacgcgtg</td>
<td>3060</td>
</tr>
<tr>
<td>cgagctggcg gcgaagctcg tcaaggggca gccaaacttc ccgggcaaca acgcgtggtg</td>
<td>3120</td>
</tr>
<tr>
<td>caactaaca agccacaaaccc cagctatcag gggacacagc cagcaagac gcacactagt</td>
<td>3180</td>
</tr>
<tr>
<td>gaaccttcag atccgcaacca acatcggagc cccggcgttg cagctcgccg acactacta</td>
<td>3240</td>
</tr>
<tr>
<td>gcagacacac ccctatgccc aagggccctg gcgtgcgccc gacacacact acctgaccc</td>
<td>3300</td>
</tr>
<tr>
<td>ccgtctgccg cttggcgaac ccccccaaca gaagccgctg cacatgcccc tcgctgagct</td>
<td>3360</td>
</tr>
<tr>
<td>eggtggcccgg gcggagcatt ttcctgcgag gaggagccag tcaagaaaaa ggcggcggc</td>
<td>3420</td>
</tr>
<tr>
<td>ctctagatcs taaccataaatt tggagttctc acccgctttt cttcggttttt aaaaaaccctc</td>
<td>3480</td>
</tr>
<tr>
<td>ccacacotcc ccotgaaact gcacaataaa ataaggtgca ttgggtgtttc ttaatttttgtt</td>
<td>3540</td>
</tr>
<tr>
<td>attgcaagtt ataattggta caaataaa gatgactca caaatccac aataaagaaca</td>
<td>3600</td>
</tr>
<tr>
<td>ttttttttccg tcggctctcg tgggttcttc tccaaacttc tcaatgtact tcattggtcctt</td>
<td>3660</td>
</tr>
<tr>
<td>tgggccacac ccaacactata cttgctgcga gcaacatgca cggcttaaat cgtatggcaca</td>
<td>3720</td>
</tr>
<tr>
<td>gaaaaaggtat gcaataaat aatctttttat tggttatcga acgaaaaaatgg ttctgcgcaaa</td>
<td>3780</td>
</tr>
<tr>
<td>cgtatgtcat atttccttatt aaaaaaggtt atctcaatttt ttttataatt aacgggtctatg</td>
<td>3840</td>
</tr>
<tr>
<td>caaatata atttctttatt cggtgatcga cggaaatttt tttgcgtcaca gatggacaata</td>
<td>3900</td>
</tr>
<tr>
<td>ttggagggatt ttgcaatgttat ttatataaaa aaggggtatgc aataaaaaat</td>
<td>3960</td>
</tr>
<tr>
<td>tatttggttt ttcggaanaag tagcatggcct ttatattttttt ataagttgtag ccaccaaaaat</td>
<td>4020</td>
</tr>
<tr>
<td>ttacctttgc gcacaataatctg aatgttttttg gcataaattt caacggccacc tttatcctgt</td>
<td>4080</td>
</tr>
<tr>
<td>gccacagcgc ctattgggatt cttcgctggcc cttccgcaga aacggccoga gggcggaaag</td>
<td>4140</td>
</tr>
<tr>
<td>gatttggttg ttcacacgcc agaagggcat tgtttttttc aatattgaac catttttttttta</td>
<td>4200</td>
</tr>
<tr>
<td>tttgacccct gcacgctgta atttttttttt cacgagaataat gtgatcacac aatatatataaa</td>
<td>4260</td>
</tr>
<tr>
<td>aaaaatatata ttatatatatct gcgctgttttt ctttttctgt gaaaaatttga ttattatgtas</td>
<td>4320</td>
</tr>
<tr>
<td>aaaaaaaatgttattat attcaatata tataaatat aatatagatgt tgggttggaaa tgaatcttgg</td>
<td>4380</td>
</tr>
<tr>
<td>egatacatttt gattaaaaat cttctttgtattc actcatcctac aagtgccgaa aattttttt</td>
<td>4440</td>
</tr>
<tr>
<td>taatttttttta atacagtggc gttgctgaagc aagagttgtc ttgctgtat cttctttttttc</td>
<td>4500</td>
</tr>
<tr>
<td>aagattttta attagctattt tataaatattc aagcagattttt gttggttgaa</td>
<td>4560</td>
</tr>
<tr>
<td>ctttttccaa aattttagttt gattgatatc tcaatagttc aataattttgtaaatttctggc</td>
<td>4620</td>
</tr>
<tr>
<td>cgcaacagcgg catttttacac aatgagggc acacagtggc caacactactt ttctacac</td>
<td>4680</td>
</tr>
</tbody>
</table>
tttcgggattg taagatat ttctacgcag cagacgatg acaagattta acaatattgta
4740
agcttctttt aacaaatatta gtatctccctt ccacagaaaa cacataaatt ctgtaattat
4800
tgactggatgt gtagtgatgc ttggctgctg tataactataa gatctatagg caggggataa
4860
cgctttaact ctctttaaaac ttaagcgcgg cggcgctag ttaacgctta ctatctcact
4920
ggttaaaccg cgtttgtcag gaatcgtat atatatctca ttgttatata aacgtgcaaaa
4980
agagcgcaca gacctgctgta cagagctgga ttcatctccaa ctcttctctt ctttttgggg
5040
tcaagcctgct gtctctcacaag gatgtagat tcgtttctata gagecccgta
5100
attgcaatgtg aatagcgtg gggtcggacc aacctcttttc tagaggttga cggccccttg
5160
tctatggttgg tgtcagagta cttagaaggct gcagggcgcc caaagttgcg cagagtaacc
5220
aagtggatcgc ttttttctgct cttttttcat tgcgcgtttc gctttttat tttttaagca
5280
ggatttccct tctctctcag ttttttcttct cccttttaca ctttctcttc aacagccgcc cttttggtcctc
5340
tgctctgatc ctcatcctag cttctttctag tttttttttc gttgtgctcca cagacacgctg ctttctctca
5400
tttctcttg gtagagccct gcagactcttcatttggtgcgg cagattcaaa gaaattaacg
5460
ttttctttg gatgagacggc atttttttccatgctttcag ttggagtctg gacttttcggg gatggtgctggttctgagctggtttccc
5520
cagttttcggc ctttttttttaatgcttcag ctttctttttc cctttttttc cttttctttct ctttttttttcttttgggg
5580
tttttttttct cttttctttttt cttttctttttt cttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
-continued

gctgggttct ctgcaagatt agcagatgtgc tcgtgaagac tggccccttgt gc gagaaaaact 7020
tgctctttct ggagatcttc ottaaaggct agagagtcat catggacacta otgoatgacag 7080
tcccggtgtaa caaactcact gacactaaga aactctgagtaa ttctcttccc acctctcttg 7140
tctcggatcc catgtgatcg ttctctctttc agcgccccct atcggatca actgggctcg 7200
ttcgtgctgc tctggtccc attggtggct gcgtgacgat aatagtgcctc ggcgttctca 7260
actctctctc ccatctgtaa caccgtaaag gtcgggcaccc tgcctctcaag aatgtcctcg 7320

tctcctctct gtcctggttt ttaatggttt cagggagcgc gttactacgt 7380
cagccggtagt cgccagacgg ctcctgctct cgcctggatc gagccagttt cacccatcgg 7440
atcctttttga aatgctcctc cttctgagtgt ttcgcgcttgtag cttctcgggtc cgctcttgag 7500
agagggctca caggtgcttt cctctctggct ctcgacagag atgcattgctt ctcctctccct 7560
ttctgtgatt attggtgact gtcgacgcttg ttaaactctg tgaagatcttc gttacacgaga 7620
gaggttctgtt gagactctgtt tccgtgaggg catttttttttt cctgctttgct ctcctctctg 7680
atgagggact gggcagagg cccctcagctt agcccctcagc cagaggtgtctc cagcaggggt 7740
gtcctctctcg attggggagt cactacagcag acactctcaata tctcctcctctgc gactgggcccct 7800
acataataggg atataccggtaa tttgggaggt tttctcatcttt tttctcctgggt atccttctcgg 7860
aaggggttga aatcctccttt ccggaagttag atagcgttgtcag gttcggccagag tgaatctcgg 7920
tggggggcgc tcctatgctgg aacacttggttg tttggtgcagct acagtctggaga ctgtaagcag 7980
ttttcgacgtg gtcgctctct cgtcctcttc ccgactaac aggtctgatgg tatctctgttt 8040
aatcctctcg cctgtgacgg ccttctccttc atacacccatat cacacactttttt cagcaggg 8100
aagaaaatta cctgggtgctt cagggactgt gtcgtctgcctg caagggcctt cgcacacacagt 8160
aatgccttgtt ggctgacactc atctgctgctg atctcactacta cacgacacggc gctctgtttggc 8220
atcctccgtgt tcaatcggcag tataactacta cagcaatttgg ggtctctactgg gtctttgtgcc 8280
gctaaaaaaaag cagctggtctag cttgctgctag cgagcaggtttaa gcatctcacta 8340
tgtgcttggtc cttctctgctgg ttaagagcttg gcctctcttcg cggagctgacca cttgatagtt 8400
aaggtgggagg ccagagagaag ctagaggtgc taagagtctatc ccaacagcccct cttctctcttc 8460
ttcgccctgg ggtggctgcag agggttctcccc cggccgaaaaa ctaggctgtctt 8520
agaggggcgc ttgctctcagc tcctggctctg tctctctcttt ggagttgggactcacaggttctg 8580

taaggggtgga cagatgagttgc aacagttttgc tggacacagcgt tggatgtctgctt ggtgacggctc 8640
ccctgcgtag gcgaagaggtcgc cccgactaccatatatcctgacagcttctagtggctcaga 8700
cgcagccctc cccggaattcgt ccctccctcag ccagactgctgac tagatgctggc 8760

tactttttct ctagtccagc ctcctcagctgc atggctgcaaa atttggggttg ggcgtcttgctc 8820

\[\text{contd.}\]
-continued

actgaagtat tcgctgact tctttatcta atataaatgt acatgtaacct aaactcgctt 9300
tgggttaaa atcgtgagc aggagcacaataaatgt gtcocgtgaaa caactgaa 9360
caacctgcaaa ggtctacaag tcctcggaga ttttggtaaagctcgccg ccaacta 9420
gttcttctt aaagggcgat aataaattt ttcgaatgatg tagatattttt 9480
tagtaaatttt gagaagggctt catcataaat tataaagaaa ttttataatgt 9540
tgcnacataaa ttcctttcaag catcataaat gtcctcataatagactgtgc ggacacccg 9600
aatggaaact gatcaactaaca cagaggacaagagcttgagtt tataaactaaac aaaaatacgas 9660
aacatccccc cccactaacag acacccacgcc gacactcggcc gtcacaaaaac aataactgca 9720
cacacccact ggaatcgtggc gacagcgtgac aacactcatac gatattgcccc gcggcaacacc 9780
ggattctcgc cccacactgat ttcgctcagta aacctacccg gcacccccagg gcgcgggacgttctttgg 9840
aacacggcgc gcggcctgaa gacgacaagct cgtattgctgttgagcgtc gtattttcgg 9900
gcggcagatt gcagctgcct attgcacagg gaggccagac cccggacactc gcgattgcgat 9960
cacacactgct gggactgcct atcactacgtgct cttacatagttt catatctttg catatcttgaga 10020
ggpacatgtgg ggtggacagt tagctgctgg gacattcttcc acaatataagactgcttc gccagagtatt 10080
cacaactgct ggggctcataaatgtgagc gaaaggtatgg caataataata atctctttatt 10140
gggagtcaaacc gcaggggatttg ttgctcagatgactcata ttcctttatattt aaagaggctta 10200
tgcaataattt cttctctaaattt ggggatagtt gccataataa ctttctttatttg ggtatatgcaac 10260
gaaatcttgg ttcgcgttaag ttcctcataat ttttatttta aagaggggtatatgatgatgc 10320
ttattaaaaa ggggtatgcga atدادaaaa stttttttttttttttttttttttttttttttttt 10380
tttttttttttt ttctcattttt gatccttatgct tttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
-continued

atctcaggca tccagcaagc gttgtgccgc cgaatcgccc atccgtgaag aagatcctgt 11580
ggaggtgac aaataatccc oggcatatgg tggcatccac gccgtagaaa aacccatcag 11640
cgagctgacc acctctgtgc tgaaccacgc ctcttcacag tccacccctt caccgaaaaga 11700
aacaaggccc gtagcgtgcag gttgtagctg cggctacgag ggcgggtcag aaaaaatcag 11760
cgtagaact cacaatatcc aaggataata acctgcatag cccacagatc tccgactgcc 11820
gtagatcag tctgtacgtgt tctgtgcggag ccaccaagag cccagaagcaca aagcagcagt 11880
gatacaggac aagtgcaccc cctgctcagta tgcctccttc ggtctactgta tttccatgc 11940
cgaaactcaggg ctagccagcgt ggtggtgtgcac gctgtgccac ccatacggaact ctctacgcgg 12000
cgttgaccc tcatcctgga tggctagtac oggagaacgca ccccttttgtt acagacacag 12060
cgttgaccc cctgctcagta tgcctccttc ggtctactgta tttccatgc 12120
tatagagaag atacgggtag gcctcctgga ggtcagcgttc cctgctcagta tgcctccttc 12180
tgctgacac ggacgtaatgc ggtgagatgg ggggctggtg gctgactata cattcgctttt 12240
cggggctggtg gctgtctttttt actgctcttt tcggaggcttg aagtgcaccc cctgctcagta 12300
tgctgacac ggacgtaatgc ggtgagatgg ggggctggtg gctgactata cattcgctttt 12360

atatgggttata ggtggaggtta atattttttt atctactgtgtg atgtcctcga cttcaataca 12420
aaacactcaat atttttctgtgc aataaagcat atatttatattatatttacta ggggaaaaaa 12480
tcttttttttt cttctgtttttt aacactataat ggggaaaaat tttttttttt tttttttttt 12540
atccgaaat tattttttttt tcgagcttggat cggagtactga agatcagtgct 12600
atcagccta atctagttttc aacactataat ggggaaaaat tttttttttt tttttttttt 12660
atcagccta atctagttttc aacactataat ggggaaaaat tttttttttt tttttttttt 12720
atcagccta atctagttttc aacactataat ggggaaaaat tttttttttt tttttttttt 12780
atcagccta atctagttttc aacactataat ggggaaaaat tttttttttt tttttttttt 12840
atcagccta atctagttttc aacactataat ggggaaaaat tttttttttt tttttttttt 13000
atcagccta atctagttttc aacactataat ggggaaaaat tttttttttt tttttttttt 13100
atcagccta atctagttttc aacactataat ggggaaaaat tttttttttt tttttttttt 13200
atcagccta atctagttttc aacactataat ggggaaaaat tttttttttt tttttttttt 13300
atcagccta atctagttttc aacactataat ggggaaaaat tttttttttt tttttttttt 13400
atcagccta atctagttttc aacactataat ggggaaaaat tttttttttt tttttttttt 13500
atcagccta atctagttttc aacactataat ggggaaaaat tttttttttt tttttttttt 13600
atcagccta atctagttttc aacactataat ggggaaaaat tttttttttt tttttttttt 13700
atcagccta atctagttttc aacactataat ggggaaaaat tttttttttt tttttttttt 13800
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 trans-
   gene; 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 endo-
   genous 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
   autosomal, X chromosome, Y chromosome, or supernumer-
   ary 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-spe-
    cific nucleases, engineered site-specific nucleases, base edit-
    ing 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 chro-
    mosome 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 germ-
    line 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 comprises
    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 trans-
   gene; 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)

* * * * *