



US 20250255283A1

(19) United States

(12) Patent Application Publication (10) Pub. No.: US 2025/0255283 A1
(43) Pub. Date: Aug. 14, 2025

(54) GENE DRIVE TARGETING FEMALE DOUBLESEX SPlicing IN ARTHROPODS

(71) Applicant: IMPERIAL COLLEGE INNOVATIONS LIMITED, London (GB)

(72) Inventors: Andrea Crisanti, London (GB); Kyros Kyrou, London (GB); Andrew Hammond, London (GB)

(73) Assignee: IMPERIAL COLLEGE INNOVATIONS LIMITED, London (GB)

(21) Appl. No.: 19/191,416

(22) Filed: Apr. 28, 2025

Related U.S. Application Data

(63) Continuation of application No. 17/253,553, filed on Dec. 17, 2020, filed as application No. PCT/GB2019/051757 on Jun. 21, 2019.

(30) Foreign Application Priority Data

Jun. 22, 2018 (GB) 1810253.3

Publication Classification

(51) Int. Cl.

A01K 67/68 (2025.01)
C12N 15/85 (2006.01)

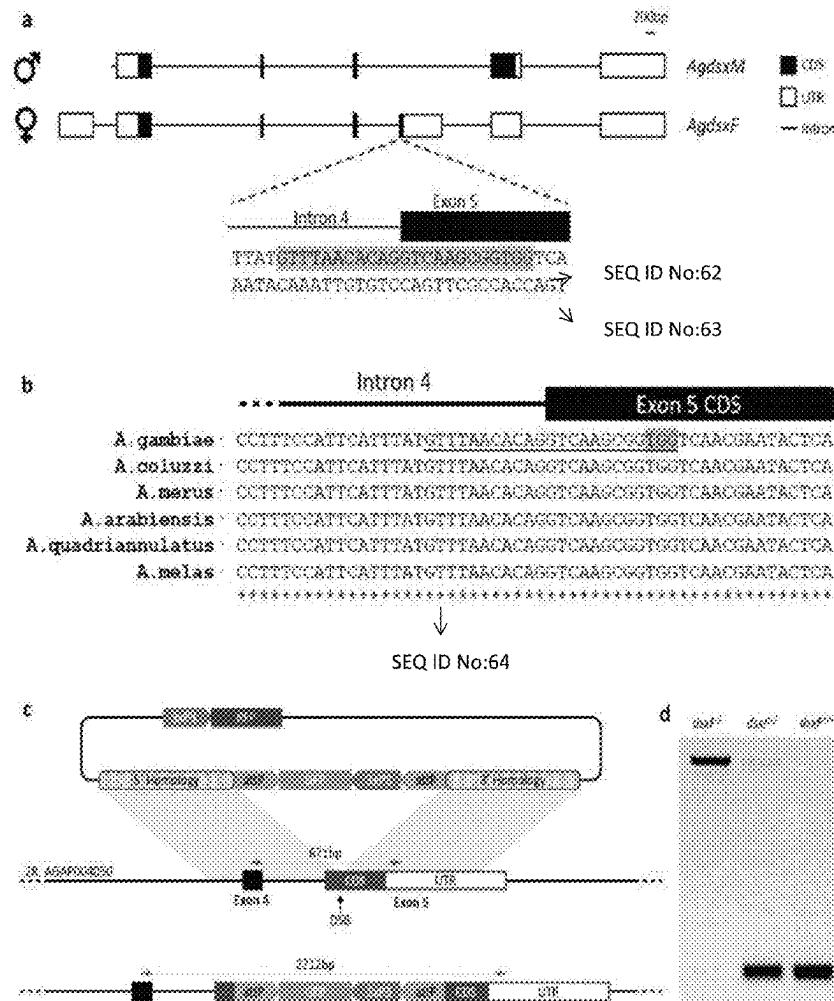
(52) U.S. Cl.

CPC *A01K 67/68* (2025.01); *C12N 15/8509* (2013.01); *A01K 2217/07* (2013.01); *A01K 2227/706* (2013.01); *A01K 2267/02* (2013.01)

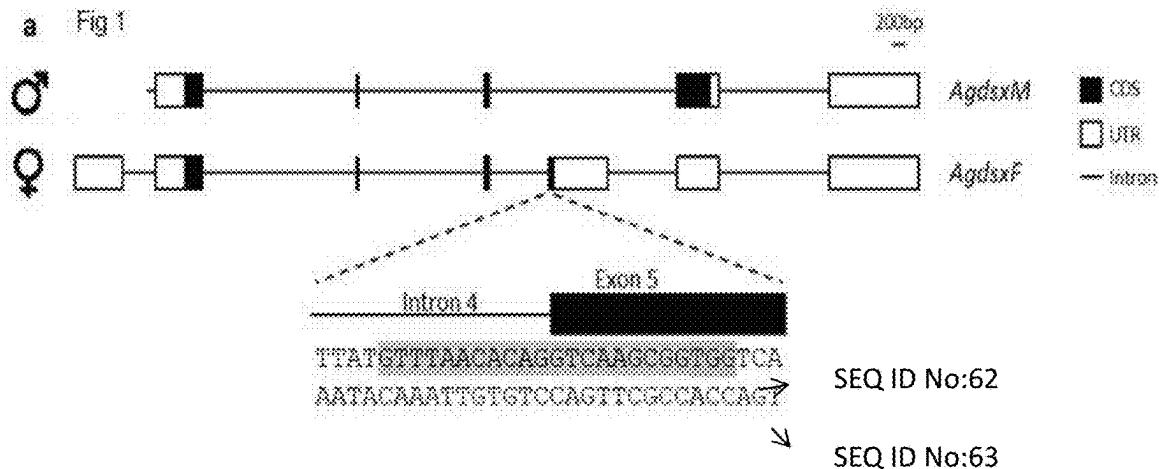
(57)

ABSTRACT

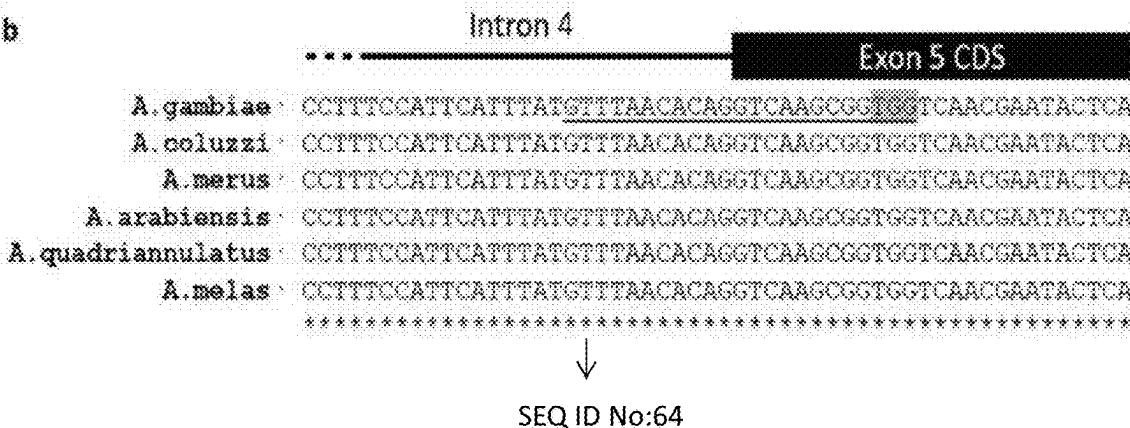
The invention relates to gene drives, and in particular to genetic sequences and constructs for use in a gene drive. The invention is especially concerned with ultra-conserved and ultra-constrained sequences for use as a gene drive target with the aim of overcoming the development of resistance to the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

Specification includes a Sequence Listing.

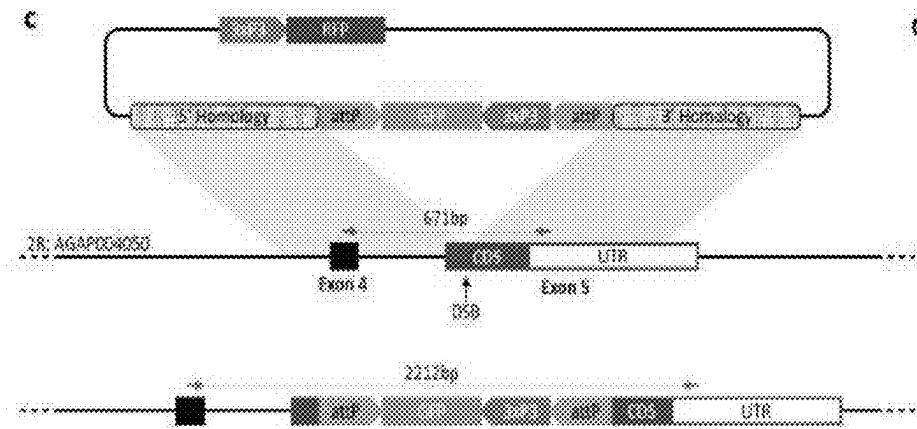
a Fig 1



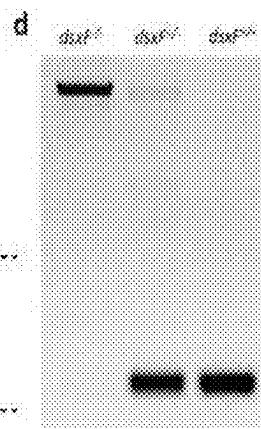
b



c



d



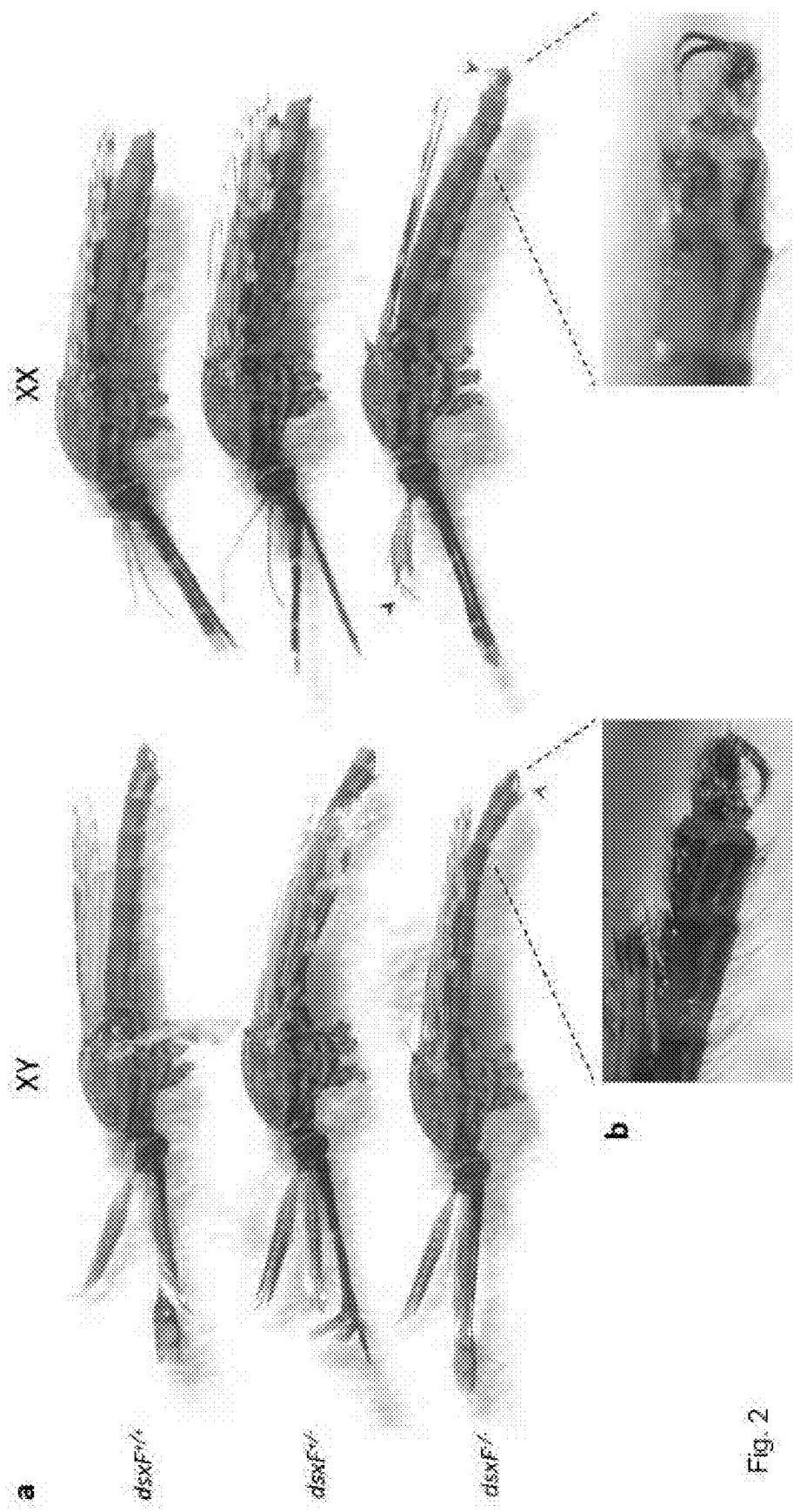


Fig. 2

Figure 3

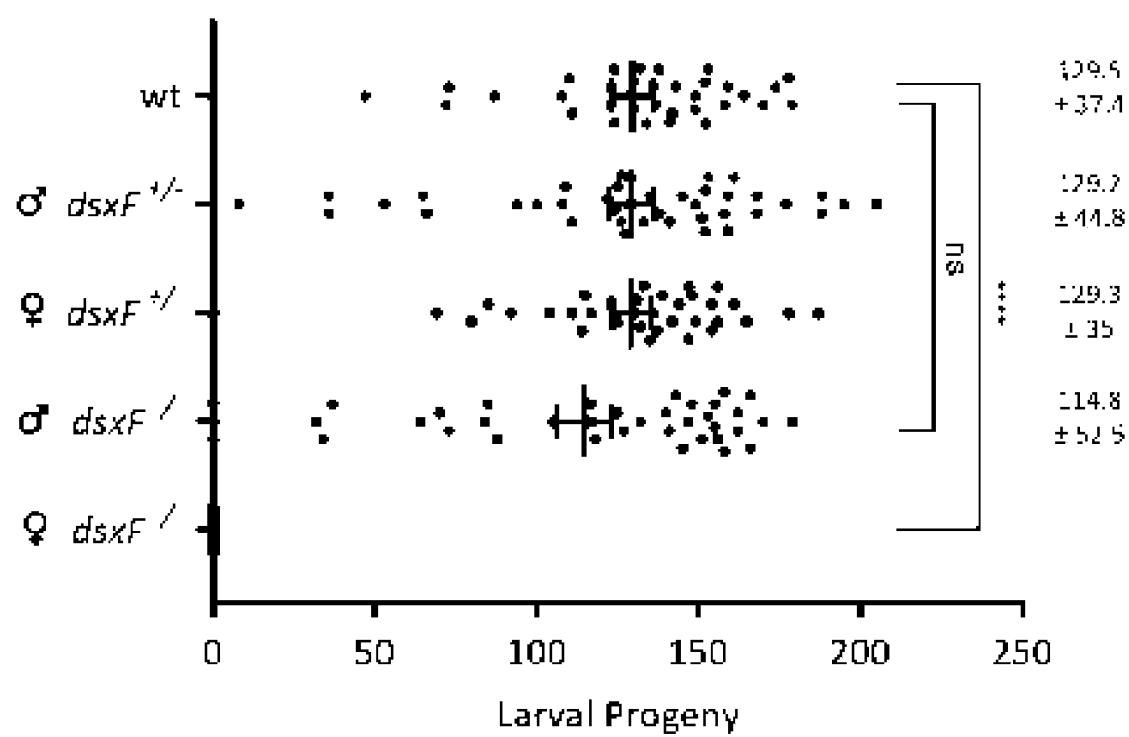


Figure 4

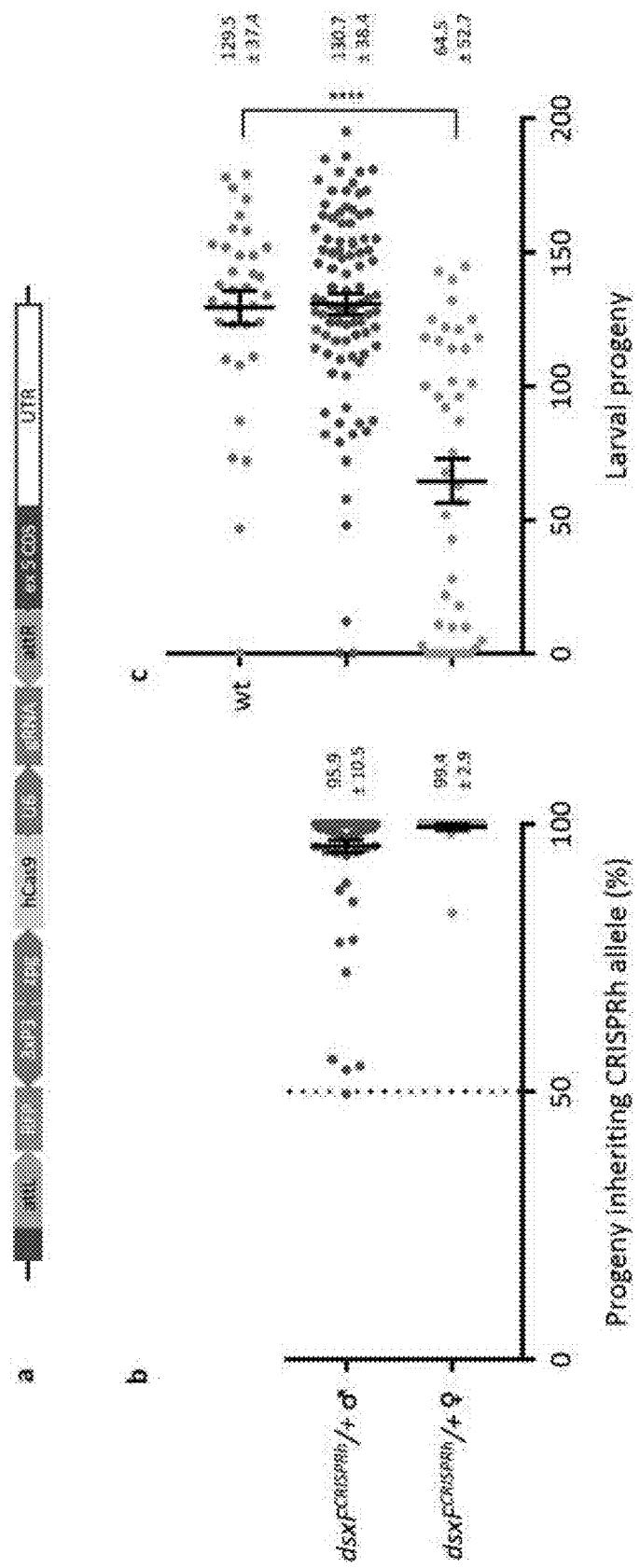


Figure 5

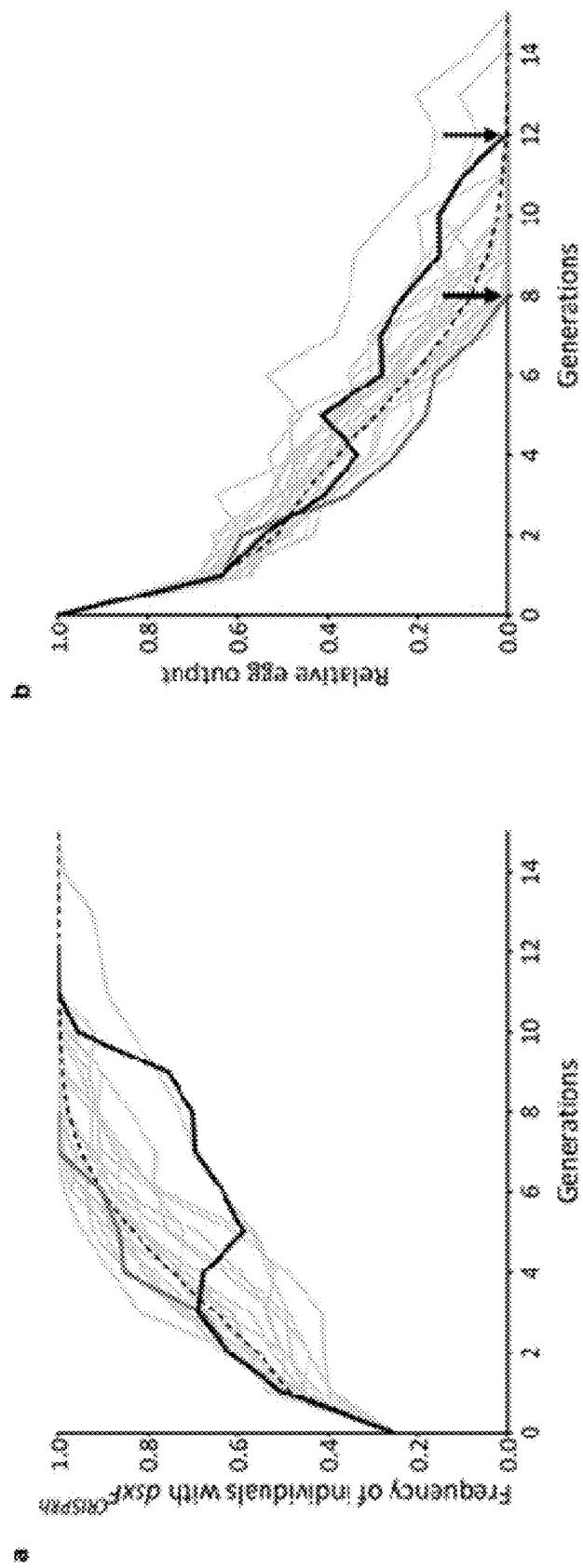


Figure 6

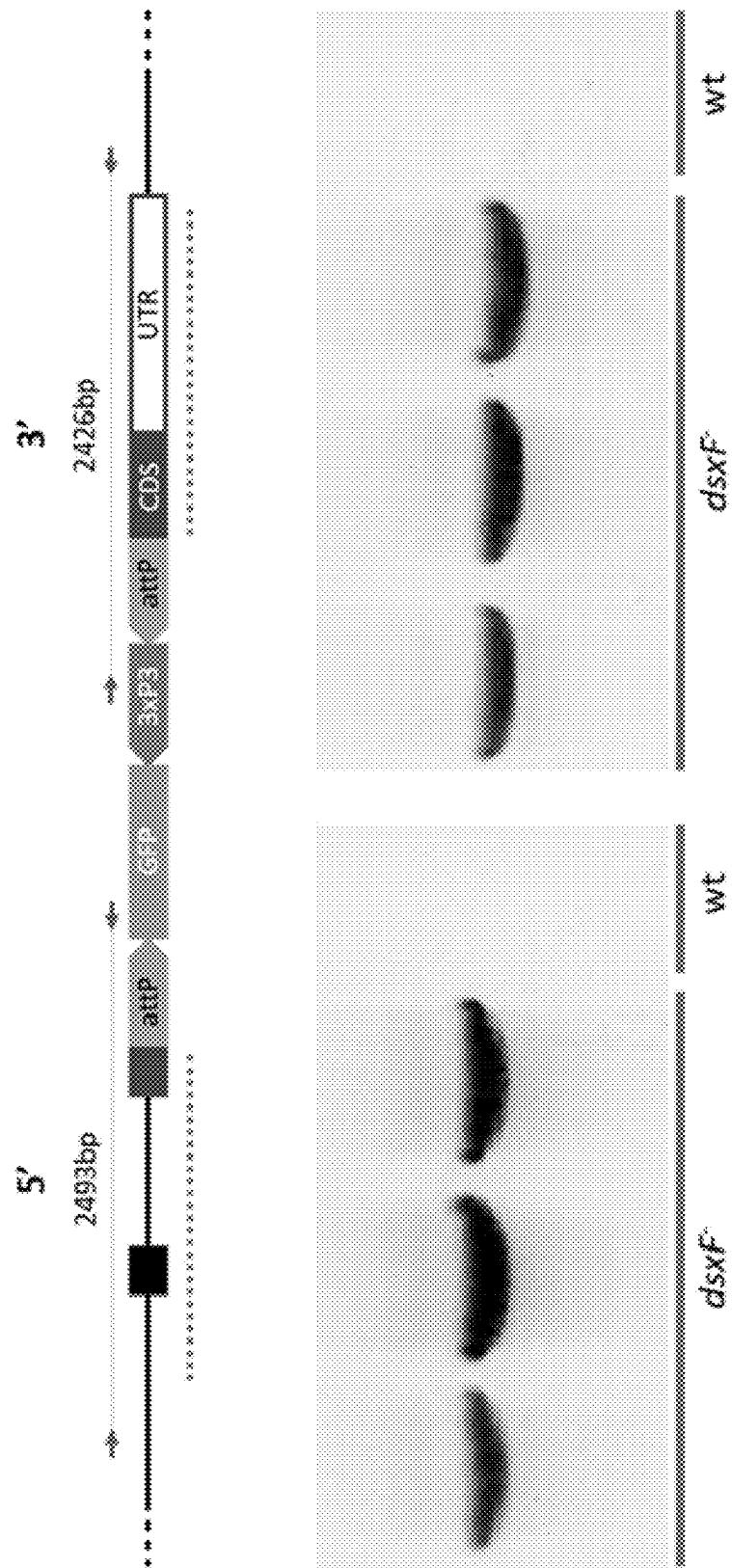


Figure 7

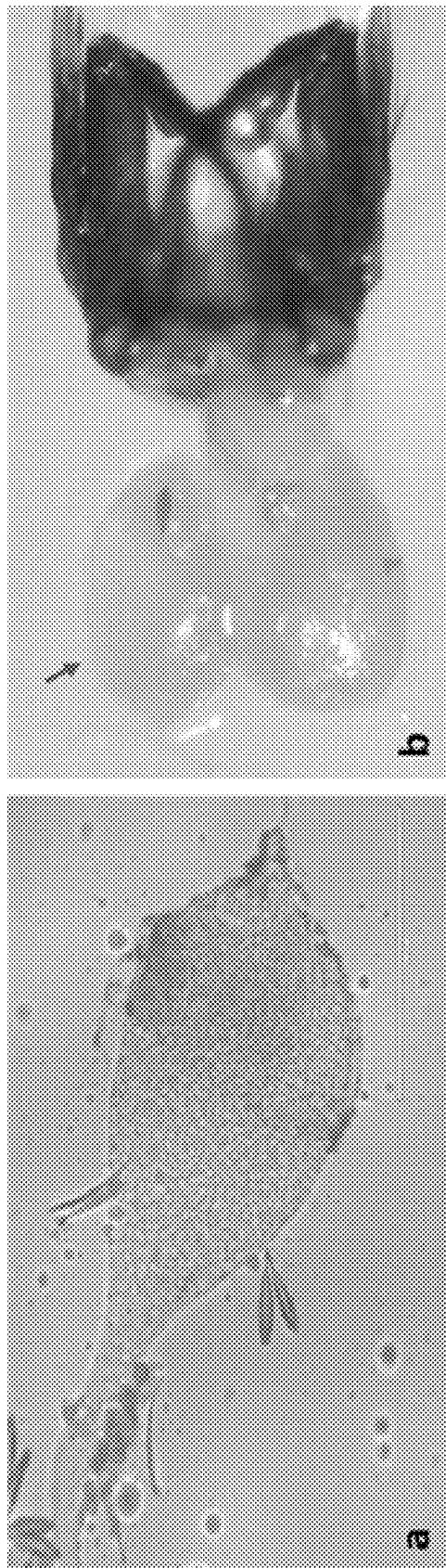


Figure 8

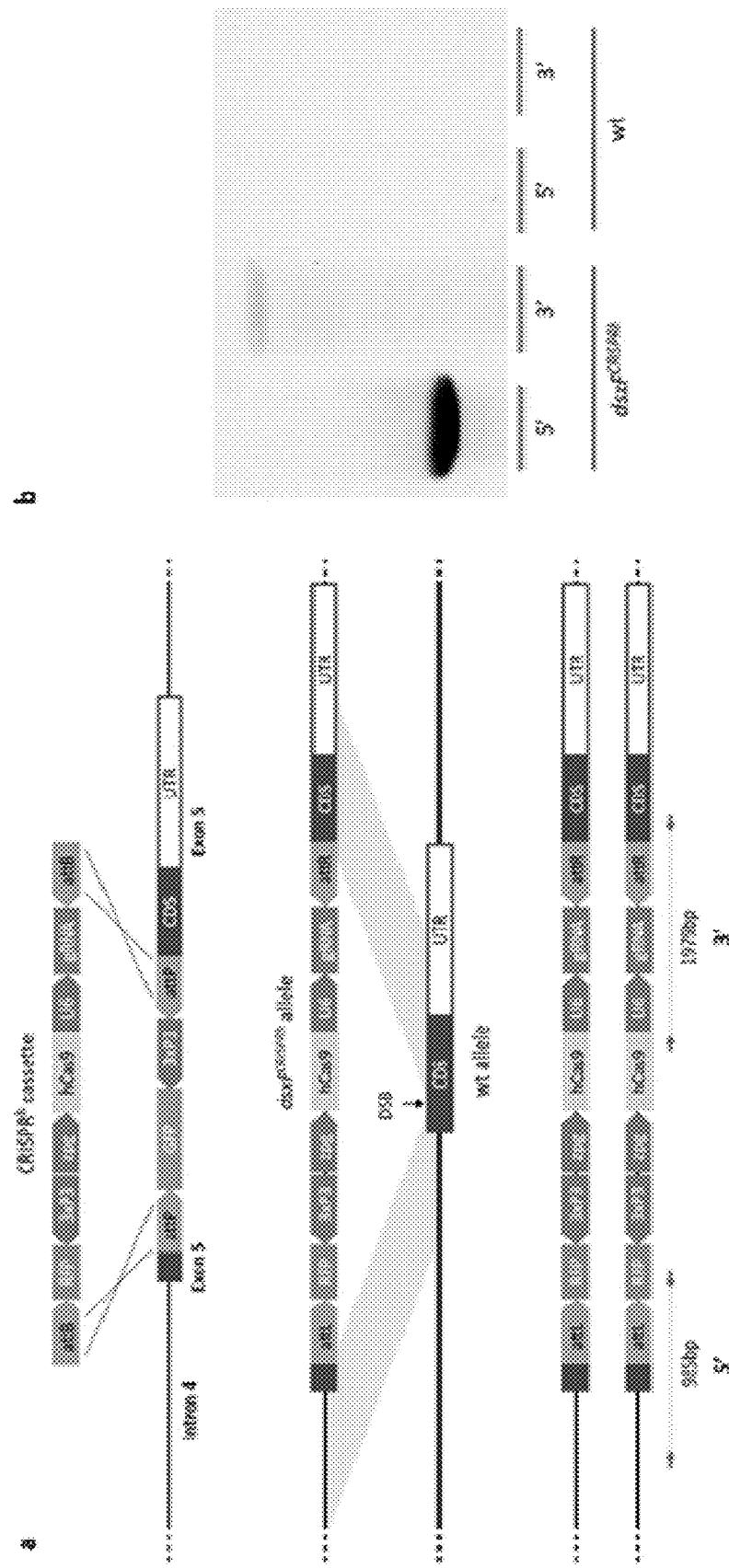


Figure 9

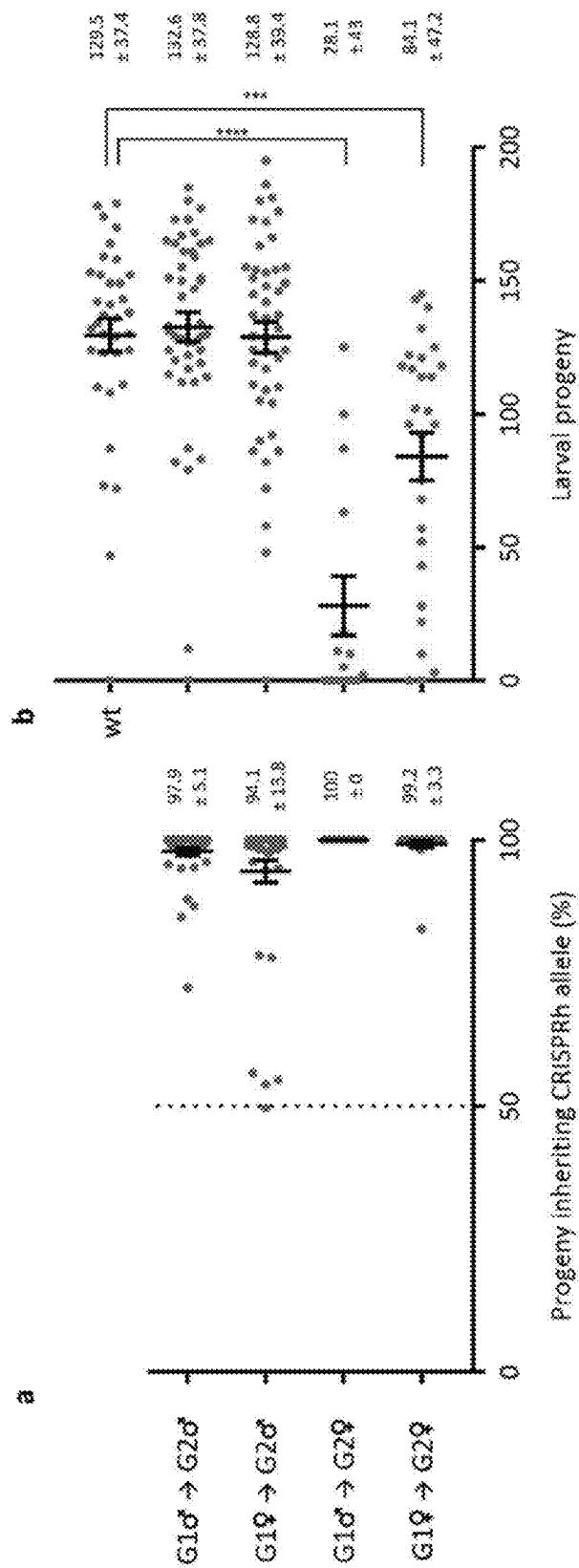


Figure 10A

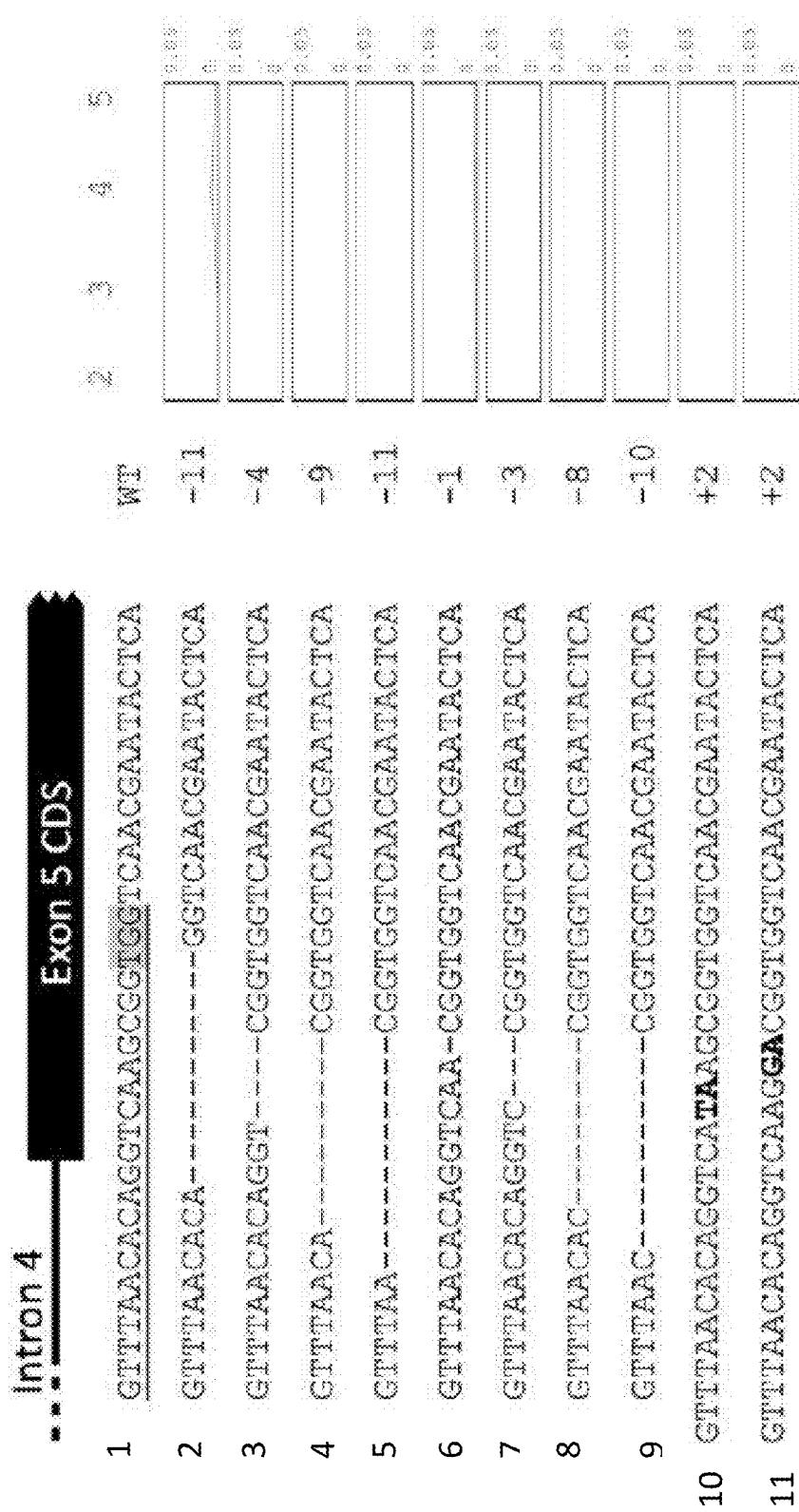


Figure 10A continued

- 1- SEQ ID No: 65
- 2- SEQ ID No: 66
- 3- SEQ ID No: 67
- 4- SEQ ID No: 68
- 5- SEQ ID No: 69
- 6- SEQ ID No: 70
- 7- SEQ ID No: 71
- 8- SEQ ID No: 72
- 9- SEQ ID No: 73
- 10- SEQ ID No: 74
- 11- SEQ ID No: 75

Figure 10B

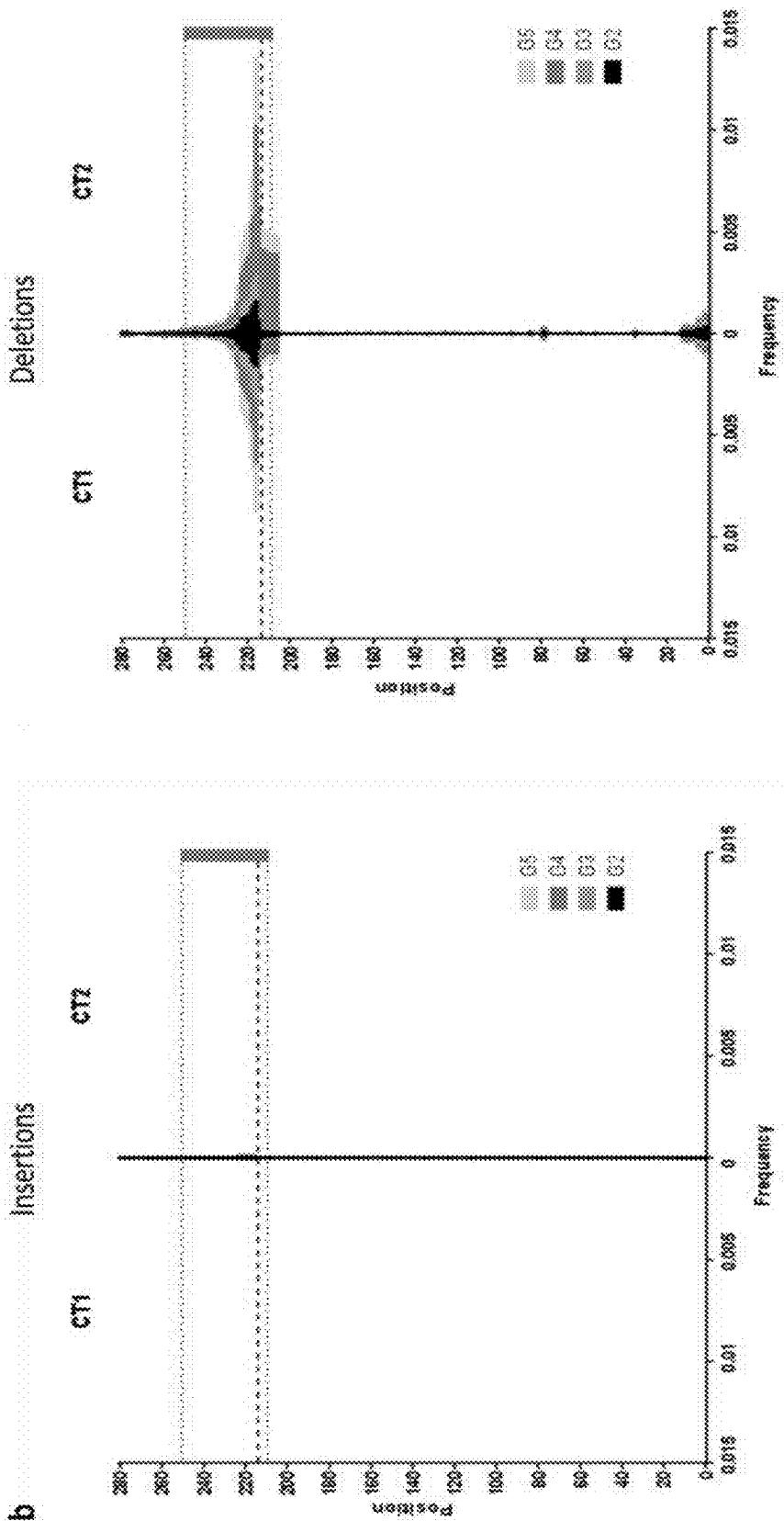


Figure 10C

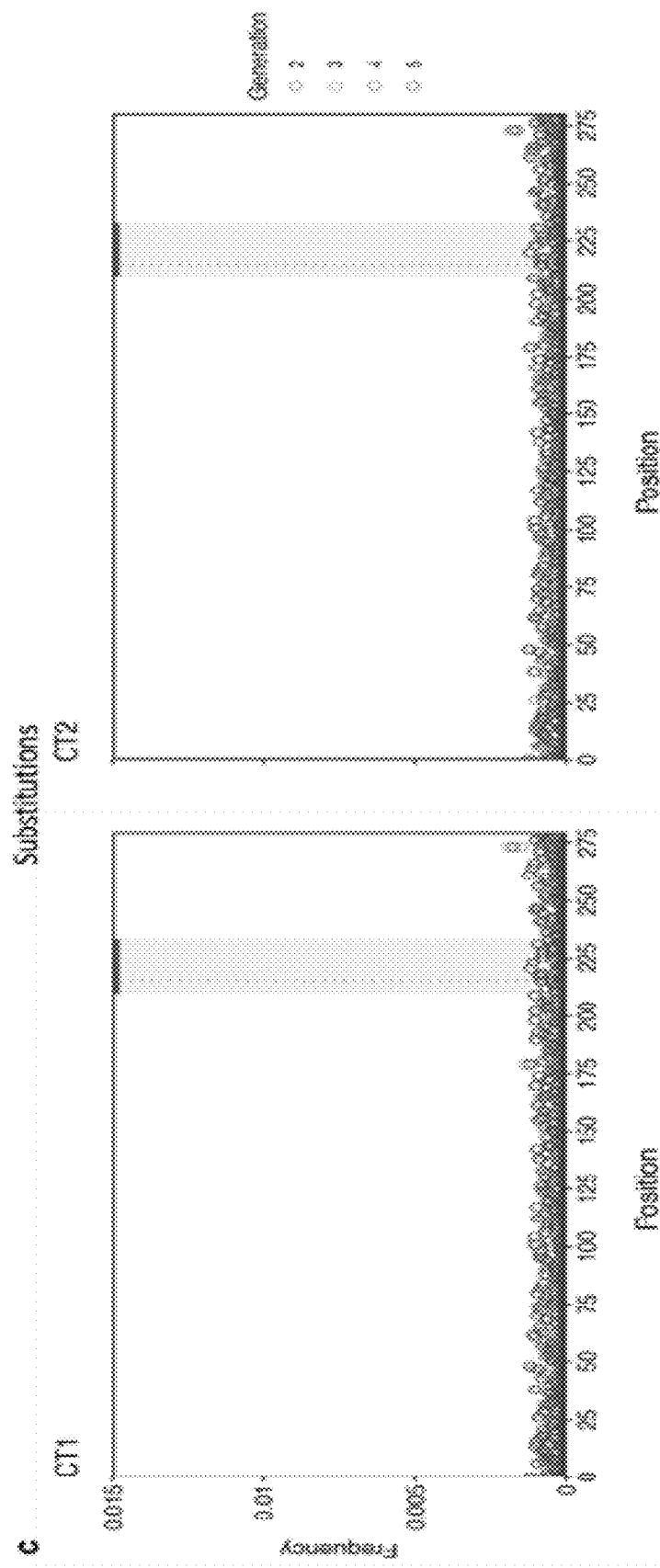
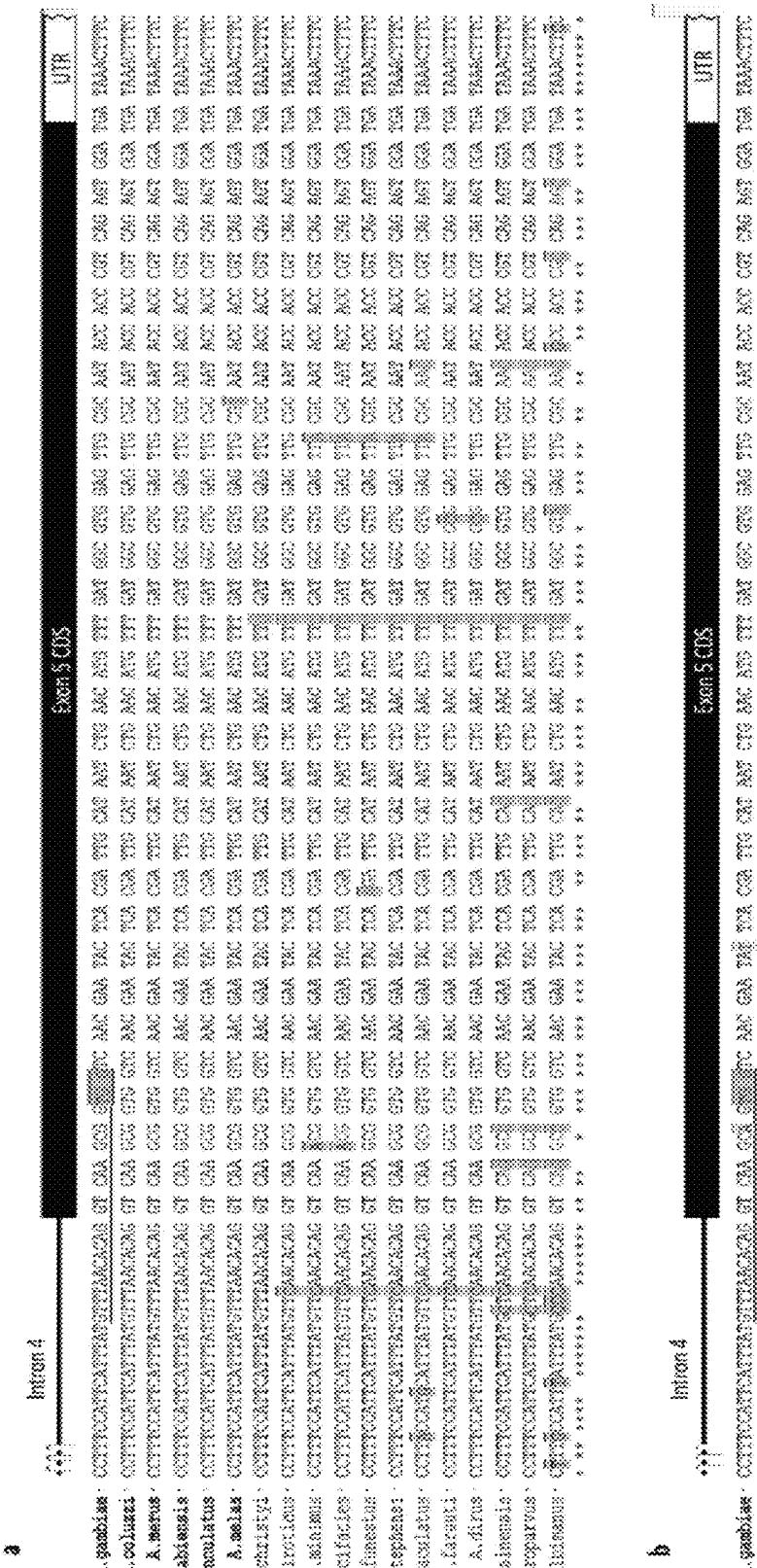


Figure 11



SEQ ID Nos: are 76-94 starting from *A.gambaie* going down the list of sequences in numerical order

Figure 12

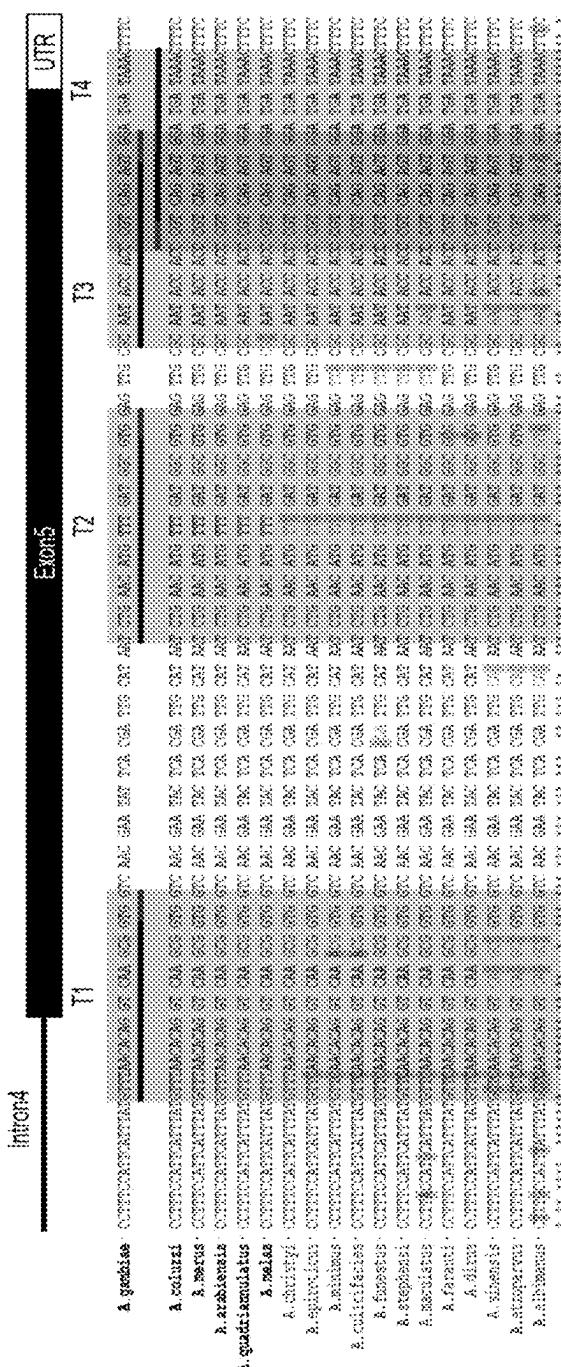


Figure 13

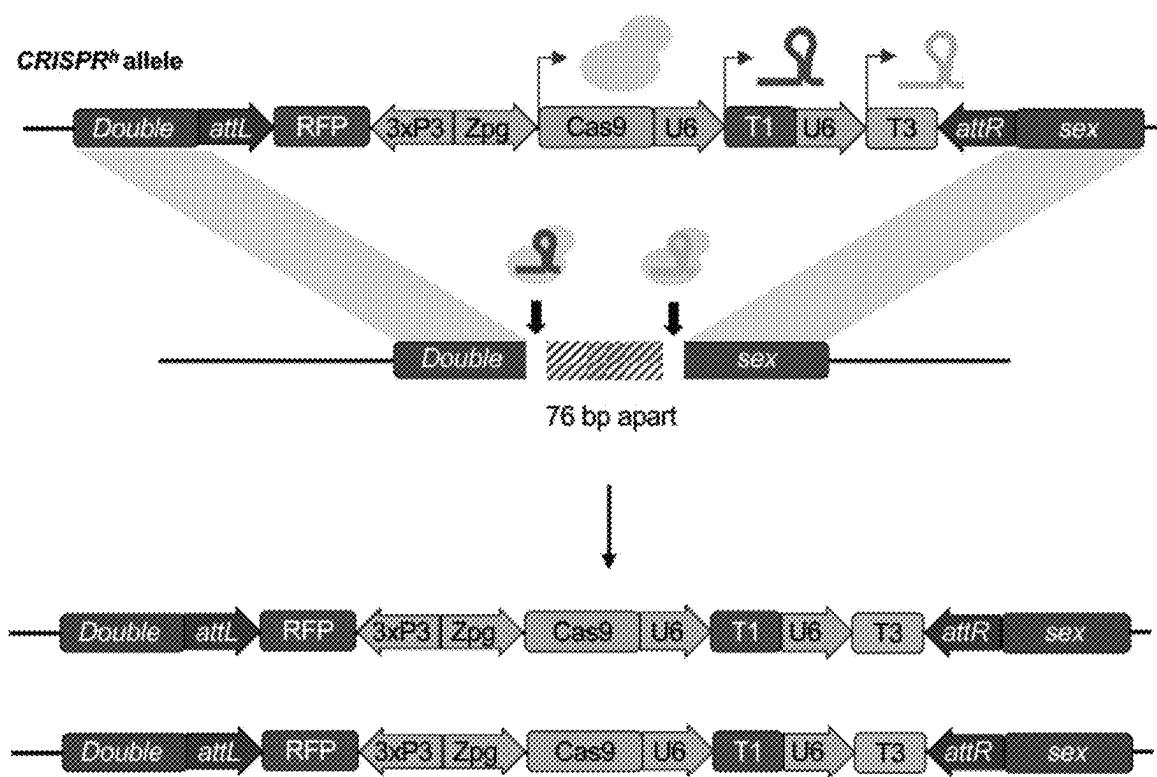


Figure 14

Original gene drive at *dsx*

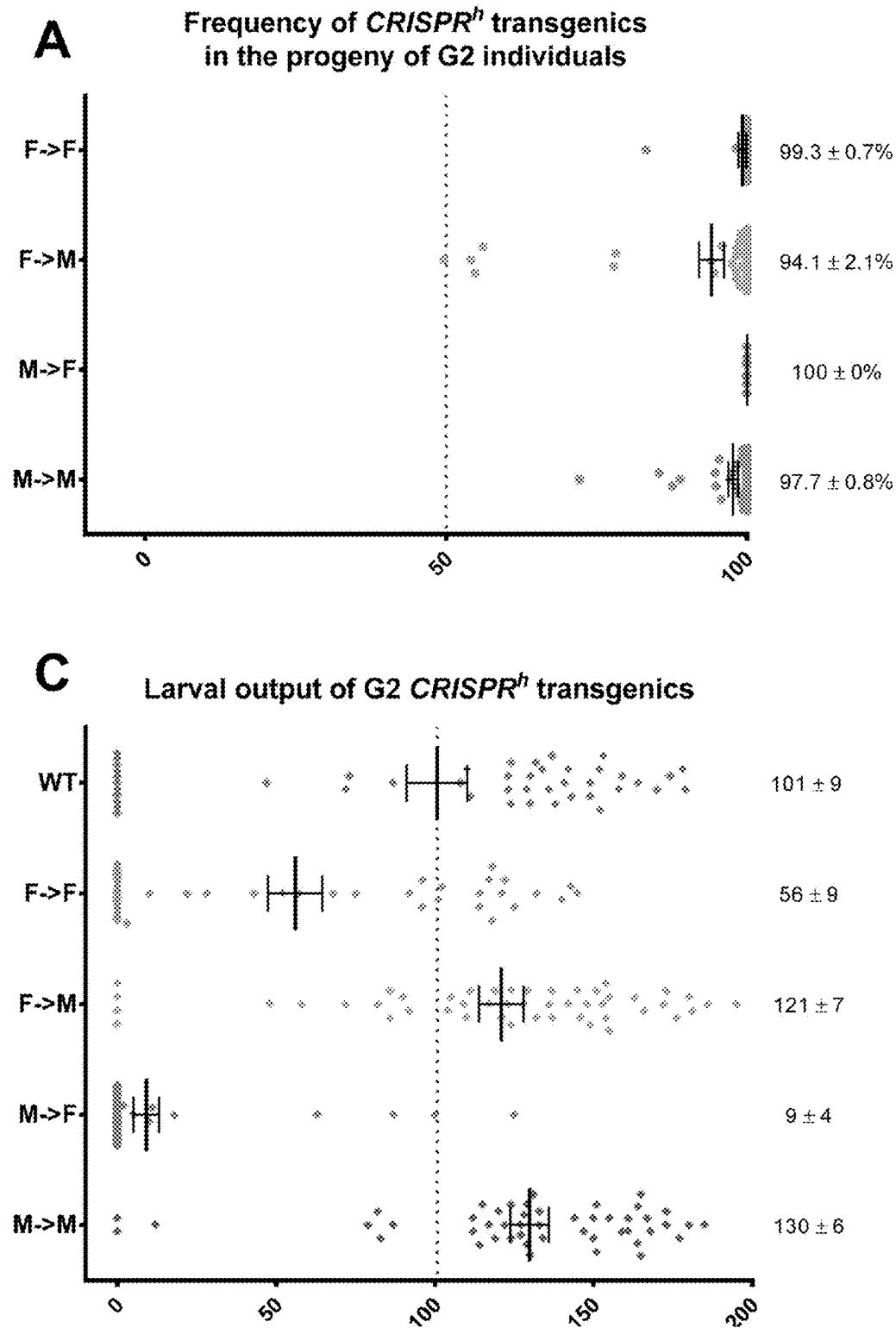
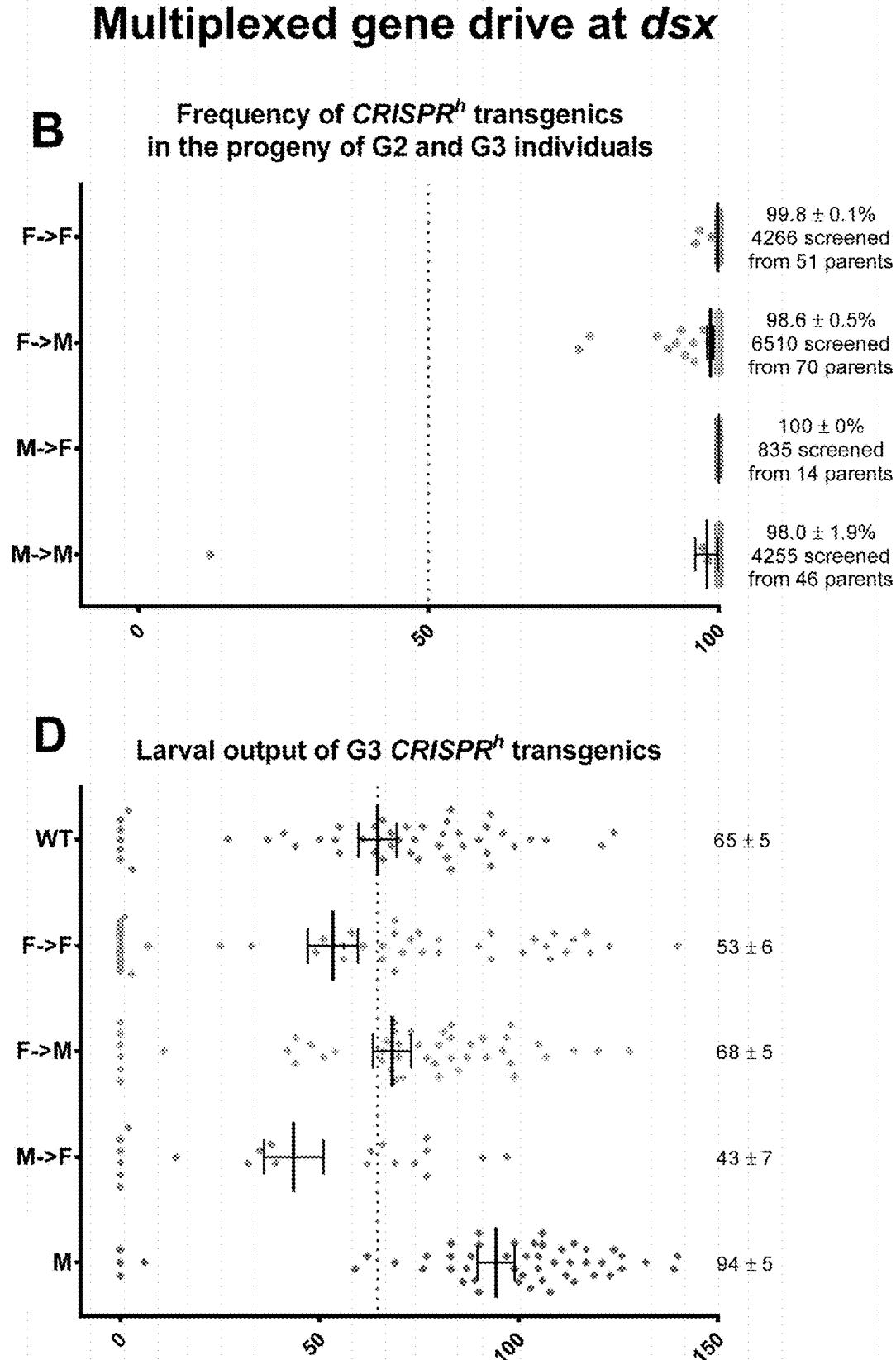


Figure 14 continued



GENE DRIVE TARGETING FEMALE DOUBLESEX SPLICING IN ARTHROPODS**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^{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^{3, 4}, or alternatively by introducing into the Y chromosome a sex distorting in the form of a nuclease designed to shred the X chromosome during meiosis, an approach known as Y-drive⁴⁻⁶. 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^{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 drive². 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 predicted³ and is regarded as the main technical obstacle for the development of an effective gene drive for vector control⁸⁻¹¹.

[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)^{12, 13}. With the exception of Yob1 acting as Y-linked male determining factor¹⁴, 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 species¹⁵. 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 mouth-parts—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:

[SEQ ID NO: 2]
CCTTCCATTCAATTGTTAACACAGGTCAAGCGGTGGCAACGAAT
ACTCACGATTGCATAATCTGAACATGTTGATGGCGTGGAGTTGCGCAA
TACCACCCGTCAAGAGTGGATGATAAACTTC

[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:

[SEQ ID NO: 3]
CCTTCCATTCAATTGTTAACACAGGTCAAGCGGTGGTAAACGAAT

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:

[SEQ ID NO: 4]
GTTAACACAGGTCAAGCGGTGG

[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:

[SEQ ID NO: 5]
GTTAACACAGGTCAAGCGG

[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:

[SEQ ID NO: 6]
GTTAACACAGGTCAAGCGGGTTTAGAGCTAGAAATAGCAAGTTAAA

TAAGGCTAGTCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT

[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 if 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:

[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:

[SEQ ID NO: 48]
GUUUAACACAGGUCAAGCGGGUUUUAGAGCUAGAAAAGCUAGUUAAA
AAGGCUAGUCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCU

[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:

[SEQ ID NO: 49]
TTTGTATGCGTGCCTGAAGGGTTGATCGAACCTTACAACAGTTGAG
CTATAACGGCTGCGTGTGGCTTCTAACGTTATCCATCGCTAGAAGTGAAC
GAATGTGCGTAGGTATATATATGAAATGGAGTTGCTCTCTGCT

[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:

[SEQ ID NO: 7]
CAGCGCTGGCGGTGGGACAGCTCCGGCTGTGGCTGTTCTGCGAGTCCT
CTTCCTGCGCACATCCCTCTGTCGACCAGTTCACTGCTGAGCGTAA
GCCTGCTGCTGTTGTCCTGCATCATCGGGACCATTGTATGGGCCATCC
GCCACCACCAACATCACCAACCGCCGTCCTATTCTAGGGGCATACCCATCA
GCATCTCCGGGGGCCATTGGCGGTGGTGCACAGGTGCCATTGTTGT
TGCTGAAAGCAAAGAAAAGCAAATTAGTGTGTTCTGCTGCACACGATA
ATTTTCGTTCTGCGCTAGACACAACAAACACTGCATCTGGAGGGAGA
AATTGACGCTAGCTGATAACTTACCTCAAAGTTATTGTCCATCGTGG
TATAATGGACCTACCGAGCCGGTTACACTACACAAAGCAAGATTATGCG
ACAAAATCACAGCGAAAACAGTAGTAATTTCATCTATCGAAAGCGGCCGAG
CAGAGAGTTGGTGTGATTGCAACTTGACATTCTGCTGCGGATAACCG
CGACGGGCTACCATGGCGCACCTGTCAGATGGCTGTCAAATTGGCCCG
TTTGCATATGGAGTGGTGAATTATCCACTCGCTGATCGTAAAAA
TAGACACCTGAAAACAATAATTGTTGTTAATTTCATTTGAAGAAC
AGCACAAGTTTGCTGACAATTTAATTACGTTCTGTTATCAACGGCAC
GGAAAGATTATCTCGCTGATTATCCCTCTCGCTCTCTGCTATCATGT

- continued

```
CCTGGTCGTTCTCGCGTACCCCGATAATCGAGAGACGCCATTTTAAT
TTGAACTACTACACCGACAAGCATGCCGTGAGCTTTCAAGTTCTCTG
TCCGACCAAAGAAACAGAGAAATACGCCCGAACAGTGCCCGAGTGATCG
ATCCATAGAAAATCGCCCATCATGTGCCACTGAGGCAGACCGCGTAGCT
TGTTCCGAATTCCAAGTGCTCCCCGTAACATCCGATATAACAAACAG
CCCAACAACAAATACAGCATCGAG
```

[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:

[SEQ ID NO: 8]

```
GTGAACCTCCATGGAATTACGTGCTTTTCGGAATGGAGTTGGCTGGTG
AAAAACACCTATCAGCACCGCCTTTCCCCGGCATTCAGGTTATACG
CAGAGACAGAGACTAAATTACCCATTACACGCACTAACCTCGCAA
TAGATTGATATTCAAAACTTCTCACCTTGCGAGTTGGATTCTGGA
TTCTGAGACTGTAAGGACTGCTACGAGCTATCATAGGGTGAAACCGAA
AACAAACAAACGTTAATGGACTGCTCCAAGTGAATCGCTCACGCAA
CAAACACACACCGCCTGGAGCGTCTGGCGCACCTTGACGATGAA
AACTGTAGCAAAACTCGCACGACCAGGCTCCGCTGGTG
TGTTTTCTTCTGAGCAAATTAGAAAACATCATTTGACGAA
AACGTCAACTGCCGAGCAGAGTGACCAGAAATACCGATGTACTGTATA
GTAGAACGTCGGTTATCGGGGGCGGATAACCGTGCACACCCAGTT
TTTGTGAGCTTGAGTGTCTAGTGGTATTCGAAATTCAATTGTTGTT
CATTAACAGTTAACCTATAGTTGATTAATAATTACTACTA
ACGATTAACCGATGGATTCAAAGTGAATAATTATGAAACTAGTGATT
TTAAATTTTATATGAATTGACATTCTGGACATTATCATCTTGGT
CTCGAGCTGCCGAATAATCGACGTTCTACTGTATTCTACCGATTTT
ATATGCCTACCGACACACAGGGGGCCCTAAACCTACCGATTTTAA
TTATCCTACCGAAAATCACAGATTGTTCTATACAGACCAAAAGTC
TGTAACCATTCCCAAACCTAAAGTGTATTAAACTCCATATGGAAATCG
CTAGCAACCAGAACCGAGAAGTCAACAGAGACACCAATTCCGTGTATG
TACTTCATGAGATGAGATTGGACGCGCTGGTAAATTATATGGATT
GACAGATAATGTAAGGCGTGCCTTTTACACGATGGAATCAATTCA
AGAGTCATTGTCAGGATTATAGAAACATCTTATTATGTTTGT
TATCGTTACAGTTACGCCCTGCTTAAGCGGCCGCGTGAAGGCCAAA
AAAAGGGAGTCCCAACGCTCAGTAGCAAATGTGCTCTATCATTG
TGGTTAGAAAAGCCTATGTGACTTCTATGAAACAAATCTAAACTATCT
CCTTAAATAGAGAATGGATGTATTTTCTGCGCACTGAACTTCTGTTG
```

- continued

```
GGAAGAGTTAGATACTCTCCCTCCCCCCCCTCCCTCAACACTCCAA
ACCTACCGAAAACCTACCGATAACAATTGTGATGTACCTACCGAAGACCGCA
AAATAATCTGGCCACACTGGCTAGATCTGATGTTTGAAACATCGCCAA
TTTACTAAATAATGCACCTGCGCTGGTGAAAGCTGCACTAAACAGAT
TAGTTGAAATTACGCTTCTGAAATGTTTATTAAACACTTGTTTTTT
AATACTCAATTAAAGCTACTTCTGGAAATGATAATTCTACCCAAAACC
AAAACCACTTACAAAGAGTGTGTTGGTGATCGGCCGGTACTGCG
ACCTGTGGTCATCGCTCATCTCACGACACATACGACACATCTGTCA
TGAAAAGCTGCACACAATCGTGTGTTGCAAAAAACCGTTCGCGCACAA
ACAGTTCGCACATGTTGCAAGCGTGCAGCAAAGGGCTTTGATGGTGA
TCCGAGTGTGTTGGTCAGCTTTTAATGTGTTTCGCTTAATCGCTTTG
TTGTGTAATGTTGCGAATAATTGTTGCGTGTACAAATGAAA
TGTACAATCCTCGCATGCTAGTGTAAACATTGCTAAATTCCGGTAAGAA
CGTTCAATTACGCTCGGATATCATCTTACGAAGCGTGTATGTGCGCTAG
TACATTGACCTTAAAGTGAATCCTTTGTTAGAAAGCAAG
```

[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:

[SEQ ID NO: 9]

```
GGAAGGTGATTGCGATTCATGTTGATGCCAATATATGATGATTTGTTG
CATATTAATAGTTGTTATGTTTATTCAAATTCAAAGATAATTAC
TTTACATTACAGTTAGTGTGACATATTACTACATAAACACATAGATC
AAACTGGTTACATAATTCAAAGTTGGATTAACCGCAGCAATTG
GTTATGAAAAATATGTGCATAACGTAATATCAAGTAAATTGTCATT
GCATATTATAGACTCCTGTTACAATTGGAAAAATGAAAATGTTAAT
TAATCAAAGAAGAAAAACAAAGAAATTAAATCATTAGGTAGCACACCA
CAAGTACATATTGTTGGCATGAATATTCTCTACACTAACATATTAA
TAGCAATTCTATTGATGCGCTTAGTATAGCGGAATTACCGAACGGCACT
ATAGTTGCTCTGTTGGCACACGCAATCATTTCATCCAGGGTTGCC
ATAGCAGTTGGCGACGGTACGTAGCATGCGAAGGATTGTCGTCAC
GGATCACTTTATTCTAACGTTGAAGAAGGCACATCTCAGTGCAGCGC
TCTGGAAAGCTGCTTACCGAACGAAACTAACTTCAAGTAAACCTCAAA
ACTTGTCTAACGACACCACTGCTACCGCGAGTTCAATTCCCGTGC
AAAGTCCCCGATTAAGCTATCATTGCGAACATTGCAAGTGTGCTAC
CCTCAGGTAAGACCATTGAGGTTACCAAGTTGCAAGAACGTGC
ACAGTAATTTCGTTCTGGTGAACACCTCTCTTGTGAGCTTGACAAA
```

[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:

[SEQ ID No: 10]
ATGTAGAACCGCAGCAAATTCTTCCATGACAGCAGCAGCTACAG
TGGGAAGCCGAACGTCAGACGTGTTGACATGCCAAGCTGGCGGGAAAA
TTACAGCGTGCCTTGTTCAAGCAAATCACAACTCGCTGCAAACAAA
ACCGTTGAGAAATTGATTGTTATAATTGTATTGATTATTGTTA
TAATAAAACTAAAAAGACATACTTTGCATATTATACATAAAAACATA
CATGCAGCATTAAAAACACATATAACCCTCCCTGAGTCCCGTATC
GAAATCTCCATCCTAGTTGACAGTACGACGGACGAGTAGGCCGTGTC
GTGCAAATTCCAGCTTTAGCAGTCTTGCCTGGAGCACTCGCGCGAG
TCCGAGGTTCTGCTGAGGTGCTTAGCGCTAAATTAGCCAATTGCTTTG
CAAGTGAATAACCAGCGAACAGTACTTCAAAACTCAGGTAAGTGA
AGTTTATAGAACAAATGTTGTTGTTAGAAGTTAGTGAAGTGTGTTG
AAAAAAATCTCTATTCCGCAAAACTAACGTAACGTGATTCAAATTGAA
TTATTGTTGTGATGTTATTATTCATCCAGTTGATTAGTATTGTTCT
TAGTTATGTTCAAAACAGTTAAATTAAATTCTATTCAATTACTCATA
AAATAATCTTGGCTTATTAATTTCGAAATTGCTTGTATTGTT
AGTAGCACGCCATTGCCCTTGTTCATTGTAACCTGCTCCCACTA
ACACACTGGCAGTGCAGAACAAAAGCCTCGCACCGGTTGCTGGTATTAG
AGTGTGTGCGTGTGTTGAGCCTGCTGTCAAAATGGCTGTTGCC
CGGTACCGAAATTGCTGTCGACGCTGTCGTAACATTCCGTGGTGT
GTATCGTGTGTCATGTTGCGCCTCCCCCTTTGATAGCAGGCT
GCCGTGGCTGCCGTGGTGTGCGCAGTTGAGTTGGATTAATTTC
TAAGGAAATGGCACGAGAACAGAGCGGGTGGCAGTGTGTTGGCTCTGTC
CCTTCCTTCTGTTGAGTGTCTTACAGCACAGCACGTATCCACC
GCACACAGAGCAGGCAAGGAAGTGGAGAACAAGTGTGCTGCGCATGC
ATGTGTGTGGGGGCATTAGCTGAGATGTCGTTATTGAGAACGGT
ATAGGGGCCAGTCGGTGTGACGTACGGAGCGGTTAGTTAATCCAA
GCGTATCCGTCGTGGAGTGGTTGTTGTCGCTCTGTCATATCAG
TTCCAGAGTGGAGTTAGTAAATCACAGTCCTGGCTTTTCGTTACAA
GATATCCAGAAGGATGGCGTTATTCACAGCTTACCATGGTCTTGT
TTGCTCGAATCAGGGGAGAAAACAGTTCTGTTGATGAAACCGCAGTT
GGCACTGGAGCGGATTCAAAAGTCTCGATATGCAATAGATAAGAGAGTC
GTTGGGGCATAGTTGGAGCAGCTTCCGAGATGTGGAGTTCCGAGAGGA
GAAATGGTGCTTCGTGACGTTCCGGACAGCGGGCCCGCGAAGAGCA
TCTCGTTGTCGTTCATCCGGCAATAATTGATGCGAAAAGCGCGCGGCCA

-continued

CTGGCTTAGCGCAGTGTACACAGTATATTACACACAGAGGCA
CACGCCTTCACACGCGCGCTGCTCAAAGGCTACTTCGGTGGCGGTGTG
TGAGGTCGCTTGCAATGGACAATGAAAATTGCTGGAAAATACCATCGT
CTCTTTAGGTGCAATGGGTCGGGAGAGCGGTGGTCGTCGATATTGGT
GGTGTAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG
GTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG
GTGCAACGGCAATTATTTTGTAATATTCGACCCTTTCTCTC
TCTCCACGTGCTGCTGCTGCTGCTGCTGCTGCATTGCATGTTCCACT
ATTCCCTCGGTTGTCGCTGCCGACCCATTGCTAGTCGAAAGAGAGTC
GCCGTTAGTCGCGCTTCGAGCAACGGACACGTTTGGTGAACCAAC
AGCTTTTCATCTCCGGAGACACACAGATCTGAATCGTACATTCCCA
TAAGGAGAATTGTCATCTCCGGTAATAAGAAAGGAAAC

[0057] Accordingly, preferably the second promoter sequence comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID No: 10, or a variant or fragment thereof.

[0058] Preferably, when transcribed, the first nucleotide sequence, which encodes a nucleotide sequence (i.e. the guide RNA) which hybridises to the intron-exon boundary, targets the nuclease to the intron-exon boundary of the doublesex gene. Preferably, the nuclease then cleaves the doublesex gene at the intron-exon boundary, such that the gene drive construct is integrated into the disrupted intron-exon boundary via homology-directed repair. The skilled person would understand that once the gene drive has been 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.

[0059] In one embodiment, the gene drive construct is inserted into the genome via recombinase-mediated cassette exchange, a technique which would be known to those skilled in the art. 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 intron-exon boundary of the doublesex (dsx) gene, the second nucleotide sequence encoding the nuclease, the first promoter sequence and the second promoter sequence.

[0060] In one preferred embodiment, the CRISPR-based gene drive is introduced into the arthropod comprising 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 that are homologous to the genomic sequences flanking the intron-exon boundary 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. 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 φC31 integrase, which is introduced into the arthropod.

[0061] Preferably, the homology arms 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.

[0062] In a preferred embodiment, the 5' homology arm is provided herein as SEQ ID No: 11, as follows:

[SEQ ID No: 11]
 CTTGTGTTAGCAGGCAGGGAGATGAGCGCAAACGTGCAAGAAGAAC
 ATCACTGTGAAGACGGCAATGCAAAGATAGTGTGCTCAACTTCGCGA
 AGATTGAAGCTAAATTAAAGCAGGAGATTAGCATGACTGAAGTGACTTT
 AAAGTGTCAAGATGGCTGCACTCGCAAACTAGCTGGATGCAGCGCAATT
 TGCCCCGGTGTGCGCGCATGCAAACGAGCAACCGCAGAGGGCAAAGGA
 GAGGATGGGAAGGAGGGAGGTGAAAGAGCAGGCTTAAGGTTGCCCTC
 GGGCATTGAAGTCGATAACAGCGGTTCTATTCCAGTGCAGTAACGATGAC
 GAAGACGATGTTGCTCTGCTGCTGTTGCTGTTGATGATGA
 TGATGATAATAGTGCACAAATATAAAATAAATCTCCGTAAGCTTGATG
 TGGTGCCTGGCTACTATAAGCCGCTGGAAGCAAGGAAGCTAGTCGGC
 AGGGTCATGCAAAGGGAGACACCTTCGGAGCTCCGGAGCTCCGCCGGC
 ACTCTCGGGGGACGTCCGTTATGCGTTGATTTATTATGGAATATT
 TTATAGTGTCTGTTGAAAAAAATCAACGGTTCGAATTTCCTAC
 ACCTCGAGATCGGGCTGGAGTGGCACACGTGGTACGGAACGGTACAGCG
 TTTGAGCGTCTGGTCTGGACTACGGATCGCAGAATGTTATTGCG
 CGCACTGATGGAAAGTCATTTTACCGAGTGGTCAGGGCGCGTAGTCC
 AGTTCGTTCTGGCTGCTGTTGCTGATGCTACGATCCTCAGGAATGATT
 GAAAACGCTGGAGATGGGGAAAAATCAAACACAAAAACGATCTAAT
 GAAACATCGTGTCTCATTGCGCACGATTGACACCTTCGATAAGAC
 GCACATAATCGAGCTAAAGGAGGGGACAGGGTCTGCTTGGCACGAG
 CGATAAGATTGCAACTCGTGAAGCGTGTGCTGGCTGAAGAAGAA
 ACGCTTCCACAGCAGTAGGTGGAAAGTGGGATTGTTGGAACGTGGCATTG
 AAAAGAACCTATTTCTAAAGCCCAGAGGCCGTTCTGAACTGGAAAAC
 CAGATGCGAGATTTTATTGTCGCCCCGCCAGGAAAACAAATGTT
 ATGCTTCTTGCCTTCCGCCCGTTCAAGACGACGAGCTAGTGAAGC
 GAGCCAATGGCTGGAGAAACTCGGCTACCGCTGGGAGATGATGCC
 CTGATGTACGTACTAAAGAGCGCCGATGGCGATGTACAAAAGCACA
 CCAGCGGATCGAGCGAAGGTAAGCTGGCGATGTTGTCGTTGACATCA
 CTTTCATCACCGTGTAGACATCTACTGTGCCTAGCACCGGGCCAGTGG

-continued

TCACAGGGTAGCAAAACGTGTTCTTTTGCGAGAGACTCTACCTCA
 TGATGCAGCTGTTAAGGAAAGGTTCAAGATGAAGGCAATTTCCTAGGA
 TAAGATGATCTTAAGTTACCTCGCTATTAGTGTAACTGTCGCTCA
 ACTCCCAGAAATGTTAATCGCTAGGGTAGTTATTATACTGTTCT
 CATTGAAATGCTGTTCAATCCAACATGTTAAGTTAGCTAGCTCAGACACG
 AGAAGTTAGGAGTATCTGCATCTGAAGGTAGCGGATATGGTGTATGC
 CACGTTCACTGACTCAAATTGATCACAAAAAAACCAAAACATCAA
 AAACCAAATTGTGAATTCCGTCAGCCAGCAGCAGTGACCTCAAAGCCTT
 ACCTTTCCATTCACTTATGTTAACACAGGTCAAG

[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:

[SEQ ID No: 12]
 CGGTGGTCAACGAATACTCAGATTGATAATCTGAACATGTTGATGGC
 GTGGAGTTGCGCAATACCACCCGTCAGAGTGGATGATAAACTTCCGAC
 CACTGTAACCTGCGTATCTTGATGTTGAGTGGTGTATGTTGG
 TGAAACGAAATTCAATAGTTCTGTGCTATTAAATCAAGCCGCGTGC
 ACTGATGCCGATAAGTCAAACACTAGTGTAAAGGAGTGGAGCAGAGAGC
 CGCACCACGGTACAGAAGGGCAGCAGAATGGTCGGCAGCCTAGCTGCAC
 TGGTGCCTGGCTCGCGTCTGGGGGAGGGCAGGAAATTCTAGTGT
 TAAATCGGAGCAGCAAACAAAACAGTGGTCGCTCCGTTCAAGAACGG
 CCTGTACACACACAGAAAACACTGAGCATGTTGATAGTGTAC
 CTAGAGCAGGTGGTCGCTCTGCAACGCTCTGGACGACGGCTTC
 GCGTATTGCGTAGCGTCCGCGATGTTGAGTGTACTGCCAACAG
 CCCGTTCTCCCATGCAACTCTGCAACCAACAAACAAACAAACAA
 AAAACCAATCGACAAATGAATCACACCCCTTTGATCATCTGTATATT
 CTTGTTCTTGCCTTCTATGTCGCCCCACGCCCCGGCGGGTAGCTA
 ATTGCGTAGAAACCCGAAAACCCGGCACATACAGTGTACATACGGTT
 TGAGGACAACTTGACCTGCACTGGCCCTCTGGGGTGGCACGTGTAGCT
 ACTTGTGAGATCGGGCCGACGGTAAAGCGCAATGGCCACACACA
 GTGTGTCACACTCAAACACTACCCCTCTGAAACTACCCGTCAGGGATGC
 ACCGGCTGGCTCATGCCCTGCAAACAGTCGGGCTCCACTGTAGTAG
 CTCCGGCTGCTCTGAGAGAAGGATGCCCTCGAAGTGTGAAAGCGT
 CATTGGCGTTCAAGTGTGTTGTTAGCGAGAACAGC
 AGCAGTTGCGTAGCTGAAAGCGAAGGAGTAATAGAGTGTACATG
 ATGAAAATGAAAATGAAGCAGGAGTAGAAGGCGGAGGAGAGCAACCTG
 TTCCACTAGTAGCGAATAGTTAGTCTAGTTGTCGTTACCAATCAAC
 CAACCATCGTCAACCAATACCTGAGTCACATCGTATCGTAC

-continued
CACAACTTATTAAAAATGAACCTTGTCCGCGCACCACGTAGGGTATCTA
AGGCACCTTCTACGGGCGCACCATGCCATCGTCACCTTCTCCA
ATCAAAACCAACAGCCTGTACCGATGGTGTGCAATTGCGTGCCTGTG
GTTATTAGCAAAAAAGAGAAAAGAGTCGACGAGAGAGATAAGATCGAGA
TCGAGAGTACAAAGAGCAGTAGAAATGTCGTTGTTGTTTCGTAAC
ACAGTTGTTAGCCAAATGGAAATTCCAATAATCCGGGGCGGGAA
ATGCGGAATACTCGTACACACATACTCAATCATAAAGAAAATCCTT
GCGTACATCACTACCGTTGCGCGGTGCTGATCTAGAGCAGACCACTTT
CCACTCCACTCTACAATCAATCTGTCGAGAAGGTATGTAAGACGG
CCTTGAGCGAGTCACGGTGCACCATAACGCCGTCGACGAGGGCTGA
ATGCGAACCTTGCTAATCGATTTCGCTTCTTTATCCACCTCCTT
TTCTCTCCCTCTCTCTTGCAGTGCCTTGTAAACCCCCAAAAGGT
AAACGACACATTAAGACCTACGAAGCGTTGTAAGTCATCGTCGATCC
GAACAGCGACCGGCTGACGGAGGACGACGACGAGGACGAGAACATCTCGG
TGACCCGCAAC

[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:

[SEQ ID No: 13]
TGCAGGGTGCAGGGCGTGCCTTGGCTCCCGGGCGCTACTCCACCTCACCCATGCGATCGCTCCGAAAGATACATTGATGAGTT
TGGACAAACACAACAGAATGCACTGAGTAAAAAAATGCTTATTGTGAAATTGTGATGCTATTGCTTATTGTAACCATTATAAGC
TGCAATAAACAACTGAGTAAACAAACAATTGCAATTCTAGTTATGTTCAAGGTCAGGGGGAGGTGTTGGAGGTTTTAAAGCAAGTAAA
ACCTCTACAAATGTGGTATGGCTGATTATGATCTAGAGTCGCGGCCGCTACAGGAACAGGTGGTGGCGGCCCTCGTGCGCTCGTACT
GCTCCACGATGGTGTAGTCCTCGTTGTGGAGGTGATGTCAGCTTGAGTCCAGCTGAGTAGTAGTAGCCGGCAGTCGACGGCTCTT
GGCCATGTAGATGGACTTGAACCTCACCAGGTAGTGGCCGCCGCTTCAGCTTCAGGGCTTGTGGATCTGCCCTCAGCACGCC
TCGCGGGGGTACAGCGCTCGTGGAGGCCTCCAGCCCAGTCTCTCTGCATTACGGGCGTCCGGAGGGAAAGTCACGCC
TGAACCTCACCTGTAGATGAAGCAGCCGCTCGCAGGGAGGAGTCTTGGTACGGTACCCAGCCGCCGTCCTCGAAGTTCATCAC
GCGCTCCCACTTGAAGCCTCGGGAGGACAGCTTGTAGTCGGGATGTCGGGGTGCCTCACGTACACCTTGGAGGCCGTA
TGGAACTGGGGGACAGGATGTCAGGCCAGGGCAAGGGCAGGGGCCCTGGTACCTTCAGCTTCAGGTGTTGTGGCCCTGTAGG
GGCGGCCCTCGCCCTCGCCCTCGATCTCGAACCTCGTGGCGCTCACGGTGCCTCCATGCGCACCTGAAGCGATGAACCTCTGTAG
GACGTTCTGGAGGAGCGCACCATGGTGGCGACCTGTGGTCCGGGCCGCGTACCGTCGACTCTAGCGGTACCCGATTGTTAG
CTTGTTCAGCTCGCTTGTATTGCTTAGCTTAGCGACGTGTTACTTGTGTTGAATTGAATTGTCGCTCCGTAGA
CGAAGCGCCTCTATTATACTCCGGCGGTGAGGGTTCGAATCGATAAGCTGGATCTAATTGAATTAGCTCTAATTGAATTAGTC
TCTAATTGAATTAGATCCCCGGCGAGCTCGAATTAAACCATTGTGGACCGGTAGCGCTGGCGTGGGACAGCTCCGGCTGTGGCTG
TTCTTGAGAGTCATCTTCGCGCACATCCCTCTCGTCGACCGAGTTCAGTTGCTGAGCGTAAGCGCTGCTGTTGCGCTCGCATC

- continued

ATCGGGACCATTGTACGGGCCATCGCCACCACCATCACCCACCGCCGCCATTCTAGGGCATACCCATCAGCATCTCCGCG
GCGCCATTGGCGGTGGTGCCTAGGTGCCATTCTGTTGCTGAAAGCAAAGAAAAGCAAATTAGTGTGTTCTGCTGCACACGATA
GTTTCGTTCTGCGCTAGACACAAACAACACTGCATCTGGAGGGAGAATTGACGCCTAGCTGTATAACTTACCTCAAAGTTAT
TGTCCATCGTGGTATAATGGACCTACCGAGCCGGTTACACTACACAAAGCAAGATTATGCCACAAAATCACAGCGAAAACAGTAAT
TTCATCATCGAAAGCGGCCAGCAGAGAGTTGGTATTGCAACTTGACATTCTCTGTGTTGGGATAAACCGCGACGGGCTACCAT
GGCGCACCTGTAGATGGCTGTCAAATTGGCCGGTTGCGATATGGAGTGGTGAATTATATCCCACCGCTGATCGTGAATA
GACACCTGAAACAATAATTGTGTGTTAATTTCACATTGAAAGAACAGCACAAGTTGCTGACAATTAAATTACGTTGTTA
TCAACGGCACGGAAAGATTATCTGCTGATTATCCCTCGCTCTCTGTCTATCATGTCCTGGTCGTTCTCGCTCACCCGGATA
ATCGAGAGACGCCATTTTAATTGAACTACTACACCGACAAGCATGCGTGAGCTCTTCAAGTCTCTGTCCGACCAAAGAAACA
GAGAATACGCCCGGACAGTGGCCGGAGTGATCCATAGAAATGCCCATATGTGCCACTGAAGCGAACGGCTAGCTGTT
CCGAATTCCAAGTGTCCCCGTAACATCCGATATAACAAGCAGCCAAACAACAAACAGCATCGAGCTCGAGATGGACTATAAG
GACCACGACGGAGACTACAAGGATCATGATATTGATTACAAGAGCATGACGATAAGATGGCCCAAAGAAGAACGGGAAGGTGGTA
TCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATGGCTGGACATGGCACCAACTCTGTGGGCTGGCGTGATCACCACGA
GTACAAGGTGCCAGCAAGAAATTCAAGGTGTGGCAACACCGACCGCACAGCATCAAGAAGAACCTGATCGAGCCCTGTT
GACAGCGCGAACAGCGAGGCCACCCGGCTGAAGAGAACGCCAGAAGAACGATACACCAGCGAACGGATCTGCTATCTGC
AAGAGATCTTCAGCAACGAGATGGCAAGGTGGACAGCAGCTCTTCCACAGACTGGAAGAGTCCTCTGGGAAGAGGATAAGAA
GCACGAGCGGCCACCCGATCTTGGCACATCGTGGACGAGGTGGCTACCACGAGACTGGAAGAGTCCTCTGGGAAGAGGATAAGAA
CTGGTGGACAGCACCGACAAGGCCGACCTGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTCCGGGCACTTCTGATCG
AGGGCGACCTGAACCCGACAACAGCGACGGACAAGCTGTCACTCCAGCTGGACACCTACAACCGACTGAGGAAAGAAA
CATCAACGCCACGGCGTGGACGCCAAGGCCATCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATGCCAGCTG
CCGGCGAGAAGAACGATGGCTGTTGGAAACCTGATTCGGCTGAGCCTGGCCCTGACCCCCAACTTCAGAGAACCTGACCTGG
CCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACCTGGACAACCTGCTGGCCAGATCGCGACCAGTACCGCA
CTGGTTCTGGCGCAAGAACCTGTCCGACCCATCTGAGCAGCTGGACACCTGAGACTGAGCTGAAACAGAGGACCTGCGGAG
AGCGCTCTATGATCAAGAGATACGACGAGCACCAGGACCTGACCCCTGCTGAAAGCTCTGTGCGGCAGCAGCTGCTGAGAAGT
ACAAAGAGATTTCTTGACCGAGCAAGAACGGTACGCCGCTACATTGACGCCGAGCCAGCAGGAAAGAGTCTACAAGTTCAT
CAAGCCCATCTGGAAAAGATGGACGGCACCGAGGAACCTGCTGAGCTGACGCCATTCTGCGGCGAGGAAGATTTACCCATTCTGAAGG
GACAACGGCAGCATCCCCACCGACATCCACCTGGAGAGCTGACGCCATTCTGCGGCGAGGAAGATTTACCCATTCTGAAGG
ACAACCGGAAAAGATCGAGAAGATCTGACCCCTCGCATCCCTACTACGTCGGCCCTCTGCCAGGGAAACAGCAGATTGCC
GATGACCAGAAAGAGCGAGGAAACCATCCCCCTGGAAACCTGAGGAAAGTGGTGGACAAGGGCCTCCGCCAGAGCTCATCGAG
CGGATGACCAACTTCGATAAGAACCTGCCAACGAGAAGGTGCTGCCAACGACAGCCTGCTGTACGAGTACTTCACCGTGATAACG
AGCTGACCAAAAGTGAATACGTGACCGAGGGATGAGAAAGCCGCCCTCTGAGCGCGAGCAGAAAAAGGCCATCGTGGACCTGCT
GTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTCAAGAAAATCGAGTGTGACTCCGTGGAAATCTCC
GGCGTGGAGATCGGTTCAACGCCCTGGCACATACACGATCTGCTGAAAATTATCAAGGACAAGGACTTCTGGACAATGAGG
AAAACGAGGACATTCTGGAAGATATCGTGCTGACCCCTGACACTGTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGC
CCACCTGTTGACGACAAAGTGATGAAGCAGCTGAGCGGGAGATAACCCGGCTGGGCAAGGCTGAGCCGAAGCTGATCAACGGC
ATCCGGGACAAGCAGTCCGGCAAGACAATCTGGATTCTGAAAGTCCGACGCCATTCTGCCAACAGAAAACCTCATGAGCTGATCC
ACGACAGCCTGACCTTAAAGAGGACATCCAGAAAGCCAGGTGTCGGCCAGGGGAGATAGCCTGACGAGCAGCTGAAAGTGA
CGGCAGCCCCGCCATTAAAGAAGGGCATCTGAGCAGCTGAGAGGTGGTGGACAGCTGTTGAGGAAAGTGA
AACATCGTGATCGAAATGGCCAGAGAGAACGACCCAGAAGGGACAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGG

- continued

GCATCAAAGAGCTGGCAGCCAGATCCTGAAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCT
GCAGAAATGGGCGGGATATGTACGTGGACCAGGAACCTGGACATCAACCGCTGCCACTACGATGTGGACCATATCGTCCTCAGAGC
TTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAGCACAAGAACCGGGCAAGAGCGACAACGTGCCCTCGAAGAGG
TCGTGAAGAAGATGAAGAACTACTGGCGCAGCTGCTGAACCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGA
GAGAGGCCCTGAGCGACTGGATAAGGCCGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAG
ATCCTGGACTCCCGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGAAAGTGAAGTGTACCCCTGAAGTCCAAGCTGG
TGTCCGATTCGGAAGGATTCAGTTTACAAAGTGCAGAGATCAACAACTACCACGCCACGACGCCACCTGAACGCCGT
CGTGGAAACGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTACGGCAGTACAAGGTGTACGACGTGCGGAAGATG
ATCGCCAAGAGCAGCAGGAATCGCAAGGCTACCGCAAGTACTTCTTACAGCAACATCATGAACCTTTCAAGACCGAGATTA
CCCTGGCAACGGCAGATCCGAAGCGGCCCTGATCGAGACAAACGGCAGAACCGGGAGATCGTGTGGATAAGGCCGGATT
TGCCACCGTGCAGAAAGTGCAGCTGAGCATGCCCAAGTGAATATCGTAAAAAGACCGAGGTGCAGACAGGCCCTCAGCAAAGAGTCT
ATCCTGCCAAGAGAACAGCGATAAGCTGATGCCAGAAAGAAGGACTGGACCTAAGAAGTACGCCGGCTCGACAGCCCCACCG
TGGCCTATTCTGTGCTGGTGGCCAAAGTGGAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGATCACCCT
CATGGAAAGAACAGCTCGAGAAAGAATCCCATCGACTTTCTGGAAGCCAAGGGTACAAAGAAGTGAAGAGGACCTGATCATCAAG
CTGCCTAAGTACTCCCTGTCAGCTGGAAACGGCCGAAGAGAATGCTGCCCTGCGGCCACTATGAGAACGCTGAAGGGCTCCCCGGAGGATAATGAGCAGAAACA
GCTGTTGTGAAACAGCACAAGCAACTACCTGGACGAGATCATCGAGCAGATCAGCAGTTCTCAAGAGAGTGATCCTGCCGACGCT
AATCTGGACAAAGTGTGCTCCGCTACAACAAGCACCGGATAAGCCCATCAGAGAGCAGGCCAGAACGACCCAGCAGGAAAGAGGTGCTGG
TGACCAATCTGGAGCCCTGCCGCTTCAGTACTTGTGACCCACCATCGACGGATCGACCTGCTCAGCTGGAGGCGACAAAGGCCGGCC
ACGAAAAGGCCGGCAGGCAAAAGAAAAGTAAATTAAAGAGGACGGCGAGAAGTAACTATGTCCGATTTGCGCAAACC
AGGCCTTAGACAATTGCGCTAACGACATTGAAATGTGAAAAGCTGAAAGCAGTGGTTCGCCAGCCCAGTTGAGCTGAGTTCTAA
TTCTCAAGTGTGCTTCAGTCTGGGGAGTGTGTTCTCCAAATGCACTCACCTGCGTGCAGTGCAAATCGTGAAGTCTAA
TTTTCATATTGTTATTACCTACCAACTAAAGTGTGTTATATATTGCGTTTACGTACGACAAATAAGTGTATTGAGAAATAT
TGCGATAAGAGAGAACTCATTGCGATGAACTCATTGTTAGCTAAGTGCCTTGATAAGTAAGCGGAACAGCAGGAATATGACA
CTCCTGGAAATACATGTAAGCGCTGTAATTAGATATATACAGCAACCAATGGCCATGGTATTAAAGCAGTGCCTGTTAG
TCGAACATTGCTATAAGCAAAATAAGCATTCAATTAACTAAATTCTCAAAGTGAACCTGACTTCAATGATGATCTTAGGCTATAGT
GAAAGCTGAAAGCTTATTGACAATGCAAGGGAAAGTGCAGCAGCTGCGTGTATGGGACGCCGCATCTATTCTCTAGCTAA
CCCTAATCATTAGTAATTGACGGCAGGATTCTGCTTCTTACTTCTTTACTTGGAGCTTTCATCAATAAACCGAGTACCATGGC
CGTACGCTCAACGGAAAGCATTCAAAAAACCGCGTCTCGTGTGATTTGTTGGGTGAGTGGCGCCATCTATTAGAGAATAGCTG
ACTACATCTCGTGGACGAAGGGTCAGAGAACGTTGATGACTGCTATCCAAGCTAGGCAGGAAGGGAGATCGCTA
GAGCAAAAGAAAAATAAGCAAATCTTTTATAACAAATCGACGTTAGCGAAATATGTTGAATCGATTAAAGGTTAGAAT
TCCCTTGGTTGCTTCAATTGCGAGGCCGCTTGTATGCGTGCCTGAGGGTTGATCGGAACCTACAAACAGTTGAGCTATA
CGGCTGCGTGTGGCTTCAACGTTATCCATCGCTAGAAGTGAACGAAATGTGCGTAGGTATATATGAAATGGAGTTGCTCTGCT
GTTAACACAGGCAAGCGGTTAGAGCTAGAAATAGCAAGTTAAAGGCTAGTCCGTTACACTGAAAAGTGGCACCGA
GTCGGTGTGTTTACGCGTGGTCCATGGGTGAGGTGGAGTACGCGCCGGGAGCCAAGGGCACGCCCTGGCACCCGA

[0072] Accordingly, preferably the gene drive construct comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID NO: 13, or a fragment or variant thereof.

[0073] The gene drive construct may for example be a plasmid, cosmid or phage and/or be a viral vector. Such recombinant vectors are highly useful in the delivery systems of the invention for transforming cells. The nucleic acid sequence may preferably be a DNA sequence. The gene drive construct may further comprise a variety of other functional elements including a suitable regulatory sequence for controlling expression of the genetic gene drive construct upon introduction of the construct in a host cell. The construct may further comprise a regulator or enhancer to control expression of the elements of the constructs required. Tissue specific enhancer elements, for example promoter sequences, may be used to further regulate expression of the construct in germ cells of an arthropod.

[0074] Thus, it will be appreciated that the inventors have developed in the human malaria vector *Anopheles gambiae* a CRISPR-based gene drive that selectively impairs mosquito embryos in producing the female splice transcript of the sex determining gene doublesex. Advantageously, the female's reproductive capacity is suppressed only in female insects homozygous for the disrupted allele, which may show an intersex phenotype characterised by the presence of male internal and external reproductive organs and complete sterility. Heterozygous females may remain fertile and may be capable of producing transformed progeny. In addition, development and fertility may be unaffected in those males heterozygous or homozygous for the disrupted allele. This has the effect of enabling the gene drive to reach a high proportion of the insect population.

[0075] Furthermore, by targeting the highly conserved and constrained doublesex intron-4-exon 5 boundary, the drive does not induce resistance, even when a variety of non-functional nuclease resistant variants are generated in each generation at the target site. Nevertheless, the inventors have carefully considered various innovative approaches that may be used to mitigate any against possible resistance to gene drive, and have successfully demonstrated that one option is to target multiple sites at the same time, because, 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. It will be appreciated that homing can also serve to remove resistant mutations generated if at least one of the multiple targeted sites is still cleavable.

[0076] The inventors have analysed the sequence of Exon 5 of doublesex and found that it surprisingly contains at least four invariant (i.e. highly conserved and constrained) target sites that are amenable to multiplexing (i.e. targeting more than one site simultaneously), which are shown in FIG. 12 as T1, T2, T3 and T4. Accordingly, the inventors generated a novel multiplexed gene drive system targeting not only the original target site at doublesex (i.e. the intron-exon boundary of the female specific splice form of the dsx gene, referred to in FIG. 12 as T1), but also one or more additional target sites selected from T2, T3 and T4, which are present at or towards the 3' end of the exon 5 coding sequence. The inheritance bias of the gene drive, and fertility of gene drive carriers was assessed through phenotype assays, and the inventors found that the novel multiplexed gene drive successfully biased its inheritance to the next generation with

transmission rates comparable to the single-guide gene drive, but with the added advantage that any resistance mutations to gene drive are significantly mitigated.

[0077] Accordingly, in an embodiment, the gene drive genetic construct of the invention may be capable of targeting (i) a first target site which comprises an 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.

[0078] The genomic nucleotide sequence of exon 5 of the doublesex (dsx) gene is provided herein as SEQ ID No: 35, as follows:

[SEQ ID No: 35]

```
GTCAAGCGGTGGTCAACGAATACTCACGATTGCATAATCTGAACATGTTT
GATGGCGTGGAGTTGCGCAATACCACCCGTAGAGTGGATGATAAACTTT
CCGCACCACTGTAACTGTCGTATCTTGATGTGGGTGTGTATGTGT
GTTTGGTAAACGAATTCAATAGTTCTGTCTATTAAATCAAGCCCG
TGCGCAACTGATGCCATAAGTCAAACTAGTGTAAAGGAGTGGAGCGA
GAGAGCCGACACCGGTACAGAAGGGCAGCAGAAATGGGTGGCAGCCTAG
CTGCACTGGTGGGTGGTCCGGCGTCTGGGGGGAGGGCGAGGAAATT
TAGTGTAAATCGGAGCAGCAGAAAAACAAAAGTGGTGTCCCCTGTCAG
AAACGGCCTGTACACACACAGAAAACACTGCAGCATGTTGTACATAG
TAGATCCTAGAGCAGGGTGTGTTGCTCTCGAACGCTCTGGACGACGG
CTTCGCGCTATTGCGTAGCGTCCGCCATCGTGGTATTGTAATGTC
ACAAGCCGCTTCTCCCAGTCACATCTGCAACCCAACAAACAA
CAACAAAAACCAATCGACAAAATGAATCACACCCCTTTGTATCATCTG
TATATTCTTGTCTTGCCTTTCTATGTGGCCCACGCCCGGGGG
TACGTAATTGCGTCGAAAACCCCGAAAACCCGGCACATACAGTGTACAT
ACGGTTTGGGACAACCTTGACCTGCGACGCCCTCTGGGGTGGCCACGTGT
AGCTATACTGTGAGATCGGGCGCCGACGGTGTAAAGCGCAATGGCCGC
CACACAGTGTGTCCACTCCAACACTACCCCTCTGGAACTACCCGTCCAG
GGATGCACCGGCTGGCTCATGCCCTGCAAACAGTCCGGCTCCACTG
TAGTAGCTCCGGCGTTGCTCTGAGAGAAGGATGCCCTCGAAGTGTGAA
AGCGTGCATTGGCGTTCAAGTGTGTGTGTAGGTTAGCGAGA
AACAGCAGCTGCGTGTGCTGAAAGAGCAAGGAGTAATAGAGTCATA
ATGAAAATGAAAATGAAGCAAAGTAGAAGCGGAGGAGAGCAA
CCTGTGTCCACTAGTAGCGAATAGTTAGTCTAGTTGTCACCAATCA
ACCTTCCAACCATCGTCAACCAATACCTGAGTCACATCGTCATCGTTA
TCGTGCCACAACCTTATTAAAATGAACCTGTCCGCCACCGTAGGGT
GATCTAAGGCGACCTTCTACGGGCGGACCCACATGCCATCGTCACCT
TCTCCAATCAAACCAACAGCCTGTACCGATGGTGTGCAATTGCGTGC
GTGTGTATTAGCAAAAAAGAGAAGAGTCGACGAGAGAGATAGA
TCCGAGATCGAGAGTACAAAAGAGCAGTAGAAATGTTGTTGTTTT
```

- continued

```
CGTAACACAGTTGTTAGCCAAATGGAAATTCCAATAATCCGGGGC
GGGAAATGCGGAATACTGCGTACACACATCAATCAAAAGAAAA
ATCCTTGCGCTACATCACTACCCTTGCAGGGTGTGATCTAGAGCAGAC
CACTTTCCACTCCACTCTACAATCAATCAATCTGTGCAGAAGGTATGGTA
AGACGGCCTTG
```

[0079] In one embodiment, therefore, the second target site comprises or consists of a nucleic acid sequence, which is disposed in the sequence substantially as set out in SEQ ID No: 35, 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: 35, 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: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:

[SEQ ID No: 36]
TCTGAACATGTTGATGGCGTGG

[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:

[SEQ ID No: 37]
GCAATACCACCGTCAGAGTGG

[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:

[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 con-

struct 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:

[SEQ ID No: 39]
TCTGAACATGTTGATGGCG

[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:

[SEQ ID No: 40]
GCAATACCACCCGTCAAGtttagagctagaaatagcaagttaaaa

[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:

[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:

[SEQ ID No: 42]
TCTGAACATGTTGATGGCGtttagagctagaaatagcaagttaaaa
taaggctagtccgttatcaactgaaaaagtggcaccgagtcgggtct

[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:

[SEQ ID No: 43]
GCAATACCACCCGTCAAGtttagagctagaaatagcaagttaaaa
aaggctagtccgttatcaactgaaaaagtggcaccgagtcgggtct

[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:

[SEQ ID No: 44]
GTTTATCATCCACTCTGAgttttagagctagaaatagcaagttaaaa
aggctagtccgttatcaactgaaaaagtggcaccgagtcgggtct

[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:

[SEQ ID No: 59]
UCUGAACAUUUUGAUGGCCGUUUUAGAGCUAGAAAUGCAAGUUAAA

[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:

[SEQ ID No: 45]
UCUGAACAUUUUGAUGGCCGUUUUAGAGCUAGAAAUGCAAGUUAAA
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:

[SEQ ID NO: 60]
GCAAUACCACCCGUAGAGA

[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:

[SEQ ID NO: 46]
GCAAUACCACCCGUAGAGGUUUUAGAGCUAGAAAUGCAAGUUAAA
AAGGCUAGUCCGUUAUCAACUUGAAAAGUGGCACCGAGUCGGUGCU

[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:

[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:

[SEQ ID NO: 47]
GUUUAUCAUCCACUCUGAGUUUUAGAGCUAGAAAUGCAAGUUAAA
AGGCUAGUCCGUUAUCAACUUGAAAAGUGGCACCGAGUCGGUGCU

[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 φC31 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:

[SEQ ID No: 50]
 GAGTGGATGATAAACTTCCGCACCACTGTAACTGTCCGTATCT
 TTGTATGTGGGTGTGTATGTGTGTTGGTGAACGAATTCAA
 TAGTTCTGTGTATTTAAATCAAGCCGCGTGCACACTGATGC
 CGATAAGTCAAAACTAGTGTGTTAAGGAGTGGAGCGAGAGGCCG
 CACCACGGTACAGAAGGGCAGCAGAATGGTCGGCAGCCTAGCT

-continued

```

GCACCTGGTGCCTGGCTCCGGCTCTGGGGGAGGGCAGGAA
ATTCTAGTGTAAATCGGAGCAGCAAAAACAAAACAGTGGTCGT
CCCGTTCAAGAAACGGCCTGTACACACACAGAAAACACTGCA
GCATGTTGTACATAGTAGATCTAGAGCAGGGTGGTCGTGCTC
CTCGAACGCTCTGGACGCACGGCTCGCGGTATTGCGTAGCG
TTCGCGGATCGGGGATTCGTAAGTGCCACAAGCCGCTTCT
CCCATGCAATCTCGCAACCAACCAACAACAACAAAAAA
CCAATGACAAAATGAATCACACCCCTTTGTATCATCTGTATA
TTCCTGTTCTTGCCTCTTCTATGTGGCCACGCCCGGCG
GGTACGTAATTGCGTCGAAACACCGAAAACCCGGCACATACA
GTGTACATACGGTTGAGGACAACTTGACCTGCAGCCCTCTG
GGGTTGCCACGTGTAGCTATACTTGTGAGATGGGCGCCGACGG
TGTAAAGCGCGAATGGCGCCACACAGTGTGTCCACTCCAACAC
TACCCCTCTGGAACACTCCCGTCAGGGATGCACGGCTCGGCT
CATGCCCTGCAAAACAGTCCGGCTCCACTGTAGCTCCGG
CGTTGCTCTGAGAGAAGGATGCCCTCGAAGTGTGAAAGCGTG
CATGGGCGTTCAAGTGTGTGTGTTAGGTTAGCGAGA
AACAGCAGCAGTGCCTGTGCTGAAAAGCGAAGGAGTAATAGAG
TGCATAATGAAAATGAAAATGAAAATGAAGCAAAAGTAGAAGGC
GGAGGAGAGCAACCTGTGTTCCACTAGTACGAAATGTTAGTC
TAGTTCTGTCACCAATCAACCTTCAACCACATCGTCAACCAATA
CCTGAGTCACATCGTCATCGTTATCGTGCACAACTTTATTAA
AAATGAACCTTGTCCCGCGACCCACATGCCATCGTCACCTTCTCCAAT
CAAACCAACAGCCTGTACCGATGGTGTGCAATTGCGTGC
GTGTGTTATTAGCAAAAAAGAGAAAGAGTCGACGAGAGAGAGA
TAGATCGAGATCGAGAGTACAAAGAGCAGTAAATGTTGTT
GTTGTTTTCTGTAACACAGTGTGTTAGCCAAATGGGAAATTC
CAATAATCCGGGGGGGGAAATGCGGAATACTGCGTACACA
CATACATCAATCAAAAGAAAAATCCTGCGCTACATCACTACC
GTTGCGCGGTGTGATCTAGAGCAGACCACTTCCACTCCACT
CTACAACTCAATCAATCGTGCAGAAGGTATGTAAGACGCCCT
TGAGCGAGTCACGGTGCACCGTCCGACCGCTCGACGAGGGCT
GAATGCGAACTTGCTAATCGATTTCCGTTCTTTATCCC
ACCTCCTTTCTCCCTCTCTCTCTGACTGCCCTTGTA
ACCCCCAAAAGGTAAACGACACATTAAGACCTACGAAGCGTTG
GTGAAGTCATCGCTCGATCCGAACAGCGACCGGCTGACGGAGGA
CGACGAGCAGGACGAGAACATCTCGGTGACCCGACC

```

[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:

-continued

[SEQ ID No: 51]
tgccccgtccaggcggtgcgccttggctccccggcgctactc
cacctcacccatgcgatcgctccggaaagatacattgtatgagtt
tqqaacaaaccacaactaqaatqcagtqaaaaaaaaatqctttttt

-continued

ggcgcacctgtcagatggctgtcaatttggccgggtttgcata
 atggagtgggtgaaatttatatatcccactcgctgatcgtaaaaata
 gacacactgaaaacaataattgttgttataatttacatttga
 gaacagcacaagtttgcataatattaattacgttgcata
 tcaacggcacggaaagattatctcgctgattatccctctcgctc
 tctctgtctatcatgtcctggctcgatcaccggata
 atcgagagacgccattttaatttgaactactacaccgacaacg
 atgcccgtgagctttcaagtttctgtccgaccaaagaaca
 gagaataccggccggacagtgccggagtatcgatccatagaa
 aatcgcccatcatgtgccactgaagcgaaccggcgtagtt
 ccaatttcaagtgtttcccgtaacatccgcatataacaacg
 agcccaacaacaatacagcatcgagctcgagatggactataag
 gaccacgacggagactacaaggatcatgatattgattacaaga
 cgatgacgataagatggccccaagaagaagcggaaaggtcgta
 tccacggagtcccgacggcacaagaagatcagcatggcctg
 gacatcgccaccaactctgtggctggccgtatcaccgacga
 gtacaagggtgcccgacgaaattcaagggtctggcaacaccg
 accggcacagcatcaagaagaacctgatcgacccctgtgtc
 gagcggcggcggaaacagccgaggccacccggctgaagagaaccgc
 cagaagaagatacaccagacggaaacccggatctgtatctgc
 aagagatcttcagcaacgagatggccaagggtggacgacgcttc
 ttccacagactggaaagacttcttggtggaaaggataagaa
 gcacgacggcggcaccatcttcggcaacatcgatggcggatgg
 cttaccacgagaagtacccaccatctaccacctgagaaagaaa
 ctgggtggacacggcggcacaaggccgacccggctgtatct
 ggccctggccacatgtcaagttccggggccacttctgtatcg
 agggcgacctgaaccccgacaacacgacgacgtggacaagctgttc
 atccagctggtgccagacctaaccagctgttgcaggaaaaacc
 catcaacgcccacggcggtggacgccaaggccatctgtgtcca
 gactgagcaagacgacggctggaaaatctgtatcgccacgt
 cccggcggcacaaggaaatggcctgttgcggaaacccgtattcc
 gagcgtggccctgtgaccccaacttcaagacgacacttcgaccc
 ccgaggatgccaactcgacgtcgacggacacccgtacgac
 gacctggacaacccgtgtggcccgatcgccgaccgtacgcca
 cctgtttctggccgccaagaacccgtccgacgccaatctgt
 gcgacatctgtgaggtgaaccccgagatcaccaaggccccctg
 agcgccctatgtatcaagagatacgacgacgacccaccggac
 gaccctgtgaaagctctgtgcggcagcagctgtgtgaaag
 acaaagagatttttcgaccagagacagaacggctacggccgc

-continued

tacattgacggcgagccagccaggaagatctacaagttcat
 caagcccatctggaaaagatggacggcaccggagactgctcg
 tgaagctgaacagagaggacactgtcgccaaagcggaccc
 gacaacggcagcatccccaccagatccacccctggagagctgca
 cgccattctgcggccggcaggaagattttaccatctgtgaagg
 acaaccggaaaagatcgagaagatctgacccctccatcccc
 tactacgtggccctctggccagggaaacagcagattccctg
 gatgaccagaaagagcggagggaaaccatccccctggaaacttc
 aggaagggtggacaaggccggctccggccagagcttcatcgag
 cggatgaccaacttcgataagaacctgcccacgagaagggtct
 gccaacgacacgcgtctgtacgacttccctgtataacg
 agctgaccaaaagtgtaaatacgtgacccggaaatgagaaggcc
 gccttcctgagcggcggcagggaaacccatctggacactgct
 gttcaagaccaaccggaaagtgtacccgtgaaagagg
 actacttcaagaaaatcgagtcttcgactccctggaaatctcc
 ggcgtggaaagatcggttcaacgcctccctggcacataccacga
 tctgtgtaaaattatcaaggacaaggacttctggacaatgagg
 aaaacgaggacatttcaagatatcgatccctggaaatctcc
 ccacccgttgcacgacaaagtgtacccgtgacccgtgaaatcc
 gatcacccggctggggcaggctgacccggaaacgtgtatcaaccc
 atccggacaacgacgtccggcaagacaatctggatttctgaa
 gtccgacggcttgcacccatcgacacttcatgcacgtatccac
 acgacacgcctgacccatccatccatccatccatccatcc
 tccggccaggccatgtatccctggcacatccatccatcc
 cggcggcccccgcattaaagggcatctgcacccatccatcc
 tggtggacacgtcgatggggcggcacaaggccgg
 aacatcgatcgaaaatggccagagagaaccacccaggaa
 gggacagaagaacacgcccggcggcggcacaaggccgg
 gcatcaaaagagactggccggcggcacaatccatccatcc
 gaaaacacccacgtcgacccatccatccatccatccatcc
 cgacaaatggccggatatcgatcgatggggcggcacaaggcc
 accggctgtccgactacgtatcgatggggcggcacaaggcc
 tttctgtatccatccatccatccatccatccatccatcc
 cgacaaacccggggcacaaggccggcacaacgtgcccctcc
 tcgtgtatccatccatccatccatccatccatccatcc
 aagctgatttccatccatccatccatccatccatccatcc
 gagaggccggcggcacaactgggataggccggatcatcaaga
 gacagctggggaaacccggcagatcacaaacgacgtggc
 acac

-continued
atcctggactccggatgaacactaagtacgacgagaatgacaa
gctgatccggaaagtgaaagtgtacccctgaagtccaagctgg
tgtccgattccggaaaggattccagtttacaaagtgcgcag
atcaacaactaccaccacgccccacgacgcccacatgaaacgcgt
cgtggaaacgcgcgtatcaaaaagtaccataagctggaaagcg
agttcggtacggcgactacaagggtgtacgacgtgcggaaagatg
atcgccaagagcgacggaaatcgcaaggctaccgcggaaatgc
cttcttctacagcaacatcatgaaactttcaagaccgagatta
ccctggccaaacggcgagatccggaaagcgccctctgatcgagaca
aacggcgaaaccggggagatcgtgtggataagggccgggatt
tgccaccgtgcggaaagtgtgagcatgccccaaagtgtaaatcg
tgaaaaagaccgaggtgcagacaggcggttcagcaaagagtct
atccctggccaaagaggaacacgcgataagctgtatcgccagaaaa
ggactgggaccctaagaagtacggcggttcgacagccccaccg
tggcttattctgtgtgggtggccaaagtggaaaaggcgaaatcg
tccaaagaactgaagagtgtgaaagagctgtgggataccat
catggaaagaagcagcttgcgagaagaaatccatgcatttcgg
aagccaagggtacaaagaagtgaaaaggacctgtatcaag
ctgcctaagtaactccctgttcgagctggaaaacggccggaaag
aatgtggccctgcggcgaactgcagaaggaaacgaaactgg
ccctgcctccaaatatgtgaaactccctgtacctggccggcc
tatgagaagctgaagggtcccccgaggataatgagcagaaaca
gctgtttgtggaaacagcacaagcactacctggacgagatcatg
agcagatcagcgagtttccaaagagactgtatccggccgacgc
aatctggacaaagtgtgtccgcctacaacaagcacccggataa
gccccatcagagagcggccgagaatatcatccacctgttaccc
tgaccaatctggagccctgccccttcaagtacttgcacacc
accatcgacccggaaagggtacaccagcaccaagagggtgtgg
cgccaccctgtatccaccagagcatcacccgcctgtacgagacac
ggatcgacccctgtctcagctggggaggcgacaaaaggccggggcc
acgaaaaaggccggccaggcaaaaaagaaaaagtaattaattaa
gaggacggcgagaagtaatcatatgtccgcatttgcgcacaaacc
aggcgcttagacaatttgcgcgtaaagcacattcgaaatgtgaaa
agctgaaagcagtgggttgcgcagcccgagttcagcgaaacgga
ttccttccaaagtgtttgcatttgcgcggagttccctcccaa
atgcactcacccctgcgtcagtgccaaatctgtgagtttcaat
ttttcatattgttattacctaccaactaaagttgtttata
tattgcgtttacgtacgacaaataagttgtattcagaaatat
ttgcgataagagagaactcatttgcgtatctcattgtatt

-continued
agctaagtgccttgcataagtaagcgacacgcggaaatgcaca
ctccttggaaatacatgtatgcgttgcattttatata
caccgcacccaaatgttccatgttgcattttatgcactgcctgtt
tgcacattgtatataagcaaaataaaagaaggcatttattatcta
aaatttcttcaaaatgtacttcaatgtatgtatctctaggctatgt
gaaagctgaaagcttatttgcataatgcaggaaatgcacgcac
gtgcgtcgatggaccgcgcgcatttctcagcttattatc
ccctaatcattgtatgcggcgcatttctgttcttactt
cccttacttggagtttcatataaaaaccatgaccatggc
cgtacgctcaacggaaaagcatttcaaaaaaccgcgttctcg
tgtgatttgcgggtgatggcgcatttcttgcatttgcatttgc
actacatctcgatggacgaaatgttgcatttgcatttgcatttgc
ttgatcgactgtatccaaatgttgcatttgcatttgcatttgc
gagcaaaagaaaaataagcaatatcttttataacaaa
tcgacgttgcatttgcatttgcatttgcatttgcatttgcatttgc
tcccttgcatttgcatttgcatttgcatttgcatttgcatttgc
gcgttgcatttgcatttgcatttgcatttgcatttgcatttgc
cggtgcgttgcatttgcatttgcatttgcatttgcatttgc
cgaatgtgcatttgcatttgcatttgcatttgcatttgcatttgc
GTTAACACAGGTCAAGCGGtttttagagctagaatagcaat
taaaaataggctatccgttatcaacttgcatttgcatttgcatttgc
gtcggtgcatttgcatttgcatttgcatttgcatttgcatttgc
cggttgcatttgcatttgcatttgcatttgcatttgcatttgc
taacgttgcatttgcatttgcatttgcatttgcatttgcatttgc
atatatgaaatggagggtgtctctgtGCAATACCACCCGTCA
AGgttttagagctagaatagcaatgttgcatttgcatttgc
ttatcaacttgcatttgcatttgcatttgcatttgcatttgc
gcgttgcatttgcatttgcatttgcatttgcatttgcatttgc
aaggcgcacccctggcaccgcac

[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 $(N/T)*100$, 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= $(N/T)*100$.

[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 3x sodium chloride/sodium citrate (SSC) at approximately 45° C. followed by at least one wash in 0.2 \times SSC/0.1% SDS at approximately 20-65° 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.

[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^{-/-}) heterozygous (dsxF^{+/-}) and wt individuals.

[0169] FIG. 2 shows morphological analysis of homozygous dsxF^{-/-} mutants. (a) Morphological appearance of genetic males and females heterozygous (dsxF^{+/-}) or homozygous (dsxF^{-/-}) 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^{-/-} 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^{-/-} 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^{-/-} and dsxF^{+/-} 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^{-/-} 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 dsxFCRISPRh driving allele and fecundity analysis of heterozygous male and female mosquitoes. Male and female mosquitoes heterozygous for the dsxFCRISPRh allele (a)

(dsxFCRISPRh⁺) were analysed in crosses with wild type mosquitoes to assess the inheritance bias of the dsxFCRISPRh 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 dsxFCRISPRh⁺ 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 dsxFCRISPRh⁺ showed a high transmission rate of up to 100% of the dsxFCRISPRh allele to the progeny. The transmission rate was determined by visual scoring among offspring of the RFP marker that is linked to the dsxFCRISPRh 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 dsxFCRISPRh⁺ 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^{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^{CRISPRh}⁺ males, seeding each cage with a dsxF^{CRISPRh} allele frequency of 12.5%. The frequency of the dsxF^{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⁻. PCRs were performed to verify the location of the dsx qC31 knock-in integration. Primers (blue arrows) were designed to bind internal of the qC31 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 qC31 knock-in line (dsxF⁻) and wild type (wt) as a negative control.

[0174] FIG. 7 shows the morphology of the dsxF^{-/-} internal reproductive organs. (a) Testis-like gonad from 3-days old female dsxF^{-/-} individual. There was no layer division between the cells and there was no evidence of sperm. (b) Dissections performed on dsxF^{-/-} genetic females revealed the presence of organs resembling accessory glands, a typical male internal reproductive organ.

[0175] FIG. 8 shows the development of dsxF^{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 3xP3 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 Φ 31 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 $dsxF^{CRISPRh}$ driving allele affect fecundity and transmission bias in heterozygotes. Male and female $dsxF^{CRISPRh}$ heterozygotes ($dsxF^{CRISPRh}/+$) 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 ($n \geq 15$) were screened for the fraction that inherited DsRed marker gene linked to the $dsxF^{CRISPRh}$ driving allele (e.g. G1 ♂ → 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 (\pm s.e.m.) is shown. (b) Counts of hatched larvae for the individual crosses revealed a fertility cost in female $dsxF^{CRISPRh}$ heterozygotes that was stronger when the allele was inherited paternally. Mean progeny count (\pm 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 Africa¹⁷. 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 protospacer 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 \pm 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 \pm 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 car-

rying 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 the homologous chromosome as 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 (wild-type), 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] $(1-d_f)(1-u_f):d_f(1-d_f)u_f$ in females

[0187] $(1-d_m)(1-u_m):d_m(1-d_m)u_m$ in females

[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_{WX}^{10} , w_{WX}^{01} , or w_{WX}^{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_{10}, w_{01}, w_{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 1

Parameters for stochastic cage model

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

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^*, D^*, R^*. 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 or 11.

$$\begin{aligned}
 e_W &= (F_{WW} + w_{WW}^{10}F_{WW}^{10} + w_{WW}^{01}F_{WW}^{01} + \\
 &\quad w_{WW}^{11}F_{WW}^{11} + (F_{WR} + w_{WR}^{10}F_{WR}^{10} + w_{WR}^{01}F_{WR}^{01} + w_{WR}^{11}F_{WR}^{11})/2)/\bar{w}_f \\
 e_R &= \frac{1}{2}(F_{WR} + w_{WR}^{10}F_{WR}^{10} + w_{WR}^{01}F_{WR}^{01} + w_{WR}^{11}F_{WR}^{11})/\bar{w}_f \\
 e_W^* &= (1 - d_f)(1 - u_f)(w_{WD}^{10}F_{WD}^{10} + w_{WD}^{01}F_{WD}^{01} + w_{WD}^{11}F_{WD}^{11})/\bar{w}_f \\
 e_D^* &= d_f(w_{WD}^{10}F_{WD}^{10} + w_{WD}^{01}F_{WD}^{01} + w_{WD}^{11}F_{WD}^{11})/\bar{w}_f \\
 e_R^* &= (1 - d_f)u_f(w_{WD}^{10}F_{WD}^{10} + w_{WD}^{01}F_{WD}^{01} + w_{WD}^{11}F_{WD}^{11})/\bar{w}_f
 \end{aligned}$$

[0192] The proportions s_i of type i alleles in sperm are:

$$\begin{aligned}
 s_W &= (M_{WW} + M_{WW}^{10} + M_{WW}^{01} + M_{WW}^{11} + (M_{WR} + M_{WR}^{10} + M_{WR}^{01} + M_{WR}^{11})/2)/\bar{w}_m \\
 s_R &= (M_{RR} + (M_{WR} + M_{WR}^{10} + M_{WR}^{01} + M_{WR}^{11})/2)/\bar{w}_m \\
 s_W^* &= (1 - d_m)(1 - u_m)(M_{WD}^{10} + M_{WD}^{01} + M_{WD}^{11})/\bar{w}_m \\
 s_D^* &= (M_{DD} + M_{DR}/2 + d_m(M_{WD}^{10} + M_{WD}^{01} + M_{WD}^{11}))/\bar{w}_m \\
 s_R^* &= (M_{DR}/2 + (1 - d_m)u_m(M_{WD}^{01} + M_{WD}^{10} + M_{WD}^{11}))/\bar{w}_m
 \end{aligned}$$

[0193] Above, \bar{w}_f and \bar{w}_m are the average female and male fitness:

$$\begin{aligned}
 \bar{w}_f &= F_{WW} + w_{WW}^{10}F_{WW}^{10} + w_{WW}^{01}F_{WW}^{01} + w_{WW}^{11}F_{WW}^{11} + w_{WD}^{10}F_{WD}^{10} + \\
 &\quad w_{WD}^{01}F_{WD}^{01} + w_{WD}^{11}F_{WD}^{11} + F_{WR} + F_{WR}^{10}w_{WR}^{10} + w_{WR}^{01}F_{WR}^{01} + w_{WR}^{11}F_{WR}^{11} \\
 \bar{w}_m &= M_{WW} + M_{WW}^{10} + M_{WW}^{01} + M_{WW}^{11} + M_{WD}^{10} + M_{WD}^{01} + \\
 &\quad M_{WD}^{11} + M_{WR} + M_{WR}^{10} + M_{WR}^{01} + M_{WR}^{11} + M_{DD} + M_{DR} + M_{RR} = 1
 \end{aligned}$$

[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 $M_{WW}=1/2$, and $M_{WD}^{01}=1/2$ 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 (t+1) are the same in males and females, $F_{ij}(t+1)=M_{ij}(t+1)$. Both are given by $G_{ij}(t+1)$ in the following set of equations in terms of the gamete proportions in the previous generation, assuming random mating:

$$\begin{aligned}
 G_{WW}(t+1) &= e_W s_W \\
 G_{WW}^{10}(t+1) &= e_W^* s_W \\
 G_{WW}^{01}(t+1) &= e_W s_W^* \\
 G_{WW}^{11}(t+1) &= e_W^* s_W^* \\
 G_{WD}^{10}(t+1) &= e_D^* s_W
 \end{aligned}$$

-continued

$$G_{WD}^{01}(t+1) = e_W s_W^*$$

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

$$G_{WR}(t+1) = e_W s_R + e_R s_W$$

$$G_{WR}^{10}(t+1) = e_W^* s_R + e_R^* s_W$$

$$G_{WR}^{01}(t+1) = e_W s_R^* + e_R s_W^*$$

$$G_{WR}^{11}(t+1) = e_W^* s_R^* + e_R^* s_W^*$$

$$G_{DD}(t+1) = e_D^* s_D^*$$

$$G_{DR}(t+1) = (e_R + e_R^*) s_D^* + e_D^* (s_R + s_R^*)$$

$$G_{RR}(e_R + e_R^*)(s_R + s_R^*)$$

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

$$f_{RFP+} =$$

$$F_{WD}^{10} + F_{WD}^{01} + F_{WD}^{11} + F_{DD} + F_{DR} + M_{WD}^{10} + M_{WD}^{01} + M_{WD}^{11} + M_{DD} + M_{DR}$$

[0196] All calculations were carried out using Wolfram Mathematica²³.

PCR

[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. form 30 seconds and elongation was performed at a temperature of 72° C. for 30 seconds per kb.

TABLE 2

Primers used in this study for Example 1	
dsxgRNA-F	TGCTGTTAACACAGGTCAAGCGG-SEQ ID No: 14
dsxgRNA-R	AAACCCGTTGACCTGTGTTAAC-SEQ ID No: 15
dsxΦ31L-F	GCTCGAATTAAACCATTGTGGACCGGTCTTGTGTTAGCAG GCAGGGGA-SEQ ID No: 16
dsxΦ31L-R	TCCACCTCACCATGGGACCCACCCGTGGTGC GGTCACC GAGATGTT-SEQ ID No: 17
dsxΦ31R-F	CACCAAGACAGTTAACGTATCCGTTACCTGACCTGTGTTA AACATAAAT-SEQ ID No: 18
dsxΦ31R-R	GGTGGTAGTCCCACACAGAGAGCTTCGCGGTGGTCAACG AATACTCACC-SEQ ID No: 19
zpgprCRISPR-F	GCTCGAATTAAACCATTGTGGACCGGTCAAGCCTGGCGGTG GGGA-SEQ ID No: 20
zpgprCRISPR-R	TCTGTGGTCCTTATAGTCCATCTCGAGCTCGATGCTGTATT GTTGT-SEQ ID No: 21
zpgteCRISPR-F	AGGCAAAAAGAAAAAGTAATTAATTAAGAGGACGGCGA GAAGTAATCAT-SEQ ID No: 22
zpgteCRISPR-R	TTCAAGCGCACGCATACAAGGCGGCCCGATAATGAA CGAACCAAAGG-SEQ ID No: 23
dsxin3-F	GGCCCTTCAACCCGAAGAAT-SEQ ID No: 24
dsxex6-R	CTTTTGTCAGCGGTACAC-SEQ ID No: 25
GFP-F	GCCCTGAGCAAAGACCCAA-SEQ ID No: 26
dsxex4-F	GCACACCACGGATCGACGAAG-SEQ ID No: 27
dsxex5-R	CCACATACAAAGATA CGGACAG-SEQ ID No: 28
dsxex6-R	GAATTTGGTGTCAAGGTTCAAGG-SEQ ID No: 29
3xP3	TATACTCCGGCGGTGAGGGTT-SEQ ID No: 30
hCas9-F	CCAAGAGACTGATCTGGCGGA-SEQ ID No: 31
dsxex5-R1	CTTATCGGCATCAGTTGCCAC-SEQ ID No: 32
dsxin4-F	GGTGTATGCCACGTTCACTGA-SEQ ID No: 33
RFP-R	CAAGTGGGAGCGCGTGTGAAC-SEQ ID No: 34

TABLE 6

Primers used in this study for Example 2

multidsxΦ31L-F	gctcgaatttaaccattgtggaccggCTTGTGTTAGCAGGCAGGGGA-SEQ ID No: 52
multidsxΦ31L-R	tgaacgattggggatccggCTTGACCTGTGTTAACATAATG-SEQ ID No: 53
multidsxΦ31R-F	agatataatcctgaacgcgtGAGTGGATGATAAACTTCCGCAC-SEQ ID No: 54
multidsxΦ31R-R	tccacactcacccatgggaccacgcgtGGTGCAGGTACCGAGATGTC-SEQ ID No: 55
4050-2U6-T1-F	gagggtctcaTGCTGTTAACACAGGTCAAGCGGgttttagagctagaaatagca agt-SEQ ID No: 56
4050-2U6-T3-R	gagggtctcaaACCTCTGACGGGTGGTATTGCagcagagacaactccattca 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⁻) 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⁻ 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 3

Ratio of larvae recovered by intercrossing heterozygous dsx ΦC31-knock-in mosquitoes			
GFP strong (dsxF ^{-/-})	GFP weak (dsxF ^{-/+})	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^{-/-} 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^{-/-} 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^{-/-} 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^{-/-} 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^{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^{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^{CRISPRh} cassette is copied into wt locus via HDR ('homing'), disrupting exon 5 in the process.

[0207] The ability of the dsxFCRISPRh 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^{CRISPRh/+} hereafter) crossed to wild type mosquitoes. Surprisingly, high dsxF^{CRISPRh} transmission rates of up to 100% were observed in the progeny of both heterozygous dsxF^{CRISPRh/+} male and female mosquitoes (FIG. 4a). The fertility of the dsxF^{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^{CRISPRh/+} males showed a fecundity rate (assessed as larval progeny per fertilised female) that did not differ from wild type males, heterozygous dsxF^{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 dsx to dsxF⁻, 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^{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^{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^{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^{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^{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^{16, 17} and the presence of highly regulated splice site critical for the mosquito reproductive biology.

[0211] Heterozygous and homozygous individuals for the dsxF⁻ 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 4

Characteristic	Genetic females homozygous for the insertion carry male-specific characteristics					
	Genetic Males			Genetic Females		
	dsxF ^{+/+}	dsxF ^{+/-}	dsxF ^{-/-}	dsxF ^{+/+}	dsxF ^{+/-}	dsxF ^{-/-}
Pupal genital lobe	male	male	male	female	female	male
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 antennae	✓	✓	✓	X	X	✓
Pilose antennae	X	X	X	✓	✓	X

[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 5

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

[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 commitment¹⁸. The development of a gene drive solutions capable of collapsing a human malaria vector population is a long sought scientific and technical achievement¹⁹. 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.

SEQUENCE LISTING

-continued

tttattgtat	tgtattctat	tgtactaaca	ccagtttaa	caccattttt	ccattccccc	4140
ctgagatct	tccaaatgtg	cgaaaatttg	tccctgagcg	gtccacttgt	ctcacccgtt	4200
atttctgtca	atgttacccg	aggcacata	acacacacac	acgccccccc	acacacacat	4260
tgtatgtca	acccttgc	gaatgattgt	aaacgcctcg	tatcaccaccc	ggggcgaccc	4320
catcccacat	tgactgc	ttgcaaaaag	aaaagagaaa	agtactca	ctatccgtc	4380
taagtgcac	agtgtgtgt	tacaatacg	gtctgtgtgt	gagtgegagt	aagcgaggt	4440
gggaaagaga	cgccaaattt	gggggtgaaa	atgtgtgat	gtgtgtgtgt	gtgtgcgtt	4500
gtggggagca	cgatcgat	tgcatac	gtctcggtcg	tctccatc	gtacagtgc	4560
cgcacgtgt	tgtgtgtgt	tgtgtgtat	tgtgtgtat	tgtgtgtat	tgtgtgtaaa	4620
acgagccgtg	aatgtcgagg	gttcgtc	gtgcata	ggggggcaca	tttagttgt	4680
gcgaaaatgt	ttgc	tcgatcaaa	gggcagcag	tgcgtacac	ataccatcgc	4740
agcgatcgac	aacagccgc	actgtcacc	ctggccgc	tacgcacgg	acgagcggca	4800
gccgacacgc	ggacagcg	ccccgtgcgg	gtatggggc	gacgcgacgc	gctgcgagtg	4860
atgtgtgt	cgggcgccg	acggagacgc	acggcgaac	gtggcgccg	acggagacgg	4920
acgattgact	tgc	tetgtgc	tgcgtgtcg	cgatgc	ggcgaactgc	4980
atgtgttcc	gcgatcg	tccatcg	tcgcac	cgctacacc	gagacgacc	5040
tagctggcc	oggacgagcg	tcgggaa	atacaacact	cctgtgtgt	ccggccgtcg	5100
cttcgaaagg	cacccaaat	gctgtcg	tctctgtgt	tgaagact	cagaagcgt	5160
cagtgcac	tgcgat	gttcggc	tgcgtgtgt	acgtcg	gtgcgtgg	5220
ccgcggtaa	aaatgttat	ctgtgtat	tcgcac	cctacta	aaattcg	5280
caatgtacca	gtgtc	tcgcgtgtc	ggagaagca	acagccac	cgaacgcac	5340
ggaaaaaaa	cg	cgcgatgt	gaagcga	agaactt	tgtgtgg	5400
aggatttctc	tgc	cgaaactt	acgtt	cggt	tgcgtgt	5460
gagaagagag	cgttgcgt	gacggctt	ctgaga	ggccgc	ccaa	5520
cggccgtgc	taa	atagaga	tagcgt	cgtgc	cggtgg	5580
aaatctgtt	cca	ataaa	tcgt	tcgt	tcgtgtgtt	5640
agoagcgc	cgt	cgtat	cggact	ggcag	tttgc	5700
gaagagtcg	gc	tcgt	acat	cgat	atagtg	5760
agggcccaaa	cg	caact	gc	caac	act	5820
aaaacgcct	g	atcccc	caa	acata	cacgc	5880
tcaccaccc	ctgg	aaaaat	tcgc	gcgc	gacag	5940
cgttgcac	catat	gagg	gtgat	gggg	ccgg	6000
ccgcgtcg	ttgt	gtatgt	tgatgt	atgg	atgg	6060
cgcgcgtgc	atgt	aaaa	gcgt	ggac	tttgc	6120
tgttgcagg	gtt	tttt	actg	ggaaa	attt	6180
gtgtatgtt	gtt	tttt	tcgtt	atgt	tttt	6240
tgcgtgtat	actt	tcgt	cgac	gat	atgt	6300
aattagtgt	atcg	tttt	tcgt	atgt	tttt	6360
aaagatagc	at	tttt	atca	ggat	tttt	6420
taaatatgc	aactt	tttt	atcc	atgg	tttt	6480
agtgatatt	ttcc	atgg	atcc	atgg	tttt	6540
taaattaaa	taa	atcc	atcc	atgg	tttt	6600
gaattattt	at	ttt	atcc	atgg	tttt	6660
taactatcg	at	ttt	atca	atgg	tttt	6720
taatacacat	cg	at	atcc	atgg	tttt	6780
cgtttgtc	at	at	atcc	atgg	tttt	6840
acgtttttt	atcc	at	atcc	atgg	tttt	6900
aagcttat	cac	at	atcc	atgg	tttt	6960
tggtgttat	ctt	at	atcc	atgg	tttt	7020
tgttttgc	tatt	at	atcc	atgg	tttt	7080
atggacaaga	taat	at	atcc	atgg	tttt	7140
gaaaatacc	cgt	at	atcc	atgg	tttt	7200
gaagttat	at	at	atcc	atgg	tttt	7260
acatataaa	at	at	atcc	atgg	tttt	7320
aatgcaaca	acat	at	atcc	atgg	tttt	7380
attatttgc	at	at	atcc	atgg	tttt	7440
aaataaaat	taa	at	atcc	atgg	tttt	7500
tctagtgt	c	taaa	atcc	atgg	tttt	7560
tttaatttca	aa	at	atcc	atgg	tttt	7620
agactgtat	gtt	at	atcc	atgg	tttt	7680
cctgtttag	gtc	at	atcc	atgg	tttt	7740
cgcctat	at	at	atcc	atgg	tttt	7800
taaatgg	ttc	at	atcc	atgg	tttt	7860
tcgaccaaa	aat	at	atcc	atgg	tttt	7920
tttcgcac	aa	at	atcc	atgg	tttt	7980
caaatctgt	ttt	at	atcc	atgg	tttt	8040
ctcgtgtc	tac	at	atcc	atgg	tttt	8100
caacgattt	cg	at	atcc	atgg	tttt	8160
atgttgcac	tcg	at	atcc	atgg	tttt	8220
tgtggtcg	gtt	at	atcc	atgg	tttt	8280
gtctcggt	gg	at	atcc	atgg	tttt	8340
ggtgcac	gg	aa	at	atcc	atgg	8400
aataacgc	gc	aa	at	atcc	atgg	8460
taatcga	gc	aa	at	atcc	atgg	8520
gtgcattcg	gg	cc	at	atcc	atgg	8580
agtttgc	at	at	atcc	atgg	tttt	8640

-continued

aggcttagcg	gagaagcgca	agttgaggag	gggtgtcgca	gggttgcgt	agacgttagag	8700
cttgggttga	agacatacgc	ggaaccacac	ggggctgtgg	tgcacatctga	atggtgtcac	8760
aggaccgtg	gacggaaagca	atgtccgact	ccgggtacga	tgcgcgcacg	gacggcaacg	8820
gtgcggccag	ctcgtaaac	aactcgctga	accggcgac	gcccgcgaac	tgcgcggctt	8880
gccgcaacca	cgggctgaag	atcggtgtga	agggccaaac	gccccgtc	aagtatcgcg	8940
ctcgccatgt	cgagaagtgc	tgccgtacgg	ccgagccggca	gcccgtgtat	gcccgtcaga	9000
cggcgctgcg	gcccggccag	accaggac	agcagccgg	actgtacgg	ggcgaggat	9060
ctcccgagcc	ggtagtetaac	attcacatac	caaagctatc	agagctaaa	gaccgtaa	9120
ataatatatgt	tcataattct	cagccgagat	cgatcgattt	cgactctcc	accggatcga	9180
tggcgctccg	accggggacc	tccacgtgc	caactgacat	acaccgacgg	tcgcggccg	9240
taccgcacca	cggtccggag	ccgcacata	tggggatgt	gtacgtat	gcccgttcat	9300
tttttcgtt	tttttacact	gcttcgtct	gttggaggat	taacacactt	tttcatacat	9360
atttaccatt	gggataaaaa	ctgaggctct	catagagctt	cttcgaatgg	ttcgaatcat	9420
gcaccgaaaa	cacttgcag	actatgattt	gtccacaat	cacgcaaaat	ggatcatctc	9480
caaagtgcgc	gcataatttaa	tgcttagatt	gcccggcc	gatcctccag	ttccacgg	9540
tttggctctgt	gtacatcttt	attgttcgc	ttaggactcc	cctcaat	gagcatctca	9600
gcacggta	cacgaggaac	ggctgcactc	agacaacgg	cgaaaatccg	tgcaatcccg	9660
ggaggggacc	gtttaat	ctgttgggt	tacgttgcct	cgctaaact	accccccgg	9720
atctctgc	catttttcgc	tcacccgtca	cttcgtttaga	ttgttagttt	cgtecgccgg	9780
acagtgcgg	gagttcggt	gagaatcg	ctaggctcc	gagggaggc	taacatg	9840
tttggatgg	cgctacacac	tcttttgc	cgtaacttcc	ccacacgtt	cctcgacat	9900
taccctgtt	gtctgtgt	cctgtgtt	ggcggttac	tctacttcg	tatgtgtt	9960
accatcacc	cttcgttgc	ttgttgggg	ctctgtttt	tttgcgtt	tttgcgtact	10020
cctctcacac	cactgtgt	tgctccac	cttcgttgc	tttttgc	tcgtattacc	10080
ataattatata	tatataatt	tctacttctt	ttcgaacgg	ggcggttgg	cccgccctc	10140
tctctctttt	ccctctttt	ccctctttt	gtctggact	gtgttgcgtt	gttttactt	10200
tttgcacgt	ttggacatgc	ttgttctt	tgcatcatc	ccccatgg	catttttag	10260
caagacgcgt	atcccttcgc	ctgcgtc	aaaccgtt	gtgcggcc	gtccggat	10320
agacgaaatt	gatcagaatt	cagacacac	tcgttatgg	gcccgtat	taccggccat	10380
ctgtcgccg	catttttttgc	gcccggcc	tgtttgcgt	ccctgtact	acaaataat	10440
gcaaacgggt	cacttgcgt	tgccgtat	tcttcgc	gggtcg	aaacggat	10500
gtaaatgtaa	aaacaacatgt	ttttgtat	tgacttgc	tttgcgtt	aaacggat	10560
ttccggctgc	agttggatct	tgttttct	gtacggat	gtccgggt	caaggcagac	10620
gttcttcgc	aggcttcgt	cgtgtat	tttttttgc	tttgcgtt	tttgcgtt	10680
atttgcacaa	tcgcgtgg	cacacat	atacacgc	acagggtacc	cgggggaaaa	10740
tgtcgtgt	ttcccttgc	cgatgtt	gtattcgg	aaagaaaaat	ttttcg	10800
tcgtgtgc	ggtggccgt	gtgtccgtt	tttgcgtt	tatgc	ccaaatgtt	10860
cttgcact	tgtcaatta	ttacttgc	tatcttttgc	tgccgggt	tttgcgtt	10920
tttccccc	cctacttcc	cttccctt	cttcgttct	tgccgc	catgc	10980
gaagctgc	cgtactat	ttactatgc	ttgtatgt	gtatgt	aaatgcgtt	11040
gcccgtat	gtatgtat	agtttgc	caatccac	tttttgc	ccgcgtgc	11100
acggtgggt	tttgcgttgc	gggcgtt	tttgcgttgc	acgaggac	gattatgt	11160
cctactgt	ccgcgggg	gaacggc	gtgtgggt	ggtaggt	tgaatg	11220
cgtaagac	gttttgc	tgttttgc	tatgttgc	aggaaac	aaaaatgt	11280
caagtgcac	atttgcattt	catttttgc	ctgtgtat	gttgc	ttgttgc	11340
tgtttttgc	aatgtat	tattagca	tactgtat	ttgtgt	gagactgt	11400
gaagtattt	atggaaat	atatgtct	aatatgtca	aatatgtcc	cagaatgtt	11460
aaaaggctcg	gagaaatgt	ttatatttgc	tgttgc	ttgtat	aaaatcttac	11520
aaacactaaat	atacatgt	aatttat	tttgcgtat	tttgc	tttgc	11580
gagtcttaac	atcgcttgc	aattgcgtt	ggtcccgt	accaacatgc	caacacgg	11640
ccggcgtac	ccaaaccgc	ctatgttgc	tcatcttgc	ttgttttgc	tttgc	11700
ccctcttcaaa	acacatccc	ggtgcggat	tttgcgttgc	tttgcgttgc	tttgcgttgc	11760
aatcgcgc	ggccgtc	gggttgc	aaaccgtc	tttcgttgc	tttgcgttgc	11820
tgtgtggct	ggcgtgttgc	cgttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	11880
agtgcgtgc	gggtgtgt	tgcgtatgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	11940
ggctgtgt	ggcttttgc	ggtatttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	12000
catgcgtcc	cttcgtat	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	12060
tccaggcaaa	gcagggtat	ggctgtgt	tttgcgttgc	tttgcgttgc	tttgcgttgc	12120
cagtgcgtcc	atcgatgtc	cctggcttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	12180
gcccgtat	ttcaat	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	12240
tgttcgtgt	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	12300
cgtcgtat	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	12360
gggggtgtat	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	12420
ttataagggt	ttgtgtgggt	gggggtgtat	tttgcgttgc	tttgcgttgc	tttgcgttgc	12480
tcggcgtac	ttgtgtggct	ttatgtgt	tttgcgttgc	tttgcgttgc	tttgcgttgc	12540
ctgcgtat	ccggccgg	gggttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	12600
gtgtgtat	atgtgtgt	atgtgtgt	tttgcgttgc	tttgcgttgc	tttgcgttgc	12660
ccgggttgc	tttttgcata	catttttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	12720
agggtatgc	atccatgtc	tttgcgttgc	aaaatcg	gtatgtat	ccgtacgc	12780
tatgtgtgt	ttgtgtgt	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	12840
ctcgacacag	tcggacatgc	gttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	12900
ctcgccgtca	gttgcgttgc	gttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	12960
gaagcaccga	tagatgtc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	13020
tttggcgtcc	gttagaagct	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	13080
ggcgagtgc	tccgacgt	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	13140
cttatttttgc	tctcgacat	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	13200

-continued

agcaaaactag	ctgtttgaga	aaagcggagg	gttggcttag	cgacaagggt	gctacataaa	13260
aaaagaaaaac	gacaaaagac	gtgttaatc	cgattgttg	gttgttccg	gttggaggaa	13320
ccgcctatgt	ctgccttcca	aacttccgca	ctaaacaaca	acttctcg	catggggact	13380
atcaactgoc	caaggcgcac	atctgaagaa	gccccaaact	cgtcgtcgaa	acaccccaa	13440
tcaaaggctc	aacatggcgg	ttactgttc	ttcttgtaag	gcccgcgtcg	tcatgtt	13500
gtgccgtaca	ttgacacac	aagtaaaaca	gaggcagccgc	tagcaggac	ttttgtatgaa	13560
cacttcgtc	ctgcgtcgat	gagggtttaga	ggcacgcga	catttcgtt	tttccctcc	13620
tgtcgaatgg	ttttcgccc	catcgaaaaa	atggttacag	tgttgcaccc	tgagtgtatg	13680
atattttaaa	agatatttca	catttactgo	tgctccctt	cctgcgtcg	gacgacgcga	13740
ctcgctgt	cattccatta	gcccgcaccc	gcccctacca	atagattgca	aatgcgcctt	13800
tctgcggccg	agtcatgtat	gagacatcta	tgacggatata	catgtggaca	aagggttaaaa	13860
aatgcacaca	aacacacaca	cacacacaca	cacacacact	tgcaactacgg	caaagatcat	13920
cttttacgcg	caccgcacac	cgatcgccgc	agcgcccaa	gtgcatacg	atgggtggagg	13980
cttgcgtt	gaaacgaccc	gcccacacgg	gcccgggtt	tgacgttg	aatttcagtt	14040
aatttagaaaa	ttatataatag	ttcttgcgo	acatgtatcg	tgccgcattt	ttttccctgg	14100
ccaaagtac	ccgggttctg	cattttcgga	gcaaggttct	cgacagggtt	tcactttccc	14160
tgtcacacgt	ttgagtgtgc	ctatgtgtgt	gtgtgtgacc	ccttctcg	tttgtccctt	14220
gggtcggtca	gaaatttcta	aaatctgtc	aatggcgcat	ccttttctc	tctgtgcgg	14280
gaacgcgtt	ccgcgaatcc	atccccctcg	cccagggttgc	tatgcata	gcccgtctt	14340
acaattttaa	acgtatttta	gatctgttgc	attaaggcgc	gcccggatg	cgatctttc	14400
cccgccgcac	cggtgtcaat	taaaagcgta	tttgaataat	ttgattattt	tatgaaaatc	14460
aaagaaaaat	gtcttaccg	gcaacaaagg	cttggcatgt	ggaaaaaccag	cacaccgaca	14520
gaacaggccct	ttggggaaac	ggagaacaca	cacccggaca	ccaaactgtt	tctttccggg	14580
tgccgcgcg	acagcagatt	acatctgttgc	acacgagata	atttcatttc	cgccatgcgt	14640
tttgcgtgt	ttgggttgc	tgctgtgttgc	cgccgaaaga	ggaggggggg	gggctttgg	14700
cagcaaatgg	ttgtttaatg	ggcttttacc	tttgagaact	gaaccgcaaa	accctgcgcg	14760
acagggttgc	gtcttgcagac	agttctatcg	cgaaggtgt	gcccgttccac	ttcttcatca	14820
cgcagctgg	cgccgcacaca	cgccctttat	tttggcaggt	tcaatcgaa	agccagcaca	14880
cacacacaca	cgttcgacag	ctaacgagaa	gcacgggttgc	gaccacccgt	tagatgttg	14940
caatccgcgc	tgtgcacttt	tgcatcgcc	acacaccccg	cgacacttt	gctcgcttt	15000
cgccccgtt	ttcttcgggttgc	atttgcgttgc	teggccgcgc	acttcgtatc	cctctatacg	15060
gtggaaacccg	aaaataatgc	gcgagtgc	ccggccacccg	cattaaatttgc	gcaccacgg	15120
ccggccgcga	gagcggcaac	actgtgcgc	ggccaaatgt	ctatttctgt	ctaattccgc	15180
acagccgcgc	ggtacgctaa	gcccgtattgc	ggcccccgc	ccgtctgtacc	cgccgatgccc	15240
gatcgcggag	caatgtgcgc	acttttgcgt	caacttaggttgc	gcaacttg	ccctgtcgta	15300
ctaacccttt	ccgtgcgcgc	tgcgctctcg	tgcgctactgt	tcttcctc	tctctcacac	15360
aacgcataa	aatgtgcagt	ttgcgggaca	gatgtgtgttgc	tgtgtgtgt	tgtgtgtgt	15420
tttgcgttgc	cggttcgttgc	cgtgtgcacgt	gtgtgcgcgc	gcccatttc	taaagcgatc	15480
gattttctc	cgggggcgct	gtttgttgc	ctttttgttgc	ttcaatttttgc	accaacccaa	15540
caacccccc	acccacccac	catgcaccc	gctgcgttgc	ccacatgtgc	atcgtgtgc	15600
agttgcgt	ctcgatgtca	gcaaaaaaaaat	gcaatgcaga	gagtgcacca	aaaacaatgc	15660
acaccatgcg	acaatgcaaa	gatgtaaaag	tcacacaccc	ccaacgaa	gcaatagat	15720
ggatggcccc	tgctggggac	ggcaacgggg	gaataggggc	agcgtatgttgc	attgtatcat	15780
tcatattcg	cgccggagac	cacccgggg	acccgtggc	cccttgggggg	ggaatatgtag	15840
catcgcgtca	cgtcgactt	aatcaacgcg	tgtgcgttat	ttgtctgc	cactcccg	15900
tgcgtatcg	tcgtgtccgt	tcgggttcgtt	cggttctcg	ttggccgttgc	cggtgtcg	15960
cacacgttgc	ggcgatgttgc	ggacatgttgc	caacacggca	cggtatgttgc	tgaagaatgt	16020
tttctttttt	ttgtgttgc	tttttttgcgtt	aaatatgttgc	tttcttc	tgccgtct	16080
gacgcgggg	actggacgaa	cacaggcgcg	gtccgcgttgc	ttttgcatttgc	ttcgtcg	16140
tgtggcgg	tgtggggcca	attgcacaca	caaaccgcgc	gagggttgc	gtatttttt	16200
acgttttaac	gggtcgatgt	tcttcgttgc	gtgcatttc	tgagggttct	tttgcctt	16260
ggggatgttgc	ttggaggatgt	ggccggaaaca	aaacggaccc	aaaaaaatgc	ccacagcaac	16320
agttcgtaaa	gcaacggacgc	aaaaaaacgc	gtatcgctcg	aaaatgtca	ctggggcga	16380
tggtgcattt	tttcacatttgc	ttttggccgt	acgaataaaa	acatgttgc	agtaccatgc	16440
aaaaattgtt	ctttaaaatgc	ccacccgttgc	cggttgcacgt	gcagagcg	cccgagggttgc	16500
acgtgcgttca	agggttgcata	aaatataat	actacaatata	tacaatata	tttgcattt	16560
ccattgcaca	gtcgaggatgt	ctgacggat	gtactatata	tctaccggca	ggtccgttgc	16620
caggctgc	caccgtcg	cagetttcc	ctcgagcagg	cagtttagtgc	gcaaaatgttgc	16680
tgtgtcgat	acgggttgc	tttgcggggat	aatcaatgttgc	agcacacaca	aacaacacgc	16740
ggatgttgc	ttttggggat	ctgttgcatttgc	ggacatgttgc	ggccgttgc	cgctgggtcg	16800
tccacccgttca	atcgcgttgc	tcgaacggaa	acacaaat	tctatataat	tttgcattt	16860
aacacaggcg	ggacgcacgc	agggacccgttgc	tatgttgc	caatttcaca	attgcacttgc	16920
acattggccg	cagaatttgcgtt	ttgcaccatgc	catccaggaa	cgatgttgc	ttgcccgtt	16980
tgtgcgttgc	gtctggatgttgc	tttacatgttgc	aattagatgttgc	cgatgttgc	actctcg	17040
acatcacaat	gtttcaataat	ttgttgcacat	gggcacatgttgc	cgcaacaa	gttaatgttt	17100
cctgttttca	ttaaaacagac	tagatgttgc	tttacatgttgc	tatgttgc	gtatgttt	17160
ggatgttgcata	atttacgttgc	aatgttgcata	tcacatgttgc	tgactgttgc	acctgttgc	17220
tatgttgcattt	tcgtgtatcc	taacgttgc	tttgcgttgc	gtgtgttgc	actgttgc	17280
agtttcttgc	accatgttgc	ctaccatgttgc	aatgttgc	taatgttgc	taatgttgc	17340
cttttcttgc	tttcgttgc	tttttgcgttgc	aaggcatgttgc	tttttgcgttgc	tttttgcgttgc	17400
ataatcgcttca	acacagcatgttgc	gcatgttgc	ggcagacgcgttgc	tactcgatgttgc	agcgttgc	17460
attatcttgc	aggcgatgttgc	caagtttgcgttgc	tttgcgttgc	atttgcatttgc	tttgcatttgc	17520
acatgacacgc	agaatagaat	acgggttgcgttgc	cttgcgcac	taccgttgc	cgccgcgcgttgc	17580
gaccccttca	atggcgatgttgc	tcatgttgcgttgc	gtctcgatgttgc	ggccgttgc	tttttgcgttgc	17640
atctcgatgttgc	tttttgcgttgc	tttttgcgttgc	tttttgcgttgc	tttttgcgttgc	tttttgcgttgc	17700
gttagacatgttgc	aggaaagatgttgc	ttatgttgcgttgc	ttatgttgcgttgc	ttatgttgcgttgc	ttatgttgcgttgc	17760

-continued

gctaaggctg	tggcagtg	gttttagattc	cttttactt	ttacttacca	ccgcacatgc	17820
tttatgtttg	atgttcaaca	ggcacgcag	acaggctgag	agcggtacag	catacacacg	17880
ccgtcttgc	tgatagacaa	ggcttcgcgg	cctggcattt	ccgtggatgt	acgtgttaagt	17940
agtgcggccaa	aggcaccact	cttcacggga	tagaatttag	tgcgttgatg	tgaacggggg	18000
gcgaggaagc	gtatgtccgg	ttgtcgctgt	agttgcagct	tctgcccggag	cagcactgtc	18060
aaaatgggtt	ttgcgttagg	ttgagaatcg	gaggaggggcc	ttcgcggtag	aagccgtacg	18120
gatcgctc	cgcgagcacg	ggacgcaatg	ttgcacaca	ttttgcggg	ttttttttt	18180
gcactcgcca	gagttacgac	ggcttc	tatggaaaggc	agcagcacat	ctcacgggt	18240
gcgtcgaaaa	tcgagcataa	ttgtatgtct	tctgatctat	ttcatttgc	gttttatgtt	18300
ttattcgact	tgcgttttc	cgccccccgg	ctcagcttcc	aggcaggggcg	ggaggctcat	18360
ttaggtttag	ggccccctt	gaegtgggg	agacagtcgg	cgatggggcg	aatatgggg	18420
gagggttgtt	accgatccct	actccatcg	gtccttc	aggactatgt	tcgcctccg	18480
acactcttga	cacttc	ccttcgtctg	atccttc	ggaaagggt	gctggggcag	18540
aaaacccatgtt	gacgcggggag	cagccagaa	ccggcttcc	ctgtcagcg	tgcaacaac	18600
aaaacacagca	aaggatctcg	gtccacact	gtgcactact	acgagagaga	aagagtgtgt	18660
gtcgcttc	gggtgttct	gtcaatgtt	aaaaagggtt	caatggaaa	agagctqaa	18720
aaacagaggc	attatgggg	gttcaaggca	ggaggattgg	tgggtgttag	gccggggcagg	18780
aaaccggatg	ggaagtgc	cgggatacgg	atgtgtctgt	tacgc	aagcggaatc	18840
gttgcggg	tccgttcaaca	ttgtttagat	ggccgtgtt	agectgcgtt	tgat	18900
actttttgt	tctttttgt	ttcatttgc	tgtgtgttcc	aatatgtt	tgcgtttgg	18960
ccggcagata	gggcttc	cgggta	ctcgagat	cg	tcgtcgctt	19020
gtcacggcat	accctcat	agtggggct	tcccg	ggatgtt	atgtgtgt	19080
atgtgttat	at	tctgggtt	ggc	atcgat	ccgtaaat	19140
aacccaacgg	agagacgac	ttgttgc	atgttgggg	ggc	caagcg	19200
gattataaa	ataaaacacg	acgtacggat	accgtggc	gggatta	ttcgttcaa	19260
tgttacatt	gtcgatcg	acaggatct	taagcggg	gaacgg	gtcaca	19320
aagggttatt	tacccat	aaggat	accat	ttgttgc	ttgttgc	19380
ccgataagct	gaccga	ggc	ggaga	gtat	ttttgtt	19440
cacaagcatc	ttgat	gggata	at	tttgc	tttgc	19500
acagttgata	aatgggac	ctcg	ggaa	tct	ttcgtac	19560
tgcatctgc	atgggat	ctgg	cttgc	att	ttcgttca	19620
gtacatgtt	aaaatgtt	tccatttgc	tgttgc	ttgttgc	tttgc	19680
tgtgttgc	tttttagt	ttaact	tttgc	act	tttgc	19740
tatgatc	tc	tc	tc	tc	tc	19800
tggtcgtc	aggat	ggc	ggc	ggc	ggc	19860
tcgacgggg	tgttgc	tgcttgc	agc	gac	gac	19920
caatcatcc	gtgg	ctc	atcg	ttc	ttc	19980
cgtcgctgac	gatcg	tttac	atcg	ttc	ttc	20040
cccgctc	gagg	ccacaa	at	ccaa	at	20100
tggtagatt	atcg	ggaa	aaa	at	ttca	20160
accggcgc	gact	ggaa	at	ttgtt	ttgtt	20220
ggtgtatgtt	g	g	g	tttgc	tttgc	20280
tttcgtc	gtgt	gggtt	tttgc	tttgc	tttgc	20340
gatata	atat	atcg	tggat	tttgc	tttgc	20400
tatcttac	tta	agg	tttgc	tttgc	tttgc	20460
ctattctaa	tgt	cct	tttgc	tttgc	tttgc	20520
atagagatc	atc	atc	atc	atc	atc	20580
tttcaaaact	tat	atc	atc	atc	atc	20640
cttcagg	tat	atc	atc	atc	atc	20700
ctctacta	cc	ctt	atc	atc	atc	20760
acaggcaaca	g	cc	atc	atc	atc	20820
aaatttac	ttt	ccat	atc	atc	atc	20880
gttta	ttt	ccat	atc	atc	atc	20940
ctccctgt	atc	gg	atc	atc	atc	21000
ccatata	atc	gg	atc	atc	atc	21060
cggaac	ttt	ccat	atc	atc	atc	21120
ggttagt	cg	aa	at	atc	atc	21180
gttaac	at	tttgc	atc	atc	atc	21240
aggggagg	aact	tttgc	atc	atc	atc	21300
attgtat	tttgc	atc	atc	atc	atc	21360
gagcgg	gg	atc	atc	atc	atc	21420
agacgg	gg	atc	atc	atc	atc	21480
gcatgtt	tttgc	atc	atc	atc	atc	21540
tgtat	tttgc	atc	atc	atc	atc	21600
gctcg	aa	at	atc	atc	atc	21660
tggat	aa	at	atc	atc	atc	21720
gtgc	at	atc	atc	atc	atc	21780
tttgc	at	atc	atc	atc	atc	21840
tggcca	at	atc	atc	atc	atc	21900
gttgg	at	atc	atc	atc	atc	21960
tcatcg	at	atc	atc	atc	atc	22020
ttact	at	atc	atc	atc	atc	22080
ctcgg	at	atc	atc	atc	atc	22140
tgtat	at	atc	atc	atc	atc	22200
aaac	at	atc	atc	atc	atc	22260
ttgc	at	atc	atc	atc	atc	22320

-continued

-continued

aatgatctca	gcattgcctc	agtcccccc	tcagtcgc	agctcgacgg	acagaaaagt	26940
ggcogatgtg	aataatgt	ttaagaatt	cattttaaa	tatgcaattt	ttattataaa	27000
ttcacccctcg	tcccttgtg	gggcaaaaaa	cgggccctcg	ggcaacgaga	ctctgcaggc	27060
tggtagcaag	gttccgtca	tctgttaat	tgttctcg	aggcggtgc	aaaaacagg	27120
ccgattttgt	ttcaggacag	aacaggaggg	ataaacat	aaagagagag	aagggttaat	27180
gtagaaaacac	aatatgaagt	tattagtgtt	atgccttc	accgcgtggca	gttagatgcc	27240
ggtggtatgca	tcaaatcat	acttcgcag	gcccataatgc	cagegcacgg	ggtgcattaa	27300
acaggcttg	atctggatc	ctttaactac	acatacagg	tcggccat	cctgaaaggc	27360
ctctacagac	aagggcataa	aatatgtatc	acgcacgaa	gatgttattg	aactcattc	27420
cttttcacaa	ggtcataat	gtccaaatgt	ggcatctaga	aatctgtatc	ccagccctga	27480
ttgatgcagg	ctagcagca	aagaatattgt	ttttccggaa	tcatctcc	gattaacat	27540
cgtgtggcat	gttaattccc	cactgtcaat	gtctgttggaa	taatagcccc	ggtgatatct	27600
cattccgcg	ggccggacag	gcacgatgc	actatggta	aaggctttt	ttcttctcac	27660
gttctcacgc	gatccctgtt	cataaagaag	tgcaactat	agtgggttgt	gcccacatgt	27720
ttgcgttccg	gacgcgcag	taatgttcc	ttttgcagg	acttccag	cgtaggggca	27780
gtacgcgtt	ttagtcctc	acggatgc	ctcgatata	taatgcatca	cctgcccgt	27840
cctgcatgt	tggttgg	tgctgcgacc	gggacggatc	aacgagcggt	aaaattactg	27900
cacagtggcg	gggggttcat	gtctgcggaa	gcaatgtc	aggatttgt	gcaattgtc	27960
gacgattgcg	tgccaggaga	gcaggatgc	aaagtgcgc	ggggacagg	ccgcgcctga	28020
ccaaagtaat	acgggggtt	tatgtttcc	ctggtaatg	tgccgttcca	cagegttaat	28080
acttcattcc	acttgacgg	agctaata	cagaatcagg	ttggctgggt	gcataagagc	28140
gaaaatcaca	aaagccgtac	acaaaacac	acaaacacgg	atggctcgg	aacgggttaa	28200
aaaagaaaga	aaaagacag	aacagctca	ggatcctt	acgtgtacac	gcaaaacacac	28260
tgcagaaag	caacaaaaaa	aatgtctt	atttcccg	gtgcggagg	accgcgtcgg	28320
agtcatcg	cagctcgat	tctgtgtgt	tgtgaacgg	ctcgca	cggaaacaaa	28380
aatgtcaacg	agagctctcc	agcagaaagg	aaaccggaa	attctccatc	gatatacgaa	28440
cagtcact	tccggcgac	gtccctact	acccccc	cactatggc	ccaaacccat	28500
ggggccgg	ggtaataat	acggggcat	acatcagcg	caagttcaag	gacaattgtc	28560
aacgcttcc	tcccaacacg	tccgcaccc	acacgtt	gggtggatgg	ggccgtcgg	28620
gaaaaaaaaa	aaagcaaccc	acgcgcacca	ccccctggaa	gctcgcggaa	aagtgtgt	28680
ggagagagag	ggggaggagg	agaaagagag	atggagagac	ggaagggg	ctcgaaaaag	28740
tgtctcgat	gtggggat	gttttacacc	gttaaccat	ggccaggcaga	tgggccatgt	28800
ggggccgat	cggttcgt	tgtgtgtca	cagctgtt	gtcategt	cggtgtcgac	28860
gtcgtcg	acgttcgt	cggttc	cacc	atacacaggc	cgccaccgaa	28920
ggaaaacat	tgtgtac	gattgtgt	gggtt	tgtgcact	caacacagcc	28980
ctccccctt	ttccctgt	tcgcgtt	tcgcgtt	tcgcgtt	tcgcgtt	29040
tacgagccgt	tggttat	tcgcgtt	tcgcgtt	tcgcgtt	tcgcgtt	29100
ctgtgtgt	gtatgtat	tcgcgtt	tcgcgtt	tcgcgtt	tcgcgtt	29160
tjacgtgtt	cagccgtt	tcgcgtt	tcgcgtt	tcgcgtt	tcgcgtt	29220
agggttagtg	ggatgggg	gaaagggtca	atgtgcac	gggaggctaa	atttggcgag	29280
acaaggaaagg	agtcgagaga	gagacgttgc	cttgggtgt	ggtgcagg	gctgtgt	29340
ggttgagcgt	tctcggt	gggtactgt	gtaaatgt	ggaaagt	ttcttctca	29400
ctttttct	ttcttcgt	ttcttcgt	ttcttcgt	ttcttcgt	ttcttcgt	29460
tgaaaat	tgcgcac	ccgcata	ttgtc	ttgtc	ttgtc	29520
tgcataccgg	cgaaaggagg	ggaaacatc	ggcca	gggaggcc	ttttcttc	29580
agtcgtt	tctgt	tcgtgcaca	cagcga	atgtgtgt	tatgtgtgt	29640
cacacaaacg	caagcgat	gacacggaa	acacgggg	ttgtgtgt	ttgtgtgt	29700
tttggacca	tcgcgtt	aaaatcggt	aaaatcgca	ttgtgtgt	atgtcgaggc	29760
cgaaggatcg	tcgtggcg	tctcggt	ccgaggac	aaaaggccc	gccccgtt	29820
ttgtgttca	gagcaacgg	gaaaggca	atatcgat	acacttagac	gagattgt	29880
tagggcat	cgctgggtt	taacagcgc	accagaaat	aatgtgtgt	gttgcgtt	29940
taatgtgt	tgtttactt	ggtttgcgt	tttgcgtt	tttgcgtt	tttgcgtt	30000
acttttatca	tccttattcg	ttataaa	ttataaaat	gtatgtat	gttgcgtt	30060
tgaacaaac	ggtttataa	catacaga	caatata	tttgcgtt	tttgcgtt	30120
agcaaaactt	caacggagg	acacattt	tttgcgtt	tttgcgtt	tttgcgtt	30180
ctgttagc	tttcgaaat	agtaaagt	tcgtt	tttgcgtt	tttgcgtt	30240
aaaatgtaaa	caaaaaaccc	gaaaggcgt	ttttgttgcgt	tttgcgtt	tttgcgtt	30300
tttcctgt	ttcgatgt	cgatggc	atttggctt	gtttaat	tgccgtt	30360
ttaggacgc	ttttgttctt	ttttgttctt	ttttgttctt	ttttgttctt	ttttgttctt	30420
attcgacag	tcgtgcgc	agacaca	tttgcgtt	tttgcgtt	tttgcgtt	30480
agggggtt	tgtttat	tcgtt	tttgcgtt	tttgcgtt	tttgcgtt	30540
ctgcaactat	ggcacat	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	30600
gaatggc	ttttgttctt	ttttgttctt	ttttgttctt	ttttgttctt	ttttgttctt	30660
acacacac	acacttctt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	30720
aatgggtt	taatgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	30780
cctaacaat	tgtatgtt	tcgtt	tttgcgtt	tttgcgtt	tttgcgtt	30840
ttgtttgt	tcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	30900
acaaaccat	cattcg	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	30960
cgaaggccca	tcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	31020
tgctagc	tcgtc	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	31080
ggtgtccgc	ttttgttctt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	31140
caccgcgtt	ttcccttccc	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	31200
tgccggccaa	tgccgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	31260
tatcatct	tctctcg	ccaccgg	tttgcgtt	tttgcgtt	tttgcgtt	31320
cggcaccg	ctgg	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	31380
accacatctc	tcgacgt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	31440

-continued

agacacacag	gaaaaggaaa	tcagtgcgt	tgggttagct	cttttgcgg	gtttcagca	31500
caatgataat	gggttatgt	tgtgtgtgt	ttagccctgt	tcttgcgtcc	acccttcct	31560
agccgtacg	ccacaatgcc	agtaagctta	acttateccc	cgggtgtgt	ctgtgttgga	31620
tttattaccc	gtggcaagta	agtgcagcc	cattgtcg	gtgcggccgg	tgcgatgtgg	31680
caatgatttc	gcatttttc	atcaagtgtt	gtgagcggcc	ggccgtcttg	gacacgcaga	31740
aaaggcttta	tcttgact	ggccgtgtgt	atgtgtgtgg	ttctgttgtt	aaagatataa	31800
tttggcac	gctttatcgc	gacccgtacg	acattgtttc	agcagegttg	cagcagcagc	31860
cgccccatcg	gaaagaaacgg	cttgcgtgg	ggcaggcgag	gtaaaaaaa	gatataaacg	31920
ccgcccgc	tgtccgttt	aatcagctgt	gtcctctgg	acagtttcc	ggtggttgg	31980
atcgagggttc	atcgactta	agtgcattgg	tgttacgc	gwgcaagaa	caatccgtg	32040
accttgcgt	gwgcaagcat	tcaaaagcga	gaaaagcage	tttctgttca	gttagctgt	32100
gatttcttga	aacgtttct	tcttttgc	gggttcttc	tcttgaaaga	tggtgaaacct	32160
tatTTTCT	tgggtgttatt	agatgtcat	taaccatgaa	gtacatttt	gcctaagata	32220
ttacgtcatt	cgtaaatat	tattagat	tgttagactt	ctgcgttgt	gattttatca	32280
cgcacacacgg	aaatttacaa	atctttcca	cacttgttta	agtgcgttg	tagttaagt	32340
aaagagaaaca	aataaaaccc	agctgtggag	cacaacacgg	caaacaaca	gggcacatct	32400
tagacatcat	tatgggtcgg	tttgcgtgg	ctgtctgc	tcataatgtat	cgggtggagg	32460
ttggagctcc	aaaacgcata	cagtcatac	gwgccgtgc	agacgtgtgt	cccggtgtg	32520
gtgaggtaaa	gcccattccgg	ccgactata	gtcaacgc	caagcagaca	ggacgagggg	32580
acacgcgtc	tggatgc	cgagatgt	ttgttcttgc	ttgggtcggc	gggtatgtt	32640
tggtagatc	aaatcgagca	aatcgatgt	gataattttc	gattattacc	gggttaccgg	32700
gcaaccggag	ggaaaatgata	tttgcgttgc	gagttgtacg	tttttattcg	ccgtgtttt	32760
tttttgcgg	cccttcttgc	taccgtgt	ttgcacccgtc	ctttcaaaac	tggaggaccc	32820
caccaaagt	gtcggtaa	attcacat	agccaggctc	gtttgcattt	ttccgcata	32880
tcaacctgtt	aattgcata	tgttagat	ttgtgtgt	ttgtgtgtgt	ggccaaggca	32940
aaggaaaaag	ggaggaaaata	cgagaaaaag	caggaacacc	aacatccaaa	tgccgttgc	33000
gttgcacgc	atttgcgc	gcaataagcg	aaagccgacag	accacggcca	gcctgtgcac	33060
ggatcgcacg	gatggggac	ggggagggca	ggggagaaag	agacatatt	gtttcagcgc	33120
accacgggct	ctcggtccgt	ttaccagac	ccccggacgt	atcggaatgc	gggtcttgc	33180
cgtggctc	cggggaaaaa	gctgataact	ttatgtgt	tgcgatgt	gaaaatcatg	33240
actgttgc	ttttatgtt	ttttatataa	ttatgttgc	ttatgttgc	ttatgttgc	33300
gggtgcgg	ttggggaaaa	cgataatcg	agtagcgtt	ccctgttac	gcagcacagg	33360
gcaatcgca	atgtggctt	ggagogaata	tgccgttgc	gtttgcacat	tgtgttttgc	33420
tttggtaat	tagtgcgt	tcaaggctgt	gtttttgtt	aaatgtatgt	cgtatTTTCA	33480
gagttgtca	gatgtttttt	gcatccgtt	aaagaaaccc	gcccgcgtat	gtcaatttaat	33540
aatggcagct	ttaaaatgt	gctgcacgtt	agtcataat	tgttattttt	tgtgcgtgt	33600
tgtgttgc	gggttgc	ggttactaga	gttactatgt	gttactatgt	gttactatgt	33660
tgacctgtc	gttagactgc	cccttccag	gcctgtgc	catatgtat	aacacccata	33720
ccactgat	ccaaactgc	gctttttat	aaatggaaag	tcatttcgt	tcatttgattt	33780
aatggatgt	tgacgtgaaa	taattgtt	cattgcgt	ctgcgttgt	caatctgcgc	33840
caccaggac	cgaccgggt	ggaccgtgt	caatggctca	atgtcatcac	agttgcgtg	33900
atatcaaatt	gcacacgg	ttcccttccag	atataatattc	ctataaacaac	acgggtcccc	33960
ggggccctt	tacggggca	cgatgtacg	aaactgtct	tttgggcgt	tccaaaataa	34020
cgcattttt	gacgcata	cgatataatc	aaatgttgt	ttggggcgtca	ttgggtgaaa	34080
ggcgatttgc	gtatctact	gcacccgt	gtttcg	ttttcgattt	ttcccgatgt	34140
tactggc	acaaacgc	cgccatcaaa	gwgccgtct	tgttccact	tatTTTCTAC	34200
gtcaataacgt	gtgcgttac	cgatgttcaaa	aaacacgcgt	ctactacgc	tccaaaggat	34260
ctgtttaagg	aaacggcttcc	gtgcgttca	aaacacgcgt	ctactacgc	tttgcgttgc	34320
gcaagggtc	ttgggttgc	atgggggtcg	aaagggttgc	tttgcgttgc	tttgcgttgc	34380
tttttcaac	atgccttgc	tagaagagag	gctctggaa	ttctcgaa	agtatgtt	34440
tgccatgtt	ttacccat	tttgcgttca	tgttccatgt	tttacccat	tttacccat	34500
gcaccaat	tgctttgg	catacgaa	cgttttttaaa	ctgtgcgt	ttgggataaac	34560
tgttggat	ataaccaggc	tcaaggatca	aaaacaactt	gtatgttgc	tttgcgttgc	34620
attgttagc	aatcttgc	catattgc	atcaactct	tacttgcgt	tttgcgttgc	34680
gaccctgt	ttggcaagaa	cttcttctt	ttctcttgc	caacttgc	tttgcgttgc	34740
tttgcgttac	aaaagaat	ggaattgtat	ttttttgtat	gagtgtat	ttatTTTCC	34800
taatttcaat	ctactgtat	tcatgttgc	tttttgcata	tttgcata	tttgcata	34860
attcttctat	cattaaacgc	gcatggcat	atcggttca	agaaacgc	tttgcgttgc	34920
acgaaaacac	acacaaacac	cactcaaaa	gacggggaa	gtacgtcga	tttgcgttgc	34980
agatacgg	tacgggtac	gacgggttc	caacatcat	atcatat	tttgcgttgc	35040
cgggttcat	atgtatgt	aaatgttgc	ccaaaggcgat	tttgcgttgc	tttgcgttgc	35100
aaatccgg	gtttggcg	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	35160
acgtatgtca	aactcactgc	cagatcact	ggggccatct	cacgggttgc	tttgcgttgc	35220
ttgttccgt	ccggaaacgg	gacacgggg	aaaggaaag	cacaatgt	tttgcgttgc	35280
tgaagctgaa	caggcaat	ggggccat	agacgtcgc	tttgcgttgc	tttgcgttgc	35340
aggaaaaacaa	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	35400
ggaagagca	aaaaaaaacg	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	35460
ggccactcggt	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	35520
cggttccgt	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	35580
cggttccgt	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	35640
acctgtatgt	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	35700
aggcaactgt	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	35760
aaccgcgaat	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	35820
cactgagagg	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	35880
tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	35940
gttcaatgt	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	36000

-continued

tcgttccgcc	ctccacagaa	cggAACGGGT	ccttcccaAG	aggggggata	ggaccggtag	36060
tggcagtgc	tccattatTA	atcgacaaGTC	aatcaacGGT	gggggtcgag	atcgaaacac	36120
acggctatcg	cgtctggatt	gggtcgatc	ggggcgatag	gcccgtctA	gggaccgtcg	36180
gctacatcg	cctattgagc	tgtctggatg	cattgtgtGA	attataata	taatttcctt	36240
tgcgcctcc	caccggtcga	gegtactga	gagcagcgtg	tgtgaacgtat	ccttggtgca	36300
tcgcacagg	atgactattg	tcctcgggcg	agaacaaggg	tgtgtcgcc	ctggatctac	36360
cttggcgtg	aaggaggagg	ttcttatgtt	tgtgtcaatTC	tgtgtcgta	atatttgcca	36420
caatagtcg	caacacgcgc	agcgttagca	gcccgtgacg	ataggcgct	gacgggggtgc	36480
tttgggtgc	gcttttgcg	agtcatgtt	tttgccat	cattctcaat	gtctcaatgg	36540
cttcgtatcg	gccaatcatca	aaagggtt	atggcagcat	cttcacagcg	tcttcgttta	36600
ctgcattcgg	attgaagggt	acctatTTT	taatttata	tgtgtatca	tccaaatgtg	36660
atttttgaag	ctgattctt	tttgcgtt	ttgtgtatcc	gcatggatgt	tttgcggga	36720
tggatgttt	tgatgtttt	aaattatttC	acattttat	ctgtaccc	tcaccgttca	36780
ccgtgacat	tgcataatctt	ttttgtcg	aataatgtat	ccgtatata	aaaaaacatta	36840
tttagaaaaag	aagtgtgtta	agggaaacata	ctaaccata	gcttta	agtcgtgaaa	36900
ataaaaatgt	ctaaaaataa	aaataaaa	ttgcacaaac	aatttgata	gtataggt	36960
tagtctgtcc	ttgcTTTAA	gactaccccA	aggggtgata	ttcgtagcat	aaattatgt	37020
tgaggtttat	tgtatgtactt	aaaatcgctc	acctgcctgt	ggccgtgt	gtggtagtat	37080
cgaccgcage	caacatcgaa	tgtccccagg	gtaacgacac	aattgtcata	aatatagaag	37140
aaccagacac	ttggctggcg	gtccggact	gcaaaataaa	gtccaaatcg	aataacgaa	37200
aatccctcta	attcaaccc	ccgtctgtt	cctcggtgc	ccgtgggtc	atggggtag	37260
agctgtgt	aaacccccc	gagaaaaagt	agggaaaaac	agtgtgt	aaaaaaaaag	37320
aaaaaaacaaat	cccgaggaaa	aaataaaaat	cccgtaaaac	gatgggttcc	gttggtagt	37380
ttcagaaagg	ttcggaaaat	agacaccca	gctgttattt	cctgcgggt	gttttgc当地	37440
tggataggat	taagtgtgt	gcagggtt	cgtatgcac	tgatcgta	tgaatcgatt	37500
taacagtgg	cagcagcagc	agcgttacca	aaacggcaac	accattctc	ctagatacac	37560
cctgtgtct	cgatgtcg	gcccgtgtt	acgctageta	tctctcgctg	taagttgtcg	37620
ggctgttaaa	cgctcggtt	acggcttgc	atgcattaa	taacgaaatg	agggcgagca	37680
gacggctgac	ggggcaggga	ccggcaatag	cggtgtgt	aaatatttgc	acattggtaa	37740
atttgcata	attgttgcg	ataaaaagaa	tgatggaa	atgtggatg	ggccgggtgg	37800
ccgggttggg	ttggatgttt	gataatgcgt	taacgtcgca	ttttagtgc	agagggtatc	37860
cgagcccaag	tcgatcattt	cgatgtcg	ttgtgtcggt	tatgtcg	tttgcgtt	37920
aactgttga	agacgacgcg	cgatgtgt	ttcgctgata	acggcgctt	aatcgctc	37980
aatccatcg	caaaactgcc	cgccgtgtat	tgctttctt	acgggtgca	cggtgggtgg	38040
cattgaattt	ctccggggcc	caggatgtt	gacgcgcact	gcacactat	cttacgcct	38100
ttgatacagc	gggtgttctt	atttgcgt	ctcgccact	cgggggtagc	cttcagttt	38160
ttgccaaccc	gttcaattt	ctccggctc	aaaccccttc	tttgcacata	gacgtgttgc	38220
ttcattatgt	ttccctttca	ccctgggtt	gcccgtgt	cacaacttt	ccgcaagcgc	38280
atcgctgtct	gtggatgt	gtgggttgc	tggtttatct	tgtactcat	gtgtttgtgt	38340
ttgttcttct	ttgttcttct	ttgtgtcg	tgtgtcaata	tgtactgtat	gtcagattt	38400
tagtcgtact	cgatTTTGC	cgcaaaacata	catacgccgt	gtctcaacgc	catgtgttgc	38460
taggtccgag	tgcaattgtt	tttgcgttgc	gctgtttttt	ccctgcattt	tctttgcgc	38520
gaatgttgc	gacttgggtt	ttgtatgt	aaatcgctt	gtatgtgt	ctcggtctg	38580
gaatagat	cttcccccata	caggcgtt	cacgcacaca	cacacacaca	cacagacgcg	38640
tggcacggc	gtgtttatgt	tgcaatgtt	atgtgtgtt	gtgtgtgtt	ccgttgcgtc	38700
aaagcaatgc	aaagcagcag	cagcgtccgg	accccgaa	acattgggt	gtgacttttg	38760
tttgcgtcc	cgccccgtt	catgcaccc	gaaatcttgc	cgcccaacgg	tgacttaggt	38820
tattgtatgg	ttaataatTT	gcactaaaa	aatcgctt	gtttttaa	tggatcatgt	38880
gcaggcgact	gttgcgttctc	tttgcgttca	ttgtgtat	cccgctttt	cccaatcaca	38940
tgctggattt	gttgcgttca	ccctccctc	gcacacacac	gtcgtgtc	tgcatcacta	39000
aagagcatgc	gaaataacga	taatgtgtat	ttgtgtatgt	acgtgtttgc	tgcttcccg	39060
ggtttgcgtt	ggccatcttca	caccgttgc	gacccttgc	ggcgatata	gcgcgtcgtc	39120
aaaaaaatata	aaataatcg	ccaaatcgaa	gaggatgtc	gtgtgtacaa	acaacacaca	39180
cacactcaca	aacatcgcc	cactaaacag	agcgcagaca	gcgcggcga	caagcggcca	39240
aagacgacaa	ctaccatctt	ccaaatccgc	gactgacaa	tctcggttgc	ttgcgttccg	39300
tttcttaat	atgcgggggg	cccaatcttca	cgatgtgttgc	acgcgttgc	cgatcttcg	39360
taactcgttt	tttcccttctt	gtgtgtgtt	ctatgtgtt	tgatgtactt	gtggactgt	39420
tgcgaaaggag	cagtatagca	accacccac	ccaaacacccc	accggggcga	cgggactaa	39480
aagtctgaca	atgtcggtt	gtcggttgc	gatgtgttgc	tgatgtgtat	cgatgtgt	39540
ccaaatcggtt	ccaaatatcg	acgtttaat	gtttccccc	ttgcgttgc	attgttacca	39600
tttgcgtatgt	tacactgtt	aatttttagt	tgatgtatgt	acgtgtgttgc	gttgcgttata	39660
ctatcggtac	gttttgcacat	gcagcggtt	gaaacggccac	aggcaggaga	gcagccaaaa	39720
cgaacggggaa	cgcataaaat	tgggttagt	gcccgttggagg	cgatgtgttgc	acgaggttgc	39780
agctgtgttgc	aaagcgatgt	gaatgtgtc	agacagacac	accacgttca	cacgaacggga	39840
cttggaaaggcg	ggagaatgtca	cggttgcata	tttgcgttgc	atgtgtatgt	cgatgtgt	39900
aaaaatgttgc	catgtcaatgc	gttataaaat	atgtgtgttgc	acgtgtgttgc	gttgcgttgc	39960
acaccggccg	cgagcggttgc	aaacatgtca	catcatactt	ggcggttactt	ggagcggttgc	40020
gttggaggttgc	tgccagcaaa	gggtgtcaaa	gtgcgttgc	tgcaacttgc	tgcaacttgc	40080
ttacttgc	tgtcgcaaga	caatactca	agaatataaa	aaatataaca	acccaaaaaaag	40140
cccgccaccaa	tttgcgttca	tttgcgttca	ctctccctc	gcagggttata	gtgtgttaca	40200
atacgaatgt	tgacagacac	ttcggggggaa	gttttgcata	gtcgttgc	atggcaaccc	40260
tttgcgttgc	actcgctgcac	actcgacagg	ggttttacac	gtgcgttgc	gacttgcac	40320
tccgttagcac	acggtaaaaca	atgtgttac	tgcaacttgc	cccttaagaa	tccttcgc	40380
cctcaatTTT	taggcgttgc	tccgttctt	tgcaacacac	ctgaaagggt	aatggggat	40440
cctatgatTA	tgctgtcagg	gagggaa	aaacgttacgc	agagccacgc	cgggccacaa	40500
ttcatcgat	cgggaccggg	aggaaaagcg	tcctcggtca	catttgaccc	tcaatgtca	40560

-continued

gcataattta	gtcaaattaa	gctgactccg	ctgggagtgg	acgacgtagg	tcgtcggtgg	40620
tggcattgtc	cgagaggact	ggtgccacgg	ttgctcaatt	gtaacaatcg	ttgactctagg	40680
tcgggtgtga	tgtgtgtggc	cattgttca	acattccact	agttcggtgg	cctcttaaaa	40740
tccactcccc	ggacggatag	ggcaacgcg	agtcacgggc	agcgaactgt	ctgtggcgag	40800
gtgtttgtgt	gttgc当地	tttgaaccga	aaactgtcact	gaccaccaact	acttgcgtgc	40860
tgtttgaac	caggagctct	gcatctccct	gactaactcg	aaaaaaagac	cgcatcccgct	40920
cacattgttt	ctatttctgc	agggacagag	aggtggctca	tggtgtccaa	agttggccac	40980
ggtgccgaa	ttcgaggccc	tacatctcc	aactaatagc	agtgccagcg	cctgttagat	41040
cctgtacta	gcacaagggt	gtgtgtgtgt	tggtgtgggg	agttcaatgt	tgaaatgttt	41100
caccgatatt	tatcccaca	ctgacccctt	ggatgagccca	cggttttgggt	gccatttctg	41160
gtgtgtttt	cgctaaacc	aaccatgtcg	acaataacc	tgatgttga	tatattcag	41220
tgtgtgtgt	tatgtgact	ttattttct	cgcttttcc	cgctgttgc	tgcatgcacat	41280
gtcgccgaa	ctgtcgacac	agattcgctc	tagtggaaat	gcatcgctc	gcattcgctg	41340
ctgcgcggc	tatcgccgtt	atctagat	acgtgtgtgg	ctagtgtagg	ccagggagta	41400
ccatcaccac	aggaaggaa	tggttcgaga	gggcgaatgc	gegcacggc	gttccaaaac	41460
acaaaaaaacg	gtttgatcc	aaactttact	gcatgtttt	cacccggact	cctgcagacg	41520
atggatccac	atggacactg	gagggaaacag	cacagggtca	gegtcagcag	taactggtca	41580
acgctgcgtt	gggttctaa	gtggggcttc	cgcttgcata	gagccttccg	cggagtgtgt	41640
gtgtgtgtgt	gtctgggtgt	ggggaaaatt	ggatgttgc	cggtatgttgc	ctgtttcgcg	41700
tgtgtgtgt	tgtgtgtgt	cagccgttga	ttgttgggg	aatatgtgt	catccatcca	41760
tgccgc当地	cgctcacggg	gtgggggtcg	cagcaccggag	agtttgttgc	gcattaagta	41820
ccttcagttg	caaaggcaat	gcaaaagaaga	atcattttatc	aaacctaacc	atcttcgtc	41880
aagggttta	tattaccctc	ggagaaccac	tttgcactat	gatccggcgt	tgagatttt	41940
tctagtttca	cacattcgac	taatttgcat	ttagcactaa	gatttgcaggc	ccggaaatgt	42000
ttacggcatt	ggcccgtaga	tcgcagaaag	ggccgc当地	aaccaaagaa	atggatgtct	42060
ttatcgaac	gaaacgtcgc	aaattttgcg	cccttttttca	ctgccccca	atagacactt	42120
gcaacaagac	ggcagcggaaa	gagttaaaaaa	ggccagagaag	gcatccggcc	aatgtgttaa	42180
aaagcaccaa	caacaacaa	accaaaaaaa	aaaaactcgca	acccaaacgc	cactcatcag	42240
taacgcgaga	ccagtgcgac	cagccaccca	tctcccttc	aacgcgcggc	tactttccca	42300
gccatataatc	atccacttca	accagattga	gtctcctgc	ggccgc当地	gctgtgaccac	42360
acgtctgtgt	cggtgtctcg	tttgcgttccg	cggtttttgtt	ggcgtgttggg	tggtgtgtgt	42420
ggggggccggg	gagaaggtaa	attatattac	acttgcacac	acgcgcgtt	caagtggggag	42480
atgcacttgt	cgctcttatt	cctcggtgt	gtcccggtct	gcatttgc	ccgttgc当地	42540
gacgcgttgg	ggttttgtgt	acgatcgta	cctttaccgc	gcttgc当地	agggttggaaa	42600
atcatcatca	tatcatcat	catcatcgta	tgtgtatcg	acggggccaca	ctcttgacgg	42660
atcgtcttca	tctcggtcc	ggtccgttcc	cgcttagccc	cctctgc当地	ttgccccgtt	42720
gcagttcggt	aaaaaaatgt	gcataaaaa	agaaatcgaa	ccctccgcac	acacccagg	42780
agggagggc	ggtatgttt	ggtcccggt	atgggtgtga	ttgtgtgggg	ctcgatgtga	42840
gtggcaatac	atttgcataa	ttatgtgtta	ttagtggta	cctgcacagg	gagcagcggc	42900
gccccatgt	ggggggccgg	acggccggaa	ggatgtcgga	tgccaaactgt	caatttttg	42960
tcttgcact	cggtgtcggt	tgtgtgttag	tgtgtctgt	ttttttttt	gcacgtgtcc	43020
tggtggcccc	agggaggag	agggggttgc	tatggggaa	tgtaaaagca	aaacaagcca	43080
cccatccccc	ttctatgtca	tctcggtcc	ttggccaaagg	ccactcccta	tccctctcgc	43140
ctttccccc	cottaatgtc	cctttgtttaa	gaaagacgt	ttgttgc当地	atctctgtct	43200
cctccctcc	catgttccac	catctctgtc	tggagaatcg	tgccacacaca	cacacacacg	43260
cacaggattt	tgacagttacc	gtccccctgt	gggggggtgg	tgaaaagaaa	cacatttccac	43320
gctgtgtgt	accctgtgt	atgtgtacat	cgatcacact	cgggccccccg	ggtgaaggccg	43380
atttgcataat	aaatgtgtact	tgccggaa	tttgc当地	ttggctgtc	cggtggccac	43440
aaccgggtcc	gagcggtgtc	catgtgtcc	gtatttttt	ttgcagocac	tttttacaact	43500
gtgtgtgtct	gttcccgctc	ccgctgtcacc	ggccagttcg	gagatccggag	cgtacgagaa	43560
gtgtatgtc	aatcaacccg	acggggaggac	accctatgtt	agctcgccgc	tggagccgt	43620
agagccaaacg	gggggggggg	ggaaggatgg	aatgtgtaa	gtcgagata	aatggcgcgt	43680
gccccaaac	cagctcgac	cgccggagaaa	ggcgtaat	tgccggccgg	tgtatgtt	43740
ttggccgggg	cgcttccccc	cttccccca	ctgccaatcg	ttctgc当地	ctggatctgg	43800
gccccggc	tgtgtactgc	taattttca	actcagttgc	ttggccgggg	tccgtaaagat	43860
gatccaaatc	actttggaa	agtaatgtgt	gcacaaactt	tctgtggaa	gttgc当地	43920
ttttaaataa	ataaaaaatgt	taattttca	ccaccaattt	cccccg	tactgttccc	43980
tagtttgcgt	tttgc当地	atcaagagga	aaagaagaag	cgatccctt	gccc当地	44040
ccttcgc当地	agtcggggag	aggggacgcgc	aaagaatcc	attgtatggaa	gtctactgc	44100
ctactactact	acacttgcgt	aattttgtta	cgatgttgc	gaatcggtgc	actatcgcc	44160
cgccggatgtt	cggtgtgtgt	ggggcgttcc	cccccttc	cgccggat	ggttaccacg	44220
gtgagcgtt	gcgcgc当地	gcgcgc当地	aatcgatgt	gccc当地	gggtggagg	44280
gccccaaat	tgtgtgtgt	gggtgtggct	gagacggct	gtctaggggaa	gataaaatgt	44340
agtcgggg	ttggggccgg	cgccgtgtcc	tttcgttgc	gccc当地	ttcccaatcg	44400
gatggggc	ggttgc当地	cggggggggg	ttgttatcg	tttgc当地	ctgc当地	44460
ggccggatct	tgaaggccca	agcatggccg	aggcaactgt	cgccaggccgg	tgccgtatgc	44520
ccggtaagaa	gggcgc当地	ggcaagctgt	taagaatgt	ccggctgcac	aggctgcgt	44580
tttcggctt	tgtcttgc	gcacggcatt	ctggagcaaa	agaagaagaa	aaaaatgtat	44640
aaaaagaaga	aatgtgtgt	ttgtgtat	ttgtgtcc	ggaccgtatgc	gtatgttgggg	44700
tttgc当地	tattttgtcg	ctaatgttgc	ttgggtttgc	cactgttgc	ttgtgttgc当地	44760
acttcgaatcg	tccgtact	aatatgtat	tgatgttgc	tttgc当地	tttttgc当地	44820
aatggacgg	actgtatcg	tttgc当地	cgatcccttcc	ttcccttcc	caccacgcac	44880
acacatgcac	acttataccca	acacagttggg	ttgtgtat	gtatgttgc	aaatgcacgt	44940
gtaaaaatgt	caacagccca	tgaatgtat	ttgtgtat	ttgtgttgc	aaaaacacta	45000
gtgggtttttt	tttacaattt	acagtgtgt	ttgtgtgt	ttgtgtgt	aaaaacacta	45060
cttacacaaa	cgcggttact	cgtgtatgc	ttgtgtat	ttgtgttgc	atgactcgcc	45120

-continued

acgattgtgc	cgtttgggtg	gatgatgaaa	agtaaataac	attctttggg	taaatagttg	45180
caacccaaag	ctagtgcac	ctgtgctgg	ttgctccctt	gctggcgtgt	tcggggctcg	45240
cgtctcgct	ccccgtac	ggacacgtaa	atggtagatg	taaaaataaa	gttgcgtc	45300
gggggtgtat	tgaacggccg	tctggggtg	ggtttgagg	ggggaaacccg	ggtatggcca	45360
ggataaaaagg	tggggtgtgt	tgagagctcc	gagggtgaaca	atcggtcg	accacggccg	45420
ggtgtgtgtc	agccaggctg	tgtgoaaact	gcagcggat	gcagggaaagg	ggttaaccgtt	45480
ttcggcgagc	cttcgttag	tttcagcacc	ctcggttacc	cacttctc	ctcctagctt	45540
caccacacgt	ctgtgtgtc	gggggttc	ttctttcttc	actgtatgtt	aaacgtttt	45600
tgaacgatgc	gttttgcgt	cgatffffa	gtttataaca	cggtgtttt	cgacatgtta	45660
acatttacat	tgtatcgt	tgatgtatgt	taatctttt	tattttttt	ctctctttt	45720
cagctactca	ctcggtgtt	tgccagaad	ctgttaatct	cctacgtgt	aagtaaatat	45780
aattaaaaaa	agggaaataat	atatttcaa	gcccgtacac	ggtgtgttag	caaacattha	45840
gtgcttcaca	ctgtacgtt	gaatatttgc	taacacgata	tgttacagcc	gacattaag	45900
catcttaaac	caactgaaac	caacatgtat	ttcttttca	gcaaataggaa	cgtcatttga	45960
aaaatgtgca	ttttagatc	atactttatg	gaatgtatgt	tggttcttcc	cgatgcatac	46020
tgtatagac	cacattgcag	gtcgtatgtt	ataataatcg	gtaacaccaa	tgcgtcacct	46080
ttttctcacc	ttacecgct	cgggacgcta	aatcttgc	gegtttgtt	tctttgcact	46140
tatccctgt	cgcttaggt	agcgtat	taagccagt	gacatgagg	actaccggct	46200
tcccttttgc	gatgtatgc	acagttaca	tcacaaggac	acacacacac	acacacagaa	46260
ataatgtcg	tatggcaat	ggaaatatt	gttattttatc	gcccacatca	ccaacccgatc	46320
gaaattgtcc	caaatacgctt	cgagttacata	attctcttat	ctgtctgcg	ctgggtggcat	46380
ttgtacaaa	acgtataaaa	tgccccgtt	ttaaggcgcac	cgccacacaa	ttgtgggcat	46440
tgagctgagg	ggcgcgcgag	actctatgtt	gtcgcgtat	catcggggc	gggggggtgg	46500
gagcagcggc	tttccgcga	cctttgtc	cctgttta	attttttca	acggacagata	46560
ccagcgtaaa	tactgtgtca	tttatacacc	gggtttaa	caggggacc	gtgtgggaca	46620
taagcagaac	gataaaat	tttgc	tttgcgtata	ctcggcggt	46680	
acgagcgtc	tggttgcata	aaaggtaaaa	caccggagac	gtgtccgc	ttggaaactg	46740
cctccaaacat	acggcccttc	gtccccctc	cctccgtt	taccacggaa	agggcaaaaa	46800
aggggtgtta	tcgtttcgct	gtggatgt	atgatggag	atcacaaga	tcaaaacgggt	46860
gctgggtgt	aaagcagcat	gtctactttt	cgacataatcg	cgctcg	gtgtgttgc	46920
gegtggacat	gttccgtat	catttcgt	attaaatgc	atcgcgatt	attttgaaat	46980
gaaaattgtat	cgccaaagaaa	atcttcaac	cttgatitt	cttccaaaaa	gaaaggagt	47040
cgaatgtcga	atacgagagt	gaaaaagaga	gcgttatgac	agtgcgtt	atggcttaatt	47100
tgcaaaacaat	ttacataggc	cgcacat	cagttcatta	cggtat	taaacaattt	47160
actttttgtc	cgatgttgc	tttttgc	ctccccgggc	gggttgcgc	atgaccggc	47220
aaaggggatc	aaagggttac	gcggcgtt	cgccgcgtc	tgcgttgc	atggagataa	47280
ggcgagcgt	aaagatgtgc	agggggagaa	gggacctcg	acaagaaac	gactacaat	47340
gcacagtacg	aaaacaggaa	gaaactcg	ggccgtat	aaactggcc	cccagggtct	47400
ggacaaaactt	ctttatccaa	gcaagact	ggaaatgggg	aggaaca	gggtccctt	47460
cctcggggcc	ttgtgtgt	gtggggc	gggacccggg	gaaataac	caattcatgt	47520
caatgtcact	gtcactcaac	cccaatcg	aaatgc	tggggccac	cgcgagg	47580
cctcggtctc	ctccgggaag	ttgggttct	ttttatcg	gtggagt	gtc gagaagg	47640
gcaggcgtca	gttgggtt	gtgtacgt	tgtaggggg	gaacgt	cg	47700
aatgtacaaat	tgatcagac	aaaggagc	aaagggttac	gtatgttgc	actgttgc	47760
tttcaggat	caatattcc	cgctgtct	caacatgtt	gtccccaaa	gttgggtgt	47820
ggggggatcg	gtggatc	atgggtg	aaatgtgt	ataaaacaaa	aaaccgtt	47880
gcaatataaata	aaataat	cttcgttct	ccagttttt	tgatgttgc	47940	
ctcgat	ttcttacatt	tagccgt	tgtgcgggg	atgtcgt	ggggcagat	48000
geggcaaaagg	gttccgtac	cgaggacaa	caaaaaatcg	tggcttga	ccgcacat	48060
ctatttcctt	ttgtgtttt	aaaaatgggt	tgcatcc	ttagtct	cttggaaat	48120
tcccttccgt	agggggacaa	ggggacac	acaggagat	cggttcc	tcggctag	48180
tcgggtggaa	tttggatc	tcacccctt	tttctttt	tagtcc	tttgcgttgc	48240
aaataatgc	gttggatc	tcgggttcc	tttcttctt	tttgcgttgc	48300	
acattgtcaac	atttgcgt	acaatgtc	tttgcgttgc	tttgcgttgc	48360	
ctccgtat	gttgggtt	gaaaggaa	accgttgc	tttgcgttgc	48420	
tcattgtca	ctccat	cgaaatcg	taaaccgggt	cggtgtgt	tttgcgttgc	48480
gcatgttcc	tttataa	aaatgtt	gtgttttgc	ggggcgtt	gtgtgttgc	48540
ggccatttgc	attcac	ccgagttc	tccaaataaa	accactgc	tgatgttgc	48600
gtatgttgc	gttgggtt	ataatcc	caaccgttcc	tttgcgttgc	48660	
tttcttgc	gatgttgc	aaacccgg	cctggatgc	cactgttgc	48720	
gggttgcac	aaaggccac	accgttac	ctggatgt	tgatgttgc	48780	
gttgcgtat	agcgactgt	gttgcgtat	tttgcgttgc	tttgcgttgc	48840	
tagacaaacc	ccaaatcg	ctgcgtcccc	gtcatgcgg	cacgaccc	cgacggcc	48900
agaccgttac	gacatgtt	atgttgcgtt	tttgcgttgc	tttgcgttgc	48960	
gcaatataac	tttccat	tttgcgttgc	tttgcgttgc	tttgcgttgc	49020	
aacgcacaca	aaacctgt	tttgcgttgc	tttgcgttgc	tttgcgttgc	49080	
acaacgcac	acgaac	tttgcgttgc	tttgcgttgc	tttgcgttgc	49140	
tgatgttgc	gttgggtt	cctcaagaa	gttggatgt	agcaatgtt	agatccgtt	49200
accgaagcga	tttccat	cacatgt	tttgcgttgc	tttgcgttgc	49260	
gttgggttgc	cgttccat	tttgcgttgc	tttgcgttgc	tttgcgttgc	49320	
atgtgtat	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	49380	
ataataatc	aaacaaat	tttgcgttgc	tttgcgttgc	tttgcgttgc	49440	
ctcggttaca	aaacaaat	tttgcgttgc	tttgcgttgc	tttgcgttgc	49500	
gccttaatta	tat	tttgcgttgc	tttgcgttgc	tttgcgttgc	49560	
atcgttgc	aatcacttac	accgttcc	agccaaatgg	tttgcgttgc	49620	
gaacgttgc	caactgt	taaacaat	gcgttcaact	gaaatactt	tgcgttct	49680

-continued

aacattgttc	atttgaact	aaacaattga	ctcctctaag	ctgatacatt	tgctcaatag	49740
agtttatcaa	tttgttttg	tttcactta	caacaataat	gogaatttag	ttgtcaataa	49800
tgtgtataga	ttgttagaaa	atttctatt	tattataact	caagatcgaa	accaattaaa	49860
acaaattcaa	ataaatttaa	tttgaataga	ttcagaatca	aacaattctg	atgcggcagc	49920
agctcggtta	atatacgatga	atgttttat	tggcggaaagc	aatgttttg	ctgcgatttg	49980
acaatgttca	aaagcacctt	agcgttgtt	agtgaaaac	tttcgaaaac	tttagttgaa	50040
aacgttggct	tgaaaaacaat	ataataactt	gcccgtcata	ccttacttta	aactcttt	50100
ctttgatgaa	ataaaacaat	cgttgatgt	caatccgatt	tatggtaac	gcaaaatttgac	50160
tttcgactat	ggtggggcg	tcaaattgaga	agaagataat	cacaattatt	tctgtacta	50220
tagccaaatg	ataatgttga	aaagacaaca	aaagataataa	caagtgtctc	aagtgtctgg	50280
atgtgtatcc	tttatgtat	aagactgtt	tcttagactgt	tctaaattt	ctacaaggagg	50340
ctttaaacat	ataaaatttg	atataatgc	cctatgtat	ttttgtcccg	agtgtccctt	50400
tttatttata	attaactatt	tattttatgt	ttattataac	ggacacaaaat	agaaaacagt	50460
tattttttca	agactgtca	tttttgcattc	gtaaaaacag	ttcctggaaa	aaagtatgca	50520
actcacagta	cagggtggaa	ataatacagc	ggttgttagag	cgtactgtt	ggacaaggat	50580
attaaatttg	acccaacgct	gttataattt	taccctgtt	ccgcgtgacg	ggcacacaca	50640
ggatcaaacc	actactgaga	aactggatct	gcttcgttcg	cactcgggg	tggaaagtcc	50700
tttccgcaca	gcacaggaca	gtgcaggatt	tgaacatca	agctctcgca	accggcgtaa	50760
cggatccat	ggggacccgg	gttctctgc	cgggatctcc	tttctccaa	gtttgtttg	50820
ctatcttgg	tcgttaatct	taacatgat	agtagttga	cagtgtatct	aaaaaggatc	50880
ggataccaaa	aaggcacgag	tagaaaggag	catgtcttaga	tgtatgttgt	gctatcattt	50940
gggtccaaat	oggatccatcg	gattgacgtc	gggtcgccgt	gtatgtgtt	tagtgaggcg	51000
attgttaggt	gcaattctcc	ctcgtgtt	tccttccgg	aatagaatgc	aaaaaggccac	51060
aatgttaatc	actatcatg	aaagacgaaa	cgggtccgtt	ccgcacccgg	aattttccgg	51120
ctcggccacag	tcgatctcg	cageccccgt	ggggacacat	aaacaagcga	ccaaacaaac	51180
ggaacacacaca	ttcttcattc	tcgttgcgt	ccactctgt	ttttgtaccc	tgctgagact	51240
gtcataaaggc	atgttagtca	aagaaaggto	tcatctgac	gottttaat	gctcacactt	51300
gccccgtccgt	ctggcattc	cgagctccgg	cagcttggg	gcaattgtt	agccgttaga	51360
ggaaaaagaca	cggtacatat	aacgcccgg	tcccagtgt	ttgaggccag	ctgcccgtgc	51420
tactgtgtt	cactgggatt	cggcaaaaca	atttcttcaa	tgtgttgcac	cgaagaacga	51480
acaagggttag	ttgttgcattt	cgctgcgtt	agaggtacgc	cactttttt	ggaaggcaagc	51540
aacccgcttag	ctccctggcc	agactgccc	aactctcaag	tacgttccgg	agattttttc	51600
ggggagcgtgt	gggtgtatg	ttggctcgtt	tcaagagggt	ggtatagcac	attttatctg	51660
ccgcactg	attcgtatgt	catacatca	ccgttgcgtt	agataatctgt	acggagatga	51720
tagacgacg	atggaaaatc	gcacagaaca	aaaggccat	acacgaggac	gaataaaagag	51780
tttgcaggcc	gcacccccc	cgaggggatg	ccacagctgt	ctcgaggagc	aagccgaat	51840
gatttgcatt	cagctgcatt	gtgcaagata	ttgaccgggt	agcatttggt	gatggagatg	51900
aacgtccacc	agagatacca	ccgaaacgcac	tgtctgggg	tgtgcgcag	gttctctgt	51960
agtgcgggtt	gtgcgcattt	aaagactgtt	gagagcttgc	cggttatttt	ttcggtctgg	52020
cacaacaggc	tttgggggtt	ttaaaacago	aaacaaacaa	tgttaatattt	gtgcacaaca	52080
tcaggcaact	tttgcgtgtc	tggtaatata	aaquaacggt	caaaaattt	cagtgcgtat	52140
gtagtgaagt	attgtttga	gaatggttt	aaaataacgg	tttgcgttgtt	atctatcaaa	52200
tttgtcatca	tcgcacataac	ttcaagggca	agttatattt	agttgtattt	agagatcaaa	52260
tacgttcc	cttcgtccat	caataaaaaaa	agccatccaa	acttgcagata	tttgcgtgtc	52320
agtgttggaa	atcgatccac	catgtgtt	tttcaacaaat	aaacaaacca	acaatacgc	52380
tatacacat	ttaaccgcac	tttcccttcc	agggtctagt	ccgcgttccc	actttattt	52440
gagcgttaatg	gtggcaacacg	tccacgttac	cgctctgtat	caagcggtgc	aatcaacgcag	52500
aaatttacagg	tttccagggt	accaatca	tcatttcagg	tcggccactt	ctgcacactc	52560
ggcgagctgt	gtttaaaacgc	atcgctcg	agcattttcc	ttacgttgg	gaaggctatg	52620
ctccatggat	ggcgaggat	cggttgcgt	cttgcgttgc	gttgcgtt	atgaacgcgt	52680
tgaacccccc	acgttgcaca	cagttgcagg	gaggttccag	cgttgggggg	agcctcttt	52740
tttccacggc	cccccccc	cccttcattca	acggcagggtt	gaggtcacta	gtgacactt	52800
aattaaaattt	aaactaatta	aaagaaacgc	tcgtataatt	ttcacaccc	accatcattt	52860
tcggggctact	ggtaatgaaa	ttaatattt	attctatttt	attattaacg	tttacatggg	52920
ggggggggcc	gggggggggg	gggcagaact	cggggcacag	ttgtttgtt	accatcgat	52980
cattgcgt	cgaccgtttc	ggagatgtt	cccttgcac	agcgttccat	tacttccgc	53040
tagtgcgtt	ttttcatacg	tcgtgggggg	tgcgttgcac	attaatctca	gaactcgaa	53100
ctgtctccct	tcgttgcgtt	ccaaatgttac	caatgtgtt	gatcaagcgc	agttagcactc	53160
cgcctcccca	gtggcacaac	atcgccgtt	tattaatgtt	tcgcgttgc	agatcgctt	53220
agcagaacaa	tttgcgttgc	cggttgcgtt	agaagcggtt	tgcccggttac	tacttccgc	53280
tttccctccat	gatgttgcgt	tcgttgcgtt	gggttcccggt	gttataaaggaa	aaagtgcgtt	53340
agtcacggac	ggcaacacat	gaatgtttca	tttgcgttgc	tttgcgggtt	ggcggtgtt	53400
aaggctgtt	gggtgttca	gcacccaaa	ttgaccgggt	cagagggtt	gtcgtttcc	53460
tttgcgttgt	tttgcgttgt	ttatccaagg	gggtgaggac	aaaactgttgc	53520	
cacgatttt	ccggcaccgg	agattttttt	ccgggttggc	tcgttgcgtt	ggtcgcaga	53580
gcaggcaca	ggttttgggt	attttttttt	acgggttac	gtgtgcgggg	atcaacgttgc	53640
agctgggtt	gtgtgttccac	atttttttgc	gattttatctt	tcgttgcgtt	tttgcgttgt	53700
agcgcgtgt	atcgacggcc	taatgtcaat	aaagcttggaa	caagcaagaa	tctggatcat	53760
gcaggcaggc	gggtgttgggg	tgacacacaa	gtgcgcacac	ttatgttacca	tttccatcc	53820
cgtgtatgg	gttaaacatctc	tatctgtgt	gtgcgttgcgt	ggccttgggtt	gaggttttac	53880
aaactcaat	gtcaagcgc	tgtaaaactgt	agaaatgttgc	gatcgttacaa	gaaatgttgc	53940
tatttcatag	tgtgcattt	cattttgtat	agcatgttca	actttggaa	ggaaatgttgc	54000
tacacggc	caataatctc	catacaactt	aaatcaggct	atagtttttt	tttccacaaa	54060
ctgtgtatgt	ttaattatcg	tgttctacc	attgttccac	gttacgttgg	aaaatgttgc	54120
acacttgc	tcggcccccatt	ttcgggggtt	tctacacact	gatataatcat	cgataccac	54180
gttggtaggt	gttaaaagga	taaagccgtt	aacaatattat	acagttcac	ggcaagagcgt	54240

-continued

caatcaaggaa	gggaaatgat	tctttcgctt	tccgttataag	cctcggcaag	gtgcacatcg	54300	
agaaaatatt	gcatgtttaat	aaatcccccc	ctccccacagt	aaacatttgc	tccaaacttcg	54360	
ggactacatgt	gttttggatgt	gttttttat	tcatttttt	gataaatcac	taaatgtgaa	54420	
tcgttactcat	ctgttgcgtt	ttatgtctgt	ggcttaccgc	tgccgaatta	accttgcgaa	54480	
actgtgtataa	aacgttgc	tttttttttt	acggctcaat	cgagggaccc	gctacatacc	54540	
acgcgaaaggc	tttgacttcta	gagtttttgc	ctgttttgc	acaacccgt	tgtaaaaaaa	54600	
aatatggtag	ggggggggaca	caaaaacac	ttggcaatgtt	tttatcggtt	ggcattgcct	54660	
tctacatgtt	agatataattt	gatcggttgc	tcqcgccgtt	tcggtaacc	tttctctgc	54720	
ctcagtgc	caagtgc	gttaatgc	caatggccgc	cggttctcg	tgccccggc	54780	
cttgcgggtt	aaatgttataa	tgatgttgc	tgtaaaaaaa	tgctaaacaa	54840		
tgttcttctt	ggggaaaaaa	ggctatgtt	atcttcggc	gacacaaattt	tggtccgaga	54900	
ccggggaccc	tcaatggcg	ctgc	ccgtttttt	cggttcc	ttgtttgacc	54960	
cttgcgcgt	ctccccctca	cgccgatca	cctgacgggt	tcctgtatgg	aggcgcagag	55020	
acagtggagt	gacatgttac	gacatgttgc	cggttgcgtt	acgcagggg	gagggtcg	55080	
gtcatttagt	gggtttgggc	ttggatgtgg	acggcgttac	acactccac	gaggagaggc	55140	
agcatatgtt	tggttatttt	ggactacata	tcagacatgt	ttgttgcgtt	ggacagaaaa	55200	
agtgttataat	gaaacgttgc	catccagcgt	ggccgcgaac	ttgtgtcccg	gggcgcgtt	55260	
gggtcgcgt	tgggaaatgtt	gggatgtatgg	agtgataagg	tggtgttgc	acaaggatgt	55320	
tgatgttgc	tcgggtatgt	gtgcgcgtt	tgcaaaatgtt	aaacaaaaaa	tattgttacc	55380	
ccaagggtt	cattttgtgc	gtatgc	tttttttttt	atcgatgtt	ggtttgcctt	acaagaacct	55440
gcgcgcgt	atttgcgtt	ataaacacaa	caaggtgtt	atgttgc	atgggggtt	55500	
gttagcgtt	tgacaaaattt	attttgcgtt	actgttttac	ataccaat	ggcatgggc	55560	
atacgttact	ttttatgtt	ttttatgtt	ttttatgtt	ttttatgtt	ttttatgtt	55620	
atccttcgtt	aaatcaattaa	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	55680	
acaatattt	actatgc	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	55740	
atactactaa	cacatgttca	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	55800	
cgaaaataaa	aatatgttca	ttccaggtt	tttttttttt	ttttatgtt	ttttatgtt	55860	
catttgcgt	gtttaaaaaacc	tctgttgc	tttttttttt	ttttatgtt	ttttatgtt	55920	
atccaaagaa	caagaagaaa	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	55980	
caagatgtt	cgctcaaaaa	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	56040	
acagcttgc	tattttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	56100	
gaaaatgtt	gaaaaatgtt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	56160	
attcgcgaca	tagtgcgtt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	56220	
atcaattaaa	aaacggccct	tcccccacag	tttttttttt	ttttatgtt	ttttatgtt	56280	
gcaaaaaaaa	gtatgcggaa	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	56340	
aaaacactcc	catgggggtt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	56400	
tattcttc	cacccatcca	tttgcgtt	tttttttttt	ttttatgtt	ttttatgtt	56460	
tgggttggcg	ggccgtgtt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	56520	
ttaatcttc	atttgcgtt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	56580	
tgaactatcc	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	56640	
aaaggaaatgt	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	56700	
agaatgttgc	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	56760	
atgggttgc	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	56820	
catttgcgt	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	56880	
gggggggggt	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	56940	
atggccaaaca	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	57000	
agcgttgcact	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	57060	
atgggttgc	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	57120	
tttgatttgc	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	57180	
gcaacgtttt	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	57240	
ggatgttgc	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	57300	
tttgcgttgc	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	57360	
gttgtcgtt	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	57420	
gaaaacttaa	atcaataaa	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	57480	
gttagcttc	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	57540	
actgttgcgc	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	57600	
tttgcgttgc	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	57660	
cagtctaa	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	57720	
accccccgtt	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	57780	
ggttcgttgc	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	57840	
ccacacatgc	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	57900	
ccaggcaccg	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	57960	
gtctcgccgg	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	58020	
ggcttcttaca	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	58080	
actatgttgc	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	58140	
ccaaagacat	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	58200	
ggcgttgcgt	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	58260	
tttgttctaa	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	58320	
cttgcgttgc	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	58380	
gtatgttgc	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	58440	
agctcgttgc	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	58500	
aaataacttt	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	58560	
tttgcgttgc	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	58620	
cttttcaac	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	58680	
aaattctaga	cagcaacaaa	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	58740	
agatgttgc	tttttttttt	tttttttttt	ttttatgtt	ttttatgtt	ttttatgtt	58800	

-continued

tggtcggcaa	catcgaccc	ccggcagaag	aagcatactt	tgtgctgacc	aatgcattc	58860
gttagggcg	cagtccccc	tttgatgttt	agcatcgat	aagtatccaa	tacactgacc	58920
atgtgtcgga	tttgtgtgt	tatgtatgt	gtctggcatg	ctctctctcc	tgtctagcga	58980
aaatttcaaa	tatcgtcaa	atgtgttcca	gcagcacatt	atcgggaccc	gtctagctag	59040
tctccacact	cacaattcc	atattttca	caccttggtc	tgaatttgc	gtctcccccg	59100
tggggcatg	gaaaattact	gtgcaactcc	ggacggtagg	tgttgcgttgc	tgcattccaaat	59160
aaacacttca	cgtgtttgc	cagggttgc	gtactgc	caacgggctt	ggcgtgcctg	59220
acgcgtacgg	ctgacaagcg	cgtgcacaa	atgttaactt	gcacacccaa	tcaacacccgt	59280
agcgttaggac	ggcgaacgg	aggcgcactc	cgccgggatt	gacatgaaat	ttcgaacgtg	59340
gttcgaacaa	tccgttcc	ccttacccaa	tgatgttgc	ccgagcg	gaacgggctt	59400
atttcagaa	ggggaaatcg	caaatggat	gtatgtttt	tccggccgt	ttatgacgaa	59460
tgtgtgcata	tccgtgtatg	tgatgtatgg	agcatggcc	ccgtgtgtt	ttggcgtgtgg	59520
caaataataa	aattcaattt	aattaaaatt	gaaataaaaa	ctggaaataa	ttacaataaa	59580
atcataat	tatctcggtt	tagatgtgt	gcaagcta	tataatccaa	tacccgcccc	59640
cgatggggac	atcgcttca	tcatatgtt	tacaataat	gccccaccc	ggaatgtctg	59700
gtacatcg	cttgcatac	ccctgtcccc	ggaaaggac	cgatataat	ttaaccacca	59760
aacctgaccc	ttgttgggc	tacgtcgcc	atcatcgctt	tgtatgtgc	ttgaactcg	59820
gccccgttgg	caagcatgg	aacggaaac	aacaaaaaaa	atcaacca	tgataaaacac	59880
ggcataaacc	gcacagaca	taaccccttgc	taccaacccg	atcgatctg	agtttcgtc	59940
tctgatecg	gtctttaatt	tttcttgc	ttttatatt	tttgcgtttt	tgcccttttgc	60000
ttttccca	ctgggtcga	tttggatgag	ccgtccgtt	cggtcgaaa	atcatgtaa	60060
ggcataat	ctgttatat	gtgcggaaat	aaaagggtcg	attgcata	ggatcgatg	60120
tttgtgcgc	cacccggccc	acactgtct	ccgtccgt	cgatgaaaag	tgcataatgg	60180
tttcaattt	gaatatggca	acgcgttttgc	gaaatgttgc	aaaatctt	cacacaagta	60240
gttccgggt	gattgagcc	atcgattdaa	actcgtttgc	gtgtgtttt	gatcgctca	60300
agctgtgaa	taatggcc	acttggtag	aatgttgc	tttttttctt	ggctacttta	60360
tgtgatgt	tctgtgttgc	gaaacgcgt	gtcgaggat	cggttttctt	aagggtgact	60420
gttgtgtcg	cgatgtgc	ttgtgtgttgc	ttatccatc	gtcaca	cacactcttca	60480
ctgtatgc	cggtcgaaat	tatgttgc	aaaacgcgt	ccaaactgaca	aagggtctca	60540
gggttattcg	gtcaaaatca	gatcaacat	ctgcaataat	cgccgtgt	agtcccccac	60600
acggageg	acttgcat	atcgatgt	cttccggaa	agcaaaaaacg	cactggggca	60660
cgtatgttgc	cgaccaac	cggtacccgt	ggccgtgtgc	caagcg	cggccca	60720
cgtcagcg	acggccac	ctgttacgt	ggatgc	tacgctc	tgatgtgc	60780
tgagttgt	cgatgggt	tttgggttgc	ccaggcgct	tttacagtg	tgctaggttgc	60840
tttgtcggt	gtgcgtgt	agccttaat	gaaacaaa	gaaatgttgc	agcaagatag	60900
aaataacaac	aacgtgcat	aaacatctt	ggtgcaata	tgtcttccc	agacgcata	60960
cccaatc	gacccccc	ccaggggttgc	gtatctt	atctagac	gaccgggt	61020
actgggtct	ccgatcg	tcagaatgtc	tttgcgc	ctcgccctt	atgtatgtt	61080
gtgcggcc	tccgatata	ctccatta	gtatgttgc	cgatgcact	cgagcaacac	61140
gcactgc	atctgtct	caaacatctt	gcttgcgt	agcaaaaa	ccgttacat	61200
gaaggagagg	gaatttcaat	agatc	cgccgtgt	gcacgggt	cgatgttaca	61260
ggttggaaat	tgtacgctc	caggaatgt	cgtacaaa	gaccctcc	agtggatgg	61320
catgtcggt	tgtatggac	cgatgtgc	gcccgttgc	tgacattac	aggaaatgg	61380
gagttcaatt	gtctgtat	tttgcata	taaaaaacat	acaacggc	gggtgtttaa	61440
gacatttca	acatgtgc	tttgcgttgc	ccggccctgt	tccatgttgc	gttaccatc	61500
tgttccctgt	cgatgttgc	ttgtgtgttgc	cgatgttgc	gtaccaaaat	ggccagcg	61560
ttcggccgc	gatgtatc	gacgtatgt	ttgttcttgc	tgatgttgc	gtatgttgc	61620
tccacccctt	cgccgtgt	gaggacgg	tttgcata	atagcggttgc	ctcgccgttgc	61680
aatgtca	cgatgtcg	accgttgc	ccggccgttgc	acgatgtc	ctgtccgt	61740
tcgttgcac	atcccgat	tttgcata	atgtatgt	cgatgttgc	gtccgaa	61800
agcagatgc	cgacacat	cttgcgttgc	accgttgc	ccgttgc	accgttgc	61860
tacaccac	cgacaaatgc	cggttgc	ccggccgttgc	cgatgttgc	ccgttgc	61920
gcgtcg	ccggccgt	cgatgttgc	tttgcata	ggcaggcc	gttgcgttgc	61980
acctgtcg	gataatctc	cagcgttgc	tcgtccgttgc	ctaccgttgc	gtgtccgc	62040
tttaccaaa	catcgccat	caagggtcg	acaaacatgt	catagaaat	gtatgtggc	62100
tgaccgcgt	ttcgatgt	gatgtatgt	aggatccgt	tccatgttgc	ccatggat	62160
tactccgtt	catggcgac	tgatgttgc	acccaaatcc	tgaaatatt	atataactatc	62220
tttgatgt	ccaaatggaa	ggaatgtcg	atcgatgt	cccgactgt	tattggat	62280
ggccgcgt	cttcgac	cgatgttgc	agcatgttgc	gttgcgttgc	gttgcgttgc	62340
acgcgtat	atcccgat	tttgcgttgc	acccaaatcc	actcggttgc	cggtgtgt	62400
gacgttac	gtacccgtt	cgatgttgc	ccatgttgc	cgatgttgc	gtccgtcg	62460
tcgcctgt	ccatgttgc	tgatgttgc	acccaaatcc	tgaaatatt	atataactatc	62520
tttgttaccgt	agaatgtcg	cttgcgttgc	atcatcttgc	cacggat	gttgcgttgc	62580
ccctaccc	gatgttgc	tttgcgttgc	cgatgttgc	ccatgttgc	gttgcgttgc	62640
ccctcggt	cgatgttgc	accgttgc	gttgcgttgc	gttgcgttgc	gttgcgttgc	62700
gtgttgcgt	ccatgttgc	tttgcgttgc	cgatgttgc	gttgcgttgc	gttgcgttgc	62760
gcatccggca	cgatgttgc	gttgcgttgc	cgatgttgc	gttgcgttgc	gttgcgttgc	62820
tttgtcgaaca	ggcttgcgttgc	ggttgcgttgc	ggcttgcgttgc	gttgcgttgc	gttgcgttgc	62880
gtgttgcgttgc	cgatgttgc	gttgcgttgc	cgatgttgc	gttgcgttgc	gttgcgttgc	62940
cgatacttgc	cgatgttgc	gttgcgttgc	cgatgttgc	gttgcgttgc	gttgcgttgc	63000
ccctcggtt	cgatgttgc	gttgcgttgc	cgatgttgc	gttgcgttgc	gttgcgttgc	63060
tgcacgtact	cgatgttgc	gttgcgttgc	atgttgcgttgc	gttgcgttgc	gttgcgttgc	63120
cccggctcg	acaggaggac	cttcttact	tacaagg	aaatgggc	aaatgttacc	63180
agagagtagc	gttgcataat	gaaatgttgc	accacacaca	cacacacaca	cttaccgttgc	63240
tttgcacat	ccccat	ccatgttgc	ggcagaaaaat	tggtaatgt	ccacgttgc	63300
cacccgttgc	atgttgcgttgc	accgttgc	gttgcgttgc	gttgcgttgc	gttgcgttgc	63360

-continued

gcaacgcct	gttttaatg	catacagcg	ttgggtgatc	gctgggggc	gagcatagac	63420
ggcgtggca	gocgcctcg	tctcggtgt	caagaagtcc	gacaccaggaa	aagccaaacag	63480
atagatcgac	atgcgcggag	tagttcaaa	gtacgttaca	acgtggcgt	ctagatcact	63540
gaatttgc	tccaaatgtt	tttgtcacaa	acacaccc	caacgtcaga	gcactcgaca	63600
atcgccatac	ccggttcgg	caaagatcg	catgttcgat	acggccttat	agctgggatg	63660
atgtttattt	cccaactcca	ccgttagcctt	caggccggc	tcgtccagac	aggggaaggc	63720
ggcgccgc	ctaatacgcc	ggaactcg	cgatgttaca	tatttgcgg	taccgttcgc	63780
atcgagatac	gagctgaggt	aaaagccatc	gtcatcgac	cgcaagtcac	cctcqaactc	63840
gagggtc	aaacgtacgagg	ccgggtcaag	cgacgcagg	atcgcaaca	cggcaactc	63900
gcccgtc	caatcggtat	agcgtcgat	ttccagaaac	gtgagggtcg	tgtttggatt	63960
ggatcgctat	agctcgatgg	aggttacg	cagtccggc	tgatgttgc	agatggttt	64020
ggcgtgtgc	cggtatgtcc	gtatgttc	caactcgca	ctgtacgatc	ggtttccgt	64080
gtgcacccgtc	gtctccagg	acagcttgc	gtgcgtcg	acgtgtac	tcggcagtc	64140
gtaccgtac	tcctcgctg	ccacttcctc	caggctgacc	ggatcgacg	tgttcagg	64200
ccgctcgata	tctgtcact	tcggatcg	cgcaatgg	cgacgttgc	gcccgtatt	64260
aaaaacacc	cacagcaat	gtacccat	gtctacaac	ttgaaggdt	attttggac	64320
tgaatctcc	ggtgcgaaat	gtcgcactca	atatcgtaa	tcgcaacagt	ttcggatgt	64380
tttacgacca	gatcgaccac	aaacagtgc	tcgtgtac	acccccccat	aaccgggtg	64440
tggggcaat	gcctttagaa	aagcaattt	tcacctgac	aatttgaattt	tccatatactt	64500
tgtatagca	ggggggctcg	tttggattga	gataagaat	cgatttgc	taataactc	64560
cgaacaagag	ctaatacgcc	ttaatcgctt	atcgctcg	agtgtgat	ttcgttaggg	64620
aataatttgc	gtttaatgc	tgatgtgt	gattttat	tgatgttgc	taatttgc	64680
ctcatctaaa	ccatcgctgt	ctaaaaac	ttgaatat	tatttgcgt	ggccgtttt	64740
ctacatcttt	gaccgtttac	ttgcgcgt	atgattgt	tttgcgtatt	gtgaatatt	64800
aatcattcc	taaatatatt	ggtgcata	tactgtata	aggatagtag	tttagtgc	64860
cagaagctt	gtacaaat	tgaacgtt	agaacacaa	actgtgttgc	tgcataatac	64920
aaatccaaag	taatcgat	aaataacgt	acgcaatgt	tatatcg	cagtgtaaa	64980
caacatgtt	ttttatgtt	aaaaatggc	cgacgcgg	aatatggatc	cgtaaaacg	65040
tctgcctaga	gacatcttct	ttcgtatgt	gtgtgcata	acacccctt	cgctcttt	65100
tagtgcgtac	cacttgcact	cccgatgc	atgtatact	agagttagg	gaaataattt	65160
atatacgat	tagggcgcg	atgtgcgt	acttgcac	accttgcata	ccgaaagggt	65220
agttgcgtt	gttcacgcgt	ttttttcc	ttatatgc	ggtgggtcg	ttcggccaa	65280
gatgtacaa	ccccacgcata	agggtggaa	gagaaacgt	ggccgagaaa	ggaacgttcc	65340
atctaagcc	gcgtggaggg	ctcttgcgt	gcgtgtac	ggcgatc	caacccaaa	65400
gagaaaggcc	gaaattatgt	ttttggcgt	tttgcacca	cagcgtc	tttgacaaa	65460
aaccaaaacg	cctcgcaaaa	tttgcac	cctccggc	cagtttttgc	ggagatgg	65520
agataatgtt	atgtgttca	aatggttaga	catttttac	acgtgaa	aagtttgc	65580
tcgtcccg	cggtttac	gtgtgttgc	tttgcgc	gggggtcg	ttcggccaa	65640
tgtgtgttgc	tgatgtgt	aaaaatgc	tttgcgt	agtcgcgc	caaaccgc	65700
aggccatcc	ctcttttcc	cagtcctt	atacattt	tttgcacca	cgagcccc	65760
ccgttagcgc	tgctgtgt	gttttttca	cggttgc	agaaataat	gcgtgtgc	65820
ctggtttgc	ttacagtcc	ctgcacaa	ccggacta	ggagggcgc	tgatgttgc	65880
tcgtgttgc	ccgtttac	gtgtgttgc	tttgcgc	ccaggaat	cgaaatgt	65940
atttataatt	cagattgt	aaaaatcc	ggcttcata	gttgcattt	agcgttgc	66000
agagtgcac	gaagagcgt	tccgtataa	atgtgc	aaaaggaa	aaacatttt	66060
gcaataata	ctgtttac	ctgtacgggt	aaagtttact	tccagtc	caattgc	66120
tgaatgtc	ttcgacccg	tttgcgt	aaactcgat	tttctatgt	atttatgc	66180
aaaataacaa	aaaaatgtt	gtctgcata	cctcttac	tcttcata	actgtat	66240
tcacgcac	gcataaaa	agatccca	ctagctc	tatgtgc	attttgc	66300
aaaaatgt	gccatattaa	tgagtac	gcaatcg	gacgtgaa	gtctgcata	66360
aaacgcgtt	aaaataggca	cagctctgg	cagtctgt	tctgcgc	gatcaggc	66420
actcataat	cgacgttgc	agcgttacta	gcatttgc	acgttgc	tcatcgat	66480
aatagcccc	cacgc	tccgtatgt	tatccggc	tttttgc	gagcggtca	66540
ggtcgtt	atcgatcg	aaaatattat	gggtatcg	tcggcgtt	tcacgtgc	66600
ttggaga	ttatggc	tcgcacat	gggtatcg	ggaggaggta	ttgttcat	66660
ttgaccac	tcgttatttgc	cgaaatgg	aaatgt	tcgtgc	actaaatgg	66720
ctgttttac	gatgtcgac	tcgtgttgc	ttcgcacgg	cagtattaa	acttccttcc	66780
taaaaatgtc	tttgcact	atgtggc	cactgc	cccgatcg	ccttcgc	66840
aagatcgacc	gttaagggg	gggggggggg	taaccgc	cgatggat	ggatatcg	66900
ggcgttgc	tgtttatgc	ttttgcgt	tcgtgc	ggcgttgc	tgccgc	66960
gatgtgttgc	tcgtgtat	aaaaatgg	cactccgttgc	ttggcgtgc	aactaaatgt	67020
cggtcaagaat	gtggcattaa	tgccgcac	gacggccgt	tttgcattt	tttgc	67080
cgtcaccgc	cgactgccc	tttgc	aaaacgtgc	gtcggtgt	cgagcggt	67140
ttgcgttgc	tgtgtgt	cacccacgc	caggcagat	gcgtgtatc	ctccaaacag	67200
taggactat	aaaaatcaat	tttgcac	tcgttgc	tttgcgtt	atcttcac	67260
ccgcgtgg	aatgtccgc	gggttgcgt	tttgcac	gat	ttttttat	67320
caaaaaataa	aaacacgaaa	taaaacac	ttaaatgtc	atgtgtgt	ttttttat	67380
tattttcgtt	ccacact	tcgtat	cgctttca	tttttgc	tttgc	67440
gttcaatggc	cgatcggtt	ggcgtc	tttgcgt	tttgcgt	tttgc	67500
gaatcatttgc	tgtgtgt	tgttctgt	tgtgtgt	tttgcgtt	tttgc	67560
gctccat	tcacaaccat	tacgacgg	ccagatcat	tcgttgc	tttgc	67620
gctgtgtgc	gtggtaattt	gtgtttt	tatttgc	tttgcgttgc	tttgc	67680
ttcccggtt	gtttttact	tgccgtgt	tgctgtgt	aaacggc	ggaaatgt	67740
tggttccat	cggaatggat	tcaaatgata	atcgat	tttgcgttgc	tttgc	67800
tgacacgac	cccat	gtcattgt	agtcgtc	tttgcgttgc	tttgc	67860
aattcaat	tcaactcacc	ggcaccggaa	gtcgat	tttgcgttgc	tttgc	67920

-continued

aagctttgtat	acattttagta	tttaaatggaa	ctgagcggaa	ttttgtgaag	tacaacgggc	67980
aatatttttattt	attttttatgtat	ttttttatgtat	tgaaatcgctt	gaaaaaccag	tcatcatctt	68040
caggaagttaa	gaaacgacgt	tttcaagatg	ctttgactca	tctgtatgcac	gtgatctcaa	68100
cacaactcc	tcacacataa	tgccaaggaa	ataagttca	ctcaatcgaa	acatgtttgt	68160
gtgtgtgtgt	gtgtgtgttt	gtcggaaaac	gctgctggaa	aatatgcgca	ttttcagttt	68220
ttactaccc	tccgaaaattt	cggtacgggtt	tcgggtgggt	gctcacccagc	ccgccccaaa	68280
gttacacgtt	gatccccc	gggggttcacg	tcactgtctc	gcacgggtgc	ggcgagagac	68340
tggccggcgt	aaagattgaa	cageggttcg	tcccaaaact	aatccgtgaa	tcatcatccg	68400
tggccggcgc	cgagcacggc	gctgcccc	ggagccaagg	ggcagtaaaa	catgtttgt	68460
tttacgacgt	tgggaaaattt	tttttcattt	tcctcgctca	accacttgc	tgtggAACGG	68520
atttgcggc	gctcgtagc	ttttcgaga	tgcgagccgt	tgccctgtt	tttgcgttcc	68580
gaaaaccact	ttgtttcgcc	ttgtttgtt	atgtgtgt	gtgtgtgt	gtgtgtgtgt	68640
gtgtgtgtgt	tttgtgtatgg	aaactaataa	gttttgcgtt	ttcccttccc	ttgtttgtctg	68700
catgtctttt	gtgtgcattt	taaaggaa	ctgactgaca	aaagccaag	tttgtgtacga	68760
cttaggtatgg	tcaaacataa	gtttgggg	gccttcatgt	gtgtatgt	gtgttttttc	68820
cacactccga	ccagtagcgt	agtcaatgt	agacatctc	ccggtaatgt	gcacccccc	68880
agegagcage	gggtgcaac	caacgaacct	tggcttgcatt	gtttttgtt	agtttaat	68940
tttggctgtat	tttggtaattt	tttgcatttt	tgtttatgaa	acgtatggaa	tgacaaaagg	69000
cacaccaggc	aaaccaggc	gaatcgacg	aaaagcaaat	cgcgtaacga	acccgacgtc	69060
caacataact	gcccacccca	tctcgaa	cggtacggc	ggggcactgc	ttcgacgtat	69120
tgcagtgat	tgatgtttt	cageagagtt	ttggcgcgc	cgccacgcgc	attgtgttgg	69180
cgaagggtgg	tgcaaatctg	cacccgaa	cggaagcagc	aaaaacggaa	tcgaaagcgc	69240
agacaccggg	aaacgtatgg	atgttttgc	gctgtatggaa	totacaaaaga	cggtcagtttgc	69300
aatgaattttgg	aaactcgcat	tttgcgttgc	caacgttcatc	ggggatgttt	cattttttt	69360
tttggggagga	tagcaa	acatcaa	cagtggccca	tcacaatgt	gatctacaag	69420
gtgtgtgtgt	tgacggcggt	gtgtgtgt	cgtttaa	acaatgt	acaatcgct	69480
agoagttgtac	gtatgttgc	atttagtggaa	tggaacccgc	ctttaa	gaccccttgc	69540
gcatgcgtgt	gtatgtccgc	catatcc	aaacgtatcc	acacgtat	gaaggcttta	69600
aaaacaaattt	gaaatcgctt	gaaacacata	ttcatagcgt	gcccggcga	gaatgggttt	69660
tacttgcgtc	ttaacgagaa	aggggtgtt	tcttcagtc	ctttcagcg	gggttagttt	69720
tgcattttgtt	gcaaaatcggt	acaaaatcg	ataatcgat	ctaattgttac	ggcgtaacga	69780
cgtgtgtgt	tacttggacc	atttggccac	agcgtgtc	ccggacaa	cgggcaacac	69840
gggggtgggg	tttagttttt	attttacatt	ttttaaatgc	ctcccttgc	tgtgccaatt	69900
gctgtgcgtat	ctgtcagggtt	tcgaacacat	ttcttcgctc	tgtgcagcga	acgcgtgc	69960
atgagcgtaa	gggtgtgtgt	atttcaat	aaaaagaggt	ccagcctgtc	ataaaacactc	70020
actccactgg	ttcccttttc	cgccgggtcg	ctcgccccatc	catcgctat	ggcatacgaaa	70080
atccactcgat	taaaccgcgaa	accacgaacc	gatccggcgc	ggggaaaggaa	caccgtgc	70140
agggggccggg	cgccgaagga	tcgtaaattt	taatatgtt	tttattat	tttagcgttag	70200
cataaaggccg	ggccggctga	gagatggc	ttatgttgc	taatgttata	ttggctttccg	70260
ttcccccggcc	gtgcacccgc	acactgggc	ccgacaagaa	atggcttgc	gtgtactgt	70320
tgtatgtgt	tgtatgcctt	tgctgttatt	tttattttt	tatttcttc	cagtcgaagg	70380
aaacgggtgt	tttggagaa	ttggggagct	ttgcacaatt	gtaccccgac	gggagactcac	70440
tctaataacg	ttcattttca	acaaaataaa	gcatgttgc	aqaaactatcg	tcaatgttgc	70500
tgtgtgtgt	tgtgtgtgt	tgtgtgtgt	tgcgataat	tctgtatgc	tttgcgtatc	70560
agtttttattt	cggtttttttt	tttttacatt	gctcgtaa	ttggcgttgc	acgcaatgtc	70620
gagccgtttt	ggcgagcaag	gaaccgcgc	caagatcggt	ttcggttccc	ttttttttgt	70680
gaatcatgtgt	tgtgtgttgc	ttgtgtgc	aaacgcgc	ctgtgttgc	attgttgc	70740
taactcgcc	actcgatgc	caaaccacaca	cacacacgc	caggcagagg	aaaaacgaag	70800
agtccggata	aaaaatttgc	ttttttgtgc	ttttatgtatc	caatttgcgt	tagacaaga	70860
accggggacga	tgccggaa	gggtttgttgc	acgcacacag	gcacactgtt	ctgggc	70920
tagtcgtatgg	aaatttgc	acgcgtatgt	cgttttgcgc	acatgttgc	tttcatctt	70980
cccttttacc	gttggggat	ggggatgttgc	ttttttatgt	gtatgttgc	ttttttccg	71040
tcaagggtt	gtttgtactgc	ttttgtgttgc	ttttttat	ttttttat	ttttttccg	71100
ggggacggaca	ggggatgtt	tccacgc	gtttttgttgc	tgttcaacaa	tcctcatat	71160
tgtatgtgtca	tttgtcaatgg	aatgttgc	tgtgtgttgc	acacgttca	tcacatttt	71220
tgttccat	tgcccaacac	agttttttt	cggttttttgc	ttttttatgt	ttttttat	71280
catttcgaaca	tgttcc	ttttttat	ttttttatgt	ttttttatgc	ttttttatgc	71340
agtttattca	caaataat	ttttttat	ttttttatgc	ttttttatgc	ttttttatgc	71400
aaaaggctgtt	ttttttatgc	ttttttatgc	ttttttatgc	ttttttatgc	ttttttatgc	71460
tatgtttttgtt	aaatgttcc	atcgatcaca	gaacgc	ttttttatgc	ttttttatgc	71520
atagaactttt	atagaacttgc	ttttttatgc	ttttttatgc	ttttttatgc	ttttttatgc	71580
aattttacatt	tgcgtatgtt	ttttttatgc	ttttttatgc	ttttttatgc	ttttttatgc	71640
aatagaatca	taacattat	ttttttatgc	ttttttatgc	ttttttatgc	ttttttatgc	71700
attttttatca	attttatgt	ttttttatgc	ttttttatgc	ttttttatgc	ttttttatgc	71760
ttttttatcaat	ttttttatgc	ttttttatgc	ttttttatgc	ttttttatgc	ttttttatgc	71820
ttttttatcaat	ttttttatgc	ttttttatgc	ttttttatgc	ttttttatgc	ttttttatgc	71880
ttttttatgc	ttttttatgc	ttttttatgc	ttttttatgc	ttttttatgc	ttttttatgc	71940
ccgttacaaga	aaaggccat	tcttcccacc	cagagcggga	atagatgc	agagaccctt	72000
gcacactttt	ccatcgccat	atgggtgtt	ttttttatgc	ttttttatgc	ttttttatgc	72060
cccttcaat	ttgtcgaggc	tttgcatttt	ttttttatgc	ttttttatgc	ttttttatgc	72120
gtgtgggcac	tccagaaat	ttttttatgc	ttttttatgc	ttttttatgc	ttttttatgc	72180
tcgagaccgg	tgccgtatgc	ttttttatgc	ttttttatgc	ttttttatgc	ttttttatgc	72240
tagttgggt	ggatctacac	ttttttatgc	ttttttatgc	ttttttatgc	ttttttatgc	72300
aaataaaatca	ataaatcaac	ttttttatgc	ttttttatgc	ttttttatgc	ttttttatgc	72360
aaagagccaa	ttttttatgc	ttttttatgc	ttttttatgc	ttttttatgc	ttttttatgc	72420
actagctatgc	cactcgaa	ttttttatgc	ttttttatgc	ttttttatgc	ttttttatgc	72480

-continued

ggtagacatg	tttcgatgtt	gcccgcattt	agcatgtttt	ggccgcacag	agcgacacgg	72540
gaagagccgc	tgcgaaacat	aaggacaggg	cggagtcAAC	cctaccat	gttgegtcTG	72600
tcggtcgttc	tgttgctcc	getctgtgg	gtggcgccgc	agcataggTC	tccgactcgc	72660
tctagcgcgt	tgcagcgtt	ccacacactt	ttttgcacgt	ggggcttgc	caccacttgc	72720
tgccgcacaa	atccgaccg	agcacgttgt	tectctatct	acatttctgc	gccaaccgg	72780
ggatgtggac	gtctctggc	acatcggtcg	aactgtgtgt	gtgtgtgcgt	agatacaaca	72840
tetcgtttag	ttgtgcetcc	gaaagccgaa	caccctgcac	ctgttcatc	gtcggtgtcg	72900
tcgcggttt	atgcctcgcc	gaaaactgtgt	cgaacgttgc	actctactc	tgtcccagt	72960
catccggcac	ggtatcttt	geatccctt	ggggtaagt	ttgggegtgt	cagcacgtat	73020
ttacatcgga	gactatcgca	aaaacgggg	gaaagaagca	gttagccga	aaatgtgtgt	73080
cggaaaattt	vaccatca	tcggggcgcc	agaggaggcc	ggtttccga	gggaatcaac	73140
aaacgatttc	gtgttattt	tgagaagca	gcaaccatct	acgaacgtt	tcttcaaaac	73200
atgaagcaca	caacgacata	ccattcggt	ctgggggaa	acatgtttt	gtgctgttt	73260
tcgcccgat	tgtcttaacc	gaaaagaagaa	aactttctt	atcaacggaa	agactatattt	73320
tttcgcctgt	tcccaaaacc	ttaccataga	aagaaggact	geatgcgcg	gatagcagac	73380
gaaaagaaacc	attnacggc	acatacttgg	gagagaagca	cgttctgtt	aaacaaggat	73440
gtttatgtt	gcfgaataa	ttcagacacg	ctctgagcgc	tttcgggtga	gattagcaat	73500
ggagcatttc	ggcaaaacgg	aagaacgtt	gegttgcga	ttggggcgtt	ttgcttgc	73560
agcgatgac	agtagctgt	ctaaaggcc	tcagctatcc	ggaaaaacgtt	gtctcgat	73620
aatgcccgtt	ggtagcatc	caacaatgc	taaaaggaca	taagacaatgt	caccggagg	73680
ctgcataaca	ccgaaaagg	aggagaaaaa	aaataacgg	cgataaaacgg	gtacaatctg	73740
agtttgttac	tgagctggg	aaaggggctg	agaaaatagg	agcagtagaa	gtcttatgt	73800
ggatgttgc	atcgatgaa	caacgtacta	aaagatcggtt	tttacacggc	ggatttatgt	73860
tggacaatag	cgtagaaatag	cgatgttgt	ttgggatgt	aaataaaaaa	aaacctcatc	73920
acttaccaag	agactaaaaa	gagatttagt	caagtagtgt	tgtagtctt	tttattagct	73980
tgggatttac	tattataact	tatatactt	atcttactt	aaaaatggc	aaaaaaagat	74040
aaatagaaag	atgtcaatac	atcaactt	ttacattgtt	ttataaactt	gttgttact	74100
acttgcgtt	tataacattt	cttatac	tttttttttac	tttataatgaa	acaaacatac	74160
aaacaatgt	ttattttca	ataatccgtt	attttagt	atatgactaa	aactaatatt	74220
gcaataaaat	gactgcactt	cttattgggt	ttgaaattcc	ctgataacgc	aaaaatgtc	74280
atattaaat	atgtgttag	taactaacta	acggcatgtt	tcaatgttg	aacaacgg	74340
tgc当地	tttttatgt	tttttatgt	acagtagaa	cagacatgt	ttttccgtt	74400
tattaaatgt	aaagtgcatt	caaacggat	atgggttac	gtttgaat	aatgtatctt	74460
atgtacatgt	ttaatcgat	gactgattgt	ttcacttatt	ggattacgtt	tgccctgaaa	74520
gttaacattt	accctgtat	gcttgcgc	catcttattt	cttgcattgt	cgctccat	74580
gchgctccata	gttccccacca	gccccccgg	aaagattgt	taacatctt	acgggttata	74640
tacttattaa	tgcgc当地	aaaaataatc	ctgccccact	atgatcgaa	cattgttacac	74700
agtgccggca	ctctccccc	atgtacggta	acaacatgt	tacccacga	aggtcatgt	74760
cgcataccgc	ccaaacatgc	agcgatcacta	agcaagtcgt	gccccaaact	taaataaaaaa	74820
taattgaatc	aatcgagac	ggcttgcgt	ttgttgcgt	aaacgatcgg	attgttgcgt	74880
cagttgcgt	atgtgttgc	cggttgcgt	ttgttgcgt	ggatgggggt	tgtgggggg	74940
tatgtacgtc	agcggtgggt	ggctacgtc	gccccacgt	cggttgcgt	aaacgaccaac	75000
cacccacccgt	tttatcgat	attaaacacg	cgaggtgtac	gctaaaaagg	gaaacccgggc	75060
gattattaa	atatttccc	ctacgatgt	ttgtacggcc	ggccgggtgc	atcgaggag	75120
ggatcttac	tcgggggtag	cgatgttgc	ttttttgt	gaaaaaaa	gataatgt	75180
ctgtgtgtt	aaaggttaga	agttatgcgt	ctgatgcgt	catccccgt	cccttccggc	75240
accggccgtt	ttggatgtcc	tgatcgatgg	gcttgcgtt	cagctgcacc	gaaaccagat	75300
gcacacacgc	gcacccgtcc	ggggccgtt	aggcaatgc	aagcaatca	catcaatag	75360
caagtttatt	tgcacaccc	ccgggttgc	ttgttgcgtt	cgatcgccg	actgttacaa	75420
attgtctgt	ctgcgcctt	agcggttgc	agatcggtt	tgccgttacc	ggtaccgc	75480
actatcaag	tatgttattt	ctgttacaa	catttccat	tacacaggcc	cggtatgtt	75540
aatcagatcc	ggaaaatgt	gtgttgcgtt	atgtgtgtt	gtgtttcggt	ggggatgt	75600
cggaaaattt	tcttttgc	gggggacgaa	gtgttgcgtt	ttgttgcgtt	gtgccaataa	75660
cgttgcgtt	cgcgatgtac	atactatgt	ctgccccact	atttgtgtt	ttgttcttgc	75960
ccgcctgtt	atatgtgtt	gttgc当地	ttgttgcgtt	ttgttgcgtt	ttgttcttgc	76020
gatgttgcgt	catggatgg	cacggaaacac	gacccataac	tgttacaaaca	aacatggatgg	76080
cacaaatgtt	ttataatgtc	ggacgcgtcg	tttacatgc	gatcaatgt	aagtatgtt	76140
tcgtatggatt	atgagcttgc	ggcggatttgc	aaggacacga	aattgtataa	cacaacaaat	76200
gaatgtgtat	tgttgc当地	aatggattt	caggatgtt	atgtctgg	aatgtttgt	76260
ccacaaatgt	catcatatca	tttgc当地	atcaatgtt	ttgttgcgtt	ttgttcttgc	76320
attgcagaaa	ggaaactgttt	tttttttgc	ataaaaaaa	atgttgcgtt	ttcaacatgt	76380
ggttggaaat	agcggaaacca	aggcccttcc	aaacccggaa	ataatgtatc	gtatgtt	76440
acgatcgat	caaaaatgtc	acaaaatatt	gtatgtt	tttcacatgt	agcaccgtatc	76500
gttttgcgt	gtatcgatgt	cggttgcgtt	ttgttgcgtt	ttgttgcgtt	ttgttcttgc	76560
attccatcc	tctccggat	ttggacgcgtt	agcttgcgtt	ttgttgcgtt	ttgttcttgc	76620
tgatgcgtt	ttactacaac	ccacggccgaa	ctgaaaaatgt	cagaaatgt	tagggaga	76680
tcgacaatgt	tcttattgt	atgttgcgt	cgccgttgc	gttgggttgc	ttgttcttgc	76740
ggatccac	cgccacaatc	tcggggacgg	cggttgcgtt	ggcttgcgtt	ttgttcttgc	76800
actttgcac	ctctgtatgc	ataccattt	ccagtcgtac	gtatgttgcgtt	ttgttcttgc	76860
aatccaccc	ccagccgggg	cggaaggagg	acgcttattt	ccaccgtatc	caaatgtatc	76920
acggccgcgt	cgcttgcgtt	tttagcaggca	ggggatgtatc	gccc当地	tgcaagaa	76980
agcatactg	tgaagacccgg	aatgtcaaa	tagtgcgtt	aacttctcc	cgaagatgt	77040

-continued

agctaaat	agcacgagat	tagcatgact	gaagtgactt	ttcaaagtgt	cagaatggct	77100
gcactcgcaa	actagctgga	tgca	ggccccgg	gtgtgtcg	gcatgc	77160
gagcaaccgc	agaggc	ggagaggat	ggaaggagg	agggagt	gaa	77220
taagggtgcc	ctcgccatt	gaagtc	cacgggtt	attccagtc	caga	77280
gacgaagacg	atgttgc	tgcgtctt	gctgtctt	ttgtt	tgatgtat	77340
aataagtgc	atataaaaata	aatcttccgt	aagctt	ttgtgt	ttgtactat	77400
aagcccgtct	ggaagcaagg	aa	ggcagggtca	ggcagggt	agacac	77460
ggagctccgg	agctcccgcc	ggcact	ctcg	cggtat	gtgtattt	77520
tatggatata	tttataatgt	gtctt	tttt	aaaaataa	ttca	77580
tacacccgtca	gatccgggt	ggatggc	cgtgtac	aacgg	ttgtt	77640
cgttcgg	tgggactac	ggatgc	atgtt	atgtt	ggcgcact	77700
cattttcc	cgagttgtca	ggggcgt	tccag	ttctgg	ttgtgt	77760
gctacgatcc	tcaggaatga	ttggaa	acgc	ctggagat	ggggaaaa	77820
aaacgc	atgc	ttgtt	ttcc	atccgt	acgttgc	77880
gacgcacata	atgagtaaa	ggaggg	cagggt	tcttgc	ccac	77940
attgcata	ctcgtgagc	tgtgtc	ggctgaa	aaaa	aaacgc	78000
aggtggaa	tgggattgt	gaacgtgg	ttgaaa	aaaa	aaagccc	78060
agccgc	caatggaa	aacc	agat	tttt	ccgcagg	78120
acaatgtat	ttatgtat	cttgc	cc	ttcag	acgc	78180
agcgagcc	atggctt	gagaa	actcg	gttgc	ggagat	78240
acgtcata	aaagagc	gatgg	gtat	ttttact	ccccc	78300
gtaa	gtatgtt	tcactt	ccacgt	gttgc	atgcac	78360
gtgcct	ccgggt	tgg	actgtt	tttt	cgag	78420
agacttacc	tcatgt	gtgtt	aaagg	tttca	at	78480
ggataa	atcttaagtt	ac	tcgt	tttt	ttgtgt	78540
agaatgtt	aatcgt	gtgtt	tttata	actt	tttcatt	78600
atcoa	atctt	ttgtt	ttgtt	actt	atgtgt	78660
ggtagccg	ttatgtt	tgcc	actgtt	tttttca	aaaaaaa	78720
accaaa	atcaat	tttgt	ccgt	ccgt	ccgt	78780
cttac	tttccat	tgtt	aaac	gggtca	gtggta	78840
attgcata	atgc	ttgtt	ggat	ccac	gtcagat	78900
atgataa	ttccgc	cc	gtgtat	ttttt	ttgtt	78960
gtgtt	ttgtt	ttgtt	ttgtt	ttttt	ttgtt	79020
tgatg	ccat	ttgtt	ttgtt	ttttt	ttgtt	79080
cagaagg	ccat	ttgtt	ttgtt	ttttt	ttgtt	79140
cgggggg	ggcaggaa	ttt	ttgtt	ttttt	ttgtt	79200
gtccgc	tttttt	ttgtt	ttgtt	ttttt	ttgtt	79260
agtagat	tttttt	ttgtt	ttgtt	ttttt	ttgtt	79320
gtat	tttttt	ttgtt	ttgtt	ttttt	ttgtt	79380
catgc	atccat	ttgtt	ttgtt	ttttt	ttgtt	79440
cacac	tttcc	ttgtt	ttgtt	ttttt	ttgtt	79500
cgc	tttcc	ttgtt	ttgtt	ttttt	ttgtt	79560
atac	ttttt	ttgtt	ttgtt	ttttt	ttgtt	79620
tttg	ttttt	ttgtt	ttgtt	ttttt	ttgtt	79680
caac	atccat	ttttt	ttgtt	ttttt	ttgtt	79740
caaa	atccat	ttttt	ttgtt	ttttt	ttgtt	79800
cgaat	ttttt	ttgtt	ttgtt	ttttt	ttgtt	79860
gaa	atccat	ttttt	ttgtt	ttttt	ttgtt	79920
gaaa	atccat	ttttt	ttgtt	ttttt	ttgtt	79980
cgaat	atccat	ttttt	ttgtt	ttttt	ttgtt	80040
ttgt	atccat	ttttt	ttgtt	ttttt	ttgtt	80100
ccac	atccat	ttttt	ttgtt	ttttt	ttgtt	80160
ttt	atccat	ttttt	ttgtt	ttttt	ttgtt	80220
tatt	atccat	ttttt	ttgtt	ttttt	ttgtt	80280
aaag	atccat	ttttt	ttgtt	ttttt	ttgtt	80340
aatt	atccat	ttttt	ttgtt	ttttt	ttgtt	80400
tcaaa	atccat	ttttt	ttgtt	ttttt	ttgtt	80460
accac	atccat	ttttt	ttgtt	ttttt	ttgtt	80520
tttg	atccat	ttttt	ttgtt	ttttt	ttgtt	80580
cta	atccat	ttttt	ttgtt	ttttt	ttgtt	80640
ca	atccat	ttttt	ttgtt	ttttt	ttgtt	80700
gaag	atccat	ttttt	ttgtt	ttttt	ttgtt	80760
cat	atccat	ttttt	ttgtt	ttttt	ttgtt	80820
ccgg	atccat	ttttt	ttgtt	ttttt	ttgtt	80880
cacca	atccat	ttttt	ttgtt	ttttt	ttgtt	80940
ctgc	atccat	ttttt	ttgtt	ttttt	ttgtt	81000
cac	atccat	ttttt	ttgtt	ttttt	ttgtt	81060
cggg	atccat	ttttt	ttgtt	ttttt	ttgtt	81120
gtcg	atccat	ttttt	ttgtt	ttttt	ttgtt	81180
ccc	atccat	ttttt	ttgtt	ttttt	ttgtt	81240
gggg	atccat	ttttt	ttgtt	ttttt	ttgtt	81300
taaaa	atccat	ttttt	ttgtt	ttttt	ttgtt	81360
gaagg	atccat	ttttt	ttgtt	ttttt	ttgtt	81420
c	atccat	ttttt	ttgtt	ttttt	ttgtt	81480
gaaca	atccat	ttttt	ttgtt	ttttt	ttgtt	81540
ctcg	atccat	ttttt	ttgtt	ttttt	ttgtt	81600

-continued

gtccacgaca	gtgcccattgt	tccacacgca	cttctcggg	tatcagccgc	cgtcgacgt	81660
gccccacgtc	gagcacattct	atcggaaagg	gcagcagcag	cagcagcagc	agcagcagg	81720
attggccgaa	ccaaaggaaac	cgacgtcg	gtttcgcgg	ggcagcaacc	ggcttacg	81780
accgaagggt	goatttttct	acggagtg	ggtgaaaat	tgcctaccc	ccca	81840
ttccattgt	accatccatt	agatccacac	tggttccact	cgctgttgc	tgcagcgtac	81900
cggggacagt	goagtgta	gctgtacaaa	aaggtaatgt	tggttagtta	gggttaggg	81960
gggatgggta	gattagacag	taggaatgt	gggatgc	tttacagcc	tttgggtc	82020
ttaacacag	acaacacaga	aggacgcta	gcacaa	tgcacaaa	atggaa	82080
gcaaattggcc	tttggacatt	cttgcattt	cacactgtt	ccggggctg	gggacgttat	82140
tagaggaaaa	acggcaataat	gttgcgtc	acatgggt	cgctccc	ctggggctg	82200
cttacttct	ggcaggatcg	atcatcgct	gttgcgtc	ggcattaaat	aatatcata	82260
tggccaaatc	aacgtttagt	tattgtatg	ggcaggagg	agcaaa	cgaaaaaaa	82320
acgggcacac	tccatcgaa	tggatactgg	aaactctgc	ccctacg	acccttatt	82380
caccctacca	gagccgat	gctgc	tctaaataaa	aataatcc	gcccgtcg	82440
aagcaataat	tttatttct	attatataat	acacacaaat	atgggtc	cat	82500
gcacgtgt	gtgtgtgt	gtgtgtgt	gtgtgacca	gtatggac	acgtggac	82560
ctgtgtgtca	aatagcggt	agcgtcg	gccgaagg	ggctatgc	acgcgttg	82620
tcgcgcgtt	ttcgcgcgt	gcctgat	caatgcct	tttgc	actgc	82680
ttgtcgccat	tgcccttggg	gggac	taaggcaat	gttgc	accgacac	82740
accgaattgc	acactgcacc	cgtcgtc	ggcagggt	gttgc	aatggcgtc	82800
tggagcaacc	aacaaacgaa	cacaaaacaa	aaaaaaaca	aatcaataga	aagaatcgag	82860
ctgtttcgat	tattcaataat	tttacacaa	aatatgcac	gtatccc	gtgggg	82920
ctcattgtcc	gacacttcc	ccccgggt	acctaacc	caccgg	aatcaatgt	82980
ataatgtttaa	agggtggcgt	gccaataact	cccg	gac	acgtagg	83040
atacagagag	cgggagctgc	agtgcac	gtgaaacaac	ctggagac	ctgcattcg	83100
caggcgaaaa	taaacaatac	aaaacaaacc	tccgc	tgc	ctgcaccca	83160
ccggcagcgc	gcaaccagg	gttccattt	ggcacttgg	cggtgt	cgat	83220
ctatgcacat	gtattgtaaa	tatgcattt	gcacagata	aaagagac	ggccgag	83280
aagggttct	gtgagcg	tagccaa	tatgcac	caaactatgc	gcgttac	83340
agatgcgtc	gttttgac	tcggcatt	cacttgg	gtat	ttcgc	83400
taactccgtc	gtgttgtgt	cggtgt	gtatgtgt	cg	ggcggag	83460
ctggcgcgt	gcgtatgc	tcggat	caggcgtc	cgccc	cc	83520
accgcgttca	tgcccaac	cagca	ccggcag	gatc	agg	83580
ttcgtgcca	ttctgtcg	tacatcg	acggac	attgtt	gac	83640
tagcaaa	ggcgccgaa	aagt	ttactat	ccat	tc	83700
tatcattaaat	ggcagcgg	cgtat	cgctatgg	cgtac	gg	83760
gtgcccgtc	gttataaaag	tgat	ttgc	tttgc	tttgc	83820
ggcacactt	ggagcga	gtttgatt	ggctcg	ttcg	tttgc	83880
catccgtg	aaggcg	tttact	tttact	tttact	tttact	83940
tctaggacat	tttcattat	ccta	tttgc	tttgc	tttgc	84000
aacttccaa	aatgaaaaaa	gtcatt	tttgc	tttgc	tttgc	84060
agctttaa	acaaacg	agaac	tttgc	tttgc	tttgc	84120
ttattgt	ccctgacat	tcg	tcg	tcg	tcg	84180
agtaatgt	cgacat	cgtt	cgat	cgat	cgat	84240
tctaattgc	tagagaat	aaaaactat	cttattt	ccgat	attat	84300
gtatcctct	gtatgt	gtat	gtat	gtat	gtat	84360
gttacgacaa	tatcgaaac	tcttctt	atatgtt	aaagac	ctc	84420
ctttgcgt	atcgtaacgt	tcc	tcg	tcg	tcg	84480
tgaaccatc	ctacatc	acgtt	tcg	tcg	tcg	84540
atgcgtgc	ggcgggt	ttt	tttgc	tttgc	tttgc	84600
acacaggat	ttacactc	ctcg	tttgc	tttgc	tttgc	84660
taacgcgc	gcccgc	acggc	tttgc	tttgc	tttgc	84720
ggcgcgt	ggacggc	tcg	tcg	tcg	tcg	84780
taggaactaa	tatata	aggc	tcg	tcg	tcg	84840
gtgtgttgc	gagtgtgt	tcg	tcg	tcg	tcg	84900
ctccagggtt	tcaaacatcg	tcc	tcg	tcg	tcg	84960
cgatactc	cgggcccc	aaagc	tcg	tcg	tcg	85020
tagacgacaa	tccgtcg	tcg	tcg	tcg	tcg	85080
gtgtggaga	aagatagg	tagc	tcg	tcg	tcg	85140
ctagaataac	actcg	tcg	tcg	tcg	tcg	85200
aaaaggcgtc	ttacaac	cgtaat	tcg	tcg	tcg	85260
aatatcttgc	ccacat	ttct	tcg	tcg	tcg	85320
ggcgcaacgg	ctggca	tcg	tcg	tcg	tcg	85380
ccggacgcgt	tttgc	tcg	tcg	tcg	tcg	85440
tgctaaat	ttcc	tcg	tcg	tcg	tcg	85500
gggtgc	tcgt	tcg	tcg	tcg	tcg	85560
ataccgtc	ctgg	tcg	tcg	tcg	tcg	85620
taatccgc	cgaa	tcg	tcg	tcg	tcg	85680
ccgcactacc	gtc	tcg	tcg	tcg	tcg	85740
gcaacggc	acgtcg	tcg	tcg	tcg	tcg	85800
ggacatctgt	gagc	tcg	tcg	tcg	tcg	85860
ctacgc	aaaa	tcg	tcg	tcg	tcg	85920
atttgc	acactgt	tcg	tcg	tcg	tcg	85980
actcatgatt	attgtt	tcg	tcg	tcg	tcg	86040
tttgtcg	aaaa	tcg	tcg	tcg	tcg	86100
tttttctaca	gtctc	acata	tcg	tcg	tcg	86160

-continued

gctggggagt	ggcgcatcaa	actccactaa	tgttgccacac	tgtgtgggg	gggatcaact	86220	
tcttcgcggc	cgtttgcgtac	cgcgggtgg	ccgatgaaa	gacgcattag	atggatttt	86280	
gccaaagaca	caccgttcca	tctggccga	acaacgggtt	caacgggtcg	ctggcagaa	86340	
ggtaatggaa	cgggttccgg	tactgategg	ccattacggg	ctagtcaatt	ttactatgtt	86400	
tcagagataa	ttttatgggt	ttccatttgc	gggaattgt	tttttatgtt	cctcaactgg	86460	
ctgtgaggc	tctcttcgtt	gcccgggtgt	tgtttgcacg	gtttcg	tttgcag	86520	
cgggttgg	cattgtgtt	gtatgatgt	aaacccaga	aaaaaa	aaaaacgat	86580	
aactacatgc	gtctggttt	tctggctgt	aatttagttt	gcagtc	tttc	aacacacaga	86640
cttacacaaa	cctcataccc	taatcattgt	gatggatata	gttc	cgtatc	acgatgttat	86700
tgagggttgt	tcacatattc	ctaataatgt	acatttttt	tttataccat	tttaaatgtat	86760	
gaataaaat	tctacaaaca	tgtataact	catataata	aactattgt	ccaaattaa	86820	
attaagtggc	gtgaaacgt	acaggttgc	cactacgc	aaattacgg	aatatgatct	86880	
aatttgcagt	gaaaatttgc	tttcttgg	tccaaatattt	ccacaac	atataatcatg	86940	
tgaattat	taaaataatgt	tatcatctt	aaaaaaaat	atcatc	aaacatcac	87000	
tagatctca	agttacatca	agccgttgc	tctgttgc	atgtttttt	cgagtgtt	87060	
aaatattta	ctttttctc	atcataacta	taactttt	ctcgat	ttccgctt	87120	
tcaaaataga	tgcattggaa	attcacgtca	atcatctgc	agcccgaa	atgctactt	87180	
gtcgccccca	gtgttgc	ctggcgat	gcaagagat	cagctt	tccttatcg	87240	
tgcacaaaag	gcccgggg	aaacaaaaca	aaaatccaa	ccactcg	acggcc	87300	
tggcggttgc	cactggagaa	actccaccc	tcgttgc	tatctaa	gg	87360	
ttcgccctt	gcccgttgc	aaaatagaag	caaataac	tttgcgg	tttgcgg	87420	
cgcgttgc	caggcgaa	ggaaa	ccctgtc	cagaaaa	cat	87480	
attttgcgt	ctgttgc	agggttac	tgttgc	aatgttgc	atgttgc	87540	
tatttagttt	ttttgttgc	aaaaaaaat	caacat	cggttgc	aaaacaa	87600	
aaaccgaaca	aaacgaaaca	gtacgcac	ggctgttgc	atgtaaatca	atctatgt	87660	
gcgtgttgt	gtgtgttgc	gtatgttgc	atgtgttgc	cgacgcgtt	cccatttt	87720	
gtcgatgttgc	ccccgttgc	atcattgtc	tttgc	aaaat	aaacccat	87780	
cgcgttgc	ggcgccgtt	ttttacttgc	cccccattt	cccacaaa	ttcaaa	87840	
tttttaat	tgtaaaatgt	aatggaaata	gttaacgggc	gtgtgttgc	tgttagat	87900	
ttcgatact	tggaaaccaa	atctcaaaaa	aaagcaat	gaaactt	tttgcgtt	87960	
cagacacacc	ggatgttgc	agttgggg	gcagatgc	tttgcgtt	gtctcg	88020	
atttctcg	tgcgtatgg	aggcgaaa	ttcagagg	catctccat	ggaaat	88080	
ggaagagat	taagaattc	gaaggcaat	ctgttacac	ggcgag	tggttat	88140	
ctagtaaaga	attgtatgt	gatacgtgc	cagtaact	tatttttca	attttgt	88200	
tgttagtgc	tcccgcc	cgtgttgc	gtccgggg	cggttgc	gactccgc	88260	
cttggagcgt	tatccatgt	tgcgtact	tcgttgc	gcctttt	tttgc	88320	
cgttttacaa	aaaaatgtat	tatattgtt	tttgc	tgcgttgc	cacacac	88380	
tgcacacgt	cgtacaatgt	aacgaatgtt	ttgttgc	ctaagt	tttgc	88440	
ttcgcttaatgt	taatgttgc	agatgttgc	ttatcg	ttatcg	tttgc	88500	
acgagatgg	aaagatgttgc	caggggcagg	gcagatgc	aaagggggt	cggttgg	88560	
cttggatgtca	aatatgttgc	gccatctgc	agtcgttgc	aaacattaa	aaacacgtt	88620	
gccttgcgt	aaacaaatacc	cccggagg	tcctgcgc	ggccc	tcgttgc	88680	
aatggatag	gtgtgttgc	ggcgttgc	agttgttgc	gcaagggtt	aatgttgc	88740	
ttgaaaagcg	aaagaaaaac	ctaatgttgc	tataagg	gtatgttgc	atatacgt	88800	
tttacgcgt	aaaaat	tgtatccca	ttttcttcc	tcatcaac	tttgc	88860	
cttcacattt	cacacaagca	agttgtact	taatgttgc	aaaatgttgc	caagcaaa	88920	
atttgcgtca	tcatcgat	ccatgttgc	aacaaat	tcaat	tgttgc	88980	
aatgttgc	tatgttgc	tcgttgc	ttttttaac	acaatcg	tatgttgc	89040	
geccgcgttgc	tgtatgttgc	tttgc	aaaat	acaacat	gtcatat	89100	
tgggtgtcg	tgcgttgc	tcgttgc	aaaaat	tcatttgc	ttgc	89160	
catgttgc	aaaatattac	tccaaacaa	aggacatgt	aaaaat	caagaaacaa	89220	
aacgtttaat	tttacaa	agcttgc	gtatgttgc	ttatgttgc	ttatgttgc	89280	
tatctcaac	aaaggggaga	aaaaacaca	tttac	tttac	tttac	89340	
aggtgtat	agcatatgt	cacaacaca	tatgttgc	tttac	tttac	89400	
ataaaaatgt	aaaaaaac	tcccgcc	tgatgttgc	ttatgttgc	ttatgttgc	89460	
aatggatgt	aaacatcaac	agttgttgc	aaaat	aaaat	ttatccgc	89520	
ccagtacata	atcttgc	gacaaatgt	gtgttgc	ttatgttgc	ttatgttgc	89580	
aagtaatgt	aacatgttgc	atgttgc	tttttgc	tttac	tttac	89640	
ctaaacgtca	tgtatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	89700	
gaaaatgttgc	tttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	89760	
ggaaacaaac	tccgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	89820	
acaaacataa	ataaattatt	aatatttata	ctgttgc	ttatgttgc	ttatgttgc	89880	
gtcgcttgc	tcgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	89940	
ccaaacacca	gtatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	90000	
tttctctgtc	tccctcc	tttgc	ttatgttgc	ttatgttgc	ttatgttgc	90060	
ctctctctcg	tttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	90120	
ctctctcttt	gatttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	90180	
ggtcgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	90240	
gcactttcc	ttttccatc	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	90300	
agtttagcaa	agcgaggag	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	90360	
ttttcccg	tttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	90420	
aaccggaaat	gtggggaga	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	90480	
gtaatcggtc	tttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	90540	
gtttttctac	cccg	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	90600	
gaagtgacagg	caatcatgt	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	90660	
atcttcactt	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	90720	

-continued

```
SEQ ID NO: 2          moltype = DNA  length = 129
FEATURE           Location/Qualifiers
source            1..129
                   mol_type = other DNA
                   organism = Anopheles gambiae
```

SEQUENCE: 2 Organism - Anopheles gambiae
cctttccatt catttatgtt taacacaggt caagcggtgg tcaacgata ctcacgattg 60

-continued

```

cataatctga acatgttga tggcgtagg ttgcgcaata ccacccgtca gagtgatga 120
taaacttc                               129

SEQ ID NO: 3      moltype = DNA  length = 54
FEATURE          Location/Qualifiers
source           1..54
                 mol_type = other DNA
                 organism = Anopheles gambiae
SEQUENCE: 3
cctttccatt catttatgtt taacacaggt caagcggtgg tcaacgaata ctca      54

SEQ ID NO: 4      moltype = DNA  length = 23
FEATURE          Location/Qualifiers
source           1..23
                 mol_type = other DNA
                 organism = Anopheles gambiae
SEQUENCE: 4
gtttaacaca ggtcaagcgg tgg                               23

SEQ ID NO: 5      moltype = DNA  length = 20
FEATURE          Location/Qualifiers
misc_feature     1..20
                 note = Nucleotide sequence encoding a nucleotide sequence
                 that is capable of hybridising to the intron-exon boundary
                 of the doublesex (dsx) gene
source           1..20
                 mol_type = other DNA
                 organism = synthetic construct
SEQUENCE: 5
gtttaacaca ggtcaagcgg                               20

SEQ ID NO: 6      moltype = DNA  length = 97
FEATURE          Location/Qualifiers
misc_feature     1..97
                 note = Nucleotide sequence encoding a nucleotide sequence
                 that is capable of hybridising to the intron-exon boundary
                 of the doublesex (dsx) gene
source           1..97
                 mol_type = other DNA
                 organism = synthetic construct
SEQUENCE: 6
gtttaacaca ggtcaagcgg gtttttagago tagaaatagc aagttaaaaat aaggcttagtc 60
cgttatcac ttagaaaaagt ggcaccgcgt cggtgct                               97

SEQ ID NO: 7      moltype = DNA  length = 1074
FEATURE          Location/Qualifiers
misc_feature     1..1074
                 note = zpg promoter
source           1..1074
                 mol_type = other DNA
                 organism = synthetic construct
SEQUENCE: 7
caacgcgtgc ggtggggaca gctccggctg tggctgttct tgcgactctt cttccgtcg 60
cacatccctc tcgtcgacca gttcgttgc ctgagcgtaa gcctgtgtt gtctgtctg 120
catcatcggg accattgtt gggccatcc gcccacccac ccacccacac ccgcgtccat 180
ttcttagggc ataccatca gcatctccgc gggcgccat ggcgtgtgtt ccaagggtgcc 240
attcgtttt tgctgaaagc aaaaagaaagc aaattatgtt tggttctgtt gcacacgata 300
attttcgttt ctggcgcta gacacaaaca acactgcata tggaggggaa aatttgcgc 360
ctagctgtat aacctacactt aaagtatttgc tccatcggtt tataatggac ctacccgagcc 420
cggttacactt acacaaacgca agattatgcg acaaataatcag akgcggaaact agtaattttc 480
atctatcgaa agccggccgg agagagttt ttttgttgc caacttgacca ttctgtctg 540
ggataaaacgg cgacgggcta ccattggcgca cctgtcgat ggctgtcaaa tttggccgg 600
tttgcgtat ttgtgtgtt aaattatatac ccactcgctg atcgtgaaaa tagacacctg 660
aaacaaataa ttgtgtgtt aatttatatac ttgtgtgtt aatttatatac ttgtgtgtt 720
tatttaatcggtt tcaacggcgc gggaaatggat ttcgtgtat tatccctctc 780
gtctctctg tttatcatgt cttgtgttgc ttgtgtgttgc cccggataat cgagacgc 840
catttttaat ttgtacttactt acacggacaa gcatggcggtt agcttttca agttttctg 900
tccgacccaa gaaacacgaga ataccggcccg gacagtgcgg gggatgtatcg atccatagaa 960
aatcggcccat catgtgccccat tgaggcgaaac cggcgtaatcttca ttccaatgtc 1020
ttccccgtaa catccgcata taacaaacag cccaaacaaaca aatacagcat cgag 1074

SEQ ID NO: 8      moltype = DNA  length = 2092
FEATURE          Location/Qualifiers
misc_feature     1..2092
                 note = nos promoter
source           1..2092

```

-continued

-continued

```

SEQ ID NO: 11          moltype = DNA    length = 1885
FEATURE                Location/Qualifiers
misc_feature           1..1885
source                 note = 5' homology arm
                      1..1885
                      mol_type = other DNA
                      organism = synthetic construct

SEQUENCE: 11
cttgcgttta gcagggcagg ggatgagcg caaactgtgc aagaagaagc atcaactgtga 60
agacggcaat gcaaagatag tgcgtcacac ttctccgcga agattgaagc taaaattaagc 120
acggagattag catgactgaa gtgacttttc aaagtgtcg aatggctgcga ctgcacaat 180
agctggatgc acgcgaattt ccgttcccggt tggtgcgcga tgcaaaacgc caacccgcaga 240
ggccaaagga gaggatggga aggagggggg ggtgtaaaga gcaggcttaa gttttccccctc 300
gggcattgaa gtcgatacag cggttctatt ccagtgcgcag taacgtgac gaagacgtat 360
tttgtcttcg tgcgtgtgtc gctgtgttg ttgtatgtat tgatgataaa atgtccaaat 420
ttaaaataat ctccgttaa ctgggtgttag ttgtgtgtgg ctactataaag cccgtctggaa 480
agaaggaaag ctatgcgggc aggggtcatgc aaaaggggaga caccttcggaa gtcggggagc 540
tcccggccgc actctcgggg ggacgtccgt tatgggttgc gattttattat ggaatattta 600
ttatagtctg ttgttttggaa aaaataactt caacgcgttcg aattttccatc acctccggat 660
cgggggcttggaa gttggcaacgt ggtacggaaac ggtacagcggt tttggccgcgt tcgggtttgg 720
gactcacggaa tcgcacaaatg ttatgtgcg cgcactgtat gggaaatgtat ttttaccggaa 780
gtggtcagggg cgctgtatccc agttcggttc ttggctgtgt tgctgtatgtc acgtatccta 840
gaaatgttggaa gaaacgcctt gggatgttgg gaaaaaaatca aacacaaaaaa cgtatccataat 900
gaacatcggt tggttcttcg cgcgtccacg attgacacctt tgcatggac gcacataatg 960
agctaaagga gaggggacacg gttttgttgcg ttggccacggat cgtatggatc gcaatcactc 1020
gtgagcgtgt gtcgtgtggc tgaagaagaa acgtttcca cagcgtatgg tggaaatgttgg 1080
gattgtggaa ctggcgtatc aaaagaaacctt attttctaa gcccgagacg cctgtctcgaa 1140
actggaaaaac cagatcgaga agttttttat tgtcggccgc caggaaaaaa aatgttattta 1200
atgttttttc tgccctttcc gggccgttcc agacgcggatc ctatgttgc gggccatgtt 1260
gtgtttggaa aactcggtt acccggtggaa gatgtatccc ctgtatgtacg tcatataaa 1320
gagccggat ggcgtatc aaaaagacaca ccagcggatc gacgaaatgttggatc agctggcgat 1380
gtatgttgcg ttgcacatca ctgcgtatc cgtgttgcacg atctactgttgc cttagccacgg 1440
ggtcccgatgt tcacagggtt tagcaaaaac gtgtttttt ttggccggatc ctgcgtatc 1500
tgatcgatc gttaaaggaa gttttcgat gggaaatttt ccgttccatggaa taatgtatc 1560
ttaaatgttccg tgcgtttagt ttgtttacat tggtgttgc acgtttccaaatg atgtttat 1620
cgatcgatggg tagtttattt atactgttgc tattttttttt cttttttttt cttttttttt 1680
agtttagtgcg ctcagacacg agaaggtagg agtatctgcg tcttgcggatc agcggccat 1740
gggtttatgcg cactgttccg gactttccaaat ttgcgtatc aaaaaccccaaaatccaaat 1800
aaaccaatat gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1860
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1885

```

- continued

-continued

gtgttgttgcacac	gatagtttc	gttcttgcc	gctagacaca	aacaacactg	1620
catactggagg	gagaaatttg	acgccttagct	gtataactta	cctcaaagg	1680
gtgggtataat	ggacatccg	agcccggtta	cactacacaa	agcaagat	1740
tcaacggaa	aacttagtaat	tttcatctat	cgaaggcgc	cgagcagaga	1800
attgcaactt	gacattctgc	tgtggataa	accgcgacgg	gtaaccatgg	1860
agatggctgt	caaatttggc	ccgggttgcg	atatggagt	ggtgaaat	1920
gctgatcg	aaaatagaca	cctgaaaaca	ataatgtt	tatcccactc	1980
gaacagcaca	agttttgt	acaatattta	attacgtt	gttatcaacg	2040
attatctcg	tgattatccc	tctegcttc	tctgtctatc	atgtctgtt	2100
taaccccgga	taatecgagag	acgcattt	taatgtt	taactacaccc	2160
cgtgagotct	tctcaagtct	tctgtcgac	caaaagaaaca	gagaatccg	2220
gccccggagt	atcgatccat	aaaaaatcg	ccatcatgt	ccactgaagc	2280
agtttgttcc	gaatttccaa	gtgttcccc	gtaacatccg	catataacaa	2340
aacaaatata	gcatecgat	cgagatggaa	tataaggacc	acgacggaga	2400
catgatattt	attacaaga	cgatcgat	aagatggcc	caaagaaaga	2460
ggtatccacg	gagtcccg	agccgacaa	aagtagacq	tcggcttgc	2520
aactctgtgg	gctggccgt	gatcaccgac	gagtacaagg	tgcccgacaa	2580
gtgctggcca	acacccgacc	gcacagcato	aagaagaacc	tgatcgagc	2640
gacagcggcc	aaacagccg	ggccacccgg	ctgaagagaa	cggccagaag	2700
agacggaaaga	acccgatctg	ctatcgaa	gagatctca	gcaacgagat	2760
gacgacagct	tcttccacag	actgaaagag	tccttcttgc	tggaagagga	2820
gagcggcacc	ccatcttcgg	caacatcg	gacgggtgc	cttaccacga	2880
accatctacc	accttgagaaa	gaaactgttgc	gacagcaccg	acaaggcgc	2940
atttatctgg	ccctggccca	catgtca	ttccggggcc	attctctgt	3000
ctgaaccccg	acaacagcga	cgtggacaag	ctgttcatcc	agctgtgc	3060
caagctgttc	aggaaaaccc	catcaacgc	agccgcgttgc	acgccaaggc	3120
gcoagactg	gcaagagcag	acggotggaa	aatctgtatc	cccaactgccc	3180
aagaatggcc	tgttccggaaa	cctgttgc	ctgacccgttgc	gcctgttgc	3240
agaacttcg	acctggccga	ggatgtccaa	ctgcagctg	gcaaggacac	3300
gaccttggaca	acctgtgtgc	ccagatcg	gaccgttgc	ccgacccgttgc	3360
agaacccatcc	cctgtgttgc	ctgtgttgc	gacatcttgc	gatgttgc	3420
aaggcccccc	tgagccctc	tatgtca	agatacgc	agcaccacca	3480
ctgctgaag	ctctcg	gcacagctg	ccttgc	gatgttgc	3540
cagagcaaga	acggctacgc	cggctacatt	gacggcgg	ccagccag	3600
aagtgtatca	accccatcc	ggaaatggatc	gacggcacc	aggactgt	3660
aacacagagg	acctgtgtgc	gaaacgttgc	acccatcg	ccccccacca	3720
atccaccc	gagagctgc	cgccattctg	cgccggcagg	agatgttttca	3780
aaggacaacc	ggggaaagat	cgagaagatc	ctgaccc	gcatccctta	3840
cctctggcca	ggggaaacac	cagatcg	tggatgc	gaaagacgca	3900
acccccttgg	acttcgagga	agtgttgc	aaggccgttgc	ccggccag	3960
cggatgacca	acttcgat	gaaacgttgc	aacggaa	tgctgccc	4020
ctgtacgat	acttcaccgt	gtataacgc	ctgaccaaa	gaccgagg	4080
atgagaaacg	ccgccttc	gagccgcg	caaaaaatgg	ccatcg	4140
aagaccaacc	ggggaaatgt	ctgaaagc	ctgaaagagg	actatccaa	4200
tggttcact	ccgttgc	ctccgg	gaagatcg	tcaacgc	4260
taccacgatc	tgctgaaaat	tatcaagg	aaggacttgc	tggacaatg	4320
gacattctgg	aaatgtatcg	gtgttgc	acactgttgc	aggacagaga	4380
gaacggctga	aaacatcg	ccactgttgc	gacgacaa	tgtatgc	4440
cggagatata	ccggcttgggg	caggctg	cggaactg	tcaacgc	4500
cagtccgca	agacaatcc	ggatttctg	aagtccg	gttgc	4560
atgcagctg	tccacgttgc	cagegttgc	ttaaaatgg	acatccg	4620
tccggccagg	cgatgttgc	gtacgttgc	attgttgc	ccggccatt	4680
aagaagggca	tcctcg	gatgttgc	agtgttgc	tggttgc	4740
cacaagccg	agaacatcg	gtacgttgc	gacgacaa	tgtatgc	4800
cagaagaa	ccgcggag	aatgttgc	atcgttgc	gcatca	4860
cagatctg	agaacacacc	cgtggaa	aaaaatgg	gttgc	4920
tactac	atgttgc	ggatgttgc	gtgttgc	gacatgttgc	4980
tccgactacg	atgttgc	tatgttgc	gtacgttgc	caacccgt	5040
aacaagggtc	tgaccc	cgacaa	ccggccaa	gacgacaa	5100
gaggctgt	agaatgt	gtacttgc	ccggccaa	gttgc	5160
cagagaaat	tccacat	gaccaagg	gagatgttgc	ccgttgc	5220
ccggcttca	tcaag	gtgttgc	accggcc	tacgttgc	5280
atccctggact	ccggat	caactgt	gacgacaa	ccgttgc	5340
aaagtgtatca	ccctgttgc	caaggttgc	ggaggat	ccatcg	5400
aaagtgtcg	atgtatca	ctaccat	ccggccac	gttgc	5460
ggaaaccc	tgtatca	gtacccttgc	ggggaaatgt	ccatcg	5520
aagggttgc	acgttgc	gtgttgc	aaatgttgc	ccatcg	5580
cccaagact	tcttctac	caacatcg	aaatgttgc	gttgc	5640
aacggcgaga	tccggaa	gtacgttgc	ggggaaatgt	ccatcg	5700
tgggataagg	ccggggat	tgccac	ggggaaatgt	ccatcg	5760
atcgatcaaa	agaccc	gttgc	ggggaaatgt	ccatcg	5820
aggtgtac	acgttgc	gtgttgc	aaatgttgc	ccatcg	5880
cccaagact	tcttctac	caacatcg	aaatgttgc	gttgc	5940
tccaagaaac	tgaagatgt	gaaatgt	ccatcg	ccatcg	6000
ttcgagaaga	atccatcg	ctttctgg	ccaaatgt	aaatgt	6060
ctgtatca	agctccct	gtactcc	ttcgttgc	aaacggcc	6120

-continued

```
SEQ ID NO: 14 mol_type = DNA length = 24
FEATURE Location/Qualifiers
misc_feature 1..24
source note = dsxgRNA-F primer
          1..24
          mol_type = other DNA
          organism = synthetic construct
SEQUENCE: 14
tgcgtttaa cacaggtaa gcgg 24
```

```
SEQ_ID_NO: 15          moltype = DNA  length = 24
FEATURE              Location/Qualifiers
misc_feature         1..24
                      note = dsxgRNA-R primer
source               1..24
                      mol_type = other DNA
                      organism = synthetic construct
```

```
aaacccgctt gacctgtttt aaac 24

SEQ ID NO: 16      moltype = DNA  length = 48
FEATURE          Location/Qualifiers
misc_feature     1..48
source           note = dsx?31L-F primer
                 1..48
                     mol_type = other DNA
                     organism = synthetic construct
```

gctcgaatta accattgtgg accgggttttgc tgtttagcag gcaggggaa 48

FEATURE	Location/Qualifiers
misc_feature	1..49 note = dsx?31L-R primer
source	1..49 mol_type = other DNA organism = synthetic construct

SEQUENCE: 17 tccacacctac ccatgggacc cacgcgtggc gcgggtcacc gagatgttc 49
SEQ ID NO: 18 moltype = DNA length = 50
FEATURE Location/Qualifiers

-continued

```

source          note = dsx?31R-F primer
               1..50
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 18
caccaagaca gttaacgtat ccgttacctt gacctgtgtt aaacataaat      50

SEQ ID NO: 19      moltype = DNA  length = 49
FEATURE          Location/Qualifiers
misc_feature     1..49
note = dsx?31R-R primer
source          1..49
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 19
ggtgttagtg ccacacagag agcttcgcgg tggtcaacga atactcagc      49

SEQ ID NO: 20      moltype = DNA  length = 44
FEATURE          Location/Qualifiers
misc_feature     1..44
note = zpgprCRISPR-F primer
source          1..44
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 20
gctcgaatta accattgtgg accggtcago gctggcggtg gggg      44

SEQ ID NO: 21      moltype = DNA  length = 46
FEATURE          Location/Qualifiers
misc_feature     1..46
note = zpgprCRISPR-R primer
source          1..46
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 21
tcgtggct tatagtccat ctcgagctcg atgctgtatt tgttgt      46

SEQ ID NO: 22      moltype = DNA  length = 50
FEATURE          Location/Qualifiers
misc_feature     1..50
note = zpgteCRISPR-F primer
source          1..50
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 22
aggcaaaaaaa gaaaaagtaa ttaattaaga ggacggcgag aagtaatcat      50

SEQ ID NO: 23      moltype = DNA  length = 51
FEATURE          Location/Qualifiers
misc_feature     1..51
note = zpgteCRISPR-R primer
source          1..51
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 23
ttcaagcgca cgcataaaaaa ggcgccctc gcataatgaa cgaaccaaag g      51

SEQ ID NO: 24      moltype = DNA  length = 20
FEATURE          Location/Qualifiers
misc_feature     1..20
note = dsxin3-F primer
source          1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 24
ggcccttcaa cccgaagaat      20

SEQ ID NO: 25      moltype = DNA  length = 20
FEATURE          Location/Qualifiers
misc_feature     1..20
note = dsxex6-R primer
source          1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 25
cttttgtac agcggtacac      20

```

-continued

```

SEQ ID NO: 26      moltype = DNA  length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
                  note = GFP-F primer
source            1..20
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 26      gccctgagca aagacccaa
                                     20

SEQ ID NO: 27      moltype = DNA  length = 22
FEATURE           Location/Qualifiers
misc_feature      1..22
                  note = dsxex4-F primer
source            1..22
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 27      gcacaccaggc ggatcgacga ag
                                     22

SEQ ID NO: 28      moltype = DNA  length = 23
FEATURE           Location/Qualifiers
misc_feature      1..23
                  note = dsxex5-R primer
source            1..23
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 28      cccacataca aagatacggc cag
                                     23

SEQ ID NO: 29      moltype = DNA  length = 22
FEATURE           Location/Qualifiers
misc_feature      1..22
                  note = dsxex6-R primer
source            1..22
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 29      gaatttggtg tcaagggttca gg
                                     22

SEQ ID NO: 30      moltype = DNA  length = 22
FEATURE           Location/Qualifiers
misc_feature      1..22
                  note = 3xP3 primer
source            1..22
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 30      tataatccgg cggtcgagggt tt
                                     22

SEQ ID NO: 31      moltype = DNA  length = 22
FEATURE           Location/Qualifiers
misc_feature      1..22
                  note = hCas9-F primer
source            1..22
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 31      ccaagagagt gatccctggcc ga
                                     22

SEQ ID NO: 32      moltype = DNA  length = 22
FEATURE           Location/Qualifiers
misc_feature      1..22
                  note = dsxex5-R1 primer
source            1..22
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 32      cttatcggca tcagttgcgc ac
                                     22

SEQ ID NO: 33      moltype = DNA  length = 22
FEATURE           Location/Qualifiers
misc_feature      1..22
                  note = dsxin4-F primer
source            1..22

```

-continued

```

mol_type = other DNA
organism = synthetic construct
SEQUENCE: 33
ggtgttatgc cacgttcaact ga 22

SEQ ID NO: 34      moltype = DNA length = 22
FEATURE           Location/Qualifiers
misc_feature      1..22
source            note = RFP-R primer
                  1..22
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 34
caagtgggag cgcggtatga ac 22

SEQ ID NO: 35      moltype = DNA length = 1712
FEATURE           Location/Qualifiers
misc_feature      1..1712
source            note = nucleotide sequence of exon 5 of the doublesex (dsx)
                  gene
                  1..1712
mol_type = other DNA
organism = unidentified
SEQUENCE: 35
gtcaaggcggt ggtcaacgaa tactcacatg tgcataatctt gaacatgttt gatggcgtgg 60
agtgcgccaa taccaccgtt cagagtggat gataaacttt ccgcaccaact gtaactgtcc 120
gtatctttat atgtgggtgt gtgtatgtt gtttgggtgaa acgaatttca tagtctgtg 180
ctatttttaaa tcaaggccgcg tgcgcactt atgcgcataa gttcaacta gtgtttaagg 240
agtggagcgcg gagaaggccgc ccacgggtaca gaaggccgcg agaatgggtc ggcagccat 300
ctgcactgtt gggtgtcgctc cggcgtctcg gggggaggccg gaggaaattc tagttaaaa 360
tgcggagcgcg aacatggatc cgggttgcgtt cccgttcaag aaacggccgt tacacacaca 420
caggaaaaac tgcagcatgt ttgtacatag tagatccatg agcagggtgt cgttgctct 480
cgaaacgtctt ggacgcacgg ctgcgcgtt atttgcgtat cggtccgcg atcgtgggtt 540
ttcgtactgc cacaaggcccg cttttccca tgcaatctt gcaaccaaac caacaacaa 600
caacaaaaaaaaa ccaatgcaca aatgtatca caccctttt gtatcatctg tatattctt 660
ttcttttgcgt tcttttctat gtggccacag ccccgccggg tagttaatgg cgtcgaaac 720
ccccaaaaacc cccgcacata cagtgatcat acgggtttagt gacaacttgc acctgcagcc 780
cttctgggggt tgccacgtgt agctataactt gtgagatcggg gcgcgcacgg tggtaaagcgc 840
gaatggccgc cacaatgcgtt gtccatccca acactacccc tctggaaacta ccccgctccag 900
ggatgcacccg gtcggctca tgccctcgaa aacatggccg ggetccactg tagtagctcc 960
ggcggttgcgt tgagagaagg atgccttcg aatgtgtcgaa acgtgttgcatt gggcggttcaa 1020
gtgtgtgtgt gtgtgttagg tttagcgaga aacagcgcga gttgcgtgtg ctgaaaagcg 1080
aaggagataat agatgtcata atgaaaatgaa aatggaaaat gaagcaaaag tagaaggccg 1140
aggagagccaa ctgtgttcc actatgtcg aatagttagt tctatgttcc tcaccaatca 1200
accttccaaac catcgatcaa ccaataacctg agtcaacatc gtcatcgatc tcgtgcacaca 1260
acttttattaa aatgtaacct tgcgtcgcc accgttaggtt gatctaaggc gaccttttt 1320
acgggcgcgaa cccacatgcg atcgatccat tctccatca aacacaaacag cctgtacccg 1380
tgggtgtgcaat ttgtgtgtc gttgtgttta tttagcaaaaa aagggaaaaga gtcgcacgaga 1440
gagagataga tccgatcgaa gatgtatcgaa aatgttcgtt gttgttttt 1500
cgtaacacag ttgttttagcc aaaatggggaa ttccaataaa tcccgccggg ggggaaatgc 1560
ggaaataactg cgtacacaca tacatcaatc aaaaagaaaaa atccctgcgc tacatcaatc 1620
ccgtttgcgtt ggtgtgtatc tagagcgcac cactttccac tccactctac aatcaatcaa 1680
tctgtgcaga aggtatggta agacggccctt tg 1712

SEQ ID NO: 36      moltype = DNA length = 23
FEATURE           Location/Qualifiers
misc_feature      1..23
source            note = T2 target site
                  1..23
mol_type = other DNA
organism = unidentified
SEQUENCE: 36
tctgaacatgttttgcgttgg 23

SEQ ID NO: 37      moltype = DNA length = 22
FEATURE           Location/Qualifiers
misc_feature      1..22
source            note = T3 target site
                  1..22
mol_type = other DNA
organism = unidentified
SEQUENCE: 37
gcaataaccac ccgtcagagt gg 22

SEQ ID NO: 38      moltype = DNA length = 21
FEATURE           Location/Qualifiers

```

-continued

```

misc_feature          1..21
                      note = T4 target site
source               1..21
                      mol_type = other DNA
                      organism = unidentified
SEQUENCE: 38
gtttatcatc cactctgacg g                                21

SEQ ID NO: 39      moltype = DNA length = 20
FEATURE
misc_feature        1..20
                      note = nucleotide sequence encoding a nucleotide sequence
                      that is capable of hybridising to T2
source              1..20
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 39
tctgaacatg tttgatggcg                                20

SEQ ID NO: 40      moltype = DNA length = 19
FEATURE
misc_feature        1..19
                      note = nucleotide sequence encoding a nucleotide sequence
                      that is capable of hybridising to T3
source              1..19
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 40
gcaataccac ccgtcagag                                19

SEQ ID NO: 41      moltype = DNA length = 18
FEATURE
misc_feature        1..18
                      note = nucleotide sequence encoding a nucleotide sequence
                      that is capable of hybridising to T4
source              1..18
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 41
gtttatcatc cactctga                                18

SEQ ID NO: 42      moltype = DNA length = 97
FEATURE
misc_feature        1..97
                      note = nd nucleotide sequence encoding a nucleotide
                      sequence that is capable of hybridising to the second
                      target site
source              1..97
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 42
tctgaacatg tttgatggcg gtttagago tagaaatagc aagttaaat aaggctagtc 60
cgttatcaac tgaaaaagt ggcacccgat cggtgct                97

SEQ ID NO: 43      moltype = DNA length = 96
FEATURE
misc_feature        1..96
                      note = second nucleotide sequence encoding a nucleotide
                      sequence that is capable of hybridising to T3
source              1..96
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 43
gcaataccac ccgtcagagg tttagagct agaaatagca agttaaaaata aggctagtc 60
gttatcaact tgaaaaagt gcacccgat cggtgct                96

SEQ ID NO: 44      moltype = DNA length = 95
FEATURE
misc_feature        1..95
                      note = second nucleotide sequence encoding a nucleotide
                      sequence that is capable of hybridising to T4
source              1..95
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 44
gtttatcatc cactctgagt tttagagcta gaaatagcaa gttaaaataa ggctagtcg 60

```

-continued

```

ttatcaactt gaaaaagtgg caccgagtcg gtgct 95
SEQ ID NO: 45 moltype = RNA length = 97
FEATURE Location/Qualifiers
misc_feature 1..97
note = second guide RNA targeting T2
source 1..97
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 45
ttcgaacatg tttgatggcg gtttagagg tagaaatagc aagttaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgct 97

SEQ ID NO: 46 moltype = RNA length = 96
FEATURE Location/Qualifiers
misc_feature 1..96
note = second guide RNA targeting T3
source 1..96
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 46
gcaataccac ccgtcagagg ttttagagct agaaatagca agttaaaata aggctagtc 60
gttatcaact tgaaaaagt gcaccgagt ggtgct 96

SEQ ID NO: 47 moltype = RNA length = 95
FEATURE Location/Qualifiers
misc_feature 1..95
note = second guide RNA targeting T4
source 1..95
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 47
gtttatcatc cactctgagt ttttagagta gaaatagcaa gttaaaataa ggctagtcg 60
ttatcaactt gaaaaagtgg caccgagtcg gtgct 95

SEQ ID NO: 48 moltype = RNA length = 97
FEATURE Location/Qualifiers
misc_feature 1..97
note = guide RNA to dsx
source 1..97
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 48
gtttaacaca ggtcaagccg gtttagago tagaaatagc aagttaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgct 97

SEQ ID NO: 49 moltype = DNA length = 143
FEATURE Location/Qualifiers
misc_feature 1..143
note = U6 promoter
source 1..143
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 49
tttgtatcg tgcgcttgaa gggttgatecg gaaccttaca acagttgttag ctatacgct 60
gcgtgtggct tctaacttta tccatcgcta gaagtgaac gaatgtgcgt aggtatatat 120
atgaaatggta gttgtctct 97 143

SEQ ID NO: 50 moltype = DNA length = 1885
FEATURE Location/Qualifiers
misc_feature 1..1885
note = 5' homology arm
source 1..1885
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 50
gagtggatga taaactttcc gcaccactgt aactgtccgt atctttgtat gtgggtgtgt 60
gtatgtgtgt ttggtaaac gattcaata gttctgtctt attttaatc aagccgcgtg 120
cgcaactgtat ggcgataagt tc当地actgt gtttaadggag tggagccgaga gagccgcacc 180
acggtagaca agggcagcag aatgggtcg cgccttagt gcactgtgc ggtgcgtccg 240
gcgtctcggg gggaggccgaa gggaaatcttca gtgttaatc ggagccgacaa aaacaaaaca 300
gtggtcgtcc cgttcaagaa acggctgtta cacacacaca gaaaacactg cagcatgtt 360
gtacatagta gatccttagag caggtggatcg ttgctccctcg aacgcgtctgg acgcacgcgt 420
tcgcgcgtat ttgcgttagcg ttccgcgtat cgtgggtatt cgtactgcga caagccgcgt 480
ttctccccatg caatctctgc aaccaaaacca acaaacaaca aaaaaaaaacc aatcgacaaa 540
atgaatcaca ccccttttgc atcatctgtt tattctgtt ctttgcgttc ttttctatgt 600

```

-continued

ggcccacgccc	ccggcggtta	cgtaaattgcg	tcgaaaaccc	cggacataca	660
gtgtacatac	gttttgagga	caacttgcac	ctgcagccct	tctgggggtg	720
ctatacttgt	gagatcgggc	gcccacgggt	taaagcgcga	atggccgcga	780
ccactccaa	actaccctc	tggactacc	ccgtccagg	atgcacccgc	840
ccccctgaaa	acagtccggg	cttccactgt	gtagctcgg	cgttgetctg	900
gccttcgaa	gtgtcgaaag	cgtgcattgg	gcgttcaagt	gtgtgtgtgt	960
tagcgagaaa	cagcagcagt	tgcggtgtct	gaaaagcga	ggagaatag	1020
gaaaatgaaa	atgaaaatga	agcaaaatgt	gaaggcggag	gagagaaccc	1080
tagtagcgaa	tagtttagtc	tagttcgct	accatcaac	cttccaacca	1140
aataccatcg	tcaacatcg	categttac	gtgcccaca	tttattaaaa	1200
tccgcgcac	cgtagggtg	tctaaaggcg	cctttttac	gggcgcgacc	1260
cgtcaccc	tccaatcaaa	accaacagcc	tgtaccatg	gtgtcaatt	1320
gtgtgttatt	agcaaaaaaa	gagaaagagt	cgacgagaga	gagatagatc	1380
gtacaaaaaa	gcagtagaaa	ttgtgttgc	ttgttttcg	taacacagg	1440
aatggaaaatt	tccaaataatc	ccggggggcg	ggaatgcgg	gaatactcgg	1500
catcaatcaa	aaagaaaaat	ccttcgcgt	catcaactac	ttttgcggg	1560
gagcagacca	ctttccactc	cactctacaa	tcaatcaatc	tgtcagaag	1620
acggcccttgc	acgcgatc	ggtcggccacc	ataacgcgt	ccgacgagg	1680
actttgtctc	tgcatttttc	gtttttct	tatccccatc	cctttctct	1740
cttttgcact	gcccttgt	acccccaaaa	aggtaacga	cacataga	1800
gttggtaag	tcategc	atccgaacag	cgaccggctg	acggaggacg	1860
cgagaacatc	tcggtgaccc	gcacc			1885

SEQ ID NO: 51 moltype = DNA length = 8251
 FEATURE Location/Qualifiers
 misc_feature 1..8251
 note = multiplex CRISPR construct
 source 1..8251
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 51

tgccgggtgcc	aggcgctgcc	cttgggctcc	ccgggcgcgt	actccaccc	acccatgcga	60
tcgcgtccgg	aagatacatt	gatgagttt	gacaaaccac	aactagaatg	cagtgaaaaa	120
aatgctttat	ttgtgaaatt	tgtgtatcta	ttgttattt	tgtacattt	ataagctgca	180
ataaaacaatg	taacaacaac	aattgcattc	attttatgtt	tcaggttca	ggggaggtgt	240
gggagggttt	ttaaagcaag	taaaaatctt	acaatatgtt	tatggctgt	tatgatctag	300
agtccgcggc	gctacaggaa	cagggtgtgg	cgccctcg	tgcgcgt	ctgctccacg	360
atggtgttagt	cctcgtgt	ggagggtgt	tccagcttg	agtccacgt	gtagttagccg	420
ggcagctgc	cgggcttctt	ggccatgtac	atggacttgc	actccacca	gtagtgccg	480
ccgtcccttc	gcttcggc	cttgcggatc	tgcgcgttca	gcacgcgtc	cggggggtac	540
aggcgcgtcg	tggaggcctc	ccagccatcg	gttcttctc	gcattacgg	gccgtcgag	600
gggaagtta	cgccgtatgaa	cttacccctt	tagatgaacg	agccgtctg	cagggaggag	660
tcttgggtca	cggttacacc	gcccgcgtcc	tgcgaagtca	tcacgcgtc	ccacttgaag	720
ccctcggggg	aggacacgtt	cttgcgtcg	gggatgtcg	cggggttctt	cacgtacacc	780
tttggagccgt	actggaaactg	ggggggacagg	atgtccagg	cgaaaggccag	ggggccccc	840
tttgcgttac	tcagtttca	gggttgcgtt	ccctcgtag	ggggccctc	gccctcgccc	900
tcgatctcg	actctgtggc	gttccacgtt	ccctccatgt	gcaccttga	gcatgttac	960
tccctgtatg	cggttttgg	ggagcggac	atggtggcga	cttgcgggtt	ccggggccgc	1020
ggtaccgtcg	actctgtgg	taccccgatt	gtttagtttgc	ttcagtcgc	tttgcgttatt	1080
tgcttagttt	tcgcgttagc	acgtgttca	tttgcgttgc	tgaattgt	tgcgcgtcc	1140
tagacgaaac	gctcttattt	atactccgg	ggtgcgggtt	tgcgaatcga	taagtttgg	1200
tccttaattt	attagtgc	atgttatttgc	tctctatttgc	aattagatcc	ccggggcgac	1260
tcgaattaaac	cattgtggac	cggttacgtc	tggcggtgg	gacagtcgc	gttgcgttgc	1320
ttcttgcgt	tcatcttcc	gcccgcacatc	cctctcg	accagttca	tttgcgttgc	1380
gtaaagctgc	tgctgtgtc	ctgtgtatc	cgggaccat	tgtacggccc	atccggccacc	1440
accacccatca	ccacccgcgt	ccattttca	gggcatacc	atcgcacat	ccggggccgc	1500
cattggcggt	ggtgcgttgc	tgcatttgc	ttgtgtgt	aagcaaaaa	aagcaattt	1560
gttgcgttgc	tgctgcacac	gatagtttgc	ttttcttgc	gttgcgttgc	gttgcgttgc	1620
catctggagg	gagaaattttgc	acgcgttgc	gtataactt	cctcaaaat	tttgcgttgc	1680
gttgcgttgc	ggacacatcc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	1740
tcacagcgaa	aactgtat	tttcatctat	cgaaagcgc	cgacgcgg	gttgcgttgc	1800
attgcacattt	gacatttgc	tgtggat	accgcgac	gttaccat	ccgcacgtc	1860
agatggctgt	caaatttggc	ccgggttgc	atatggat	gttgcgttgc	tttgcgttgc	1920
gtctgtatgc	aaaatagaca	cctggaaaaca	ataatttgc	tgttatttgc	tttgcgttgc	1980
gaacagcaca	agttttgcgt	caataat	tttgcgttgc	tttgcgttgc	tttgcgttgc	2040
attatctcg	tgattatcc	tctcg	cttc	tttgcgttgc	tttgcgttgc	2100
tcaccccggt	taatcgag	acgcacattt	taatttgc	taatgcac	ccaaacatgc	2160
cgtgagatct	ttcaagttt	tctgtcc	gac	tttgcgttgc	tttgcgttgc	2220
gcccggagtg	atcgatccat	agaaaaatcg	ccatcatgt	ccactgt	tttgcgttgc	2280
agcttgcgt	gaatttccaa	gttgcgttgc	gttgcgttgc	gttgcgttgc	tttgcgttgc	2340
aacaatatac	gatcgatgt	cgatgttgc	tataaggac	acgcgcgg	tttgcgttgc	2400
catgatatttgc	attacaaaga	cgatgtat	aaatgtgc	tttgcgttgc	tttgcgttgc	2460
ggtatccacg	gagtcccg	acgcacaa	aaatgtac	tttgcgttgc	tttgcgttgc	2520
aactctgtgg	gttgcgttgc	gttgcgttgc	gttgcgttgc	gttgcgttgc	tttgcgttgc	2580
gttgcgttgc	acaccgc	gttgcgttgc	gttgcgttgc	gttgcgttgc	tttgcgttgc	2640
gacagcgccg	aaacagcc	gttgcgttgc	gttgcgttgc	gttgcgttgc	tttgcgttgc	2700

-continued

agacggaaga	accggatctg	ctatctgcaa	gagatcttca	gcaacgagat	ggccaagggtg	2760
gacgacgact	tcttccacag	actggaagag	tccttcctgg	tggaagagga	taagaagcac	2820
gagccgcacc	ccatcttcgg	caacatctg	gacgggttgg	cctaccacga	gaagtacccc	2880
accatctacc	acctgagaaa	gaaactggtg	gacagcaccc	acaaggccga	cctggggctg	2940
atctatctgg	ccctggccca	catgtatcaag	ttccggggcc	acttctgtat	cgagggcgac	3000
ctgaaccccg	acaacagcg	cgtggacaag	ctgttcatcc	agctgggtca	gaccataaac	3060
cagctgttcg	aggaaaaccc	catcaacgc	agcggcgtgg	acgcaagggc	catectgtct	3120
gccccactg	gcaagagcg	acggctggaa	aatctgtat	cccagctggc	cggcgagaa	3180
aagaatggcc	tgttcgaaaa	cctgattgc	ctgagcctgg	gcctgacccc	caacttcaag	3240
agcaacttcg	acctggccg	ggatgccaag	ctgcaagctgg	gcaaggacac	ctacqacgac	3300
gacctggaca	acctgtggc	ccagatcg	gaccatgtac	ccgacatgtt	tctggccg	3360
aagaacctgt	ccgacgccc	cctgtgtac	gacatcttgc	gagtgaacac	cgagatcacc	3420
aaggcccccc	tgagegcctc	tatgtatca	agatacgac	agcaccacca	ggacctgacc	3480
ctgctgaaag	ctctctgtgc	gcagcagctg	cctgagaatg	acaaaagat	tttcttcgac	3540
cagagcaaga	acggctacgc	cggcttacatt	gacggcggag	ccagccggag	agagttctac	3600
aagttcatca	acccatctct	ggaaaagatg	gacggcaccg	aggaaactgc	cgtgaactgt	3660
aacagagagg	acctgtgcg	gaagcagcg	accttcgaca	acggcagcat	ccccccacca	3720
atccacccgt	gagactgtca	cgccattctc	cggccggcagg	aatgttttta	ccccattctg	3780
aaggacaacc	ggggaaagat	cgagaaatgc	ctgacatttc	geatccccca	ctacgtggc	3840
cctctggcca	ggggaaacag	cattgtcc	tggatgtac	gaaagagcga	ggaaaccatc	3900
accccttgg	acttcgagg	agtggtggac	aaggggcgtt	ccgcccagag	cttcatcgag	3960
cggatgacca	atccgatata	gaatctggcc	aacggagaagg	tgctgc	ccaa	4020
ctgtacgat	acttcaccgt	gtataacgc	ctgacatttc	tgaaataatgc	gaccggg	4080
atgagaaagc	ccgccttct	gagccgc	cgaaaaaa	ccatctgtga	cctgtgtt	4140
aagaccaacc	ggaaaatgtac	cgtgaagcg	ctgaaagagg	actacttca	gaaaatcgag	4200
tgtttcgact	ccgtggaaat	ctccggcgt	gaagatcggt	tcaacgc	ctc	4260
taccacgatc	tgctgaaaat	tatacaggad	aaggacttcc	tggacaatga	ggaaaacagag	4320
gacattctgg	aaagatctgt	gctgaccctg	acactgtttg	aggacagaga	gatgtcgag	4380
gaacggctga	aaacctatgc	ccacccgtt	gacgacaaag	tgatgtac	gctaaggcg	4440
cggagatata	ccggctgggg	caggctgac	cggaagctga	tcaacgc	ccgggaca	4500
cagtccggca	agacaatctc	ggatcttcgt	aatgtccgac	gcttccgac	gcttcc	4560
atcgactg	tccacgac	cgtgttac	tttaaagagg	acatccagaa	agcccagg	4620
tccggccagg	gogatagcc	gcacggac	attgcca	atc	tggccggc	4680
aagaaggc	tcctgcagac	agtgttgc	gtggac	gac	tcgt	4740
cacaaggcc	agaacatctg	gtatgttgc	gcccag	gac	ccagaa	4800
cagaagaaca	gcccgcgag	atgtatgt	gttgc	gat	gttgc	4860
cagatctg	agaacacacc	cgtggaaa	acccatgtc	agaac	gatgtt	4920
tactactgc	agaatggcg	ggatatgtac	gtggacc	acttgg	gat	4980
tccgactac	atgtggac	tatctgtc	ca	gat	ggcc	5040
aacaaggatc	tgaccagaa	cgacaa	gggg	gat	ccccc	5100
gaggtctg	agaagatgaa	gtactatgt	cgccatgtc	tga	acgc	5160
cagagaaatg	tgcataatc	gaccaagg	gatgttgc	gatgtt	ccgg	5220
ccggcgtca	tcaagac	gctgtggaa	acc	ccgg	ccaa	5280
atctctgact	cccggtgaa	caactgt	gacg	gatgt	ccgg	5340
aaagtatca	coctgtact	caatgttgc	gttgc	gatgtt	ccat	5400
aaagtgcgc	agatcaacaa	ctaccac	gcccac	gatgtt	ccgg	5460
ggaaaccgc	tgatcaaaaa	gtatcttgc	ctg	gatgtt	ccgg	5520
aaagtgtac	acgtgcgg	gtatgtc	aa	gatgtt	ccgg	5580
cccaatgt	tcttcac	caatcat	acttttca	agatcc	gatgtt	5640
aaacggc	tccgaaagcg	gcctctgt	gagacaa	ccg	gatgtt	5700
tgggataa	ccgggattt	tgc	ccatgt	gatgtt	ccgg	5760
atctgtt	aaacggat	gacac	ggat	ccat	gttgc	5820
cccaatgt	ataaatgt	ccggc	aa	gatgtt	ccgg	5880
ttcgac	ccaccgttgc	ctattctgt	ctgtgtt	ccaaatgt	ggatgtt	5940
tccaa	aaatgt	tgaaatgt	ctgggtt	ggatgtt	ccgg	6000
ttcgat	aaatctat	cttctgt	ccaaatgt	aaatgt	ccgg	6060
ctgtat	actgttct	gtatcttgc	ttcgatgt	aaatgt	ccgg	6120
ctggctct	ccggcgtact	gcagaagg	aaatgt	ccat	atgt	6180
acttctgt	actgtggcc	ccatgt	acttttca	gatgtt	ccgg	6240
cagaaac	ccatgt	caatgt	acttttca	gatgtt	ccgg	6300
ccggatgt	tccaaatgt	gtatctgt	ccatgt	gatgtt	ccgg	6360
tacaaca	accggat	gcccac	gttgc	gatgtt	ccgg	6420
accctgac	atctggg	ccctgtcc	ttcaatgt	ttgtac	ccatgt	6480
aaaggt	taca	ccatgt	actgt	ccatgt	ccatgt	6540
ccggatgt	ccatgt	caatgt	actgt	ccatgt	ccatgt	6600
ccggatgt	ccatgt	caatgt	actgt	ccatgt	ccatgt	6660
aatcatatgt	ccgcattt	cgcaaa	actgt	ccatgt	ccatgt	6720
cgaaatgt	aaatgt	ccatgt	actgt	ccatgt	ccatgt	6780
ttccaatgt	ttgtat	ccatgt	actgt	ccatgt	ccatgt	6840
gtgccaat	ttgtat	ccatgt	actgt	ccatgt	ccatgt	6900
ttgttatata	ttgtat	ccatgt	actgt	ccatgt	ccatgt	6960
agagagaact	catttgc	gtatctt	ttatgt	ccatgt	ccatgt	7020
ggaacagc	gatatgaca	tttacat	ttatgt	ccatgt	ccatgt	7080
tatac	accatgt	tttacat	ttatgt	ccatgt	ccatgt	7140
ataagcaaa	taaagaagca	tttacat	ttatgt	ccatgt	ccatgt	7200
atctcttag	ttatgt	ccatgt	ttatgt	ccatgt	ccatgt	7260

-continued

gtgcgtcgta	tgggaccgcg	cgcacatttctt	ctctcagctaa	atccccctaa	tcatttagtaa	7320
ttgacggcac	gattttctgt	tcctacttcc	ttttacttttgc	gagctttca	tcaataaaaac	7380
cagtaccatg	gcgcgtacgt	caacggaaaa	gcattcaaaa	aaacccgcgt	tcctcggttg	7440
atttgtgggt	gagtggcgcc	atctattaga	gaatgcgtt	actacatctc	gtggacgaaag	7500
gggtcagaga	agttgaaaga	gagttgtatc	gactgtatc	caagctaggc	gaggaaggga	7560
gtacgtcgt	gcaaaaagaaa	aaaaataaago	aaatatcttt	ttttataaca	aatcgacgtt	7620
agcgaatata	gtttaatcg	atthaacgtt	tagaattccc	tttgggtcg	tcattatgcg	7680
aggcgcgcgt	ttgtatgcgt	gcccgttgcg	ggttgcgtgg	aaccttacaa	cagtgttagc	7740
tatacggctg	ctgtggctt	ctaacgttat	ccatcgctag	aagtggaaacg	aatgtgcgt	7800
gttatataata	tgaaatggag	ttgtctctgt	ctgtttaacca	caggtaaagc	gggttttaga	7860
getagaataata	gcaagttaaa	ataaggcttag	tccgttataca	acttgaaaaa	gtggcacccg	7920
gtcgggtgtt	tttttttttg	tatgcgtcg	cttgaagggt	tgatcgaaac	cttacaacag	7980
ttttagctat	acggctgcgt	gtggcttcta	acgttatacca	tcgttgcgtt	tgaaacgaat	8040
gtgcgttagt	atataatata	aatggagtat	ctctctgtgt	caataccacc	cgtcagagg	8100
ttttagagctt	gaaatagcaa	gtttaaataaa	ggtctgttcc	ttatcaactt	gaaaaaggtgg	8160
caccggatcg	gtgctttttt	ttacgcgtgg	gtcccatggg	tgagggtgg	tacgcgcgg	8220
gggagccca	gggcacggcc	tggcacccgc	a			8251

SEQ ID NO: 52	moltype = DNA length = 48				
FEATURE	Location/Qualifiers				
misc_feature	1..48				
source	note = multidsx?31L-F primer				
	1..48				
	mol_type = other DNA				
	organism = synthetic construct				
SEQUENCE: 52					
gctcgaatta	accattgtgg	accggcttgc	tgttttagcag	gcagggga	48

SEQ ID NO: 53	moltype = DNA length = 44				
FEATURE	Location/Qualifiers				
misc_feature	1..44				
source	note = multidsx?31L-R primer				
	1..44				
	mol_type = other DNA				
	organism = synthetic construct				
SEQUENCE: 53					
tgaacgatttgcgtt	gggttaccgggt	cttgacactgt	gtttaaacata	aatgtt	44

SEQ ID NO: 54	moltype = DNA length = 44				
FEATURE	Location/Qualifiers				
misc_feature	1..44				
source	note = multidsx?31R-F primer				
	1..44				
	mol_type = other DNA				
	organism = synthetic construct				
SEQUENCE: 54					
agatataatc	ctgaacgcgt	gagtggatga	taaacttcc	gcac	44

SEQ ID NO: 55	moltype = DNA length = 49				
FEATURE	Location/Qualifiers				
misc_feature	1..49				
source	note = multidsx?31R-R primer				
	1..49				
	mol_type = other DNA				
	organism = synthetic construct				
SEQUENCE: 55					
tccacacctac	ccatgggacc	cacgcgtgg	gcgggtcacc	gagatgttc	49

SEQ ID NO: 56	moltype = DNA length = 58				
FEATURE	Location/Qualifiers				
misc_feature	1..58				
source	note = 4050-2U6-T1-F primer				
	1..58				
	mol_type = other DNA				
	organism = synthetic construct				
SEQUENCE: 56					
gagggtctca	tgctgtttaa	cacaggtcaa	gcgggttta	gagctgaaat	58

SEQ ID NO: 57	moltype = DNA length = 56
FEATURE	Location/Qualifiers
misc_feature	1..56
source	note = 4050-2U6-T3-R primer
	1..56
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 57	

-continued

gggggtctca aaacctctga cgggtggtat tgcagcagag agcaactcca tttcat	56
SEQ ID NO: 58	moltype = RNA length = 20
FEATURE	Location/Qualifiers
misc_feature	1..20
	note = guide RNA component
source	1..20
	mol_type = other RNA
	organism = synthetic construct
SEQUENCE: 58	
gtttaacaca ggtcaaggcg	20
SEQ ID NO: 59	moltype = RNA length = 20
FEATURE	Location/Qualifiers
misc_feature	1..20
	note = second guide RNA targeting T2 component
source	1..20
	mol_type = other RNA
	organism = synthetic construct
SEQUENCE: 59	
tctgaacatg tttgatggcg	20
SEQ ID NO: 60	moltype = RNA length = 19
FEATURE	Location/Qualifiers
misc_feature	1..19
	note = second guide RNA targeting T3 component
source	1..19
	mol_type = other RNA
	organism = synthetic construct
SEQUENCE: 60	
gcaataccac ccgtcagag	19
SEQ ID NO: 61	moltype = RNA length = 18
FEATURE	Location/Qualifiers
misc_feature	1..18
	note = second guide RNA targeting T4 component
source	1..18
	mol_type = other RNA
	organism = synthetic construct
SEQUENCE: 61	
gtttatcatc cactctga	18
SEQ ID NO: 62	moltype = DNA length = 30
FEATURE	Location/Qualifiers
misc_feature	1..30
	note = Intron 4 Exon 5 boundary
source	1..30
	mol_type = other DNA
	organism = unidentified
SEQUENCE: 62	
ttatgttaa cacaggtcaa gcgggtgtca	30
SEQ ID NO: 63	moltype = DNA length = 30
FEATURE	Location/Qualifiers
misc_feature	1..30
	note = Intron 4 exon 5 boundary
source	1..30
	mol_type = other DNA
	organism = unidentified
SEQUENCE: 63	
aatacaaatt gtgtccagtt cgccaccagt	30
SEQ ID NO: 64	moltype = DNA length = 54
FEATURE	Location/Qualifiers
misc_feature	1..54
	note = intron 4 exon 5 boundary
source	1..54
	mol_type = other DNA
	organism = unidentified
SEQUENCE: 64	
ccttccatt catttatgtt taacacaggt caagcgggtg tcaacgaata ctca	54
SEQ ID NO: 65	moltype = DNA length = 37
FEATURE	Location/Qualifiers
misc_feature	1..37
	note = intron 4 exon 5 boundary

-continued

```

source          1..37
               mol_type = other DNA
               organism = unidentified
SEQUENCE: 65
gtttaacaca ggtcaagcgg tggtaaacga atactca                                37

SEQ ID NO: 66      moltype = DNA  length = 26
FEATURE
misc_feature       Location/Qualifiers
1..26
note = Intron 4 exon 5 boundary
1..26
source
mol_type = other DNA
organism = unidentified
SEQUENCE: 66
gtttaacaca ggtcaacgaa tactca                                              26

SEQ ID NO: 67      moltype = DNA  length = 33
FEATURE
misc_feature       Location/Qualifiers
1..33
note = intron 4 exon 5 boundary
1..33
source
mol_type = other DNA
organism = unidentified
SEQUENCE: 67
gtttaacaca ggtcggtggt caacgaatac tca                                33

SEQ ID NO: 68      moltype = DNA  length = 28
FEATURE
misc_feature       Location/Qualifiers
1..28
note = intron 4 exon 5 boundary
1..28
source
mol_type = other DNA
organism = unidentified
SEQUENCE: 68
gtttaaacacg gtggtaacgc aatactca                                              28

SEQ ID NO: 69      moltype = DNA  length = 26
FEATURE
misc_feature       Location/Qualifiers
1..26
note = intron 4 exon 5 boundary
1..26
source
mol_type = other DNA
organism = unidentified
SEQUENCE: 69
gtttaacggt ggtcaacgaa tactca                                              26

SEQ ID NO: 70      moltype = DNA  length = 36
FEATURE
misc_feature       Location/Qualifiers
1..36
note = intron 4 exon 4 boundary
1..36
source
mol_type = other DNA
organism = unidentified
SEQUENCE: 70
gtttaacaca ggtcaacggt ggtcaacgaa tactca                                36

SEQ ID NO: 71      moltype = DNA  length = 34
FEATURE
misc_feature       Location/Qualifiers
1..34
note = intron 4 exon 5 boundary
1..34
source
mol_type = other DNA
organism = unidentified
SEQUENCE: 71
gtttaacaca ggtccgggtgg tcaacgaata ctca                                34

SEQ ID NO: 72      moltype = DNA  length = 29
FEATURE
misc_feature       Location/Qualifiers
1..29
note = intron 4 exon 5 boundary
1..29
source
mol_type = other DNA
organism = unidentified
SEQUENCE: 72
gtttaacacc ggtggtaac gaatactca                                              29

```

-continued

```

SEQ ID NO: 73      moltype = DNA length = 27
FEATURE          Location/Qualifiers
misc_feature     1..27
note = intron 4 exon 5 boundary
source           1..27
mol_type = other DNA
organism = unidentified

SEQUENCE: 73
gtttaaccgg tggtaaacgca atactca                                27

SEQ ID NO: 74      moltype = DNA length = 39
FEATURE          Location/Qualifiers
misc_feature     1..39
note = intron 4 exon 5 boundary
source           1..39
mol_type = other DNA
organism = unidentified

SEQUENCE: 74
gtttaacaca ggtcataaagc ggtggtaac gaatactca                                39

SEQ ID NO: 75      moltype = DNA length = 39
FEATURE          Location/Qualifiers
misc_feature     1..39
note = Intron 4 exon 5 boundary
source           1..39
mol_type = other DNA
organism = unidentified

SEQUENCE: 75
gtttaacaca ggtcaaggac ggtggtaac gaatactca                                39

SEQ ID NO: 76      moltype = DNA length = 129
FEATURE          Location/Qualifiers
misc_feature     1..129
note = Intron 4 exon 5 boundary
source           1..129
mol_type = other DNA
organism = unidentified

SEQUENCE: 76
cctttccatt catttatgtt taacacaggt caagcggtgg tcaacgaata ctcacgattg  60
cataatctga acatgtttga tggcggtggag ttgcgcataa ccacccgtca gagtggatga 120
taaactttc                                              129

SEQ ID NO: 77      moltype = DNA length = 129
FEATURE          Location/Qualifiers
misc_feature     1..129
note = intron 4 exon 5 boundary
source           1..129
mol_type = other DNA
organism = unidentified

SEQUENCE: 77
cctttccatt catttatgtt taacacaggt caagcggtgg tcaacgaata ctcacgattg  60
cataatctga acatgtttga tggcggtggag ttgcgcataa ccacccgtca gagtggatga 120
taaactttc                                              129

SEQ ID NO: 78      moltype = DNA length = 129
FEATURE          Location/Qualifiers
misc_feature     1..129
note = intron 4 exon 5 boundary
source           1..129
mol_type = other DNA
organism = unidentified

SEQUENCE: 78
cctttccatt catttatgtt taacacaggt caagcggtgg tcaacgaata ctcacgattg  60
cataatctga acatgtttga tggcggtggag ttgcgcataa ccacccgtca gagtggatga 120
taaactttc                                              129

SEQ ID NO: 79      moltype = DNA length = 129
FEATURE          Location/Qualifiers
misc_feature     1..129
note = intron 4 exon 5 boundary
source           1..129
mol_type = other DNA
organism = unidentified

SEQUENCE: 79
cctttccatt catttatgtt taacacaggt caagcggtgg tcaacgaata ctcacgattg  60
cataatctga acatgtttga tggcggtggag ttgcgcataa ccacccgtca gagtggatga 120

```

-continued

taaaacttcc	129
SEQ ID NO: 80	moltype = DNA length = 129
FEATURE	Location/Qualifiers
misc_feature	1..129
	note = intron 4 exon 5 boundary
source	1..129
	mol_type = other DNA
	organism = unidentified
SEQUENCE: 80	
cctttcatt catttatgtt	taacacaggtaaagcggtgg tcaacgaata ctcacgattg 60
cataatctga acatgttgc	tggcgtggag ttgcgcaata ccacccgtca gagtggatga 120
taaaacttcc	129
SEQ ID NO: 81	moltype = DNA length = 129
FEATURE	Location/Qualifiers
misc_feature	1..129
	note = SEQ ID No: 82
source	1..129
	mol_type = other DNA
	organism = unidentified
SEQUENCE: 81	
cctttcatt catttatgtt	taacacaggtaaagcggtgg tcaacgaata ctcacgattg 60
cataatctga acatgttgc	tggcgtggag ttgcgcaata ccacccgtca gagtggatga 120
taaaacttcc	129
SEQ ID NO: 82	moltype = DNA length = 129
FEATURE	Location/Qualifiers
misc_feature	1..129
	note = intron 4 exon 5 boundary
source	1..129
	mol_type = other DNA
	organism = unidentified
SEQUENCE: 82	
cctttcatt catttatgtt	taacacaggtaaagcggtgg tcaacgaata ctcacgattg 60
cataatctga acatgttgc	tggcgtggag ttgcgcaata ccacccgtca gagtggatga 120
taaaacttcc	129
SEQ ID NO: 83	moltype = DNA length = 129
FEATURE	Location/Qualifiers
misc_feature	1..129
	note = intron 4 exon 5 boundary
source	1..129
	mol_type = other DNA
	organism = unidentified
SEQUENCE: 83	
cctttcatt catttatgtt	caacacaggtaaagcggtgg tcaacgaata ctcacgattg 60
cataatctga acatgttgc	tggcgtggag ttgcgcaata ccacccgtca gagtggatga 120
taaaacttcc	129
SEQ ID NO: 84	moltype = DNA length = 129
FEATURE	Location/Qualifiers
misc_feature	1..129
	note = intron 4 exon 5 boundary
source	1..129
	mol_type = other DNA
	organism = unidentified
SEQUENCE: 84	
cctttcatt catttatgtt	caacacaggtaaagcggtgg tcaacgaata ctcacgattg 60
cataatctga acatgttgc	tggcgtggag ttacgcataa ccacccgtca gagtggatga 120
taaaacttcc	129
SEQ ID NO: 85	moltype = DNA length = 128
FEATURE	Location/Qualifiers
misc_feature	1..128
	note = intron 4 exon 5 boundary
source	1..128
	mol_type = other DNA
	organism = unidentified
SEQUENCE: 85	
cctttcatt catttatgtt	caacacaggtaaagcggtgg tcaacgaata ctcacgattg 60
cataatctga acatgttgc	tggcgtggag ttacgcataa ccacccgtca gagtggatga 120
taaaacttcc	128
SEQ ID NO: 86	moltype = DNA length = 129
FEATURE	Location/Qualifiers

-continued

```

misc_feature          1..129
                      note = Intron 4 exon 5 boundary
source               1..129
                      mol_type = other DNA
                      organism = unidentified
SEQUENCE: 86
cctttccatt catttatgtt caacacaggt caagcggtgg tcaacgaata ctcacgattg  60
cataatctga acatgttcga tggcggtggag ttacgcata ccacccgtca gagtggatga 120
taaacttcc          129

SEQ ID NO: 87          moltype = DNA length = 129
FEATURE
misc_feature          1..129
                      note = intron 4 exon 5 boundary
source               1..129
                      mol_type = other DNA
                      organism = unidentified
SEQUENCE: 87
cctttccatt catttatgtt caacacaggt caagcggtgg tcaacgaata ctcacgattg  60
cataatctga acatgttcga tggcggtggag ttacgcata ccacccgtca gagtggatga 120
taaacttcc          129

SEQ ID NO: 88          moltype = DNA length = 129
FEATURE
misc_feature          1..129
                      note = intron 4 exon 5 boundary
source               1..129
                      mol_type = other DNA
                      organism = unidentified
SEQUENCE: 88
ccttaccatg catttatgtt caacacaggt caagcggtgg tcaacgaata ctcacgattg  60
cataatctga acatgttcga tggcggtggag ttacgcaca ccacccgtca gagtggatga 120
taaacttcc          129

SEQ ID NO: 89          moltype = DNA length = 129
FEATURE
misc_feature          1..129
                      note = intron 4 exon 5 boundary
source               1..129
                      mol_type = other DNA
                      organism = unidentified
SEQUENCE: 89
cctttccatt catttatgtt caacacaggt caagcggtgg tcaacgaata ctcacgattg  60
cataatctga acatgttcga tggcggtggag ttacgcata ccacccgtca gagtggatga 120
taaacttcc          129

SEQ ID NO: 90          moltype = DNA length = 129
FEATURE
misc_feature          1..129
                      note = intron 4 exon 5 boundary
source               1..129
                      mol_type = other DNA
                      organism = unidentified
SEQUENCE: 90
cctttccatt catttatgtt caacacaggt caagcggtgg tcaacgaata ctcacgattg  60
cataatctga acatgttcga tggcggtggag ttacgcata ccacccgtca gagtggatga 120
taaacttcc          129

SEQ ID NO: 91          moltype = DNA length = 129
FEATURE
misc_feature          1..129
                      note = intron 4 exon 5 boundary
source               1..129
                      mol_type = other DNA
                      organism = unidentified
SEQUENCE: 91
cctttccatt catttatgtt caacacaggt caggcggtgg tcaacgaata ctcacgattg  60
cacaatctga acatgttcga tggcggtggag ttacgcata ccacccgtca gagtggatga 120
taaacttcc          129

SEQ ID NO: 92          moltype = DNA length = 129
FEATURE
misc_feature          1..129
                      note = intron 4 exon 5 boundary
source               1..129
                      mol_type = other DNA

```

-continued

```

SEQUENCE: 92
cctttccatt catttatgct caacacaggt caggccgtgg tcaacgaata ctcacgattg 60
cacaatctga acatgttcga tggcgtagag ttgcgcaca ccacccgtca gagtggatga 120
taaaacttcc 129

SEQ ID NO: 93      moltype = DNA length = 129
FEATURE
misc_feature       Location/Qualifiers
1..129
note = intron 4 exon 5 boundary
source             1..129
mol_type = other DNA
organism = unidentified

SEQUENCE: 93
cttgcatt tatttatgcc caacacaggt caggccgtgg tcaacgaata ctcacgattg 60
cacaatctga acatgttcga tggcgtagag ttgcgcaca ccacccgtca gagtggatga 120
taaaacttcc 129

SEQ ID NO: 94      moltype = DNA length = 129
FEATURE
misc_feature       Location/Qualifiers
1..129
note = intron 4 exon 5 boundary
source             1..129
mol_type = other DNA
organism = unidentified

SEQUENCE: 94
cctttccatt catttatgtt taacacaggt caagcagtgg tcaacgaata ttcacgattg 60
cataatctga acatgtttga tggcgtagag ttgcgcaca ccacccgtca gagtggatga 120
taaaacttcc 129

```

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 to 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).

* * * * *