

# US Patent & Trademark Office

## Patent Public Search | Text View

United States Patent Application Publication	20250255283
Kind Code	A1
Publication Date	August 14, 2025
Inventor(s)	Crisanti; Andrea et al.

### GENE DRIVE TARGETING FEMALE DOUBLESEX SPLICING IN ARTHROPODS

#### Abstract

The invention relates to gene drives, and in particular to genetic sequences and constructs for use in a gene drive. The invention is especially concerned with ultra-conserved and ultra-constrained sequences for use as a gene drive target with the aim of overcoming the development of resistance to the drive. The invention is also concerned with methods of suppressing wild type arthropod populations by use of the gene drive construct described herein.

Inventors:	Crisanti; Andrea (London, GB), Kyrou; Kyros (London, GB), Hammond; Andrew (London, GB)
Applicant:	IMPERIAL COLLEGE INNOVATIONS LIMITED (London, GB)
Family ID:	63042589
Assignee:	IMPERIAL COLLEGE INNOVATIONS LIMITED (London, GB)
Appl. No.:	19/191416
Filed:	April 28, 2025

#### Foreign Application Priority Data

GB	1810253.3	Jun. 22, 2018
----	-----------	---------------

#### Related U.S. Application Data

parent US continuation 17253553 20201217 PENDING WO continuation PCT/GB2019/051757 20190621 child US 19191416

#### Publication Classification

Int. Cl.:	A01K67/68 (20250101); C12N15/85 (20060101)
U.S. Cl.:	
CPC	A01K67/68 (20250101); C12N15/8509 (20130101); A01K2217/07 (20130101); A01K2227/706 (20130101); A01K2267/02 (20130101)

#### Background/Summary

CROSS REFERENCE TO RELATED APPLICATIONS [0001] The present application is a continuation of co-pending application having U.S. Ser. No. 17/253,553, filed on Dec. 17, 2020, which is a 371 of International application having Serial No. PCT/GB2019/051757, filed on Jun. 21, 2019, which claims priority to Application GB 1810253.3, filed on Jun. 22, 2018, the contents of all of which are incorporated herein by reference.

#### INCORPORATION BY REFERENCE OF SEQUENCE LISTING

[0002] The Sequence Listing in the XML file, named as 41041Z\_SequenceListing of 247,167 bytes, created on Apr. 28, 2025 and submitted to the United States Patent and Trademark Office via Patent Center, is incorporated herein by reference.

[0003] The invention relates to gene drives, and in particular to genetic sequences and constructs for use in a gene drive. The invention is especially concerned with ultra-conserved and ultra-constrained sequences for use as a gene drive target with the aim of overcoming the development of resistance to the drive. The invention is also concerned with methods of suppressing wild type arthropod populations by use of the gene drive construct described herein.

[0004] A gene drive is a genetic engineering approach that can propagate a particular suite of genes throughout a target population. Gene drives have been proposed to provide a powerful and effective means of genetically modifying specific populations and even entire species. For example, applications of gene drive include either suppressing or eliminating insects that carry pathogens (e.g. mosquitoes that transmit malaria, dengue and zika pathogens), controlling invasive species, or eliminating herbicide or pesticide resistance.

[0005] CRISPR-CAS9 nucleases have recently been employed in gene drive systems to target endogenous sequences of the human malaria vector *Anopheles gambiae* and *Anopheles stephensi* with the objective to develop genetic vector control measures.sup.1, 2. These initial proof-of-principle experiments have demonstrated the potential of gene drive approaches and translated a theoretical hypothesis into a powerful genetic tool potentially capable of modifying the genetic makeup of a species and changing its evolutionary destiny either by suppressing its reproductive capability or permanently modifying the outcome of the mosquito interaction with the malaria parasites they transmit.

[0006] According to mathematical modelling, suppression of *A. gambiae* mosquito reproductive capability can be achieved using gene drive systems targeting haplosufficient female fertility genes.sup.3, 4, or alternatively by introducing into the Y chromosome a sex distorter in the form of a

nuclease designed to shred the X chromosome during meiosis, an approach known as Y-drive.sup.4-6. Both strategies are anticipated to cause a progressive decrease of the number of fertile females to the point of population collapse. However, a number of technical and scientific issues need to be addressed in order to progress from proof-of-principle demonstration to the availability of an effective gene drive system for vector population suppression. The development of a Y-drive has so far proven difficult because of the complete transcriptional shut down of the sex chromosomes during meiosis that prevents the expression of a Y-linked sex distorter during gamete formation.sup.6, 7.

[0007] A gene drive system designed to destroy the *A. gambiae* fertility gene AGAP007280, after an initial increase in frequency, induced in the span of a few subsequent generations the selection of nuclease-resistant functional variants that completely blocked the spread of the drive.sup.2. These variants comprised small insertions or deletions (i.e. indels) of differing length generated by non-homologous end joining repair following nuclease activity at the target site. The development of resistance to the gene has been largely predicted.sup.3 and is regarded as the main technical obstacle for the development of an effective gene drive for vector control.sup.8-11.

[0008] As described in the Examples, the inventors have developed novel genetic constructs for use in a gene drive approach which targets a key sequence of the doublesex gene of *Anopheles gambiae* essential for the maturation of female specific transcript of this gene. The doublesex gene has been shown to be ultra-conserved and ultra-constrained, and so represents a robust target gene for a gene drive approach.

[0009] Accordingly, in a first aspect of the invention, there is provided a gene drive genetic construct capable of disrupting an intron-exon boundary of the female specific splice form of the doublesex (dsx) gene in an arthropod, such that when the construct is expressed, the intron-exon boundary is disrupted and at least one exon is spliced out of a doublesex precursor-mRNA transcript, wherein a female arthropod, which is homozygous for the construct, exhibits a suppressed reproductive capacity.

[0010] Sex differentiation in insect species follows a common pattern where a primary signal activates a key gene that in turn induces a cascade of molecular events that ultimately control the alternative splicing of the gene doublesex (dsx).sup.12, 13. With the exception of Yob1 acting as Y-linked male determining factor.sup.14, the molecular mechanisms and the genes involved in regulating sex differentiation in *A. gambiae* are not well understood.

[0011] However, without wishing to be bound to any particular theory, the inventors hypothesise that the gene dsx is key in determining the sexual dimorphism in this mosquito species.sup.15. In *A. gambiae*, dsx (i.e. Agdsx) consists of seven exons, distributed over an 85-kb region on chromosome 2R, with similarities in gene structure to *D. melanogaster* dsx (Dmdsx) and orthologues from other insects, and is alternatively spliced in the two sexes to produce the female and male transcripts AgdsxF and AgdsxM, respectively. The female transcript consists of a 5' segment common with males, a highly conserved female-specific exon (exon 5) and a 3' common region, while the male transcript comprises only the 5' and 3' common segments. The male-specific region is transcribed as non-coding 3' UTR in females, as shown in FIG. 1a.

[0012] The inventors have surprisingly identified that this female-specific exon (i.e. exon 5) of dsx is ultra-conserved across the *Anopheles gambiae* species complex and even throughout the wider Anophelinae subfamily, as shown in FIGS. 1b and 11a, and 12. This type of ultra-conservation is very rare because even proteins that are highly constrained show some variation at the level of the DNA sequence because "silent" variation does not alter the composition of the final encoded protein. The inventors carefully assessed the ultra-conserved sequence in the doublesex gene and, without wishing to be bound to any particular theory, believe that it is the splice acceptor site at the 5' boundary of exon 5 that is required for sex-specific splicing of dsx into the female form, as this sequence may represent the target of RNA binding proteins that direct the alternative splicing of this important exon.

[0013] The inventors were especially surprised to observe that targeting an intron-exon boundary of the female specific splice form of the doublesex (dsx) gene resulted in suppressed reproductive capacity in females which were homozygous for the construct. This was because their previous studies had strongly suggested that intron 4 was spliced mainly in males, as indicated by a fluorescent reporter construct designed to be activated by the splicing of intron 4.

[0014] The inventors generated the gene drive construct of the first aspect such that it targets the splice acceptor site at the 5' boundary of exon 5 of dsx, and were surprised to observe that, in stark contrast to all previous demonstrations of gene drive, no resistance was selected after release into caged populations of the mosquito. Moreover, additional experiments that were designed to reveal rare instances of resistance that were not selected in caged experiments also surprisingly failed to detect putative resistant mutations, thereby indicating that all mutations that were generated did not restore dsx function. The inventors have demonstrated that disruption of a female-specific exon (exon 5) of dsx leads to incomplete sexual dimorphism in females, but not males. When female mosquitoes carry this mutation in homozygosity, they display a range of mutant attributes including the inability to produce ovaries and biting mouthparts—an advantageous outcome that is optimally suited for a gene drive aimed at population suppression.

[0015] The inventors have therefore demonstrated that the gene drive construct of the invention can be used to spread through, replace and ultimately suppress any arthropod population by using the ultra-conserved, ultra-constrained sites found in different species at the intron/exon boundary of the female specific exon. The development of the gene drive construct of the invention which is capable of collapsing a human malaria vector population is a long sought scientific and technical achievement. The inventors describe herein a gene drive solution that shows a number of desired efficacy features for field applications in term of inheritance bias, fertility of heterozygous carrier individuals, phenotype of homozygous females and lack of nuclease-resistant functional variants at the target site. Advantageously, these results open a new phase in the effort to develop novel vector control measures and will stimulate unprecedented interest in the scientific community as well as among both policy makers and the general public.

[0016] Furthermore, the inventors believe that the results disclosed herein will have implications well beyond the field of malaria vector control, i.e. *A. gambiae*. The highly conserved functional role of dsx for sex determination in all insect species so far analysed and the high degree of sequence conservation amongst members of the same species in regions involved in sex specific splicing suggests that these sequences represent an Achilles heel for similar gene drive solutions aimed at targeting other vector species and agricultural pests.

[0017] It will be appreciated that suppression of a female's reproductive capacity can relate to a reduced ability of the female of the specific to procreate, or complete sterility of the female. Preferably, the reproductive capacity of the female homozygous for the construct is reduced by at least 5%, 10%, 20% or 30% compared to the corresponding wild type female. More preferably, the reproductive capacity of the female homozygous for the construct is reduced by at least 40%, 50% or 60% compared to the corresponding wild type female. Most preferably, the reproductive capacity of the female homozygous for the construct is reduced by at least 70%, 80%, 90% or 95% compared to the corresponding wild type female. Most preferably, suppression of a female's reproductive results in complete sterility of the female.

[0018] The skilled person will appreciate that the gene drive construct of the invention may relate to a construct comprising one or more genetic elements that biases its inheritance above that of Mendelian genetics, and thus increases in its frequency within a population over a number of generations.

[0019] Suitable arthropods which may be targeted using the gene drive genetic construct of the invention include insects, arachnids, myriapods or crustaceans. Preferably, the arthropod is an insect. Preferably, the arthropod, and most preferably the insect, is a disease-carrying vector or pest (e.g. agricultural pest), which can infect, cause harm to, or kill, an animal or plant of agricultural value, for example, Anopheline species, *Aedes* species (as a disease vector), *Ceratitis capitata*, or *Drosophila* species (as an agricultural pest).

[0020] Preferably, the insect is a mosquito. Preferably, the mosquito is of the subfamily Anophelinae. Preferably, the mosquito is selected from a group consisting of: *Anopheles gambiae*; *Anopheles coluzzi*; *Anopheles merus*; *Anopheles arabiensis*; *Anopheles quadriannulatus*; *Anopheles stephensi*; *Anopheles arabiensis*; *Anopheles funestus*; and *Anopheles melas*.

[0021] Most preferably, the mosquito is *Anopheles gambiae*.

[0022] The sequence of the doublesex gene in various arthropods, insects, and mosquito species are publicly available and so known to the skilled

person. However, in a preferred embodiment, the doublesex gene is from *Anopheles gambiae* (referred to as AGAP004050), which is provided herein as SEQ ID No: 1. SEQ ID No:1 is the whole AGAP004050 gene, plus about 3000 bp upstream of its putative promoter and about 4000 bp downstream of its putative terminator.

[0023] Accordingly, preferably the doublesex gene comprises a nucleic acid sequence substantially as set out in SEQ ID NO: 1, or a fragment or variant thereof.

[0024] Preferably, however, the intron-exon boundary targeted by the genetic construct of the invention is the boundary between intron 4 and exon 5 of the doublesex gene. In an embodiment, the intron 4-exon 5 boundary of the doublesex gene is provided herein as SEQ ID No: 2, as follows:

TABLE-US-00001 [SEQ ID No: 2] CCTTTCCATTTCATTTATGTTTAACACAGGTCAAGCGGTGGTCAACGAAT  
ACTCACGATTGCATAATCTGAACATGTTTGATGGCGTGAGTTGCGCAA TACCACCCGTCAGAGTGGATGATAAACTTTC

[0025] Accordingly, preferably genetic construct targets a nucleic acid sequence comprising or consisting of the nucleotide sequence substantially as set out in SEQ ID NO: 2, or a fragment or variant thereof. In some embodiments, the genetic construct targets a nucleic acid sequence comprising or consisting of the nucleotide sequence substantially as set out in SEQ ID NO: 2, or a fragment or variant thereof. The target sequence may include up to 1, 2, 3, 4, 5, 10 or 15 nucleotides 5' and/or 3' of SEQ ID No:2.

[0026] In a preferred embodiment, the intron 4-exon 5 boundary of the doublesex gene targeted by the gene drive construct is provided herein as SEQ ID No: 3, as follows:

TABLE-US-00002 [SEQ ID No: 3] CCTTTCCATTTCATTTATGTTTAACACAGGTCAAGCGGTGGTCAACGAAT ACTCA

[0027] Accordingly, preferably the genetic construct targets a nucleic acid sequence comprising or consisting of the nucleotide sequence substantially as set out in SEQ ID NO: 3, or a fragment or variant thereof. In some embodiments, the genetic construct targets a nucleic acid sequence comprising or consisting of the nucleotide sequence substantially as set out in SEQ ID NO: 3, or a fragment or variant thereof. The target sequence may include up to 1, 2, 3, 4, 5, 10 or 15 nucleotides 5' and/or 3' of SEQ ID No:3.

[0028] In a most preferred embodiment, the intron 4-exon 5 boundary of the doublesex gene targeted by the gene drive construct is provided herein as SEQ ID No: 4, as follows:

TABLE-US-00003 [SEQ ID No: 4] GTTTAACACAGGTCAAGCGGTGG

[0029] Accordingly, most preferably the genetic construct targets a nucleic acid sequence comprising or consisting of the nucleotide sequence substantially as set out in SEQ ID NO: 4, or a fragment or variant thereof. In some embodiments, the genetic construct targets a nucleic acid sequence comprising or consisting of the nucleotide sequence substantially as set out in SEQ ID NO: 4, or a fragment or variant thereof. The target sequence may include up to 1, 2, 3, 4, 5, 10 or 15 nucleotides 5' and/or 3' of SEQ ID No:4.

[0030] The concept of gene drive genetic constructs is known to those skilled in the art. Preferably, the gene drive genetic construct is a nuclease-based genetic construct. The gene drive genetic construct may be selected from a group consisting of: a transcription activator-like effector nuclease (TALEN) genetic construct; Zinc finger nuclease (ZFN) genetic construct; and a CRISPR-based gene drive genetic construct. Preferably, the genetic construct is a CRISPR-based gene drive construct, most preferably a CRISPR-Cpf1-based or CRISPR-Cas9-based gene drive genetic construct. However, it will be appreciated that other nucleases used in CRISPR-based genomic engineering methods are known and may be used in accordance with the invention.

[0031] Accordingly, in an embodiment in which the genetic construct is a CRISPR-based gene drive genetic construct, the genetic construct comprises a first nucleotide sequence encoding a nucleotide sequence that is capable of hybridising to the intron-exon boundary of the doublesex (dsx) gene, preferably with the objective to disrupt or destroy the female specific splice form. Preferably, the nucleotide sequence encoded by the first nucleotide sequence which is capable of hybridising to the intron-exon boundary of the doublesex (dsx) gene is a guide RNA. Preferably, the guide RNA is at least 16 base pairs in length. Preferably, the guide RNA is between 16 and 30 base pairs in length, more preferably between 18 and 25 base pairs in length.

[0032] Preferably, the CRISPR-based gene drive genetic construct further comprises a second nucleotide sequence encoding a CRISPR nuclease, preferably a Cpf1 or Cas9 nuclease, and most preferably a Cas9 nuclease. The sequences of the CRISPR nuclease and encoding nucleotides are known in the art. The first and second nucleotide sequences may be on separate nucleic acid molecules forming two genetic constructs, which act in tandem (i.e. in trans) as the gene drive genetic construct of the invention. Preferably, however, the first and second nucleotide sequences are on, or form part of, the same nucleic acid molecule, thereby creating the gene drive genetic construct of the invention. Preferably, the second nucleotide sequence encoding the nuclease is disposed 5' of the first nucleotide sequence encoding a nucleotide sequence that is capable of hybridising to the intron-exon boundary of the doublesex (dsx) gene.

[0033] In a preferred embodiment, the first nucleotide sequence encoding a nucleotide sequence that is capable of hybridising to the intron-exon boundary of the doublesex (dsx) gene (i.e. a guide RNA component) is provided herein as SEQ ID No: 5, as follows:

TABLE-US-00004 [SEQ ID No: 5] GTTTAACACAGGTCAAGCGG

[0034] Accordingly, preferably the first nucleotide sequence encoding a nucleotide sequence that is capable of hybridising to the intron-exon boundary of the doublesex (dsx) gene comprises a nucleic acid sequence substantially as set out in SEQ ID NO: 5, or a fragment or variant thereof.

[0035] The part of the nucleotide sequence that is capable of hybridising to the intron-exon boundary (i.e. the guide RNA) is known as a protospacer. In order for the nuclease to function, it also requires a specific protospacer adjacent motif (PAM) that varies depending on the bacterial species of the nuclease encoding gene. The most commonly used Cas9 nuclease recognizes a PAM sequence of NGG that is found directly downstream of the target sequence in the genomic DNA on the non-target strand. Recognition of the PAM by the nuclease is believed to destabilise the adjacent sequence, allowing interrogation of the sequence by the guide RNA, and resulting in RNA-DNA pairing when a matching sequence is present. The PAM is not present in the guide RNA sequence, but needs to be immediately downstream of the target site in the genomic DNA.

[0036] The skilled person would understand that the nucleotide sequence (i.e. guide RNA) that is capable of hybridising to the intron-exon boundary of the doublesex (dsx) gene may further comprise a CRISPR nuclease binding sequence, preferably a Cpf1 or Cas9 nuclease binding sequence, and most preferably a Cas9 nuclease binding sequence.

[0037] The CRISPR nuclease binding sequence creates a secondary binding structure which complexes with the nuclease, for example a hairpin loop. The PAM on the host genome is recognised by the nuclease.

[0038] Accordingly, in a preferred embodiment, the first nucleotide sequence encoding a nucleotide sequence that is capable of hybridising to the intron-exon boundary of the doublesex (dsx) gene (i.e. a guide RNA) is provided herein as SEQ ID No: 6, as follows:

TABLE-US-00005 [SEQ ID No: 6] GTTTAACACAGGTCAAGCGGTTTTAGAGCTAGAAATAGCAAGTTAAAA  
TAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT

[0039] Accordingly, preferably the first nucleotide sequence encoding a nucleotide sequence that is capable of hybridising to the intron-exon boundary of the doublesex (dsx) gene comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID NO: 6, or a fragment or variant thereof. The underlined sequence denotes the spacer, which encodes the nucleotide which hybridises to the dsx target site (i.e. SEQ ID No:5), and the rest of the gRNA backbone necessary for complexing with the nuclease, i.e. it encodes the CRISPR nuclease binding sequence

[0040] In one embodiment, the nucleotide sequence which is encoded by the first nucleotide sequence and which is capable of hybridising to the intron-exon boundary of the doublesex (dsx) gene (i.e. a guide RNA component) is provided herein as SEQ ID No: 58, as follows:

TABLE-US-00006 [SEQ ID No: 58] GUUUAACACAGGUCAAGCGG

[0041] Accordingly, preferably the nucleotide sequence which is encoded by the first nucleotide sequence and which is capable of hybridising to the intron-exon boundary of the doublesex (dsx) gene (i.e. a guide RNA) comprises a nucleic acid sequence substantially as set out in SEQ ID NO: 58,

or a fragment or variant thereof.

[0042] In one embodiment, the nucleotide sequence which is encoded by the first nucleotide sequence and which is capable of hybridising to the intron-exon boundary of the doublesex (dsx) gene (i.e. a guide RNA) is provided herein as SEQ ID No: 48, as follows:

TABLE-US-00007 [SEQ ID No: 48] GUUUAACACAGGUCAAGCGGUUUUAGAGCUAGAAUAGCAAGUUAUUAAU  
AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCU

[0043] Accordingly, preferably the nucleotide sequence which is encoded by the first nucleotide sequence and which is capable of hybridising to the intron-exon boundary of the doublesex (dsx) gene (i.e. a guide RNA) comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID NO: 48, or a fragment or variant thereof.

[0044] The CRISPR-based gene drive genetic construct further comprises at least one promoter sequence, which drives expression of the first and second nucleotide sequence. In other words, expression of the first and second nucleotide sequences is under the control of the same promoter.

[0045] Preferably, however, the CRISPR-based gene drive genetic construct comprises at least two promoter sequences, such that expression of the first and second nucleotide sequence is under the control of separate promoters. Preferably, therefore, the construct comprises a first promoter sequence operably linked to the first nucleotide sequence and a second promoter sequence operably linked to the second nucleotide sequence. The first and second promoter sequence may be any promoter sequence that is suitable for expression in an arthropod, and which would be known to those skilled in the art. Accordingly, the guide RNA is preferably expressed under control of the first promoter, and the nuclease is expressed under control of the second promoter.

[0046] Preferably, the first promoter is a polymerase III promoter, and most preferably a polymerase III promoter which does not add a 5'cap or a 3'polyA tail. More preferably, the promoter is a U6 promoter.

[0047] One embodiment of a nucleotide sequence of a U6 promoter is provided herein as SEQ ID No: 49, as follows:

TABLE-US-00008 [SEQ ID No: 49] TTTGTATGCGTGCGCTTGAAGGGTTGATCGGAACCTTACAACAGTTGTAG  
CTATACGGCTGCGTGTGGCTTCTAACGTTATCCATCGCTAGAAGTGAAC  
GAATGTGCGTAGGTATATATGAAATGGAGTTGCTCTCTGCT

[0048] Accordingly, preferably the first promoter sequence comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID No: 49, or a variant or fragment thereof.

[0049] Preferably, the second promoter sequence is a promoter sequence that substantially restricts expression of the second nucleotide sequence to germline cells of the arthropod. For example, the second promoter sequence may be selected from a group consisting of: zpg; nos; exu; and vasa2.

[0050] In one preferred embodiment, the second promoter sequence is referred to as “zero population growth” or “zpg”, and is provided herein as SEQ ID No: 7, as follows:

TABLE-US-00009 [SEQ ID No: 7] CAGCGCTGGCGGTGGGGACAGCTCCGGCTGTGGCTGTTCTTGCGAGTCCT  
CTTCCTGCGGCACATCCCTCTCGTCGACCAGTTCAGTTTGCTGAGCGTAA  
GCCTGCTGCTGTTTCGTCCTGCATCATCGGGACCATTTGTATGGGCCATCC  
GCCACCACCACCATCACCACCGCCGTCATTTCTAGGGGCATACCCATCA  
GCATCTCCGCGGGCGCCATTGGCGGTGGTGCCAAGGTGCCATTCTGTTTGT  
TGCTGAAAGCAAAGAAAGCAAATTAGTGTGTTTCTGCTGCACACGATA  
ATTTTCGTTTCTTGCCGCTAGACACAAACAACACTGCATCTGGAGGGAGA  
AATTTGACGCCTAGCTGTATACTTACCTCAAAGTTATTGTCCATCGTGG  
TATAATGGACCTACCGAGCCCGGTTACACTACACAAAGCAAGATTATGCG  
ACAAAATCACAGCGAAACATAGTAATTTTCATCTATCGAAAGCGGCCGAG  
CAGAGAGTTGTTTGGTATTGCAACTTGACATTCTGCTGCGGGATAAACCG  
CGACGGGCTACCATGGCGCACCTGTCAGATGGCTGTCAAATTTGGCCCGG  
TTTGCGATATGGAGTGGGTGAAATTATATCCCACTCGCTGATCGTGAAAA  
TAGACACCTGAAAACAATAATTGTTGTGTTAATTTACATTTTGAAGAAC  
AGCACAAGTTTTCGCTGACAATATTTAATTACGTTTCGTTATCAACGGCAC  
GGAAAGATTATCTCGCTGATTATCCCTCTCGCTCTCTGTCTATCATGT  
CCTGGTCTGTTCTCGCGTACCCCGGATAATCGAGAGACGCCATTTTAAAT  
TTGAACTACTACCCGACAAGCATGCCGTGAGCTCTTTCAAGTTCCTCTG  
TCCGACCAAAGAAACAGAGAATACCGCCCGGACAGTGCCCGGAGTGATCG  
ATCCATAGAAAATCGCCCATCATGTGCCACTGAGGCGAACCGGCGTAGCT  
TGTTCCGAATTTCCAAGTGCTTCCCGTAACATCCGCATATAACAAACAG CCAACAACAATAACAGCATCGAG

[0051] Accordingly, preferably the second promoter sequence comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID No: 7, or a variant or fragment thereof.

[0052] In another preferred embodiment, the second promoter sequence is referred to as “nanos” or “nos”, and is provided herein as SEQ ID No: 8, as follows:

TABLE-US-00010 [SEQ ID No: 8] GTGAACTTCCATGGAATTACGTGCTTTTTTCGGAATGGAGTTGGGCTGGTG  
AAAAACACCTATCAGCACCGCACTTTTCCCCGGCATTTACAGTTATACG  
CAGAGACAGAGACTAAATATTCACCCATTCATCAGCACTAAGTTCGCAA  
TAGATTGATATCCAAAACCTTTCTTACCTTTGCCGAGTTGGATTCTGGA  
TTCTGAGACTGTAAAAAGTCGTACGAGCTATCATAGGGTGTAACCGGAA  
AACAAACAAACGTTTAAATGGACTGCTCCAAGTGAATCGCTTCACGCAAA  
CAACACACACGCGCTGGGAGCGTTCCTGGCGTCACCTTTGCACGATGAA  
AACTGTAGCAAAACTCGCAGCACCGAAGGCTCTCCGTCCCTGCTGGTGTG  
TGTTTTTTCTTTTCTGACGCAAAATTAGAAAACATCATCATTTGACGAA  
AACGTCAACTGCGCGAGCAGAGTGACCAGAAATACCGATGTATCTGTATA  
GTAGAACGTCGGTTATCCGGGGGCGGATTAACCGTGCGCACACCAGTTT  
TTTGTGCAGCTTTGTAGTGCTAGTGATTTTTCGAAATTCATTTTGT  
CATTACAGTTGTAAACCTATAGTTATTGATTAATAATATCTACTA  
ACGATTAACCGATGGATTCAAAGTGAATAAATTATGAACTAGTGATTTT  
TTTAAATTTTATATGAATTTGACATTTCTTGACCATTCATCTTGGT  
CTCGAGCTGCCCGAATAATCGACGTTCTACTGTATTCCTACCGATTTT  
ATATGCCTACCGACACACAGGTGGGCCCCCTAAACTACCGATTTTAAAT  
TTATCCTACCGAAAATCACAGATTGTTTCATAATACAGACCAAAAAGTCA  
TGTAACCATTTCCCAAATCACTTAATGTATTAACCTCCATATGGAAATCG  
CTAGCAACCAGAACCAGAAGTTCAACAGAGACAACCAATTTCCGTGTATG  
TACTTCATGAGATGAGATTGGACGCGCTGGTAAAATTTATATGGGATTT  
GACAGATAATGTAAGGCGTGCGATTTTTTTCATACGATGGAATCAATTCA

AGAGTCAATTGTGTCAGGATTTATGAGAAAACAATCTCTTATTTATGTTTTGT  
TATCGTTACAGTTACAGCCCTGTCTTAAGCGGCCGCGTGAAGGCCCAAAA  
AAAAGGGAGTCCCCAACGCTCAGTAGCAAATGTGCTTCTCTATCATTTCTG  
TGGGTTAGAAAAGCCTCATGTGACTTCTATGAACAAAATCTAAACTATCT  
CCTTTAAATAGAGAATGGATGTATTTTTTCGTGCCACTGAACCTTCGTTG  
GGAAGATTAGATACCTCTCCCTCCCCCCCCCTCCCTTTCAACACTTCAAA  
ACCTACCGAAAACTACCGATACAATTTGATGTACCTACCGAAGACCGCCA  
AAATAATCTGGCCACACTGGCTAGATCTGATGTTTTGAAACATCGCCAAA  
TTTTACTAAATAATGCACTTGCGCGTTGGTGAAGCTGCACTTAAACAGAT  
TAGTTGAATTACGCTTTCTGAAATGTTTTTATTAACACTTGTTTTTTTT  
AATACTTCAATTTAAAGCTACTTCTTGGAATGATAATTCTACCCAAAACC  
AAAACCACTTTACAAAGAGTGTGTGGTTGGTGATCGCGCCGGCTACTGCG  
ACCTGTGGTCATCGCTCATCTCACGCACACATACGCACACATCTGTCAAT  
TGAAAAGCTGCACACAATCGTGTGTTGTGCAAAAAACCGTTTCGCGCACA  
ACAGTTTCGCACATGTTTGCAAGCCGTGCAGCAAAGGGCTTTTGATGGTGA  
TCCGCAGTGTGTTGGTCAGCTTTTTAATGTGTTTTCGCTTAATCGCTTTTG  
TTTGTGTAATGTTTTGTGCGAATAATTTTTATGCGTCGTTACAAATGAAA  
TGTACAATCCTGCGATGCTAGTGTAACCATTTGCTAATTCCCGGTAAGAA  
CGTTCATTACGCTCGGATATCATCTTACGAAGCGTGTGTATGTGCGCTAG  
TACATTGACCTTTAAAGTGATCCTTTTGTTCTAGAAAGCAAG

[0053] Accordingly, preferably the second promoter sequence comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID No: 8, or a variant or fragment thereof.

[0054] In a further preferred embodiment, the second promoter sequence is referred to as “exuperantia” or “exu”, and is provided herein as SEQ ID No: 9, as follows:

TABLE-US-00011 [SEQ ID No: 9] GGAAGGTGATTGCGATTCCATGTTGATGCCAATATATGATGATTTTGTG  
CATATTAATAGTTGTTGTTATGTTTTATTCAAATTTCAAAGATAATTTAC  
TTTACATTACAGTTAGTGAGCATATTATCTACTACATAAACACATAGATC  
AAACTGGTTTACATAAATTTAAAAAGTTTGGATTAAAATCGCAGCAATTG  
GTTATGAAAAAATATGTGCATAACGTAAATATCAAGTAAATTTTTGCATT  
GCATATTTATAGACTCCTGTTTACAATTTTCGGAAAAATGAAAAATGTTAAT  
TAATCAAAGAAGAAAAAACAAAGAAATTAATCATTAGGTAGCACAAACCA  
CAAGTACATATTTTTATGGCATGAATATTCCTCTACACTAACATATTTTA  
TAGCAATTCTATTGATCGCCTTAGTATAGCGGAATTACCAGAACGGCACT  
ATAGTTGTCTCTGTTTGGCACACGCAATCATTTTCATCCCAGGGTTGCC  
ATAGCAGTTTGGCGACGGTCACGTAGCATGCGAAGGATTTTCGTTTCGCACA  
GGATCACTTTTATTCTAACGTTTGAAGAAGGCACATCTCAGTGCAAGCGC  
TCTGGAAGCTGCTTTTACCGAACGAACCTAATTTTCAAGTAACCTCAAAA  
ACTTGTCTCTAACGCAACCGACGTCGTATCCGCGAGTTTCATTTCCCGTGC  
AAAGTTCCCCGATTTAGCTATCATTTCGTGAACATTTTCGTAGTGCTCTAC  
CCTCAGGTAAGACCATTTCGAGGTTTACCAAGTTTTGTGCAAAGAACGTGC  
ACAGTAATTTTCGTTCTGGTGAAACCTTCTCTTGTGTAGCTTGTACAAA

[0055] Accordingly, preferably the second promoter sequence comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID No: 9, or a variant or fragment thereof.

[0056] In a still further preferred embodiment, the second promoter sequence is referred to as “vasa2”, and is provided herein as SEQ ID No: 10, as follows:

TABLE-US-00012 [SEQ ID No: 10] ATGTAGAACGCGAGCAAATTCCTTTTCCCTCCATGACAGCAGCAGCTACAG  
TGGGAAGCCGAACGTCAGACGTGTTTGACATGCCGAACCTGGGCGGGAAAA  
TTACAGCGTGCGCTTTGTTTTCAAGCAAATCACAACCTCGCTGCAACAAA  
ACCGTTGAGAAATTGATTGTTTTATAATTTGTATTGATTTTATTTGTTA  
TAATAAACTAAAAAGACATACTTTTGCATATTTTATACATAAAAAACATA  
CATGCAGCATTATAAAACACATATAAACCCCTCCCTGTAGAGTCCCGTATC  
GAAATCTTCCATCCTAGTTGCACAGTACGACGGACGAGTAGGCCGTGTCC  
GTGCAAAATCCAGCTTTTAGCAGTCTTTTGCTCGGAGCACTCGCGGCGAG  
TCGGAGGTTTCTGCTGAGGTGCTTAGCGCTAAATTAGCCAATTGCTTTTG  
CAAGTGAAATAACCAGCCGAATAGTACTTCAAAAACCTCAGGTAAGTGAAC  
AGTTTTATAGAACAAATGTTTGTTTGTTAGAAGTAGTGAAGTGTGTTGTG  
AAAAAAATCTCTCATTTTCGGCAAAACTAACGTAAGTGAATTTCAAATTGAA  
TTATTGTTTTGTGATGTTATATTATTCATCCAGTTGATTAGTATTTTCT  
TAGTTATGTTCAAAATACAGTTAAATTAATTTCAATTTCAATTTACTCATA  
AAATAATCTCTTGGCTTATTTAATTTTTCTCGAATTCGCTTGTATTGTTT  
AGTAGCACGCGCATTCGCCCTTTGTTTCATTTGTACCTGCTCCCACTA  
ACACAGTGGCAGTGCGAAGCAAAAAGCCTTCGCACGCGTTGCTGGTATTAG  
AGTGTGTGCGTGTGTGTGTTGAGCGCTCTGTCAAAAATCGGCTGTTGCCGC  
CGGTACCGAAATTGCCTGTTTCGCACGCTGTTTCGTAAACATTCCGTGGTGT  
GTATCGTGTGTTGTGCATGTTGCGCGCCTCCCCCTTTTGATAGCAGGCT  
GCCGTGGCTGCCGTGGTGTGTGGCGCAGTTGAGTTTTTGGAATTAATTTT  
TAAGGAAATGGCACGAGAAGAGCGGTGGCAGTGTGTTGGTTTGCTCTGTC  
CCTTCCTTTCTGTGTGAAGTGTCTTACAGCACAGCACGTATCCACCACC  
GCACACAGAGCAGGCAAGGAAGTGGAAGTGAACAAGTGTGCTGCGCATGC  
ATGTGTGTGGGGGCAATTTAGCTGAGATCGTCGTTATTTGAGAAGCGGT  
ATAGGGGCCAGTCGGTGTGACGTACGGAAGCGGTTTAGTTTTAATCCAA  
GCGTATCCCGTCGTGGAGTGGTTGTGTGGCTCTGTGTGCTCTCATATCAG  
TTCCAGAGTGAGGTTAGTAGAATCACAGTCCTTGGCCTTTTTTCGTTACAA  
GATATCCAGAAGGATGGCGTTATTTCCACAGCTTACCATGGTGCTCTTGT  
TTGCTCGAATCAGGGGAGAAAAACAGTTTCGTGTTTCATGAACCGCAGTT

ATCCTGTGTAAGACGGCAATGCAAAGATAGTGTGCTCAACTTCTCCGCGA  
AGATTGAAGCTAAATTAAGCACGAGATTAGCATGACTGAAGTGACTTTTC  
AAAGTGTGAGAATGGCTGCACTCGCAAAGCTAGCTGGATGCAGCGCAATTT  
TGCCCCGGTGTGTGCGCGCATGCAAACGAGCAACCGCAGAGGGCAAAGGA  
GAGGATGGGAAGGAGGGGAGGGAGTGAAAGAGCAGGCTTAAGGTTGCCCTC  
GGGCATTGAAGTCGATACAGCGGTTCTATTCCAGTGCCAGTAACGATGAC  
GAAGACGATGTTGCTTCTGCTGCTGTTGCTGCTGTTGTTGTGATGATGA  
TGATGATAATAGTGCAAATATAAAATAAATCTTCCGTAAGCTTTGTGTAG  
TGGTGCGTGGCTACTATAAGCCCGTCTGGAAGCAAGGAAGCTAGTCGGGC  
AGGGTCATGCAAAAGGGAGACACCTTCGGAGCTCCGGAGCTCCCGCCGGC  
ACTCTCGGGGGGACGTCCGTTATGCGTTGTGATTTATTATGGAATATTTA  
TTATAGTGTCTTGTTTTGAAAAATAAATTC AACGGTTCGAATTTCTCTAC  
ACCTCGAGATCGGGGCTGGAGTGGCAACGTGGTACGGAACGGTACAGCGG  
TTTGAGCCGTTCCGTCCTTGGGACTCACGGATCGCAGAATGTTATTGTGCG  
CGCACTGATGGGAAAGTCAATTTTACCAGAGTGGTACGGCCGCTAGTCC  
AGTTCGTTTCTGGCTGCTGTTGCTGATGCTACGATCCCTCAGGAATGATTG  
GAAACGCCTGGAGATGGTGGGAAAAAATCAAACACAAAAACGATCCTAAT  
GAACATCGTGTGTTCTCATTGCTGCCACGATTGACACCTTCGATAAGAC  
GCACATAATGAGCTAAAGGAGAGGGGACAGGGTCTTGCTTTTGCCACGAG  
CGATAAGATTGCAATCACTCGTGAGCGTGTGCTGCTGGGCTGAAGAAGAA  
ACGCTTTCCACAGCAGTAGGTGGGAAGTGGGATTGTGGAACGTGGCATTG  
AAAAGAACCTATTTTCTAAAGCCCGAGAGCCCGTTCTCGAACTGGAAAAC  
CAGATGCAGAAAGTTTTTTATTGTCCCCCGCCAGGAAAAACAATGTATTTA  
ATGCTTTCTTTGCTTTTTTCCCGCGTTTCAGACGACGAGTGTGTAAGC  
GAGCCCAATGGCTGTTGGAGAAACTCGGCTACCCGTGGGAGATGATGCC  
CTGATGTACGTCATACTAAAGAGCGCCGATGGCGATGTACAAAAAGCACA  
CCAGCGGATCGACGAAGGTAAGCTGGCGATGATGGTGTGCTTCGACATCA  
CTTTCATCACCGTGTGACACATCTACTGTGCCTAGCACCGGGTCCAGTGG  
TCACAGGGTGTAGCAAAAACGTGTTCTTTTTTGCAGAGAGACTCTACCTCA  
TGATGCAGCTGTTAAGGAAAGGTTTCAGATGAAGGCAATTTTTCTAGGA  
TAAGATGATCTTAAGTTAACTGCGTATTAGTGTTTAAACATTGTCGTCTCA  
ACTCCCAAGAATGTTTTAACTGCTAGGGCTAGTTTATTATACATGTTCT  
CATTGAAATGTCGTTCAATCCAACATGTTAAGTTAGCTAGCTCAGACAGC  
AGAAGTTAGGAGTATCTGCATCTTGAAGGTAGCGGCATATGGTGTATGC

CACGTTCACTTCAATTCGATACAAAAAACCACATCAA  
AAACCAAATTGTGAATTCGTCAGCCAGCAGCAGTGACCTCAAAGCCTT ACCTTTCCATTCAATTTATGTTTAAACACAGGTCAAG  
[0063] Accordingly, preferably the 5' homology arm comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID No: 11, or a variant or fragment thereof.

[0064] In a preferred embodiment, the 3' homology arm is provided herein as SEQ ID No: 12, as follows:

TABLE-US-00014 [SEQ ID No: 12] CGGTGGTCAACGAATACTCAGATTGCATAATCTGAACATGTTTGATGGC  
GTGGAGTTGCGCAATACCACCCGTCAGAGTGGATGATAAACTTTCCGCAC  
CACTGTAAGTGTCCGTATCTTTGTATGTGGGTGTGTGTATGTGTGTTTGG  
TGAAACGAATTCAATAGTTCTGTGCTATTTTAAATCAAGCCGCGTGCGCA  
ACTGATGCCGATAAGTTCAAAGTGTGTTTAAAGGAGTGGAGCGAGAGAGC  
CGCACCACGGTACAGAAGGGCAGCAGAATGGGTCGGCAGCCTAGCTGCAC  
TGGTGGGTGCGTCCGGCGTCTCGGGGGGAGGGCGAGGAAATTCTAGTGT  
TAAATCGGAGCAGCAAAAAACAAACAGTGGTCGTCCCGTTCAAGAAACGG  
CCTGTACACACACACAGAAAACACTGCAGCATGTTTGTACATAGTAGATC  
CTAGAGCAGGTGGTCGTTGCTCCTCGAACGCTCTGGACGCACGGCTTCGC  
GCGTATTTGCGTAGCGTTCCGCCGATCGTGGGTATTCGTAAGTCCACAAG  
CCCGCTTTTCTCCCATGCAATCTCTGCAACCAACCAACAAACAACAACAA  
AAAACCAATCGACAAAATGAATCACACCCCTTTTGTATCATCTGTATATT  
CTTGTTCTTTGCGTTCCTTTCTATGTGGCCACGCCCGGGGTACGTA  
ATTGCGTCGAAAACCCCGAAAACCCCGGCACATACAGTGTACATACGGTT  
TGAGGACAACCTTTGACCTGCAGCCCTTCTGGGGTTGCCACGTGTAGCTAT  
ACTTGTGAGATCGGGCGGACGGTGTAAAGCGCGAATGGCCGCCACACA  
GTGTGTCCACTCCAACACTACCCCTCTGGAAGTACCCCGTCCAGGGATGC  
ACCGGCTCGGCTCATGCCCTGCAAAACAGTCCGGGCTCCACTGTAGTAG  
CTCCGGCGTTGCTCTGAGAGAAGGATGCCCTTCGAAGTGTGAAAGCGTG  
CATTGGGCGTTCAAGTGTGTGTGTGTGTGTAGGTTTAGCGAGAAACAGC  
AGCAGTTGCGTGTGCTGAAAAGCGAAGGAGTAATAGAGTGCATAATGAAA  
ATGAAAATGAAAATGAAGCAAAAGTAGAAGGCGGAGGAGAGCAACCTGTG  
TTCCACTAGTAGCAATAGTTTAGTCTAGTTTCGTCACCAATCAACCTTC  
CAACCATCGTTCAACCAATACCTGAGTCAACATCGTCATCGTTATCGTGC  
CACAACCTTTATTAATAATGAACCTTGTCCGCGCCACCGTAGGGTGATCTA  
AGGCGACCTTTCTTACGGGCGCGACCCACATGCCATCGTCACCTTCTCCA  
ATCAAAACCAACAGCCTGTACCGATGGTGTGCAATTGTGCGTGCGTGTGT  
GTTATTAGCAAAAAAGAGAAAGAGTCGACGAGAGAGATAGATCGAGA  
TCGAGAGTACAAAAGAGCAGTAGAAAATGTTCTGTTGTTTTCGTAAC  
ACAGTTGTTTAGCCAAAATGGGAATTTCCAATAATCCCGGGGGCGGGGAA  
ATGCGGGAATACTGCGTACACACATCAATCAAAAAGAAAAATCCTT  
GCGCTACATCACTACCGTTTTCGCGGGTGTGATCTAGAGCAGACCACTT  
CCACTCCACTCTACAATCAATCAATCTGTGCAGAAGGTATGGTAAGACGG  
CCTTTGAGCGAGTCACGGTCGCCACCATAACGCCGTCCGACGAGGGCTGA  
ATGCGAATTTGCTAATCGATTTTCCGCTTTCTTTTATCCACCTCCTT  
TTCTCTCCCTCTCTCTTTTGCAGTGGCCCTTGTAAACCCCAAAAAGGT  
AAACGACACATTAAGACCTACGAAGCGTTGGTGAAGTCATCGCTCGATCC  
GAACAGCGACCGGCTGACGGAGGACGACGAGGACGAGAACATCTCGG TGACCCGCACC

[0065] Accordingly, preferably the 3' homology arm comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID No: 12, or a variant or fragment thereof.

[0066] In another embodiment, the CRISPR-based gene drive construct may instead be inserted into the genome by homology-directed repair, i.e. without the use of a docking construct, as described above. Accordingly, preferably, the CRISPR-based gene drive genetic construct further comprises third and fourth nucleotide sequences which, respectively, flank the first nucleotide sequence encoding the nucleotide sequence that is capable of hybridising to the intron-exon boundary of the doublesex (dsx) gene, the second nucleotide sequence encoding the nuclease, the first promoter sequence and the second promoter sequence, wherein the third and fourth nucleotides are homologous to the genomic sequences flanking the intron-exon boundary, such that the gene drive construct is integrated into the genome via homology-directed repair.

[0067] Preferably, the third and fourth nucleotide sequences are at least 100 bp in length, at least 200 bp in length, at least 400 bp in length, at least 600 bp in length, at least 800 bp in length, at least 1000 bp in length at least 1200 bp in length, at least 1400 bp in length, at least 1600 bp in length, at least 1800 bp in length, at least 2000 bp in length. Preferably, the third and fourth nucleotide sequences are up to 4000 bp in length, up to 3000 bp in length, up to 2000 bp in length. Preferably, the third and fourth nucleotide sequences are between 100 and 4000 bp in length, more preferably between 150 and 3000 bp in length and most preferably between 200 and 2000 bp in length. Preferably, the third and fourth nucleotide sequences are about 2000 bp in length.

[0068] Accordingly, preferably the third nucleotide sequence comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID No: 11, or a variant or fragment thereof.

[0069] Accordingly, preferably the fourth nucleotide sequence comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID No: 12, or a variant or fragment thereof.

[0070] Preferably, the CRISPR-based gene drive construct targets the intron-4-exon 5 boundary of the doublesex gene.

[0071] In a preferred embodiment, the gene drive construct is provided herein as SEQ ID No: 13, as follows:

TABLE-US-00015 [SEQ ID No: 13]  
TGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCCACCTCACCCATGCGATCGCTCCGGAAGATACATTGATGAGTT  
TGGACAAACCACAAGTAAACAACAACAATTGCATTCAATTTATGTTTCAGGTTACAGGGGAGGTGTGGGAGGTTTTTAAAGCAAGTAAA  
ACCTCTACAAATGTGGTATGGCTGATTATGATCTAGAGTCGCGGCCGTACAGGAACAGGTGGTGGCGGCCCTCGGTGCGCTCGTACT  
GCTCCACGATGTTGTAGTCTGCTCGTTGTGGGAGGTGATGTCCAGCTTGGAGTCCACGTAGTAGCCGGGCGAGCTGCACGGGCTTCTT  
GGCCATGTAGATGGACTTGAAGTCCACCAGGTAGTGCCGCCGTCCTTCAGCTTCAGGGCCTTGTGGATCTCGCCCTTCAGCACGCCG  
TCGCGGGGTACAGGCGCTCGGTGGAGGCCTCCAGCCCATGGTCTTCTTCTGCATTACGGGGCGCTCGGAGGGGAAGTTCACGCCGA  
TGAAGTTCACCTTGTAGATGAAGCAGCGTCTCTGCAGGGAGGAGTCTTGGGTCACGGTCACCACGCCGCCGCTCTCGAAGTTCATCAC  
GCGCTCCCACTTGAAGCCCTCGGGGAAGGACAGCTTCTTGTAGTCGGGGATGTGCGCGGGGTGCTTCACGTACACCTTGGAGCCGTAC  
TGGAAGTGGGGGACAGGATGTCCAGGCGAAGGGCAGGGGGCCGCCCTTGGTCACCTTCAGCTTCACGGTGTGTGGCCCTCGTAGG

CGGCGCCCTCGCCCTCGCCCTCGATCTCGAAGCTCGTGCCCGTTACAGGTGCCCTCCATGCGCACCTTGAAGCGCATGAACCTCTTGAT  
GACGTTCTTGGAGGAGCGCACCATGGTGGCGACCTGTGGGTCCCGGGCCCCGGGTACCGTTCGACTCTAGCGGTACCCCGATTGTTTAG  
CTTGTTACAGCTGCGCTTGTATTGCTTAGCTTTCGCTTAGCGACGTGTTCACTTTGCTTGTGTTGAATTGAATTGTCGCTCCGTTAGA  
CGAAGCGCCTCTATTATATCTCCGGCGGTGAGGGTTCCGAATTCGATAAGCTTGGATCCTAATTGAATTAGCTCTAATTGAATTAGTC  
TCTAATTGAATTAGATCCCCGGGCGAGCTCGAATTAACCAATTGTGACACCGGTACGCGTGGCGGACAGCTGGCGGTGTGGCTG  
TCTTGTAGAGTCACTTTCCTGCGGACATCCCTCTCGTCACCAAGTTCAGTTTGTGAGCGTAAGCGTGTGCTGTTCTGCTCATC  
ATCGGGACCAATTGTACGGGGCCATCCGCCACCACCACCATCACCACCGCCGTCCATTTCTAGGGGCATACCCATCAGCATCTCCGCGG  
GCGCCATTGGCGGTGGTGCCAAGGTGCCATTTCGTTTGTGCTGAAAGCAAAAGAAAGCAAATTAGTGTGTTTCTGCTGCACACGATA  
GTTTTCGTTTCTTGGCGCTAGACACAAACAACACTGCATCTGGAGGGAGAAATTTGACGCCTAGCTGTATAACTTACCTCAAAGTTAT  
TGTCATCGTGGTATAATGGACCTACCGAGCCCGGTTACACTACACAAAGCAAGATTATGCGACAAAATCACAGCGAAAACCTAGTAAT  
TTTCATCTATCGAAAGCGGCCGAGCAGAGAGTTCGTTTGGTATTGCAACTTGACATTCTGCTGTGGGATAAACCCGCGACGGGCTACCAT  
GGCGACCTGTTCAGATGGCTGTCAAATTTGGCCGTTTTCGATATTGAGTGGGTGAAATTATATATCCCATCTGCTGATCGTGAATAA  
GACACCTGAAAACAATAATTGTTGTGTTAATTTTACATTTTGAAGAACAGCACAAGTTTTGCTGACAATATTTAATTACGTTTTCGTTA  
TCAACGGCACGGAAGATTATCTCGCTGATTATCCCTCTCGCTCTCTCTGTCTATCATGTCTCGTGGTCTCTCGCGTCACCCCGGATA  
ATCGAGAGACGCCATTTTAAATTTGAACTACTACACCGACAAGCATGCCGTGAGCTCTTTCAAGTTCTTCTGTCCGACCAAGAAACA  
GAGAATACCGCCCCGGACAGTGCCCGGAGTGATCGATCCATAGAAAATCGCCCATCATGTGCCACTGAAGCGAACC GGCGTAGCTTGT  
CCGAATTTCCAAAGTGCTTCCCCGTAACATCCGCATATAACAAGCAGCCCCAACAAATAACAGCATCGAGCTCGAGATGGACTATAAG  
GACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTTCGGTA  
TCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGA  
GTACAGGTTGCCAGCAAGAAATCAAGGTTGCTGGGCAACCCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTT  
CACAAGCGGCAAAACAGCCGAGGCCACCCGGCTGGAAGAGAACC GCCCAAGAAATACACAGCAAGGAAGAACCCGATCTGCTATCTGC  
AAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCTTCTGTTGGAAGAGGATAAGAA  
GCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAAA  
CTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTCCGGGGCCACTTCTGATCG  
AGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAAAACCC  
CATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGGTGAAAAATCTGATCGCCCGAGCTG  
CCCGCGAGAAGAAGATGGCTGTTTCGAAACCTGATTGCCCTGAGCTGGGCTGACCCCACTTCAAGAGCAACTTCGACCTGCG  
CCGAGGATGCCAACTGCAGCTGAGCAAGGACACCTACGACGACCTGGACAACCTGCTGGCCAGATCGGCCAGCAGTACGCGGA  
CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAGGCCCCCTG  
AGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGGCGCAGCAGCTGCCTGAGAAGT  
ACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCAT  
CAAGCCCATCCTGGAAGATGGACGGCACCGAGGAAGTCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTC  
GACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTACCCATTCTGAAAGG  
ACAACCGGGAAGATCGAGAAGATCCTGACCTTCCGCATCCCTTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTG  
CGTGACCAAGAAGAGCGAGGAACACCTACCCCTGGAATCTCGAGGAAGTGGTGACAAAGGCGCTTCCGCCAGAGCTTCATCGAG  
CGGATGACCAACTTCGATAAGAACCTGCCAACAGGAAGGTGCTGCCAACAGCAGCCTGTGTACAGTACTTCCCGGTATATAACG  
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAAGCCGCCCTTCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCT  
GTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGTTCGACTCCGTGGAATCTCC  
GGCGTGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCTGGACAATGAGG  
AAAACGAGGACATTCTGGAAGATATCGTGCTGACCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGC  
CCACCTGTTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGC  
ATCCGGGACAAGCAGTCCCGGACAAGACAATCCTGGATTTCTGAAGTCCGACGGCTTCGCCAACAGCAAACTTCATGCAGCTGATCCACG  
ACGACGCTGACCTTTAAAGAGGACATCTCAGAAAGCCAGGTGTCCGGCCAGGGCAGATACCTGACGACGACATTGCCAATTCGGC  
CGGACGCCCCGCCATTAAAGAAGGCGATCCTGCAGACAGTGAAGGTGGTGACGAGCTCGTGAAAGTGATGGGCGGCAACAGCCCGAG  
AACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCAGAAAGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGG  
GCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCT  
GCAGAATGGGCGGGATATGTACGTGGACCAGGAAGTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGC  
TTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACC GGGGCAAGAGCGACAACGTGCCCTCCGAAGAGG  
TCGTGAAGAAGATGAAGAACTAGTGGCGGCGAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGA  
GAGAGGCGGCTGAGCGCAATGGATAAGGCCGGCTTCATCAAGAGCAGCTGGTGGAAACCCGGCAGATCACAAGCAGCTGGCACAG  
ATCTTGACTCCCGGATGAACACTAAGTACGACGAGATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGG  
TGTCGATTTCCGGAAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAATAACCACCACGCCACGACGCTACCTGAACGCCGT  
CGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATG  
ATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTCAAGACCGAGATTA  
CCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATT  
TGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCT  
ATCCTGCCCAAGAGGAACAGCGTAAAGTGTACGCCAGAAAGAAGGACTGGGACCTTAAGAAGTACGGCGGCTTCGACAGCCCCACCG  
TGGCTATTCTGTGCTGGTGTGGCCAAAGTGGAAAGGGCAAGTCCAAGAACTGAAGAGTGTGAAGAGCTGTGGGATCCCATCCAT  
CATGGAAGAAGACAGCTGTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAGGAGGACCTGATCATCAAG  
CTGCCATAAGTACTCCCTGTTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACCTG  
CCCTGCCCTCCAAATATGTGAACCTTCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACA  
GCTGTTTGTGGAACAGCACAAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCT  
AATCTGGACAAAGTGCTGTCCGCTACAACAAGCACCGGGATAAGGCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCC  
TGACCAATCTGGGAGCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGA  
CGCCACCTTGATCCACCAGAGCATACCGGCCCTGTACGAGACCGGATGCACCTGTCTAGCTGGGAGGCGCAAAAAGCGGCGCGCC  
ACGAAAAAGCGCGCGGACGCAAAAAAGTAAATTAATTAAGAGGACGGCGAGGTAATCATATGTCCGATTTTGGCGCAAAAC  
AGCGCTTAGACAATTTGCGCGTAAGCACATTCGAAATGTGAAAAGCTGAAAGCAGTGAGTTTTCGCCAGCCCGAGTTTCAGCGAAACGGA  
TTCCTTCCAAGTGTTTGCATTCTTGGCGGAGTGTTCTTCCCAAATGCACTCACCTGCGTGCAGTGCCAAATCGTGAGTTTCTTAAT  
TTTTTCATATTGTTTATTACCTACCAACTAAAGTTGTTGTTATATATTGCGTTTTACGTACGACAAATAAGTTTCGTATTTCAGAAATAT  
TTGCGATAAGAGAGAACTCATTGCGATGAATCTCATTGTATTTAGCTAAGTGCCTTGATAAGTAAGCGGAACAGCAGGAATATGACA  
CTCCTTGGGAAATACATGTAAGCGTCTGTAATTAGATATATATACACGCAACCAAATGGTCCATGGTTGATTTAAGCACTGCCTGTTG  
TCGAACATTGCTATAAGCAAAATAAAGAAGCATTATTAATCTAAAAATTTCTTCAAAGTGACTTCAATGATGATCTCTAGGCTATAGT  
GAAAGCTGAAAGCTTTTGTGACAAATGCAAGGGAAGTGAAGCAGCTGCGTGCATGGGACCGCGCATCTATTCTCAGCTAATTC  
CCCTATCATATTGATAATTGACGGCAGGATTTCTGCTTCTTACTTCTTTACTTTTGGAGCTTTTCATCAATAAAACCACTACCATGGC  
CGTACGCTCAACGGAAGCATTCAAAAAAACCCGCGTTTCTCGTGTGATTTGTGGGTGAGTGGCGCCATCTATTAGAGAATAGCTGT





site may include up to 1, 2, 3, 4, 5, 10 or 15 nucleotides 5' and/or 3' of SEQ ID No:35.

[0080] As shown in FIG. 12, the second target site may be the sequence shown as T2, which is provided herein as SEQ ID No: 36, as follows:

TABLE-US-00017 [SEQ ID No: 36] TCTGAACATGTTTGATGGCGTGG

[0081] In one embodiment, therefore, the second target site comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID No: 36, or a variant or fragment thereof. In some embodiments, the genetic construct targets a second target site comprising or consisting of the nucleotide sequence substantially as set out in SEQ ID NO: 36, or a fragment or variant thereof. The second target site may include up to 1, 2, 3, 4, 5,

10 or 15 nucleotides 5' and/or 3' of SEQ ID No:36. As is shown in FIG. 12, T2 is wholly contained within exon 5.

[0082] The second target site may be the sequence shown as T3, which is provided herein as SEQ ID No: 37, as follows:

TABLE-US-00018 [SEQ ID No: 37] GCAATACCACCCGTCAGAGTGG

[0083] In one embodiment, therefore, the second target site comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID No: 37, or a variant or fragment thereof. In some embodiments, the genetic construct targets a second target site comprising or consisting of the nucleotide sequence substantially as set out in SEQ ID NO: 37, or a fragment or variant thereof. The second target site may include up to 1, 2, 3, 4, 5,

10 or 15 nucleotides 5' and/or 3' of SEQ ID No:37. As is shown in FIG. 12, T3 is wholly contained within exon 5.

[0084] The second target site may be the sequence shown as T4, which is provided herein as SEQ ID No: 38, as follows:

TABLE-US-00019 [SEQ ID No: 38] GTTTATCATCCACTCTGACGG

[0085] In one embodiment, therefore, the second target site comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID No: 38, or a variant or fragment thereof. In some embodiments, the genetic construct targets a second target site comprising or consisting of the nucleotide sequence substantially as set out in SEQ ID NO: 38, or a fragment or variant thereof. The second target site may include up to 1, 2, 3, 4, 5,

10 or 15 nucleotides 5' and/or 3' of SEQ ID No:38. As is shown in FIG. 12, T4 is partially in the 3' end of exon 5 and extends into the untranslated region of exon 5.

[0086] The gene drive construct of the invention may target one or more of a second target site selected from a group consisting of T2, T3 and T4.

Most preferably, the gene drive genetic construct of the invention targets T1 and one or more of T2, T3 and T4. For example, the construct may target T1 and T2, or T1 and T3, or T1 and T4, or T1, T2 and T3, T1, T2 and T4, or T1 and T3 and T4, or any combination thereof.

[0087] However, as described in the Examples and as shown in FIG. 13, preferably the gene drive genetic construct of the invention targets T1 and T3, which has been shown to be very effective.

[0088] Accordingly, in this embodiment in which the genetic construct is a CRISPR-based gene drive genetic construct, the construct comprises: (i) a first nucleotide sequence encoding a first guide RNA which is capable of hybridising to a first target site which is an intron-exon boundary of the female specific splice form of the doublesex (dsx) gene, and (ii) a fifth nucleotide sequence encoding a second guide RNA which is capable of hybridising to a second target site disposed in exon 5 of the female specific splice form of the doublesex (dsx) gene.

[0089] Preferably, the first and/or fifth nucleotide sequence encodes a guide RNA, most preferably separate guide RNA molecules. Preferably, each guide RNA is at least 16 base pairs in length. Preferably, each guide RNA is between 16 and 30 base pairs in length, more preferably between 18 and 25 base pairs in length.

[0090] As discussed herein, the second nucleotide sequence encodes a CRISPR nuclease, preferably a Cpf1 or Cas9 nuclease, most preferably a Cas9 nuclease, though other nuclease are known in the art.

[0091] The first, second and fifth nucleotide sequences may be on separate nucleic acid molecules. Preferably, however, the first, second and fifth nucleotide sequences are on, or form part of, the same nucleic acid molecule. Most preferably, the first, second and fifth nucleotide sequences are expressed separately. Preferably, the first nucleotide sequence is disposed 5' of the fifth nucleotide sequence. Preferably, the second nucleotide sequence encoding the nuclease is disposed 5' of the first and fifth nucleotide sequences.

[0092] In one embodiment, the fifth nucleotide sequence encoding a nucleotide sequence that is capable of hybridising to the second target site (i.e. T2 shown in FIG. 12) disposed in exon 5 of the female specific splice form of the doublesex (dsx) gene (i.e. the second guide RNA component) is provided herein as SEQ ID No: 39, as follows:

TABLE-US-00020 [SEQ ID No: 39] TCTGAACATGTTTGATGGCG

[0093] Accordingly, preferably the fifth nucleotide sequence encoding a nucleotide sequence that is capable of hybridising to the second target site comprises a nucleic acid sequence substantially as set out in SEQ ID NO: 39, or a fragment or variant thereof.

[0094] In another embodiment, the fifth nucleotide sequence encoding a nucleotide sequence that is capable of hybridising to the second target site (i.e. T3 shown in FIG. 12) disposed in exon 5 of the female specific splice form of the doublesex (dsx) gene (i.e. the second guide RNA component) is provided herein as SEQ ID No: 40, as follows:

TABLE-US-00021 [SEQ ID No: 40] GCAATACCACCCGTCAGAG

[0095] Accordingly, preferably the fifth nucleotide sequence encoding a nucleotide sequence that is capable of hybridising to the second target site comprises a nucleic acid sequence substantially as set out in SEQ ID NO: 40, or a fragment or variant thereof.

[0096] In yet another embodiment, the fifth nucleotide sequence encoding a nucleotide sequence that is capable of hybridising to the second target site (i.e. T4 shown in FIG. 12) disposed in exon 5 of the female specific splice form of the doublesex (dsx) gene (i.e. the second guide RNA component) is provided herein as SEQ ID No: 41, as follows:

TABLE-US-00022 [SEQ ID No: 41] GTTTATCATCCACTCTGA

[0097] Accordingly, preferably the fifth nucleotide sequence encoding a nucleotide sequence that is capable of hybridising to the second target site comprises a nucleic acid sequence substantially as set out in SEQ ID NO: 41, or a fragment or variant thereof.

[0098] The skilled person would understand that the nucleotide sequence (i.e. guide RNA) that is capable of hybridising to the second target site in the doublesex (dsx) gene may further comprise a CRISPR nuclease binding sequence, preferably a Cpf1 or Cas9 nuclease binding sequence, and most preferably a Cas9 nuclease binding sequence.

[0099] The CRISPR nuclease binding sequence creates a secondary binding structure which complexes with the nuclease, for example a hairpin loop.

[0100] Accordingly, in one preferred embodiment, the second nucleotide sequence encoding a nucleotide sequence that is capable of hybridising to the second target site (i.e. the second guide RNA targeting T2) is provided herein as SEQ ID No: 42, as follows:

TABLE-US-00023 [SEQ ID No: 42] TCTGAACATGTTTGATGGCGgttttagagctagaatagcaagttaaaa  
taaggctagtcggttatcaacttgaaaaagtgccaccgagtcggtgct

[0101] Accordingly, preferably the fifth nucleotide sequence encoding a nucleotide sequence that is capable of hybridising to the second target sequence comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID NO: 42, or a fragment or variant thereof.

[0102] In another preferred embodiment, the second nucleotide sequence encoding a nucleotide sequence that is capable of hybridising to the second target site (i.e. the second guide RNA targeting T3) is provided herein as SEQ ID No: 43, as follows:

TABLE-US-00024 [SEQ ID No: 43] GCAATACCACCCGTCAGAGgttttagagctagaatagcaagttaaaa  
aaggctagtcggttatcaacttgaaaaagtgccaccgagtcggtgct

[0103] Accordingly, preferably the fifth nucleotide sequence encoding a nucleotide sequence that is capable of hybridising to the second target sequence comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID NO: 43, or a fragment or variant thereof.

[0104] In a further preferred embodiment, the second nucleotide sequence encoding a nucleotide sequence that is capable of hybridising to the second target site (i.e. the second guide RNA targeting T4) is provided herein as SEQ ID No: 44, as follows:

TABLE-US-00025 [SEQ ID No: 44] **TTTATCATCCACTCTGA**gttttagagctagaatagcaagttaaata

aggtagtcggttatcaactgaaagtggcaccgagtcggtgct

[0105] Accordingly, preferably the fifth nucleotide sequence encoding a nucleotide sequence that is capable of hybridising to the second target sequence comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID NO: 44, or a fragment or variant thereof.

[0106] In one embodiment, the nucleotide sequence which is encoded by the fifth nucleotide sequence and which is capable of hybridising to the second target site (i.e. the second guide RNA targeting T2 component) is provided herein as SEQ ID No: 59, as follows:

TABLE-US-00026 [SEQ ID No: 59] **UCUGAACAUUUUGAUGGCG**

[0107] Accordingly, preferably the nucleotide sequence which is encoded by the fifth nucleotide sequence and which is capable of hybridising to the second target site (i.e. the second guide RNA targeting T2) comprises nucleic acid sequence substantially as set out in SEQ ID NO: 59, or a fragment or variant thereof.

[0108] In one embodiment, the nucleotide sequence which is encoded by the fifth nucleotide sequence and which is capable of hybridising to the second target site (i.e. the second guide RNA targeting T2) is provided herein as SEQ ID No: 45, as follows:

TABLE-US-00027 [SEQ ID No: 45] **UCUGAACAUUUUGAUGGCGUUUUAGAGCUAGAAAUAGCAAGUAAAA**  
**UAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCU**

[0109] Accordingly, preferably the nucleotide sequence which is encoded by the fifth nucleotide sequence and which is capable of hybridising to the second target site (i.e. the second guide RNA targeting T2) comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID NO: 45, or a fragment or variant thereof.

[0110] In another embodiment, the nucleotide sequence which is encoded by the fifth nucleotide sequence and which is capable of hybridising to the second target site (i.e. the second guide RNA targeting T3 component) is provided herein as SEQ ID No: 60, as follows:

TABLE-US-00028 [SEQ ID No: 60] **GCAAUACCACCCGUCAGAG**

[0111] Accordingly, preferably the nucleotide sequence which is encoded by the fifth nucleotide sequence and which is capable of hybridising to the second target site (i.e. the second guide RNA targeting T3) comprises nucleic acid sequence substantially as set out in SEQ ID NO: 60, or a fragment or variant thereof.

[0112] In another embodiment, the nucleotide sequence which is encoded by the fifth nucleotide sequence and which is capable of hybridising to the second target site (i.e. the second guide RNA targeting T3) is provided herein as SEQ ID No: 46, as follows:

TABLE-US-00029 [SEQ ID No: 46] **GCAAUACCACCCGUCAGAGGUUUUAGAGCUAGAAAUAGCAAGUAAAAU**  
**AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCU**

[0113] Accordingly, preferably the nucleotide sequence which is encoded by the fifth nucleotide sequence and which is capable of hybridising to the second target site (i.e. the second guide RNA targeting T3) comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID NO: 46, or a fragment or variant thereof.

[0114] In a further embodiment, the nucleotide sequence which is encoded by the fifth nucleotide sequence and which is capable of hybridising to the second target site (i.e. the second guide RNA targeting T4 component) is provided herein as SEQ ID No: 61, as follows:

TABLE-US-00030 [SEQ ID No: 61] **GUUUAUCAUCCACUCUGA**

[0115] Accordingly, preferably the nucleotide sequence which is encoded by the fifth nucleotide sequence and which is capable of hybridising to the second target site (i.e. the second guide RNA targeting T4) comprises a nucleic acid sequence substantially as set out in SEQ ID NO: 61, or a fragment or variant thereof.

[0116] In a further embodiment, the nucleotide sequence which is encoded by the fifth nucleotide sequence and which is capable of hybridising to the second target site (i.e. the second guide RNA targeting T4) is provided herein as SEQ ID No: 47, as follows:

TABLE-US-00031 [SEQ ID No: 47] **GUUUAUCAUCCACUCUGAGUUUUAGAGCUAGAAAUAGCAAGUAAAAUA**  
**AGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCU**

[0117] Accordingly, preferably the nucleotide sequence which is encoded by the fifth nucleotide sequence and which is capable of hybridising to the second target site (i.e. the second guide RNA targeting T4) comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID NO: 47, or a fragment or variant thereof.

[0118] The CRISPR-based gene drive genetic construct further comprises at least one promoter sequence, such that expression of the first, second and fifth nucleotide sequence is under the control of the same promoter.

[0119] In a preferred embodiment, however, the gene drive genetic construct comprises more than one promoter sequence, such that expression of the first, second and fifth nucleotide sequences are under the control of separate promoters. Preferably, the construct comprises a first promoter sequence operably linked to the first nucleotide sequence, a second promoter sequence operably linked to the second nucleotide sequence, and a third promoter sequence operably linked to the fifth nucleotide sequence.

[0120] The first, second and third promoter sequence may be any promoter sequence that is suitable for expression in an arthropod, and which would be known to those skilled in the art. Accordingly, the first guide RNA for targeting the first target site is expressed under control of the first promoter, the nuclease is expressed under control of the second promoter, and the second guide RNA for targeting the second target site (either T2, T3 or T4) is expressed under the control of the third promoter. Accordingly, in use, the first guide RNA targets the T1 target site, and the second guide RNA targets one or more of T2, T3 and/or T4, as described above.

[0121] Preferably, the first and/or third promoter sequence is a polymerase III promoter, and most preferably a polymerase III promoter which does not add a 5' cap or a 3' polyA tail. More preferably, the first and/or third promoter is a U6 promoter, for example as shown in SEQ ID No: 49, as described herein. Preferably, the first promoter is a U6 promoter and the third promoter is a U6 promoter. In other words, preferably expression of the two guide RNAs is achieved using two separate transcription units, each one preferably containing a U6 promoter.

[0122] Preferably, the second promoter sequence is a promoter sequence that substantially restricts expression of the second nucleotide sequence to germline cells of the arthropod. For example, the second promoter sequence may be selected from a group consisting of: zpg (SEQ ID No: 7); nos (SEQ ID No: 8); exu (SEQ ID No: 9); and vasa2 (SEQ ID No: 10), as described herein. Most preferably, the second promoter is zpg (SEQ ID No: 7).

[0123] Preferably, when transcribed, the first nucleotide sequence, which encodes a nucleotide sequence (i.e. the first guide RNA) which hybridises to the first target site of the doublesex gene (i.e. T1 in FIG. 12), targets the nuclease to the first target site. Preferably, the nuclease then cleaves the doublesex gene at the first target site, such that the gene drive construct is integrated into the disrupted first target site via homology-directed repair. In addition, when transcribed, the fifth nucleotide sequence, which encodes a nucleotide sequence (i.e. the second guide RNA) which hybridises to the second target site of the doublesex gene (i.e. T2, T3 or T4), targets the nuclease to the second target site. Preferably, the nuclease then cleaves the doublesex gene at the second target site, wherein the gene drive construct is integrated into the disrupted second target site via homology-directed repair. Preferably, when both the first and fifth nucleotide sequences are transcribed, they encode nucleotide sequences (i.e. the first and second gRNAs) that hybridise to both the target sites, such that the doublesex gene is cleaved in two sites at once, removing a 76 bp region of exon 5, which is replaced by the CRISPR gene drive construct (for example, see FIG. 13). The skilled person would understand that once the gene drive construct is inserted into the genome of the arthropod, it will use the natural homology found at the site in which it is inserted in the genome

[0124] Preferably, in one embodiment, the CRISPR-based gene drive is introduced into the arthropod via a docking construct, wherein the docking construct comprises integrase attachment sites, preferably attP integrase attachment sites, that are flanked by 5' and 3' homology arms (sixth and seventh nucleotide sequences, respectively) that are homologous to the genomic sequences flanking the two cut-sites which are disposed in exon 5 of the arthropod, such that when the docking construct is introduced into the arthropod, it is integrated into the arthropod's genome by homology

directed repair.

[0125] In one preferred embodiment, therefore, the gene drive construct is inserted into the genome via recombinase-mediated cassette exchange. Accordingly, preferably, the CRISPR-based gene drive genetic construct further comprises integrase attachment sites, preferably attB integrase attachment sites, which, respectively, flank the first nucleotide sequence encoding the nucleotide sequence that is capable of hybridising to the first target site which is an intron-exon boundary of the female specific splice form of the doublesex (dsx) gene, and the fifth nucleotide sequence capable of hybridising to a second target site disposed in exon 5 of the female specific splice form of the doublesex (dsx) gene, the second nucleotide sequence encoding the nuclease, the first promoter sequence, the second promoter sequence and the third promoter sequence. Preferably, an attB site is disposed at the 5' end, and an attB site is disposed at the 3' end of the construct. The CRISPR-based gene drive construct is preferably inserted into the arthropod genome via recombinase-mediated cassette exchange, wherein the docking construct is exchanged for CRISPR-based gene drive construct through the action of an integrase, preferably  $\phi$ C31 integrase, which is introduced into the arthropod.

[0126] Preferably, the homology arms (i.e. the sixth and seventh nucleotide sequences) are at least 100 bp in length, at least 200 bp in length, at least 400 bp in length, at least 600 bp in length, at least 800 bp in length, at least 1000 bp in length at least 1200 bp in length, at least 1400 bp in length, at least 1600 bp in length, at least 1800 bp in length, at least 2000 bp in length. Preferably, the homology arms are up to 4000 bp in length, up to 3000 bp in length, up to 2000 bp in length. Preferably, the homology arms are between 100 and 4000 bp in length, more preferably between 150 and 3000 bp in length and most preferably between 200 and 2000 bp in length. Preferably, the homology arms are about 2000 bp in length.

[0127] In a preferred embodiment, the 5' homology arm (i.e. the sixth nucleotide sequence) is provided herein as SEQ ID No: 11, as described herein. Accordingly, preferably the 5' homology arm comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID No: 11, or a variant or fragment thereof.

[0128] In a preferred embodiment, the 3' homology arm (i.e. the seventh sequence) is provided herein as SEQ ID No: 50, as follows:

TABLE-US-00032 [SEQ ID No: 50] GAGTGGATGATAAACTTTCCGCACCACTGTAAGTGTCCGTATCT  
TTGTATGTGGGTGTGTGTATGTGTTGGTGAACGAATTCAA TAGTCTGTGCTATTTTAAATCAAGCCGCGTGCGCAACTGATGC  
CGATAAGTTCAAACCTGAGTGTAAAGGAGTGGAGCGAGAGCCG  
CACCACGGTACAGAAGGGCAGCAGAATGGGTCGGCAGCCTAGCT  
GCACTGGTGCCTGCGTCCGGCGTCTCGGGGGGAGGGCGAGGAA  
ATTCTAGTGTTAAATCGGAGCAGCAAAAACAAAACAGTGGTCTGT  
CCCCTTCAAGAAACGGCCTGTACACACACACAGAAAACACTGCA GCATGTTTGTACATAGTAGATCCTAGAGCAGGTGGTTCGTTGCTC  
CTCGAACGCTCTGGACGCACGGCTTCGCGCGTATTTGCGTAGCG TTCCGCCGATCGTGGGTATTCGTACTGCCACAAGCCCGCTTTCT  
CCCATCGCAATCTCTGCAACCAACCAACAAACAAACAAACAAACCAATCGACAAAATGAATCACACCCCTTTTGTATCATCTGTATA  
TTCTTGTCTTTTGGCTTCTTTTCTATGTGGCCCAACGCCCGGGG GGTACGTAATTGCGTCGAAAACCCCGAAACCCCGGCACATACA  
GTGTACATACGTTTGGAGCAACTTTGACCTGCAGCCCTCTG GGGTTGCCACGTGTAGCTATACCTTGTGAGATCGGGCGCCGACCG  
TGTAAGCGCGAATGGCCGCCACACAGTGTGTCCACTCCAACAC TACCCCTCTGGAACCTACCCCGTCCAGGGATGCACCGGCTCGGCT  
CATGCCCTGCAAAACAGTCCGGGCTCCACTGTAGTAGCTCCG  
CGTTGCTCTGAGAGAAGGATGCCCTTCGAAGTGTGAAAGCGTG CATTGGGCGTTCAAGTGTGTGTGTGTGTGTAGGTTAGCGAGA  
AACAGCAGCAGTTGCGTGTGCTGAAAAGCGAAGGAGTAATAGAG  
TGCATAATGAAAATGAAAATGAAAATGAAGCAAAAGTAGAAGGC  
GGAGGAGAGCAACCTGTGTTCCACTAGTAGCGAATAGTTTAGTC TAGTTTCGTCACCAATCAACCTTCCAACCATCGTTCAACCAATA  
CCTGAGTCAACATCGTCATCGTTATCGTGCACAACTTTATTA AAATGAACCTTGTCCGCGCCACCGTAGGGTGATCTAAGGCGACC  
TTTCTTACGGGCGCGACCCCATGCCATCGTCACTTCTCCAAT CAAAACCAACAGCCTGTACCGATGGTGTGCAATTGTGCGTGCCT  
GTGTGTTATTAGCAAAAAAAGAGAAAGAGTCGACGAGAGAGAGA  
TAGATCGAGATCGAGAGTACAAAAGAGCAGTAGAAATGTTTCGTT GTTTGTTTTTCGTAACACAGTTGTTTAGCCAAAATGGGAATTTT  
CAATAATCCCGGGGGCGGGGAAATGCGGGAATACTGCGTACACA CATAATCAATCAAAAAGAAAATCCTTGCCTACATCACTACC  
GTTTGCCTGCGGTGCTGATCTAGAGCAGACCACTTTCCACTCCACT CTACAATCAATCAATCTGTGCAGAAGGTATGGTAAGACGGCCTT  
TGAGCGAGTCACGGTCGCCACCATAACGCCGTCGACGAGGGCT GAATGCGAACTTTGCTAATCGATTTTCCGCTTTCTTTTATCCC  
ACCTCCTTTTCTCTCCCTCTCTCTTTTGCCTGCCCCCTTGTA ACCCCCAAAAAGGTAAACGACACATTAAGACCTACGAAGCGTTG  
GTGAAGTCATCGCTCGATCCGAACAGCGACCGGCTGACGGAGGA CGACGACGAGGACGAGAACATCTCGGTGACCCGCACC

[0129] Accordingly, preferably the 3' homology arm used in this embodiment comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID No: 50, or a variant or fragment thereof.

[0130] In another preferred embodiment, however, the CRISPR-based gene drive construct may be inserted into the genome by homology directed repair, i.e. without the use of a docking construct. Accordingly, preferably, the CRISPR-based gene drive genetic construct further comprises of the two homology arms noted above, sixth and seventh nucleotide sequences, which, respectively, flank the first nucleotide sequence encoding the nucleotide sequence that is capable of hybridising to the intron-exon boundary of the doublesex (dsx) gene (i.e the first gRNA), the fifth nucleotide sequence encoding the nucleotide sequence that is capable of hybridising to the second target site in exon 5 of the doublesex (dsx) gene (i.e the second gRNA), the second nucleotide sequence encoding the nuclease, the first promoter sequence and the second and third promoter sequence, wherein the sixth and seventh nucleotides are homologous to the genomic sequences flanking upstream of the first target site and downstream of the second target site (preferably T3 shown in FIG. 12), such that the gene drive construct is integrated into the genome via homology-directed repair.

[0131] Preferably, the homology arms (i.e. the sixth and seventh nucleotide sequences) are at least 100 bp in length, at least 200 bp in length, at least 400 bp in length, at least 600 bp in length, at least 800 bp in length, at least 1000 bp in length at least 1200 bp in length at least 1400 bp in length, at least 1600 bp in length, at least 1800 bp in length, at least 2000 bp in length. Preferably, the third and fourth nucleotide sequences are up to 4000 bp in length, up to 3000 bp in length, up to 2000 bp in length. Preferably, the third and fourth nucleotide sequences are between 100 and 4000 bp in length, more preferably between 150 and 3000 bp in length and most preferably between 200 and 2000 bp in length. Preferably, the third and fourth nucleotide sequences are about 2000 bp in length.

[0132] Accordingly, preferably the sixth nucleotide sequence comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID No: 11, or a variant or fragment thereof.

[0133] Accordingly, preferably the seventh nucleotide sequence comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID No: 50, or a variant or fragment thereof.

[0134] Preferably, the CRISPR-based gene drive construct targets the intron-4-exon 5 boundary of the doublesex gene (i.e. the first target site) and one of T2, T3 and/or T4 (i.e. the second target site). Most preferably, the CRISPR-based gene drive construct targets the intron-4-exon 5 boundary of the doublesex gene (i.e. the first target site) and T3 (i.e. the second target site)

[0135] In a preferred embodiment, the full DNA sequence of the multiplex CRISPR construct is provided herein as SEQ ID No: 51, as follows:

TABLE-US-00033 [SEQ ID No: 51] tgcgggtgccaggcgctgcccttgggctccccggggcgctact caccacccatgcatcgctccggaagatacattgatgagtt  
tggaacaaccacaactagaatgcagtgaaaaaatgcttattt gtgaatttggatgctattgctttatttgaaccattataagc tgcaataacaagttaacaacaacaattgcattcatttatgtt  
tcagggttcagggggaggtgtgggaggttttttaagcaagtaaa acctctacaatgtggtatggctgattatgatctagagtcgagg cgcctacaggaaacaggtgtggcgccctcggtgcgctgact  
gctccacgatggttagtctcgttgggaggtgatgtccagc ttggagtcacgtagtagtagcggggcagctgcaggggttctt ggccatgtagatggactgaactccaccaggtagtgccgcccgt  
ccttcagcttcagggccttggatctcgccttcagcagccg tcgcggggttacaggcgctcggtggaggctccagcccatggt cttctctgcattacggggccgtcggaggggaggttcacgccga  
tgaacttcacctttagatgaagcagccgtcctgcaggaggag tctgggtcacggtcaccacgcgcgctcctgaagtctcatc gcgctcccactgaagccctcggggaaggacagcttctgtagt

[0136] Accordingly, preferably the gene drive construct comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID NO: 51, or a fragment or variant thereof.

[0137] In a second aspect, there is provided the use of the gene drive genetic construct of the first aspect, to disrupt an intron-exon boundary of the female specific splice form of the doublesex gene in an arthropod, such that when the construct is expressed, the exon is spliced out of a doublesex precursor-mRNA transcript, wherein the female arthropod's reproductive capacity is suppressed when females are homozygous for the construct.

[0138] Preferably, the doublesex gene, the intron-exon boundary, the gene drive genetic construct, and the arthropod are as defined in the first aspect. The gene drive genetic construct may be capable of additionally targeting a second target site, which is disposed in exon 5 of the female specific splice form of the doublesex (dsx) gene, as described in relation to the first aspect. Preferably, the use comprises multiplexed genome targeting. In other words, preferably, T1 shown in FIG. 12 is targeted together with T2, T3 and/or T4, most preferably T1 and T3.

[0139] In a third aspect, there is provided a method for preventing or reducing the inclusion of at least one exon into the female specific splice form of arthropod doublesex mRNA, when said mRNA is produced by splicing from a precursor mRNA transcript, the method comprising contacting one or more cells of an arthropod, preferably one or more cells of an arthropod embryo, in vitro or ex vivo, under conditions conducive to uptake of the gene drive genetic construct of the first aspect by such a cell, and allowing splicing to take place.

[0140] Preferably, the doublesex gene, the intron-exon boundary, the gene drive genetic construct, and the arthropod are as defined in the first aspect. The gene drive genetic construct may be capable of additionally targeting a second target site, which is disposed in exon 5 of the female specific splice form of the doublesex (dsx) gene, as described in relation to the first aspect. Preferably, the method comprises multiplexed genome targeting. In other words, preferably, T1 shown in FIG. 12 is targeted together with T2, T3 and/or T4, most preferably T1 and T3.

[0141] In a fourth aspect, there is provided a method of producing a genetically modified arthropod, the method comprising introducing into an arthropod a gene drive genetic construct capable of disrupting an intron/exon boundary of the female specific splice form of the doublesex gene in an arthropod, such that when the gene-drive construct is expressed, an exon is spliced out of a doublesex precursor-mRNA transcript, wherein a female arthropod, which is homozygous for the construct, exhibits a suppressed reproductive capacity.

[0142] Preferably, the doublesex gene, the intron-exon boundary, the gene drive genetic construct, and the arthropod are as defined in the first aspect. The gene drive genetic construct may be capable of additionally targeting a second target site, which is disposed in exon 5 of the female specific splice form of the doublesex (dsx) gene, as described in relation to the first aspect. Preferably, the method comprises multiplexed genome targeting. In other words, preferably, T1 shown in FIG. 12 is targeted together with T2, T3 and/or T4, most preferably T1 and T3.

[0143] The gene drive genetic construct may be introduced directly into an arthropod host cell, preferably an arthropod host cell present in an arthropod embryo, by suitable means, e.g. direct endocytotic uptake. The construct may be introduced directly into cells of a host arthropod (e.g. a mosquito) by transfection, infection, electroporation, microinjection, cell fusion, protoplast fusion or ballistic bombardment. Alternatively, constructs of the invention may be introduced directly into a host cell using a particle gun.

[0144] Preferably, the construct is introduced into a host cell by microinjection of arthropod embryos, preferably an insect embryo and most preferably mosquito embryos.

[0145] Preferably, the gene drive genetic construct is introduced into freshly laid eggs, within 2 hours of deposition. More preferably, the gene drive genetic construct is introduced into an arthropod embryo at the start of melanisation, which the skilled person would understand takes place within 30 minutes after egg laying. Preferably, the mosquito is of the subfamily Anophelinae. Preferably, the mosquito is selected from a group consisting of: *Anopheles gambiae*; *Anopheles coluzzi*; *Anopheles merus*; *Anopheles arabiensis*; *Anopheles quadriannulatus*; *Anopheles stephensi*, *Anopheles funestus* and *Anopheles melas*.

[0146] In a fifth aspect, there is provided a genetically modified arthropod obtained or obtainable by the method of the fourth aspect.

[0147] The genetically modified arthropod may be targeted for target site T1, and one or more of target sites T2, T3 and/or T4, most preferably T1 and T3.

[0148] In a sixth aspect, there is provided a genetically modified arthropod comprising a disrupted intron-exon boundary of the female specific splice form of the doublesex gene, such that the exon is spliced out of a doublesex precursor-mRNA transcript, and wherein a female arthropod, which is homozygous for the disrupted intron-exon boundary, exhibits a suppressed reproductive capacity.

[0149] Preferably, the intron-exon boundary has been disrupted by a gene drive genetic construct as defined in the first aspect. Preferably, the doublesex gene, the intron-exon boundary, the gene drive genetic construct, and the arthropod is as defined in the first aspect. The genetically modified arthropod may be targeted for target site T1, and one or more of target sites T2, T3 and/or T4, most preferably T1 and T3.

[0150] In a seventh aspect, there is provided a method of suppressing a wild type arthropod population, the method comprising breeding a genetically modified arthropod comprising an intron-exon boundary of the female specific splice form of the doublesex gene that has been disrupted by a gene drive genetic construct, such that the exon is spliced out of a doublesex precursor-mRNA transcript, with a wild type population of the arthropod, such that when the gene drive construct is expressed in offspring of the genetically modified arthropod and wild type arthropod, it disrupts the doublesex gene contributed by the wild type population, and wherein when the offspring is a female arthropod homozygous for the disrupted intron-exon boundary, it has suppressed reproductive capacity, such that female reproductive output in the population is reduced, and the wild type arthropod population is suppressed.

[0151] Preferably, the doublesex gene, the intron-exon boundary, the gene drive genetic construct, and the arthropod is as defined in the first aspect. The gene drive genetic construct may be capable of additionally targeting a second target site, which is disposed either wholly or partially in exon 5 of the female specific splice form of the doublesex (dsx) gene, as described in relation to the first aspect. Preferably, the method comprises multiplexed genome targeting. In other words, preferably, T1 shown in FIG. 12 is targeted together with T2, T3 and/or T4, most preferably T1 and T3.

[0152] In an eighth aspect, there is provided a nucleic acid comprising or consisting of a nucleotide sequence substantially as set out as any one of SEQ ID No: 6-34, 42-48, 50-57 or a fragment or variant thereof.

[0153] In a ninth aspect, there is provided a guide RNA comprising any one of SEQ ID No: 58 to 61 and a nuclease binding region.

[0154] The nuclease binding region may bind to, or complex with, a CRISPR nuclease, which may be a Cas endonuclease. For example, the nuclease binding region may bind or complex with Cas9 or Cpf1. The guide RNA may comprise trans-activating CRISPR RNA (tracrRNA) and a CRISPR RNA (crRNA). Alternatively, the guide RNA may comprise a single guide RNA (sgRNA).

[0155] In a tenth aspect, there is provided the nucleic acid according to the eighth aspect or the guide RNA of the ninth aspect, for use in a genome editing method, preferably for suppressing a wild type arthropod population.

[0156] The genome editing method or technique may be carried out in vivo, in vitro or ex vivo.

[0157] Preferably, the nucleic acid according to the eighth aspect or the guide RNA of the ninth aspect is used in the method of the seventh aspect.

[0158] It will be appreciated that the invention extends to any nucleic acid or peptide or variant, derivative or analogue thereof, which comprises substantially the amino acid or nucleic acid sequences of any of the sequences referred to herein, including variants or fragments thereof. The terms “substantially the amino acid/nucleotide/peptide sequence”, “variant” and “fragment”, can be a sequence that has at least 40% sequence identity with the amino acid/nucleotide/peptide sequences of any one of the sequences referred to herein, for example 40% identity with the sequence identified as SEQ ID Nos: 1-94 and so on.

[0159] Amino acid/polynucleotide/polypeptide sequences with a sequence identity which is greater than 65%, more preferably greater than 70%, even more preferably greater than 75%, and still more preferably greater than 80% sequence identity to any of the sequences referred to are also envisaged. Preferably, the amino acid/polynucleotide/polypeptide sequence has at least 85% identity with any of the sequences referred to, more preferably at least 90% identity, even more preferably at least 92% identity, even more preferably at least 95% identity, even more preferably at least 97% identity, even more preferably at least 98% identity and, most preferably at least 99% identity with any of the sequences referred to herein.

[0160] The skilled technician will appreciate how to calculate the percentage identity between two amino acid/polynucleotide/polypeptide sequences. In order to calculate the percentage identity between two amino acid/polynucleotide/polypeptide sequences, an alignment of the two sequences must first be prepared, followed by calculation of the sequence identity value. The percentage identity for two sequences may take different values depending on:—(i) the method used to align the sequences, for example, ClustalW, BLAST, FASTA, Smith-Waterman (implemented in different programs), or structural alignment from 3D comparison; and (ii) the parameters used by the alignment method, for example, local vs global alignment, the pair-score matrix used (e.g. BLOSUM62, PAM250, Gonnet etc.), and gap-penalty, e.g. functional form and constants.

[0161] Having made the alignment, there are many different ways of calculating percentage identity between the two sequences. For example, one may divide the number of identities by: (i) the length of shortest sequence; (ii) the length of alignment; (iii) the mean length of sequence; (iv) the number of non-gap positions; or (v) the number of equivalenced positions excluding overhangs. Furthermore, it will be appreciated that percentage identity is also strongly length dependent. Therefore, the shorter a pair of sequences is, the higher the sequence identity one may expect to occur by chance.

[0162] Hence, it will be appreciated that the accurate alignment of protein or DNA sequences is a complex process. The popular multiple alignment program ClustalW (Thompson et al., 1994, Nucleic Acids Research, 22, 4673-4680; Thompson et al., 1997, Nucleic Acids Research, 24, 4876-4882) is a preferred way for generating multiple alignments of proteins or DNA in accordance with the invention. Suitable parameters for ClustalW may be as follows: For DNA alignments: Gap Open Penalty=15.0, Gap Extension Penalty=6.66, and Matrix=Identity. For protein alignments: Gap Open

Penalty=10.0, Gap Extension Penalty=0.2, and Matrix=Gonnet. For DNA and Protein alignments: ENDGAP=-1, and GAPDIST=4. Those skilled in the art will be aware that it may be necessary to vary these and other parameters for optimal sequence alignment.

[0163] Preferably, calculation of percentage identities between two amino acid/polynucleotide/polypeptide sequences may then be calculated from such an alignment as  $(N/T) \times 100$ , where N is the number of positions at which the sequences share an identical residue, and T is the total number of positions compared including gaps and either including or excluding overhangs. Preferably, overhangs are included in the calculation. Hence, a most preferred method for calculating percentage identity between two sequences comprises (i) preparing a sequence alignment using the ClustalW program using a suitable set of parameters, for example, as set out above; and (ii) inserting the values of N and T into the following formula:—  
Sequence Identity= $(N/T) \times 100$ .

[0164] Alternative methods for identifying similar sequences will be known to those skilled in the art. For example, a substantially similar nucleotide sequence will be encoded by a sequence which hybridizes to DNA sequences or their complements under stringent conditions. By stringent conditions, the inventors mean the nucleotide hybridises to filter-bound DNA or RNA in  $3 \times$  sodium chloride/sodium citrate (SSC) at approximately  $45^\circ \text{C}$ . followed by at least one wash in  $0.2 \times \text{SSC}/0.1\%$  SDS at approximately  $20\text{--}65^\circ \text{C}$ . Alternatively, a substantially similar polypeptide may differ by at least 1, but less than 5, 10, 20, 50 or 100 amino acids from the sequences shown in, for example, SEQ ID Nos:1 to 94.

[0165] Due to the degeneracy of the genetic code, it is clear that any nucleic acid sequence described herein could be varied or changed without substantially affecting the sequence of the protein encoded thereby, to provide a functional variant thereof. Suitable nucleotide variants are those having a sequence altered by the substitution of different codons that encode the same amino acid within the sequence, thus producing a silent (synonymous) change. Other suitable variants are those having homologous nucleotide sequences but comprising all, or portions of, sequence, which are altered by the substitution of different codons that encode an amino acid with a side chain of similar biophysical properties to the amino acid it substitutes, to produce a conservative change. For example small non-polar, hydrophobic amino acids include glycine, alanine, leucine, isoleucine, valine, proline, and methionine. Large non-polar, hydrophobic amino acids include phenylalanine, tryptophan and tyrosine. The polar neutral amino acids include serine, threonine, cysteine, asparagine and glutamine. The positively charged (basic) amino acids include lysine, arginine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid. It will therefore be appreciated which amino acids may be replaced with an amino acid having similar biophysical properties, and the skilled technician will know the nucleotide sequences encoding these amino acids.

[0166] All of the features described herein (including any accompanying claims, abstract and drawings), and/or all of the steps of any method or process so disclosed, may be combined with any of the above aspects in any combination, except combinations where at least some of such features and/or steps are mutually exclusive.

---

## Description

[0167] For a better understanding of the invention, and to show how embodiments of the same may be carried into effect, reference will now be made, by way of example, to the accompanying Figures, in which:

[0168] FIG. 1 shows targeting the female-specific isoform of doublesex. (a) Schematic representation of the male- and female-specific dsx transcripts and the gRNA sequence used to target the gene (shaded in grey). The gRNA spans the Intron4-Exon5 boundary. The proto-spacer adjacent motive (PAM) of the gRNA is highlighted in blue. The scale bar indicates a 200 bp fragment. Introns are not drawn to scale. (b) Sequence alignment of the dsx Intron4-Exon5 boundary in 6 of the species from the *Anopheles gambiae* complex. The sequence is highly conserved within the complex suggesting tight functional constraint at this region of the dsx gene. The gRNA used to target the gene is underlined and the PAM is highlighted in blue. (c) Schematic representation of the HDR knockout construct specifically recognising exon 5 and the corresponding target locus. (d) Diagnostic PCR using a primer set (blue arrows in panel (c)) to discriminate between the wild type and dsxF allele in homozygous (dsxF.sup.-/-) heterozygous (dsxF.sup.+/-) and wt individuals.

[0169] FIG. 2 shows morphological analysis of homozygous dsxF.sup.-/- mutants. (a) Morphological appearance of genetic males and females heterozygous (dsxF.sup.+/-) or homozygous (dsxF.sup.-/-) for exon 5 null allele. This assay was performed in a strain containing dominant RFP marker linked to the Y chromosome, whose presence permits unambiguous determination of male or female genotype. Anomalies in sexual morphology were observed only in dsxF.sup.-/- genetic female mosquitoes. This group of XX individuals showed male-specific traits including a plumose antenna and claspers (arrows). This group also showed anomalies in the proboscis and accordingly they could not bite and feed on blood. Representative samples of each genotype are shown. (b) Magnification of the external genitalia. All dsxF.sup.-/- females carried claspers, a male-specific characteristic. The claspers were dorsally rotated rather than in the normal ventral position.

[0170] FIG. 3 shows the reproductive phenotype of dsxF mutants. Males and females dsxF.sup.-/- and dsxF.sup.+/- individuals were mated with the corresponding wild type sexes. Females were given access to a blood meal and subsequently allowed to lay individually. Fecundity was investigated by counting the number of larval progeny per lay ( $n \geq 43$ ). Using wild type (wt) as a comparator the inventors saw no significant differences ('ns') in any genotype other than dsxF.sup.-/- females, which were unable to feed on blood and therefore failed to produce a single egg (\*\*\*\*,  $p < 0.0001$ ; Kruskal-Wallis test). Vertical bars indicate the mean and the s.e.m.

[0171] FIG. 4 shows the transmission rate of the dsxF<sup>CRISPRh</sup> driving allele and fecundity analysis of heterozygous male and female mosquitoes. Male and female mosquitoes heterozygous for the dsxF<sup>CRISPRh</sup> allele (a) (dsxF<sup>CRISPRh</sup>+) were analysed in crosses with wild type mosquitoes to assess the inheritance bias of the dsxF<sup>CRISPRh</sup> drive construct (b) and for the effect of the construct on their reproductive phenotype (c). (b) Scattered plot of the transgenic rate observed in the progeny of dsxF<sup>CRISPRh</sup> female or male mosquitoes ( $n \geq 42$ ) crossed to wild type individuals. Each dot represents the progeny derived from single females. Both male and female dsxF<sup>CRISPRh</sup> showed a high transmission rate of up to 100% of the dsxF<sup>CRISPRh</sup> allele to the progeny. The transmission rate was determined by visual scoring among offspring of the RFP marker that is linked to the dsxF<sup>CRISPRh</sup> allele. The dotted line indicates the expected Mendelian inheritance. Mean transmission rate ( $\pm$ s.e.m.) is shown (c) Scattered plot showing the number of larvae produced by single females from crosses of dsxF<sup>CRISPRh</sup> mosquitoes with wild type individuals after one blood meal. Mean progeny count ( $\pm$ s.e.m.) is shown. (\*\*\*\*,  $p < 0.0001$ ; Kruskal-Wallis test).

[0172] FIG. 5 shows the dynamics of the spread of the dsxF.sup.CRISPRh allele and effect on population reproductive capacity. Two cages were set up with a starting population of 300 wild type females, 150 wild type males and 150 dsxF.sup.CRISPRh/+ males, seeding each cage with a dsxF.sup.CRISPRh allele frequency of 12.5%. The frequency of the dsxF.sup.CRISPRh mosquitoes was scored for each generation (a). The drive allele reached 100% prevalence in both cage 2 (grey) and cage 1 (black) at generation 7 and 11 in agreement with a deterministic model (dotted line) that takes into account the parameter values retrieved from the fecundity assays. 20 stochastic simulations were run (light grey lines) assuming a max population size of 650 individuals. (b) Total egg output deriving from each generation of the cage was measured and normalised relative to the output from the starting generation. Suppression of the reproductive output of each cage led the population to collapse completely (black arrows) by generation 8 (cage 2) or generation 12 (cage 1). Parameter estimates included in the model are provided in Table 1.

[0173] FIG. 6 shows molecular confirmation of the correct integration of the HDR-mediated event to generate dsxF.sup.-. PCRs were performed to verify the location of the dsx  $\phi$ C31 knock-in integration. Primers (blue arrows) were designed to bind internal of the  $\phi$ C31 construct and outside of the regions used for homology directed repair (HDR) (dotted gray lines) which were included in the Donor plasmid K101. Amplicons of the expected sizes should only be produced in the event of a correct HDR integration. The gel shows PCRs performed on the 5' (left) and 3' (right) of 3 individuals for the dsx  $\phi$ C31 knock-in line (dsxF.sup.-) and wild type (wt) as a negative control.

[0174] FIG. 7 shows the morphology of the dsxF.sup.-/- internal reproductive organs. (a) Testis-like gonad from 3-days old female dsxF.sup.-/-

individual. There was no layer division between the cells and there was no evidence of sperm. (b) Dissections performed on dsxF.sup.-/- genetic females revealed the presence of organs resembling accessory glands, a typical male internal reproductive organ.

[0175] FIG. 8 shows the development of dsxF.sup.CRISPRh drive construct and its predicted homing process and molecular confirmation of the locus. (a) The drive construct (CRISPRh cassette) contained the transcription unit of a human codon-optimised Cas9 controlled by the germline-restrictive zpg promoter, the RFP gene under the control of the neuronal 3×P3 promoter and the gRNA under the control of the constitutive U6 promoter, all enclosed within two attB sequences. The cassette was inserted at the target locus using recombinase-mediated cassette exchange (RMCE) by injecting embryos with a plasmid containing the cassette and a plasmid containing a Φ31 recombination transcription unit. During meiosis the Cas9/gRNA complex cleaves the wild type allele at the target locus (DSB) and the construct is copied across to the wild type allele via HDR (homing) disrupting exon 5 in the process. (b) Representative example of molecular confirmation of successful RMCE events. Primers (blue arrows) that bind components of the CRISPRh cassette were combined with primers that bind the genomic region surrounding the construct. PCRs were performed on both sides of the CRISPRh cassette (5' and 3') on many individuals as well as wild type controls (wt).

[0176] FIG. 9 shows the maternal or paternal inheritance of the dsxF.sup.CRISPRh driving allele affect fecundity and transmission bias in heterozygotes. Male and female dsxF.sup.CRISPRh heterozygotes (dsxF.sup.CRISPRh/+) that had inherited a maternal or paternal copy of the driving allele were crossed to wild type and assessed for inheritance bias of the construct (a) and reproductive phenotype (b). (a) Progeny from single crosses (n≥15) were screened for the fraction that inherited DsRed marker gene linked to the dsxF.sup.CRISPRh driving allele (e.g. G1 ♂.fwdarw.G2 ♀ represents a heterozygous female that received the drive allele from her father). Levels of homing were similarly high in males and females whether the allele had been inherited maternally or paternally. The dotted line indicates the expected Mendelian inheritance. Mean transmission rate (±s.e.m.) is shown. (b) Counts of hatched larvae for the individual crosses revealed a fertility cost in female dsxF.sup.CRISPRh heterozygotes that was stronger when the allele was inherited paternally. Mean progeny count (±s.e.m.) is shown. (\*\*\*, p<0.001;\*\*\*\*, p<0.0001; Kruskal-Wallis test).

[0177] FIGS. 10A-C show resistance plots variants and deletions in sequence. Pooled amplicon sequencing of the target site from 4 generations of the cage experiment (generations 2, 3, 4 and 5) revealed a range of very low frequency indels at the target site (FIG. 10A), none of which showed any sign of positive selection. Insertion, deletion and substitution frequencies per nucleotide position were calculated, as a fraction of all non-drive alleles, from the deep sequencing analysis for both cages. Distribution of insertions and deletions (FIG. 10B) in the amplicon is shown for each cage. Contribution of insertions and deletions arising from different generations is displayed with the frequency in each generation represented by a different colour. Significant change (p<0.01) in the overall indel frequency was observed in the region around the cut-site (dotted area+/-20 bp) for both cages. No significant changes were observed in the substitution frequency (FIG. 10C) around the cut-site (shaded area+/-20 bp) when compared with the rest of the amplicon, confirming that the gene drive did not generate any substitution activity at the target locus and that the laboratory colony is devoid of any standing variation in the form of SNPs within the entire amplicon.

[0178] FIG. 11 shows a sequence comparison of the dsx female-specific exon 5 across members of the *Anopheles* genus and SNP data obtained from *Anopheles gambiae* mosquitoes in Africa. (a) Sequence comparison of the dsx Intron4-Exon5 boundary and the dsx female-specific exon 5 within the 16 Anopheline species. The sequence of the Intron4-Exon5 boundary is completely conserved within the six species that form the *Anopheles gambiae* complex (noted in bold). The gRNA used to target the gene is underlined and the PAM is highlighted in blue. Changes in the DNA sequence are shaded grey and codon silent and missense substitutions are noted in blue and red respectively. (b) SNP frequencies obtained from 765 *Anopheles gambiae* mosquitoes captured across Africa. Across the dsx female-specific Exon 5 there are only 2 SNP variants (noted in yellow) with frequencies of 2.9% (the SNP in the gRNA-complementary sequence) and 0.07%.

[0179] FIG. 12 shows a sequence comparison of the dsx female-specific exon 5 across members of the *Anopheles* genus and SNP data obtained from *Anopheles gambiae* mosquitoes in Africa. It shows a further three invariant target sites (referred to as T2, T3 and T4) in addition to the original target site (referred to as T1), which have been identified in exon 5 of the *Anopheles gambiae* doublesex gene. A sequence alignment in the coding sequence of AgdsxF exon 5 (including part of intron 4, and the 3' untranslated region (UTR) of exon5) amongst all available mosquito species in which a doublesex homologue could be identified is shown. Species names are shown on the left, and species in bold belong to the *Anopheles gambiae* species complex. Nucleotides that are variable compared to the *Anopheles gambiae* sensus stricto reference sequence on the top are shaded in dark grey. Nucleotides are shown in light blue or red, depending on whether a variation causes a synonymous or non-synonymous amino acid change in the exon 5 coding sequence. Asterisks denote the nucleotide positions that remained unchanged in all species. gRNA binding sites are shaded in light grey and underlined in black, the proto-spacer adjacent motives (PAMs) required for Cas9 cleavage are underlined in red. The 3' splicing acceptor CAGG is shaded in green. In yellow, a single nucleotide polymorphism that has been identified in wild *Anopheles gambiae* populations, is highlighted.

[0180] FIG. 13 shows one embodiment of a novel multiplexed gene drive at doublesex. This embodiment contains a visible marker (the RFP marker), a germline-expressed Cas9 nuclease and two ubiquitously expressed gRNAs targeting target sites T1 and T3. The CRISPR construct was knocked in between the T1 and T3 cut sites. Homing analysis of the new multi-guide gene drive is shown. Promoter sequences are shown as light grey arrows.

[0181] FIG. 14 shows a comparison of the transmission rates and fertility of heterozygous gene drive carriers when the gene drive contained a single target, i.e. T1 (FIGS. 14A & C) or two targets, i.e. T1 and T3 (Figures B & D). Female or male gene drive carriers that inherited the drive from a female or male transgenic individual (F->F, F->M, M->F, M->M) were crossed to wild-type mosquitoes. Females were allowed to lay individually. The reproductive output of females was determined by counting eggs and hatched larvae and transmission rates were determined by screening the progeny for RFP fluorescence, indicative of carrying the gene drive. Figures A & B show that the transmission rates correspond to the total number of RFP+ progeny over the total number of screened progeny per female. Mean transmission rates±s.e.m. (standard error of mean) are shown. Figures C & D show that the larval output of each class is shown, including a wild-type control, as the standard for comparison (red line). Mean larval outputs±s.e.m. are shown. Note that females with zero larval output that showed no evidence of mating were all included in the analysis, since mating competence can be affected by carrying mutations at doublesex. The results from Kyrou et al. (2018) shown on the left were adapted to also include unmated individuals in the analysis.

#### EXAMPLES

[0182] The invention described herein relies on inserting site-specific nuclease genes into a locus of choice, in formations that both confer some trait of interest on an individual and lead to a biased inheritance of the trait. The approach relies on “homing” leading to suppression. The invention is focused on population suppression, whereby the gene drive construct is designed to insert within a target gene in such a way that the gene product, or a specific isoform thereof, is disrupted. To build the nuclease-based gene drive of the invention, the nuclease gene is inserted within its own recognition sequence in the genome such that a chromosome containing the nuclease gene cannot be cut, but chromosomes lacking it are cut. When an individual contains both a nuclease-carrying chromosome and an unmodified chromosome (i.e. heterozygous for the gene drive), the unmodified chromosome is cut by the nuclease. The broken chromosome is usually repaired using the nuclease-containing chromosome as a template and, by the process of homologous recombination, the nuclease is copied into the targeted chromosome. If this process, called “homing”, is allowed to proceed in the germline, then it results in a biased inheritance of the nuclease gene, and its associated disruption, because sperm or eggs produced in the germline can inherit the gene from either the original nuclease-carrying chromosome, or the newly modified chromosome.

[0183] Due to the negative reproductive load the gene drive imposes, selection can be expected to occur for resistant alleles. The most likely source of such resistance is sequence variation at the target site that prevents the nuclease cutting yet at the same time permits a functional product from the target gene. Such variation can pre-exist in a population or can be created by activity of the nuclease itself—a small proportion of cut chromosomes,



rather than using a template, can instead be repaired by end-joining (EJ), which can introduce small insertions or deletions (“indels”) or base substitutions during the repair of the target site. In-frame indels or conservative substitutions might be expected to show selection in the presence of a gene drive. The inventors have previously observed target site resistance in cage experiments (data not shown) and found that end-joining in chromosomes of the early embryo, due to parentally-deposited nuclease, was likely to be the predominant source of the resistant alleles at the target site.

[0184] In mitigating and preventing the emergence of resistant alleles, the strategy being investigated by the inventors involves carefully selecting target sites in regions of the target gene that are so functionally constrained and conserved that most variation is unlikely to restore function to the gene, meaning that the majority of variants will simply not confer any selective advantage. The inventors therefore investigated whether *Anopheles gambiae* doublesex gene (dsx) is a suitable target for a gene drive approach aimed at suppressing population reproductive capacity to eradicate malaria. To do this, they disrupted the intron 4-exon 5 boundary of dsx (referred to as target site “T1”) with the primary objective to prevent the formation of functional AgdsxF while leaving the AgdsXM transcript unaffected. They also disrupted target sites (referred to as T2, T3 and T4) in addition to the original target site, T1.

## Materials and Methods

### Population Genetics Model

[0185] To model the results of the cage experiments, the inventors used discrete-generation recursion equations for the genotype frequencies, treating males and females separately.  $F_{ij}(t)$  and  $M_{ij}(t)$  denote the frequency of females (or males) of genotype  $i/j$  in the total female (or male) population. The inventors considered three alleles, W (wildtype), D (driver) and R (non-functional resistant), and therefore six genotypes.

Homing Adults of genotype W/D produce gametes at meiosis in the ratio W:D:R as follows: [0186]  $(1-d_{sub.f})(1-u_{sub.f}):d_{sub.f}(1-d_{sub.f}):u_{sub.f}$  in females [0187]  $(1-d_{sub.m})(1-u_{sub.m}):d_{sub.m}(1-d_{sub.m}):u_{sub.m}$  in males

[0188] Here,  $d_f$  and  $d_m$  are the rates of transmission of the driver allele in the two sexes and  $u_f$  and  $u_m$  are the fractions of non-drive gametes that are non-functional resistant (R alleles) from meiotic end-joining. In all other genotypes, inheritance is Mendelian.

[0189] Fitness. Let  $w_{ij} \leq 1$  represent the fitness of genotype  $i/j$  relative to  $w_{WW}=1$  for the wild type homozygote. The inventors assume no fitness effects in males. Fitness effects in females are manifested as differences in the relative ability of genotypes to participate in mating and reproduction. The inventors assume the target gene is needed for female fertility, thus D/D, D/R and R/R females are sterile; there is no reduction in fitness in females with only one copy of the target gene (W/D, W/R).

### Parental Effects

[0190] The inventors consider that further cleavage of the W allele and repair can occur in the embryo if nuclease is present, due to one or both contributing gametes derived from a parent with one or two driver alleles. The presence of parental nuclease is assumed to affect somatic cells and therefore female fitness but has no effect in germline cells that would alter gene transmission. Previously, embryonic EJ effects (maternal only) were modelled as acting immediately in the zygote [1,2]. Here, the inventors consider that experimental measurements of female individuals of different genotypes and origins show a range of fitnesses, suggesting that individuals may be mosaics with intermediate phenotypes. The inventors therefore model genotypes W/X (X=W, D, R) with parental nuclease as individuals with an intermediate reduced fitness  $w_{sub.WX.sup.10}$ ,  $w_{sub.WX.sup.01}$ , or  $w_{sub.WX.sup.11}$  depending on whether nuclease was derived from a transgenic mother, father, or both. The inventors assume that parental effects are the same whether the parent(s) had one or two drive alleles. For simplicity, a baseline reduced fitness of  $w_{sub.10}$ ,  $w_{sub.01}$ ,  $w_{sub.11}$  is assigned to all genotypes W/X (X=W, D, R) with maternal, paternal and maternal/paternal effects, with fitness estimated as the product of mean egg production values and hatching rates relative to wild type in Table 1 in the deterministic model. In the stochastic version of the model, egg production from female individuals with different parentage is sampled with replacement from experimental values.

TABLE-US-00034 TABLE 1 Parameters for stochastic cage model Method of Parameter Estimate estimation Mating 0.85 for heterozygotes; 0 for Estimated probability D/D, D/R and R/R homozygotes from Hammond et al. 2017 Egg production Mean 137.4. Sampling with From assays from wildtype replacement of observed values of mated female (no (10, 61, 96, 98, 111, 111, 113, females parental 127, 128, 129, 132, 132, 134, nuclease) 135, 137, 138, 138, 139, 142, 142, 146, 146, 149, 152, 152, 152, 158, 160, 162, 164, 170, 179, 186, 189, 191) Egg production Mean 118.96. Sampling with From assays from W/D replacement of observed values of mated heterozygote (12, 31, 76, 90, 96, 100, 106, females female (nuclease 106, 107, 113, 117, 118, 119, from ♀) 130, 133, 136, 136, 136, 137, 138, 139, 142, 143, 145, 146, 148, 157, 174) Egg production Mean 59.67. Sampling with From assays from W/D replacement of observed values of mated heterozygote (0, 0, 0, 0, 0, 34, 47, 50, 65, females female (nuclease 105, 113, 115, 115, 125, 126) from ♂) Hatching 0.941 From assays probability, of mated wildtype female females (no parental nuclease) Hatching 0.707 From assays probability, of mated W/D heterozygote females female (nuclease from ♀) Hatching 0.47 From assays probability, of mated W/D heterozygote females female (nuclease from ♂) Probability 0.8708 Average of of emergence observations from pupa over all (survival generations from larva) and both cage experiments Drive in 0.9985 Observed W/D females fraction transgenic from assays Drive in 0.9635 Observed W/D males fraction transgenic from assays Meiotic EJ 0.4685 Estimated parameter from Hammond (fraction et al. 2016 non-drive alleles that are resistant)

### Recursion Equations

[0191] The inventors firstly considered the gamete contributions from each genotype, including parental effects on fitness. In addition to W and R gametes that are derived from parents that have no drive allele and therefore have no deposited nuclease, gametes from W/D females and W/D, D/R and D/D males carry nuclease that is transmitted to the zygote, and these are denoted as  $W\{\circlearrowleft\}$ ,  $D\{\circlearrowleft\}$ ,  $R\{\circlearrowleft\}$ . The proportion of type  $i$  alleles in eggs produced by females participating in reproduction are given in terms of male and female genotype frequencies below. Frequencies of mosaic individuals with parental effects (i.e., reduced fitness) due to nuclease from mothers, fathers or both are denoted by superscripts 10, 01 and 11.

[00001]

$$e_W = (F_{WW} + w_{WW}^{10} F_{WW}^{10} + w_{WW}^{01} F_{WW}^{01} + w_{WW}^{11} F_{WW}^{11} + (F_{WR} + w_{WR}^{10} F_{WR}^{10} + w_{WR}^{01} F_{WR}^{01} + w_{WR}^{11} F_{WR}^{11}) / 2) / \bar{w}_f e_R = \frac{1}{2} (F_{WR} + w_{WR}^{10} F_{WR}^{10} + w_{WR}^{01} F_{WR}^{01} + w_{WR}^{11} F_{WR}^{11})$$

[0192] The proportions  $s_{sub.i}$  of type  $i$  alleles in sperm are:

[00002]

$$s_W = (M_{WW} + M_{WW}^{10} + M_{WW}^{01} + M_{WW}^{11} + (M_{WR} + M_{WR}^{10} + M_{WR}^{01} + M_{WR}^{11}) / 2) / \bar{w}_m s_R = (M_{RR} + (M_{WR} + M_{WR}^{10} + M_{WR}^{01} + M_{WR}^{11}) / 2) / \bar{w}_m s_W^* = (1 - d_m)(1$$

[0193] Above,  $w_{sub.f}$  and  $w_{sub.m}$  are the average female and male fitness:

[00003]

$$\bar{w}_f = F_{WW} + w_{WW}^{10} F_{WW}^{10} + w_{WW}^{01} F_{WW}^{01} + w_{WW}^{11} F_{WW}^{11} + w_{WD}^{10} F_{WD}^{10} + w_{WD}^{01} F_{WD}^{01} + w_{WD}^{11} F_{WD}^{11} + F_{WR} + F_{WR}^{10} w_{WR}^{10} + w_{WR}^{01} F_{WR}^{01} + w_{WR}^{11} F_{WR}^{11} \bar{w}_m = M_{WW} + M_{WW}^{10} + M_{WW}^{01} + M_{WW}^{11} + (M_{WR} + M_{WR}^{10} + M_{WR}^{01} + M_{WR}^{11}) / 2$$

[0194] To model cage experiments, the inventors started with an equal number of males and females, with an initial frequency of wildtype females in the female population of  $F_{WW}=1$ , wildtype males in the male population of  $M_{sub.WW}=1/2$ , and  $M_{sub.WD.sup.01}=1/2$  heterozygote drive males that inherited the drive from their fathers. Assuming a 50:50 ratio of males and females in progeny, after the starting generation, genotype frequencies of type  $i/j$  in the next generation ( $t+1$ ) are the same in males and females,  $F_{sub.ij}(t+1)=M_{sub.ij}(t+1)$ . Both are given by  $G_{sub.ij}(t+1)$  in the following set of equations in terms of the gamete proportions in the previous generation, assuming random mating:

[00004]

$$G_{WW}(t+1) = e_W s_W G_{WW}^{10}(t+1) = e_W^* s_W G_{WW}^{01}(t+1) = e_W s_W^* G_{WW}^{11}(t+1) = e_W^* s_W^* G_{WD}^{10}(t+1) = e_D^* s_W G_{WD}^{01}(t+1) = e_W s_W^* G_{WD}^{11}(t+1) = e_W^* s_D^* + e_D^*$$

[0195] The frequency of transgenic individuals can be compared with experiment (fraction of RFP+ individuals):

$$[00005] f_{RFP+} = F_{WD}^{10} + F_{WD}^{01} + F_{WD}^{11} + F_{DD} + F_{DR} + M_{WD}^{10} + M_{WD}^{01} + M_{WD}^{11} + M_{DD} + M_{DR}$$

[0196] All calculations were carried out using Wolfram Mathematica.sup.23.

[0197] The PCR reactions were performed using Phusion High Fidelity Master Mix. Initial denaturation was performed in 98° C. for 30 seconds. Primer annealing was performed at a temperature range of 60-72° C. for 30 seconds and elongation was performed at a temperature of 72° C. for 30 seconds per kb.

TABLE-US-00035 TABLE 2 Primers used in this study for Example 1 dsxgRNA-F TGCTGTTTAAACACAGGTCAAGCGG-SEQ ID No: 14 dsxgRNA-R AAACCCGCTTGACCTGTGTTAAAC-SEQ ID No: 15 dsxΦ31L-F GCTCGAATTAACCATTTGTTGACCGGTCTTGTGTTTAGCAG GCAGGGGA-SEQ ID No: 16 dsxΦ31L-R TCCACCTCACCCATGGGACCCACGCGTGGTGC GGGTCAACG GAGATGTTC-SEQ ID No: 17 dsxΦ31R-F CACCAAGACAGTTAACGTATCCGTTACCTTGACCTGTGTTA AACATAAAT-SEQ ID No: 18 dsxΦ31R-R GGTGGTAGTGCCACACAGAGAGCTTCGCGGTGGTCAACG AATACTCACG-SEQ ID No: 19 zpgprCRISPR-F GCTCGAATTAACCATTTGTTGACCGGTGAGCGCTGGCGGTG GGA-SEQ ID No: 20 zpgprCRISPR-R TCGTGGTCTTATAGTCCATCTCGAGCTCGATGCTGTATTT GTTG-SEQ ID No: 21 zpgteCRISPR-F AGGCAAAAAAGAAAAAGTAATTAATTAAGAGGACGGCGA GAAGTAATCAT-SEQ ID No: 22 zpgteCRISPR-R TTCAAGCGCACGCATACAAAGGCGCGCTCGCATAATGAA CGAACCAAGG-SEQ ID No: 23 dsxin3-F GGCCCTTCAACCCGAAGAT-SEQ ID No: 24 dsxex6-R CTTTTGTACAGCGGTACAC-SEQ ID No: 25 GFP-F GCCCTGAGCAAAGACCCCAA-SEQ ID No: 26 dsxex4-F GCACACCAGCGGATCGACGAAG-SEQ ID No: 27 dsxex5-R CCCACATACAAAGATACGGACAG-SEQ ID No: 28 dsxex6-R GAATTTGGTGTCAAGGTTTCAGG-SEQ ID No: 29 3xP3 TATACTCCGGCGGTGAGGGTT-SEQ ID No: 30 hCas9-F CCAAGAGAGTGATCCTGGCCGA-SEQ ID No: 31 dsxex5-R1 CTTATCGGCATCAGTTGCGCAC-SEQ ID No: 32 dsxin4-F GGTGTTATGCCACGTTCACTGA-SEQ ID No: 33 RFP-R CAAGTGGGAGCGCGTGATGAAC-SEQ ID No: 34

TABLE-US-00036 TABLE 6 Primers used in this study for Example 2 multidxΦ31L-F gctcgaataaccattgtggaccggtCTTGTTGTTAGCAGGCAGGGGA-SEQ ID No: 52 multidxΦ31L-R tgaacgattgggtaccggtCTTGACCTGTGTTAAACATAAATG-SEQ ID No: 53 multidxΦ31R-F agataaatctcgaacggtGAGTGGATGATAAACTTTCCGCAC-SEQ ID No: 54 multidxΦ31R-R tccacctcacccatgggaccacggtGGTGC GGGTCAACGAGATGTTC-SEQ ID No: 55 4050-2U6-T1-F gaggggtctcaTGCTGTTTAAACACAGGTCAAGCGGgttttagctagaataagca agt-SEQ ID No: 56 4050-2U6-T3-R gaggggtctcaAAACCTCTGACGGGTGGTATTGCagcagagagcaactcattca t-SEQ ID No: 57

#### Example 1

[0198] To investigate whether dsx represented a suitable target for a gene drive approach aimed at suppressing population reproductive capacity, the inventors disrupted the intron 4-exon 5 boundary of dsx with the objective to prevent the formation of functional AgdsxF while leaving the AgdsxM transcript unaffected. The inventors injected *A. gambiae* embryos with a source of Cas9 and gRNA designed to selectively cleave the intron 4-exon 5 boundary in combination with a template for homology directed repair (HDR) to insert an eGFP transcription unit (FIG. 1c). Transformed individuals were intercrossed to generate homozygous and heterozygous mutants among the progeny.

#### Results

[0199] HDR-mediated integration was confirmed by diagnostic PCR using primers that spanned the insertion site, producing a larger amplicon of the expected size for the HDR event and a smaller amplicon for the wild type allele, and thus allowing easy confirmation of genotypes (FIG. 1d).

[0200] The knock-in of the eGFP construct resulted in the complete disruption of the exon 5 (dsxF.sup.-) coding sequence and was confirmed by PCR and genomic sequencing of the chromosomal integration (FIG. 6 and data not shown). Crosses of heterozygote individuals produced, wild type, heterozygous and homozygous individuals for the dsxF.sup.- allele at the expected Mendelian ratio 1:2:1, indicating that there was no obvious lethality associated with the mutation during development (Table 3).

TABLE-US-00037 TABLE 3 Ratio of larvae recovered by intercrossing heterozygous dsx ΦC31-knock-in mosquitoes GFP strong (dsxF.sup.-/-) GFP weak (dsxF.sup.-/+) no GFP (+/+) Total 262 (24.9%) 523 (49.7%) 268 (25.5%) 1053

[0201] Larvae heterozygous for the exon 5 disruption developed into adult male and female mosquitoes with a sex ratio close to 1:1. On the contrary half of dsxF.-/- individuals developed into normal males whereas the other half showed the presence of both male and female morphological features as well as a number of developmental anomalies in the internal and external reproductive organs (intersex).

[0202] To establish the sex genotype of these dsxF.sup.-/- intersex, the inventors introgressed the mutation into a line containing a Y-linked visible marker (RFP) and used the presence of this marker to unambiguously assign sex genotype among individuals heterozygous and homozygous for the null mutation. This approach revealed that the intersex phenotype was observed only in genotypic females that were homozygous for the null mutation. The inventors saw no effect in heterozygous mutants, suggesting that the female-specific isoform of dsx is haplosufficient.

[0203] Examination of external sexually dimorphic structures in dsxF.sup.-/- genotypic females showed several phenotypic abnormalities including: the development of dorsally rotated male claspers (and absent female cerci), longer flagellomeres associated with male-like plumose antennae (FIG. 2). The analysis of the internal reproductive organs of these individuals failed to reveal the presence of fully developed ovaries and spermathecae; instead they were replaced by male-accessory glands (MAGs) and in some cases (~20%) by rudimentary pear-shaped organs resembling unstructured testes (FIG. 7).

[0204] Males carrying the dsxF- null mutation in heterozygosity or homozygosity showed wild type levels of fertility as measured by clutch size and larval hatching per mated female, as did heterozygous dsxF- female mosquitoes. On the contrary, intersex XX dsxF.sup.-/- female mosquitoes, though attracted to anaesthetised mice were unable to take a bloodmeal and failed to produce any eggs (FIG. 3).

[0205] The surprisingly drastic phenotype of dsxF.sup.-/- in females is proof of key functional role of exon 5 of dsx in the poorly understood sex differentiation pathway of *A. gambiae* mosquitoes and suggested that its sequence could represent a suitable target for gene drive approaches aimed at population suppression.

[0206] The inventors employed recombinase-mediated cassette exchange (RMCE) to replace the 3xP3::GFP transcription unit with a dsxF.sup.CRISPRh gene drive construct that consists of an RFP marker gene, a transcription unit to express the gRNA targeting dsxF, and the Cas9 gene under the control of the germline promoter of zero population growth (zpg) and its terminator sequence (FIG. 8). The zpg promoter has shown improved germline restriction of expression and specificity over the vasa promoter used in previous gene drive constructs (Hammond and Crisanti unpublished). Successful RMCE events that incorporated the dsxF.sup.CRISPRh into its target locus were confirmed in those individuals that had swapped the GFP for the RFP marker. During meiosis the Cas9/gRNA complex cleaves the wild type allele at the target sequence and the dsxF.sup.CRISPRh cassette is copied into wt locus via HDR ('homing'), disrupting exon 5 in the process.

[0207] The ability of the dsxF<sup>CRISPRh</sup> construct to home and bypass Mendelian inheritance was analysed by scoring the rates of RFP inheritance in the progeny of heterozygous parents (referred to as dsxF.sup.CRISPRh/+ hereafter) crossed to wild type mosquitoes. Surprisingly, high dsxF.sup.CRISPRh transmission rates of up to 100% were observed in the progeny of both heterozygous dsxF.sup.CRISPRh/+ male and female mosquitoes (FIG. 4a). The fertility of the dsxF.sup.CRISPRh line was also assessed to unravel potential negative effects due to ectopic expression of the nuclease in somatic cells and/or parental deposition of the nuclease into the newly fertilised embryos (FIG. 4b). These experiments showed that while heterozygous dsxF.sup.CRISPRh/+ males showed a fecundity rate (assessed as larval progeny per fertilised female) that did not differ from wild type males, heterozygous dsxF.sup.CRISPRh/+ female showed reduced fecundity overall (mean fecundity 49.8%±6.3% S.E., p<0.0001).

[0208] Surprisingly, the inventors noticed a more severe reduction in the fertility of heterozygous females when the drive allele was inherited from their father (mean fecundity 21.7%±8.6%) rather than their mother (64.9%±6.9%) (FIG. 9). Without wishing to be bound to any particular

theory, the inventors believe that this could be explained assuming a paternal deposition of active Cas9 nuclease into the newly fertilized zygote that stochastically induces conversion to of dsxF.sup.–, either through end-joining or HDR, in a significant number of cells resulting in a reduced fertility in females. Consistent with this hypothesis, some heterozygous females receiving a paternal dsxF.sup.CRISPRh allele showed a somatic mosaic phenotype that included, with varying penetrance, the absence of spermatheca and/or the formation of an incomplete clasper set. A mathematical model built considering the inheritance bias of the construct, the fecundity of heterozygous individuals, the phenotype of intersex as well as the paternal deposition of the nuclease on female fertility, indicated that the dsxF.sup.CRISPRh had the potential to reach 100% frequency in caged population in a span of 9-13 generations depending on starting frequency and stochasticity (FIG. 5a).

[0209] To test this hypothesis, caged wild type mosquito populations were mixed with individuals carrying the dsxF.sup.CRISPRh allele and subsequently monitored at each generation to assess the spread of the drive and quantify its effect on reproductive output. To mimic a hypothetical release scenario, the inventors started the experiment in two replicate cages putting together 300 wild type female mosquitoes with 150 wt male mosquitoes and 150 dsxF.sup.CRISPRh/+ male individuals and allowed them to mate. Eggs produced from the whole cage were counted and 650 eggs were randomly selected to seed the next generations. The larvae that hatched from the eggs were screened for the presence of the RFP marker to score the number of the progeny containing the dsxF.sup.CRISPRh allele in each generation. During the first three generations, the inventors observed in both caged populations an increase of the drive allele from 25% up to ~69% and thereafter they diverged. In cage 2 the drive reached 100% frequency by generation 7; in the following generation no eggs were produced and the population collapsed. In cage 1 the drive allele reached 100% frequency at generation 11 after drifting around 65% for two generations. This cage population also failed to produce eggs in the next generation. Though the two cages showed some apparent differences in the dynamics of spreading both curves fall within the prediction of the model (FIG. 5b). A summary of the cage trials is shown in table 5.

[0210] The inventors also monitored at different generations the occurrence of mutations at the target site to identify the occurrence of nuclease resistant functional variants. Amplicon sequencing of the target sequence from pooled population samples collected at generation 2, 3, 4 and 5 revealed the presence of several low frequency indels generated at the cleavage site, none of which appeared to encode for a functional AgdsxF transcript (FIGS. 10A-C). Accordingly, none of the variants identified showed any signs of positive selection as the drive progressively increased in frequency over generations, thus indicating that the selected target sequence has rigid functional and structural constraints. This notion is supported by the high degree of conservation of exon 5 in *A. gambiae* mosquitoes.sup.16, 17 and the presence of highly regulated splice site critical for the mosquito reproductive biology.

[0211] Heterozygous and homozygous individuals for the dsxF.sup.– allele were separated based on the intensity of fluorescence afforded by the GFP transcription unit within the knockout allele. Homozygous mutants were distinguishable as recovered in the expected Mendelian ratio of 1:2:1 suggesting that the disruption of the female-specific isoform of Agdsx is not lethal at the L1 larval stage.

TABLE-US-00038 TABLE 4 Genetic females homozygous for the insertion carry male-specific characteristics

Genetic Males	Genetic Females
Characteristic dsxF.sup.+/+ dsxF.sup.+/- dsxF.sup.-/- dsxF.sup.+/+ dsxF.sup.+/- dsxF.sup.-/-	Pupal genital male male male female female male
lobe Claspers ✓ ✓ ✓ X X ✓ Cercus X X X ✓ ✓ X Spermatheca X X X ✓ ✓ X MAGs ✓ ✓ ✓ X X ✓	Feed on blood X X X ✓ ✓ X Can lay eggs
X X X ✓ ✓ X Plumose ✓ ✓ ✓ X X ✓ antennae Pilose X X X ✓ ✓ X antennae	

[0212] The inventors assume that parental effects on fitness (egg production and hatching rates) for non-drive (W/W, W/R) females with nuclease from one or both parents are the same as observed values for drive heterozygote (W/D) females with parental effects. For combined maternal and paternal effects (nuclease from both parents), the minimum of the observed values for maternal and paternal effect is assumed.

TABLE-US-00039 TABLE 5 Summary of values obtained from the cage trials

Cage Trial 1	Cage Trial 2	Transgenic	Hatching	Egg Repr.	Transgenic
Hatching	Egg Repr.	Rate	Rate	Output	Load
Generation (%)	(%)	(N)	(%)	(%)	(%)
G0 25 — 27462 — 25 —					
26895 — (150/600)	(150/600)	G1 49.65	88.62	17405	36.62
50	86.15	16578	38.36	(268/576)	(576/650)
(280/560)	(560/650)	G2 62.01	74.92	14957	45.54
61.79	80.92	15565	42.13	(302/487)	(487/650)
(325/526)	(526/650)	G3 68.94	76.77	11249	59.04
68.05	74.15	9376	65.14	(344/499)	(499/650)
(328/482)	(482/650)	G4 67.67	71.85	9170	66.61
85.41	71.69	6514	75.78	(316/467)	(467/650)
(398/466)	(466/650)	G5 58.67	69.23	11364	58.62
86.5	61.54	4805	81.13	(264/450)	(450/650)
(346/400)	(400/650)	G6 63.3	70	7727	71.86
90.09	52.77	4210	84.35	(288/455)	(455/650)
(309/343)	(343/650)	G7 69.47	78.62	7785	71.65
100	55.85	1668	93.8	(355/511)	(511/650)
(363/363)	(363/650)	G8 70.07	70.92	6293	77.08
100	42.77	0	100	(323/461)	(461/650)
(278/278)	(278/650)	G9 75.58	66.15	4107	85.04
—	—	—	—	(325/430)	(430/650)
G10 95.71	57.38	4146	84.90	(357/373)	(373/650)
G11 100	57.54	2645	90.37	(374/374)	(374/650)
G12 100	38.92	0	100	(253/253)	(253/650)

[0213] Transgenic rate, hatching rate, egg output and reproductive load at each generation during the cage experiment. The reproductive load indicates the suppression of egg production at each generation compared to the first generation.

## Conclusions

[0214] In the human malaria vector, *Anopheles gambiae*, the gene doublesex (Agdsx) encodes two alternatively spliced transcripts dsx-female (AgdsxF) and dsx-male (AgdsxM) that, in turn, regulate the activation of distinct subordinate genes responsible for the differentiation of the two sexes. The female transcript, unlike AgdsxM, contains an exon (exon 5) whose coding sequence is highly conserved in all *Anopheles* mosquitoes so far analysed. CRISPR-Cas9 targeted disruption of the intron 4-exon 5 sequence boundary aimed at blocking the formation of functional AgdsxF did not affect male development or fertility, whereas females homozygous for the disrupted allele showed an intersex phenotype characterised by the presence of male internal and external reproductive organs and complete sterility, as summarised in table 4. A CRISPR-Cas9 gene drive construct targeting this same sequence was able to spread rapidly in caged mosquito populations reaching 100% prevalence within a span of 8-12 generations while progressively reducing the egg production to the point of total population collapse. Notably, this drive solution did not induce resistance. A variety of non-functional Cas9 resistant variants were generated in each generation at the target site, they all failed to block the spread of the drive.

[0215] Hence, these data all together provide important functional insights on the role of dsx in *A. gambiae* sex determination while demonstrating substantial progress towards the development of effective gene drive vector control measures aimed at population suppression. Without wishing to be bound to any particular theory, the intersex phenotype of dsxF–/– genetic females demonstrates that exon 5 is critical for the production of a functional female transcript. Furthermore, the observation that heterozygous dsxFCRISPRh/+ females are fertile and produce nearly 100% transformed progeny would indicate that the majority of the germ cells in these females are homozygous and, unlike somatic cells, do not undergo autonomous dsx-mediated sex commitment.sup.18. The development of a gene drive solutions capable of collapsing a human malaria vector population is a long sought scientific and technical achievement.sup.19. The gene drive dsxFCRISPRh targeting exon 5 of dsx showed a number of desired efficacy features for field applications, in term of inheritance bias, fertility of heterozygous individuals, phenotype of homozygous females and apparent lack of nuclease-resistant functional variants at the target site.

## Example 2

[0216] A promising approach to mitigate resistance to gene drive is to target multiple sites at the same time in a strategy analogous to combinational drug therapy. For resistance to get selected against the gene drive, resistant mutations would have to be simultaneously present at all target sites, and co-operatively restore the targeted gene's original function. Note that homing will also serve to remove resistant mutations generated if at least one of the targeted sites is still cleavable.

[0217] Exon 5 of doublesex that was targeted with a gene drive as described in Example 1 contains a total of four invariant target sites that are amenable to multiplexing (FIG. 12). Accordingly, the inventors then generated a novel multiplexed gene drive targeting the original target site at doublesex (T1) and a new target site (T3) present at the 3' end of the exon 5 coding sequence. The transgenic line that was obtained contains a CRISPR construct bearing a 3×P3::RFP marker, Cas9 expressed under the zpg promoter and two multiplexed U6::gRNA expression cassettes as shown in FIG. 13.

[0218] The inheritance bias of the gene drive, and fertility of gene drive carriers was assessed through phenotype assays. Gene drive heterozygotes of both sexes that had inherited the drive from either males or females were crossed to wild-type individuals and females of each cross were allowed to lay eggs individually. The same was done with a wild-type cage, as a control. Egg and larval output of each female was counted, as soon as they laid and hatched respectively. Larvae were then screened for RFP fluorescence indicative of gene drive presence. The mating status of females that did not give offspring was determined by dissecting their spermathecae and examining it under an EVOS cell imaging microscope for the presence of spermatozoa. Females that showed no evidence of mating were all included in the analysis as having given 0 progeny, since mating competence can be affected by carrying the doublesex gene drive. The results from Kyrou et al. (2018) were adapted to also include unmated individuals in the analysis.

[0219] The results revealed that the novel multiplexed gene drive can successfully bias its inheritance to the next generation with transmission rates comparable to the single-guide gene drive we previously developed ( $p > 0.05$ ) or higher ( $p = 0.04$ ), when the gene drive was transmitted by a male carrier who inherited it maternally (F->M class) (FIGS. 14A and 14B). As with the original doublesex gene drive, the fertility of gene drive carrier females descended from transgenic males (M->F class) was decreased compared to all other classes (FIGS. 14C and 14D). The total and relative number of average larval progeny of females that inherited the gene drive from males (M->F class), is surprisingly higher for the multiplexed gene drive (FIGS. 14C and 14D).

## REFERENCES

[0220] 1. Gantz, V. M. et al. Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*. Proc Natl Acad Sci USA 112, E6736-6743 (2015). [0221] 2. Hammond, A. et al. A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. Nat Biotechnol 34, 78-83 (2016). [0222] 3. Burt, A. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. Proc Biol Sci 270, 921-928 (2003). [0223] 4. Deredec, A., Godfray, H. C. & Burt, A. Requirements for effective malaria control with homing endonuclease genes. Proc Natl Acad Sci USA 108, E874-880 (2011). [0224] 5. Hamilton, W. D. Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. Science 156, 477-488 (1967). [0225] 6. Galizi, R. et al. A synthetic sex ratio distortion system for the control of the human malaria mosquito. Nat Commun 5, 3977 (2014). [0226] 7. Magnusson, K. et al. Demasculinization of the *Anopheles gambiae* X chromosome. BMC Evol Biol 12, 69 (2012). [0227] 8. Champer, J. et al. Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. PLoS Genet 13, e1006796 (2017). [0228] 9. Hammond, A. M. et al. The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. PLoS Genet 13, e1007039 (2017). [0229] 10. Marshall, J. M., Buchman, A., Sanchez, C. H. & Akbari, O. S. Overcoming evolved resistance to population-suppressing homing-based gene drives. Sci Rep 7, 3776 (2017). [0230] 11. Unckless, R. L., Clark, A. G. & Messer, P. W. Evolution of Resistance Against CRISPR/Cas9 Gene Drive. Genetics 205, 827-841 (2017). [0231] 12. Burtis, K. C. & Baker, B. S. *Drosophila* doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. Cell 56, 997-1010 (1989). [0232] 13. Graham, P., Penn, J. K. & Schedl, P. Masters change, slaves remain. Bioessays 25, 1-4 (2003). [0233] 14. Krzywinska, E., Dennison, N. J., Lycett, G. J. & Krzywinski, J. A maleness gene in the malaria mosquito *Anopheles gambiae*. Science 353, 67-69 (2016). [0234] 15. Scali, C., Catteruccia, F., Li, Q. & Crisanti, A. Identification of sex-specific transcripts of the *Anopheles gambiae* doublesex gene. J Exp Biol 208, 3701-3709 (2005). [0235] 16. Neafsey, D. E. et al. Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes. Science 347, 1258522 (2015). [0236] 17. *Anopheles gambiae* Genomes, C. et al. Genetic diversity of the African malaria vector *Anopheles gambiae*. Nature 552, 96-100 (2017). [0237] 18. Murray, S. M., Yang, S. Y. & Van Doren, M. Germ cell sex determination: a collaboration between soma and germline. Curr Opin Cell Biol 22, 722-729 (2010). [0238] 19. Curtis, C. F. Possible use of translocations to fix desirable genes in insect pest populations. Nature 218, 368-369 (1968). [0239] 20. National Academies of Sciences, E. & Medicine Gene Drives on the Horizon: Advancing Science, Navigating Uncertainty, and Aligning Research with Public Values. (The National Academies Press, Washington, DC; 2016). [0240] 21. Papathanos, P. A., Windbichler, N., Menichelli, M., Burt, A. and Crisanti, A. The vasa regulatory region mediates germline expression and maternal transmission of proteins in the malaria mosquito *Anopheles gambiae*: a versatile tool for genetic control strategies. BMC Mol Biol 10, 65, (2009). [0241] 22 Hammond, A. M. et al. The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. PLoS Genet 13, e1007039 (2017). [0242] 23. Wolfram Research, Inc., 2017 Mathematica 11.2, Champaign, IL.

## Claims

1. A gene drive genetic construct capable of disrupting an intron-exon boundary of the female-specific splice form of the doublesex gene in an arthropod.
2. The gene drive genetic construct according to claim 1, wherein the arthropod is an insect.
- 3.-47. (canceled)
48. The gene drive genetic construct according to claim 2, wherein the insect is selected from the group consisting of an Anopheline species, an *Aedes* species, *Ceratitis capitata* and a *Drosophila* species.
49. The gene drive genetic construct according to claim 48, wherein the insect is an Anopheline species, selected from a group consisting of *Anopheles gambiae*, *Anopheles coluzzi*, *Anopheles merus*, *Anopheles arabiensis*, *Anopheles quadriannulatus*, *Anopheles stephensi*, *Anopheles funestus*, and *Anopheles melas*.
50. The gene drive genetic construct according to claim 49, wherein the arthropod is *Anopheles gambiae*.
51. The gene drive genetic construct according to claim 1, wherein the doublesex gene comprises a nucleic acid sequence as set out in SEQ ID NO: 1, or a fragment or variant thereof.
52. The gene drive genetic construct according to claim 1, wherein the intron-exon boundary is the boundary between intron 4 and exon 5 of the doublesex gene.
53. The gene drive genetic construct according to claim 1, wherein the intron-exon boundary comprises the nucleotide sequence as set out in SEQ ID NO: 2, 3 or 4, or a fragment or variant thereof.
54. The gene drive genetic construct according to claim 52, wherein the intron-exon boundary is targeted by a genetic construct having a target sequence including up to 1, 2, 3, 4, 5, 10 or 15 nucleotides 5' and/or 3' of SEQ ID No:2, 3 or 4 and said gene drive genetic construct is a nuclease-based genetic construct.
55. The gene drive genetic construct according claim 54, wherein the nuclease-based genetic construct is selected from a group consisting of a transcription activator-like effector nuclease (TALEN) genetic construct, Zinc finger nuclease (ZFN) genetic construct, and a CRISPR-based gene drive genetic construct.
56. The gene drive genetic construct according to claim 55, wherein the gene drive genetic construct is a CRISPR-based gene drive construct selected from the group consisting of a CRISPR-Cpf1-based or a CRISPR-Cas9-based gene drive genetic construct.
57. The gene drive genetic construct according to claim 55, wherein the genetic construct comprises a first nucleotide sequence encoding a nucleotide sequence that is capable of hybridising to the intron-exon boundary of the doublesex gene.
58. The gene drive genetic construct according to claim 57 wherein the genetic construct comprises a first nucleotide sequence encoding a nucleotide sequence that is capable of hybridising to the intron-exon boundary of the doublesex gene is a guide RNA.
59. The gene drive genetic construct according to claim 57, wherein the first nucleotide sequence encoding a nucleotide sequence that is capable of

hybridising to the intron-exon boundary of the doublesex (dsx) gene comprises a nucleic acid sequence as set out in SEQ ID NO: 5 or 6, or a fragment or variant thereof.

**60.** The gene drive genetic construct according to claim 57, wherein the nucleotide sequence which is encoded by the first nucleotide sequence and which is capable of hybridising to the intron-exon boundary of the doublesex (dsx) gene comprises a nucleic acid sequence as set out in SEQ ID NO: 58 or 48, or a fragment or variant thereof.

**61.** The gene drive genetic construct according to claim 55, wherein the gene drive genetic construct further comprises a second nucleotide sequence encoding a CRISPR nuclease.

**62.** The gene drive genetic construct according to claim 55, wherein the gene drive genetic construct further comprises a second nucleotide sequence encoding a CRISPR nuclease, wherein the second nucleotide sequence encodes a Cpf1 or Cas9 nuclease.

**63.** The gene drive genetic construct according to claim 55, wherein the gene drive genetic construct further comprises at least one promoter sequence, which drives expression of the first and second nucleotide sequence.

**64.** The gene drive genetic construct according to claim 63, wherein the gene drive genetic construct comprises a first promoter sequence operably linked to the first nucleotide sequence and a second promoter sequence operably linked to the second nucleotide sequence.

**65.** The gene drive genetic construct according to claim 64, wherein the first promoter is a polymerase III promoter, or wherein the first promoter sequence comprises a nucleic acid sequence substantially as set out in SEQ ID No: 49, or a variant or fragment thereof.

**66.** The gene drive genetic construct according to claim 64, wherein the second promoter sequence is a promoter sequence that restricts expression of the second nucleotide sequence to germline cells of the arthropod.

**67.** The gene drive genetic construct according to claim 66 wherein the second promoter sequence is: (i) zpg, wherein the second promoter sequence comprises a nucleic acid sequence as set out in SEQ ID No: 7, or a variant or fragment thereof; (ii) nos, wherein the second promoter sequence comprises a nucleic acid sequence as set out in SEQ ID No: 8, or a variant or fragment thereof; (iii) exu, wherein the second promoter sequence comprises a nucleic acid sequence as set out in SEQ ID No: 9, or a variant or fragment thereof; or (iv) vasa2, wherein the second promoter sequence comprises a nucleic acid sequence as set out in SEQ ID No: 10, or a variant or fragment thereof.

**68.** The gene drive genetic construct according to claim 55, wherein the construct further comprises attB integrase attachment sites which flank the first nucleotide sequence encoding the nucleotide sequence that is capable of hybridising to the intron-exon boundary of the doublesex gene, the second nucleotide sequence encoding the nuclease, the first promoter sequence and the second promoter sequence.

**69.** The gene drive genetic construct according to claim 55, wherein the construct further comprises third and fourth nucleotide sequences which flank the first nucleotide sequence encoding the nucleotide sequence that is capable of hybridising to the intron-exon boundary of the doublesex gene, the second nucleotide sequence encoding the nuclease, the first promoter sequence and the second promoter sequence, wherein the third and fourth nucleotides are homologous to the genomic sequences flanking the intron-exon boundary, such that the gene drive construct is integrated into the disrupted intron-exon boundary via homology-directed repair.

**70.** The gene drive genetic construct according to claim 69, wherein the third nucleotide sequence comprises a nucleic acid sequence as set out in SEQ ID No: 11, or a variant or fragment thereof and/or wherein the fourth nucleotide sequence comprises a nucleic acid sequence as set out in SEQ ID No: 12, or a variant or fragment thereof.

**71.** The gene drive genetic construct according to claim 1, wherein the gene drive construct comprises a nucleic acid sequence as set out in SEQ ID NO: 13, or a fragment or variant thereof.

**72.** The gene drive genetic construct according to claim 1, wherein the construct is capable of targeting (i) a first target site which comprises the intron-exon boundary of the female specific splice form of the doublesex (dsx) gene, and (ii) a second target site disposed in exon 5 of the female specific splice form of the doublesex (dsx) gene.

**73.** The gene drive genetic construct according to claim 72, wherein the second target site comprises a nucleic acid sequence, which is disposed in the sequence as set out in SEQ ID No: 35, 36 (T2), 37 (T3) or 38 (T4) or a variant or fragment thereof, or wherein the second target site includes up to 1, 2, 3, 4, 5, 10 or 15 nucleotides 5' and/or 3' of SEQ ID No:35, 36, 37 or 38.

**74.** The gene drive genetic construct according to claim 72, wherein the gene drive construct targets: (i) SEQ ID NO: 4, or up to 1, 2, 3, 4, 5, 10 or 15 nucleotides 5' and/or 3' of SEQ ID No:4; and (ii) SEQ ID No:3, or up to 1, 2, 3, 4, 5, 10 or 15 nucleotides 5' and/or 3' of SEQ ID No:3.

**75.** The gene drive genetic construct according to claim 72, wherein the construct comprises: (i) a first nucleotide sequence encoding a first guide RNA which is capable of hybridising to a first target site which is an intron-exon boundary of the female specific splice form of the doublesex (dsx) gene, and (ii) a fifth nucleotide sequence encoding a second guide RNA which is capable of hybridising to a second target site disposed in exon 5 of the female specific splice form of the doublesex (dsx) gene.

**76.** The gene drive genetic construct according to claim 75, wherein the first and/or fifth nucleotide sequence encodes a guide RNA.

**77.** The gene drive genetic construct according to claim 75, wherein the fifth nucleotide sequence encoding a nucleotide sequence that is capable of hybridising to the second target site comprises a nucleic acid sequence as set out in any one of SEQ ID NO: 39-44, or a fragment or variant thereof.

**78.** The gene drive genetic construct according to claim 73, wherein the nucleotide sequence which is encoded by the fifth nucleotide sequence and which is capable of hybridising to the second target site comprises a nucleic acid sequence as set out in any one of SEQ ID NO: 45, 46, 47, 59, 60 or 61 or a fragment or variant thereof.

**79.** The gene drive genetic construct according to claim 75, wherein the construct comprises a first promoter sequence operably linked to the first nucleotide sequence, a second promoter sequence operably linked to the second nucleotide sequence, and a third promoter sequence operably linked to the fifth nucleotide sequence.

**80.** The gene drive genetic construct according to claim 75, wherein the construct further comprises att integrase attachment sites which flank the first nucleotide sequence encoding the nucleotide sequence that is capable of hybridising to the first target site which is an intron-exon boundary of the female specific splice form of the doublesex (dsx) gene, and the fifth nucleotide sequence capable of hybridising to a second target site disposed in exon 5 of the female specific splice form of the doublesex (dsx) gene, the second nucleotide sequence encoding the nuclease, the first promoter sequence, the second promoter sequence and the third promoter sequence.

**81.** The gene drive genetic construct according to claim 75, wherein the construct further comprises sixth and seventh nucleotide sequences which flank the first nucleotide sequence encoding the nucleotide sequence that is capable of hybridising to the first target site which is an intron-exon boundary of the female specific splice form of the doublesex (dsx) gene, and the fifth nucleotide sequence capable of hybridising to a second target site disposed in exon 5 of the female specific splice form of the doublesex (dsx) gene, the second nucleotide sequence encoding the nuclease, the first promoter sequence, the second promoter sequence and the third promoter sequence, wherein the sixth and seventh nucleotides are homologous to the genomic sequences flanking the two cut-sites which are disposed in exon 5 of the arthropod, such that when the docking construct is introduced into the arthropod, it is integrated into the arthropod's genome by homology directed repair.

**82.** The gene drive genetic construct according to claim 81, wherein the sixth nucleotide sequence comprises a nucleic acid sequence as set out in SEQ ID No: 11, or a variant or fragment thereof and/or wherein the seventh nucleotide sequence comprises a nucleic acid sequence as set out in SEQ ID No: 50, or a variant or fragment thereof.

**83.** The gene drive genetic construct according to claim 72, wherein the gene drive construct comprises a nucleic acid sequence substantially as set out in SEQ ID NO: 51, or a fragment or variant thereof.

**84.** A method for preventing or reducing the inclusion of at least one exon into the female specific splice form of arthropod doublesex mRNA, when said mRNA is produced by splicing from a precursor mRNA transcript, the method comprising contacting one or more cells of an arthropod, one or

more cells of an arthropod embryo, in vitro or ex vivo, under conditions conducive to uptake of the gene drive genetic construct according to claim by such cell, and allowing splicing to take place.

**85.** A method of producing a genetically modified arthropod, the method comprising introducing into an arthropod a gene drive genetic construct capable of disrupting an intron/exon boundary of the female specific splice form of doublesex gene in an arthropod.

**86.** An arthropod obtained or obtainable by the method according to claim 84.

**87.** A genetically modified arthropod comprising a disrupted intron-exon boundary of the female specific splice form doublesex gene, wherein the genetically modified arthropod is targeted using the gene drive genetic construct as defined in claim 1.

**88.** A method of suppressing a wild type arthropod population comprising breeding a genetically modified arthropod comprising an intron-exon boundary of the female exon of doublesex gene that has been disrupted by a gene drive genetic construct, with a wild type population of the arthropod, such that when the gene drive construct is expressed in offspring of the genetically modified arthropod and wild type arthropod, it disrupts the doublesex gene contributed by the wild type population.

**89.** The method according to claim 88, wherein the genetic construct is as defined in claim 1.

**90.** A nucleic acid comprising a nucleotide sequence substantially as set out as any one of SEQ ID No: 6-34, 42-48, 50-57 or a fragment or variant thereof.

**91.** A guide RNA comprising any one of SEQ ID No:58 to 61 and a nuclease binding region.

**92.** A guide RNA according to claim 91, wherein the nuclease binding region binds to, or complexes with, a CRISPR nuclease, which is a Cas endonuclease, Cas9 or Cpf1.

**93.** A guide RNA according to claim 92, wherein the guide RNA comprises trans-activating CRISPR RNA (tracrRNA) and a CRISPR RNA (crRNA), or a single guide RNA (sgRNA).

---