US 20250255283A1

(19) United States

(12) Patent Application Publication

(10) Pub. No.: US 2025/0255283 A1

(43) Pub. Date: Aug. 14, 2025

Crisanti et al.

(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)

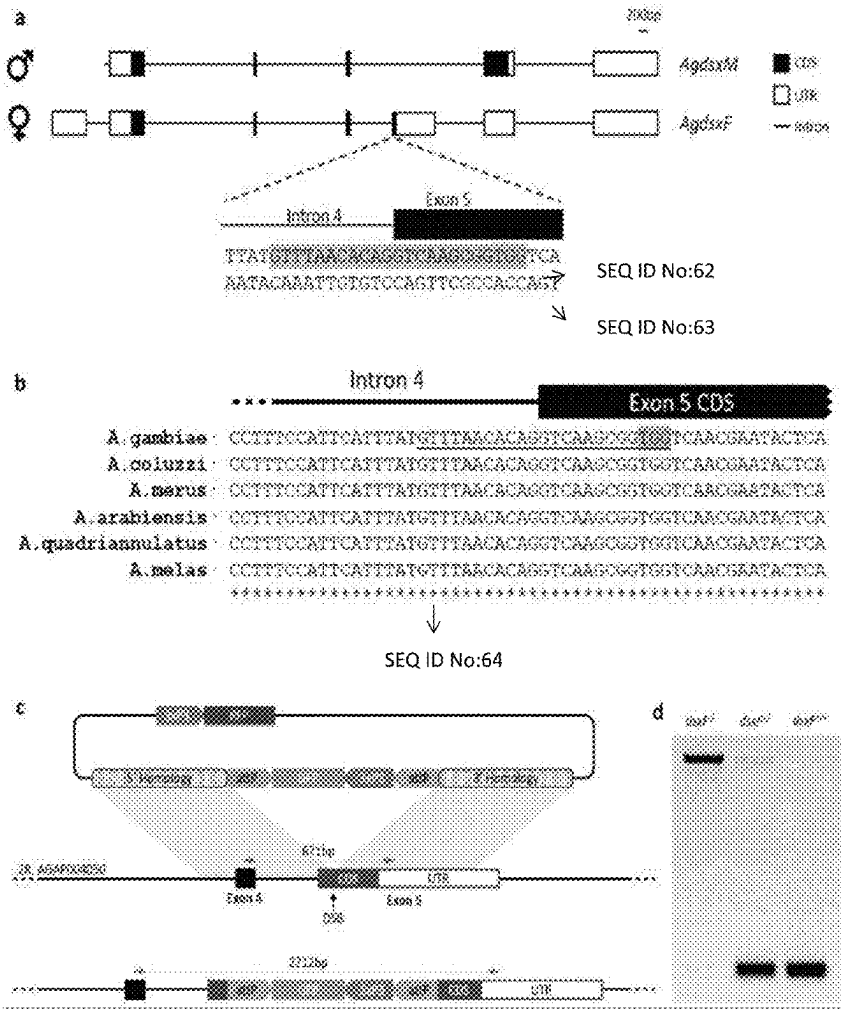
(52) **U.S. Cl.**

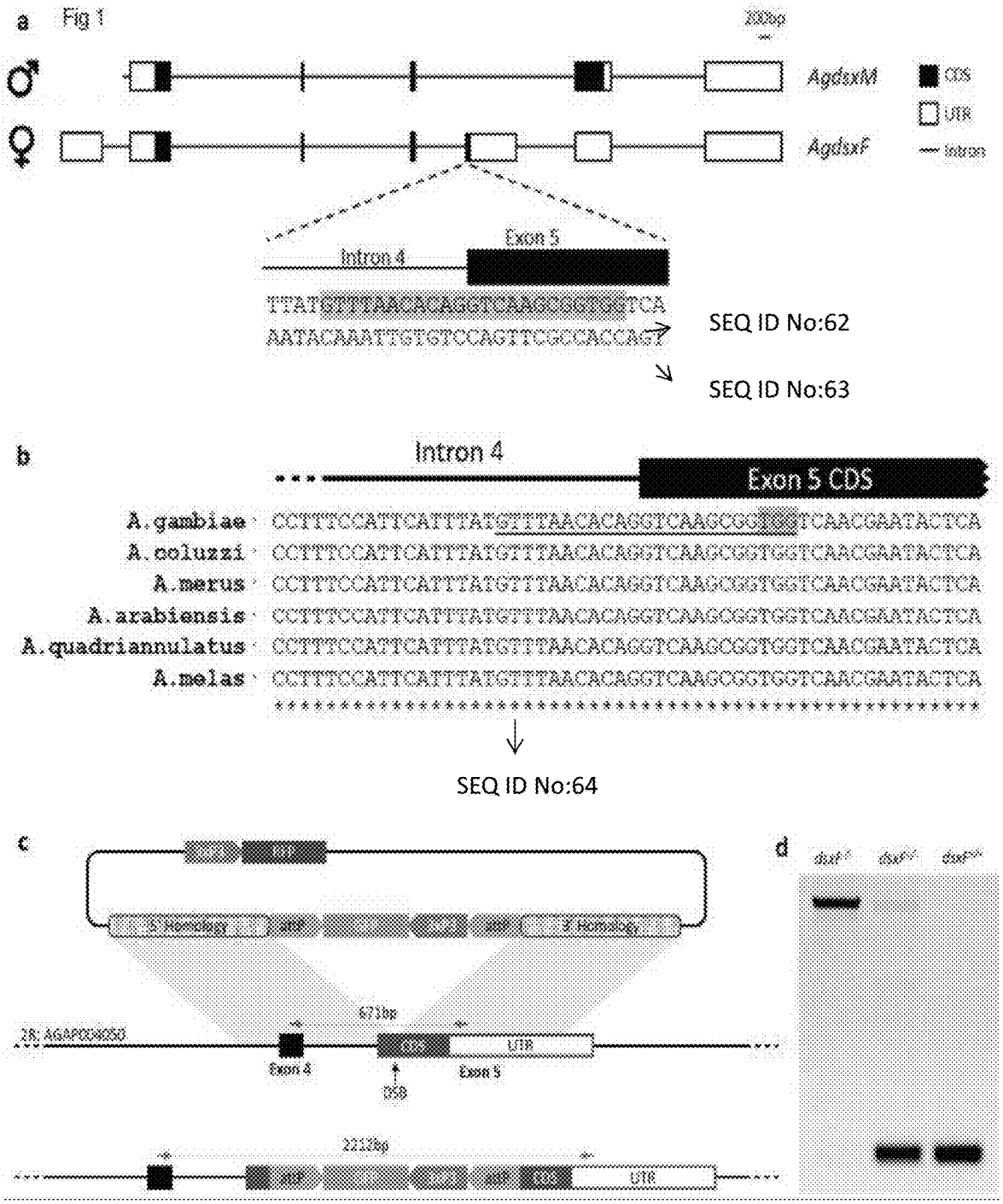
CPC ***A01K 67/68*** (2025.01); ***C12N 15/8509*** (2013.01); *A01K 2217/07* (2013.01); *A01K 2227/706* (2013.01); *A01K 2267/02* (2013.01)

(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
AgdsxM
AgdsxF
CDS
UTR
Intron
Exon 5
Intron 4
SEQ ID No:62
SEQ ID No:63
b
Intron 4
Exon 5 CDS
A.gambiae
A.coluzzi
A.merus
A.arabiensis
A.quadriannulatus
A.melas
SEQ ID No:64
c
d

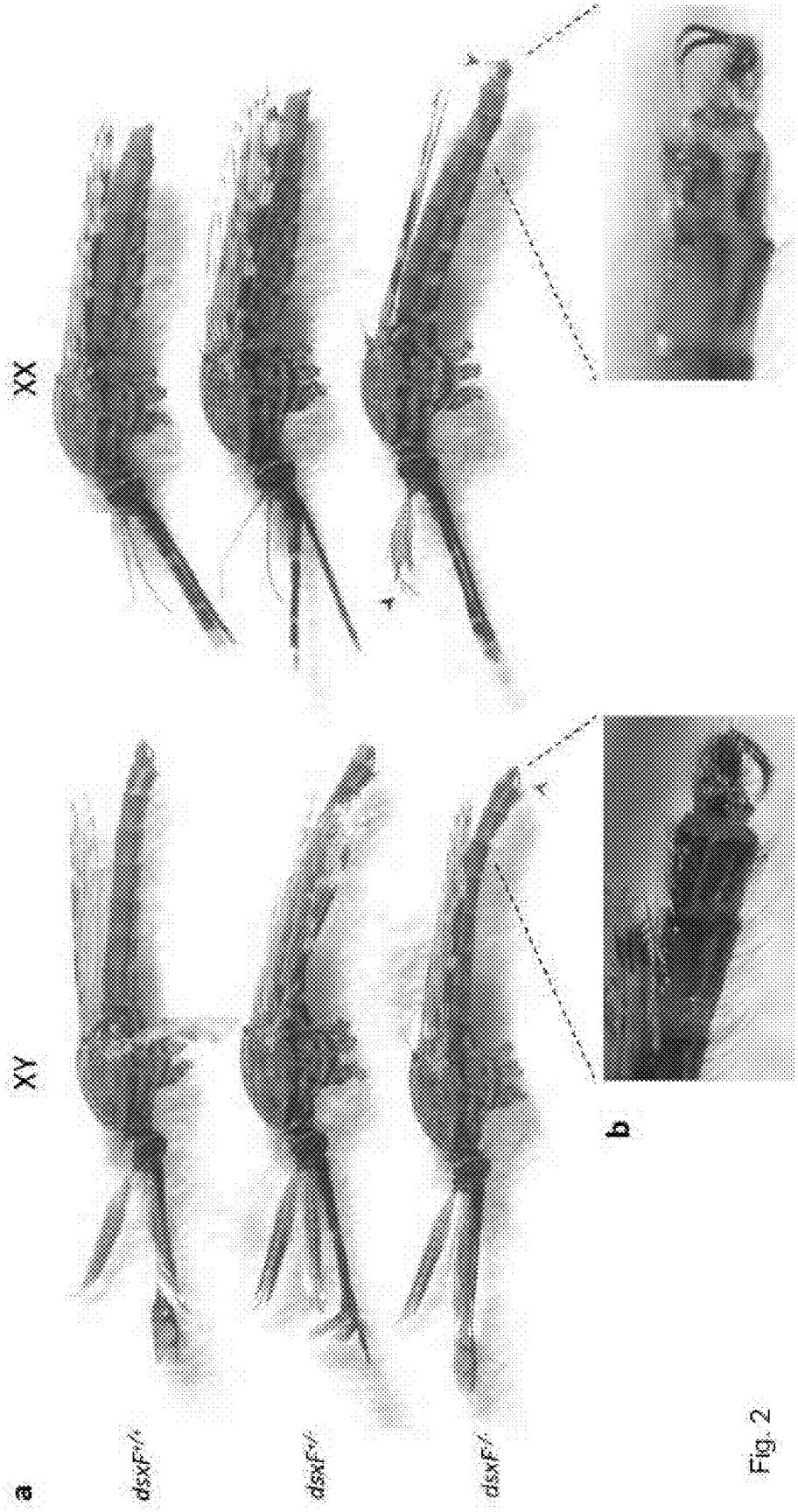


Fig. 2

Figure 3

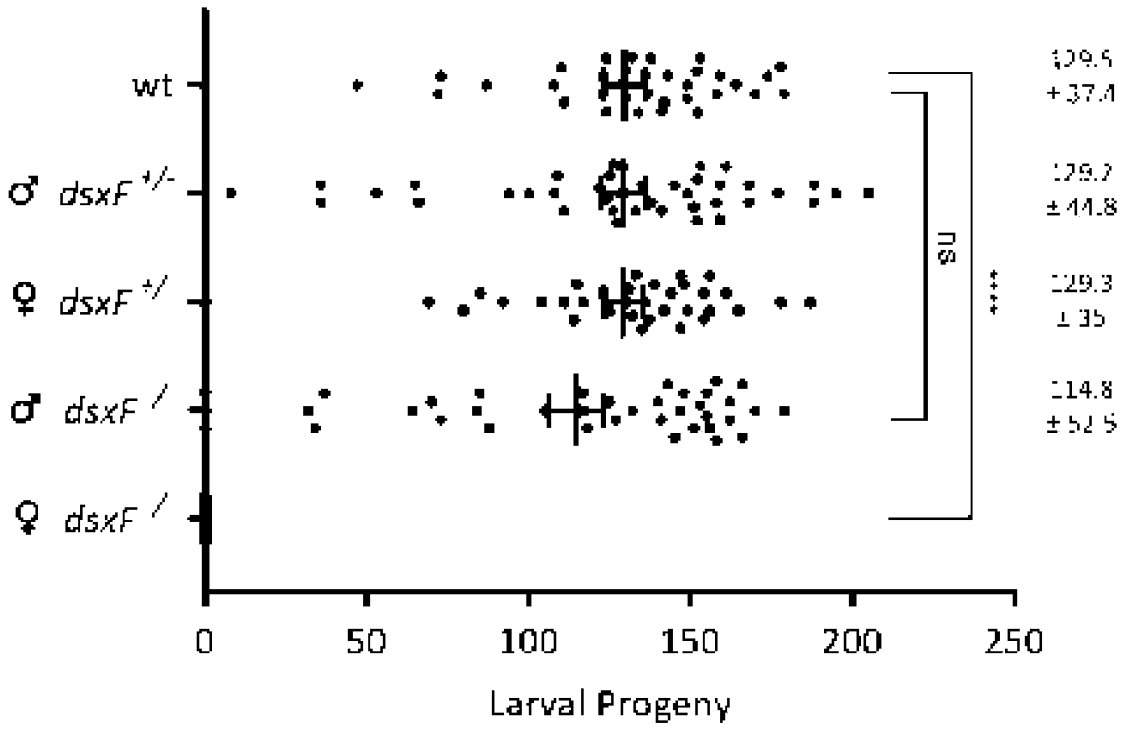


Figure 4

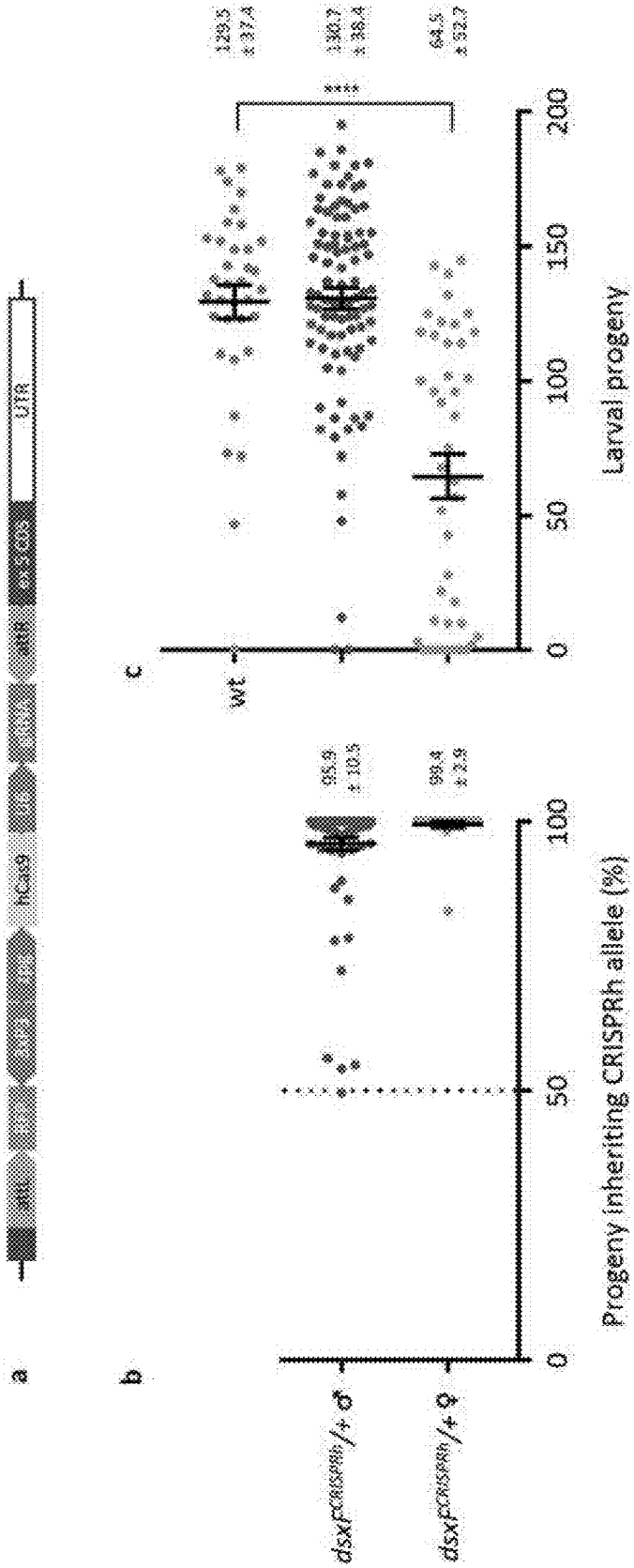
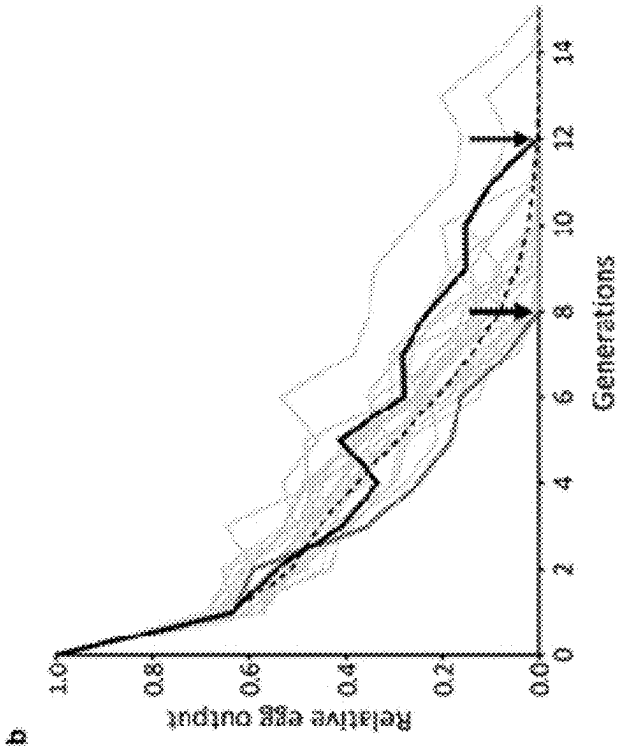

a
b
c
hCas9
UTR
dsx^CRISPRh/+ ♂
dsx^CRISPRh/+ ♀
95.9 ± 10.5
99.4 ± 2.9
wt
129.5 ± 37.4
130.7 ± 38.4
64.5 ± 32.7
0
50
100
150
200
Progeny inheriting CRISPRh allele (%)
Larval progeny

Figure 5


b
Relative egg output
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
12
14
Generations

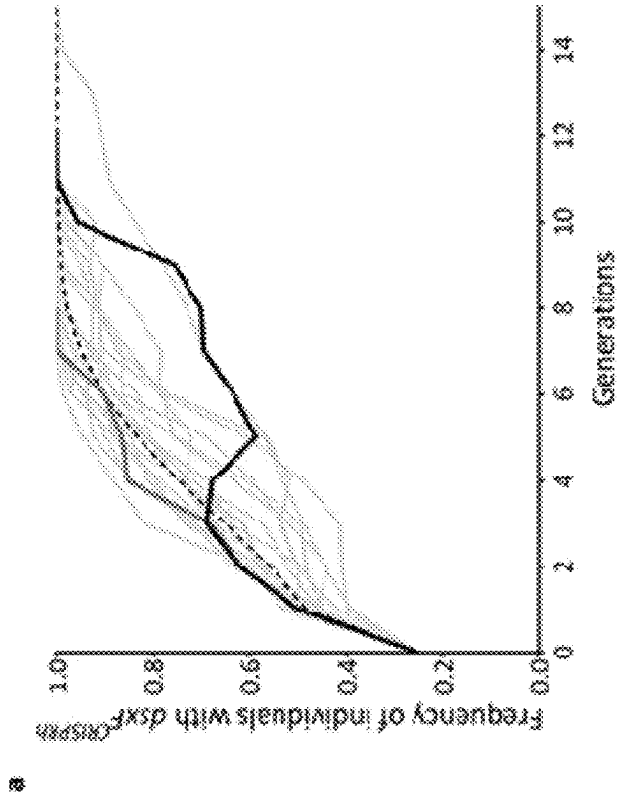

a
Frequency of individuals with dsxF^CRISPR
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
12
14
Generations

Figure 6

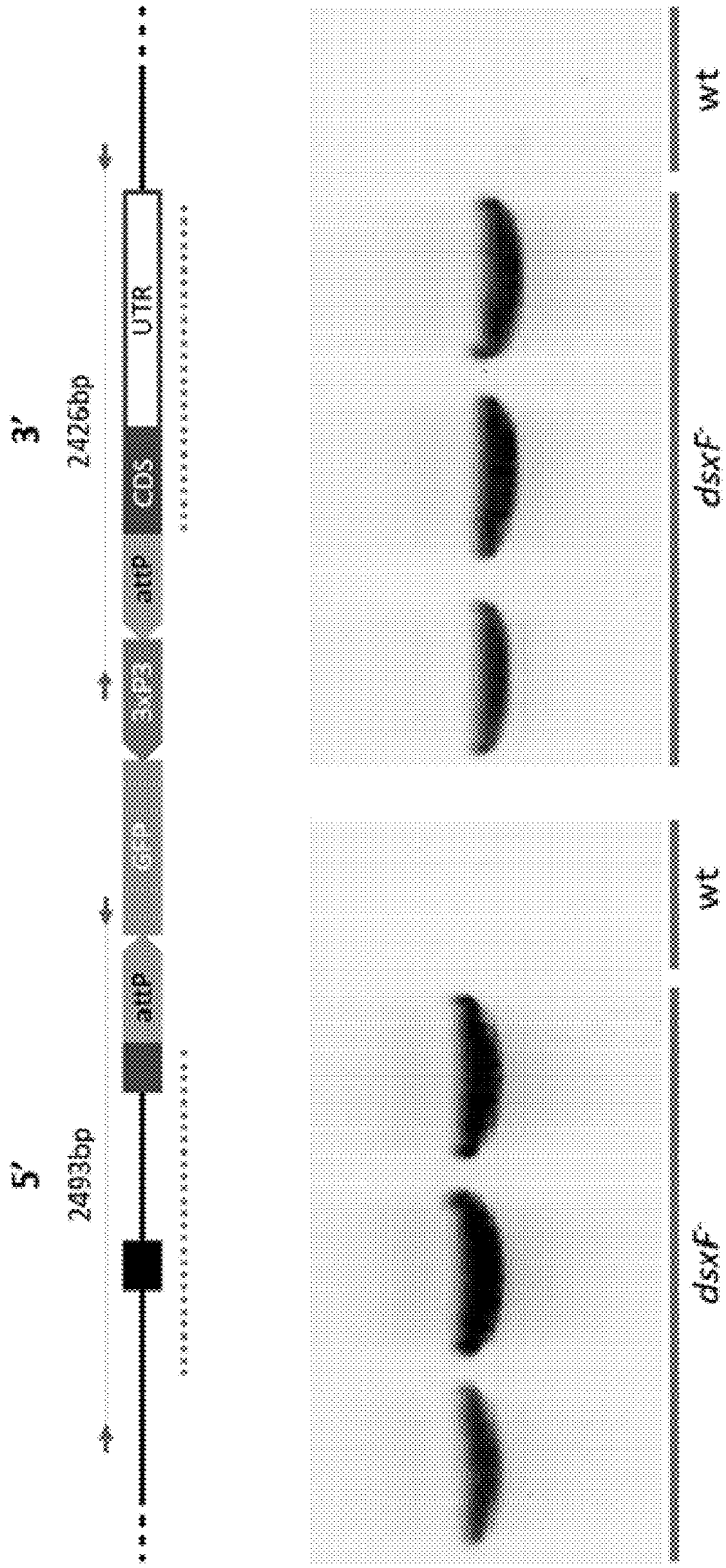


Figure 7

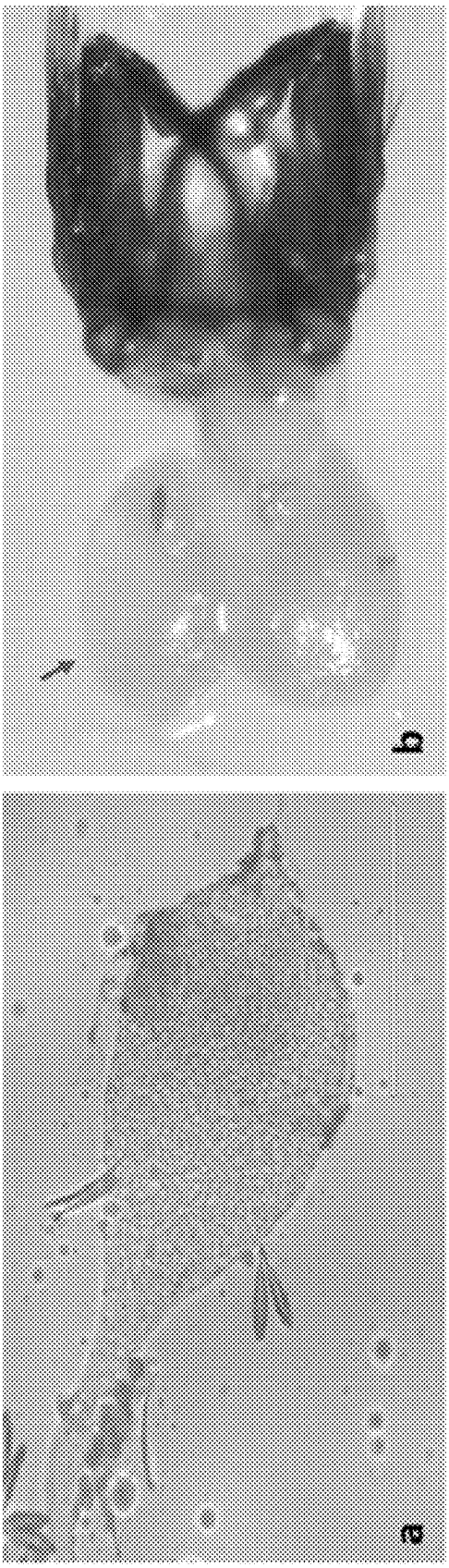


Figure 8

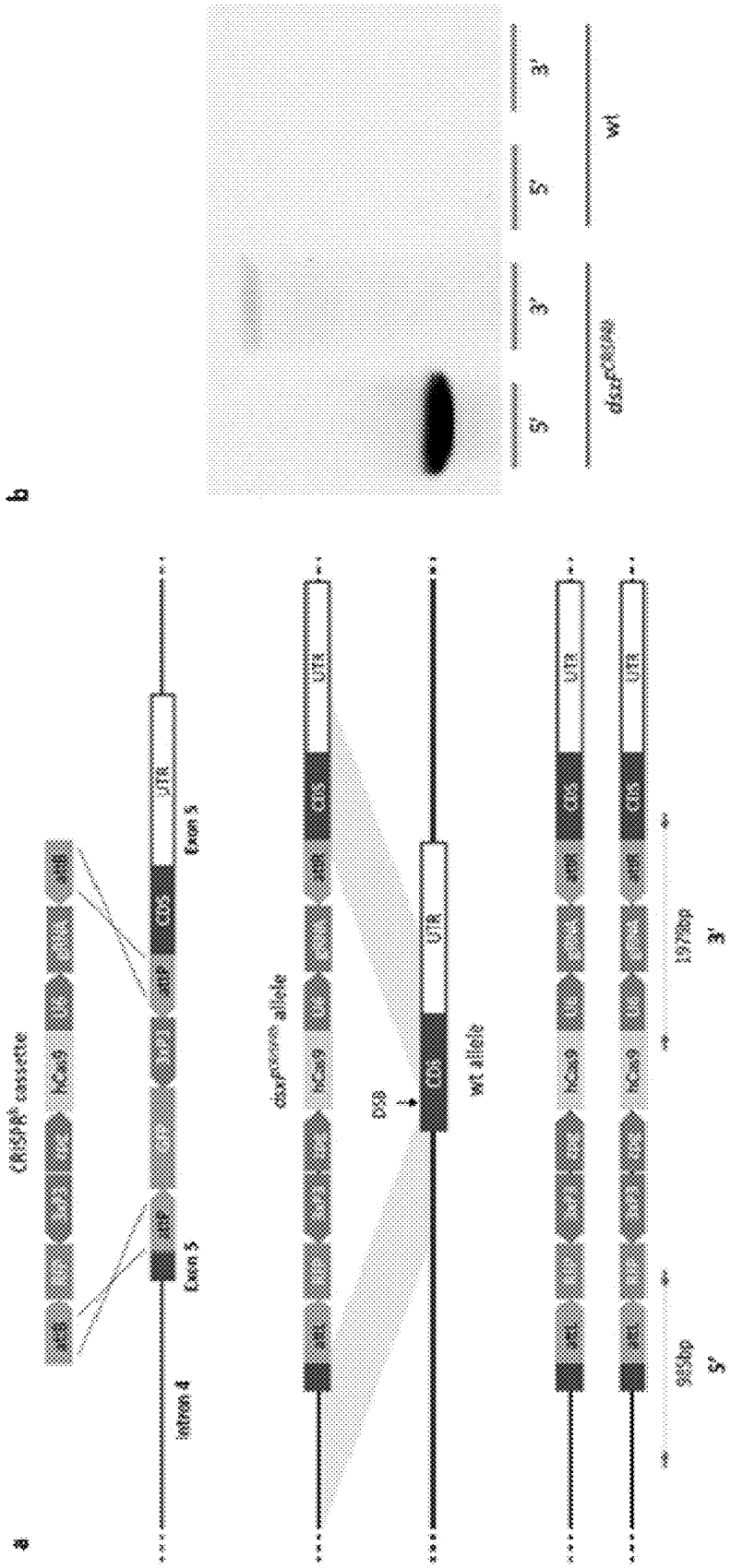

a
CRISPR cassette
intron 4
Exon 5
Exon 5
UTR
CDS
DSB
wt allele
5'
3'
5'
3'
wt
b

Figure 9

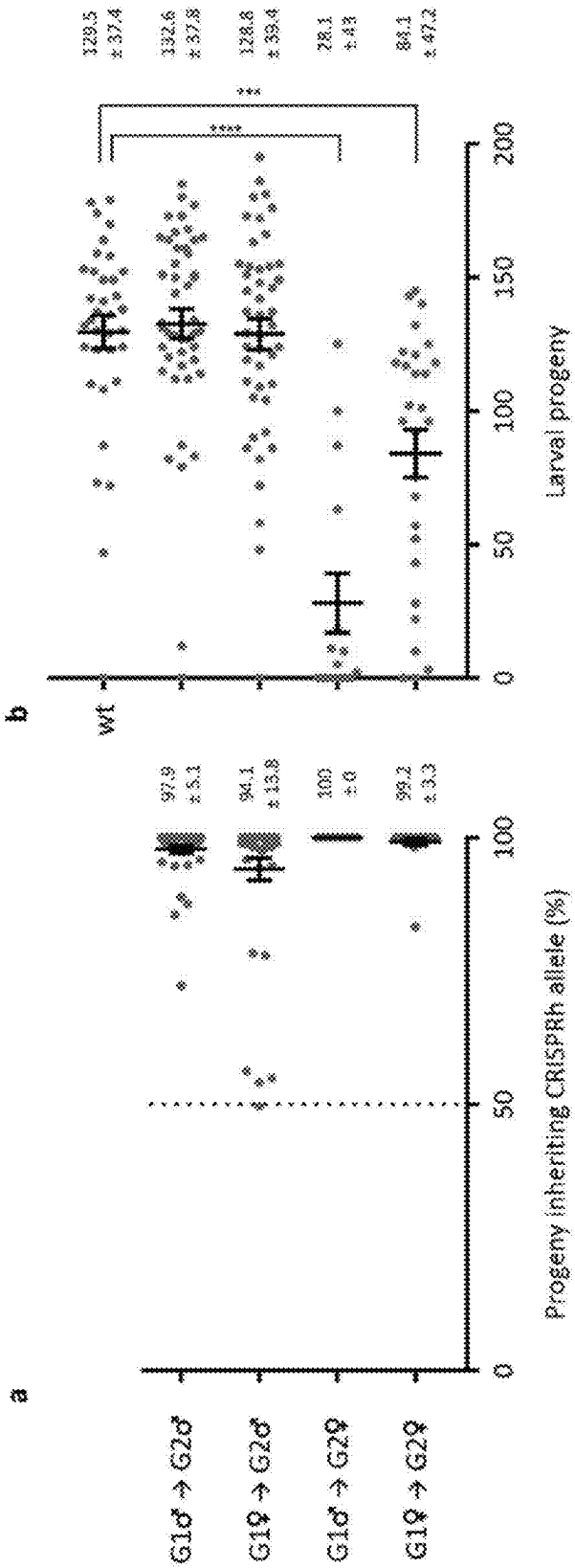

a
G1♂ → G2♂
G1♀ → G2♂
G1♂ → G2♀
G1♀ → G2♀
97.9 ± 5.1
94.1 ± 13.8
100 ± 0
99.2 ± 3.3
0
50
100
Progeny inheriting CRISPRh allele (%)
b
wt
329.5 ± 37.4
332.6 ± 37.8
328.8 ± 39.4
28.1 ± 43
84.1 ± 47.2
***
0
50
100
150
200
Larval progeny

Figure 10A

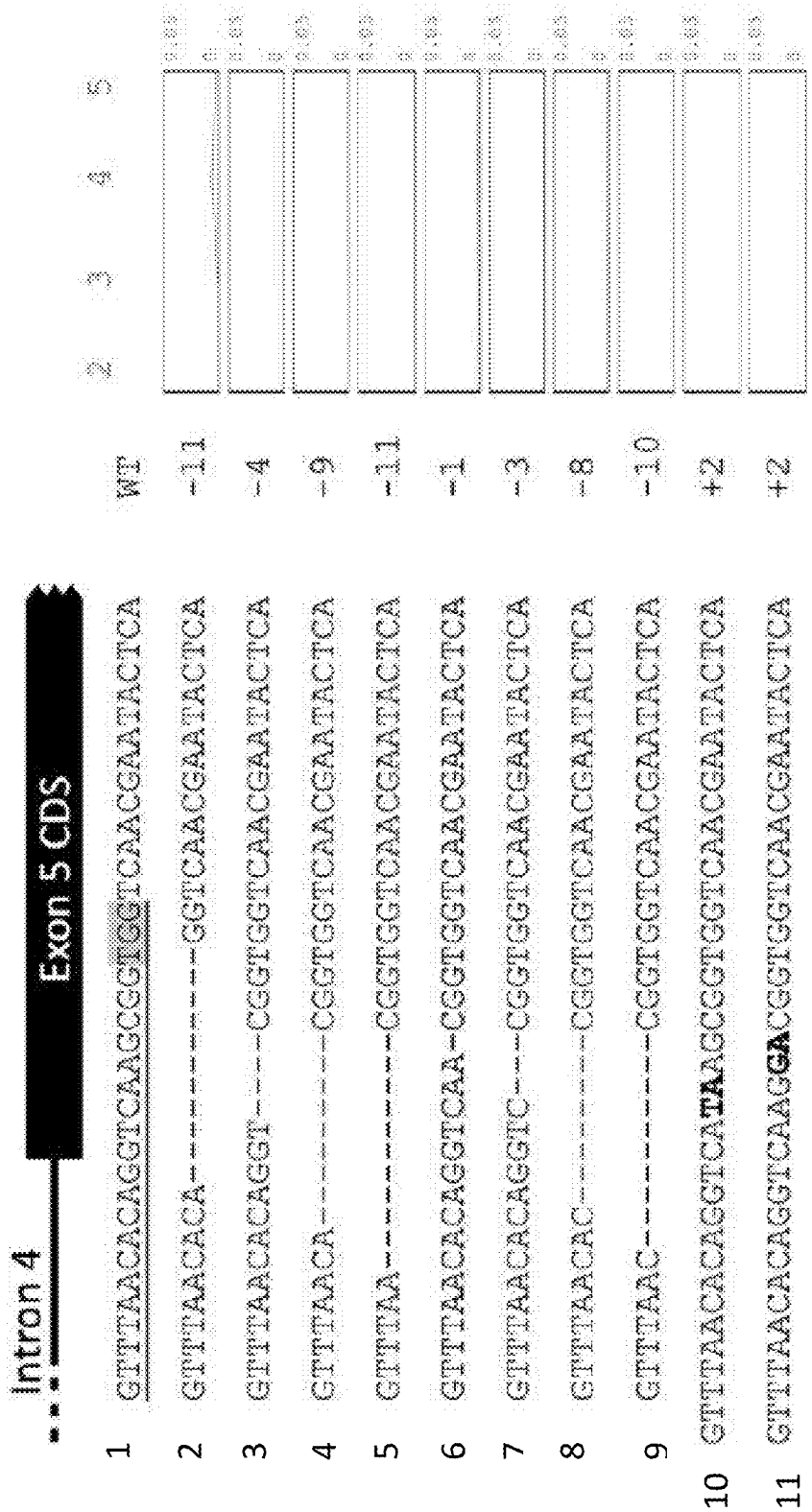

Intron 4
Exon 5 CDS
1 GTTTAACACAGGTCAAGCGGTCGGTCAACGAATACTCA
2 GTTTAACACA----------GGTCAACGAATACTCA
3 GTTTAACACAGGT----CGGTGGTCAACGAATACTCA
4 GTTTAACA----------CGGTGGTCAACGAATACTCA
5 GTTTAA----------CGGTGGTCAACGAATACTCA
6 GTTTAACACAGGTCAA-CGGTGGTCAACGAATACTCA
7 GTTTAACACAGGTC---CGGTGGTCAACGAATACTCA
8 GTTTAACAC----------CGGTGGTCAACGAATACTCA
9 GTTTAAC----------CGGTGGTCAACGAATACTCA
10 GTTTAACACAGGTCATAAGCGGTGGTCAACGAATACTCA
11 GTTTAACACAGGTCAAGGACGGTGGTCAACGAATACTCA
WT
-11
-4
-9
-11
-1
-3
-8
-10
+2
+2
2
3
4
5

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

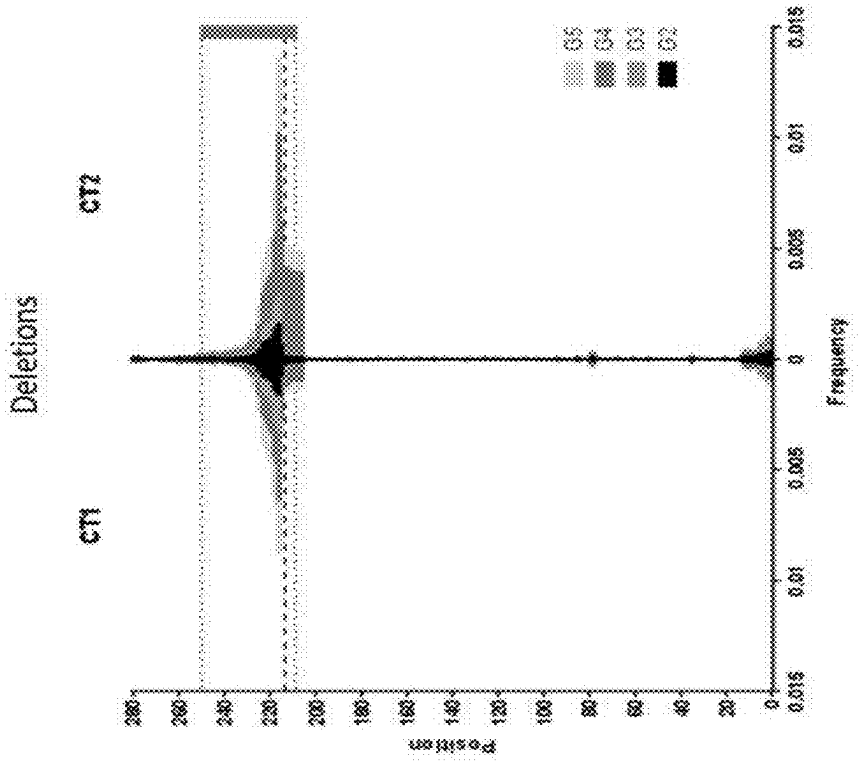
8- SEQ ID No: 72

9- SEQ ID No: 73

10- SEQ ID No: 74

11- SEQ ID No: 75

Figure 10B


Deletions
CT1
CT2
Position
Frequency
C5
C4
C3
C2

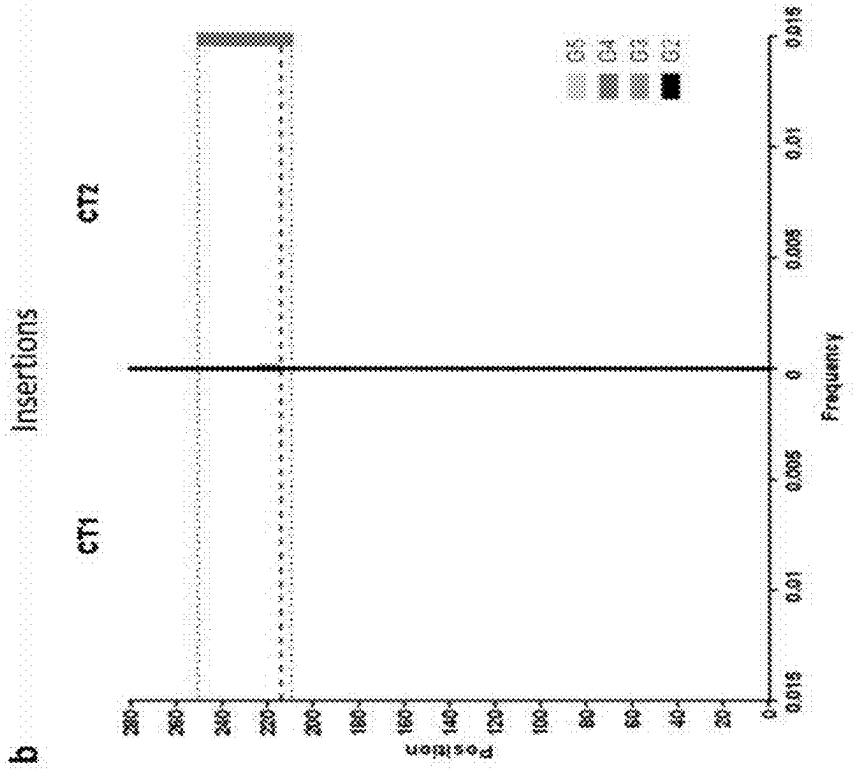

b
Insertions
CT1
CT2
Position
Frequency
C5
C4
C3
C2

Figure 10C

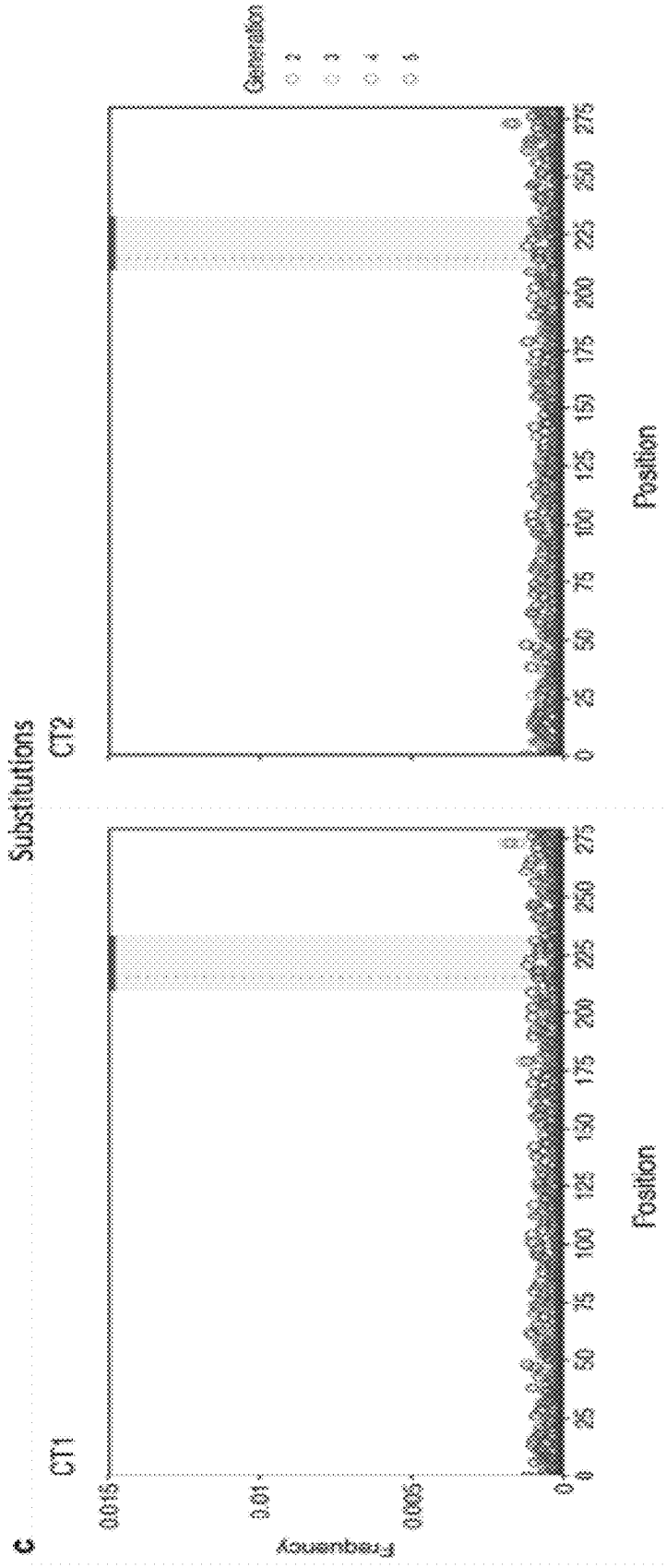
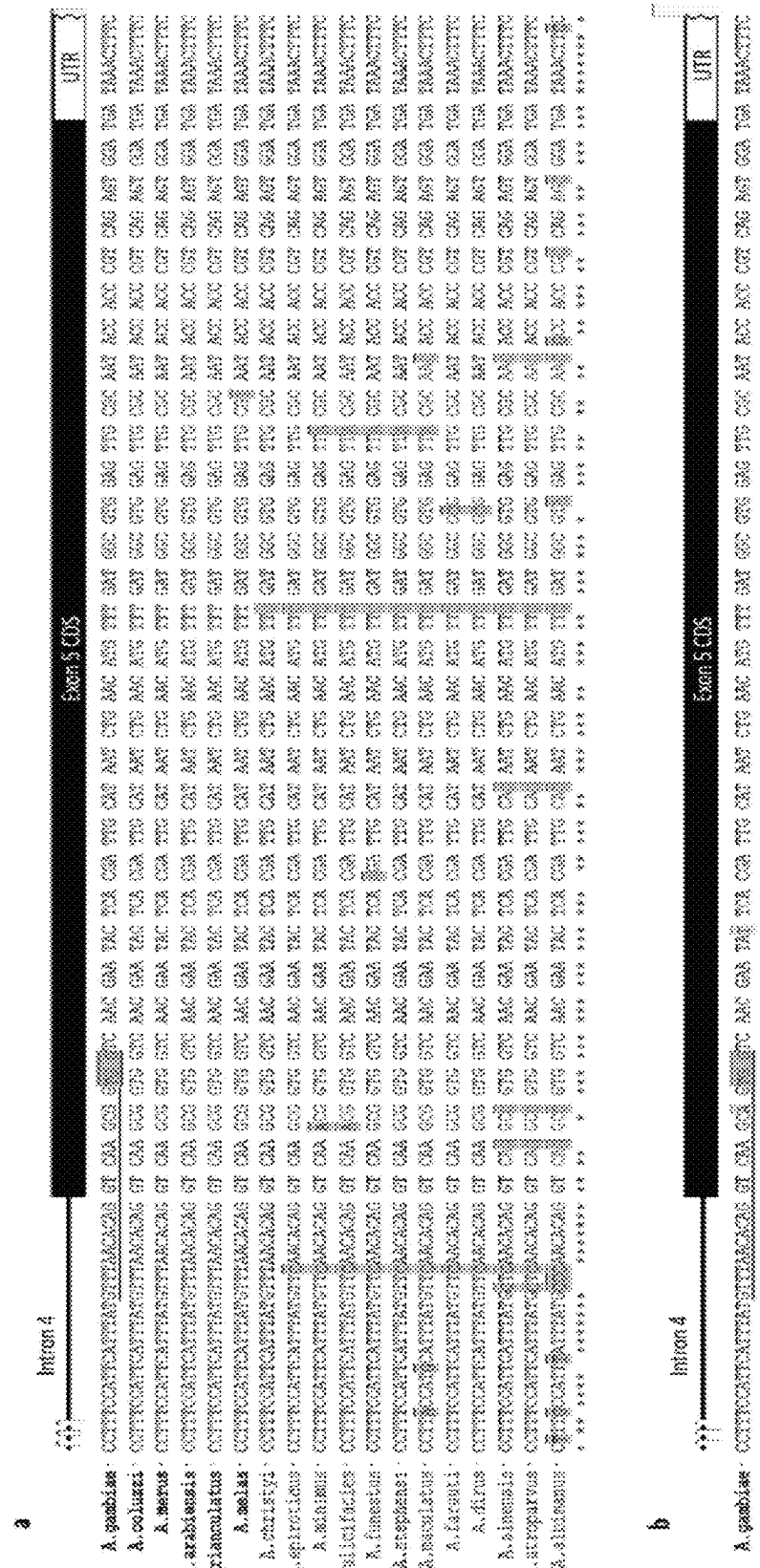

C
Substitutions
CT1
CT2
Frequency
0.015
0.01
0.005
0
Position
0
25
50
75
100
125
150
175
200
225
250
275
Generation
2
3
4
5

Figure 11



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

Figure 12

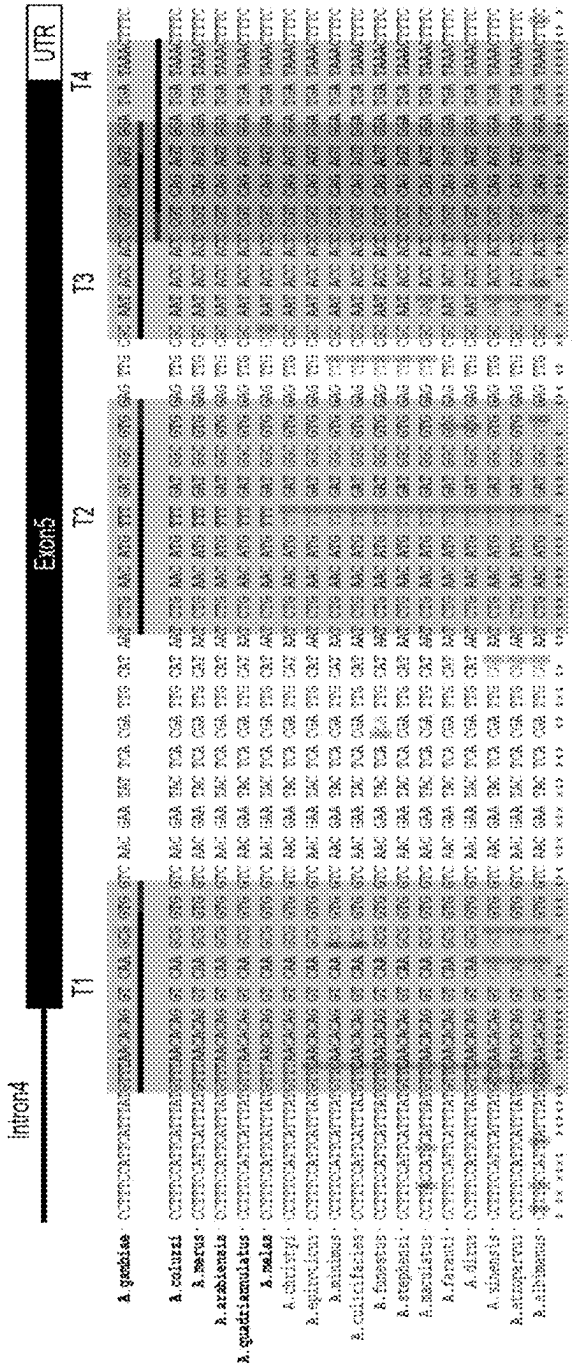

intron4
Exon5
UTR
T1
T2
T3
T4
A. gambiae
A. coluzzii
A. merus
A. arabiensis
A. quadriannulatus
A. melas
A. christyi
A. epiroticus
A. minimus
A. culicifacies
A. funestus
A. stephensi
A. maculatus
A. farauti
A. dirus
A. sinensis
A. atroparvus
A. albimanus

Figure 13

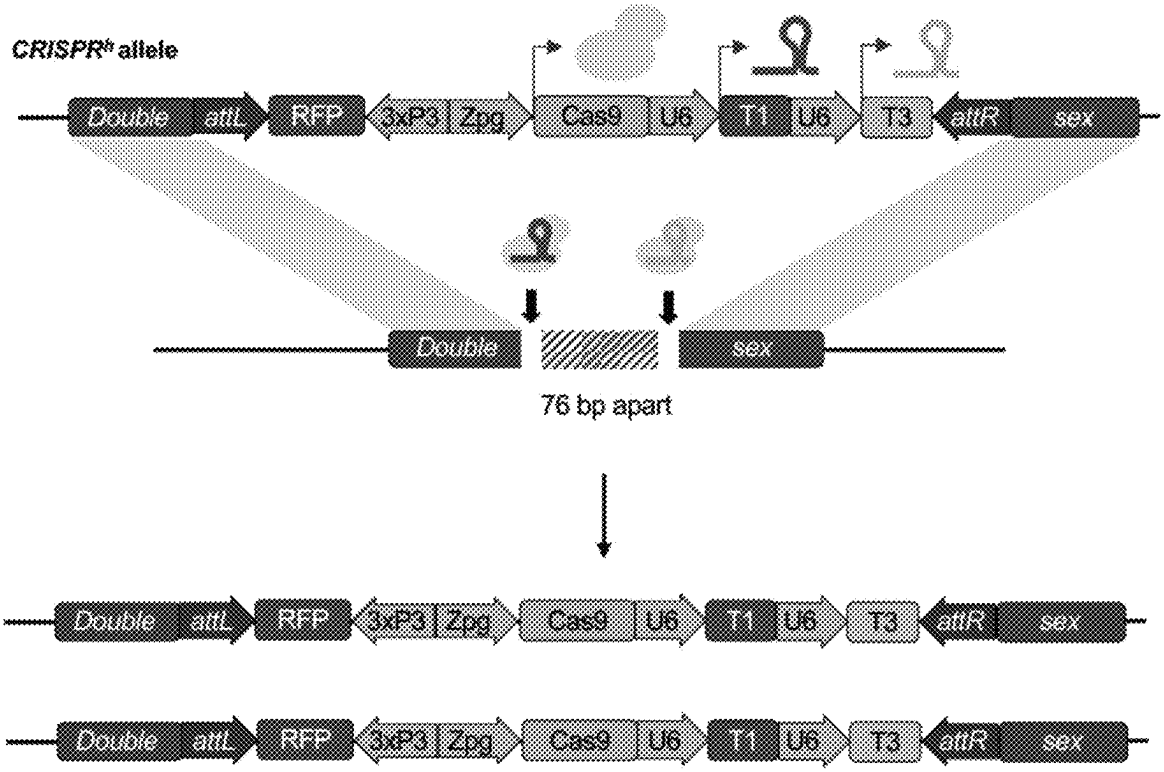

CRISPR^h allele
Double
attL
RFP
3xP3
Zpg
Cas9
U6
T1
U6
T3
attR
sex
Double
sex
76 bp apart
Double
attL
RFP
3xP3
Zpg
Cas9
U6
T1
U6
T3
attR
sex
Double
attL
RFP
3xP3
Zpg
Cas9
U6
T1
U6
T3
attR
sex

Figure 14

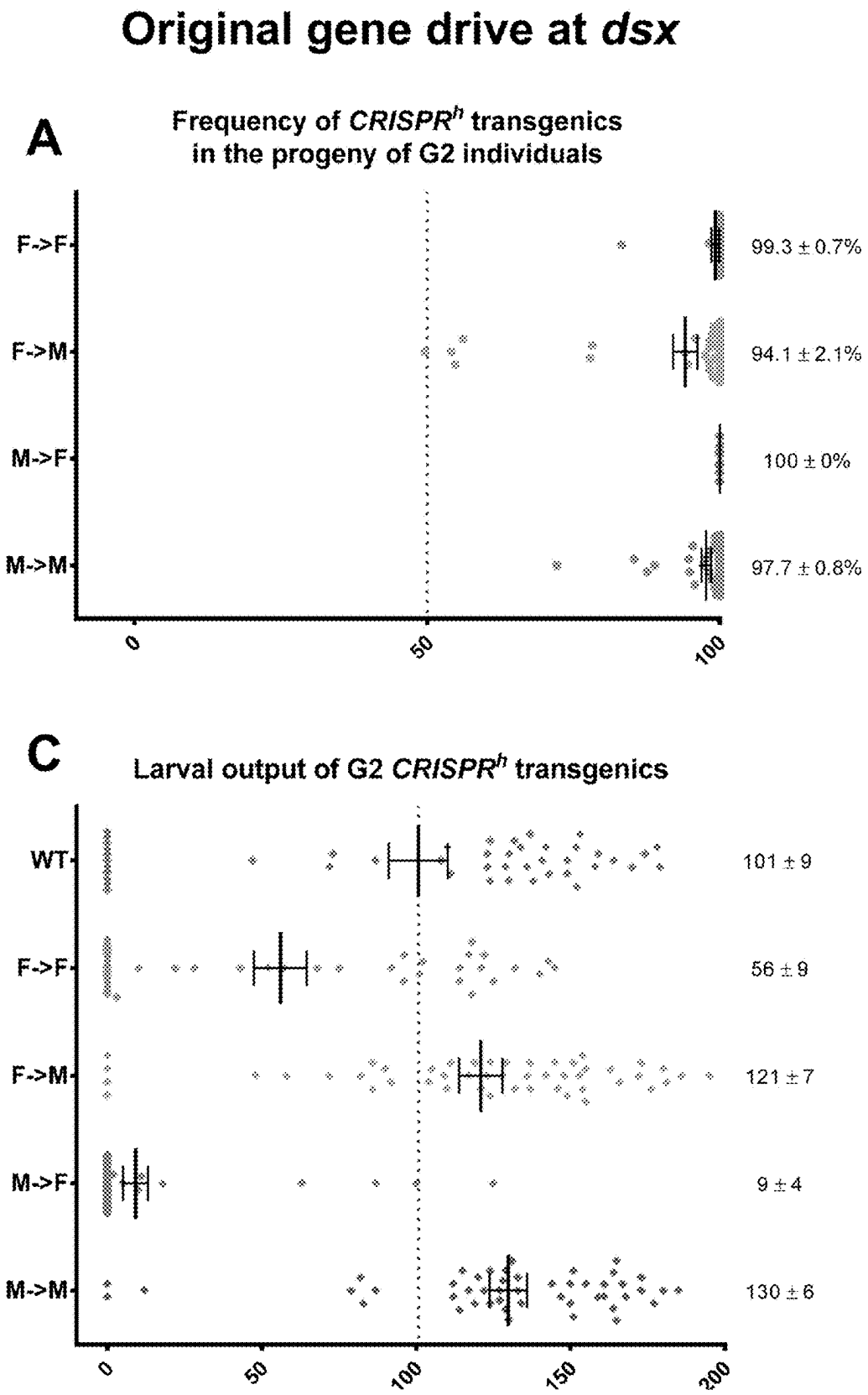
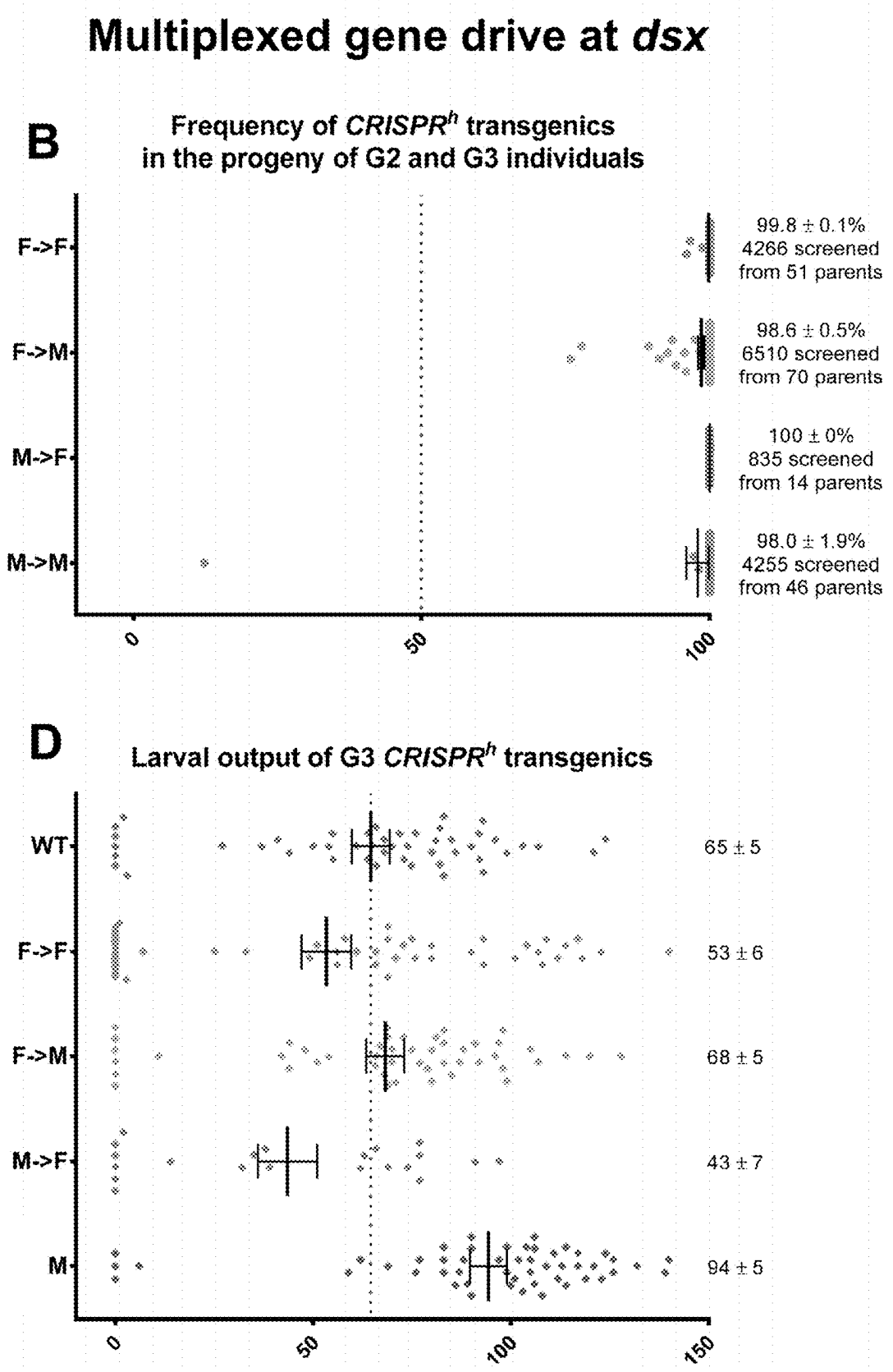

Original gene drive at dsx
A
Frequency of CRISPRh transgenics in the progeny of G2 individuals
F->F
F->M
M->F
M->M
99.3 ± 0.7%
94.1 ± 2.1%
100 ± 0%
97.7 ± 0.8%
0
50
100
C
Larval output of G2 CRISPRh transgenics
WT
F->F
F->M
M->F
M->M
101 ± 9
56 ± 9
121 ± 7
9 ± 4
130 ± 6
0
50
100
150
200

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

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

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

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

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

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

[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. **1***b* and **11***a*, and **12**. This type of ultra-conservation is very rare because even proteins that are highly constrained show some variation at the level of the DNA sequence

because "silent" variation does not alter the composition of the final encoded protein. The inventors carefully assessed the ultra-conserved sequence in the doublesex gene and, without wishing to be bound to any particular theory, believe that it is the splice acceptor site at the 5' boundary of exon 5 that is required for sex-specific splicing of dsx into the female form, as this sequence may represent the target of RNA binding proteins that direct the alternative splicing of this important exon.

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

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

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

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

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

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

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

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

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

[0022] The sequence of the doublesex gene in various arthropods, insects, and mosquito species are publicly available and so known to the skilled person. However, in a preferred embodiment, the doublesex gene is from *Anopheles gambiae* (referred to as AGAP004050), which is provided herein as SEQ ID No: 1. SEQ ID No:1 is the whole AGAP004050 gene, plus about 3000 bp upstream of its putative promoter and about 4000 bp downstream of its putative terminator.

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

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

[SEQ ID No: 2]

```
CCTTTCCATTCATTTATGTTTAACACAGGTCAAGCGGTGGTCAACGAAT

ACTCACGATTGCATAATCTGAACATGTTTGATGGCGTGGAGTTGCGCAA

TACCACCCGTCAGAGTGGATGATAAACTTTC
```

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

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

[SEQ ID No: 3]

CCTTTCCATTCATTTATGTTTAACACAGGTCAAGCGGTGGTCAACGAAT

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]

GTTTAACACAGGTCAAGCGGTGG

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

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

GTTTAACACAGGTCAAGCGG

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

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

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

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

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

[SEQ ID No: 6]

GTTTAACACAGGTCAAGCGGGTTTTAGAGCTAGAAATAGCAAGTTAAAA

TAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT

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

GUUUAACACAGGUCAAGCGGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAU

AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCU

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

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

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

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

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

[SEQ ID No: 49]

TTTGTATGCGTGCGCTTGAAGGGTTGATCGGAACCTTACAACAGTTGTAG

CTATACGGCTGCGTGTGGCTTCTAACGTTATCCATCGCTAGAAGTGAAAC

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]

CAGCGCTGGCGGTGGGGACAGCTCCGGCTGTGGCTGTTCTTGCGAGTCCT

CTTCCTGCGGCACATCCCTCTCGTCGACCAGTTCAGTTTGCTGAGCGTAA

GCCTGCTGCTGTTCGTCCTGCATCATCGGGACCATTTGTATGGGCCATCC

GCCACCACCACCATCACCACCGCCGTCCATTTCTAGGGGCATACCCATCA

GCATCTCCGCGGGCGCCATTGGCGGTGGTGCCAAGGTGCCATTCGTTTGT

TGCTGAAAGCAAAAGAAAGCAAATTAGTGTTGTTTCTGCTGCACACGATA

ATTTTCGTTTCTTGCCGCTAGACACAAACAACACTGCATCTGGAGGGAGA

AATTTGACGCCTAGCTGTATAACTTACCTCAAAGTTATTGTCCATCGTGG

TATAATGGACCTACCGAGCCCGGTTACACTACACAAAGCAAGATTATGCG

ACAAAATCACAGCGAAAACTAGTAATTTTCATCTATCGAAAGCGGCCGAG

CAGAGAGTTGTTTGGTATTGCAACTTGACATTCTGCTGCGGGATAAACCG

CGACGGGCTACCATGGCGCACCTGTCAGATGGCTGTCAAATTTGGCCCGG

TTTGCGATATGGAGTGGGTGAAATTATATCCCACTCGCTGATCGTGAAAA

TAGACACCTGAAAACAATAATTGTTGTGTTAATTTTACATTTTGAAGAAC

AGCACAAGTTTTGCTGACAATATTTAATTACGTTTCGTTATCAACGGCAC

GGAAAGATTATCTCGCTGATTATCCCTCTCGCTCTCTCTGTCTATCATGT

-continued

CCTGGTCGTTCTCGCGTCACCCCGGATAATCGAGAGACGCCATTTTTAAT

TTGAACTACTACACCGACAAGCATGCCGTGAGCTCTTTCAAGTTCTTCTG

TCCGACCAAAGAAACAGAGAATACCGCCCGGACAGTGCCCGGAGTGATCG

ATCCATAGAAAATCGCCCATCATGTGCCACTGAGGCGAACCGGCGTAGCT

TGTTCCGAATTTCCAAGTGCTTCCCCGTAACATCCGCATATAACAAACAG

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]

GTGAACTTCCATGGAATTACGTGCTTTTTCGGAATGGAGTTGGGCTGGTG

AAAAACACCTATCAGCACCGCACTTTTCCCCCGGCATTTCAGGTTATACG

CAGAGACAGAGACTAAATATTCACCCATTCATCACGCACTAACTTCGCAA

TAGATTGATATTCCAAAACTTTCTTCACCTTTGCCGAGTTGGATTCTGGA

TTCTGAGACTGTAAAAAGTCGTACGAGCTATCATAGGGTGTAAAACGGAA

AACAAACAAACGTTTAATGGACTGCTCCAACTGTAATCGCTTCACGCAAA

CAAACACACACGCGCTGGGAGCGTTCCTGGCGTCACCTTTGCACGATGAA

AACTGTAGCAAAACTCGCACGACCGAAGGCTCTCCGTCCCTGCTGGTGTG

TGTTTTTTTCTTTTCTGCAGCAAAATTAGAAAACATCATCATTTGACGAA

AACGTCAACTGCGCGAGCAGAGTGACCAGAAATACCGATGTATCTGTATA

GTAGAACGTCGGTTATCCGGGGGCGGATTAACCGTGCGCACAACCAGTTT

TTTGTGCAGCTTTGTAGTGTCTAGTGGTATTTTCGAAATTCATTTTTGTT

CATTAACAGTTGTTAAACCTATAGTTATTGATTAAAATAATATTCTACTA

ACGATTAACCGATGGATTCAAAGTGAATAAATTATGAAACTAGTGATTTT

TTTAAATTTTTATATGAATTTGACATTTCTTGGACCATTATCATCTTGGT

CTCGAGCTGCCCGAATAATCGACGTTCTACTGTATTCCTACCGATTTTTT

ATATGCCTACCGACACACAGGTGGGCCCCCTAAAACTACCGATTTTTAAT

TTATCCTACCGAAAATCACAGATTGTTTCATAATACAGACCAAAAAGTCA

TGTAACCATTTCCCAAATCACTTAATGTATTAAACTCCATATGGAAATCG

CTAGCAACCAGAACCAGAAGTTCAACAGAGACAACCAATTTCCGTGTATG

TACTTCATGAGATGAGATTGGACGCGCTGGTAAAATTTTATATGGGATTT

GACAGATAATGTAAGGCGTGCGATTTTTTTCATACGATGGAATCAATTCA

AGAGTCAATTGTGCAGGATTTATAGAAACAATCTCTTATTTATGTTTTGT

TATCGTTACAGTTACAGCCCTGTCCTAAGCGGCCGCGTGAAGGCCCAAAA

AAAAGGGAGTCCCCAACGCTCAGTAGCAAATGTGCTTCTCTATCATTCGT

TGGGTTAGAAAAGCCTCATGTGACTTCTATGAACAAAATCTAAACTATCT

CCTTTAAATAGAGAATGGATGTATTTTTTCGTGCCACTGAACTTTCGTTG

-continued

GGAAGATTAGATACCTCTCCCTCCCCCCCCCTCCCTTTCAACACTTCAAA

ACCTACCGAAAACTACCGATACAATTTGATGTACCTACCGAAGACCGCCA

AAATAATCTGGCCACACTGGCTAGATCTGATGTTTTGAAACATCGCCAAA

TTTTACTAAATAATGCACTTGCGCGTTGGTGAAGCTGCACTTAAACAGAT

TAGTTGAATTACGCTTTCTGAAATGTTTTTATTAAACACTTGTTTTTTTT

AATACTTCAATTTAAAGCTACTTCTTGGAATGATAATTCTACCCAAAACC

AAAACCACTTTACAAAGAGTGTGTGGTTGGTGATCGCGCCGGCTACTGCG

ACCTGTGGTCATCGCTCATCTCACGCACACATACGCACACATCTGTCATT

TGAAAAGCTGCACACAATCGTGTGTTGTGCAAAAAACCGTTCGCGCACAA

ACAGTTCGCACATGTTTGCAAGCCGTGCAGCAAAGGGCTTTTGATGGTGA

TCCGCAGTGTTTGGTCAGCTTTTTAATGTGTTTTCGCTTAATCGCTTTTG

TTTGTGTAATGTTTTGTCGGAATAATTTTTATGCGTCGTTACAAATGAAA

TGTACAATCCTGCGATGCTAGTGTAAAACATTGCTAATTCCCGGTAAGAA

CGTTCATTACGCTCGGATATCATCTTACGAAGCGTGTGTATGTGCGCTAG

TACATTGACCTTTAAAGTGATCCTTTTGTTCTAGAAAGCAAG

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

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

[SEQ ID No: 9]

GGAAGGTGATTGCGATTCCATGTTGATGCCAATATATGATGATTTTGTTG

CATATTAATAGTTGTTGTTATGTTTTATTCAAATTTCAAAGATAATTTAC

TTTACATTACAGTTAGTGAGCATATTATCTACTACATAAACACATAGATC

AAACTGGTTTACATAAATTCAAAAAGTTTGGATTAAAATCGCAGCAATTG

GTTATGAAAAAATATGTGCATAACGTAAATATCAAGTAAATTTTTGCATT

GCATATTTATAGACTCCTGTTACAATTTCGGAAAAATGAAAAATGTTAAT

TAATCAAAGAAGAAAAAACAAAGAAATTAAATCATTAGGTAGCACAACCA

CAAGTACATATTTTTATGGCATGAATATTCCTCTACACTAACATATTTTA

TAGCAATTCTATTGATCGCCTTAGTATAGCGGAATTACCAGAACGGCACT

ATAGTTGTCTCTGTTTGGCACACGCAATCATTTTTCATCCCAGGGTTGCC

ATAGCAGTTTGGCGACGGTCACGTAGCATGCGAAGGATTTCGTTCGCACA

GGATCACTTTTATTCTAACGTTTGAAGAAGGCACATCTCAGTGCAAGCGC

TCTGGAAGCTGCTTTTACCGAACGAACTAACTTTTCAAGTAACCTCAAAA

ACTTGTCTCTAACGACACCACGTGCTATCCGCGAGTTTCATTTCCCGTGC

AAAGTTCCCCGATTTAGCTATCATTCGTGAACATTTCGTAGTGCCTCTAC

CCTCAGGTAAGACCATTCGAGGTTTACCAAGTTTTGTGCAAAGAACGTGC

ACAGTAATTTTCGTTCTGGTGAAACCTTCTCTTGTGTAGCTTGTACAAA

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

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

[SEQ ID No: 10]

ATGTAGAACGCGAGCAAATTCTTTTCCTTCCATGACAGCAGCAGCTACAG

TGGGAAGCCGAACGTCAGACGTGTTTGACATGCCGAACTGGGCGGGAAAA

TTACAGCGTGCGCTTTGTTTTCAAGCAAATCACAACTCGCTGCAAACAAA

ACCGTTGAGAAATTGATTGTTTTATAATTTGTATTGTATTTTATTTGTTA

TAATAAACTAAAAAGACATACTTTTTGCATATTTTATACATAAAAACATA

CATGCAGCATTATAAAACACATATAAACCCTCCCTGTAGAGTCCCGTATC

GAAATCTTCCATCCTAGTTGCACAGTACGACGGACGAGTAGGCCGTGTCC

GTGCAAATTCCAGCTTTTAGCAGTCTTTTGCTCGGAGCACTCGCGGCGAG

TCGGAGGTTTCTGCTGAGGTGCTTAGCGCTAAATTAGCCAATTGCTTTTG

CAAGTGAAATAACCAGCCGAATAGTACTTCAAAACTCAGGTAAGTGAACT

AGTTTTATAGAACAAATGTTTGTTTGTTAGAAGTTAGTGAAGTGTTTGTG

AAAAAAATCTCTCATTTCGGCAAAACTAACGTAACTGATTTCAAATTGAA

TTATTGTTTTGTGATGTTATATTATTTCATCCAGTTGATTAGTATTTTCT

TAGTTATGTTCAAAATACAGTTAAATTAAATTTCATTTCATTTACTCATA

AAATAATCTCTTGGCTTATTTAATTTTTCTCGAATTCGCTTGTATTGTTC

AGTAGCACGCGCCATTCGCCCTTTGTTTCATTTTGTACCTGCTCCCACTA

ACACACTGGCAGTGCGAAACAAAAGCCTTCGCACGCGTTGCTGGTATTAG

AGTGTGTGCGTGTGTGTGTTGAGCGCTCTGTCAAAATCGGCTGTTGCCGC

CGGTACCGAAATTGCCTGTTCGCACGCTGTTCGTAAACATTCCGTGGTGT

GTATCGTGTGTTGTGCATGTTGCGCGCCTCCCCCCTTTTGATAGCAGGCT

GCCGTGGCTGCCGTGGTGTGTGGCGCAGTTGAGTTTTTGGATTAATTTTC

TAAGGAAATGGCACGAGAAGAGCGGTGGCAGTGTGTTGGTTTGCTCTGTC

CCTTCCTTTCTGTGTGAAGTGTTCTTACAGCACAGCACGTATCCACCACC

GCACACAGAGCAGGCAAGGAAGTGGAAGTGAACAAGTGTGCTGCGCATGC

ATGTGTGTGGGGGGCATTTTAGCTGAGATCGTCGTTATTTGAGAAGCGGT

ATAGGGGCCAGTCGGTGTCGACGTACGGAAGCGGTTTAGTTTTAATCCAA

GCGTATCCCGTCGTGGAGTGGTTGTGTGGCTCTGTGTGCTCTCATATCAG

TTCCAGAGTGAGGTTAGTAGAATCACAGTCCTTGGCCTTTTTCGTTACAA

GATATCCAGAAGGATGGCGTTATTTCCACAGCTTACCATGGTGCTCTTGT

TTGCTCGAATCAGGGGAGAAAAACAGTTTCGTGTTTCATGAACCGCAGTT

GGCACTGGAGCGGATTCAAAAGTCTTCGATATGCAATAGATAAGAGAGTC

GTTGGGGCATAGTTGGGAAGCCTTTCCGAGATGTGGAGTTTCCGAGAGGA

GAAATGGTGCTTTCGTGCACGTTCCGGGACAGCGGGCCCCGCGAAGAGCA

TCTCGTTGTCGTTCATCCGGCAATAATTGATGCGAAAAGCGCGCGCGCCA

-continued

CTGGCTTAGCGCAGTGTACACAGTGATATTCACCTACACACACAGAGGCA

CACGCCTTCACACGCGCGCGTGCTTCAAAGGCTACTTCGGTGGCGGTGTG

TGAGGTCGCTTGCAATGGACAATGAAAATTTCGCTGGAAAATACCATCGT

CTCTTTAGGTTGCAATGGGTGCGGGTAGAGCGGTGGTCGTCGATATTGGT

GGTGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT

GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT

GTGCAACGGCAATTATTTTTTGTAATATTTCGACCATCTTTCTTTCTCTC

TCTCCACGTGCTGCTGCTGTTGCTGCTGCTGCTGCATTGCATGTTCCACT

ATTCCTCTCGGTTTGTGCCTGCGGACGCCATTGCTAGTCGAAAGAGAGTC

GCCGTTAGTCGCGCTTCGAGCAACGGACACGTTTTTTGGTTGAAACCAAC

AGCTTTTTTCATCTTCGGGAGACACACAGATCTCGAATCGTACATTCCCA

TAAGGAGAATTGTCATCTTCCGGTGAATAAAGAAAGGAAAC

[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]

```
CTTGTGTTTAGCAGGCAGGGGAGATGAGCGCAAACTGTGCAAGAAGAAGC
ATCACTGTGAAGACGGCAATGCAAAGATAGTGTGCTCAACTTCTCCGCGA
AGATTGAAGCTAAATTAAGCACGAGATTAGCATGACTGAAGTGACTTTTC
AAAGTGTCAGAATGGCTGCACTCGCAAACTAGCTGGATGCAGCGCAATTT
TGCCCCGGTGTGTGCGCGCATGCAAACGAGCAACCGCAGAGGGCAAAGGA
GAGGATGGGAAGGAGGGAGGGAGTGAAAGAGCAGGCTTAAGGTTGCCCTC
GGGCATTGAAGTCGATACAGCGGTTCTATTCCAGTGCCAGTAACGATGAC
GAAGACGATGTTGCTTCTGCTGCTGTTGCTGCTGTTGTTGTTGATGATGA
TGATGATAATAGTGCAAATATAAAATAAATCTTCCGTAAGCTTTGTGTAG
TGGTGCGTGGCTACTATAAGCCCGTCTGGAAGCAAGGAAGCTAGTCGGGC
AGGGTCATGCAAAAGGGAGACACCTTCGGAGCTCCGGAGCTCCCGCCGGC
ACTCTCGGGGGGACGTCCGTTATGCGTTGTGATTTATTATGGAATATTTA
TTATAGTGTCTTGTTTTGAAAAAATAACTTCAACGGTTCGAATTTCCTAC
ACCTCGAGATCGGGGCTGGAGTGGCAACGTGGTACGGAACGGTACAGCGG
TTTGAGCCGTTCGGTCTTGGGACTCACGGATCGCAGAATGTTATTGTGCG
CGCACTGATGGGAAAGTCATTTTTCACCGAGTGGTCAGGGCGCGTAGTCC
AGTTCGTTTCTGGCTGCTGTTGCTGATGCTACGATCCTCAGGAATGATTG
GAAACGCCTGGAGATGGTGGGAAAAAATCAAACACAAAAACGATCCTAAT
GAACATCGTGTGTTCTCATTCGCTGCCACGATTGACACCTTCGATAAGAC
GCACATAATGAGCTAAAGGAGAGGGGACAGGGTCTTGTCTTTGCCACGAG
CGATAAGATTGCAATCACTCGTGAGCGTGTGCTGCTGGGCTGAAGAAGAA
ACGCTTTCCACAGCAGTAGGTGGGAAGTGGGATTGTGGAACGTGGCATTG
AAAAGAACCTATTTTCTAAAGCCCGAGAGCCCGTTCTCGAACTGGAAAAC
CAGATGCAGAAGTTTTTTATTGTCCCCCGCCAGGAAAACAAATGTATTTA
ATGCTTTCTTTGCCTTTTCCGCCCCGTTTCAGACGACGAGCTAGTGAAGC
GAGCCCAATGGCTGTTGGAGAAACTCGGCTACCCGTGGGAGATGATGCCC
CTGATGTACGTCATACTAAAGAGCGCCGATGGCGATGTACAAAAAGCACA
CCAGCGGATCGACGAAGGTAAGCTGGCGATGATGGTGTCGTTCGACATCA
CTTTCATCACCGTGTCAGACATCTACTGTGCCTAGCACCGGGTCCAGTGG
```

-continued

```
TCACAGGGTGTAGCAAAAACGTGTTCTTTTTTGCGAGAGACTCTACCTCA
TGATGCAGCTGTTAAGGAAAGGTTTCAGATGAAGGCAATTTTTCCTAGGA
TAAGATGATCTTAAGTTACCTGCGTATTAGTGTTTAACATTGTCGTCTCA
ACTCCCAAGAATGTTTTAATCGTCTAGGGCTAGTTTATTTATACTGTTCT
CATTGAAATGTCGTTCAATCCAACATGTTAAGTTAGCTAGCTCAGACACG
AGAAGTTAGGAGTATCTGCATCTTGAAGGTAGCGGCATATGGTGTTATGC
CACGTTCACTGACTTCAAAATTCGATACAAAAAAAAAACCAAAACATCAA
AAACCAAATTGTGAATTCCGTCAGCCAGCAGCAGTGACCTTCAAAGCCTT
ACCTTTCCATTCATTTATGTTTAACACAGGTCAAG
```

[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]

```
CGGTGGTCAACGAATACTCACGATTGCATAATCTGAACATGTTTGATGGC
GTGGAGTTGCGCAATACCACCCGTCAGAGTGGATGATAAACTTTCCGCAC
CACTGTAACTGTCCGTATCTTTGTATGTGGGTGTGTGTATGTGTGTTTGG
TGAAACGAATTCAATAGTTCTGTGCTATTTTAAATCAAGCCGCGTGCGCA
ACTGATGCCGATAAGTTCAAACTAGTGTTTAAGGAGTGGAGCGAGAGAGC
CGCACCACGGTACAGAAGGGCAGCAGAATGGGTCGGCAGCCTAGCTGCAC
TGGTGCGGTGCGTCCGGCGTCTCGGGGGGAGGGCGAGGAAATTCTAGTGT
TAAATCGGAGCAGCAAAAACAAAACAGTGGTCGTCCCGTTCAAGAAACGG
CCTGTACACACACACAGAAAACACTGCAGCATGTTTGTACATAGTAGATC
CTAGAGCAGGTGGTCGTTGCTCCTCGAACGCTCTGGACGCACGGCTTCGC
GCGTATTTGCGTAGCGTTCCGCCGATCGTGGGTATTCGTACTGCCACAAG
CCCGCTTTCTCCCATGCAATCTCTGCAACCAAACCAACAAACAACAACAA
AAAACCAATCGACAAAATGAATCACACCCCTTTTGTATCATCTGTATATT
CTTGTTCTTTGCGTTCTTTTCTATGTGGCCCACGCCCCGGCGGGTACGTA
ATTGCGTCGAAAACCCCGAAAACCCCGGCACATACAGTGTACATACGGTT
TGAGGACAACTTTGACCTGCAGCCCTTCTGGGGTTGCCACGTGTAGCTAT
ACTTGTGAGATCGGGCGCCGACGGTGTAAAGCGCGAATGGCCGCCACACA
GTGTGTCCACTCCAACACTACCCCTCTGGAACTACCCCGTCCAGGGATGC
ACCGGCTCGGCTCATGCCCCTGCAAAACAGTCCGGGCTCCACTGTAGTAG
CTCCGGCGTTGCTCTGAGAGAAGGATGCCCTTCGAAGTGTCGAAAGCGTG
CATTGGGCGTTCAAGTGTGTGTGTGTGTGTTAGGTTTAGCGAGAAACAGC
AGCAGTTGCGTGTGCTGAAAAGCGAAGGAGTAATAGAGTGCATAATGAAA
ATGAAAATGAAAATGAAGCAAAAGTAGAAGGCGGAGGAGAGCAACCTGTG
TTCCACTAGTAGCGAATAGTTTAGTCTAGTTTCGTCACCAATCAACCTTC
CAACCATCGTTCAACCAATACCTGAGTCAACATCGTCATCGTTATCGTGC
```

-continued

CACAACTTTATTAAAAATGAACCTTGTCCGCGCCACCGTAGGGTGATCTA

AGGCGACCTTTCTTACGGGCGCGACCCACATGCCATCGTCACCTTCTCCA

ATCAAAACCAACAGCCTGTACCGATGGTGTGCAATTGTGCGTGCGTGTGT

GTTATTAGCAAAAAAAGAGAAAGAGTCGACGAGAGAGAGATAGATCGAGA

TCGAGAGTACAAAAGAGCAGTAGAAATGTTCGTTGTTTGTTTTTCGTAAC

ACAGTTGTTTAGCCAAAATGGGAATTTCCAATAATCCCGGGGGCGGGGAA

ATGCGGGAATACTGCGTACACACATACATCAATCAAAAAGAAAAATCCTT

GCGCTACATCACTACCGTTTGCGCGGTGCTGATCTAGAGCAGACCACTTT

CCACTCCACTCTACAATCAATCAATCTGTGCAGAAGGTATGGTAAGACGG

CCTTTGAGCGAGTCACGGTCGCCACCATAACGCCGTCCGACGAGGGCTGA

ATGCGAACTTTGCTAATCGATTTTCCGCTTTCTTTTTATCCCACCTCCTT

TTCTCTCCCTCTCTCTCTTTTGCACTGCCCCTTGTAACCCCCAAAAAGGT

AAACGACACATTAAGACCTACGAAGCGTTGGTGAAGTCATCGCTCGATCC

GAACAGCGACCGGCTGACGGAGGACGACGACGAGGACGAGAACATCTCGG

TGACCCGCACC

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

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

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

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

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

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

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

[SEQ ID No: 13]

TGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCCACCTCACCCATGCGATCGCTCCGGAAAGATACATTGATGAGTT

TGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGC

TGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAA

ACCTCTACAAATGTGGTATGGCTGATTATGATCTAGAGTCGCGGCCGCTACAGGAACAGGTGGTGGCGGCCCTCGGTGCGCTCGTACT

GCTCCACGATGGTGTAGTCCTCGTTGTGGGAGGTGATGTCCAGCTTGGAGTCCACGTAGTAGTAGCCGGGCAGCTGCACGGGCTTCTT

GGCCATGTAGATGGACTTGAACTCCACCAGGTAGTGGCCGCCGTCCTTCAGCTTCAGGGCCTTGTGGATCTCGCCCTTCAGCACGCCG

TCGCGGGGGTACAGGCGCTCGGTGGAGGCCTCCCAGCCCATGGTCTTCTTCTGCATTACGGGGCCGTCGGAGGGGAAGTTCACGCCGA

TGAACTTCACCTTGTAGATGAAGCAGCCGTCCTGCAGGGAGGAGTCTTGGGTCACGGTCACCACGCCGCCGTCCTCGAAGTTCATCAC

GCGCTCCCACTTGAAGCCCTCGGGGAAGGACAGCTTCTTGTAGTCGGGGATGTCGGCGGGGTGCTTCACGTACACCTTGGAGCCGTAC

TGGAACTGGGGGGACAGGATGTCCCAGGCGAAGGGCAGGGGGCCGCCCTTGGTCACCTTCAGCTTCACGGTGTTGTGGCCCTCGTAGG

GGCGGCCCTCGCCCTCGCCCTCGATCTCGAACTCGTGGCCGTTCACGGTGCCCTCCATGCGCACCTTGAAGCGCATGAACTCCTTGAT

GACGTTCTTGGAGGAGCGCACCATGGTGGCGACCTGTGGGTCCCGGGCCCGCGGTACCGTCGACTCTAGCGGTACCCCGATTGTTTAG

CTTGTTCAGCTGCGCTTGTTTATTTGCTTAGCTTTCGCTTAGCGACGTGTTCACTTTGCTTGTTTGAATTGAATTGTCGCTCCGTAGA

CGAAGCGCCTCTATTTATACTCCGGCGGTCGAGGGTTCGAAATCGATAAGCTTGGATCCTAATTGAATTAGCTCTAATTGAATTAGTC

TCTAATTGAATTAGATCCCCGGGCGAGCTCGAATTAACCATTGTGGACCGGTCAGCGCTGGCGGTGGGGACAGCTCCGGCTGTGGCTG

TTCTTGAGAGTCATCTTCCTGCGGCACATCCCTCTCGTCGACCAGTTCAGTTTGCTGAGCGTAAGCCTGCTGCTGTTCGTCCTGCATC

-continued

ATCGGGACCATTTGTACGGGCCATCCGCCACCACCACCATCACCACCGCCGTCCATTTCTAGGGGCATACCCATCAGCATCTCCGCGG

GCGCCATTGGCGGTGGTGCCAAGGTGCCATTCGTTTGTTGCTGAAAGCAAAAGAAAGCAAATTAGTGTTGTTTCTGCTGCACACGATA

GTTTTCGTTTCTTGCCGCTAGACACAAACAACACTGCATCTGGAGGGAGAAATTTGACGCCTAGCTGTATAACTTACCTCAAAGTTAT

TGTCCATCGTGGTATAATGGACCTACCGAGCCCGGTTACACTACACAAAGCAAGATTATGCGACAAAATCACAGCGAAAACTAGTAAT

TTTCATCTATCGAAAGCGGCCGAGCAGAGAGTTGTTTGGTATTGCAACTTGACATTCTGCTGTGGGATAAACCGCGACGGGCTACCAT

GGCGCACCTGTCAGATGGCTGTCAAATTTGGCCCGGTTTGCGATATGGAGTGGGTGAAATTATATCCCACTCGCTGATCGTGAAAATA

GACACCTGAAAACAATAATTGTTGTGTTAATTTTACATTTTGAAGAACAGCACAAGTTTTGCTGACAATATTTAATTACGTTTCGTTA

TCAACGGCACGGAAAGATTATCTCGCTGATTATCCCTCTCGCTCTCTCTGTCTATCATGTCCTGGTCGTTCTCGCGTCACCCCGGATA

ATCGAGAGACGCCATTTTTAATTTGAACTACTACACCGACAAGCATGCCGTGAGCTCTTTCAAGTTCTTCTGTCCGACCAAAGAAACA

GAGAATACCGCCCGGACAGTGCCCGGAGTGATCGATCCATAGAAAATCGCCCATCATGTGCCACTGAAGCGAACCGGCGTAGCTTGTT

CCGAATTTCCAAGTGCTTCCCCGTAACATCCGCATATAACAAGCAGCCCAACAACAAATACAGCATCGAGCTCGAGATGGACTATAAG

GACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTA

TCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGA

GTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTC

GACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGC

AAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAA

GCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAA

CTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCG

AGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCC

CATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTG

CCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGG

CCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGA

CCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTG

AGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGT

ACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCAT

CAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTC

GACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGG

ACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTG

GATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAG

CGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG

AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCT

GTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCC

GGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGG

AAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGC

CCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGC

ATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACG

ACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGC

CGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAG

AACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGG

-continued

GCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCT

GCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGC

TTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGG

TCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGA

GAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAG

ATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGG

TGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGT

CGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATG

ATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTA

CCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTT

TGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCT

ATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCG

TGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCAT

CATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAG

CTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGG

CCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACA

GCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCT

AATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCC

TGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGA

CGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCC

ACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGTAATTAATTAAGAGGACGGCGAGAAGTAATCATATGTCCGCATTTTGCGCAAACC

AGGCGCTTAGACAATTTGCGCGTAAGCACATTCGAAATGTGAAAAGCTGAAAGCAGTGGTTTCGCCAGCCCGAGTTCAGCGAAACGGA

TTCCTTCCAAGTGTTTGCATTCCTGGCGGAGTGTTCCTCCCAAAATGCACTCACCCTGCGTGCAGTGCCAAATCGTGAGTTTCCTAAT

TTTTTCATATTGTTTATTACCTACCAACTAAAGTTGTTGTTATATATTGCGTTTTACGTACGACAAATAAGTTCGTATTCAGAAATAT

TTGCGATAAGAGAGAACTCATTTGCGATGAATCTCATTGTATTTAGCTAAGTGCCTTGATAAGTAAGCGGAACAGCAGGAATATGACA

CTCCTTGGGAAATACATGTAAGCGTCTGTAATTAGATATATATACACGCAACCAAATGGTCCATGGTTGATTTAAGCACTGCCTGTTG

TCGAACATTGCTATAAGCAAAATAAAGAAGCATTCATTAATCTAAAATTTCTTCAAAGTGACTTCAATGATGATCTCTAGGCTATAGT

GAAAGCTGAAAGCTTATTTGACAATGCAAGGGAAAGTGACGCACGTGCGTCGTATGGGACCGCGCGCATCTATTCTCTCAGCTAATTC

CCCTAATCATTAGTAATTGACGGCACGATTTCTGCTTCTTACTTCCTTTTACTTTGGAGCTTTTCATCAATAAAACCAGTACCATGGC

CGTACGCTCAACGGAAAAGCATTCAAAAAAACCCGCGTTCCTCGTGTGATTTGTGGGTGAGTGGCGCCATCTATTAGAGAATAGCTGT

ACTACATCTCGTGGACGAAGGGGTCAGAGAAGTTGAAAGAGAGCTTGATCGACTGCTATCCAAGCTAGGCGAGGAAGGGAGATCGCTA

GAGCAAAAGAAAAAAAATAAGCAAATATCTTTTTTTATAACAAATCGACGTTAGCGAAATATGTTTGAATCGATTTAACGGTTAGAAT

TCCCTTTGGTTCGTTCATTATGCGAGGCGCGCCTTTGTATGCGTGCGCTTGAAGGGTTGATCGGAACCTTACAACAGTTGTAGCTATA

CGGCTGCGTGTGGCTTCTAACGTTATCCATCGCTAGAAGTGAAACGAATGTGCGTAGGTATATATATGAAATGGAGTTGCTCTCTGCT

GTTTAACACAGGTCAAGCGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA

GTCGGTGCTTTTTTTTACGCGTGGGTCCCATGGGTGAGGTGGAGTACGCGCCCGGGGAGCCCAAGGGCACGCCCTGGCACCCGCA

[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
GATGGCGTGGAGTTGCGCAATACCACCCGTCAGAGTGGATGATAAACTTT
CCGCACCACTGTAACTGTCCGTATCTTTGTATGTGGGTGTGTGTATGTGT
GTTTGGTGAAACGAATTCAATAGTTCTGTGCTATTTTAAATCAAGCCGCG
TGCGCAACTGATGCCGATAAGTTCAAACTAGTGTTTAAGGAGTGGAGCGA
GAGAGCCGCACCACGGTACAGAAGGGCAGCAGAATGGGTCGGCAGCCTAG
CTGCACTGGTGCGGTGCGTCCGGCGTCTCGGGGGGAGGGCGAGGAAATTC
TAGTGTTAAATCGGAGCAGCAAAAACAAAACAGTGGTCGTCCCGTTCAAG
AAACGGCCTGTACACACACACAGAAAACACTGCAGCATGTTTGTACATAG
TAGATCCTAGAGCAGGTGGTCGTTGCTCCTCGAACGCTCTGGACGCACGG
CTTCGCGCGTATTTGCGTAGCGTTCCGCCGATCGTGGGTATTCGTACTGC
CACAAGCCCGCTTTCTCCCATGCAATCTCTGCAACCAAACCAACAAACAA
CAACAAAAAACCAATCGACAAAATGAATCACACCCCTTTTGTATCATCTG
TATATTCTTGTTCTTTGCGTTCTTTTCTATGTGGCCCACGCCCCGGCGGG
TACGTAATTGCGTCGAAAACCCCGAAAACCCCGGCACATACAGTGTACAT
ACGGTTTGAGGACAACTTTGACCTGCAGCCCTTCTGGGGTTGCCACGTGT
AGCTATACTTGTGAGATCGGGCGCCGACGGTGTAAAGCGCGAATGGCCGC
CACACAGTGTGTCCACTCCAACACTACCCCTCTGGAACTACCCCGTCCAG
GGATGCACCGGCTCGGCTCATGCCCCTGCAAAACAGTCCGGGCTCCACTG
TAGTAGCTCCGGCGTTGCTCTGAGAGAAGGATGCCCTTCGAAGTGTCGAA
AGCGTGCATTGGGCGTTCAAGTGTGTGTGTGTGTGTTAGGTTTAGCGAGA
AACAGCAGCAGTTGCGTGTGCTGAAAAGCGAAGGAGTAATAGAGTGCATA
ATGAAAATGAAAATGAAAATGAAGCAAAAGTAGAAGGCGGAGGAGAGCAA
CCTGTGTTCCACTAGTAGCGAATAGTTTAGTCTAGTTTCGTCACCAATCA
ACCTTCCAACCATCGTTCAACCAATACCTGAGTCAACATCGTCATCGTTA
TCGTGCCACAACTTTATTAAAAATGAACCTTGTCCGCGCCACCGTAGGGT
GATCTAAGGCGACCTTTCTTACGGGCGCGACCCACATGCCATCGTCACCT
TCTCCAATCAAAACCAACAGCCTGTACCGATGGTGTGCAATTGTGCGTGC
GTGTGTGTTATTAGCAAAAAAAGAGAAAGAGTCGACGAGAGAGAGATAGA
TCGAGATCGAGAGTACAAAAGAGCAGTAGAAATGTTCGTTGTTTGTTTTT
```

-continued

CGTAACACAGTTGTTTAGCCAAAATGGGAATTTCCAATAATCCCGGGGGC

GGGGAAATGCGGGAATACTGCGTACACACATACATCAATCAAAAAGAAAA

ATCCTTGCGCTACATCACTACCGTTTGCGCGGTGCTGATCTAGAGCAGAC

CACTTTCCACTCCACTCTACAATCAATCAATCTGTGCAGAAGGTATGGTA

AGACGGCCTTTG

[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]
TCTGAACATGTTTGATGGCGTGG

[0081] In one embodiment, therefore, the second target site comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID No: 36, or a variant or fragment thereof. In some embodiments, the genetic construct targets a second target site comprising or consisting of the nucleotide sequence substantially as set out in SEQ ID NO: 36, or a fragment or variant thereof. The second target site may include up to 1, 2, 3, 4, 5, 10 or 15 nucleotides 5' and/or 3' of SEQ ID No:36. As is shown in FIG. **12**, T2 is wholly contained within exon 5.

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

[SEQ ID No: 37]
GCAATACCACCCGTCAGAGTGG

[0083] In one embodiment, therefore, the second target site comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID No: 37, or a variant or fragment thereof. In some embodiments, the genetic construct targets a second target site comprising or consisting of the nucleotide sequence substantially as set out in SEQ ID NO: 37, or a fragment or variant thereof. The second target site may include up to 1, 2, 3, 4, 5, 10 or 15 nucleotides 5' and/or 3' of SEQ ID No:37. As is shown in FIG. **12**, T3 is wholly contained within exon 5.

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

[SEQ ID No: 38]
GTTTATCATCCACTCTGACGG

[0085] In one embodiment, therefore, the second target site comprises or consists of a nucleic acid sequence substantially as set out in SEQ ID No: 38, or a variant or fragment thereof. In some embodiments, the genetic construct targets a second target site comprising or consisting of the nucleotide sequence substantially as set out in SEQ ID NO: 38, or a fragment or variant thereof. The second target site may include up to 1, 2, 3, 4, 5, 10 or 15 nucleotides 5' and/or 3' of SEQ ID No:38. As is shown in FIG. **12**, T4 is partially in the 3' end of exon 5 and extends into the untranslated region of exon 5.

[0086] The gene drive construct of the invention may target one or more of a second target site selected from a group consisting of T2, T3 and T4. Most preferably, the gene drive genetic construct of the invention targets T1 and one or more of T2, T3 and T4. For example, the construct may target T1 and T2, or T1 and T3, or T1 and T4, or T1, T2 and T3, T1, T2 and T4, or T1 and T3 and T4, or any combination thereof.

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

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

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

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

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

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

[SEQ ID No: 39]
TCTGAACATGTTTGATGGCG

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

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

[SEQ ID No: 40]
GCAATACCACCCGTCAGAG

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

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

[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]
TCTGAACATGTTTGATGGCGgttttagagctagaaatagcaagttaaaa

taaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgct

[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]
GCAATACCACCCGTCAGAGgttttagagctagaaatagcaagttaaaat

aaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgct

[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]
GTTTATCATCCACTCTGAgttttagagctagaaatagcaagttaaaata

aggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgct

[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]
UCUGAACAUGUUUGAUGGCG

[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]
UCUGAACAUGUUUGAUGGCGGUUUUAGAGCUAGAAAUAGCAAGUUAAAA

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]
GCAAUACCACCCGUCAGAG

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

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

[SEQ ID No: 46]
GCAAUACCACCCGUCAGAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAU

AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCU

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

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

[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]
GUUUAUCAUCCACUCUGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUA

AGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCU

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

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

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

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

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

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

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

[0124] Preferably, in one embodiment, the CRISPR-based gene drive is introduced into the arthropod via a docking

construct, wherein the docking construct comprises integrase attachment sites, preferably attP integrase attachment sites, that are flanked by 5' and 3' homology arms (sixth and seventh nucleotide sequences, respectively) that are homologous to the genomic sequences flanking the two cut-sites which are disposed in exon 5 of the arthropod, such that when the docking construct is introduced into the arthropod, it is integrated into the arthropod's genome by homology directed repair.

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

GAGTGGATGATAAACTTTCCGCACCACTGTAACTGTCCGTATCT

TTGTATGTGGGTGTGTGTATGTGTGTTTGGTGAAACGAATTCAA

TAGTTCTGTGCTATTTTAAATCAAGCCGCGTGCGCAACTGATGC

CGATAAGTTCAAACTAGTGTTTAAGGAGTGGAGCGAGAGAGCCG

CACCACGGTACAGAAGGGCAGCAGAATGGGTCGGCAGCCTAGCT

-continued

GCACTGGTGCGGTGCGTCCGGCGTCTCGGGGGGAGGGCGAGGAA

ATTCTAGTGTTAAATCGGAGCAGCAAAAACAAAACAGTGGTCGT

CCCGTTCAAGAAACGGCCTGTACACACACACAGAAAACACTGCA

GCATGTTTGTACATAGTAGATCCTAGAGCAGGTGGTCGTTGCTC

CTCGAACGCTCTGGACGCACGGCTTCGCGCGTATTTGCGTAGCG

TTCCGCCGATCGTGGGTATTCGTACTGCCACAAGCCCGCTTTCT

CCCATGCAATCTCTGCAACCAAACCAACAAACAACAACAAAAAA

CCAATCGACAAAATGAATCACACCCCTTTTGTATCATCTGTATA

TTCTTGTTCTTTGCGTTCTTTTCTATGTGGCCCACGCCCCGGCG

GGTACGTAATTGCGTCGAAAACCCCGAAAACCCCGGCACATACA

GTGTACATACGGTTTGAGGACAACTTTGACCTGCAGCCCTTCTG

GGGTTGCCACGTGTAGCTATACTTGTGAGATCGGGCGCCGACGG

TGTAAAGCGCGAATGGCCGCCACACAGTGTGTCCACTCCAACAC

TACCCCTCTGGAACTACCCCGTCCAGGGATGCACCGGCTCGGCT

CATGCCCCTGCAAAACAGTCCGGGCTCCACTGTAGTAGCTCCGG

CGTTGCTCTGAGAGAAGGATGCCCTTCGAAGTGTCGAAAGCGTG

CATTGGGCGTTCAAGTGTGTGTGTGTGTGTTAGGTTTAGCGAGA

AACAGCAGCAGTTGCGTGTGCTGAAAAGCGAAGGAGTAATAGAG

TGCATAATGAAAATGAAAATGAAAATGAAGCAAAAGTAGAAGGC

GGAGGAGAGCAACCTGTGTTCCACTAGTAGCGAATAGTTTAGTC

TAGTTTCGTCACCAATCAACCTTCCAACCATCGTTCAACCAATA

CCTGAGTCAACATCGTCATCGTTATCGTGCCACAACTTTATTAA

AAATGAACCTTGTCCGCGCCACCGTAGGGTGATCTAAGGCGACC

TTTCTTACGGGCGCGACCCACATGCCATCGTCACCTTCTCCAAT

CAAAACCAACAGCCTGTACCGATGGTGTGCAATTGTGCGTGCGT

GTGTGTTATTAGCAAAAAAAGAGAAAGAGTCGACGAGAGAGAGA

TAGATCGAGATCGAGAGTACAAAAGAGCAGTAGAAATGTTCGTT

GTTTGTTTTTCGTAACACAGTTGTTTAGCCAAAATGGGAATTTC

CAATAATCCCGGGGGCGGGGAAATGCGGGAATACTGCGTACACA

CATACATCAATCAAAAAGAAAAATCCTTGCGCTACATCACTACC

GTTTGCGCGGTGCTGATCTAGAGCAGACCACTTTCCACTCCACT

CTACAATCAATCAATCTGTGCAGAAGGTATGGTAAGACGGCCTT

TGAGCGAGTCACGGTCGCCACCATAACGCCGTCCGACGAGGGCT

GAATGCGAACTTTGCTAATCGATTTTCCGCTTTCTTTTTATCCC

ACCTCCTTTTCTCTCCCTCTCTCTCTTTTGCACTGCCCCTTGTA

ACCCCCAAAAAGGTAAACGACACATTAAGACCTACGAAGCGTTG

GTGAAGTCATCGCTCGATCCGAACAGCGACCGGCTGACGGAGGA

CGACGACGAGGACGAGAACATCTCGGTGACCCGCACC

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

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

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

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

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

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

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

[SEQ ID No: 51]

```
tgcgggtgccagggcgtgcccttgggctccccgggcgcgtactc

cacctcacccatgcgatcgctccggaaagatacattgatgagtt

tggacaaaccacaactagaatgcagtgaaaaaaatgctttattt
```

-continued

```
gtgaaatttgtgatgctattgctttatttgtaaccattataagc

tgcaataaacaagttaacaacaacaattgcattcattttatgtt

tcaggttcagggggaggtgtgggaggttttttaaagcaagtaaa

acctctacaaatgtggtatggctgattatgatctagagtcgcgg

ccgctacaggaacaggtggtggcggccctcggtgcgctcgtact

gctccacgatggtgtagtcctcgttgtgggaggtgatgtccagc

ttggagtccacgtagtagtagccgggcagctgcacgggcttctt

ggccatgtagatggacttgaactccaccaggtagtggccgccgt

ccttcagcttcagggccttgtggatctcgcccttcagcacgccg

tcgcgggggtacaggcgctcggtggaggcctcccagcccatggt

cttcttctgcattacggggccgtcggaggggaagttcacgccga

tgaacttcaccttgtagatgaagcagccgtcctgcagggaggag

tcttgggtcacggtcaccacgccgccgtcctcgaagttcatcac

gcgctcccacttgaagccctcggggaaggacagcttcttgtagt

cggggatgtcggcggggtgcttcacgtacaccttggagccgtac

tggaactgggggacaggatgtcccaggcgaagggcagggggcc

gcccttggtcaccttcagcttcacggtgttgtggccctcgtagg

ggcggccctcgccctcgccctcgatctcgaactcgtggccgttc

acggtgccctccatgcgcaccttgaagcgcatgaactccttgat

gacgttcttggaggagcgcaccatggtggcgacctgtgggtccc

gggcccgcggtaccgtcgactctagcggtaccccgattgtttag

cttgttcagctgcgcttgtttatttgcttagctttcgcttagcg

acgtgttcactttgcttgtttgaattgaattgtcgctccgtaga

cgaagcgcctctatttatactccggcggtcgagggttcgaaatc

gataagcttggatcctaattgaattagctctaattgaattagtc

tctaattgaattagatccccgggcgagctcgaattaaccattgt

ggaccggtcagcgctggcggtggggacagctccggctgtggctg

ttcttgagagtcatcttcctgcggcacatccctctcgtcgacca

gttcagtttgctgagcgtaagcctgctgctgttcgtcctgcatc

atcgggaccatttgtacgggccatccgccaccaccaccatcacc

accgccgtccatttctaggggcatacccatcagcatctccgcgg

gcgccattggcggtggtgccaaggtgccattcgtttgttgctga

aagcaaaagaaagcaaattagtgttgtttctgctgcacacgata

gttttcgtttcttgccgctagacacaaacaacactgcatctgga

gggagaaatttgacgcctagctgtataacttacctcaaagttat

tgtccatcgtggtataatggacctaccgagcccggttacactac

acaaagcaagattatgcgacaaaatcacagcgaaaactagtaat

tttcatctatcgaaagcggccgagcagagagttgtttggtattg

caacttgacattctgctgtgggataaaccgcgacgggctaccat
```

-continued

ggcgcacctgtcagatggctgtcaaatttggcccggtttgcgat
atggagtgggtgaaattatatcccactcgctgatcgtgaaaata
gacacctgaaaacaataattgttgtgttaattttacattttgaa
gaacagcacaagttttgctgacaatatttaattacgtttcgtta
tcaacggcacggaaagattatctcgctgattatccctctcgctc
tctctgtctatcatgtcctggtcgttctcgcgtcaccccggata
atcgagagacgccatttttaatttgaactactacaccgacaagc
atgccgtgagctctttcaagttcttctgtccgaccaaagaaaca
gagaataccgcccggacagtgcccggagtgatcgatccatagaa
aatcgcccatcatgtgccactgaagcgaaccggcgtagcttgtt
ccgaatttccaagtgcttccccgtaacatccgcatataacaagc
agcccaacaacaaatacagcatcgagctcgagatggactataag
gaccacgacggagactacaaggatcatgatattgattacaaaga
cgatgacgataagatggccccaaagaagaagcggaaggtcggta
tccacggagtcccagcagccgacaagaagtacagcatcggcctg
gacatcggcaccaactctgtgggctgggccgtgatcaccgacga
gtacaaggtgcccagcaagaaattcaaggtgctgggcaacaccg
accggcacagcatcaagaagaacctgatcggagccctgctgttc
gacagcggcgaaacagccgaggccacccggctgaagagaaccgc
cagaagaagatacaccagacggaagaaccggatctgctatctgc
aagagatcttcagcaacgagatggccaaggtggacgacagcttc
ttccacagactggaagagtccttcctggtggaagaggataagaa
gcacgagcggcaccccatcttcggcaacatcgtggacgaggtgg
cctaccacgagaagtaccccaccatctaccacctgagaaagaaa
ctggtggacagcaccgacaaggccgacctgcggctgatctatct
ggccctggcccacatgatcaagttccggggccacttcctgatcg
agggcgacctgaaccccgacaacagcgacgtggacaagctgttc
atccagctggtgcagacctacaaccagctgttcgaggaaaaccc
catcaacgccagcggcgtggacgccaaggccatcctgtctgcca
gactgagcaagagcagacggctggaaaatctgatcgcccagctg
cccggcgagaagaagaatggcctgttcggaaacctgattgccct
gagcctgggcctgacccccaacttcaagagcaacttcgacctgg
ccgaggatgccaaactgcagctgagcaaggacacctacgacgac
gacctggacaacctgctggcccagatcggcgaccagtacgccga
cctgtttctggccgccaagaacctgtccgacgccatcctgctga
gcgacatcctgagagtgaacaccgagatcaccaaggcccccctg
agcgcctctatgatcaagagatacgacgagcaccaccaggacct
gaccctgctgaaagctctcgtgcggcagcagctgcctgagaagt
acaaagagattttcttcgaccagagcaagaacggctacgccggc

-continued

tacattgacggcggagccagccaggaagagttctacaagttcat
caagcccatcctggaaaagatggacggcaccgaggaactgctcg
tgaagctgaacagagaggacctgctgcggaagcagcggaccttc
gacaacggcagcatcccccaccagatccacctgggagagctgca
cgccattctgcggcggcaggaagatttttacccattcctgaagg
acaaccgggaaaagatcgagaagatcctgaccttccgcatcccc
tactacgtgggccctctggccaggggaaacagcagattcgcctg
gatgaccagaaagagcgaggaaaccatcacccccctggaacttcg
aggaagtggtggacaagggcgcttccgcccagagcttcatcgag
cggatgaccaacttcgataagaacctgcccaacgagaaggtgct
gcccaagcacagcctgctgtacgagtacttcaccgtgtataacg
agctgaccaaagtgaaatacgtgaccgagggaatgagaaagccc
gccttcctgagcggcgagcagaaaaaggccatcgtggacctgct
gttcaagaccaaccggaaagtgaccgtgaagcagctgaaagagg
actacttcaagaaaatcgagtgcttcgactccgtggaaatctcc
ggcgtggaagatcggttcaacgcctccctgggcacataccacga
tctgctgaaaattatcaaggacaaggacttcctggacaatgagg
aaaacgaggacattctggaagatatcgtgctgaccctgacactg
tttgaggacagagagatgatcgaggaacggctgaaaacctatgc
ccacctgttcgacgacaaagtgatgaagcagctgaagcggcgga
gatacaccggctggggcaggctgagccggaagctgatcaacggc
atccgggacaagcagtccggcaagacaatcctggatttcctgaa
gtccgacggcttcgccaacagaaacttcatgcagctgatccacg
acgacagcctgacctttaaagaggacatccagaaagcccaggtg
tccggccagggcgatagcctgcacgagcacattgccaatctggc
cggcagccccgccattaagaagggcatcctgcagacagtgaagg
tggtggacgagctcgtgaaagtgatgggccggcacaagcccgag
aacatcgtgatcgaaatggccagagagaaccagaccacccagaa
gggacagaagaacagccgcgagagaatgaagcggatcgaagagg
gcatcaaagagctgggcagccagatcctgaaagaacaccccgtg
gaaaacacccagctgcagaacgagaagctgtacctgtactacct
gcagaatgggcgggatatgtacgtggaccaggaactggacatca
accggctgtccgactacgatgtggaccatatcgtgcctcagagc
tttctgaaggacgactccatcgacaacaaggtgctgaccagaag
cgacaagaaccggggcaagagcgacaacgtgccctccgaagagg
tcgtgaagaagatgaagaactactggcggcagctgctgaacgcc
aagctgattacccagagaaagttcgacaatctgaccaaggccga
gagaggcggcctgagcgaactggataaggccggcttcatcaaga
gacagctggtggaaacccggcagatcacaaagcacgtggcacag

-continued

atcctggactcccggatgaacactaagtacgacgagaatgacaa
gctgatccgggaagtgaaagtgatcaccctgaagtccaagctgg
tgtccgatttccggaaggatttccagttttacaaagtgcgcgag
atcaacaactaccaccacgcccacgacgcctacctgaacgccgt
cgtgggaaccgccctgatcaaaaagtaccctaagctggaaagcg
agttcgtgtacggcgactacaaggtgtacgacgtgcggaagatg
atcgccaagagcgagcaggaaatcggcaaggctaccgccaagta
cttcttctacagcaacatcatgaactttttcaagaccgagatta
ccctggccaacggcgagatccggaagcggcctctgatcgagaca
aacggcgaaaccggggagatcgtgtgggataagggccgggattt
tgccaccgtgcggaaagtgctgagcatgccccaagtgaatatcg
tgaaaaagaccgaggtgcagacaggcggcttcagcaaagagtct
atcctgcccaagaggaacagcgataagctgatcgccagaaagaa
ggactgggaccctaagaagtacggcggcttcgacagccccaccg
tggcctattctgtgctggtggtggccaaagtggaaaagggcaag
tccaagaaactgaagagtgtgaaagagctgctggggatcaccat
catggaaagaagcagcttcgagaagaatcccatcgactttctgg
aagccaagggctacaaagaagtgaaaaaggacctgatcatcaag
ctgcctaagtactccctgttcgagctggaaaacggccggaagag
aatgctggcctctgccggcgaactgcagaagggaaacgaactgg
ccctgccctccaaatatgtgaacttcctgtacctggccagccac
tatgagaagctgaagggctcccccgaggataatgagcagaaaca
gctgtttgtggaacagcacaagcactacctggacgagatcatcg
agcagatcagcgagttctccaagagagtgatcctggccgacgct
aatctggacaaagtgctgtccgcctacaacaagcaccgggataa
gcccatcagagagcaggccgagaatatcatccacctgtttaccc
tgaccaatctgggagcccctgccgccttcaagtactttgacacc
accatcgaccggaagaggtacaccagcaccaaagaggtgctgga
cgccaccctgatccaccagagcatcaccggcctgtacgagacac
ggatcgacctgtctcagctgggaggcgacaaaaggccggcggcc
acgaaaaaggccggccaggcaaaaaagaaaaagtaattaattaa
gaggacggcgagaagtaatcatatgtccgcattttgcgcaaacc
aggcgcttagacaatttgcgcgtaagcacattcgaaatgtgaaa
agctgaaagcagtggtttcgccagcccgagttcagcgaaacgga
ttccttccaagtgtttgcattcctggcggagtgttcctcccaaa
atgcactcaccctgcgtgcagtgccaaatcgtgagtttcctaat
tttttcatattgtttattacctaccaactaaagttgttgttata
tattgcgttttacgtacgacaaataagttcgtattcagaaatat
ttgcgataagagagaactcatttgcgatgaatctcattgtattt

-continued

agctaagtgccttgataagtaagcggaacagcaggaatatgaca
ctccttgggaaatacatgtaagcgtctgtaattagatatatata
cacgcaaccaaatggtccatggttgatttaagcactgcctgttg
tcgaacattgctataagcaaaataaagaagcattcattaatcta
aaatttcttcaaagtgacttcaatgatgatctctaggctatagt
gaaagctgaaagcttatttgacaatgcaagggaaagtgacgcac
gtgcgtcgtatgggaccgcgcgcatctattctctcagctaattc
ccctaatcattagtaattgacggcacgatttctgcttcttactt
ccttttactttggagcttttcatcaataaaaccagtaccatggc
cgtacgctcaacggaaaagcattcaaaaaaacccgcgttcctcg
tgtgatttgtgggtgagtggcgccatctattagagaatagctgt
actacatctcgtggacgaaggggtcagagaagttgaaagagagc
ttgatcgactgctatccaagctaggcgaggaagggagatcgcta
gagcaaaagaaaaaaaataagcaaatatctttttttataacaaa
tcgacgttagcgaaatatgtttgaatcgatttaacggttagaat
tccctttggttcgttcattatgcgaggcgcgcctttgtatgcgt
gcgcttgaagggttgatcggaaccttacaacagttgtagctata
cggctgcgtgtggcttctaacgttatccatcgctagaagtgaaa
cgaatgtgcgtaggtatatatatgaaatggagttgctctctgct
GTTTAACACAGGTCAAGCGGgttttagagctagaaatagcaagt
taaaataaggctagtccgttatcaacttgaaaaagtggcaccga
gtcggtgctttttttttttgtatgcgtgcgcttgaagggttgat
cggaaccttacaacagttgtagctatacggctgcgtgtggcttc
taacgttatccatcgctagaagtgaaacgaatgtgcgtaggtat
atatatgaaatggagttgctctctgctGCAATACCACCCGTCAG
AGgttttagagctagaaatagcaagttaaaataaggctagtccg
ttatcaacttgaaaaagtggcaccgagtcggtgctttttttac
gcgtgggtcccatgggtgaggtggagtacgcgcccggggagccc
aagggcacgccctggcacccgca

[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 3× sodium chloride/sodium citrate (SSC) at approximately 45° C. followed by at least one wash in 0.2×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 (±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 (±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 φC31 knock-in integration. Primers (blue arrows) were designed to bind internal of the φC31 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 φC31 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 3×P3 promoter and the gRNA under the control

of the constitutive U6 promoter, all enclosed within two attB sequences. The cassette was inserted at the target locus using recombinase-mediated cassette exchange (RMCE) by injecting embryos with a plasmid containing the cassette and a plasmid containing a Φ31 recombination transcription unit. During meiosis the Cas9/gRNA complex cleaves the wild type allele at the target locus (DSB) and the construct is copied across to the wild type allele via HDR (homing) disrupting exon 5 in the process. (b) Representative example of molecular confirmation of successful RMCE events. Primers (blue arrows) that bind components of the CRISPRh cassette were combined with primers that bind the genomic region surrounding the construct. PCRs were performed on both sides of the CRISPRh cassette (5' and 3') on many individuals as well as wild type controls (wt).

[0176] FIG. **9** shows the maternal or paternal inheritance of the dsxF^{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≥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 (±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 (±s.e.m.) is shown. (***, $p<0.001$;****, $p<0.0001$; Kruskal-Wallis test).

[0177] FIGS. **10**A-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. **10**A), 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. **10**B) 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. **10**C) around the cut-site (shaded area+/−20 bp) when compared with the rest of the amplicon, confirming that the gene drive did not generate any substitution activity at the target locus and that the laboratory colony is devoid of any standing variation in the form of SNPs within the entire amplicon.

[0178] FIG. **11** shows a sequence comparison of the dsx female-specific exon 5 across members of the *Anopheles* genus and SNP data obtained from *Anopheles gambiae* mosquitoes in Africa. (a) Sequence comparison of the dsx Intron4-Exon5 boundary and the dsx female-specific exon 5 within the 16 Anopheline species. The sequence of the Intron4-Exon5 boundary is completely conserved within the six species that form the *Anopheles gambiae* complex (noted in bold). The gRNA used to target the gene is underlined and the PAM is highlighted in blue. Changes in the DNA sequence are shaded grey and codon silent and missense substitutions are noted in blue and red respectively. (b) SNP frequencies obtained from 765 *Anopheles gambiae* mosquitoes captured across Africa[17]. 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. **14**A & C) or two targets, i.e. T1 and T3 (Figures B & D). Female or male gene drive carriers that inherited the drive from a female or male transgenic individual (F->F, F->M, M->F, M->M) were crossed to wild-type mosquitoes. Females were allowed to lay individually. The reproductive output of females was determined by counting eggs and hatched larvae and transmission rates were determined by screening the progeny for RFP fluorescence, indicative of carrying the gene drive. Figures A & B show that the transmission rates correspond to the total number of RFP+ progeny over the total number of screened progeny per female. Mean transmission rates±s.e.m. (standard error of mean) are shown. Figures C & D show that the larval output of each class is shown, including a wild-type control, as the standard for comparison (red line). Mean larval outputs±s.e.m. are shown. Note that females with zero larval output that showed no evidence of mating were all included in the analysis, since mating competence can be affected by 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≤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, 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, 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
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.

$$e_W = \left(F_{WW} + w_{WW}^{10}F_{WW}^{10} + w_{WW}^{01}F_{WW}^{01} + w_{WW}^{11}F_{WW}^{11} + \left(F_{WR} + w_{WR}^{10}F_{WR}^{10} + w_{WR}^{01}F_{WR}^{01} + w_{WR}^{11}F_{WR}^{11}\right)/2\right)/\overline{w}_f$$

$$e_R = \frac{1}{2}\left(F_{WR} + w_{WR}^{10}F_{WR}^{10} + w_{WR}^{01}F_{WR}^{01} + w_{WR}^{11}F_{WR}^{11}\right)/\overline{w}_f$$

$$e_W^* = (1-d_f)(1-u_f)\left(w_{WD}^{10}F_{WD}^{10} + w_{WD}^{01}F_{WD}^{01} + w_{WD}^{11}F_{WD}^{11}\right)/\overline{w}_f$$

$$e_D^* = d_f\left(w_{WD}^{10}F_{WD}^{10} + w_{WD}^{01}F_{WD}^{01} + w_{WD}^{11}F_{WD}^{11}\right)/\overline{w}_f$$

$$e_R^* = (1-d_f)u_f\left(w_{WD}^{10}F_{WD}^{10} + w_{WD}^{01}F_{WD}^{01} + w_{WD}^{11}F_{WD}^{11}\right)/\overline{w}_f$$

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

$$s_W = \left(M_{WW} + M_{WW}^{10} + M_{WW}^{01} + M_{WW}^{11} + \left(M_{WR} + M_{WR}^{10} + M_{WR}^{01} + M_{WR}^{11}\right)/2\right)/\overline{w}_m$$

$$s_R = \left(M_{RR} + \left(M_{WR} + M_{WR}^{10} + M_{WR}^{01} + M_{WR}^{11}\right)/2\right)/\overline{w}_m$$

$$s_W^* = (1-d_m)(1-u_m)\left(M_{WD}^{10} + M_{WD}^{01} + M_{WD}^{11}\right)/\overline{w}_m$$

$$s_D^* = \left(M_{DD} + M_{DR}/2 + d_m\left(M_{WD}^{10} + M_{WD}^{01} + M_{WD}^{11}\right)\right)/\overline{w}_m$$

$$s_R^* = \left(M_{DR}/2 + (1-d_m)u_m\left(M_{WD}^{01} + M_{WD}^{10} + M_{WD}^{11}\right)\right)/\overline{w}_m$$

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

$$\overline{w}_f = F_{WW} + w_{WW}^{10}F_{WW}^{10} + w_{WW}^{01}F_{WW}^{01} + w_{WW}^{11}F_{WW}^{11} + w_{WD}^{10}F_{WD}^{10} + w_{WD}^{01}F_{WD}^{01} + w_{WD}^{11}F_{WD}^{11} + F_{WR} + F_{WR}^{10}w_{WR}^{10} + w_{WR}^{01}F_{WR}^{01} + w_{WR}^{11}F_{WR}^{11}$$

$$\overline{w}_m = M_{WW} + M_{WW}^{10} + M_{WW}^{01} + M_{WW}^{11} + M_{WD}^{10} + M_{WD}^{01} + M_{WD}^{11} + M_{WR} + M_{WR}^{10} + M_{WR}^{01} + M_{WR}^{11} + M_{DD} + M_{DR} + M_{RR} = 1$$

[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, genotype frequencies of type i/j in the next 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:

$$G_{WW}(t+1) = e_W s_W$$

$$G_{WW}^{10}(t+1) = e_W^* s_W$$

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

$$G_{WW}^{11}(t+1) = e_W^* s_W^*$$

$$G_{WD}^{10}(t+1) = e_D^* s_W$$

-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 Mathematica[23].

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	TGCTGTTTAACACAGGTCAAGCGG-SEQ ID No: 14
dsxgRNA-R	AAACCCGCTTGACCTGTGTTAAAC-SEQ ID No: 15
dsxΦ31L-F	GCTCGAATTAACCATTGTGGACCGGTCTTGTGTTTAGCAG GCAGGGGA-SEQ ID No: 16
dsxΦ31L-R	TCCACCTCACCCATGGGACCCACGCGTGGTGCGGGTCACC GAGATGTTC-SEQ ID No: 17
dsxΦ31R-F	CACCAAGACAGTTAACGTATCCGTTACCTTGACCTGTGTTA AACATAAAT-SEQ ID No: 18
dsxΦ31R-R	GGTGGTAGTGCCACACAGAGAGCTTCGCGGTGGTCAACG AATACTCACG-SEQ ID No: 19
zpgprCRISPR-F	GCTCGAATTAACCATTGTGGACCGGTCAGCGCTGGCGGTG GGGA-SEQ ID No: 20
zpgprCRISPR-R	TCGTGGTCCTTATAGTCCATCTCGAGCTCGATGCTGTATTT GTTGT-SEQ ID No: 21
zpgteCRISPR-F	AGGCAAAAAAGAAAAAGTAATTAATTAAGAGGACGGCGA GAAGTAATCAT-SEQ ID No: 22
zpgteCRISPR-R	TTCAAGCGCACGCATACAAAGGCGCGCCTCGCATAATGAA CGAACCAAAGG-SEQ ID No: 23
dsxin3-F	GGCCCTTCAACCCGAAGAAT-SEQ ID No: 24
dsxex6-R	CTTTTTGTACAGCGGTACAC-SEQ ID No: 25
GFP-F	GCCCTGAGCAAAGACCCCAA-SEQ ID No: 26
dsxex4-F	GCACACCAGCGGATCGACGAAG-SEQ ID No: 27
dsxex5-R	CCCACATACAAAGATACGGACAG-SEQ ID No: 28
dsxex6-R	GAATTTGGTGTCAAGGTTCAGG-SEQ ID No: 29
3xP3	TATACTCCGGCGGTCGAGGGTT-SEQ ID No: 30
hCas9-F	CCAAGAGAGTGATCCTGGCCGA-SEQ ID No: 31
dsxex5-R1	CTTATCGGCATCAGTTGCGCAC-SEQ ID No: 32
dsxin4-F	GGTGTTATGCCACGTTCACTGA-SEQ ID No: 33
RFP-R	CAAGTGGGAGCGCGTGATGAAC-SEQ ID No: 34

TABLE 6

Primers used in this study for Example 2	
multidsxΦ31L-F	gctcgaattaaccattgtggaccggtCTTGTGTTTAGCAGGCAGGGGA-SEQ ID No: 52
multidsxΦ31L-R	tgaacgattggggtaccggtCTTGACCTGTGTTAAACATAAATG-SEQ ID No: 53
multidsxΦ31R-F	agatataatcctgaacgcgtGAGTGGATGATAAACTTTCCGCAC-SEQ ID No: 54
multidsxΦ31R-R	tccacctcacccatgggacccacgcgtGGTGCGGGTCACCGAGATGTTC-SEQ ID No: 55
4050-2U6-T1-F	gagggtctcaTGCTGTTTAACACAGGTCAAGCGGgttttagagctagaaatagcaagt-SEQ ID No: 56
4050-2U6-T3-R	gagggtctcaAAACCTCTGACGGGTGGTATTGCagcagagagcaactccatttcat-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. **1***c*). 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. **1***d*).

[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 3×P3::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. **4***a*). 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. **4***b*). 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. **5***a*).

[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. **5***b*). 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. **10**A-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

Genetic females homozygous for the insertion carry male-specific characteristics

	Genetic Males			Genetic Females		
Characteristic	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

Summary of values obtained from the cage trials

	Cage Trial 1				Cage Trial 2			
Generation	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 commitment[18]. The development of a gene drive solutions capable of collapsing a human malaria vector population is a long sought scientific and technical achievement[19]. The gene drive dsxFCRISPRh targeting exon 5 of dsx showed a number of desired efficacy features for field applications, in term of inheritance bias, fertility of heterozygous individuals, phenotype of homozygous females and apparent lack of nuclease-resistant functional variants at the target site.

Example 2

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

[0217] Exon 5 of doublesex that was targeted with a gene drive as described in Example 1 contains a total of four

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

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

[0219] The results revealed that the novel multiplexed gene drive can successfully bias its inheritance to the next generation with transmission rates comparable to the single-guide gene drive we previously developed ($p>0.05$) or higher ($p=0.04$), when the gene drive was transmitted by a male carrier who inherited it maternally (F->M class) (FIGS. **14**A and **14**B). 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. **14**C and **14**D). 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. **14**C and **14**D).

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 CRTSPR/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

```
Sequence total quantity: 94
SEQ ID NO: 1            moltype = DNA  length = 94797
FEATURE                 Location/Qualifiers
source                  1..94797
                        mol_type = other DNA
                        organism = Anopheles gambiae
SEQUENCE: 1
gctaatttcc aagtcccaaa tgttctggtg gtatattcat ttcttataac aagaacccgt  60
tgtttatgaa taattttgtt aaattactat aattttatcc gatgcaaata gtaagaacag  120
atttttggtt tgcagtgctt acagcacttc tcaaaatatt ctcgcgggcc gcattcatta  180
tccacgtggg ccgtatgcgg cccgcgggcc gccagtttga catacctgca ttaaaagaac  240
cgtagcgttc ttctcttgta aaccggttca ttcatttttt tcacgtgaac caaatgaacg  300
gttctgattc atttggcaca cttctagtac agacaaactt taatcgacaa cagttgttgt  360
gccaatgaag aaaaataata ataattataa tattaataac aataataaaa agtaagtagg  420
gattgtctgt aagagtattt tttctgttta tttattcgta ttgaaataat ctaaaaacta  480
ttttcaactt ctttatggtt taaattctta cctcttcctt ttcaataaac aaagaaaaaa  540
cagttcaaaa taatatttta tttacaaata ataaccaacc attataacga aagcgtacag  600
atctcttcct aatgccatcg gtttgacgcg catattgtta cttgggaccc ttgcctcacg  660
catacataac aagcgagcgc gtaaggctgt gctctagcat atggaaccgt gcgtcgaaca  720
ctctatcgcc catattgtgc tgcgttggga aacaacctat cttggccttt ggaaaaccgc  780
tttctggctg ctcccggaag aacaccactc aaacatgcat cgcgagcaaa taaacaccca  840
atcgcacact ctacaacatg cacgtgtttg aaaaagaaac tcgagccgta cgacagtctc  900
tagttacagc acagcctcag taacaatgtt gtgaatgtat tgcagggacg ttgtgttgtg  960
gcgcagtctt ttttttaaac aaaaccgaac ccttagtgta aaccgaacgt ggttgtgggg  1020
atagagcgtt agaggggtgg gcagggaagg gtggaaaaat caaaaacttg ttgcacactc  1080
cgccggacca gaccgttgcg atgtgtgtgc tgacctacaa caactttcct ttcccagccc  1140
tactgcccca tcctaccgaa ccgtccgctc cggtgaggca gcgtgctcat cgatgtgtgc  1200
gagctgaaaa gggccgtgcg cgtgtgtttg tgcgaaacgt atgtgtgtgt gtgtgagtgt  1260
gtttgcgtaa atgcacattt atcagtgcag ttccgcgtac tcgccgcttc gcaatcgcaa  1320
tctggtcttt aatcgaggag gcaacatttg accatcgctc gttggcagtt gccgtttact  1380
actggggcgg gtgtaacgag gcccacaaca gcagcacgga tcttgtgctt taacggtgag  1440
acgacggtaa aggtagcgca aaaaataata cacaatgtgt gcaaagtgca gtgaaaacaa  1500
aagcgttatg taggtgtttt aagcaaaggt tctacaagtg cgtataccaa agttgacaaa  1560
gtgcgcgaaa tcggactctg ccaagaagtg ccgggaacaa aacaaaacag ctacaacaac  1620
acaagcaatc gacacacaca cacagagatg tgtcgtcgtg agtggtaaag ggcagtgaaa  1680
gaatacgaac gtaaagtgcg caaaaaaaac attcaatttt cagtgcgaat ttgattattc  1740
aacgatgcaa ttgtatttga atgtactgcc ggttttgcac ttcccaatac acacaaacac  1800
acacacacac acacacacac acacacacac acacacacac acacacacac acacacacac  1860
acacacacac acacacacac acacacacac acacacccca cactgtcgtt cgttctgttc  1920
cctttttgt gaagtcgaga cgagccactc gagccgtcaa atggcgagga cacgcacgtg  1980
tgaaggggaa gagcggtgta atggtaatga gactgttgta gcgaggggcg ggaggggagg  2040
gtagatgaga gtagaaaggg ggaggaaggg cgagtgctcc attggcgtcg ctgcatccgc  2100
tgcagcgcgc ggtgtgtgca tccaagacgt tttcgcttcg gtcgttcaat aataaaaagt  2160
gtgcatcgaa accgcacaca cctttcctct cctctcctac gatcaacttc tctcacacac  2220
tccctctctc tctcttacac acacacatcc actcgggcga atcagctcca tggggcgcag  2280
acggctcttc gatggtgtgt atgcgttgcg cgccaccttc acgcacacaa cgaacccgct  2340
ccttataatt aatgcaacaa tgttgctccg ttttcattac ctgttttgct tcccaccgac  2400
agcaccgcgc tgtgcctctc ccttcgcacg ccctctcccc cccccccccct tttttgcatc  2460
gttacccctt tttgcgtcga tgcacttcca tcctctctct ctcacacacg cactggtatt  2520
tctttctccc ctcccgttgc tgcaacccac ctcaatcacc cccccccaca ccctttcgca  2580
cacttcgcct acagcccatc caactgctct aatgctacca tttccccgtt tttcgcgtac  2640
tgctgctgct tcggttggag agccgcgtgt tgtcatggta gcgtttgcgt ttggccgtct  2700
tttttgcctt catcttttgc gcccgcgtgt ttgtatgcgt gtttgtcacg catgtggtgt  2760
gtgtgtgcgt ctatgtgtga ccataaaaaa gcataacgcg acgaagtgtt tgctagcagg  2820
cggcggcggc ggctcgctgg gcagtgtcgg ttcgttttcg cgttttcgtt ttgacggctt  2880
gttagggcgc tgttcggtgt tgttgtggtg gcgccgtcgg tgtacgaaaa tcaaaacaac  2940
aaaacatatg tttttcggaa agttccaccc caaagggttg tgcgcgcacg gagcgccgct  3000
cggtggagcg cattgtgtat ctgtgtgtga gagaaacaga gagagagaga gagtggaaga  3060
gagggggata gagtgtgtgt gtgtgtggga ggcagaggct tgccgccaaa tattgttgca  3120
ttctgcgtgg cattgcgtgg ggttttgcgg actggtgaat atcggtgtga gcgagcgatc  3180
gtgtgtggga gggggttgcc ggacggccgg tacatttatc aaacgtgaga cacgtgcgtt  3240
tttttgttgt cgttgttgcg cttcatgtta tctgtgtgtc gcagtgataa ggttcgagca  3300
gctcagcacc aattgcactg cagagtggtg tgcaaaaatc atgttcgtta tacctacgat  3360
gaagttatca gtctggagag aaagatgcaa ttatgttgga taatgttgat tatttatcta  3420
acgagtcgtg tgacgatcag agctgataaa aaacactagc agactatcat ttcaatcagc  3480
ttaatttatt tcatttctca ctgttgctag ggctgtttag tatctcttct atttgtacat  3540
ttgtcagtgt agtgattgta acgaatgatt taatcaatga taaatgattg aaggaaagaa  3600
tcgaaaatga aattattttt tcttacaagt atgttaccct ttttcatcgt catttcgctc  3660
gcttggatta cagtcttact ctttggtata gttatacaaa ctattataac tattgattat  3720
aaattgaaat tagcataata gtattattta tcatttttct gcaaatattc tttggataga  3780
tttttttat cttactttga tgaattatgt tttgctcatt cattatttga aaatgtggca  3840
acagcttgta acagccgtta acttgttgca tagcaattca attctatact ttacaaaagg  3900
gtaagattgt ggcattaaaa tctatgtacg gtactcgcaa accgaaaaat ttaaaatcat  3960
ttcgattgta caaagtacgc aattacactc ttttttattc ctttacataa cttcctatca  4020
ttttcgtccg tttcatttca ttgcttgtta aatataggtt aacacttcgc tcaggatccg  4080
```

-continued

```
tttattgtat tgtattctat tgtactaaca ccagttttaa caccattttt ccattccttc  4140
ctgagatcct tcgaatagtg cgaaatttga tccttgagcg gtccacttgt ctcaccgttt  4200
atttctgcta atgttcaccg aggcacatat acacacacac acgcccccgg acacacacat  4260
tgatagttca acccttgtct gaatgattgt aaacgcctcg tatcaccacc ggggcgaccc  4320
catcccacat tgactgccct ttgcaaaaag aaaagagaaa agtactcact ctatccgtgc  4380
taagtgcaac agtgtgtgtg tacaatacgt gtcctggtgt gagtgcgagt aagcgagagt  4440
gggaaagaga cggcaaattg ggggtgcaaa atgtgtgagt gtgtgtgtgt gtgtgcgttt  4500
gtggggagca cgatcgtaca tgcatacacg tgctcggtcg tctccatcac gtacagtgcg  4560
cgcatgcttg tgtgtgtgtg tgtgtgtgtg tgtgtgtatg tgtgtgatgg tgtgtgtaaa  4620
agcagccgtg aagatgcagg gttcgctgcc gatgcaatga ggggggcaca ttgagtttgt  4680
gcgaaaatgt ttgccaaagc tcgatcaaaa gggcagcagt tcgttcacac ataccatcgc  4740
agcgttagca aacagccgcc actgctcacc ctgcccgccc tacgacggag acgagcggca  4800
gccgacacgc ggacagcgtt ccccgtgcgg gtatggggcc gacgcgacgc gctgcgagtg  4860
tatgtgtgta cgggcgcgcg agcgagacgg acggcgaacg gtggcgcgcg agcgagacgg  4920
acgattgact tcgcctcaac tctgttgcat tgcgtgtcgg cgatgcactt ggcgaactgc  4980
agtttgttcc gcagcatcgt tcccatcgca tcgcatcgcg cgctacaacc gagacgaccg  5040
tagctggcca cggacgagcg tcgggaacac atacaacact cctgtgctgt ccgccgtcga  5100
cttcgaaagg cacccaaatc gcgctcgctc tctctgtgtg tgaagcactg cagaagcgtg  5160
cagtcgacat tcgagcatcc gttcgggcag tgcgtgtggt acgtgcggca gtgcagtggg  5220
ccgccggtaa aagtgtatat cgttgctatg tcgacgatcg cctactaagg aaattgcgtc  5280
caatgtacca gtgtcagtaa cgcgcgtgtc ggagaagcaa acagccacgg cgaacgcaac  5340
ggaaaaaaaa cgtttgtaac cgcgttagtt gaagcgaacg agaactttag tgtgttgggc  5400
aggatttctc tgctaaaacc cggaaacttt acgttcggat cggtgagctg tgccgtgtgt  5460
gagaagagag ccttggcggt gacggcttgg ctgagaaagg ggccgcccaa taatcctgaa  5520
cggccgtgcg taaatagaga tagccgtgcg cgtgccggtg cggtggaatt tcgtgtggtt  5580
aaatctgctt ccaataaaac tcgttgacgg cgcttgacaa aatacagccg cccaatcggt  5640
agcagcggcc cagtcagtat cggactgcaa aaaaaaaact gccagttttg atagtgtgag  5700
gaagagtgcg gcctacgcgc acacgtgtag tttacgccag ctgataacgg tttcggcggc  5760
aggccccaaa cgcacaactc gcaggcggta cgcaacacag ttccaagtca aaaagcgtga  5820
aaaaacgcct gcatccccaa caaacacata cacgcatgcg gccgatagaa aagtaaatat  5880
tcaccaccgc ctggggaaat tgcgataagt gaagggcggt gaagacacgg cacagatatt  5940
cgattgaccg catatagagg cgcgaaaagt gtagaattaa atgggtagaa aataaacact  6000
ccgcgttgcg ttgtgatgtg tgatgtgcgg attggagcga gtcacaatcc tctggccctg  6060
cgcccgttgc agtgaaaccc gcgtggacgg aatgcaattt ttatctatct cgtgtgtgtg  6120
tgttgaaggg gtttgttgaa actggaaaat caattgtgaa acaaaaaatt atcagtgatt  6180
gtgatggtgt gtttttgttg tcgttaacag tgtgctggga atgagattaa gatttacgtg  6240
tgcgtgtagt acttgcctgg cgagcaagaa gatatgagat acccgctcat tcagtaacaa  6300
aattagtgtg atcgtgtgtg ttttatgtga ttgtgcagtg atgattgtcc aattaacgta  6360
aagatagcag atttaagaat tttatcaaaa ggagtgcttc aaaaatatat atttggtaag  6420
taaatatgca aacttttgtg aaatcctcct aaggacagtc aggccgtgtc gcttgaaaaa  6480
agtgtatatt ttccagggaa atcattagtc atttaatgat tgctagtttt ttttttaatg  6540
taaaattaaa taaattctat taataaataa attaaatgtg cagcatataa atgagataac  6600
gaaattattt attttctcct gacatgaaat tttgtaattt ttttttgctt ttcgtaacct  6660
taactatcga gaattttttt ttacaagacg ttgactaact ctaacgtttg tctaagatcg  6720
taatacacat cgcaatagaa tttggtcaaa atattccaca gtgatttaaa tttatgaatg  6780
cgttttgctg atacaattct ttaattgttg ttaattctat aagtattcca agtcgtacta  6840
acgttttatt atccataata attccgttaa tttggtttca atgcttttgg aatttcaaat  6900
aagctatatc cagcattaat gaactgaaaa attcaataac acaattttca ttattttcaa  6960
tggtgttatg ctttggtcat cctagcagaa gtgaaaaaat gctaatttta aatgttccaa  7020
tgttttgaaa tattacagga aatcaaatta atgtatatta tgtcttaaat aagatgttaa  7080
atggacaaga taataattag ccaaaatatt gcattacttc aaataaaata tgagatcttt  7140
gaaaataccc ccgtgcaggc aattggctac agcaagaagc aattgcggtt ctttgtcatt  7200
gaagttatat atatttaaaa gatatatcaa caaaaatatg ctttttaaca tttgttagat  7260
acatataaac attcgagaac aatacaaaat tatgtaattt tgaattttaa caccataaca  7320
aatgcaacaa acatagcctg tgtgttttgt tttcttaaca tttttttgtc atagtattaa  7380
attatttgaa atgatgtata tgatcccttc gatcgaattc taatgacact tgatcgaaac  7440
aaataaaata taaaatatat atagctaggc ttgtttaaaa tgttttatgg tgagcgaaga  7500
tctagtgtga ccttaaatta taaaacagct atttccatat caaatttcat tgtttttttt  7560
tttaatttca aagatcggcc atattgctat tcaaattttc ttttattctg aagaaatgcc  7620
agactgtaat gttcttactt acattaatta tcatgttcat tatcttactg tcatctgtta  7680
cctgtattag gtccggttat ttaggtatat tgaaatgtta aatgtaattt tacgttggaa  7740
cgcctatatc atcttaatga attaagttta atatgacaaa aattaagacc ataaaatttc  7800
taaatggttc tttcggtacg tttgattgca gatctcccaa accctagcac catcgcttcc  7860
tcgaccaacc aataccgaca gcccgagaac gatcgtaccc gagtggaaaa cacattgtat  7920
tttcgcagca aaaacaacac agaaatcttt aaatatttta agataaactc catgtcccga  7980
caaatctgct tttttgcgat tacatagtaa agaaacacag tagtgaggag cttacttttg  8040
ctcgtgctcg taccaccttt taaaaaaacc cggagggaca atgccgtcac gcaccacggc  8100
caacgatttg cgcgagctcg atgtagcgcc ggcaagtgta acgttagatc aagcttccag  8160
atgttgagag tcggagtcac aatacgtcca caactgtcgg ttcgtccaat ctgtacattg  8220
tgtggtcggt gtttggtggg aatgacaacg gtgtgtcctc ttcgaaggtg ctaaaaggaa  8280
gctcgctgac gaggcggtag ggtgtgagag tttggccagt ttgttgttgc gcttgtgtgg  8340
ggtgcagcag ggaaagcatt agccgagagg tagagacaca caagctattt gggaccgtga  8400
aatacgccgc gcgcaacagt aataacataa cgtaccgtaa gccgaagcga tcgaatcgtg  8460
taatcgaagc ggtctcgtgt ttttttcctc ctatatcgag aggccaaccg atacatccag  8520
gtgcattcgg cggcatagat aacgcagcat taagagtcgg aattggctct cgaacgcaac  8580
agtttgattg atatataggc aaggcgtagt cagagaggtg ctgtaaacga gaagaaagta  8640
```

-continued

```
aggctagcag gagaagcgca agttgaggag gggtgtcgca gggttgacgt agacgtagag  8700
cttgtttgga agacatacgc ggaaccacac gggcgtgtgg tgcatcttga atggtgtcac  8760
aggaccgctg gacggaagca atgtccgact ccgggtacga ttcgcgcacg gacggcaacg  8820
gtgcggccag ctcgtgcaac aactcgctga acccgcggac gccgccgaac tgcgcccgtt  8880
gccgcaacca cgggctgaag atcgggctga agggccacaa gcggtactgc aagtatcgcg  8940
cctgccagtg cgagaagtgc tgcctgacgg ccgagcggca gcgcgtgatg gccctgcaga  9000
cggcgctgcg gcgcgcccag acccaggacg agcagcgggc actgaacgag ggcgaggtac  9060
ctcccgagcc ggtagctaac attcacatac caaagctatc agagctgaaa gacctgaagc  9120
ataatatgat tcataattct cagccgagat cgttcgattg cgactcctcc accggatcga  9180
tggcgtccgc accggggacc tccagcgtgc cactgacgat acaccgacgg tcgccgggcg  9240
taccgcacca cgttcccgag ccgcagcata tgggaggtaa gtacgatcat gcgtcttcat  9300
ttcttcgttt ttttacaact gcttcagtct gttgaggatt taacacactt tttcatacat  9360
atttaccatt gggatacaaa ctgaggctct catagagctt cttcgaatgg ttcgaatcat  9420
gcaccgaaaa cacttgcaag actatgattt gctccaacat cacgcaaagt ggatcatctc  9480
caaagtgagc gcatctttaa tgcttagatt gcgcaccaga gatcctccag ttcccacgga  9540
ttgggcctgt gctacatttt attggttcgc ttaggcactg cctcaaattg gagcatctca  9600
gcacggtacg cacgaggaac ggctgcactc agacaacggt cggaaatccg tgcaatcccg  9660
ggaggggacc ggttttaatg ctgtttggtc tacgttgcct cgctaaacct accttccggg  9720
atctctgcaa catttttcgc tcacctgcca cttcgttaga ttgtagttcc cgtcgcgagg  9780
acagtgccgg gagttcggtg gagcaatgcg ctaggctcca gagaggaggc tacgaatgcc  9840
ttggaatgga cgctacacac tctttttgtg cgtacttcca ccacacgtta cctcgacgat  9900
taccctggtg gcctggtgtg cctggtgttt ggcgtttacg tctcacttcg tatgtgtttc  9960
acccatcacc cttcgtttcg ttgttggggg ctctgctttt tttctgcttc tttcgtactc  10020
cctctcacac cactgctgct tgctccagca cgtccgattc ttttttcgca tcgtattacc  10080
ataattatat tatttaatta tctacttctt ttcgaacggt ggcgttggag cccgtccctc  10140
tctctctttt tccctctttt ccctctcttt gtctggcact gtgttcgttt gttttacttg  10200
tttgcacgct tggacaatgc ttgtttctta tgcatcatcc cccattggta cattctttag  10260
caagacgcgt atcctttcgc ctgcatgcag aaccgtttaa gtgcgcccag gtccggagtg  10320
agacgaaatt gatcagaatt cagacacacc tcgttatggg gccgatgatg taccgccatg  10380
ctgtcggacg cattggtttg gcgacgaagg tgtttcggtg ccctggtact acaaataatg  10440
gcaaacggtg cactggcgta tgcgtatgct tcttcgcccc ggttcgtttt aaacggatcg  10500
gtaatagtaa aacaacacgt aaaagcgata ttttgtagtg gactttggta aacaataagg  10560
ttccggctgc agttggatct tgtttttcta gctacggaat gtccggtgtg caaggcagac  10620
gttcttcagc aggtcctgtg cgtgataaaa cacaaaggga caaacttttc atttgctcct  10680
atttgtacaa ctgcgtggaa cacacctcat atacacgcac acagggtacc cggggaaaaa  10740
tgtcgtgtcg cttccttgga cgattggtat gtattcggaa aaagaaaata cttttcgagc  10800
tcgtgtgccg ggtggcggtg gctgccgttg ttggaacggt tatcgccaaa ttgctcttaa  10860
ctttgccact tgtgcaatta ttacttgtta tatcttttcc tgccggctgg cttctctcta  10920
tttcccccaa cctactctcc ctttcccttc ctttcctcta tcgccgccat catgccaaag  10980
gaagctgcag tcagcactcc ctactatcgg ttgaatgtgt gtagtcaaag attaagcgtt  11040
gcccgtatat gctaaataaa agtttgcacg caattccacg cttttcctcg ccgcctgcga  11100
acggtggggt tttggtggcg gggcaatgtt ttcttcctgc acgagaggac gattagttga  11160
ccttactgag cgcacggagg gaacgcagga gtgtgggtag ggtaggttac tgaatgacca  11220
cgtaagagac gtttttgctt tgttattgat tatttttcag aggaaacaga acaaaatgag  11280
caagttgaac atttgattta cattcttggg ctgtgagatt gcattagatt tgtgttgagc  11340
tgtttttga aatgtaaaat tattagcaat tactgaaggt ttgctgaaag gagagctgaa  11400
gaagtattct attgggaaat atatgtctat aaatgtgcaa aatactttcc cagaagattc  11460
aaaaggctcg gagaaagatc ttacattttg tgttgtaaat gtgatcattg aaaacctcac  11520
aacactaaat atacctagta aatttaaatt tttaacgata ttgcctacat aaaacatcta  11580
gagtcttaac atcgcttaga aatgccgttt ggtcccagct accaacatgc caacacgggt  11640
ccggtcagca ccaaacccgc ctatggaagc tcatctttgg cttgttttta ttgttttcat  11700
cccctctaaa acacattccc ggtgcggcat gttaaaactg tcattagaag ctttggcgcg  11760
aatcgcgcgc gcccgctcag gggtcttgca aacccgttcg cttcagcttc tggctgtgtg  11820
tgtgtggctg ggcgtaggta cgaatttgcg gaatgttgca gaatgtgtcg ccagcaggac  11880
agtgcggtgc ggtgtgcatt tgctagaaca ggtttcgcga aggaagaacg tttgctagct  11940
ggctgtgtaa ggcttttgaa ggtatttgat tgattacgac cgccaacgtt catcgttaat  12000
catgcgcccg ctcagaatag cctaccagtc atgggtggag gagttcgcgg tggagttctt  12060
tccaggcaaa gcagggagct gcgtgtgacc cggacccgct tgcacattgt tcgacagccg  12120
cagtcgctcc atcgaatgtc cctggctttg ctggccggct ttgcgcaccg gctcgctctg  12180
gcgcaatgag ttcaattttc gttgcgatcg tgaaaagatc gcccgaatca tccggtagtc  12240
tgctccggtg ctgcaactac ttattaagca gcattatgta tcttacagct cattaggcgg  12300
cgtcgaagga gcacatcagc aaacaaccgt accgtaatgt cttaaatgcg cgtttatgat  12360
ggggtgacgg acctgacggc atggcggccg ttgcttttgt tttgattttg tttttggcac  12420
ttataaggtg tggtggggtt gggcggatgg ggtcccccaa acaggtaacg actttgaccg  12480
tcgccgtaac tggtcgctgg tcacatgtcg aaaggtggag ggctgcacta tcaaatgtca  12540
ctgcatcgaa acgacgggag gtgttgtatg tgtaccatgt tactgtttgt gtgtgtgtgt  12600
gtgtgagtgt atgctggcca atgttgcaga ggtttttgcg cgcgtacgat cgccctgtaa  12660
ccggtttgaa tttttgcaca catttttttg tgtatttcca gcatcaggtc gcgctggaaa  12720
aggtgattcg atcccatttc tcttcgctcc aaaatcgagc gcatgcacct cggtacgcgg  12780
tatgtgtgtg tgtgtgtgtg cttacgtgtt tgatgggtcc ggttactgcg cacataaatc  12840
ctcgacacag tcggacaagg gctctcgtgt ctctagtttt tggcgatggc ttttcggccg  12900
ctcgcgcgca gctcctgacg gctccgagcg gcgatggtgt tgattgagtc atttactacc  12960
gaagcaccga tagagatctc gttggtggtg gtgtgcgcca cagatcttga cgacagattt  13020
tttggcgtcc gtagaagctc atttcacggt gcgatgaaga cgaatggccg gctagagagc  13080
gccgagtcgc tccgagcggt attgtggtca gagtgagtag ctttgtcaag gcgtcgttac  13140
cctttatttc tctcgcgatc ttcgtttttt ttggttaatc aagaagggga aaagaatgac  13200
```

-continued

```
agcaaactag ctgtttgaga aaagcggagg gttggcttag cgacaagggt gctacataaa  13260
aaaagaaaca gacaaagagc gtgtttaatc cgattgttgt gttgtttccg gttgagggaa  13320
ccgccatgct ctgccttcca aacttccgca ctaaacaaca acttcctgcg catgaggact  13380
atcactgccg caaggcgcac atctgaagaa gcccaaaact cgtcgtcgaa acaccccaaa  13440
tcaaaggtca aacatggcgg ttactgcttc ttcttgtaag gccgccgtcg tcatgctttt  13500
gtgccgtaca ttgacacctc aagtaaaaca gagcagcggc tagcagggac ttttgatgaa  13560
cactttcgtc ctcgcctgat gagtggtaga ggcacgcaag catttcagtt tttcccctcc  13620
tgtcgaatgg tttttcgccc catgcgaaaa atggttacag tgttcgaccg tgagtgagtg  13680
atattttaaa agatatttca catttactgc tgctcccttt cctgcgctgc gacgagcgca  13740
ctcgctcgta catcccatta gcgagcacgc ggccctacca atagattgca aatgcgcctt  13800
tctgcgggcg agtcatgagt gagacatcta tgacggatac catgtggaca aagcgtaaaa  13860
aatgcacaca aacacacaca cacacacaca cacacacact tgcactacgg caaagatcat  13920
cttttacgcg caccgcacac cgatcgcggc agcgcccaaa gtgcatagcg atggtggagg  13980
cttgcgtttt ggaacagacc gcgcacacgg gccgccggtg tgacgtgtgg aatttcagct  14040
aattagaaaa ttattaatag ttccttgcgc acatgatcgg tgcgccattc ttcttcctgg  14100
ccaaagtcac ccgggttctg catttccgga gcagagtcct cgacaggttt tcactttccc  14160
tgtcacacgt ttgagtgtgc ctatgtgtgt gtgtgtgacc ccttctcgtc ttgtgccttg  14220
gggtcggcta gcaatttcta aaacttgctc aatggcgcat ccttttcctc tctgtgcgga  14280
gaacgttttt ccgcgaatcc atcccctcgc cccaggtgct tatgcaatca gcgctgcttt  14340
acaaattaaa acgtaattta gatcctgttc attaaggcgc gcgcccgatg cgatcctttc  14400
cccgcgccac gcggtgcaat taaaagcgta tttgaataat ttgattattg tatgaaaatc  14460
aaagaaattt gtctttaccg gcaacaaagg cttggcatgt ggaaaaccag cacaccgaca  14520
gaacaggcct gtgggaaaac ggagaacaca caccggcaca ccaaactggt tctttccggg  14580
tgcgcgcgcg acagcagatt acatctggtg acacgagata atttccattc cgcgatgcgt  14640
tttgcgctgt ttggttgttg tgcgtgtgtt cggccgaaga ggagggggg gggctttgga  14700
cagcaaatgg cttgttaatg ggcttttacc tttgagaact gaaccgcaaa accctgccga  14760
acaggggtga gtcttgagac agtctatcgt cgaagctgct gcgcgttcac ttcctcatca  14820
cgcaagctgg cgcgcgcaca cggcctttat tttggcagct tcaatcggaa agccagcaca  14880
cacacacaca cgttcgacag ctaacgagaa gcagggttgg gaccaccgat tagagatgtg  14940
caatccgcgc tgtgcacttt tgcatcgtcc acacaccccg cggacacttt gctcgctttt  15000
cgccccgttg ttctcggttg atttcgccgt tcggccgccg acttcgattc cctcatacgg  15060
gtggaaaccg aaaataatgc gcgagttgcg ccgccacccg cctaaattta gcaccacgag  15120
ccggccgcga gagcggcaac actgttgcgc ggccaaatgt ctattttcgt ctaattccgc  15180
acagcccgtc ggtacgctaa gccgtattgc ggccccgccc ccgctgtacc cgccgatgcc  15240
gatcgcggag caatgtgcgc acttcttgag caactagggt gcacttgcac ccctgtcgta  15300
ctaacctttt ccgtgcgccg tgcgctctcg tgcgcactgt tcttcctctc tctctcacac  15360
aagcgcataa aatgtgcagt ttgcgggaca gatgtgtgtg tgtgtgtgtg tgtgtgtgtg  15420
ttgcgctttc cggttcgtta cgtgtgacgt gtgtgcgcgc gcgccattgc taaagcgatc  15480
gattatcctc cgggagcgct gttctgttcg ctcttgttct ttcaatttta accaaccaag  15540
caacccaccc acccacccac catgcacccc gctgcctgtt ccacatgtgc atcagtggtc  15600
agcttgcatg ctcgaatgca gcaaaaaagt gcaatgcaga gagtgcagca aaaacaaagc  15660
acaccatgcg acaatgcaaa gatgtaaaag tcacacacct ccaacgaacc gcaatagatg  15720
ggatggcccc tgctgggacg ggcaacggga gaataggggc agcgatgatg attgatacat  15780
tcatattcgt cgccggagac cacccgggcc accgtggcag cccttggggg ggaatatgag  15840
catcgcgtca cgtcgtactt aatcaacgcg tgtgcgttat ttgtctgcgg cacttccgcg  15900
tgcgtatctg tcgtgtccgt tcggttcggt cggttctcgg ttggccgtcc cggtgctgga  15960
cacacgcttt gcgcgattgc ggacagtctg caaacggcaa cggtatggtg tgaagaagtg  16020
gttctttttt gtgtgcttct tttctttcgg aaatatgaaa tttcttccgc tgcctgcctg  16080
gacgccggga actggacgaa cacaggcgcg gtccgccgta ttttgccatt ttcgctcgga  16140
tgtggtcgga tgtggggcca attgcacaca caaaccgcgc gaggtggaat gtatttattt  16200
acgttttaac ggtgcagctg tctcctgccg gtgcatttcg tgaggttcct tttgcccatc  16260
gggagtgttg tgagaggagt ggccgaaaca aaacggaccg aaaaaaactg ccacagcaac  16320
agttcgaaaa gcacggacgc acaaaaacga gatcgctcgg aaaagtgcaa ctggtggcga  16380
tggtgcatta tttcacattc ttttggccgt acgaataaaa acatgaagca agtaccatgc  16440
gaaaattgaa cttaaaagat ccacccgtaa cggttgcacg gcagagcgtg cccgagtggg  16500
acgtgcgtta aggtgaaata aaataaatta actacaaatt tacaattaaa ttgattccat  16560
ccattgcaca gtcgaggtct ctgagcagga gtactaatat tctaccggca ggtccgtttg  16620
caggctgcaa caccgtcgtg cagctttccc ctcgagcagg cagttagtag gcaaagttta  16680
tgtgctagat agcggtggtt ttgcggggag aatcaagtct agcacacaca aacaaacacg  16740
ggtatgtaaa ggttgaaagg ctgtctcagg ggaccgagtt gccgattggg cgctggttcg  16800
tccaccgtcc atcgcgcgtc ctgaacggaa acaataacac tcataataat gtttcaatta  16860
aacacaggcg ggacgacgac aggaaccggt tatgatggga caatttcaca attgcacttg  16920
acattgggcg cagaattggt ttgcaccagc catccaggga cagttgagca ttgcccagtt  16980
tgagcctttg gtctggagct tttacatgct aattagattt cagttagaca actctgcgca  17040
acatacgaat gctttcaata tgttgcacaa gggcacaatg ccgcaacaag gtaaatgttt  17100
cctgtttcta taaaacagac tagacgtact ttaaccaagc tatggacaga gtctattttc  17160
ggatgtcata atttacgttt gaatgatcaa tcacatttag tgactgctaa acctgcttgt  17220
tatgcttatc ctgtgtatcc taacgcttaa ttgttccgtt gtgtcgttaa actagcttaa  17280
agcttcttga accattgaag ctaccattat gaatgcagta taagcatgca agatttattt  17340
cttttcttcg tttcgattat tctttcgtaa aaggcatctt gatttaatga atcttttgcg  17400
ataatcggct acacagcatg gcatctgcgg ggcagaacgg tactcgatcg agcagtcgcc  17460
attatctagg agtgcgtaat caagtttagg ttgccacgtg attcgattca tttcacaccg  17520
acatgacagc agaatagaat acgggtgcgc cttgccgcac taccgttgac cgtcgcgcga  17580
gaccttctca atggctgcat tcatctcgct gctcgcaagt gcgccgtgag tggagcataa  17640
atctcgacaa acgttattgc atttcatcga ctgtcttcga tcgggtttgg gggggctgg  17700
gtagacattt aggaagcaat aacaactgtc ttatcgtgca aggaaacaca ccggcacgcg  17760
```

-continued

```
gctaagcctg tggtgcagtg gtttagattc ctttttactt ttacttacca ccgcacatgc  17820
tttatgttgg atgttcaaca ggcagcgcag acaggctgag agcggtacag catacacacg  17880
ccgtcttgct tgatagacaa ggcttcgcgg cctggcattg ccgtggagtg acgtgtaagt  17940
agtgccccaa aggcaccact cttcacggga tagaattgag tgcgttgatg tgaacggggg  18000
gcgaggaagc gtagtgccgg ttgtcgtcgt agttgcagct tctgcccgag cagcactgtc  18060
aaaatgggtt ttgcgctagg ttgagaatcg gaggagggcc ttcgccgtag aagccgtagc  18120
gatcgtcctc cgcgagcacg ggacgcaatg ttgccacaca ttttgccgcg ctttttttt  18180
gcactcggca gagttacgac ggctctccgg tatggaagcg agcagcacat ctcacgggct  18240
gcgtcgaaaa tcgagcataa ttgtatgctg tctgatctat ttcatttcgc gttttatgtt  18300
ttattcgact tgctgttttc cgccgcccgg ctcagcttcc aggcagggcg ggaggctcat  18360
tgtaggttag ggccccgttt gacgtgggcc agacagtcgg cgatggggcg aatatgggga  18420
gaggttggtg accgatccct actccatcgt gtcctccttg aggactagtt tcgctctccg  18480
acactcttga cacttctctt ccttcgtctg atcctctcca gggaaaggct gctgggcgag  18540
aaaaccttga gacgcgggag cagccagaaa ccggctcctc ctgtgcagcg tgcaacaaac  18600
aaaacagcaa aagattctag gctccacact gtgcactact acgagagaga aagagtgtgt  18660
gtgcgtcctg gggtagttct gtcaatgttg aaaaaggtgg caatggaaga agagctagaa  18720
aaacagaggc attatggggt gtttcaggca ggaggattgg tgggtgttag gccgggcagg  18780
aaaccggatg ggaagtcgaa cgggatacgg atgctgctgt tacgccactg aagcggaatc  18840
gtttgcggaa tcggtcaaca ttgttgagat ggccgtgttc agcctgcggt tgatttagtt  18900
actttttgat tctttttga ttcatttcgt ttgtgtgtcc aaatgaagtg tgctgttggg  18960
ccggcagata gggctttcgg cgggtacgca ctcgagagtt cgtgcgcgta tttctcgaac  19020
gtcacggcat accctcatca agtgaggctg tcccgcgata ggtcttgtgt atgtgtgtgt  19080
atgtgtatat atttttaaat tctggtttgg ggcatcagga ccctgaaaat gtaccaccga  19140
aacccaacgg agagacgagc ttgtctgaga atggttggga gcgcaagcag tggtgcttac  19200
gatttataaa ataaacaacg acgtacggat accgtgcgac gggattaagg tcacgttcaa  19260
tgttacgatt gtcgatcgag acaggcatct taagcgggct gaacggcttg gtcacactgg  19320
aagggattat ttaccgatat aagcgatttc accattggcg ttgtccgtaa tgcgagggcg  19380
ccgataagct gaccgaagca ggcgcgaaga gtatttttgt aacttggttg aagaaacaat  19440
cacaagcatc ttgatgataa gggataatga attaaacata attgcatcac ctgtgatgag  19500
acagttgata aatgggacgt ctcgcgaaat tctggaaagc gagcaatatc ttcgtacagc  19560
tgcatctgac attgacgtgg ctgccggttg cattgcgaaa cgtcaaaggt ggcgctaaaa  19620
gtacatgttt aaaattagtt tccattttgt ttgtttgtaa tgcgctccgg tttgtgtgca  19680
tgtgttcggg tttttagcta ttaactgcaa tttctgcact gcaaaatgta gccgttccgg  19740
tatgatcagc tgcagacacg tggtggacgg atcttctgct tcgcgcaaag tgcacttaaa  19800
tggtcgtcga aggagtggac agcgcccgcg tctgagctca taatcggcag gccaattatg  19860
tcgacgggaa tgtggaagga tgcttgctgc agcgaacaag atgcattaag catgggcaat  19920
caatcatccc gtggctctgc aatcgaggtt tccgtgacac acacgcgcgt ccccgggtgt  19980
cgtcgctgac gatcgcgtgt tttacaagtg cgtccgtgcg ttccgtacgt ccgctgcgtc  20040
gccgtcgtcc gagccacaac atgcccacgg ccaataatca gtataattcg gtttaacgtt  20100
tggttagatt atcgggaaag aaaataagcc gaggtaaaaa cggatcactt ttcaaaccga  20160
accgagcgca ggactgcaaa gatgggaaat gtgtgttcac gtgttgcgtg cgtgatccag  20220
ggtgtatgtt gcgagaaatt attggaatca ttccaaagtt atgtcggtaa cctcagcgtt  20280
tttcgtgcgg tgtgtcggtt ttatgcagaa agcagagatc ttaaagcgag ctggcatttt  20340
gatatagcac atatattcga tggatgtagc attgaggtat cctcaatgac cattctaaat  20400
tatcttatcc ttaaggctgt ttttgggccg agtcctgcaa gactagaaaa agtccgatac  20460
ctattctaac tgtcctccca tgtacacgtt tctgcatcgt tcctggaagt catggaagtc  20520
atagagagtc attcagtttc atcacagaaa cgaacagaac attgccatca aattggacag  20580
tttcaaaact tcattcaagc aaagattaaa ttctagcgtt agctccataa gatattcgac  20640
ctccaggtta agttatattg gtctctagct aaggttgatg tattgatatg gtcttcaaac  20700
ctctactaca ccctaaatat ctttgtcaaa gtcgttaact ctcacctggc atgtagagga  20760
acaggcaaca gaccaatgat tgaaaagcca cgctcatgtc ttcagaccat aacctcggcc  20820
aaatttacct tccaatccat cgataaaacc tcatcgttaa tgtcattaac cttttgcaaa  20880
gcttttactc cagtgccacc aacaaacatt gcgtcaaaaa acgaccagtg tcacgttctc  20940
ctccctgtgt atcggagcat ctacgaaaaa aataccaaaa gcctccctta aactgggagg  21000
cccataattc cagctgaacg cttagattgg aacggaactg gcggtgtctt tcgtagggct  21060
cggaacgttt tcctaccagc ttctgtttgc tcgaacccga agcagagcac aaaccgtcta  21120
ggttagctga cagaagaaat tgcaagatgc acaaaaaatc gcacacacat acacacagac  21180
gttaacagtg tattgcgacc gaacgggcag caaaacgctg tggctattgt gccagaccag  21240
aagggaggag aactcaaaaa cggtaaagct aataaacctg tttctttcca ttttttgcgc  21300
attgattcat ttcttgcgcc ggcgagagct gcccggcagt tcctgttgca tacatgcagg  21360
gagcgcgggt ttctcgatgt gcgccacctc tgccgccggc atcgccacca ccgtcaccac  21420
agaccggctc gaaggctgcg ggatgcaagc gcggcaacca ctggaaggta acctctcggg  21480
gcgattgttg tatttaccaa tcgtgatgca tgatcaatgt tgtgcggagt attttatttc  21540
ttgtaagcag cagtttgagg atcggccaga ggtttgggta aacatttcag tcgctcagtc  21600
gctcgcgaaa cagaataaaa aaaacgcaca cagcgttcaa gagaaaggcg cgcatggcgg  21660
tggatgtaaa atgcctcatt tgtggcgtct tttcccctgc gcgcagcaga acgtgaatgt  21720
gtgcagagca tggtgtagcg tcggacgagg agcatgaatt ttgagcaagc ggagatggtt  21780
ttgagtaaat cggtttctat gcagccaagg caacggcagc cgcatagaac tagagcactg  21840
tgggccaagt cgcagtcgag gcacggaagc agggcagaat cgcgactctc tatcgccctt  21900
gttggacgac ggataggacc gatgccggtg cgggtcaagt tcagttggct taccgatgca  21960
tcatcggaag ccatcttaag taaatggaga gctggttggc gatggagcat ggggctcgct  22020
ttactctttt gagtgggcac aggagtgttg tgctagaaat agattcggct caaattacgg  22080
ctcgggcttg cctagagaaa gggcaatgaa ggattgaaca catcaaagtt aagtattttt  22140
tgtatttgtg gttgctgtcg ttaaatggtt tattgaagcg tttccattat aaaagttgtg  22200
aaacagttgg aggatgaaca gaaaagcgtg gatgtggaat tatatttcaa tacaaacaca  22260
ttgcacatga tcacatggat caacggtata taatttagtt ggatataaaa atgcacatcc  22320
```

-continued

```
agcattgagg atggtatttt gccatcctcc acagctcatt atgttcacaa ggtgatggtg  22380
gcgatggttt cacagtaaaa gtttctcagg caaaacggct gcgaggcatt gtgcgaaagt  22440
ttgcagtacc gtgttctatg ttcacaattg ggttttaaat gccccaaact gttcgaaccc  22500
ttctcacatg gagtgtgtgt gtgtagctgt gtgtgtcaag gaccgcaaac aggaagggtc  22560
aagggacaag ggagggcttg tgatcggaag cgcaacagaa tcatgatgag cgcagactgg  22620
caccgggcat aatttgcccg ttttttatc gtgtgttgcg cattacggcc ctatgttgaa  22680
ggagatcgtt ttcctcccca catacataca cacacacaca tcgatcgtaa ggtatgcaag  22740
aggaatgttg ccttaacact gcgcgagttc ggttgcagtc gatagaattc ggtggtttcg  22800
agtgcgtgca gcgcatatta acgccaaggt tggtcaagtc gtttttcaac gccccttgaa  22860
ctttggtgat gcgagtcaag gaataagagc aagaaaacaa acactccaca gaactttagg  22920
atgcatggac gctgctgcag tggcggtgat ggtgctgttg tttcgtgtgt cactgtaaca  22980
cggctcatta acggctgcag acacagcgat tgtgtcgtct gacgagttta ctttaaatta  23040
gcgatggcaa aatcaataga aactttcgtc gccgccgccg ccgccgtctt ttgtattgat  23100
ctcactgtcc agcgaaacaa ggtattagca cgtcacgatc ttatcccgat tcctgatcgt  23160
gtaaggttta cttactttta atgagcctaa aacaaatagg aacaatgctc gtcggaatgc  23220
tctgcagcag ctgcgtactg tttactgtta gtgttcgctt gtcttgcgat gttttgcttg  23280
atcttaatta ttaataaggg cgcggtacta tttgtttgca aaaagtcttc tataatgatc  23340
gattgtattt tttaaatgag atgtaaagtt aaaatatttg cacaatataa acatcaaatg  23400
caaaacatgc taaggaagaa cgtaaatatt tcgtgtggaa tagttccttt ttatttgaag  23460
ttttcaatat gagtaatttt taaaaggcac tttgacatat ttgttttcac caatgttaca  23520
gacaatctat caaatatgcc tataatttta tcagataacc tgaaatcttt tgcaagatgc  23580
tgttcagaca atcacttcaa agtttctagt gatatttgag atttagattt gcatttaaaa  23640
tcgtgcacag catagccttt tatgcatttt atgtaaatcg caatcaccac accaaacaga  23700
ggcgaaacag attgtaatat tttcatttaa ataacatccc ccgaccaccc atatgtgtgt  23760
gtaatcgagt gaccttgatg cattcagcga tgcatggctt ggcatagagg ggaccacaaa  23820
atcgggacgg gcggtagggc agtgctagca caagcgcaga aaattgcctt atcaaataac  23880
aaacccttc tcctcatggt tgcatccgca ctgccctacc gcgtcgaccg atgcatccga  23940
tcgttttcat gcctgaatca gttggaaaaa cttctctctc gtcggcgtcg cgaatggaaa  24000
agcgtttcac aattgcttcc tactgtgacg ctcgacggcg tatgtggaaa aagggtgcgg  24060
tgggaggcgg gatgtggaga ggcttatcgt cactcactct tgggtgtatg cgtgtgtgtg  24120
ttgttcgcgg gaaagcccat atcgtaatcg atatgcttgt tagagatccg ttttgatgca  24180
atggaaaaac taacgctcca gtctagagac caacaaacac acacacacat cgaaagagaa  24240
agggaaatgt gtgggaggaa gggagaggag gggtgagagt ggaaatgcaa tgtagtgtga  24300
aagtgtggct gactggttaa atggatggga aaacaaggaa atggatggaa aggaaggaaa  24360
aaaaaaccgt ccgacggtta cagaaagacg caaaagtgct cgtacgaatc gtcgtatcgt  24420
cgttggcgaa caaacaggcg aagccagagc ctgccagcaa cggagttcta cggagctgac  24480
gggacggcca gtccgccggt gtggtggatt tgtttggaca gaaaaagatc ggaacaggag  24540
aaaaaaacgc acgccttcat aatgaaatga tagacacgtg cacgtttcca gtttcaaatc  24600
aatttcacac tcgaagtgag aacaaacctc ggaaacagtc gcacatacac acatacacat  24660
tgggatggtt ggctggtggg tggttttggt tcactttgct ctccactaca tgtccaacgc  24720
tgctgttgct gcgtatttca tctgcccttg tgaaacgaat caccagaagc ggtttgggtt  24780
tcgggagctc atgttgtgtg cgatgcgtcg ccagtaagca ttctcgcgga aacgataaca  24840
aatgtgtgtg tgtgttgggt gggagtgaga gagaacatga ggttgggggc gaccatgaca  24900
ctgacctagg acaattagaa actgattgac ggaaacgata tgcatcgaaa gcgagacgca  24960
ggttttcttc gttttatcag acgcaggccg gccttagaca cgtttactct agggagtcat  25020
tttgctgagg acagtgagca cagcactatg taggttagat gggggggcgtg gtgggagctt  25080
ggtggtccgt tggatttgaa gttgccagag gacaacgatg aaagtaatgg ccaaggatca  25140
gtgcgaataa aactcatcct tgcacttaca tacacacaca tacggtcctg tgttggattt  25200
cgcaggacat tgcgaaatgt cttcggtgga ggttttactg gccacgtttg atgaccttcg  25260
gcattgctgc cctggctgtc ggtttcggtt gcccggttcc acatttccgg tggctggctg  25320
gagataatga acatcaattt caagaacggc aataatcgta aaatgcaggg aaatatttct  25380
tgatgcattc ccgggctgga tcttgaagaa cgcgccgcac attggagttg atttgagcat  25440
gggaaaactc ggagcgccgc ccgtgccagt acggctgtcc tccgctccgc gttgttacag  25500
atcctggcag ttcatacatt ttcatcgaac caaccagaag catcaagcca ttcagccacc  25560
accacgtacc acgagatgga tgcaaaggaa ggacaaaaac aaatgtaaag tcgcccagaa  25620
caatgtgcac tgctcgcgcg agtcctgctt ttcgtctccg gtgcgtctgc tgcctgcgtc  25680
ttgccgaggt cgggaggaag ccagcacaca cacagagtct tatgccagtg atgatgcacc  25740
acaatcaatc ccttctatgc agaccgaggg gatcaatcta ggttggtttc atttttgtt  25800
tctctctccc ccttcatact cgttttatga ttagagagct tttccgctgc ttttcgttgt  25860
gcgccgtgct gtattttgtc atgcttttgt tcgacgttcc cttgtcactg gaccgctttt  25920
tttctttcct ccttccttcc gcttgtttcc cgtggcaggt tgtttttgtt ttcgaacgac  25980
tcggatttgc catgtataga tgcgctcagc ttttacaaaa aaagacaaat aaaacacgaa  26040
catacgagct aaaaacaatg cttttgatgc acaacaatca caactaccag cgctcacaca  26100
cacacagaga cactctctga cgcacatttg tcgcttacgc aaagggaagg aaagaaaatg  26160
ctcgaatgct gctgcagctg ctgcctggga aaagaaattg gatggtcgta aatttcgggt  26220
tcggtagaag gaaagctctt ccttgtttca tttacagtgt aacagtcgca cacgttggca  26280
ccacgctgcc atggtggtgg cgtgtggatc gaaaattgag atgaggtttg gaatttttcg  26340
ctacataaac tttatcctgt gctggtgtgg actgtttgtt tctgttgccc agttttatga  26400
cgtcccggaa acgcggacaa gcgaaccgtg cgaccggcta attggtctca tccgcctcgt  26460
gatttttccg accaaccggc tgcaatacaa tttgtccaac catcgtgttc cgccggtggc  26520
tgctgggata agcagaagaa cataaatctg attgaatgcc atttcaatgc aacaaatttt  26580
aggaaaaatg gctaaacaac tccttggcaa gcttctggcc aagagtaaag gtaaacaact  26640
tgccagtact ggtcactctt ttgtccaccc acctttccgg ttgtatgtgg attgatgcat  26700
tttaagcata atacattatt aactccacag acaaacaacc ccgaaatggc ttcagctcag  26760
cttaaccagg cggcaaactg atttcgatcc gcacgacatc atcttgcacg ggacgagaaa  26820
ttgcctccga tacctccagc gcggcgtcag tcagccatct ctcatatttg ctctcttaca  26880
```

-continued

```
aatgatctca gcattgcctc agtcgggccc tcagtcgcgc agctcgacgg acagaaaagt  26940
ggcgatgtga aatattaatg ttaaagaatt catttttaaa tatgcaaatt ttaattaata  27000
ttcaccctcg ttcccttgtg gggcaaaaac gcgggcctcg ggcaacgaga ctctgcaggc  27060
tggtagcaag gtttcggtca tctgtaaatg tgttctcgtt aggcggttgc gaaaaacagg  27120
ccgattttgt ttcaggacag aacaggaggg ataaacatat aaagagagag aagggttaat  27180
gtagaaacac aatatgaagt tattagtgtt attgctttcg accgatggca gtagatgccc  27240
ggtggatgca tcaaatcatg acttcgacag gcccaatgtc cagcgacagg ggtgcattaa  27300
aacaggcttg attctggatc ctttaactac acatacaggg tcggccagat cctgaaaggc  27360
ctctacagac aagggcataa aatatgtatc acgcacgaac gatgttattg aactcatttc  27420
cttttcacaa ggtcaattta gtccaaagct ggcatctaga aatctgatct ccagccctga  27480
ttgatgcagg ctagcagcaa aagaaattgt tttcccggaa tcattcctcc gattaaccat  27540
cgtgtggcat gtaaattccc cactgtcaat gctgtttgaa taatagcccc ggtgatatct  27600
cattcccgca gggcggacag gcacgatggc actatggtga aagccttttt ttcttctcac  27660
gttctcacgc gatcctgttg cataaagaag tgcactaatg agtggtggct gcgcacatgt  27720
ttgcgttcgg gacgccgcag taagtcctcg ttttgcagtt acttccagct cgtagggcca  27780
gtagcgctgc ttagtccttc acggattgcg ctcgatgata taatgcatca cctgccctgt  27840
cctgccatgt tggttgttgt tgctgcgacc gggacggatc aacgagcggt aaaattactg  27900
cacagtggcg gcggtttcat gctcgcaaag gcgaatgcac aggattgtgt gcaattgtgc  27960
gacgattgcg tgcaggaaga gcaggagctg aaagtgcgca gggggacagg ccgcgctcga  28020
ccaaagtaat agcgggggtg tatgttttcc ctggtgaatg tgcggtccca cagcgttact  28080
acttcattcc acttgacgga agctaatgag cagaatcagg ttggctgggt gcataagagc  28140
gaaaatcaca aaagccgtac acaaaaacac acaaacagcg atgggctcgg aacgggttaa  28200
aaaagaaaga aaaaagacag aacagctcca ggatcctttc acgtgtacac gcaaaacaac  28260
tgcagaaaag caacaaaaaa aaatgctcct attttccggt gtgccgagtt accgcgtcgg  28320
agtcatcgtg cagctcgatg tctgtgtgtg tgtgaacggt ctcgcagtaa cggaacaaaa  28380
aatgtcaacg agagctctcc agcagaaagg aaaccggaaa attctccatc gatatagcaa  28440
cagctccact tcggcgcaca gtccctacct accttcccct cactattgcc ccaacccatt  28500
gggcggcggt ggtaaatcgg aacggggcat acatcagcgt caagttcaag gacaattgtc  28560
aacgcttccg tccacaacga tccgccaccc acacgtcttg gggtggatgg ggcggtcggg  28620
gaaaaaaata gaagcaaccg acgcgcacca ccccctggaa gctcgcggaa aagtgtgcta  28680
ggagagagag agggaggcag agaaagagag atggagagac ggaagggagt ctcggaaaag  28740
tgtctcggat gtgggaaatc ggtttacacc gttaaccgat gccagccaga tgggccatgt  28800
ggggccgatg ccgttcgatg tgtgcgtgca cagcgtgttt gtcatcgttg cgttgtcgac  28860
gtcgtcgtcg acgttcgtgc cggctcaccc atacacaggc cgcaccgaag caagcagttg  28920
ggaaaacatg tggctacgac gattcgtgcc gggtttttcc tcgtgcactg caacacagcc  28980
ctccccccttg tttccctgtc ctgcgttgag tcgcatggcg cacgaagctg tttgtttggg  29040
tacgagccgt tgttatgacg cggcacggca aacgcgtttt ccactccggg ggccggggcg  29100
ctgtgtgtgt gtatgtatgt gcgcggggtt aggttacgtt tccgcgcgcg cgattcggcc  29160
tgacgctgtt cagccagtgg ccgcaacatt gttgctaacc gggctgattt tgtggccgaa  29220
agggtaggtg ggatgggagg gaagggtgca atgtgcagac gggctaaagg atttggcgag  29280
acaaggaagg agtcgagaga gagacgtgtc cttggtgtgt ggtgcaggtc gcgctgtgta  29340
ggttgagccg tctcgtgtac ggttgactgt gtaagtaagt ggaaagttct ctctttctca  29400
ctttttctct ttctttctgt ttctctctct ctctctctct ctctctctct ttctatcggt  29460
tgaaaattat ctcgcgccac ccgcatacac ttgtcacggg ggagtgtggg gcagtgaaaa  29520
tgcataccgg cgaaaggagg ggaaaacctc ggccaagaaa gggaggccag tttttctctc  29580
agctgttggt tctgtcgact cggctgcaca cagcgaaagg atgtgtgttg tatgccgccg  29640
cacacaaagc caagcgtacc gacacggaac acacgggcgt ttgtgcatgt gggtgagcgc  29700
tttggacgca tgcgatgtgg aaaatcggtg aaaatgcaag attgttgctg agtgcaggcc  29760
cgaaagtcag tcgtggcgct tctcgcgtac ccgaaggacg caaaaggccc gcccggtttg  29820
ttgctgttca gagcaagcgg gaaaggcaag atatcgtatg acacttagac gagattgagt  29880
tagggcatgg cgctggggtg taacagcggc accagacaat aatgctcgta ggtatcgcat  29940
taatgctgct tgtttacttg ggtttgagtg cttgaagagg tgtagcaggt ttttgtttca  30000
acttttatca ctcttattcg taaataagaa ttattaaaat gtaatgttag gtatttctgt  30060
tgaacaaaac ggttttataa catacagaag caattaatgc attgaaatag tcttatagaa  30120
agcaaaactt caacgaggaa acacattttg gatgtttcag aaaaaacata ccatcaacaa  30180
ctgtagagct tttcagaaag agtaaagttc ctgcccagtt ttgattggcc ccgttatcaa  30240
aaaagtgaaa caaaaacctt gaaagcagct tgtttgttcg tttgtcccta atttatgttc  30300
tttccttgct ttcgatgatg cgatggcacg attttggctt gctttaatga tgcgttctga  30360
ttaaggaccg attagacgtt ttttttcttc cttttctcct cgctcgccag cttcctctag  30420
attcgcagag catcggtgcg agacacaacc aacgttagcg ttgataaata acaaactcca  30480
agggggttgt tgttgttatg cgttcctttt ttgccacaat ctccaaatga tagcgtaaac  30540
ctgcaactat ggcacatcat aacgtcccgc ttgagagaga aaataggcaa attaaaatgc  30600
gaatgggcca tttttgcttt cgttcattct gctaccgatc ggtacgattt tagtgttcac  30660
acacacacac acacttcttg atgatcgctt cattcatcgg ggcaacagag gggtggccgg  30720
aatggtgtta taacgtataa tttgtgctaa tggttatggg gtggctttat ttatcattac  30780
cctaacaaat tgatagattc cgttgactgg ctcacacttt gctgcggccc tgtgagacct  30840
ttgctttgat cagtcggcgg cagtgtgttc tgggtgcgat aggttccagt tgttgcctcc  30900
acaaaccgat cattcgtcga tcgttgatcg cgcatcccag gtacataact catccaattg  30960
cgaagcccca gcgtgtggtg atgaaggaag tggcgcagtc gccgctgtta cgacctcttc  31020
tgctagcatc gggccacggc accgggtggc actgggggct caacgacgtt tgcctcatcc  31080
ggtgtccggc tgtttggctg ccaaacccgc gagcaaacat aagcagacaa acaaaacgcg  31140
caccgctcgg tccccctccc agccaggcca ggttcacaca caataagccg gcaccgcgcg  31200
tgcggccgaa tgccgcaact gttgaatgca tgtcgtaaaa taaaaattta tgattgtaat  31260
tatcatctct tctctcgcac ccaccggctc cgagcgagga tgggagggat gtggcgaacg  31320
cggcaccgag ctggagcaaa tcttcgcaca cccgtctgca tcccattttc ttcggatctc  31380
accacatctc tcgagcgctg gtgcaaccgg agatttaaag acaaaaggca aaccatacac  31440
```

-continued

```
agacacacag gaaaaggaaa tcagttcgct tggggtagct ctttttcgcg gtttgcagca  31500
caatgataat gggttatgta tgtgcttgtg ttagccctgt tcttgctccc acctttctct  31560
agccgtaacg ccacaatgcc agtaagctta acttatcccc cggttgctgt ctgtgttgga  31620
tttattaccg gtggcaagta agttgcagcc cattgctgcg gtgcgcgcgg tgcgttatgg  31680
caatgatttc gcatcttttc atcaagtggt gtgagcggcg ggccgtcttg gacacgcaga  31740
aaaggtctta tcttgtgact ggccgtgtgt atgtgtgtgg ttctgcgctt aaagatataa  31800
tttgtggcac gctttatcgc gacccgtacg acattgtttc agcagcgttg cagcagcacg  31860
cgccccatcg gaaagaacgg cttgatggac ggcaggcgag gtaaataaaa gatataaacg  31920
ccgcccgcca tgtccagttt aatcagctgt gtcctctgga acagttttcc ggtggtttgg  31980
atgaggttgc atcgttacta agtgcattgg tgttacgcat gcgcgaagaa caattccgtg  32040
accttgtcgt gcgcaagcat tcaaaagcga gaaaagcagc tttctgttca gttagctgat  32100
gatttcttga aacgctttct tctttttgac gggttctttc tcttggaaga tggtgaacct  32160
tatttttcat tggtgttatt agatgtcatg taaccatgaa gtacattctt gcctaagata  32220
ttacgtcatt cgtaaatatt tattagacat tgtagaactt ctgctcagat gatttattca  32280
cgcaacacgg aaatttacaa atcttttcca cacttgttaa agtgcttgag tagttaagtg  32340
aaagagaaca aataaaaccc agctgtggag cacaacagcc caaacgaaca gggcatcctt  32400
tagacatcat tatgggtcgg ttctgcaggg ctgtctgcaa tcataatgat cggttggagg  32460
ttggagctcc aaaacgcaat cagtccatac gcgcggtgca agacgtgtgt cccggtgctg  32520
gtgaggtaaa gccattccgg ccgactatca gtcaacgcag caagcagaca ggacgagggg  32580
acacgctgga tggatgcctc cagagtgtga tgttctttgg tggggtcggc gggtatgttg  32640
tggtagcatc aaatcgagca aatcgagatg gataattttc gattattacc gggtaccgag  32700
gcaaaccgag ggaaatgata ttgttttctc gagttgtacg tttttattcg ccgtgtttta  32760
tttttcgcca tccctcctgg tacccgttgc tgtcaccgtc ctttcaaaac tggaaggacc  32820
caccaaagtc gtcggtaagc attcacatgc agccaggctc gcttgcatct ttccgctata  32880
tcaacctggt aattgcatag tgtgagtatg gtggtggtgc tggtggtggt ggccaagcca  32940
aagggaaagg ggaggaaata cggagaaaag caggaacacc aacatccaaa tgcgctttgc  33000
gcttgcaggc atttcgcgca gcattaagcg aagccgacag accacggcca gcctgtgcac  33060
ggatcgcacg gattgggcac gggaagggca cggggagaag agacatgatt gcttcacgcc  33120
accacgggct ctcggtccgt gtaccagacg ccccggacgt atcggaatgc gggctctggg  33180
cgtggctcac ccggggaaaa gctgataact ttatgatgtg tcgaagatga gaaaatcatg  33240
actgttgtat ttttatgtgt ttttaaataa tacaattgac gttatgttaa cgggcggtta  33300
ggctgccggt tggaggaaaa cgaataatcg agtacagtcc ccctgtacac gcagcacagg  33360
gcaaatgcga atgtggcttt ggagcgaata tgcggttgcg gtttgcacat tgttgtttgg  33420
tttggtgaat tagttcggct tcaaggtctg gcttttgttt aagttaatgt cgtattttga  33480
gagtttgcat gatagttttt gcatcctgtt aagaaccttc gcccgccgat gtcaattaat  33540
aatggcagct ttaaaaatgt gctgcacgtt agctcaatca tgctatttgt tgtgcgtgtg  33600
tgtgcttggc gcgttgcaga atgtatttgc ggtaactaga gtacaatgct gcatctgcac  33660
tgacctagtc gtagagctgc ccttctccag gccttgcgca cacatgctat aacacctaca  33720
ccactgagta ccaactgagc gcttctttat aaatgggaag tcatttcgat tcattgattg  33780
aatggatgag tgacgtgaaa taattgcatt cattgcagct ctcgcagtag caatctgcgc  33840
caccaggaac cgaccgggtg ggacctagct caatggctca atgtcatcac agttgcgtga  33900
atatcaaatt gcacacggtt tcccttccag atatatattc ctataacaac acggtgcccc  33960
gcggtccttt tacggaggca cgatgtacgc aaactgctcg tttgggcagt tccaaaaata  34020
cgcatttttc gacgcaatga cgatataatc caaagtttgt tgggagcgca cggggtgaaa  34080
ggcgatttga gtattctact gcaccgtagc gtttcgtttt gtagccaatt ttccagtcga  34140
tactggcgca acaaacgcaa cggcatcaaa gcgcgtgtct tgtacccact tattttctac  34200
gtcaatacgt gctgcgaatc cgttgtcaaa aacacgcgta ctactacgcc tccaaaggat  34260
ctgcttaagg aacggcttcc gtgcgaagtc ggcactgctt cttggatggt ttctttcgag  34320
gcaaaggctc tggttctggc atgggggtcg aaggtggttg aagaaagttg cacggctatt  34380
tgtttcaaac atgccctaga tagaagagag gctctggaag ttctcgaaga agtatgctta  34440
tgcagatgtt ttacctttttt ttcgttccat tgctacctgt cttaaacagc taccaatagt  34500
gcaccaatag tgctttggtg catacgagaa cgtttttaaa cgtgcactga cggggataac  34560
tgatggagat ataaccaggc tcaaggatca aaaacaactt gatagtccag agtttagcgt  34620
attgtagcag aatcttgaag catattgcca atcaactctg tacttgcgct ctgagaagat  34680
gacctggtga tggacaagaa ctctttcttt ttctctttcg caactcacat tcactcataa  34740
tttgcttcac aaaagaatat ggaattgatc tgttttgatt gagtgtattc atatctttcc  34800
taatttcaat ctactgactc tcatctgttg ctttataacg gaagcggaag aaaatgatcg  34860
attcttctag cattaaacga gcatcggcat atcggtccag agaaacgcca aagacaaaag  34920
acgaaaacag acacaaacaa cactcaaaac gaccggggaa gtacgatcga caaggggcga  34980
agatacggga tacggtgtac gacgagttcc caacatcatt atcatcatta ctgaagtgat  35040
cgcgtcattt atgatctgct aaagttatga ccaaggcgat cgaaagcaaa aaaaaacgaa  35100
aaatccggtg gtttgggcgt agccgtgctc ccgaacgacc tcgagaaatg cataaattgg  35160
acgatgtcca aactcacgag cagatcactg ggggccatct cacggtgtgc tcgataccgg  35220
tgttccctgt ccgaagcgaa gacacgggcg aaagggaaag cacaagctgc cggtagataa  35280
tgaagctgaa caggcaatgg gggccgatga agagctcgcg taccgaagag attgcaacta  35340
aggaaaacaa ttctgaagat tgatcgtgtg acgaacacaa cttggggcgc tcactcgtac  35400
ggaagagcaa aaaaaaaacg gttaggcgaa gcgaacgaaa ctatgaaggt accacttgag  35460
gccactcggt ggtgcatcag tccctccttc ccctcggggc gaagggaacc atttggatgg  35520
cggctggaga ggaccgtttc aaatcgccac aaatcgatca acgactgtcg aagaatcgtc  35580
gcgtcgtgtg gacggaggta caggggtggt gtgtgtggtg tatggtacga ccattgtctc  35640
acctgagcgc agcagctcag ctcagttggc tgttgttcgg ggtgttgcca gccgctgcag  35700
aggcaactgt aggcgcactg tctggcggcg gtacaggcag cttctttaaa aattgatttc  35760
aaccgcgaat tgcggctcga gggggccgct ggcgagccgg cgatgcgcaa aacaaaggct  35820
cactgagagg gatccaataa aatcgacaaa tgaacgatct ttctctcggc tcgtgggttt  35880
tttgttgttg tggttgatgt tgtagtgcct tctttagcaa tcttcgtgtg aaggctgttc  35940
gcttaagtca cggcgatggt caatgatgca ctgcacactc aaccgtaatc atcttcgtca  36000
```

-continued

```
tcgtttcgcc ctccacagaa cggaacgggt ccttcccaag agggggggata ggaccggtag  36060
tggcagtgca tccactatta atgcagaatc aatcaacggt gggggtcgag atcgaaacac  36120
acggctatcg cgtctggatt gggtgcgatc gggccgatag gccggctcta gggaccgctg  36180
gctacatcgt cctattgagc tgtctggatg cattgtgtga attatataat taatttcctt  36240
tgcgccctcc caccggtcga gcgtcactga gagcagcgtg tgtgaacgat ccttggtgca  36300
tcgcacgatt atgactattg tcctcgggcg agaacaaggg tgtgctgcgc ctggatctac  36360
cttgggcgtg aaggaggagg ttcttatgtg tgtgctaatc tgtcggtcga atatttgcca  36420
caatagtcgg caacagcagc agcagtagca gccgtgacga ataggcgcct gacggggtgc  36480
ttttggtgtc gctttttgcg agtcagttgt tttgcctcat cattctcaat gtctcaatgg  36540
cttcgatgcg gccaacatca aaagggtttg atggcagcat cttcacagcg tcttcgttta  36600
ctgcattcgg attgaaggtg acctattttt taattattta tggtatttca tccaaatgtg  36660
atttttgaag ctgattcttg tttgtgttct ttgtgtatct gcatggatgt tttgtgcgga  36720
tggatgtgtt tgatgtgttg aaattatttc acatttattg ctgtaacctt tcaccgttca  36780
ccgtgacgat tgcatatctt tttttgtgca aataatgtat ccgtaatatc aaaaacatta  36840
ttagaaaaag aagtgttgta aggaaacata ctaaccaata gctttgaatt agtctgagaa  36900
ataaaatagt ctaaaaataa aaataaaata ttgcacaaac aatttgtata gctataggct  36960
tagtctgtcc ttgctttaaa gactacccca agggttgata ttcgtagcat aaattatgta  37020
tgagagttat tgattgactt aaaatcgctc acctgcctgt ggccgtggct gtggtagtat  37080
cgaccgcagc caacatgcaa tgtcccaggt gtaacgacac aattgcatac aatatagaag  37140
aaccagacac tggctggccg gctcgggact gcaaatgaaa ggcaaaatcg aataacgaag  37200
aatccttcta atttcaaccc ccgtcctgtt cctcgtggcc ccgtggggtc atggggtgac  37260
agctgtgtgt aaacctcccg gagaaaagta aggaaaaacg agtgagtgag aaaaaaaaag  37320
aaaaaacaat cccaggaaaa aaataaaatc cccgtcaaac gatggtgtcc gttgttgctg  37380
ttgcagaagg ttcgaaaaat agacaccaga gcgtttattg cctgccggtg gctttgcaaa  37440
tggataggat taagtgttgt gcaggttagc cgtatgcaac tgattcgtac tgaatcgatt  37500
tacagtggag cagcagcagc agcagtacca aacaggcaag accattcctg ctagatacac  37560
cctgttgctg cagtttcgag gccaggcttg acgctagcta tctctcgctg taagctgtcg  37620
ggctgttaaa cgctcgtgtt accgtttgcg atgcattaat taacgaagtg agggcgagca  37680
gacggctgac ggggcaggga ccggcaatag cggagctgtg aaaatcattg acattggtaa  37740
atttgcatat attgttcgcg ataaaagaaa tgattaagaa atgtggagtg ggccgggtgg  37800
ccggtttggg tggctgttac gataagcgtt taacgtcgca ttaattagtc agagggtatc  37860
cgagcccaag tcgatcattt cgtgctgccc tggtcacggt tatgatgcgg tttgacgttc  37920
aactgtttga agacgacgcg cgttgtgact ttcgctgata acgccgtctt aatcgtgctc  37980
aatcacatcg caaaactgcc gcggtgtatg tgcgtttcta agcggtgcaa cggtgggtgg  38040
cattgaattc ctcccaggcc caggcattgt gacgcgcact gcacactaat cttatcgcct  38100
ttgatacacg ggtgtcctct attctggtca ctcgccactc cgggggtagc ctttcagttt  38160
ttgccaaccc gcttcaattc ctccggtctc aacaccctcc cttgcacata gacgtgcttg  38220
ttcattagtg ttcctcttca ccctggtggt gccatgaacg cacaactctt ccgcaagcgc  38280
atcgtcgtct gtggatgagt gtgggttgtg tggtttacat tgtactcatg gtgtttgagt  38340
ttgctttttt tgttcttcct ttgcttgcgt tgtgcaatac tgctacgaat gtcagatttc  38400
tagtcgtact cgattttggc cgcaaacaca catacgcgct gctctaacgc catggtctgg  38460
taggtccgag tgcaattgtg ttatcagctg gcgatttttg ccctgcattt tctttgccgc  38520
gagtgacctc gacttgggat ttgctatgta aacataacgt gtacgtgtag ctcgtgcctg  38580
gaatagattg cctccccata cagccagtga cacgcacaca cacacacaca cacagacgcg  38640
tggcacggct gtgtttatgt tgcaaagatt agtttgtgtt ggtgcagtcc ccgttcgctc  38700
aaagcaatgc aaagcagcag cagcgacggc accccggaac acattggctg gtgactttgg  38760
ttttgtgccc cgtccccgtg catgccaccc ggaaatctag ccgccaacgg tgactaggtg  38820
tattgatgaa tttaaatttt gcactacaaa aatgcgcttt gctttttaaa tggtacatgt  38880
gcaggcgact ggttgctctc ctttccttca ttgctgcatt gccgcttttt cccaatcaca  38940
tgctggattt ggttgtctta cccctccctc gcacacacac gctcgctcgc tgcatcacta  39000
aagagcatgc gaaataacga taagtgacag ttgaatgttc agctgtttgc tgctacccgg  39060
ggtttcgtaa agccatcttc caccgtgccc gacccttgtt ggcgataaac gcgcgctcgc  39120
gaaaaataaa atcaaatacg ccaactggaa gagcagttcg gctgtacaac acaacacaca  39180
cacactcaca aacctagccg cactaaacag agcgcagaca gcgacggcga caagcggcca  39240
aagacgacaa ctaccctatc ccaaccccgc gactgacaag tctcgggctc ttgcgttccg  39300
cttctaatta agcgcggagg cccaccttca gcgtacagcg acgacggtgg cagtccttcg  39360
tactcgtttt tttccttcct gtgctgtgcc ctactatgtg gtagcactat gtggcactgt  39420
tgcgaaggag cagtatagca accacccacg ccaacacccc accgggccga cgggagctaa  39480
aagtctgaca agttcaggca gctcgcacgg gagtcgggaa tcgattgtat cgatagcagc  39540
ccaagcgtcc ccaataatcg acgttaaatt gtttcccccg ttcgcgttgg attgttacca  39600
tttgcgtagt tacactgctt aatttttagg cgtaatagta ccgcatcaca gtgtcgtaaa  39660
ctatcggtac gttttgacat gcagcgcgtt gaaacggcac aggcaggaga gcagccaaaa  39720
cgaacgggaa cgcataaaat tgggttagct gcggtggagg cgtcacggta acgagctgga  39780
agctggcgta aagcgtagat gaagctgcac agacagacag accacgtcca cacgaacgga  39840
ctgggaagcg ggagaatgca cgttgcaatc tttgaatctg atttgcacgc agatcgatgc  39900
aaaaatgttg catgtcaagc gttaataaag attggtgttt acgagtgttc gttttggctg  39960
acaccggccg gcagcgggtg aaacatgcga catcatacct ggcggtactt ggagcggaga  40020
gttggagctg tgccagcaaa ggtgtcaaac gtgcagctta tcgaaagggt aatgaggcat  40080
ttacttgctc tgtcgcaaga caattactca agaatagaat aaatacaaca accaaaaaag  40140
cccgcaccaa tttgtaagga ttcattccag ctctcccctc gcagggtaat gtgtgtaaca  40200
atacgaagtg tgacagacac ttcgggggaa gtttttgaca gctcctggga atggcaaccc  40260
ttgcggctgc actgctgcac actcgacagg ggttttacac gtgcatgcgc gactggtcac  40320
tccgtagcac acggtaaaca atgttgtaac tgcaactcgc cccttaagaa tcctttcgcc  40380
cctcaatttg taggcaagtt tccgtctctt tgcacacacg ctgaaggaac agaacgtcgt  40440
cctatgatta tgctgtcagg gagaggaaga aacagtacgc agagccacgc cggggcacaa  40500
ttcattcgat cgggaccggg aggaaaagcg tcctcgtgca catttgcacc tcaatagcga  40560
```

-continued

```
gcataattta gtcaaattaa gcgtactccg ctgggagtgg acgacgtagg tcgtcggtgg  40620
tggcattgtc cgagaggact ggtgccacgg ttgctcaatt gtaacaatcg ttgacctagg  40680
tcggtggtga tgtgtgtggc cattgtttca acattccact agcttcgggt cctcctaaaa  40740
tccactcccc ggacggatag ggcgaacgca agtcacgggc agcgactgct ctgtggcgag  40800
gtgtttgtgt gttgcaaact tttgaaccga aaactgctac gaccaccact acttcgctgc  40860
tgttttgaac caggagctct gcatctcctc gactaactga caaaaaagac cgcatccgct  40920
cacattgttt ctatttctgc agggacagag aggtggtcta gtggtgccaa agttgcccac  40980
ggtggccgaa ttcgaggccc tacatcctcc aactaatagc agtgccagcg cctgctagat  41040
cctgctacta gcacaagtgt gtgtgtgtgt gtgggtggga agttcaatgt tgaaatgttt  41100
caccgatatt tatcccgaca ctgacccctt ggatgagcca gcgttttggt gccatttctg  41160
gctgtgtttt cgctcaaacc aaccagttcg acaataacca gtgatgttga tatattcacg  41220
tgtgtgtgtg tatgtgaact ttatttttct cgcgttttcc cgctggaatg tgcatgacat  41280
gtcgccgcaa ctgtcgacac agattcgctc tagtggaagt gcatcgtcgc gcattcgctg  41340
ctgcgcgggc tatcgcgggt atctagacat acgtgtgtgg ctagtgtagg ccagggagta  41400
ccatcaccac aggaaggaag tggttcgaga gggcgaatgc gcgccacggc gttccaaaac  41460
acaaaaagcg gtttggatcc aaactttact gcatgttttc caccggcagt cctgcagacg  41520
atggatccac atggacactg gagggaacag cacagggtca gcgtcagcag taactggtca  41580
acgctgcgtt gcgttctaat gtggggcttc cgcttgtcta gagccttccg cggagtgagt  41640
gtgtgtgtgt gtctggctgt cctgaaaatt ggattcagag cggatgttga ctgtttcgcg  41700
tgtgtgtgtg tgtgtttgtc cagccgtgga ttgttgggag aatatgtgct catccatcca  41760
tgcggcaagt cgctcacggg gtggaggtcg cagcaccgag agtttgtttg gcattaagta  41820
ccttcagttg caaaggcaat gcaaagaaga atcatttatc aaacctaacc atcttcgctc  41880
aagggtttga tattaccctc ggagaaccac tttgactcat gatccggcgt tgagcatttt  41940
tctagtttca cacattgcag taattgtcat tagcacttaa gattgaaagc ccggaatgct  42000
ttacggcatt ggcccgtaga tcgcagaaag gccgcgagca aaccaaagaa atggatgtct  42060
ttatcgcaac gaaacgtcgc aaattttgcg cccttttta ctgccccgca atagacactt  42120
gcaacaagac ggcagcgaaa gagtaaaaaa gccagagaag gcattccgcc aatgctgtaa  42180
aaagcaccaa caacaacaac accaacaaaa aaaaactcga accaaacgca cactcatcag  42240
taacgcgaga ccagtgcgac caggcaccca tctcccttcg aacgcgcggc tactttccca  42300
gccataaatc atccacttca accagattga gtctcctgcc gccgcaccag gcgtgaccac  42360
acgtctggtg cggtgtctcg tttgttccgc cgtttttgtt ggcgtgtggg tggtggtggt  42420
gggggcgggg gagaaggtaa attaatttac acttgcacac agcgcagctt caagtgggag  42480
atgcacttgt cgtctcattg cctcgttgct gctccggcct gcattgcccg ccgtgccaat  42540
gacgcagtgg ggttttggtg acgatcgcta cctttaccgc gcttgatata agggttgaaa  42600
atcatcatca tcatcatcat catcatcgga tgctgatcgg acgggccaca ctcttgacgg  42660
atcgtctcca tctcgttgcc ggtccgcttt cgcctagccc cctcgtcgcc ttgcccgtta  42720
gcagttcgtg aagaaaatgt gcataaaatt agaaatcgaa ccctccgcac acaccccagg  42780
agggaggggc ggtatgattg ggtcccgtgt atgggtgtga tggtgtgggg ctcgatgtga  42840
gtggcaatac atttgcaata ttagtggtta gattccattt cctgcacagg gagcagcgca  42900
gcggaatgta gaaaaacaaa acgccggcaa gaagtgcgga tgcaaacttg caattgttgg  42960
ttctgcagct cgggtgcggg tgtgtgtgag tgtgtctgtt tgttttcttt gcacgctgcc  43020
tggtggcccc agggaaggag agggcgttgt tatgggagaa tgtaaaagca aaacaagcca  43080
cccatccccg ttctattgca tctcgtctcg tggtccaaga ccactcccta tccctctcgc  43140
ctcttcccgc ccttaatgtc cctctgtaaa gaaagacgat ttgttctcac attcctgctt  43200
cctccttccc catgtaccac catctctgtc tggagaatcg tgcgcacaca cacacacagc  43260
cacaggattg tgacagtacc gtcccctgct gggaggtgag tgaaaagaaa cacatttcac  43320
gcgtgtgtgt accctgtgta atgtcacagt cgatcacact cgggcccccg ggtgaagccg  43380
attgaatcat aaattgcact tacggaagca cttgttcgca ctggcctgtc cggtggccac  43440
aaccgggtcc gagcggtgtc catgtgtgcc gcattttatt ttgcagccac ttttacaact  43500
gtgctgctct gctcccgctc ccgctgcacc gccagttcga gagatccgag cgtacgagaa  43560
gtgatgatgc aatcaaccgg acgggaggca acccatcgtt agctcgccgc tggagccgat  43620
agagccaacg gggccgggag ggaaggatgg aatgtgtaac gctgcagcta aatggcgcgt  43680
gcaccaacac cagctcgcag cggcgagaaa ggcgtaaatt gtgcggcgcg tgtatgattc  43740
ttggccgggg cgcgttctcc ctttccccca ctgccaatcg ttctgccctt ctggatctgg  43800
gcgggcggca tgtgactagc taattttcca actcagtggc tggccggcgg tccgtaagat  43860
gatcacaatc actttggaac agtaatgtgg gcacaaactt tcgttggaag gttgagtttt  43920
ttttaaataa ataaaattgt taaatttcca ccaccaattt cccccgtttt cactgttccc  43980
tagtttgagt ttgaaggtca atcaagagga aaagaagaag cgaattccct gcgcaatcac  44040
ccttcgcgag agtcggagga agggacgcgc aaagaatcct attgatagaa gctactgcag  44100
ctactacact acacttgcgt aattgtttaa cgtgcagaat gaatcggtgc actatgcggc  44160
cgggaagtgg ccgtgtggtg gggcagctct cccccgttcc cgcggcattg ggttaccagc  44220
gtgagcgtga gcgcgcgcgc gcgcgcgaag aatcgatgat gccgtggagg ttgtcgcgcg  44280
gcgcaaacat tgtggtgtgt ggtgtggcct gagaccggct gctaggggaa gataaaatgt  44340
agctcgggtt tgggtggcgg cgcgtgctgg tttcgtgatc gcggctcacc ttcccaatcg  44400
gatgggcggc ggttgatggt cgggcgggga gtagtatctg gtgttcattg ctgcagttcg  44460
gggcagaatc tgaaggccca agcatgggcg aggcaagtga cgcaggcggg tgccgatgca  44520
ccggtaagaa gggcgcgcga ggcaagctga taagaatgtg ccggctgcac aggctgcagt  44580
tttcggtctt tgtctttgtc gcacggcatt ctggagcaaa agaagaagaa gaaaatgatg  44640
aaaaagaaga aagatgcgtg tgttggatga ttgtagccga ggaccgatgc gatggtgcgg  44700
ttggtggtgt tattggtcag ctaatggtga gccggtttgc cactgtaaaa ggtaatcgcg  44760
actcgaatcg tcgcgagact aaatatagag cacttcctga gttcatgcca agtggcggaa  44820
aatggacgga actgcatcgc ttgcccctcc cgtaccctcc ttccccttc caccagccac  44880
acacatgcac acttatacca acacagtggg gttgaacagt gcattggaca aaatgcacgt  44940
gtaaaaaatg caacagccca tgaatgtagt tgtgtgatat ggtgcactca ttgtgtacgt  45000
gtggtttttt tttacaaatt acagtgtgtg tgtttgtgtg tgtttgtata aaaaacacta  45060
cttacacaaa cgcgtttact cgtgaagatc aattcattgc aacgcgccga atgactcgcg  45120
```

-continued

```
acgattgtgc cgtttgggtg gatgatgaaa agtaaataac attctttggg taaatagttg  45180
caacccgaag ctagtgccaa ctgtgctggc ttgctccttt gctggcgtgt tcgggcctcg  45240
cgtctcgtct cccgttacac ggacacgtaa atggtagatg taaaaataaa gtttcgcgtc  45300
ggggttgtat tgaacggccg tctggggtgg ggttttgagg ggggaacgcg ggtatggcca  45360
ggataaaagg tgggtgtgtg tgagagctcc gaggtgaaca atcggtcgtg accacggccg  45420
ggtgttgtgc agccaggctg tgtgcaaact gcagcgagat gcaggaaagg ggtaaccgtt  45480
ttcggcgagc cttcttgtag tttcagcacc ctcggttacc cacttctcct ctcctagctt  45540
caccacacgt ctgttgttgc gggcgttctg ttcttctttc actgatgttt aaacgtttct  45600
tgaacgatgc gttttgcgta cgatttttga gtttataaca cgtggttttg cgacatgtta  45660
acatttacat tgtaatcagt tgattgatgt taatcttttt tatttatttg ctctcctttt  45720
cagctactca ctcgtgcgtt tcgccagaac ctgtaaatct cctacctggt aagtaaatat  45780
aattaaaaaa aggaaataat atatttcaaa gcggtacaac ggtgttgtag caaacattta  45840
gtgcttcaca ctgtacgttt gaatatttgc taacacgata tgttacagcc gacattaaag  45900
catcttaaac caactgaacc caacatgtag ttctttgcaa gcaaatagga cgtcatttga  45960
aaaatgtgca tttatagctc atactttatg gaatgatgta tgttcttgcc cgatgcaatc  46020
tgctatagac cacattgcag gctgcatgtt ataaatatcg gctaacacaa tgcgtcacct  46080
ttttctcacc ttaccgcgct cggacgctta aatcttgtgg gcgtttgctt tctttgacct  46140
tatccttgtg cgctaggcta agcgtatttc taagccagtg gacatgaggt actaccggct  46200
tccctttttc gatatgtaac acagttaaca tcacaagcac acacacacac acacacagaa  46260
ataatgtcgg tatggcaatt ggacaatatt gttatttatc gccacattca ccaaccgatc  46320
gaaattgtcc caaatcgctt cgagtacata attctcctat ctgtctgccg ctggtggcat  46380
ttgtacgaaa acgtataaaa tgccccgttc ttaaggcgac cgccacacaa ttgtgggcat  46440
tgagctgagg ggcgcgcgag actcatgttt gtcgcatgca catcgcggcg gcggcggtgg  46500
gagcagcggc ttttcgcgca cctttgtcgc cctgttaagc atttttctag acgacagata  46560
ccagcgcaaa tactgttgca ttatacaccg ggtgtttaag cagggacccg gtggtggaca  46620
taagcagaac gataaaatat ttgcaaaacc gatgtttctt tgcgctgata ctcggcggat  46680
acgagcgctg tgtttgtaca aaggtacaaa caccgagagc gtgtccgcca tgggaaactg  46740
cctcaaacat acgcccttcc gtccccctcg cctcgccttt taccaccgaa agggcaaaaa  46800
agggtgttaa tcgtttcgct gtgcgatgtg atgattggag atcacgaaga tcaaacgggt  46860
gctggggtga aaagcacgat gctacttttg cgacataatg cgctcgcttc gatgtgttgc  46920
gcgtggacat gttcggcatg cattcttcgc attaaatgca atacgcgatt attttgaaat  46980
gaaaattgat cgcaaagaaa atctcaaacg cttgatttta cttccaaaaa gaaaggagtg  47040
cgcaatgcga atacgagagt gaaaaagaga gcgttatgac agtgcgcttg atggctaatt  47100
tgcaaacaat ttacataggc cgcatcagaa cagttcatta cggatcaaaa taaacaattt  47160
actttttgct cgtatttgct ttttttgttg ctccccgggc ggttgttgcg atgacccgtc  47220
aaaggggatc agcggtaaca gcggcgaatt cggcgcgctc tcgtggccgt atggagataa  47280
ggcgagcgta aagagtgcga aggggaggaa gggacctcga acaagaacac gactacaatc  47340
gcacagtacg aaaacaggaa gaaactcgga ggccgatgta aaactggccg cccagggtct  47400
ggacaaaact ctttatccaa gcaagcactg ggaatggggg aggaacaagg gcgctccttt  47460
cctcggggcc ttgctggctg gtgggcggca gggaccgggg gaaataacac caattcatgt  47520
caatgtcact gtcactcaac cccaacatgc aactgcatca tgggggcacg cgcgaggttc  47580
cctcgttctc ctccgggaag ttggtttcct tttttaatcg gtggagtgtc gagaaggggt  47640
gcaggcacga ggtttgggta ggtacagtga tgtaggggga gaacgatgcg tgtgcagtgc  47700
aatgatcaaa tgatacaggc aaggagagcg aagaggtcac gaatggtgga agtacttgat  47760
tttcaggaat caatattcct cgctgtctgt caaccgttct gtccccaaaa gctggcggtg  47820
gggggatccg gtggatcacg atgggtgaga aaatgagtga ataaaacaaa aaacccgatt  47880
gcaatactaa taataaaata aaataaatct cctgcctcgt ccagcttttt tgattgtgag  47940
cctgattttt ctctacattg tagccgatcg tgtgcggggg atgtcagcct ggggcagatg  48000
gcgcaaaagg gttgccgtac gcaggacaag cagaaaatcg tggcttgaag cccgcacaat  48060
ctatttcctt tggttgtttt aaaaatgggt tgcatccagc ttagtctgag ctggaagttg  48120
tctcacccgt aggggcaaca gggaacacga acaggagact cgtttccgca tcggctagct  48180
tcggtggaaa ttgaaggcat tcacccctttt tttctttttc tagtccataa ttgcgggtga  48240
aaataatgcc gcagttttcg tgccgtccag gggacaggtt ttcttcctac aacatgatta  48300
acattgcaac atttgttgta acaatgcgat tgtgtgtccc agtgcgtaaa acgcacgagc  48360
ctccgatcat gatgggcatg ggaaggaaaa accgttcgac ggtacatttg ttgcgttcga  48420
tcattgtcaa ctccattaaa cgaacctgaa taaaccggtg cgtgtgtgtc tgcggtgatg  48480
gcgatctttc tttatcaaac aaacgtgttt gagtgttctg gaggcgtttg agtgagcagc  48540
ggccatttgc attcacgaag ccgagttgca tcccaataaa accaactgca tgagatgatt  48600
gatgttggga gatgagctgc aatacattcc caaccgtccc gtttggtgtt tgattgattt  48660
ttcttgcacc gagctgctgc aaaccgggcc cctggatgcg cactgatttg tttgcttgct  48720
ggttgcaaca aagccacacc accgttaaac ctggtgatgg tgatgcacct gtggcggatc  48780
gttgcgatgg agcgactgat ggtgtgagct ttgtaaatgg aatttcacgc gtagcgcgtc  48840
tagacaaacc ccaattgcgg ctgcagcccc gtcatgcggg cacgaccgac cggacggccg  48900
agaccggtaa gacagtgtta agtggaaatg agctgcggaa tggctggcat ggtcgtcgtg  48960
gcaaataacg ttggccatgt tagggacaca agaagatgcc ggtatttggc agaaggtgca  49020
aacgcacaca aacctacgtg aatgcgatgt cttctgaaat taactgtatc gtttgatgac  49080
acaacgcaaa acgaaccagt ttgtcgttac tttgagagaa gaggatcatg atgatgatga  49140
tgatggcggt ggtggtggtt cctcaagaaa gatggagtga agcaagtgtt agatccggtt  49200
accgaagcga ttttcaaacg cacagtaatg attagcgaac gggcccctta ctgtttgcct  49260
gttggtggtg cagtcttcaa tcatggaaca cgctgggctc ataaggaaac atggggcata  49320
atggtcatgt gaataatttt gctcttttga taaatcatta attatctttca aaatcgttga  49380
ataataattc aacaaaaatt ggtgctttaa ctctagattc atggtacaac atgaactgca  49440
ctcgtttaca aacaaaatca gtttaaaaaa atgtcagaca aaattgcaag ttgcaaaatt  49500
gccttaatta tattttttat aatgatgcga agccaaatgg taatcggccg atcccgtcag  49560
atcagttgtc aatcacttac accggtttcg agcccaagta aattatgtaa agctgcttta  49620
gaacgttgtt caactgtaag taaacaatta gcgtccaact gaaatactta tgcgtttctg  49680
```

-continued

```
aacattgttc atttgtaact aaacaattga ctcctctaag ctgatacatt tgctcaatag  49740
agtttatcaa tttgtttttg ttttcactta caacaataat gcgaatttag ttgtcaataa  49800
tgtgtataga ttgctagaaa atttctcatt tattataact caagatcgaa accaattaaa  49860
acaatttcaa aataatttaa tttgaataga ttcagaatca aacaattctg atgcccgacg  49920
agctcgggta atatagatga atgtttatat tggcgaaagc aaatgttttg ctgcgatttg  49980
acaatgttca aaagcacctt agcgttgttt agttgaaaac tttcgaaaac tttagttgaa  50040
aacgttggct tgaaaacaat ataataactt gcccgtcata ccttacttta aactctcttt  50100
ctttgagtaa ataaacaaat cgttgatagt caatccgatt tatggttaac gcaaattgac  50160
tttcgactat ggtgtttgcg tcaaatgaga agaagataat cacaattatt tctgtaacta  50220
tagccaaatg ataatggtaa aaagacaaca aagataataa caagtgtctc aagtgtctgg  50280
atgtgtatcc tttatttgat aagactgttt tctagactgt tctaataatt ctacaagagg  50340
ctttaaacat ataaatttgt atatattgac cctatgatga ttttgctccg agtgtcctta  50400
ttatttatta attaactatt tatttatgat ttattataac ggacacaaat agaaaacagt  50460
tatttttgca agactgtgca tttttgatcc gtaaaaacag ttcctggaaa aaagtatgca  50520
actcacagta caggtgaaac ataatacagc ggttgtagag cgtactgttt ggacaagtta  50580
attaaattgc acccaagcgt gtattaattg tacccgtgtt cggcgtgacg ggcacacaca  50640
ggatcaaacc actactgaga aactggatct gcttcgttcg cactcggcgg tggaaagtcc  50700
tttccgcaca gcacaggaca gtgcagattt tgaaacatta agctctcgca accggcgtaa  50760
ccgaatccat aaaaacggag gttcctcgtc cgggatctcc tttcttccaa gtttgtgttg  50820
ctatcttggg tcgtaaatct taacagtagc agtagttgga cagtgtatct aaaaaggtac  50880
ggataccaaa aaggcacgag tagaaaggag catgtctaga tgatgctggt gctatcattt  50940
ggctccaatt cggacatccg gattgacgtc ggctcgcggt gtatgtgctt tagtgaggcg  51000
attgtaggta gcaattctcc ctcgtgttgc tcctttccgg aatagaatgc aacaaggcac  51060
aatgttaatc actcatcaga aaagacgaaa cgggtccgtt ccgcaccggc aattttccgg  51120
ctcggcacag tcgatttctg cagcccccgt ggggacacat aaacaagcga ccaaacaaac  51180
ggaacacaca ttcttcattc tcgttgcgct ccactcgtcg ttttgtaccg tgctggagct  51240
gtcataaagc atgtagtgca aagaaagttc tcatctgagc gcttcttaat gctcacactt  51300
gcggtcccgt ctggccttcg gcagctccgg cagctttggg gcaattgttg agccgtagga  51360
ggaaaagaca cggtacatat aacgcccgcc tcccagtgtg ttgagggcag ctgcccgtgc  51420
tactgtgctg cactgggatt cggcaaaaca atttcctaaa tgtggtcgac cgaagaacga  51480
acaaggttag tgtgtacctt cgctgcatcg agaggtacgc cacttctttg ggaagcaagc  51540
aaccgctcag ctcctggtcc agactgccga aactctcaag tacgtttcgg agattccttc  51600
gggagcgtgt gggttgtatg tggcctcggt tcaagaggtg ggtatagcac attttatctg  51660
ccgcactgcc attcgtgatg catacatcaa ccgttgctgg aagtaatcgt acggagatga  51720
tagacgagcg atgaaaaatc gcacagaaca aaaggccatg acacgaggac gaataaagag  51780
ttgccagggc gccatcccac cgaggggatg ccacagctgt ctcgaggagc aagccgaaat  51840
gatttgcatt cagctgcatc gtgcaagata tggaccggtg agcattggct gatggagatg  51900
aacgtccacc agagatacca ccgaacgcac tgtctggtgg tgtgcgcaag gttctctgtg  51960
agtgcggttt gctgcgatca aaagactgcc gagagcctgt cggcttattt ttcggctcgg  52020
cacaacaggc tttggggttg taaaacaagc aacaaacaaa tgtaaatatc gtgcacaaca  52080
tcaggcactg tttgagtgtc tggttaaata aagaaacggt ccaaaattta cagtgcgatg  52140
gtagtgaagt attgctttga gaatggtttg aaaataacgg tttgtaagtt atctatcaaa  52200
tttgtcatca tgcacataac ttacaagcca agttatatgt agttgatttt agagatcaaa  52260
tacgttcctc cctgccaatg caataaaaaa agccatccaa acttgagaca tttgctgtgc  52320
agtgttggga atcgatccac catgttgtaa tttcaacaat aacaaaccga acaatacgcc  52380
tatacaccat tttaaccgac tttccccttc agggctcagt cccgcttccc actcttattg  52440
gagcgtaagt gcagcaaacg tccaagcatt cgctctgtag caagcggtgc aatcaacgag  52500
aaattacagg cttccaggct accaatacga tcatttcagc tgccacctct ctgccacctc  52560
gccgagtgta ggtaaaacgc atcgcctcga agcatttccc ttacgtcgga gaaggctatg  52620
ctccatggat gccgagttgc cgtggatgcg cttgtgttgc gttgttcttt atgaacgcgt  52680
tgaaccttcc acgttgaaca cagctgaggc gagcttccag cgttggggcg agcctctttt  52740
tttcaccgcc tccccttta cccttcatca acggcagggc gagtgcacta gtgagcactt  52800
aattaaaatt aaactaatta agaaagctcg tcgtataatt ttcacaccac accatcattt  52860
tcgggctact ggtaatgaaa ttaatatttc attctatttt attattaacg tttacatggg  52920
gggggggcg gggggggggg gggcagaact cggggcacag ttgtttggta accatcgtac  52980
cattgcagct cgaccgtttc ggagatgtga cccttgcaac agcgtttctt tacttaccat  53040
tagtgcgaga ttttcatacg cgcggggagc tctgcaccac attaatctca gaactcggaa  53100
ctgctcccct tcgtcctcgg ccaatgttac caatgctgtt gatcaagcgc agtagcacgc  53160
cgccctccca gtagcacacg atcgcgcgtc tattaagtgt tcgcatgtgc agatcgcttt  53220
agcagaacaa tttatggtgc cggctgtttg agaagcgggc tgccggctac ttacttccgc  53280
ttcctccgat gattaccagg ctggtagctg gggtcccggt ggtataagaa aaagtcgctc  53340
agtcacggac ggcaacacat gaatgtttca ttgaactctt ttgccgggtg ggcggtggct  53400
aaggctgaaa gggtgcttca gcaccaaaac tggaccggtt cagaggtttc gtcgttttcc  53460
cttagaacgt gtgtgtgtgt ttgtgtgtgt ttatccaaga ggtgaggacg aaaactgctg  53520
cacgattctt cggcaccgag agattcttac ccgggttggc ctcgtagtag ggtcgcaaga  53580
gcaggccaag ggtttgggtc aatttaaaaa acgggataaa gtgtgcgagg atcaagctga  53640
agctggtggt gtgtgtccac attgtttgat gatttatctt ctgttgctgt ttgcgattgg  53700
agcgcgtgca atcgaagccg taatgctaat aaagctggaa caagcaagaa tctggatcag  53760
gcaggcaggc gggtgtcggg tgacacacaa gtgcgccaca ttatgaatta ttcatcctca  53820
cgtgatggaa gttaaacctc tatcgtgctg gtgcgagtac ggcctgggtg gagagtttac  53880
aaactcaaat gtcaagcgca tgtaaactgt agaaagtgta gatcgctaca gaaatgtctc  53940
tatttcatag tgtgaccttc cattttgtag agcatgtcaa actttggaag ggaaattgtg  54000
tacacggcca caatatctgc catacaactc aaatcaggct atagtttttt tttccacaaa  54060
ctgctgatgt ttaattatcg tgttctaccc attgcttcac gtaacgttgg aaaatgcttt  54120
acacttgcaa tccgcccatt ttcgggcgtt tctacacact gattaatcat cgataccaac  54180
gctggtaggt gttaaaagga taaagccggt aacaattaat acagtttcac ggcaagagcg  54240
```

-continued

```
caatcaagga gggaaatgat tctttcgctt tccgttatag cctcggcaag gtgcatcggg  54300
agaaaatatt gcatggtaat aaattccccc ctcccacagt aaacattgca tccaacttcg  54360
ggactacagt gtaaaggagt gcatttttat tcattttttt gataaatcac taaatgtgaa  54420
tcgtactcat cgtggatgct ttatgctgat ggctaccgct tgccgaatta acctgcgaag  54480
actgtgataa aacgttgctt acggctcaat cgaggaaccg gctacatacc cactaactcc  54540
acgcgaaggc ttgacctcta gagtgctttc cgtgttcagc acaaccgaat tgtacaaaag  54600
aatatggtag gcgggggaca caaaaacacg ttggcaatga tttatcggtt ggcattgcct  54660
tctacattga agatacaatt gatcggtcgg tcgcgccggt tcggtcaacc tttctcttgc  54720
ctcagtgcat caagtgcagc gtaaatgcaa caatgccgcg cgtttcctcg tgcccccggc  54780
cttgcgggta aagtacaaat gcagtttatt tccaaattaa ttagatccgc tgctaaacaa  54840
tgttctcctc gagcaaaaaa gcctaatgag atcttcggcc gcacgaaatt tgtgccgaga  54900
ccgcggaccc tacaatggcg ctgcaaatta ccgctttttc cgttcccttt ttgtttgacc  54960
cttgcgacgt cctcccctca cgccgatcaa cctgacgggt tcctgatggg aggcgcagag  55020
acagtggagt gacagttatc gacacttgca cggtgagcaa acgcagggag gaggtcgctg  55080
gtcattagtg ggttttgggc tggagatggg acggcgtcac acactccacg gaggagaggc  55140
agcatagtga tgttcatttt ggactacaat tcagacagtc gttcgcggtc ggacagaaaa  55200
agtgctaatc gaacgcattg catccagcgt ggccgcgaac ttgtgtcccg gggcagtttg  55260
ggtcgcgcat tggaaagtta ggagtaatgg agtgataagg gtgagtgtgg acaaggatga  55320
tgatgttgct tcgggtatga gtgcgcgagt tgcaaagtgg caaaaccaaa tattgtaccg  55380
ccaagggatg catttggtgc gatgcaccaa atcgagctgt ggttgcctct acaagaacct  55440
gcgcgctgcc attagcgcct ataaacacaa caaggtgtga atgttcgaat tgggaggtga  55500
gttagcagtg tgacaaattg atttgaaatg actgtttaac ataccaatac ggcatgggca  55560
atacgtactg attacaacaa gtttaatgag ttaaacaata tacttaattt gttgcattca  55620
atcctcagct aacaattaaa agtttttttt gtgtgacgaa acaacaaccc atcttaacaa  55680
acaatatttc actagccaac tagaagaata aaacaaaaaa acaatgcgaa tgaaagctag  55740
atactactaa cacagttcaa ctgtttgggt atggtcccgt agtaaagtcg atataacgga  55800
cgaaataaca aaatgttcca tccaggtgta ggcgccataa gacacaatgg tacatcaatc  55860
cattgctgat gattaaaccc tctagttgct taggcatgtc ttgatcaact acgcttgtta  55920
atccaaagaa caagaagaaa aagtgttaat ccaaagaaca agaagaacaa gtggttaatt  55980
caagatgtat cgctcaaaaa aaccaactga gttgactgca gtacaggaaa acaaaatctt  56040
acagcttgaa tatttttatt attattatta ttattactat tacaccattt agcagctgtt  56100
gaaaatgtat gaaaaaatgt gtacaaacac tgtgtcaaac ataattccaa cgtgtcatca  56160
attcgcgaca tagctgtccc gcaaatggca gtaaaacccc ttgaaacggt ttttaaatcc  56220
atcaattaaa aacgagccct tccccaacag aagaaacaga gagacaatca aaaacaatat  56280
gcaaaaaaaa gatgacggaa agcaaaaatt ttatcaaaaa agaaaaaaaa atgcaacaga  56340
aaaacactcc catgggggta aaaaaaggaa acaaaacatg cacattgtac gaaaacgtgt  56400
tattctcttc caccttacca ttgcgtgaac gatatgttat gccaaaccgc tcgaggccga  56460
tgggtaggcg gccgtgtgta cgtatgagtg agttaccacc accatacctg tcggcggatg  56520
ttcaatttcg attctgtgaa tggatttact tccgggtgga attgcaccgt ttgaaccgtt  56580
tgaactaccc cagaatgccg gggcggtttt gtttttcttt ccgttccgaa cgccgtatgg  56640
aaaggaaatg gattgttgtt agcacgtagc gcaagccaaa aaaagcaaaa agagttggaa  56700
agaatgaagg catgaaacga agagcacaga acagcagtag cagcaaatac gattcggcaa  56760
agtaaattta catattcgac gatcgacggc tggttttcct ctgcccagcg atttgctatc  56820
cattgccgcg gtgtttggcg tggggaaaca gcatcggcac aaggaaattg gccacccatg  56880
gggggagggt actgcttcgc ttgtccatcg taatcggtgc ccatttgcac tcactggtac  56940
atggccaaca cagagaggga gagagaccgg ggtggcatta tttgggggag ttggtgtcgg  57000
agcgtgcact tgccaagggt gtcatcatgt gccttgaacg ttgcatttcc gattccccag  57060
aatggctgcg atacggcgag caagaatggt tagcgtgaaa caaaacagtc gtttgatgat  57120
tttgattccg tttcgatcgg aagagttggt gtgcgatatt gaatgtgtgg gacgggggtg  57180
gcgaacgttt ttgttccctg tacagatgga ctgtcacaaa tttatgcaaa atgtattaaa  57240
ggatgacgtt tcgagtgatg gagccagttc gtgttgtttt ttcgcgcaag ctctaccatt  57300
ttcggtggtc gaatttttgc gccacgttta ctaaatcgcc aaacaacgcg atccaaaaat  57360
gtgtcagctc tctttgtttt gattttggct ggcgttggag gtaaaaccaa caagaaaaaa  57420
gaaaacttaa atcaaataaa taaaacctct tggccggcac tggcgggaga acgggccacg  57480
gctagctctg ctaaattaaa cactttgtta tgttttgctg caacttatta tattataagc  57540
actgctcggc cgacaggaaa cgtattgaaa tttacgattg caacaatgta gagctgttcg  57600
tttgcagcac cccatttgtg aatggcactt gtgcgctgga agtacaaatt tgaatgttta  57660
cagtctaagc tgtgcgcaca agaattgtca cccgcgaaga aacaatcatt tcgacacttt  57720
acccccggtt ccctttttctt cggctttctc tctctccctt gccgctgctg gttcgtcgct  57780
ggttcggttc ccacagctgc aaaccattta aacacttacg caaaacgcgc gttccacttc  57840
cagggcaccg ggaacaacgc ccagaacgaa atatcgttaa tctccttcgg gcgtgtcctt  57900
gcctcgcggg tacttgtctc ttggtttgcc cagcgagatc tgtacggccg cgtgtacaca  57960
ggctcttaca atgttgcgtg tgtgtgcgga gaaaatgtgt aatcgattta gtggcgcaac  58020
actatgcgca acgtttttct attaatgcac gtctgtgcgt tttgtcctgc ccgaagacgc  58080
ccaagacact cttcccaagg aatgtgtgtg cacaggaagt gtcaactcgt caaaccaaac  58140
gcggtggagt gtgtgtgtaa ggtgtcgtaa atgtcatgcc agcaaggata gggtatttgt  58200
tgttcttaaa atttacgatt acccgttcta cgctagtgcg caattcgttt tgggcatgtg  58260
cttgttggac atgttgtggc gggcagtata tgcaaagcaa acagagagca taattgttat  58320
gatgactgcg ctcctttcac ggacggagcg gtttcagctg gaagggccca caacactccc  58380
agctcagaag caaaacaatt taatgacgaa tcgtggaaaa agaaaccaat taatggaaat  58440
aaatactttg ttgcgagcag tagagggctg tttagaaatt ttggtaacta gcgattgcgt  58500
gtgtttacaa tgtattaaaa tgtttataag ccgtataact atcgagcagg aagcattgat  58560
tctttcaaac aaagattcgg attcaatgtc gcgtcgttgg atgaacgaac aatattcttc  58620
aaattctaga cagcaacaaa atcgcgctgc aatacaacta taccgttgat cggcgttaaa  58680
aagtatgcag acacaaagta aggcaacaat aattacatta attcatcagc gaagaacata  58740
atcaagcata gctggagtgt tacactggtt acatgccaat cggtagaatt cattaggaat  58800
```

-continued

```
tggtcggcaa catcgtacct ccggcagaag aagcatactt tgtgctgacc aatgcaattc  58860
gttaggcgag cagtctccct ttgatgtttt agcatcgatg aagtgatcaa tacactgacc  58920
atgtgtcgga tttgtgtgtg tatgtatgta gtctggcatg ctctctctcc tgtctagcga  58980
aaatttcaaa tatcagtcaa atgtgttcca gcagcacatt atcgggaccc gtctagctag  59040
tctccacact cacactttcc atatttttca caccttggtc tgaatttgta gtcgtccccg  59100
tgcgggcatg gaaaattact gtgcaactcc ggacggtagg tgttgatgta tgcatccaat  59160
aaacacttca cgtgttttgc caggtttcgc gtactgcaaa cacgggcttt ggcgtgccgt  59220
acgcgtacgg ctgacaagcg cgtgcgacaa atgttaactc gccacctcaa tcaacaccgt  59280
agcgtaggac ggcgaacggt aggcgcactc cgccgggatt gacatgaaat ttcgaacgtg  59340
gttcgaacaa tcgacctcac ccttacccaa tgatttcgcg ccgagcgttc gaacgggcta  59400
attttcagaa gggaaatcgg caaatggatg gatgtgtttt tccggccgta ttatgacgaa  59460
tgtgtgcata tccgtgtatg tgagtatggg agcatgcccg cggtggtggt tggcggtggg  59520
caaataataa aattcaattt aattaaaatt gaaattaaaa ctggaaataa ttacaaataa  59580
atcataatta tatctgcggt tagattgtgt gcaagctaat tataaatcaa tacccgcccg  59640
cgattgggac attcgcttca tcattaatgg tcacaataat gcgggacacc ggaatgctcg  59700
gtagcatcgg cctggcatac ccctgtcccc ggaaggacag gcgatacaat ttaaccacca  59760
aacctgaccg ttgttcgggc tacgatcgcc atcatcgctt tgatgtgcac ttgaactgcg  59820
gcggcgttgg caagcattgg aacggaacga aacaaaaaaa atcaaccaag tgataaacac  59880
ggcataacca gcacagaaca taacctccag taccaaccgg atcagtactg agtttcgctc  59940
tctgatccgt gtctttaatt ttctttgctt ttttatcatt ttgcttttgt tgcctttttg  60000
tttttcccag cgtggctcga ttggaatgag ccgtccggtt cggtcggaaa atcatgtaac  60060
ggcataatta ctgttaatat gtgcgcaaat aaaaggtgcg attgcatagc ggatcgagtg  60120
ttgttgccgc caccggggcc acactgtcta ccgtccgctg cgatgaaaag tgcataatgg  60180
tttcaaaatt gaatatggca acgcgtttgg ggaatgaatg gaaatctctt cacacaagta  60240
gtttccggtt gattgagcca atcgattaac actcgtttgt gtgtgctttt gattcgctca  60300
agctgtgaaa taatgcgcca actttggtag aatgttgtag ttttttcttc ggctacttta  60360
tgtgagctga tctgattgct gaaacgcgct gctgaggatg ccgttttctc aagggtgact  60420
gtgttgtgcg gcagtgtgac tgtgtggtag taatccctac gtcacacaca cacactccta  60480
ctgtatgcag cggcgaaggt tatgtttagc aaaacgcgtc ccaactgaca aagggcttca  60540
gggttattcg gtcaaattca gatcaacatg ctgcaataat cgcgctgata agtcccgcac  60600
acggagcgcc acttgcatgc atcgttgaat cttccggaac agcaaaacga cactggggca  60660
cgtatgtttg cagcaacacg gctgacccgt ggccgtgtgc caagcgtgcg cggcccagta  60720
cgtcagcgac acggccacag ctggtacgat ggatgctcag tacgctcagt tgatatgcgc  60780
tgagttgtgt cagttgggtg gttgggttga ccaggcgcta gtttacagtg tgctaggtgg  60840
ttggtcgggt gtgcctgtga agcctaaatg gaaccaaaaa gaaggttcgg agcaagatag  60900
aaataacaac aacgtgccat aaacagctcc ggtgcaaata tgtctcctcc agacgcgata  60960
cccaatcagc gcaccccagc ccagcgggta gtatcacttt atctagagcg gaccggtgct  61020
actggtgctg ccgatacgtg tcagaatgtc gtttcgcgcg ctcgcgccct atgatgcttc  61080
gtgcgcccag tcggcataca ctcctaattc gtatggataa cgttacgact cgagcaacac  61140
gcactgcacg atctgtctga caaacactct gccttgctag agcaaaccgc tttattctta  61200
gaaggagagg gaatttcaat agatcacgcg tcgtgctgca gcacggtgtc cgattgtaca  61260
ggttggaaat tgtaacgctc caggaagtag cgtagcaaaa gaccctcccg agtggatggc  61320
catgctaggt tgatggacgc cgtagtgcga gcgcttgcac tgacattagc aggaagtacc  61380
gagttcaatt gctctagtaa tgcaatcagc taaaaacagt acaagaaggc gggtgttaaa  61440
gacatttcaa acatgctgca gttgcggtgt gcggcctcgt tccattgtat gcttaccatc  61500
tgttcctcgt cgagcgtatt ggtgctggtg gcgatcgatt gcaccaaatt ggccagcgcg  61560
ttcggaccga gcagactcac gacgtacgtg tagttctcgg tgaggaattc gatcaatgcg  61620
tccacccctt ggcggctgct gaggacggac tgtagaatgg atagccgttc ctcggcgttg  61680
aagttcacct gcagctcgcc accgatggca gccagcaggt acgagctcag ctgctccgta  61740
tcgttggcac atcccagtgc attgatcagc agttgccgtt cacctcggtt gtccgaaccc  61800
agcagcttgc cgaacagata ctggaaggcg accgttggcg cggttcgcaa accgtaacag  61860
tacaccaccg ccgaaacgtc cgggtgcaca ggttccgcgt cgaacacttc ccgttccagg  61920
gcgtcgcggg tcgccgtcat gcagctttct atttccattc ggcaggccca gctggagatt  61980
acctgtcgga gatacttctc cagcagtctc tcgtccggtg ctaccgttgt gatgtccagc  62040
gttacaaaca catcgccaat caaggtgtcg acaaacagct catagagaat gtaatcgggc  62100
tgaccgcgca ttcgaccgtg gaagtagctg aggacccgat tagccgcttc ccatggagga  62160
tactcccgtt catggcgcac gtagcccagc agctcgagcg caatctccag atcgagccga  62220
tttgagcgag ccaaatggaa ggaatcgtcg atcagctgcg cccgactgtg cattggaatg  62280
gccgccgtgt cctcgagcag cgtccgaatc agcatgtacc agttcgaggg atcatagttg  62340
acgcgataga atcccgtctg attgacgttg accaaaatcc actcgttgtt cggtgtgctg  62400
gacggtacac gtaccgcttt cgaagtcatc cactgccact cgagcagagc gtcctgcgca  62460
tcgccctgct ccatcatcgt gtacggtatt acccaaaccg tgaaatcatt attaactatc  62520
ttgttaccgt agaatcggtc ctgcgagagg atcatctctc cacggtatga gcggcgaact  62580
tccagcacgg gatagccggc ttgattgacc cagctatgaa caaaccgctc cacatcggtc  62640
ccctcgggca gcgatacgac accgtcgaac gcttccgtca gtgcggccac gaagttatcc  62700
gtgttgaccg tgccgaactc gttgccctgc acgtacgtgc gcaacatctg ccgccaggcg  62760
gcatccggca gcagcagccg gaacatctga agtaccgagc cacccttgga gtacgccacg  62820
ttgtcgaaca ggctgaggat ggcattaaac gttgcgccgc ggctgaaagt catcgggcgc  62880
gtgctttccg cggcgtctgt gatgagaaca cgctgcacca cctgaacgtt gaacaggtcc  62940
cgatactggc gctccggata agccatatcg gcccccagga actcgtacag cgtcgcgaag  63000
ccctcgttaa gccagagata gctccaccac tcgttggtga taacgttgcc gaaccactgg  63060
tgcacgtact cgtgcgcgat gattgtggtg atggtcgttt gcgctcgata cgtcgtaacg  63120
cccggctcga acaggaggac ctcttcactg tacaagcagg aaatgggcgc aaatgttacc  63180
agagagtagc gttgacaaat gaaatgattc accacacaca cacacacaca ctcaccgata  63240
tttgcacagt ccccagtttt ccatggcacc ggcagaaaat tgggtaagtg ccacctgatc  63300
caccttgggc atgtaggagc gatagggtag accgatgtgc tcgtccagcg cgtccattac  63360
```

-continued

```
gcgaacgcct gcttctaatg catacagcgt ttggttgatc gcgttggggc gagcatagac  63420
gcgctgggca gccgcctcgt tctcggtgta caagaagtcc gacaccagga aagccaacag  63480
atagatcgac atgcgcggag tagtttcaaa gtacgtaaca acgttgccgt ctagatcact  63540
gaaattgcaa tcgaaagtta tttgtcacaa acacacctcg caacgtcaga gcactcgaca  63600
atcgccatac ccggcttcgg caaagatcgg catgttcgat acggccttat agctgggatg  63660
atgtttaatt cccaactcca ccgtagcctt cagggccggc tcgtccagac aggggaaggc  63720
ggcgcgcgca ctaatcgcct ggaactgcgt cgatgctaca tatttgcgcg taccgttcgc  63780
atcgagatac gagctgaggt aaaagccatc gtcatcgacg cgcagctcac cctcgaaatc  63840
gaggtgcaaa acgtacgagg ccggtgcaag cgcacgacgg atcgcgaaca cggcaaactc  63900
gcgctcagca tcctcggtat agcgcagagt ttccagaaac gtgaggttcg tgttgggatt  63960
ggatgcgtat agctcgttgg aggtaatgcg cagtccgcgc tgatgcacgt agatggtttt  64020
ggcctgctgc cggatgtcca gatgtatgtc cacactgcca ctgtacgatc ggtttccggt  64080
gtgcacctgc gtctccaggt acagcttgta gtgcgtcggc acgatgtagc tcggcagtcg  64140
gtaccgtagc tcctgcgctg ccacttcctg caggctgacc ggatcgagcg tgttcagttt  64200
ccgctcgcta tgctgcacct tcggatgcgc cgcaatggct gcagagtgca gcccgattag  64260
aaaaacaccg cacagcaaat gtagccgcat gtctacaaac ttgaaggttg attttgggac  64320
tgaaatctcc ggtgcgaaat gtcgactcca atatccgtaa tcgcaacagt ttcggattgt  64380
tttacgacca gatcgaccac aaacagttgc tcgtgtacgt acccccccgat aaccgaggtg  64440
tggggcaaat gccttaggaa aagcaatttc tcacctgagc aattgaatta tccatacctt  64500
tgtatagcaa gcggggctcg tttggattga gataagaagt cgattgagtg taataactgc  64560
cgaacaagag ctaatcggcc ttaatcgctt atcgctcgct agtgagtaaa ttcgtagggg  64620
aataattgac gtttactcaa tgacttgtgt gatttatatt tgatgtttga taattcgcat  64680
ctcatctaaa ccaatgctgt ctaaaaacga ttgaatatct tattgacgtg ggccgttttt  64740
ctacattttt gaccgtttac ttgcgcagtc atgattgaat ttggctgatt gtgaatcatt  64800
aatcattccg taaatatatt ggtgctatac tactgtataa aggatagtag cttagtagct  64860
cagaagctta gtacaatatt tgaacgttaa agaaaccaaa actgagtttg tgcatataac  64920
aaatcccaag tactagcgat aaataacgct acgcaagtaa tctatctgtc cagttgtaaa  64980
caacatgtaa taaaatggtt caaaatggcg cgacgaccgg aaatggatcg cgttaaaacg  65040
tctgcctaga gacatcttct ttcgtatggt gtgtgccata acacctctct cgctcttttg  65100
tagttcgtac cacttagact cccgatgccg atgtaatact agagtaggag gaaataatta  65160
atatcacagt tagggcacga atgcttgcgt acttcacgaa accttatgta ccgaaggtgg  65220
agttgcgatt gctcacgcgt tgttgccccg ttatatgcga ggtgggtcgt ttcgggccaa  65280
gatgtaacaa ccccagcata aggtgggaac gagaaaccgt gcccgagaaa ggaacgttcc  65340
atctaagcca gcgtggaggg ctctttgtgg gcatgtgtac ggcgatacgg caacccaaaa  65400
gagaaagggc gaaattaatg tgtttggctc gttggccaaa cagcagtcgg tttgcacaaa  65460
aaccaaagcg cctgcgaaaa ttagtcacac cctcccgggc cagcttttgg ggagagtggg  65520
agataatgtt atgtgtctaa aatggttaga catttttac acgtgaagca aagtttgcat  65580
tcgctccgag cgggagcagg ttgtgccatg tcggcttagg gtgggtggaa tgcgcgtgtt  65640
tgtgtgtgtt tgatgtgatg aaaaatgcaa ttgcgagcaa agtacgcgca caaaccccgc  65700
aggccaatcc ctctttttc cagctccttt atacatttaa ttccagccaa gcagagcccg  65760
ccgttagccg tgctgtgtga gctttttaca cgcttgagat agaaataatg gcgtagtgcg  65820
ctggttttcg ttacagtccg ctgcacaaac ccggactaag ggagggcggc tgatggtgga  65880
tcgctggtgc cgcgtttacg gtgtgttgca ttaacgaggc ccaggaatag gcagaaatgt  65940
atttataatt cagattagta acaaaatggt ggctctcaaa gtgcgattga agcgcgaaga  66000
agagtgcaac gaagagcgtg tccgtaataa atgtgcaaaa aaaaggaacc aaacattttt  66060
gcaataaata ctgtttacag ctgacggggt aaagtttact tccagcgttg caattgcgct  66120
tgaatgctcg ttcgacccgg ttgtgtgccg aactcgaagc tttctagttt attttatgac  66180
aaaataacaa acaaaatggt gtctgtcaca ccctgtaacc tctctattaa actgatgatg  66240
tcacgcagca gccataaaac agacatccca ctaagctctc tatgatcgta atttgtagtg  66300
caaaaatgta gccatattaa tgagtacctt gcaatcggac gacagtgaag gtctgccata  66360
aaagcgttac aaaataggca cagctctggg cagtctagtt tctgcgcagc gatcaggcac  66420
actcataagt gcagctttga agcgtaaact gcacttacta acgtcctgat tcatcgatcg  66480
aatagcccgg cacgccccca tccgtaggct tatccgggct gttttgctac gagcggttca  66540
ggtcgttaaa atcgatcgtt aaaatattat gggatctgtc ctcggctctt ctcacgtgca  66600
ttggagaagg tatggcgcgg tgcagatgaa gggatgccga ggaggaggta tggttcatat  66660
ttgaccacag tgcgtatttg cgaaacccga aaggtgcatc agctaaatgg tggaatgttt  66720
ctgcttttac gagtcgacag ctgtggctcc ttcgacgggg cagtcattaa actctcctcc  66780
taaaatgtcg tttgcactca atagtggcag cactgcctgg cccgatcgag ccttcgccaa  66840
aagatcgacc gttaagggag gggggagggg taaccgcgag cgatggataa ggatatcggt  66900
ggcatcgatt tcgtttaatg ttttgcctgc tgcatcgcag gccgtcgtta tgagccctcc  66960
gattagtgca tcgtgataat aagggcaaaa cactccgttg gtggcgctgc aactaactgt  67020
cggcaagaat gtggcattaa tgccggcaac gacgggccgt tttgtttaat ttcttttcgt  67080
cgtcaccggc cgactgcccg ctttgccaat aaaaccgtgc gtcgcgtgtg cgagcgtgtg  67140
ttgcctggct tgtagcagtg caccccagcc cagccagagt gcgctgatcg ctccaaacag  67200
taggactatt aaaaatcaat tttccaccga tcctcacgca gtcgtttttt atctctacct  67260
ccgctggggg aatgatccgc gggcttgtct ttacgcaggc gattaaaatg caagtgaaaa  67320
caaaaaataa aaacacgaaa taaaacacga ttaaaatgtc agtgagtgat ctttttttat  67380
tattttcgtt ccacactgca tgcatgcgta cgctttttca gttttgtaag ttcagaattg  67440
gttcaatggc cgatacggtt ggcgctcggt ttgaagtaac gaccccgcag cataaaatgt  67500
gaatcatttg tgtgcgtgtc tgtctctgtg tgtgatggca ttctggtttt tcaatgatgc  67560
gctcctattt tcacaaccat tacggaaggg ccagattcat tagccgttaa tcggaaattt  67620
gcgtggtgac gtggtaattt gtagtttatt tatttgtgat tgctttcgga cgatgccctt  67680
ttcccggttt gtttttact gcggatgtgg tgcgtgtgcg aaacggcagg aaaggtcgac  67740
tggttcccat cggaatggat tcaaatgata atctgattta tttagcaatg gcactgaggc  67800
tgacacgagc cccattttgt gtcacattgt agctgcagtg gtaagttgcc gtaaaacttt  67860
aattcaattt tcaactcacc ggcaccggaa gctcgtacag ccttgacaag gaagaaaaaa  67920
```

-continued

```
aagctttgat acatttagta tttaaatgga ctgagcggaa ttttgtgaag tacaacgggc  67980
aatatttatt atttatttta gtacttttat tgaatcgctt gcaaaaccag tcatcatctt  68040
caggaagtaa gaaacgacgt tttcaagatg ctttgactca tctgatgcac gtgatctcaa  68100
cacaacttcc tcacacataa tgccaaggaa ataagtttca ctcaatcgaa acatgtttgt  68160
gtgtgtgtgt gtgtgtgctt gtcgaaaaac gctgctggaa aatatgcgca ttttcagttt  68220
ttactacctc tccgaaaatt cggtacggtt tcggtgcggt gctcaccagc ccgcccaaaa  68280
gttacacgtt gattcccctc ggaggtcacg tcactgtcta gcacggtggc ggcgagagac  68340
tggcgggctg aaagattgaa cagcggttcg tcccaaaact aatccgtgaa tcatcatccg  68400
tggccgagcg cgagcacggc gctgcccccg ggagccaagg ggcagtaaaa catgtttggt  68460
tttacgagct tggaaaagtt tttctcattt tcctcgctca accactttgc tgtggaacgg  68520
attgcgcggc gctcgttagc gttttcgaga tgcgagccgt tgcctctgtt cttcgtcttc  68580
gaaaccactg ttgtttcgcc tgtttgattt atgtgtgtgt gtgtgtgtgt gtgtgtgtgt  68640
gtgtgtgtag tttgtgatgg aaactaataa gttttgatgc ttcctttccc tgtttgtctg  68700
catgctcttt ggtggcattt taagaaagca ctgactgaca aaagccaagt ttgtgtacga  68760
cttaggatgg tcaaaccata gtttgggagg gccttcatgt gtgtatgtgt gtgtttttc  68820
cacactccga ccagtacgct agtgcaatgt agacatcctc ccggtaagat gcatcttccc  68880
agcgagcagc ggttgcgaac caacgaacct tggcttgcat gtttttgatg agttttaaat  68940
tttggctgat ttggtaaatt tttacgactt tgtttatgaa acgatggaac tgacaaaagg  69000
cacaccaggc aaaccagcag gaatcgagcg aaaagcaaat cgcgtaacga accgcacgtc  69060
caacataact gcgcacccca tctcgaacgg tggacggtgc ggggcacgtc ttcgcagcat  69120
tgcagtggat tgatgtcttc cagcagagtt ttggcgccgc cgtccagcgc attgtgctgg  69180
cgaaggtcgg tgcaaatctg caccggaaca cggaagcacg aaaaacggaa tcgaaagcgc  69240
agacaccggg aacgataaag atgtttgaat gcgtcataaa tctacaaaga cggtcagtga  69300
aatgaattgg aaactcgcat ttgtcgtcgt caacgtcatc gggagttgtt catttttttt  69360
tttgggagga tagcaaacgc acatcaaatg cagtggccca tcacaagtgt gatctacaag  69420
gtggtggtga tgacggcggt ggtcttgctc cgtttaaacg acaatgtaac caatacgtct  69480
agcagttgac gatgcatatg attagtgaag tggaaccgcg ctttaaagac acctttgctt  69540
gcatgcgtgt gtatgtccgc cagatcgcac aattcatccc aacgacatgt gaaggcttta  69600
aaaacaaatt gaaatcgctt gaaacacata ttcatagcgt gcccggccga gaatgggttt  69660
tacttgctcg ttaacgagaa agagggtgtt tcttcagctg ctcttcagcg gggttagttt  69720
tgcatttgaa gcaaatcgtt acaaaatgca ataaaatcgt ctaatggtac ggcgtaacga  69780
cgtgtagttg tacttggacc aattggccac agcgtgttcg ccgcggaaca cgggcaacac  69840
ggggtggggt tttagttttt attttacatt ttttaaatgc ctcccttcgt tgtgccaatt  69900
gctgtgcgat ctgtcaggtt tcgaacacat ttcttcgctc tgtgcagcga acgcgtgcaa  69960
atgagcgtaa gcgtgagtga atttcaattc caaaagaggt ccagcctgtc ataaaacctc  70020
actccactgg ttccctttc cgcgcggtcg ctcgcccatc catcgctgat ggcatcgaaa  70080
atccactcgt taaacgcgaa accacgaacc gatcggcgcg gggaaaggga caccggtgcc  70140
agcggccggg cgcgcaagga tcgtaaatta taatatgatt tttattacat tttagcgtag  70200
cataagccga ggccggctga gagacgttcg taatttgtta taatgttata tggctttccg  70260
ttcccgagcc gtgcaccgac acactgggcg ccgacaagaa atggctcagg gtgtactgtg  70320
tgtatgtgtg tgtatgcctt tgctgctatt gttattttta tatttccttc cagtcgaagg  70380
aaacgggtgt ctttggagaa tggggaagct ttgcacaatt gtaccccagc ggagactcac  70440
tctaataacg ttcattttca acaaataaaa gcattgcatc agaactatcg tcagagtgtg  70500
tgtgtgtgtg tgtgtgtgtt tgtgtgctgc tgcgataatt tctgtatcgc tttcgtcatc  70560
agttttattt cgttatttta ttttacaatt gctcgtgaag tggcgtgcaa acgcaattgc  70620
gagccgcttt ggcgagcaag gaaccgcgcc caagatcggt ttcggttccc ttttctttgt  70680
gaatcatggt tgtgaagatt tgttgtgcaa aaacgccaag ctagtaacga attggtaaaa  70740
taactgcgcc actgcatgca caaacacaca cacacacgca caggcagagg aaaaacgaag  70800
agtccggata caaaattgcg gttttgtagc ttttatgatc caattagctg tagaacaaga  70860
accgggacga tgcgaaaggg gtgttgtaag acgcacacag gcacactggt ctgggcatgc  70920
tagtcgatgg aaattgaatc agcggatatg cgttttgcgc acatgccttt tttcatcctt  70980
ccctttacc gttgaggcat gggaagtgtc ataaactcgt gtatgcgatt tgttgttccg  71040
tcaaggtttc gtttgactga gttgctgtaa atcaaaataa aataaagtgg ccaaagggcc  71100
gggacgagca gaggaatgtt tccaacgcat gtcttggtgg tgtccaacaa tcctcattta  71160
tgatgctgca ttgtcaatgg aatggtctca tgtggtccgg acacgtccaa tcacatttat  71220
tgcttcatta tgccgaacga agttttattt cggaagtgtg gaaagtatgt ttttttaact  71280
cattcgaaca tgttcctttt caatataatt ttgtatagct tcgacaagga attcgctagc  71340
agttattcaa caaataatta cgcatgcaat aatttgtcgc atgcaaattc cggtttcagc  71400
aaaagctggt ttttaaaagc tcgagtaaat gtgttcaaca tcctgctatg taaaattaac  71460
tatgttttgt aagtgttcca atcagtcaca gaacgccaag ctgaggaaga gtatagtgtt  71520
atagaacttt actagaagcc agttggattt tgttcatccc cacactaata agacagacac  71580
aatttacatt tgcgtagttt gtgcttttgc ataatacatt taaatgtaga aatttaaata  71640
aatagaatca taacattatg cttctggggt aaagtacagc tagcttccat ccttccctac  71700
attaaaatca attgaatgct gccatataat tacgtgaaaa gaagaagaaa tagtttattg  71760
cggtgtttta ccgctattat tgcattaccc gcagcaccgt cagtaggagt agtgctatgc  71820
ttttacctaa tcataaaact agttattata taccttctgc acacccaagt ggcatgattc  71880
gttgtgttgc cctttctccc catgctttgt gccgattccc aacagcgagt gtgagaacac  71940
ccgtacaaga aaagccctat tcttcccacc cagagcggga atagtatacg agagaccctt  72000
gcacactttt ccatcgcgat atgggtgtaa tggtcggtgt tggggtgaat tttccagatc  72060
ccctcaatat tgctcgaggc tttcgattgg ctcgggctgc tgtaatagtg tgtaatgggt  72120
gtgtgggcac tccagaagat ggaaaccatt tcgtataaaa caaaagaaac caccccatgc  72180
tcgagaccgg tgcgatcgct cgaatcgctg aaactccacc gtcacgagca cgacgttgtc  72240
tagttgggct ggatctacac caacctgtgc tagtgcgcgc gactagatgt gcatgtaaaa  72300
aaataaacat ataaatcaac aatgctcggc gtggcaagca tcaaagcaag taacggatag  72360
aaagagcaaa ctcgagggag caaacttcga cgccaacaaa ccctcccgcg cgcgcccagc  72420
actagctatg cactcgaagc gcatagcgaa agatttacgc ggggggatac ggttggtgtt  72480
```

-continued

```
ggtgagcatg tttcgatgtt gcgccccatg agcatgtttt gggccgccag agcgagacgg  72540
gaagagcgcg tgcgaaacat aagacagagg cggagtcaac cctaccattg gttgcgctcg  72600
tcggtcgttc tgttgctccc gctctgatgg gtggcgcgcg agcataggtc tccgactcgc  72660
tctagcgcgt tgcagccgtt ccacacacct ttttgcacgt gcggctttgc caccactggc  72720
tgcggcacaa attccgaccg agcacgtggt tcctctatct acatttctgc gccaaccggt  72780
ggatgtggac gtctcctggc acatcggtcg aactgtgtgt gtgtgtgcgt agatacaaca  72840
tctcgttatg ttgtgcctcc gaaagccgaa caccctcgac cgtcgtcatc gtcggtgtcg  72900
tcgcggtttt atgctccggc gaaactgctg cgaacgtttc actctcactc tgtcccagtg  72960
catccggcac ggtatctttt gcatcccttc ggcggtaagt ttgggcgttg cagcacgatg  73020
ttacatcgga gcactccgca aaaagcaggc ggaagaagca gctagcccga aaatgtgtgt  73080
cggaaaattt caccatcagt tcgggagcgg agaggaggcc gcttttccga gggaatcaac  73140
aaacgatttc gctgcttatt tgaagaagca gcaaccatct acgaacggtt tcttcaaacg  73200
atgaagcaca caacgacata ccattcggct ctgggggaaa acatgtttta gtgctgcttt  73260
tcgccacgta tgtctaaacc gaaaaagaag aactttctct atcaacggaa agactatttt  73320
tttcgcctgt ttcccaaacc ttaccataga aagaaggact gcaatgcgcg gatacgacag  73380
gaaaagaacc atttagcggc acatacttgg gagagaagca cgttcgtagg aaacaaggat  73440
gtttatgtta gcgcgaataa ttcagacacg ctctgagcgc tttcgggtga gattagcaat  73500
ggagcattcg ggcaaacgaa aagaacgttt gcgtttcgaa tggggcgttt ttgcttgtgc  73560
agcgatgacg agtacctcgt ctaaaggcag tcagctatcc ggaaaacgtt gctctcgatt  73620
aatgcccgtt ggtagcatcg cacaatagca taaaagcaca taagacaagt caccggaagg  73680
ctgcataaca ccgaaaggtt aggagaaaaa aaataacgga cgataaacgg gtacaatctg  73740
agttggtatc tgagctggga aaagggctga agaaaatagg agcagtagaa gctttatgta  73800
ggatttgctc atcgaatgaa caacgtacta aagatcgttt tttacacggc ggatttatgt  73860
tggaacaagt cgttaaatag cgagctttgt tgggagtatc aaataaaaga aaacctcatc  73920
acttaccaag agcactaaaa gagatttagt caagtagtgt tgttagtctt tttattagct  73980
tgggatttac tatttatact tatatatctt atctttactt aaaaaatggc aaaaaaagat  74040
aaatagaaag atgtcaaatc atcaaacttg ttacattgtt ttataaactt gtttgttact  74100
acttgtttgt tataacattg ctttatacac ttgtttttac tttaatgaaa acaaacatac  74160
aaacaagtat ttatttttca ataatccgtt atttttagtt atatgactaa aactaatatt  74220
gcaataaaat gactgcactt cttattggtg ttgaaattcc ctgataacgc aaaaaatgtc  74280
attaaaaatt atgtgttagc taactaacta acgccatgtt tcaatgttga aacaagcaga  74340
tgccaaaagt tttttatgat tttttatagt acagtagaaa cagacgatat ttttccgatt  74400
tattaaagtt aaagtgcatt caaacggcat attggtttac gtttgaattg aatgtatctt  74460
tatgtacagt ttaatcagtc gactgattgt ttcactcatt ggattacgtt tgccttgaaa  74520
gtaacatttc aacctgtatg gcattgcgca catctattta cttgtcatgt cgctcctatg  74580
gcgctccata gttcccacca gccccaccga aaagattgat taacatcttg acgggtcata  74640
tacttattaa tgccgcccat aaaattaatc ctgcccgact atgaatcgga cattgtacac  74700
agtgcagcga ctctcctccc atgtacggta acaaccatgt tacctcacga aggtcatgtc  74760
cgcatacgcg ccaaacatga agcgtaccta agcaagtcgt gcaccaaact taaataaaaa  74820
taattgaatc aatcgagcac ggcttgtgat aaacgatccg attgattcgt tagccggatg  74880
cagttgcagt agttgtcttg cggttgtgga gttgcagtag ggatgggggt tgtggagggg  74940
tatgtacgtc agcgttgggt ggctacgatc gcgccacgtg cgttcgcgaa aacgaccaac  75000
cagcaccggt cttatctgag attaaacgaa cgaggtgaca gctaaaagga gaaaccgggc  75060
gattatttaa attagttccc ctacgaatgt tgtacggcgc ggcgggctgc atcggaggag  75120
ggatcttatc tcgggggtag cgttatttgc gttattgtag gcaaaaaaag gataagtatg  75180
ctgctggtaa gaaggtaaga agtatgcgcg ctgcaataag catccccgtg ccctttcggc  75240
acccggcgtg tggagctcgg tgcatcggaa gctcggattt cagctgcacc gaaaccagat  75300
gcacacacgc gcaccgctcc gggggcgttg aggcaatcga aagcaatcaa catcaattag  75360
caagtttatt tgcaacaccg ccggtttcga tggattcttc cgcatcggcg actggtacaa  75420
attgctgctg ctgcgccttt agcgggtggc agatcggttt tgccgctacc ggtaccgcat  75480
actatgaaag tatgatttat cgtctacaat catttcccat tacacaggcg cggatcgtaa  75540
aatcagctcc ggaaatatgt gtgtgggttt atgtgtgtgt gtgtttcggt ggggatgaat  75600
cgaaaattca tcttttgcta gcgggacgaa gctgttggtg tggagtgccc gtgccaaata  75660
cgttgaaggt cgcgatgtac gcgattctct agccttgctt agtcattcag cgggaatggg  75720
ttggttgttg cgctcgcatt ggaaaggtgc attctgcacc gaagcattcc agtagcgcac  75780
gccgatcgtt tgctcgatta tggtttgttt agtctggatg aataaaatat tgctcaatta  75840
ttcaatttat cgcgggcctg ggcccggcag tggcaaacag gactgaaacc gccgttctgt  75900
gcaggtctgt tccgcgatcg atactatcgt ctgccagtgc atttgtgtgt ttgttctggc  75960
ccgcttgttg atatgttgtg gttgcccgct tggcaaatgt gcaacgcatc cgcgaatcga  76020
gatgttgcag catggatgga cacgaaacac gagccataac tgtacaaaca aacgattggc  76080
ccaagttggt ttataattgc gaagcgtgcg ttaacatggc gatcaagaat aagttcataa  76140
tcgatggatt atgagcttga gcggaattgc aaggacacga aattgataag cacaaacaat  76200
gaatgtgtat tgtgaaagtg aatggaattt caggtgattc atgtctggga aatgtttgta  76260
ccacaaattg catcatacca ttgagaagct acaattacgc agattaattt tacgcacaga  76320
attgcagaaa ggaactgttt ttttttgcaa ataaaaaaaa agattgaata ttcaacagtt  76380
ggttggaact agcgaaacca agggcccttc aacccgaaga ataatgatac gtaatttttc  76440
acgatcgatg caaaacatgc acaaaatatt gcatttaatt cttcacagct agcaccgatc  76500
gttttgtcat gatcagcgat cggtcgatgt gtgccgctgc ttgcaagtta ctattctggt  76560
attcccattc tctccggtac tggagcagcc agcttcgtgt catcgacaaa gcgcttcaag  76620
tgatgccctt ttactacaac ccacggcgaa ctgaaaatgc cagaaataga tagaggaaga  76680
tcgacaatga tctattgact agttcaggcg cgcgcgtctc gctaggattt gcttttcgga  76740
ggatccacct cggcacaatc tcggagacgg cggtgatggc ggctctaccg gtggattgac  76800
actttgacag ctctgatgca atacccattt ccagtcgacg gatgacgcga aatcgcacaa  76860
aatccaccct ccagccgggg cggaaggagg acgcttattt ccaccgtgat caaatgacaa  76920
acgggcgcgt gcgcttgtgt ttagcaggca ggggagatga gcgcaaactg tgcaagaaga  76980
agcatcactg tgaagacggc aatgcaaaga tagtgtgctc aacttctccg cgaagattga  77040
```

-continued

```
agctaaatta agcacgagat tagcatgact gaagtgactt ttcaaagtgt cagaatggct  77100
gcactcgcaa actagctgga tgcagcgcaa ttttgccccg gtgtgtgcgc gcatgcaaac  77160
gagcaaccgc agagggcaaa ggagaggatg ggaaggaggg agggagtgaa agagcaggct  77220
taaggttgcc ctcgggcatt gaagtcgata cagcggttct attccagtgc cagtaacgat  77280
gacgaagacg atgttgcttc tgctgctgtt gctgctgttg ttgttgatga tgatgatgat  77340
aatagtgcaa atataaaata aatcttccgt aagctttgtg tagtggtgcg tggctactat  77400
aagcccgtct ggaagcaagg aagctagtcg ggcagggtca tgcaaaaggg agacaccttc  77460
ggagctccgg agctcccgcc ggcactctcg gggggacgtc cgttatgcgt tgtgatttat  77520
tatggaatat ttattatagt gtcttgtttt gaaaaaataa cttcaacggt tcgaatttcc  77580
tacacctcga gatcggggct ggagtggcaa cgtggtacgg aacggtacag cggtttgagc  77640
cgttcggtct tgggactcac ggatcgcaga atgttattgt gcgcgcactg atgggaaagt  77700
catttttcac cgagtggtca gggcgcgtag tccagttcgt ttctggctgc tgttgctgat  77760
gctacgatcc tcaggaatga ttggaaacgc ctggagatgg tgggaaaaaa tcaaacacaa  77820
aaacgatcct aatgaacatc gtgtgttctc attcgctgcc acgattgaca ccttcgataa  77880
gacgcacata atgagctaaa ggagagggga cagggtcttg tctttgccac gagcgataag  77940
attgcaatca ctcgtgagcg tgtgctgctg ggctgaagaa gaaacgcttt ccacagcagt  78000
aggtgggaag tgggattgtg gaacgtggca ttgaaaagaa cctattttct aaagcccgag  78060
agcccgttct cgaactggaa aaccagatgc agaagttttt tattgtcccc cgccaggaaa  78120
acaaatgtat ttaatgcttt ctttgccttt tccgccccgt ttcagacgac gagctagtga  78180
agcgagccca atggctgttg gagaaactcg gctacccgtg ggagatgatg cccctgatgt  78240
acgtcatact aaagagcgcc gatggcgatg tacaaaaagc acaccagcgg atcgacgaag  78300
gtaagctggc gatgatggtg tcgttcgaca tcactttcat caccgtgtca gacatctact  78360
gtgcctagca ccgggtccag tggtcacagg gtgtagcaaa aacgtgttct tttttgcgag  78420
agactctacc tcatgatgca gctgttaagg aaaggtttca gatgaaggca atttttccta  78480
ggataagatg atcttaagtt acctgcgtat tagtgtttaa cattgtcgtc tcaactccca  78540
agaatgtttt aatcgtctag ggctagttta tttatactgt tctcattgaa atgtcgttca  78600
atccaacatg ttaagttagc tagctcagac acgagaagtt aggagtatct gcatcttgaa  78660
ggtagcggca tatggtgtta tgccacgttc actgacttca aaattcgata caaaaaaaaa  78720
accaaaacat caaaaaccaa attgtgaatt ccgtcagcca gcagcagtga ccttcaaagc  78780
cttacctttc cattcattta tgtttaacac aggtcaagcg gtggtcaacg aatactcacg  78840
attgcataat ctgaacatgt ttgatggcgt ggagttgcgc aataccaccc gtcagagtgg  78900
atgataaact ttccgcacca ctgtaactgt ccgtatcttt gtatgtgggt gtgtgtatgt  78960
gtgtttggtg aaacgaattc aatagttctg tgctatttta aatcaagccg cgtgcgcaac  79020
tgatgccgat aagttcaaac tagtgtttaa ggagtggagc gagagagccg caccacggta  79080
cagaagggca gcagaatggg tcggcagcct agctgcactg gtgcggtgcg tccggcgtct  79140
cgggggggagg gcgaggaaat tctagtgtta aatcggagca gcaaaaacaa aacagtggtc  79200
gtcccgttca agaaacggcc tgtacacaca cacagaaaac actgcagcat gtttgtacat  79260
agtagatcct agagcaggtg gtcgttgctc ctcgaacgct ctggacgcac ggcttcgcgc  79320
gtatttgcgt agcgttccgc cgatcgtggg tattcgtact gccacaagcc cgctttctcc  79380
catgcaatct ctgcaaccaa accaacaaac aacaacaaaa aaccaatcga caaaatgaat  79440
cacacccctt ttgtatcatc tgtatattct tgttctttgc gttcttttct atgtggccca  79500
cgccccggcg ggtacgtaat tgcgtcgaaa accccgaaaa ccccggcaca tacagtgtac  79560
atacggtttg aggacaactt tgacctgcag cccttctggg gttgccacgt gtagctatac  79620
ttgtgagatc gggcgccgac ggtgtaaagc gcgaatggcc gccacacagt gtgtccactc  79680
caacactacc cctctggaac taccccgtcc agggatgcac cggctcggct catgcccctg  79740
caaaacagtc cgggctccac tgtagtagct ccggcgttgc tctgagagaa ggatgccctt  79800
cgaagtgtcg aaagcgtgca ttgggcgttc aagtgtgtgt gtgtgtgtta ggtttagcga  79860
gaaacagcag cagttgcgtg tgctgaaaag cgaaggagta atagagtgca taatgaaaat  79920
gaaaatgaaa atgaagcaaa agtagaaggc ggaggagagc aacctgtgtt ccactagtag  79980
cgaatagttt agtctagttt cgtcaccaat caaccttcca accatcgttc aaccaatacc  80040
tgagtcaaca tcgtcatcgt tatcgtgcca caactttatt aaaaatgaac cttgtccgcg  80100
ccaccgtagg gtgatctaag gcgacctttc ttacgggcgc gacccacatg ccatcgtcac  80160
cttctccaat caaaaccaac agcctgtacc gatggtgtgc aattgtgcgt gcgtgtgtgt  80220
tattagcaaa aaaagagaaa gagtcgacga gagagagata gatcgagatc gagagtacaa  80280
aagagcagta gaaatgttcg ttgtttgttt ttcgtaacac agttgtttag ccaaaatggg  80340
aatttccaat aatcccgggg gcggggaaat gcgggaatac tgcgtacaca catacatcaa  80400
tcaaaaagaa aaatccttgc gctacatcac taccgtttgc gcggtgctga tctagagcag  80460
accactttcc actccactct acaatcaatc aatctgtgca gaaggtatgg taagacggcc  80520
tttgagcgag tcacggtcgc caccataacg ccgtccgacg agggctgaat gcgaactttg  80580
ctaatcgatt ttccgctttc tttttatccc acctcctttt ctctccctct ctctcttttg  80640
cactgcccct tgtaaccccc aaaaaggtaa acgacacatt aagacctacg aagcgttggt  80700
gaagtcatcg ctcgatccga acagcgaccg gctgacggag gacgacgacg aggacgagaa  80760
catctcggtg acccgcacca actccactat tcggtcgagg tccagctcgc tgtcgcggtc  80820
ccggtcctgc tcgcgccagg ccgaaactcc ccgggccgac gatcgggccc tgaaccttga  80880
caccaaattc aaaccatctg ccagcagcag cagcaccggc tgcgatcggg acgacggtga  80940
ctgcagcgcg ttcgacgaca gtgcctcggt ggtgcggggg cacgggcgga cggcccacag  81000
caccggtagc aggggccgca gccactcgaa acggtaccac accctcccgg ccgagcacat  81060
cgggagccac atggcggccg cccagagtcg atcgcccgcc ccggacgacg agccggtggt  81120
gtcggtgtcc gtgtacgaga gcctggtcga agcggccagc aaaaagacgc gcaccttcag  81180
cccgccccgg ggggaggcgg aagatttgca tgccgcacgg aaagcatcgc cccacgacga  81240
gcgggacgag ccgaccccgg cccagcccta cgaagcgtac ctggagtcgg tgcggcggag  81300
taaaaagtgc ttcgcgctca aggacagcga ggcgccgggc gaggagccga cgggctacga  81360
gaaggagaag gagccgcgca ttccgtactc gctgccgaag agcaccttcg agcggctcga  81420
cctgctgaag aaaccgaacg ggctgacgtt tccgatgtac aagtacagcg ggatcgagcc  81480
gaacaacttt gccctgccgc tgctgctgcc cgggctggag gcggtcaacc ggacgctcta  81540
ctcgacgccc ttcccggccc agctcctgcc gtccagtctg tatccgtccg ttagcagcga  81600
```

-continued

```
gtccacgaca gtgcccatgt tccacacgca ctttctcggg tatcagccgc cgctgcagct  81660
gccccacgtc gagcacttct atcggaagga gcagcagcag cagcagcagc agcagcaggg  81720
attggccgaa ccaaaggaac cgacgtcgtc gtcttcgccg ggcagcaacc ggcttacgcc  81780
accgaagggt gcatttttct acgcgagtgc ggtggaaaat tcgctcaccg cccagcaggc  81840
ttccattgct accatccatt agatccacac tgcgtccact cgctgtttgc tgcagcgtac  81900
cgcggacagt gcagtgtacc gctgtacaaa aaggtaagtg tgggtagtaa gcggtagggt  81960
gggatgggta gattagacag taggcaagtg gggatgcaaa tttacagccc ttttggtcac  82020
tttaacagac acaacagaca agggacgcta gcacgaatca tcgcaacaaa atggaatgaa  82080
gcaaatggcc tttggacatt ctttgatctt cacactgttt ccgcgggctg gggacgttat  82140
tagaggaaaa acgccaatat gttgtcgtca acattggttc cgctcccagc ctgggggctg  82200
ctttacttct gccagtatcg atcatcgcct ggtatcgctc ggcattaaat aaatcattca  82260
tggccaaatc aacgtttagt tattgatatg ggcaggagga agcaaacaaa cgaaaaaaaa  82320
acgggcacac tccatcgaac tggatactgg aaactctgca ccctacgctc accctcattg  82380
caccctacca gagccgatat gctgcaaaat tctaaataaa aataatccat gcgggtcgcg  82440
aagcaaataa tttatttcct atttatattt atttttaatc acacacaaat atgggtgcat  82500
gcacgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgaccga gtatggacgg acgatggaca  82560
ctgtggtgca aatagcggtg agcggtcgtg gccgaaggtt ggctaatgca acgcgttgtg  82620
tcgcccgttt ttccgagcgt gcctgatttc caatgcctat ttttcactcc actgccgctt  82680
tggtcgccat tgccttcggg gggacctttt taaggcaaat gttgatttgc accgacacac  82740
accgaattgc acactgcacc cagtcagtca ggcaggtggt gttgtttgaa aatggcgctc  82800
tggagcaacc aacaaacgaa cacaaaacaa aaaaaaaaca aatcaataga aagaatcgag  82860
ctgtttcgat tattcaaaat ttatacacaa aatatgcaac gtattccccg gtggggtacc  82920
ctcattgtcc gacctactcc cccccggtgc acctcaaacc caccggcagc aatcaatgta  82980
ataatggtaa agggtggcgt gccaaatact cccggaccat tccgcgctcg acgtagggac  83040
atacagagag cgggagctgc agtgacacga gtgaaacaac ctggagaccc ctgcattcgt  83100
caggcggaaa taaacaaatc aaaacaaacc tcccgtctga tctcgcgacc ctgccaccca  83160
ccggcagccg gcaaccagtc gtccaatttc ggcactttgg cggtgtgcaa ctttagcagt  83220
ctatgcacat gcattgtaaa tatgcatatt gcacgagata aagagagacg ggccgagaga  83280
aagggtctct gtgagcgggg tagccagaag tatcgaacga caaactatgc gcgtattacg  83340
agatgcgatc ggtttgacac tcggcattcg cactttggtg gctattttta ttcgcctgct  83400
taactccgtc gctgtttgtg cgtggctgcg tgtatgtggc cgggcgagcg tttgtttaat  83460
ctggcacggt gcagtatgca gttcggatgc cagcgctcgc cgccccctgc accactgacc  83520
acccgttcca tgcccaacga cagcaacgtc ccggcagagt gatcagcaga agaaaggcgt  83580
ttcgtgccaa ttctgtcgta tacatcgtgc acggacgcgg attgttgacg aaaggttttg  83640
tagcaaaccg ggcggcgaac aagttatgaa taaatttact ccattcgtta tccactgatg  83700
tatcattaat ggcagccggt cagctatggg gcgctatggg cagtacagtc ggtcccgggt  83760
gtgccgatcg gtaaataaag tgatttttgc attccgcttc cgtggtagct aattttgtgt  83820
ggcacacttt ggagcgaatt gtttgattag ggctcgtttg ttcgcttgac tgtaagctat  83880
catccgatga aagcgggctt aaatgctaga tttactaggc cgatcatttt gacaggtagc  83940
tctaggagct tttcattatg cctaattata ttgtaaatat ttagttgtgc atttaatgca  84000
aacttccaac aaatgaaaaa gtcattctgc tcttttaagt attttaatca gtattttcaa  84060
agctttaagc acaaacgctt agaacgtttg atgtttttag tattttatct acttatttgt  84120
ttattgagtg cccctgacat tcgtcgctca caaacaataa atatttttgg acctggatct  84180
agtaaatgta cgacatagct cgaattgaaa atcaacgtca atatctctct aattttatgg  84240
tctaattgca tagagaagat aaaaaactat ctattattta ccgattagaa attaattcta  84300
gtatcctcct gctagtgctc gaatcgaatt catttgcatt ccttctgctt gctagccgca  84360
ggtacagcaa tatcggaaac tctttcttta atataggttt aaagagcctc taatgtgcat  84420
ctttgcgctg atcgtaacgt ttcaccgaat catcaacgag tgttgttttg ccttctgcaa  84480
tgaaaccatc ctacactctc acgtgtttga aagaggtcca cggcacaccg ggaatgcatt  84540
atgcgctgac ggcggtggtg ttttgttcga agttcgtgat gcaaccgccg gggaagttgc  84600
acacagggat ttaacgactc ctcgtaaaac ggtattatat atcgaggccg cagcgaaagg  84660
taacgccgca gccgcagcaa acggctacac aaaagtaaac ccctctctgc cgcactcgtt  84720
gcgcagtgcc ggaccgcatg gcgcacatct tcgaccagtt cgcgaggtcg ctcaatacat  84780
taggaactaa tatatattcc aggcaataat aattttctat tttactgccc ttcgtgggga  84840
gatgctttgc gagtggtgct ctgtgccagg agaggcagag aaggcatacc caccaaccac  84900
ctccagggtt tcaaacacgt tccctgcgct tatcgtgaat cttttgcatc ttttgatgat  84960
cgatactcct cgggcccggg acaagaccaa cgccaaggtg caccgtgtgg accaacatcg  85020
tagacgacaa tccgtgcgtt gcgttttggc aaggaggagc tgtacgaggt gagatagagt  85080
gtgtgggaga aagataggga tagcacaaaa gagtgtgtga gagagagaga gagagcgcac  85140
ctagaataac agctcgcctg actgacttga ctgactggca ggccatagaa tggtggcgag  85200
aaaaagcgtc ttacaagacg cgctaaatgc aactttacaa cggtcgtaaa ctaggtcgta  85260
aatatctttg ccagcatacc ttctgcaaaa gagcagatcc cgcaaacaca cactgcgtac  85320
ggcgcaacgg ctgccactcg tgatgcactt gtagtagacc ggggcccgat ccgaaccgtc  85380
ccggacgcgt tttgctgacc gaaacagaca cgcacacagg gtgcattttg ctaattttta  85440
tgctaaattt ttccaccacc gacatgggat agtttccagc tgagagtgca agtgcacttg  85500
gggtgcaagt tgtcgcatgg agcgcgataa cggacgcagt ccactgctca tcttagcctt  85560
atacctgctc ctggaagatc cgatatgtct ccaatcagta tcgtcggcag tattttacga  85620
taatccgcag cgaacgggaa ccggccgcct tggtagcggt ttgtcaaacg gatctgcact  85680
ccgcactacc gtcatgacgc gattagaggt agagcagcat gccgtactac gctaccactt  85740
gcaacggcaa acgtcgcgga gcaacattgt ggccgcagcg ccgaagcaat aaaagttgga  85800
ggacatctgt gagcagataa tttacaagct actttgtata atgaaaaacg cattaaaaaa  85860
ctacgcctgg caaaagttcc tagttgttct tagggggggag gaagttggag gggggcaatc  85920
atttgcgaac cagactgcga aactgttaca agacaaaccc ggagcatttc cgggcgatca  85980
actcatgatt attgttagac tcgcggtgac gagctgtgaa gcgtcctgcc ttttcggacg  86040
ttgtgcgaaa tgtttcgcac tgcagcacgg cgggtgttcg atgccgtggt gtagttgcgg  86100
tttttctaca gctctcacat acacataacc ggcatgaaac acggaatgcg agcgatgcga  86160
```

-continued

```
gctgggagtt ggcgcatcaa actccactaa tgttgcacac tgtgtggggt gggatcaact  86220
tcttcgccgg cgtttgttac cgcggtggtg ccgatgaaaa gacgccatag atggatttta  86280
gccaaagaca caccgttcca tcgtggccga acaacggttg caacggtgcg ctgggcagaa  86340
ggtaatggaa ccggttccgg tactgatcgg ccattacggg ctagtgaatt ttactagttt  86400
tcagagataa ttttatgggt ttccatttgt gggaattgct ttttttattg cctcaactgg  86460
ctgtgaggtc tctcttctgg gccggtgtgt tgtttcagca gtttcgttcc tttgttcgag  86520
cggttttgtg cattgtgctt gatgatatga caaacccaga aaacaaaaca aaaaaacgat  86580
aactacatgc gtctggttta tctggctgta aatttagttt gcagtccttc aacacacaga  86640
cttacacaaa cctcataccc taatcattgt gatggatatc gttcagtatc acgatgttat  86700
tgaggtgtgt tcacatattc ctaatgaatt acattttttg ttttatccat tttaaatgat  86760
gaataaatat tctacaaaca tgtataaact catattaata aacctattgt ccaaattaat  86820
attaagtggc gtgaaacgat acagcttatg cactacgcaa atattacgag aatatgatct  86880
aatttgcagt gaaaatttgt tttccttggt tccaatattt ccacaacctt atatatcatg  86940
tgaattattt taaaataagt tatcatctta gaaaaaaatc atcatcagat caaacatcac  87000
tagatctcaa agttacatca agccgttcgc tctgaattgt agttttattt cgagtgtttc  87060
aaataattta ctttttctc atcatactta tacacttttt ctcgatttct ttccgcttcc  87120
tcaaaataga tcgattggaa attcacgtca atcatctgca agcccgaaag atgctaccta  87180
gtcgtcccca gctgttgcta ctggagcttt gcaagagatc cagctttcgt tccttatcga  87240
tgcacaaaag gcgcacccgg aaacaaaaca aaaatccaac ccactcgtca acggcccaca  87300
tggcgggttg cactggagaa actcccaccc tcgtaagtgc tatctaagcg ttaaattacc  87360
ttcgcccttt gcggtagaac aaaatagaag caaatgaaac aaaaaaatca ttgccggagg  87420
cgcaagtgaa cagcggaaag ggaaagaaac ccctgtcgaa cagaaaacat gattattgat  87480
atttttcgat cgtgcaacga aggtctacac tgtgatacaa aatgttgtgt acaggataaa  87540
tattagattt ttttgtttgg aaaacaaaaa cacagctaaa cggtaggaac aaaacaaggc  87600
aaaccgaaca aaacgaaaca gtacgcacac ggctcgttgt atgtaaatca atctatgtga  87660
gcgtgtgtgt gtgtgtgatc gtatgtgatt atgtgtgtgg cgaacggttt cccattttct  87720
gtgagtaacg ccccgttacg atcattgctg ttggaaaaaa agctaaaacc aaaccttcat  87780
cgaaacgaat ggcgcgcgtt ctttacttgg cgcccaattt cccaccaaaa ttcaaacctg  87840
tttttaatag tgtaaaacgt aatgaaaata gtaaacgggc gtgtgttgtg tgtagcatgg  87900
ttcgatcact tggaaccaaa atctcaaaaa aaagcaaaca gaaactcatt ggcagaaagg  87960
cagacacacc ggaattgcga agttgggaaa gcagatcact ttcttgttat gtctgcgttt  88020
atttctcgtg tgcgaatgga aggcaggaaa ttcagaggtt catctcccat ggaagatgac  88080
ggaaagagat taagaaattc gaaggcaaat ctgttacaac ggcgagcgat tgtgttatgg  88140
ctagtaaaga attgaattgt gatacgtgcg cagtactgca tatttgttca atttgtagct  88200
tgtaggtaga tcgccgtcct cgtgttccgt gatccggggg cgggatgata gactccgcca  88260
cttggagcga tatcccatgt tgctgtactc tcgtttcggt gcctttttt cttgctcttt  88320
cgttttacaa aaaaagtaat tatattgctt ttgttttatg tgcgcacccg cacacacagc  88380
tgcacacgat cgtacaagtt aacgaatggt ttagtttgcg ctaagtttga ttggttctag  88440
ttcgctaagt tagtctgtag agagattcgt ttatcgttat gttcagcagc agtgtcagga  88500
acgagattgg aagataatta caggggcagg gcagatgagc aaaggggta cggttagggg  88560
ctggaagtca aaatgcttta gccatcctgc agtcgaattt aaacattaaa aaacaggtcc  88620
gccttgacga aacaaatacc cccgaggagt tcctgcgccc ggcccctcga atgtgcacga  88680
aatggaatag gtgttgtaca ggcagaagac agttgtagaa gcaagggtgt aatgttccaa  88740
ttgaaaagcg aagagaaaac ctaatgtaac tacaaggcag atatacagct gagagctata  88800
ttttacgcag cgaaatacaa tgtaatccca ttttctccac tcatcaaacc ttcattagtc  88860
cttcacattt cacacaagca agttgtacta taatgtagaa aaaagtagaa caagcaaacc  88920
atttgatgca tcatcgtcat ccagcttgaa aacaaataga tcaaattaca tagaactggc  88980
aatgtctatt gatacgctgt tcgagagact tttttttaac acaaccgtaa catcagtggt  89040
gccgcgtgaa tgtatgttta tttctgagta taaagaaaaa acaacaatgt gcatatatac  89100
tggtgtgcag tcagctcttt ctgagagaat aaaaacctta acatttcgct ttgcacaaac  89160
catgtcttgt aaaatattac tccaacaaga aggacagtca aagaaagaaa caagaaacaa  89220
aacgttaaac ttaaatcaaa agctagaaat gcacatgtac catacattat tgcccagaaa  89280
ttatctcaac aaaggggaga acaaaacaca gttacagcca acagaaaaca gttacagcaa  89340
aggtgtacat agcatagagt cacaacacaa tatgtacatt ttacccggtt caatatcaaa  89400
ataaaatgaa aaaaaaaacg tcccgtccgc tgatgacgga gtaatgagac gaggcgtgaa  89460
aatgaaaatg caacatcaac agttaagaat caaaataaca aaaaacaccc ttatccggct  89520
ccagtacaca atctattgat gacgaaacgt gtgctgcgaa taatgtttta acaaaagatg  89580
aagtaagtag aacgtgtttg atggaagcga tgggcagcaa aggtaacgaa aacacacatg  89640
ctaaacgtca tgtgtagcat gtgtataata gcaagaagaa atttcagagc aagacccaag  89700
gaaaagtatc tttgattcgt caaacgccgc aaaacgctgt tttactgctg taagtttgag  89760
ggaaacaacc tccggtaaaa gagaaataaa gtggaacaaa gcaaacaaac aaacaaacaa  89820
acaaacataa ataaattatt aatattatta ctgaactccg tcgtgcgtgc tgtatttcga  89880
gtcgctttgc tcgccaatgt atgcgtccga aacgatgtgt ttatttagtt atttttacca  89940
ccaacaacca gatggtggtg aagttcaaga aaaaagtagc tgaacgcaac gctgcgtcaa  90000
tttctctgtc tccccaccgc ctttctctct ctctctctct ctctctctct ctctctctct  90060
ctctctctcg ctctctctcg ctctctctct cactctctct ctcactctct ctctctctct  90120
ctctctcttt gatttcatcg gatcagtctg aactttgcca tccaaacaac atttaattac  90180
ggtcgtcggt attgaggcat agttttatca atcctggcag cgggactcga atagagagat  90240
gcacttttcc cttttccatc ggagtaagga cgttgtgagg atggcaaaat taggttgact  90300
agtttagcaa agcggaggag aagagttttc aatggtttca ccgttcttag acgcgatttc  90360
ttcttcccag ctggatgagc cacagtttga gccggtcgca ttgtactgtg caaggatatg  90420
aaccggaatg gtggcggaga tgagtcgtgc tgatgcggtt ccatccagtc tccagacccg  90480
gtaatcggtc cttggccctc tacctttctg aaacggtcct ctgcaaggta gaaaataggt  90540
ggttttctac cccgttttgt cttctctcac tcttgcgtcg ttgtgtgcaa agtactacca  90600
gaagtacagg caatcatgat gctgagatcg tgatgctgca tatccgtggc gcgagaacga  90660
atcttcactt tgcactgtac gggggaaatt gccataaaat gcgacaagcg gtacggtgga  90720
```

-continued

```
aaacaaaact gtgcattgta cgcttcaccg aaagatgcca gcgaacgcgg gcttgatgct  90780
ttcgtacttc gggaagtttt ctttttttta tttctctctc aattggagtc tgtccttcgt  90840
gccgtggaaa ccccgtaatc atgcagcacg gtaccgagag cgtggctcag gcacgaaccg  90900
tcgcaaacgt gagcatgtgt gtgggtgctg tgaaatggga agcatcgata cgataagaaa  90960
ctccagcaat cgattgtgcc agggcgcaaa gccggagcaa acataaacat gcagctcatc  91020
aaggatgggt taaaggagtc ggcaactaac cggctacaga acgaaacagt gaagcgcgaa  91080
gaagcaattg ctaaccgtgc ggtcccttgc ctgaccgaac aatagtgaag ctcattttcc  91140
aagcgacgtt ggttggctgt gtgggctatg gggtaaattt taaaacttct tttggggaag  91200
tttttggaag gaaaatttca ttacgtttca ccctattcct ttgcaagagc gggtcgtgat  91260
aagatctctc gatggggacg tgctgcgaga caggttgata gtggcgagaa aacgtttgac  91320
gagcgatatc attgaaaact atctgcaaaa tgcttcacca gcggtgtgca cttagatgct  91380
agagtttagt tttcgttgct aggtgtgcaa gtgtgcaaaa aatattctta caatcgcttg  91440
ttacttaaat tttattacag atagcgaaca aagaggatgt tatgtttcag ctacataaat  91500
ttcattcaat aagtacattt caatggtaaa acatctccct tgtgttaaaa tctgtacaat  91560
tgttgagaaa tttcaatgaa gtttataggt tactaattac cgtttattat tcataaaata  91620
acaacttagc ccctggacaa ttcacggata ctaggatgtc caagggtatg tgtgtaactt  91680
tatcatagaa taatttgtta tcctaattac ttcgttttaa cagtgtatcg ctcagttcta  91740
cgtcaactat ccgtggttca gtagctgaat tcccgcgttg gaatcgcgtt ggttctaggt  91800
tagtatctca tatgcagatt ggttaacatg atagtcaata atgtttaaat ccatgactga  91860
acattgaaga atatgataca ttttatgcta ttgctatttt ttttaattca tcacatacca  91920
cacggtacat tattgatttc agaaaggcat attttgatta ttatataatt aaaaattaca  91980
gctatttttc aagtaaacac caagctcatg cattaaacca caataaaatt gatttttaa  92040
ttacactcaa cacgctaaca ttttttcaaa aaataacatt acatccatta catgccgttg  92100
atgaatacat aaattacgcc ttgtttttga tgcacgataa tttttatttt gcgcaccttt  92160
tgcccccggt cctatacaac attaccatga ttcgtacgtg ttcccgctcg gcaaatctcg  92220
ctaatcaacc gttcaacaat ccatacatac ccgacgttga tcgcacacga tgtaacgcgg  92280
accggctgga gcgattttgg cttgcccgac tcgacacaac cgatcgacat caattgcagg  92340
gattaccggc acgccatcat caaccgacat cgcctcggca aacgcagctc caatcagcag  92400
gggctaatca ctcgaagcag ggatgcccgg ggagcagaga gaccagaaac gctacattat  92460
ccacgcggct gctattaagt ttcgcccaca accagcgcgc acacaataat cgtcattgat  92520
cggcaccggc aaaattaaac attggcaaac acaacggcaa ctacaaaaac tccgatcaaa  92580
cggtcacggt ctgaattgag ctcaaggggg atggagagcg agtgagaaag aggtgagata  92640
tcatattcca atcgatttta ttcaaattct taaataacat ttatcttccc gatagctgat  92700
tcattgccgt cgctcacgcc tgcttgtctg cttccgctcc gttcgcgttc tatttgctac  92760
tgcattattt ctgctgatgc acccaatcat cctatctccc accctctcta tctgtactga  92820
gcaccgggca gggcgaaaaa gggggagcgg cagcaaaatg cattccccgg agaggaacaa  92880
gaagaagaag gcggtgcaac aaaaaagcaa acccggatca tcccggctcg gtggaaaata  92940
gattacatta tttgtgtttc attttgtagt atatacgtgt gtgtgtgggt gtgagtgttt  93000
gtagtttgcc ttaaattgtt ttataattac tcttgtgcga caaaacgccc ctgactagag  93060
tgggttggga gcgaacacca caatcgtgaa ctggacggga gaacataatc cgatgtcctc  93120
gggtgatttg atgtacgcca gggaaagcgg atcatcaaat ggtgtatact ggcaaatatg  93180
caaaaacttc ggaaaagggg aactggaaca ttgaaacaag ctattatgca ccttgcactt  93240
tgtcccacca actgtccagc aattcgaaat aaaatgacag aagcgaccgt acattacact  93300
cccatttttt tgtcttattc tacatttcaa tacttttcgc cgggtgtttg acgggaatgg  93360
aaaaggtgtg aagcgcgttc aatcttcatc atcctttgcc cacatctcga cctgcggacc  93420
tggcgggcca tgtccatcaa cgggcaagct gcagcgccca tcaccgccgc tttttgttac  93480
ccgtcgactc atcttccggt gcgggccagt gcagtctttt cctttttac gctcgctctc  93540
tctcttaaac gcttccaata tttgtgttta attattcgaa cggaatcctc tctgcgacag  93600
cacatccgta cggggtgcca gtagtgtgtg cgagtccgtg tttgtgtgta gccgtaatta  93660
tgttgtgatt gtcattgtca ctcgatgcgc gataaacaat ctacctacaa tttatgcacc  93720
cactgggcgg cctcgcctcg tgatccagtc cggtttgcaa gtcgccgcaa ctccaattca  93780
atgtcatccg ttctcacagc gaacgaacag aacggagggg acacgaacgc caacaacagc  93840
aacagcggca aaaaatgcac ccaaagtcct ggatgctggg gatgacaaga gccgccgatc  93900
cggcctccca ccacacacca aacgcacaat cgcagttgga attgcacggt ttaaatatat  93960
acatgttgtt gctgtttttt tgttttgttt ttggcgtgca actgtgctgc tcctgctcct  94020
atcgtgcgct atcgtggctg gatcccgcgg ggctactcgg tgcacggtct aacgcatccg  94080
gacgagcgtt tggtttggtt ccaatgttgc agttgcagtt ggagttcggg tcggggacaa  94140
aaaatcactt acttccactc gagcgccacc gcgccggaac gaacgcggaa acccgttcca  94200
cggtccatca tactctcttt cctccctccc caaccgtcgc tcagttcaac atatggccgt  94260
ggggatcggg attgggagct gtcaggtcca ggtgccgcgg gaagggatcc tgcagggaag  94320
tatcaagcgc cggaactgga agcacccgat gacagatggt gctcgaaagt gaactgtaaa  94380
actggacgcc catcaccaac aacatcacac cggcatgcag tgcgacaaaa aaaacacacc  94440
cacactgaga gagaaacaaa aatcacatcc acgcccgtcg tcatcagggg cgaaaaaaca  94500
acaaaccaca caaccggctg agccaacaga aactaacaca gcgcgcactg ggctggccac  94560
aaaatgtagt actaactaaa tccaatccaa ataattatat ttcaattgtt tatgaacggc  94620
attatgcgac cggaccggaa agtcgctggc tcgactcgtc cgtccagtcc cagcaacaat  94680
atcaacaata acacatgctc ccggcctgga acggtgggta tgcgtcggcg gcgtatgctg  94740
accaacataa tcaacgtatc ctttgtggtg ggattccggg attccggcag gatccgc     94797
```

```
SEQ ID NO: 2            moltype = DNA  length = 129
FEATURE                 Location/Qualifiers
source                  1..129
                        mol_type = other DNA
                        organism = Anopheles gambiae
SEQUENCE: 2
cctttccatt catttatgtt taacacaggt caagcggtgg tcaacgaata ctcacgattg  60
```

-continued

```
cataatctga acatgtttga tggcgtggag ttgcgcaata ccacccgtca gagtggatga  120
taaactttc                                                          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 gttttagagc tagaaatagc aagttaaaat aaggctagtc  60
cgttatcaac ttgaaaaagt ggcaccgagt 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
cagcgctggc ggtggggaca gctccggctg tggctgttct tgcgagtcct cttcctgcgg  60
cacatccctc tcgtcgacca gttcagtttg ctgagcgtaa gcctgctgct gttcgtcctg  120
catcatcggg accatttgta tgggccatcc gccaccacca ccatcaccac cgccgtccat  180
ttctaggggc atacccatca gcatctccgc gggcgccatt ggcggtggtg ccaaggtgcc  240
attcgtttgt tgctgaaagc aaaagaaagc aaattagtgt tgtttctgct gcacacgata  300
attttcgttt cttgccgcta gacacaaaca acactgcatc tggagggaga aatttgacgc  360
ctagctgtat aacttacctc aaagttattg tccatcgtgg tataatggac ctaccgagcc  420
cggttacact acacaaagca agattatgcg acaaaatcac agcgaaaact agtaattttc  480
atctatcgaa agcggccgag cagagagttg tttggtattg caacttgaca ttctgctgcg  540
ggataaaccg cgacgggcta ccatggcgca cctgtcagat ggctgtcaaa tttggcccgg  600
tttgcgatat ggagtgggtg aaattatatc ccactcgctg atcgtgaaaa tagacacctg  660
aaaacaataa ttgttgtgtt aattttacat tttgaagaac agcacaagtt ttgctgacaa  720
tatttaatta cgtttcgtta tcaacggcac ggaaagatta tctcgctgat tatccctctc  780
gctctctctg tctatcatgt cctggtcgtt ctcgcgtcac cccggataat cgagagacgc  840
catttttaat ttgaactact acaccgacaa gcatgccgtg agctctttca agttcttctg  900
tccgaccaaa gaaacagaga ataccgcccg gacagtgccc ggagtgatcg atccatagaa  960
aatcgcccat catgtgccac tgaggcgaac cggcgtagct tgttccgaat ttccaagtgc  1020
ttccccgtaa catccgcata taacaaacag cccaacaaca 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

```
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 8
gtgaacttcc atggaattac gtgctttttc ggaatggagt tgggctggtg aaaaacacct  60
atcagcaccg cacttttccc ccggcatttc aggttatacg cagagacaga gactaaatat  120
tcacccattc atcacgcact aacttcgcaa tagattgata ttccaaaact ttcttcacct  180
ttgccgagtt ggattctgga ttctgagact gtaaaaagtc gtacgagcta tcatagggtg  240
taaaacggaa aacaaacaaa cgtttaatgg actgctccaa ctgtaatcgc ttcacgcaaa  300
caaacacaca cgcgctggga gcgttcctgg cgtcaccttt gcacgatgaa aactgtagca  360
aaactcgcac gaccgaaggc tctccgtccc tgctggtgtg tgtttttttc ttttctgcag  420
caaaattaga aaacatcatc atttgacgaa aacgtcaact gcgcgagcag agtgaccaga  480
aataccgatg tatctgtata gtagaacgtc ggttatccgg gggcggatta accgtgcgca  540
caaccagttt tttgtgcagc tttgtagtgt ctagtggtat tttcgaaatt catttttgtt  600
cattaacagt tgttaaacct atagttattg attaaaataa tattctacta acgattaacc  660
gatggattca aagtgaataa attatgaaac tagtgatttt tttaaatttt tatatgaatt  720
tgacatttct tggaccatta tcatcttggt ctcgagctgc ccgaataatc gacgttctac  780
tgtattccta ccgatttttt atatgcctac cgacacacag gtgggccccc taaaactacc  840
gatttttaat ttatcctacc gaaaatcaca gattgtttca taatacagac caaaaagtca  900
tgtaaccatt tcccaaatca cttaatgtat taaactccat atggaaatcg ctagcaacca  960
gaaccagaag ttcaacagag acaaccaatt tccgtgtatg tacttcatga gatgagattg  1020
gacgcgctgg taaaatttta tatgggattt gacagataat gtaaggcgtg cgatttttt  1080
catacgatgg aatcaattca agagtcaatt gtgcaggatt tatagaaaca atctcttatt  1140
tatgttttgt tatcgttaca gttacagccc tgtcctaagc ggccgcgtga aggcccaaaa  1200
aaaagggagt ccccaacgct cagtagcaaa tgtgcttctc tatcattcgt tgggttagaa  1260
aagcctcatg tgacttctat gaacaaaatc taaactatct cctttaaata gagaatggat  1320
gtattttttc gtgccactga actttcgttg ggaagattag atacctctcc ctcccccccc  1380
ctccctttca acacttcaaa acctaccgaa aactaccgat acaatttgat gtacctaccg  1440
aagaccgcca aaataatctg gccacactgg ctagatctga tgttttgaaa catcgccaaa  1500
ttttactaaa taatgcactt gcgcgttggt gaagctgcac ttaaacagat tagttgaatt  1560
acgctttctg aaatgttttt attaaacact tgtttttttt aatacttcaa tttaaagcta  1620
cttcttggaa tgataattct acccaaaacc aaaaccactt tacaaagagt gtgtggttgg  1680
tgatcgcgcc ggctactgcg acctgtggtc atcgctcatc tcacgcacac atacgcacac  1740
atctgtcatt tgaaaagctg cacacaatcg tgtgttgtgc aaaaaaccgt tcgcgcacaa  1800
acagttcgca catgtttgca agccgtgcag caaagggctt ttgatggtga tccgcagtgt  1860
ttggtcagct ttttaatgtg ttttcgctta atcgcttttg tttgtgtaat gttttgtcgg  1920
aataattttt atgcgtcgtt acaaatgaaa tgtacaatcc tgcgatgcta gtgtaaaaca  1980
ttgctaattc ccggtaagaa cgttcattac gctcggatat catcttacga agcgtgtgta  2040
tgtgcgctag tacattgacc tttaaagtga tccttttgtt ctagaaagca ag          2092

SEQ ID NO: 9            moltype = DNA  length = 849
FEATURE                 Location/Qualifiers
misc_feature            1..849
                        note = exu promoter
source                  1..849
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 9
ggaaggtgat tgcgattcca tgttgatgcc aatatatgat gattttgttg catattaata  60
gttgttgtta tgttttattc aaatttcaaa gataatttac tttacattac agttagtgag  120
catattatct actacataaa cacatagatc aaactggttt acataaattc aaaaagtttg  180
gattaaaatc gcagcaattg gttatgaaaa aatatgtgca taacgtaaat atcaagtaaa  240
tttttgcatt gcatatttat agactcctgt tacaatttcg gaaaaatgaa aaatgttaat  300
taatcaaaga agaaaaaaca aagaaattaa atcattaggt agcacaacca caagtacata  360
tttttatggc atgaatattc ctctacacta acatatttta tagcaattct attgatcgcc  420
ttagtatagc ggaattacca gaacggcact atagttgtct ctgtttggca cacgcaatca  480
tttttcatcc cagggttgcc atagcagttt ggcgacggtc acgtagcatg cgaaggattt  540
cgttcgcaca ggatcacttt tattctaacg tttgaagaag gcacatctca gtgcaagcgc  600
tctggaagct gcttttaccg aacgaactaa cttttcaagt aacctcaaaa acttgtctct  660
aacgacacca cgtgctatcc gcgagtttca tttcccgtgc aaagttcccc gatttagcta  720
tcattcgtga acatttcgta gtgcctctac cctcaggtaa gaccattcga ggtttaccaa  780
gttttgtgca aagaacgtgc acagtaattt tcgttctggt gaaaccttct cttgtgtagc  840
ttgtacaaa                                                          849

SEQ ID NO: 10           moltype = DNA  length = 2291
FEATURE                 Location/Qualifiers
misc_feature            1..2291
                        note = Vasa2
source                  1..2291
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 10
atgtagaacg cgagcaaatt cttttccttc catgacagca gcagctacag tgggaagccg  60
aacgtcagac gtgtttgaca tgccgaactg ggcgggaaaa ttacagcgtg cgctttgttt  120
tcaagcaaat cacaactcgc tgcaaacaaa accgttgaga aattgattgt tttataattt  180
gtattgtatt ttatttgtta taataaacta aaaagacata ctttttgcat attttataca  240
taaaaacata catgcagcat tataaaacac atataaaccc tccctgtaga gtcccgtatc  300
```

-continued

```
gaaatcttcc atcctagttg cacagtacga cggacgagta ggccgtgtcc gtgcaaattc  360
cagcttttag cagtcttttg ctcggagcac tcgcggcgag tcggaggttt ctgctgaggt  420
gcttagcgct aaattagcca attgcttttg caagtgaaat aaccagccga atagtacttc  480
aaaactcagg taagtgaact agttttatag aacaaatgtt tgtttgttag aagttagtga  540
agtgtttgtg aaaaaaatct ctcatttcgg caaaactaac gtaactgatt tcaaattgaa  600
ttattgtttt gtgatgttat attatttcat ccagttgatt agtattttct tagttatgtt  660
caaaatacag ttaaattaaa tttcatttca tttactcata aaataatctc ttggcttatt  720
taatttttct cgaattcgct tgtattgttc agtagcacgc gccattcgcc ctttgtttca  780
ttttgtacct gctcccacta acacactggc agtgcgaaac aaaagccttc gcacgcgttg  840
ctggtattag agtgtgtgcg tgtgtgtgtt gagcgctctg tcaaaatcgg ctgttgccgc  900
cggtaccgaa attgcctgtt cgcacgctgt tcgtaaacat tccgtggtgt gtatcgtgtg  960
ttgtgcatgt tgcgcgcctc cccccttttg atagcaggct gccgtggctg ccgtggtgtg  1020
tggcgcagtt gagtttttgg attaattttc taaggaaatg gcacgagaag agcggtggca  1080
gtgtgttggt ttgctctgtc ccttcctttc tgtgtgaagt gttcttacag cacagcacgt  1140
atccaccacc gcacacagag caggcaagga agtggaagtg aacaagtgtg ctgcgcatgc  1200
atgtgtgtgg ggggcatttt agctgagatc gtcgttattt gagaagcggt ataggggcca  1260
gtcggtgtcg acgtacggaa gcggtttagt tttaatccaa gcgtatcccg tcgtggagtg  1320
gttgtgtggc tctgtgtgct ctcatatcag ttccagagtg aggttagtag aatcacagtc  1380
cttggccttt ttcgttacaa gatatccaga aggatggcgt tatttccaca gcttaccatg  1440
gtgctcttgt ttgctcgaat caggggagaa aaacagtttc gtgtttcatg aaccgcagtt  1500
ggcactggag cggattcaaa agtcttcgat atgcaataga taagagagtc gttggggcat  1560
agttgggaag cctttccgag atgtggagtt tccgagagga gaaatggtgc tttcgtgcac  1620
gttccgggac agcgggcccc gcgaagagca tctcgttgtc gttcatccgg caataattga  1680
tgcgaaaagc gcgcgcgcca ctggcttagc gcagtgtaca cagtgatatt cacctacaca  1740
cacagaggca cacgccttca cacgcgcgcg tgcttcaaag gctacttcgg tggcggtgtg  1800
tgaggtcgct tgcaatggac aatgaaaatt tcgctggaaa ataccatcgt ctctttaggt  1860
tgcaatgggt gcgggtagag cggtggtcgt cgatattggt ggtgtagtgt gtgtgtgtgt  1920
gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt  1980
gtgtgtgtgt gtgtgtgtgt gtgcaacggc aattattttt tgtaatattt cgaccatctt  2040
tctttctctc tctccacgtg ctgctgctgt tgctgctgct gctgcattgc atgttccact  2100
attcctctcg gtttgtgcct gcggacgcca ttgctagtcg aaagagagtc gccgttagtc  2160
gcgcttcgag caacggacac gtttttggt tgaaaccaac agctttttc atcttcggga  2220
gacacacaga tctcgaatcg tacattccca taaggagaat tgtcatcttc cggtgaataa  2280
agaaaggaaa c                                                       2291

SEQ ID NO: 11          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: 11
cttgtgttta gcaggcaggg gagatgagcg caaactgtgc aagaagaagc atcactgtga  60
agacggcaat gcaaagatag tgtgctcaac ttctccgcga agattgaagc taaattaagc  120
acgagattag catgactgaa gtgacttttc aaagtgtcag aatggctgca ctcgcaaact  180
agctggatgc agcgcaattt tgccccggtg tgtgcgcgca tgcaaacgag caaccgcaga  240
gggcaaagga gaggatggga aggagggagg gagtgaaaga gcaggcttaa ggttgccctc  300
gggcattgaa gtcgatacag cggttctatt ccagtgccag taacgatgac gaagacgatg  360
ttgcttctgc tgctgttgct gctgttgttg ttgatgatga tgatgataat agtgcaaata  420
taaaataaat cttccgtaag ctttgtgtag tggtgcgtgg ctactataag cccgtctgga  480
agcaaggaag ctagtcgggc agggtcatgc aaaagggaga caccttcgga gctccggagc  540
tcccgccggc actctcgggg ggacgtccgt tatgcgttgt gatttattat ggaatattta  600
ttatagtgtc ttgttttgaa aaaataactt caacggttcg aatttcctac acctcgagat  660
cggggctgga gtggcaacgt ggtacggaac ggtacagcgg tttgagccgt tcggtcttgg  720
gactcacgga tcgcagaatg ttattgtgcg cgcactgatg ggaaagtcat ttttcaccga  780
gtggtcaggg cgcgtagtcc agttcgtttc tggctgctgt tgctgatgct acgatcctca  840
ggaatgattg gaaacgcctg gagatggtgg gaaaaaatca aacacaaaaa cgatcctaat  900
gaacatcgtg tgttctcatt cgctgccacg attgacacct tcgataagac gcacataatg  960
agctaaagga gaggggacag ggtcttgtct ttgccacgag cgataagatt gcaatcactc  1020
gtgagcgtgt gctgctgggc tgaagaagaa acgctttcca cagcagtagg tgggaagtgg  1080
gattgtggaa cgtggcattg aaaagaacct attttctaaa gcccgagagc ccgttctcga  1140
actggaaaac cagatgcaga agtttttat tgtcccccgc caggaaaaca aatgtattta  1200
atgctttctt tgccttttcc gccccgtttc agacgacgag ctagtgaagc gagcccaatg  1260
gctgttggag aaactcggct acccgtggga gatgatgccc ctgatgtacg tcatactaaa  1320
gagcgccgat ggcgatgtac aaaaagcaca ccagcggatc gacgaaggta agctggcgat  1380
gatggtgtcg ttcgacatca ctttcatcac cgtgtcagac atctactgtg cctagcaccg  1440
ggtccagtgg tcacagggtg tagcaaaaac gtgttctttt ttgcgagaga ctctacctca  1500
tgatgcagct gttaaggaaa ggtttcagat gaaggcaatt tttcctagga taagatgatc  1560
ttaagttacc tgcgtattag tgtttaacat tgtcgtctca actcccaaga atgttttaat  1620
cgtctagggc tagtttattt atactgttct cattgaaatg tcgttcaatc caacatgtta  1680
agttagctag ctcagacacg agaagttagg agtatctgca tcttgaaggt agcggcatat  1740
ggtgttatgc cacgttcact gacttcaaaa ttcgatacaa aaaaaaaacc aaaacatcaa  1800
aaaccaaatt gtgaattccg tcagccagca gcagtgacct tcaaagcctt acctttccat  1860
tcatttatgt ttaacacagg tcaag                                        1885
```

-continued

```
SEQ ID NO: 12          moltype = DNA  length = 1961
FEATURE                Location/Qualifiers
misc_feature           1..1961
                       note = 3′ homology arm
source                 1..1961
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 12
cggtggtcaa cgaatactca cgattgcata atctgaacat gtttgatggc gtggagttgc  60
gcaataccac ccgtcagagt ggatgataaa ctttccgcac cactgtaact gtccgtatct  120
ttgtatgtgg gtgtgtgtat gtgtgtttgg tgaaacgaat tcaatagttc tgtgctattt  180
taaatcaagc cgcgtgcgca actgatgccg ataagttcaa actagtgttt aaggagtgga  240
gcgagagagc cgcaccacgg tacagaaggg cagcagaatg ggtcggcagc ctagctgcac  300
tggtgcggtg cgtccggcgt ctcgggggga gggcgaggaa attctagtgt taaatcggag  360
cagcaaaaac aaaacagtgg tcgtcccgtt caagaaacgg cctgtacaca cacacagaaa  420
acactgcagc atgtttgtac atagtagatc ctagagcagg tggtcgttgc tcctcgaacg  480
ctctggacgc acggcttcgc gcgtatttgc gtagcgttcc gccgatcgtg ggtattcgta  540
ctgccacaag cccgctttct cccatgcaat ctctgcaacc aaaccaacaa acaacaacaa  600
aaaaccaatc gacaaaatga atcacacccc ttttgtatca tctgtatatt cttgttcttt  660
gcgttctttt ctatgtggcc cacgccccgg cgggtacgta attgcgtcga aaaccccgaa  720
aaccccggca catacagtgt acatacggtt tgaggacaac tttgacctgc agcccttctg  780
gggttgccac gtgtagctat acttgtgaga tcgggcgccg acggtgtaaa gcgcgaatgg  840
ccgccacaca gtgtgtccac tccaacacta cccctctgga actaccccgt ccagggatgc  900
accggctcgg ctcatgcccc tgcaaaacag tccgggctcc actgtagtag ctccggcgtt  960
gctctgagag aaggatgccc ttcgaagtgt cgaaagcgtg cattgggcgt tcaagtgtgt  1020
gtgtgtgtgt taggtttagc gagaaacagc agcagttgcg tgtgctgaaa agcgaaggag  1080
taatagagtg cataatgaaa atgaaaatga aaatgaagca aaagtagaag gcggaggaga  1140
gcaacctgtg ttccactagt agcgaatagt ttagtctagt ttcgtcacca atcaaccttc  1200
caaccatcgt tcaaccaata cctgagtcaa catcgtcatc gttatcgtgc cacaacttta  1260
ttaaaaatga accttgtccg cgccaccgta gggtgatcta aggcgacctt tcttacgggc  1320
gcgacccaca tgccatcgtc accttctcca atcaaaacca acagcctgta ccgatggtgt  1380
gcaattgtgc gtgcgtgtgt gttattagca aaaaaagaga aagagtcgac gagagagaga  1440
tagatcgaga tcgagagtac aaaagagcag tagaaatgtt cgttgtttgt ttttcgtaac  1500
acagttgttt agccaaaatg ggaatttcca ataatcccgg gggcggggaa atgcgggaat  1560
actgcgtaca cacatacatc aatcaaaaag aaaaatcctt gcgctacatc actaccgttt  1620
gcgcggtgct gatctagagc agaccacttt ccactccact ctacaatcaa tcaatctgtg  1680
cagaaggtat ggtaagacgg cctttgagcg agtcacggtc gccaccataa cgccgtccga  1740
cgagggctga atgcgaactt tgctaatcga ttttccgctt tctttttatc ccacctcctt  1800
ttctctccct ctctctcttt tgcactgccc cttgtaaccc ccaaaaaggt aaacgacaca  1860
ttaagaccta cgaagcgttg gtgaagtcat cgctcgatcc gaacagcgac cggctgacgg  1920
aggacgacga cgaggacgag aacatctcgg tgacccgcac c                      1961

SEQ ID NO: 13          moltype = DNA  length = 8005
FEATURE                Location/Qualifiers
misc_feature           1..8005
                       note = Gene Drive construct
source                 1..8005
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 13
tgcgggtgcc agggcgtgcc cttgggctcc ccgggcgcgt actccacctc acccatgcga  60
tcgctccgga aagatacatt gatgagtttg gacaaaccac aactagaatg cagtgaaaaa  120
aatgctttat ttgtgaaatt tgtgatgcta ttgctttatt tgtaaccatt ataagctgca  180
ataaacaagt taacaacaac aattgcattc attttatgtt tcaggttcag ggggaggtgt  240
gggaggtttt ttaaagcaag taaaacctct acaaatgtgg tatggctgat tatgatctag  300
agtcgcggcc gctacaggaa caggtggtgg cggccctcgg tgcgctcgta ctgctccacg  360
atggtgtagt cctcgttgtg ggaggtgatg tccagcttgg agtccacgta gtagtagccg  420
ggcagctgca cgggcttctt ggccatgtag atggacttga actccaccag gtagtggccg  480
ccgtccttca gcttcagggc cttgtggatc tcgcccttca gcacgccgtc gcgggggtac  540
aggcgctcgg tggaggcctc ccagcccatg gtcttcttct gcattacggg gccgtcggag  600
gggaagttca cgccgatgaa cttcaccttg tagatgaagc agccgtcctg cagggaggag  660
tcttgggtca cggtcaccac gccgccgtcc tcgaagttca tcacgcgctc ccacttgaag  720
ccctcgggga aggacagctt cttgtagtcg gggatgtcgg cggggtgctt cacgtacacc  780
ttggagccgt actggaactg gggggacagg atgtcccagg cgaagggcag ggggccgccc  840
ttggtcacct tcagcttcac ggtgttgtgg ccctcgtagg ggcggccctc gccctcgccc  900
tcgatctcga actcgtggcc gttcacggtg ccctccatgc gcaccttgaa gcgcatgaac  960
tccttgatga cgttcttgga ggagcgcacc atggtggcga cctgtgggtc ccgggcccgc  1020
ggtaccgtcg actctagcgg taccccgatt gtttagcttg ttcagctgcg cttgtttatt  1080
tgcttagctt tcgcttagcg acgtgttcac tttgcttgtt tgaattgaat tgtcgctccg  1140
tagacgaagc gcctctattt atactccggc ggtcgagggt tcgaaatcga taagcttgga  1200
tcctaattga attagctcta attgaattag tctctaattg aattagatcc ccgggcgagc  1260
tcgaattaac cattgtggac cggtcagcgc tggcggtggg gacagctccg gctgtggctg  1320
ttcttgagag tcatcttcct gcggcacatc cctctcgtcg accagttcag tttgctgagc  1380
gtaagcctgc tgctgttcgt cctgcatcat cgggaccatt tgtacgggcc atccgccacc  1440
accaccatca ccaccgccgt ccatttctag gggcataccc atcagcatct ccgcgggcgc  1500
cattggcggt ggtgccaagg tgccattcgt ttgttgctga aagcaaaaga aagcaaatta  1560
```

-continued

```
gtgttgtttc tgctgcacac gatagttttc gtttcttgcc gctagacaca aacaacactg  1620
catctggagg gagaaatttg acgcctagct gtataactta cctcaaagtt attgtccatc  1680
gtggtataat ggacctaccg agcccggtta cactacacaa agcaagatta tgcgacaaaa  1740
tcacagcgaa aactagtaat tttcatctat cgaaagcggc cgagcagaga gttgtttggt  1800
attgcaactt gacattctgc tgtgggataa accgcgacgg gctaccatgg cgcacctgtc  1860
agatggctgt caaatttggc ccggtttgcg atatggagtg ggtgaaatta tatcccactc  1920
gctgatcgtg aaaatagaca cctgaaaaca ataattgttg tgttaatttt acattttgaa  1980
gaacagcaca agttttgctg acaatattta attacgtttc gttatcaacg gcacggaaag  2040
attatctcgc tgattatccc tctcgctctc tctgtctatc atgtcctggt cgttctcgcg  2100
tcaccccgga taatcgagag acgccatttt taatttgaac tactacaccg acaagcatgc  2160
cgtgagctct ttcaagttct tctgtccgac caaagaaaca gagaataccg cccggacagt  2220
gcccggagtg atcgatccat agaaaatcgc ccatcatgtg ccactgaagc gaaccggcgt  2280
agcttgttcc gaatttccaa gtgcttcccc gtaacatccg catataacaa gcagcccaac  2340
aacaaataca gcatcgagct cgagatggac tataaggacc acgacggaga ctacaaggat  2400
catgatattg attacaaaga cgatgacgat aagatggccc caaagaagaa gcggaaggtc  2460
ggtatccacg gagtcccagc agccgacaag aagtacagca tcggcctgga catcggcacc  2520
aactctgtgg gctgggccgt gatcaccgac gagtacaagg tgcccagcaa gaaattcaag  2580
gtgctgggca acaccgaccg gcacagcatc aagaagaacc tgatcggagc cctgctgttc  2640
gacagcggcg aaacagccga ggccacccgg ctgaagagaa ccgccagaag aagatacacc  2700
agacggaaga accggatctg ctatctgcaa gagatcttca gcaacgagat ggccaaggtg  2760
gacgacagct tcttccacag actggaagag tccttcctgg tggaagagga taagaagcac  2820
gagcggcacc ccatcttcgg caacatcgtg gacgaggtgg cctaccacga gaagtacccc  2880
accatctacc acctgagaaa gaaactggtg gacagcaccg acaaggccga cctgcggctg  2940
atctatctgg ccctggccca catgatcaag ttccggggcc acttcctgat cgagggcgac  3000
ctgaaccccg acaacagcga cgtggacaag ctgttcatcc agctggtgca gacctacaac  3060
cagctgttcg aggaaaaccc catcaacgcc agcggcgtgg acgccaaggc catcctgtct  3120
gccagactga gcaagagcag acggctggaa aatctgatcg cccagctgcc cggcgagaag  3180
aagaatggcc tgttcggaaa cctgattgcc ctgagcctgg gcctgacccc caacttcaag  3240
agcaacttcg acctggccga ggatgccaaa ctgcagctga gcaaggacac ctacgacgac  3300
gacctggaca acctgctggc ccagatcggc gaccagtacg ccgacctgtt tctggccgcc  3360
aagaacctgt ccgacgccat cctgctgagc gacatcctga gagtgaacac cgagatcacc  3420
aaggcccccc tgagcgcctc tatgatcaag agatacgacg agcaccacca ggacctgacc  3480
ctgctgaaag ctctcgtgcg gcagcagctg cctgagaagt acaaagagat tttcttcgac  3540
cagagcaaga acggctacgc cggctacatt gacggcggag ccagccagga agagttctac  3600
aagttcatca agcccatcct ggaaaagatg gacggcaccg aggaactgct cgtgaagctg  3660
aacagagagg acctgctgcg gaagcagcgg accttcgaca acggcagcat cccccaccag  3720
atccacctgg gagagctgca cgccattctg cggcggcagg aagattttta cccattcctg  3780
aaggacaacc gggaaaagat cgagaagatc ctgaccttcc gcatccccta ctacgtgggc  3840
cctctggcca ggggaaacag cagattcgcc tggatgacca gaaagagcga ggaaaccatc  3900
acccccctgga acttcgagga agtggtggac aagggcgctt ccgcccagag cttcatcgag  3960
cggatgacca acttcgataa gaacctgccc aacgagaagg tgctgcccaa gcacagcctg  4020
ctgtacgagt acttcaccgt gtataacgag ctgaccaaag tgaaatacgt gaccgaggga  4080
atgagaaagc ccgccttcct gagcggcgag cagaaaaagg ccatcgtgga cctgctgttc  4140
aagaccaacc ggaaagtgac cgtgaagcag ctgaaagagg actacttcaa gaaaatcgag  4200
tgcttcgact ccgtggaaat ctccggcgtg gaagatcggt tcaacgcctc cctgggcaca  4260
taccacgatc tgctgaaaat tatcaaggac aaggacttcc tggacaatga ggaaaacgag  4320
gacattctgg aagatatcgt gctgaccctg acactgtttg aggacagaga gatgatcgag  4380
gaacggctga aaacctatgc ccacctgttc gacgacaaag tgatgaagca gctgaagcgg  4440
cggagataca ccggctgggg caggctgagc cggaagctga tcaacggcat ccgggacaag  4500
cagtccggca agacaatcct ggatttcctg aagtccgacg gcttcgccaa cagaaacttc  4560
atgcagctga tccacgacga cagcctgacc tttaaagagg acatccagaa agcccaggtg  4620
tccggccagg gcgatagcct gcacgagcac attgccaatc tggccggcag ccccgccatt  4680
aagaagggca tcctgcagac agtgaaggtg gtggacgagc tcgtgaaagt gatgggccgg  4740
cacaagcccg agaacatcgt gatcgaaatg gccagagaga accagaccac ccagaaggga  4800
cagaagaaca gccgcgagag aatgaagcgg atcgaagagg gcatcaaaga gctgggcagc  4860
cagatcctga aagaacaccc cgtggaaaac acccagctgc agaacgagaa gctgtacctg  4920
tactacctgc agaatgggcg ggatatgtac gtggaccagg aactggacat caaccggctg  4980
tccgactacg atgtggacca tatcgtgcct cagagctttc tgaaggacga ctccatcgac  5040
aacaaggtgc tgaccagaag cgacaagaac cggggcaaga gcgacaacgt gccctccgaa  5100
gaggtcgtga agaagatgaa gaactactgg cggcagctgc tgaacgccaa gctgattacc  5160
cagagaaagt tcgacaatct gaccaaggcc gagagaggcg gcctgagcga actggataag  5220
gccggcttca tcaagagaca gctggtggaa acccggcaga tcacaaagca cgtggcacag  5280
atcctggact cccggatgaa cactaagtac gacgagaatg acaagctgat ccgggaagtg  5340
aaagtgatca ccctgaagtc caagctggtg tccgatttcc ggaaggattt ccagttttac  5400
aaagtgcgcg agatcaacaa ctaccaccac gcccacgacg cctacctgaa cgccgtcgtg  5460
ggaaccgccc tgatcaaaaa gtaccctaag ctggaaagcg agttcgtgta cggcgactac  5520
aaggtgtacg acgtgcggaa gatgatcgcc aagagcgagc aggaaatcgg caaggctacc  5580
gccaagtact tcttctacag caacatcatg aactttttca agaccgagat taccctggcc  5640
aacggcgaga tccggaagcg gcctctgatc gagacaaacg gcgaaaccgg ggagatcgtg  5700
tgggataagg gccgggattt tgccaccgtg cggaaagtgc tgagcatgcc ccaagtgaat  5760
atcgtgaaaa agaccgaggt gcagacaggc ggcttcagca aagagtctat cctgcccaag  5820
aggaacagcg ataagctgat cgccagaaag aaggactggg accctaagaa gtacggcggc  5880
ttcgacagcc ccaccgtggc ctattctgtg ctggtggtgg ccaaagtgga aaagggcaag  5940
tccaagaaac tgaagagtgt gaaagagctg ctggggatca ccatcatgga aagaagcagc  6000
ttcgagaaga atcccatcga ctttctggaa gccaagggct acaaagaagt gaaaaaggac  6060
ctgatcatca agctgcctaa gtactccctg ttcgagctgg aaaacggccg gaagagaatg  6120
```

-continued

```
ctggcctctg ccggcgaact gcagaaggga aacgaactgg ccctgccctc caaatatgtg  6180
aacttcctgt acctggccag ccactatgag aagctgaagg gctcccccga ggataatgag  6240
cagaaacagc tgtttgtgga acagcacaag cactacctgg acgagatcat cgagcagatc  6300
agcgagttct ccaagagagt gatcctggcc gacgctaatc tggacaaagt gctgtccgcc  6360
tacaacaagc accgggataa gcccatcaga gagcaggccg agaatatcat ccacctgttt  6420
accctgacca atctgggagc ccctgccgcc ttcaagtact ttgacaccac catcgaccgg  6480
aagaggtaca ccagcaccaa agaggtgctg gacgccaccc tgatccacca gagcatcacc  6540
ggcctgtacg agacacggat cgacctgtct cagctgggag gcgacaaaag gccggcggcc  6600
acgaaaaagg ccggccaggc aaaaaagaaa aagtaattaa ttaagaggac ggcgagaagt  6660
aatcatatgt ccgcattttg cgcaaaccag gcgcttagac aatttgcgcg taagcacatt  6720
cgaaatgtga aaagctgaaa gcagtggttt cgccagcccg agttcagcga aacggattcc  6780
ttccaagtgt ttgcattcct ggcggagtgt tcctcccaaa atgcactcac cctgcgtgca  6840
gtgccaaatc gtgagtttcc taattttttc atattgttta ttacctacca actaaagttg  6900
ttgttatata ttgcgtttta cgtacgacaa ataagttcgt attcagaaat atttgcgata  6960
agagagaact catttgcgat gaatctcatt gtatttagct aagtgccttg ataagtaagc  7020
ggaacagcag gaatatgaca ctccttggga aatacatgta agcgtctgta attagatata  7080
tatacacgca accaaatggt ccatggttga tttaagcact gcctgttgtc gaacattgct  7140
ataagcaaaa taaagaagca ttcattaatc taaaatttct tcaaagtgac ttcaatgatg  7200
atctctaggc tatagtgaaa gctgaaagct tatttgacaa tgcaagggaa agtgacgcac  7260
gtgcgtcgta tgggaccgcg cgcatctatt ctctcagcta attcccctaa tcattagtaa  7320
ttgacggcac gatttctgct tcttacttcc ttttactttg gagcttttca tcaataaaac  7380
cagtaccatg gccgtacgct caacggaaaa gcattcaaaa aaacccgcgt tcctcgtgtg  7440
atttgtgggt gagtggcgcc atctattaga gaatagctgt actacatctc gtggacgaag  7500
gggtcagaga agttgaaaga gagcttgatc gactgctatc caagctaggc gaggaaggga  7560
gatcgctaga gcaaaagaaa aaaaataagc aaatatcttt ttttataaca aatcgacgtt  7620
agcgaaatat gtttgaatcg atttaacggt tagaattccc tttggttcgt tcattatgcg  7680
aggcgcgcct ttgtatgcgt gcgcttgaag ggttgatcgg aaccttacaa cagttgtagc  7740
tatacggctg cgtgtggctt ctaacgttat ccatcgctag aagtgaaacg aatgtgcgta  7800
ggtatatata tgaaatggag ttgctctctg ctgtttaaca caggtcaagc gggttttaga  7860
gctagaaata gcaagttaaa ataaggctag tccgttatca acttgaaaaa gtggcaccga  7920
gtcggtgctt ttttttacgc gtgggtccca tgggtgaggt ggagtacgcg cccggggagc  7980
ccaagggcac gccctggcac ccgca                                        8005
```

```
SEQ ID NO: 14          moltype = DNA  length = 24
FEATURE                Location/Qualifiers
misc_feature           1..24
                       note = dsxgRNA-F primer
source                 1..24
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 14
tgctgtttaa cacaggtcaa 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
SEQUENCE: 15
aaacccgctt gacctgtgtt aaac                                         24

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

SEQ ID NO: 17          moltype = DNA  length = 49
FEATURE                Location/Qualifiers
misc_feature           1..49
                       note = dsx?31L-R primer
source                 1..49
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 17
tccacctcac ccatgggacc cacgcgtggt gcgggtcacc gagatgttc              49

SEQ ID NO: 18          moltype = DNA  length = 50
FEATURE                Location/Qualifiers
misc_feature           1..50
```

-continued

```
                        note = dsx?31R-F primer
source                  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
ggtggtagtg ccacacagag agcttcgcgg tggtcaacga atactcacg             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 accggtcagc gctggcggtg ggga                  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
tcgtggtcct 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
aggcaaaaaa 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 cgcatacaaa ggcgcgcctc 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
ctttttgtac 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 aagaccccaa                                                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
gcacaccagc 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 aagatacgga 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 tcaaggttca 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
tatactccgg cggtcgaggg 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 gatcctggcc 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 cacgttcact ga                                          22

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

SEQ ID NO: 35           moltype = DNA  length = 1712
FEATURE                 Location/Qualifiers
misc_feature            1..1712
                        note = nucleotide sequence of exon 5 of the doublesex (dsx)
                         gene
source                  1..1712
                        mol_type = other DNA
                        organism = unidentified
SEQUENCE: 35
gtcaagcggt ggtcaacgaa tactcacgat tgcataatct gaacatgttt gatggcgtgg  60
agttgcgcaa taccacccgt cagagtggat gataaacttt ccgcaccact gtaactgtcc  120
gtatctttgt atgtgggtgt gtgtatgtgt gtttggtgaa acgaattcaa tagttctgtg  180
ctattttaaa tcaagccgcg tgcgcaactg atgccgataa gttcaaacta gtgtttaagg  240
agtggagcga gagagccgca ccacggtaca gaagggcagc agaatgggtc ggcagcctag  300
ctgcactggt gcggtgcgtc cggcgtctcg gggggagggc gaggaaattc tagtgttaaa  360
tcggagcagc aaaaacaaaa cagtggtcgt cccgttcaag aaacggcctg tacacacaca  420
cagaaaacac tgcagcatgt ttgtacatag tagatcctag agcaggtggt cgttgctcct  480
cgaacgctct ggacgcacgg cttcgcgcgt atttgcgtag cgttccgccg atcgtgggta  540
ttcgtactgc cacaagcccg ctttctccca tgcaatctct gcaaccaaac caacaaacaa  600
caacaaaaaa ccaatcgaca aaatgaatca cacccctttt gtatcatctg tatattcttg  660
ttctttgcgt tcttttctat gtggcccacg cccggcggg tacgtaattg cgtcgaaaac  720
cccgaaaacc ccggcacata cagtgtacat acggtttgag gacaactttg acctgcagcc  780
cttctggggt tgccacgtgt agctatactt gtgagatcgg gcgccgacgg tgtaaagcgc  840
gaatggccgc cacacagtgt gtccactcca acactacccc tctggaacta ccccgtccag  900
ggatgcaccg gctcggctca tgcccctgca aaacagtccg ggctccactg tagtagctcc  960
ggcgttgctc tgagagaagg atgcccttcg aagtgtcgaa agcgtgcatt gggcgttcaa  1020
gtgtgtgtgt gtgtgttagg tttagcgaga aacagcagca gttgcgtgtg ctgaaaagcg  1080
aaggagtaat agagtgcata atgaaaatga aaatgaaaat gaagcaaaag tagaaggcgg  1140
aggagagcaa cctgtgttcc actagtagcg aatagtttag tctagtttcg tcaccaatca  1200
accttccaac catcgttcaa ccaatacctg agtcaacatc gtcatcgtta tcgtgccaca  1260
actttattaa aaatgaacct tgtccgcgcc accgtagggt gatctaaggc gacctttctt  1320
acgggcgcga cccacatgcc atcgtcacct tctccaatca aaaccaacag cctgtaccga  1380
tggtgtgcaa ttgtgcgtgc gtgtgtgtta ttagcaaaaa aagagaaaga gtcgacgaga  1440
gagagataga tcgagatcga gagtacaaaa gagcagtaga aatgttcgtt gtttgttttt  1500
cgtaacacag ttgtttagcc aaaatgggaa tttccaataa tcccgggggc ggggaaatgc  1560
gggaatactg cgtacacaca tacatcaatc aaaaagaaaa atccttgcgc tacatcacta  1620
ccgtttgcgc ggtgctgatc tagagcagac cactttccac tccactctac aatcaatcaa  1680
tctgtgcaga aggtatggta agacggcctt tg                               1712

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

SEQ ID NO: 37           moltype = DNA  length = 22
FEATURE                 Location/Qualifiers
misc_feature            1..22
                        note = T3 target site
source                  1..22
                        mol_type = other DNA
                        organism = unidentified
SEQUENCE: 37
gcaataccac 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                 Location/Qualifiers
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                 Location/Qualifiers
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                 Location/Qualifiers
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                 Location/Qualifiers
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 gttttagagc tagaaatagc aagttaaaat aaggctagtc  60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgct                           97

SEQ ID NO: 43           moltype = DNA  length = 96
FEATURE                 Location/Qualifiers
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 ttttagagct agaaatagca agttaaaata aggctagtcc  60
gttatcaact tgaaaaagtg gcaccgagtc ggtgct                            96

SEQ ID NO: 44           moltype = DNA  length = 95
FEATURE                 Location/Qualifiers
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 ggctagtccg  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
tctgaacatg tttgatggcg gttttagagc tagaaatagc aagttaaaat 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 aggctagtcc  60
gttatcaact tgaaaaagtg gcaccgagtc 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 tttagagcta gaaatagcaa gttaaaataa ggctagtccg  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 ggtcaagcgg gttttagagc tagaaatagc aagttaaaat 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
tttgtatgcg tgcgcttgaa gggttgatcg gaaccttaca acagttgtag ctatacggct  60
gcgtgtggct tctaacgtta tccatcgcta gaagtgaaac gaatgtgcgt aggtatatat  120
atgaaatgga gttgctctct gct                                          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 ttggtgaaac gaattcaata gttctgtgct attttaaatc aagccgcgtg  120
cgcaactgat gccgataagt tcaaactagt gtttaaggag tggagcgaga gagccgcacc  180
acggtacaga agggcagcag aatgggtcgg cagcctagct gcactggtgc ggtgcgtccg  240
gcgtctcggg gggagggcga ggaaattcta gtgttaaatc ggagcagcaa aaacaaaaca  300
gtggtcgtcc cgttcaagaa acggcctgta cacacacaca gaaaacactg cagcatgttt  360
gtacatagta gatcctagag caggtggtcg ttgctcctcg aacgctctgg acgcacggct  420
tcgcgcgtat ttgcgtagcg ttccgccgat cgtgggtatt cgtactgcca caagcccgct  480
ttctcccatg caatctctgc aaccaaacca acaaacaaca acaaaaaacc aatcgacaaa  540
atgaatcaca cccctttgt atcatctgta tattcttgtt ctttgcgttc ttttctatgt  600
```

-continued

```
ggcccacgcc ccggcgggta cgtaattgcg tcgaaaaccc cgaaaacccc ggcacataca  660
gtgtacatac ggtttgagga caactttgac ctgcagccct tctggggttg ccacgtgtag  720
ctatacttgt gagatcgggc gccgacggtg taaagcgcga atggccgcca cacagtgtgt  780
ccactccaac actacccctc tggaactacc ccgtccaggg atgcaccggc tcggctcatg  840
cccctgcaaa acagtccggg ctccactgta gtagctccgg cgttgctctg agagaaggat  900
gcccttcgaa gtgtcgaaag cgtgcattgg gcgttcaagt gtgtgtgtgt gtgttaggtt  960
tagcgagaaa cagcagcagt tgcgtgtgct gaaaagcgaa ggagtaatag agtgcataat 1020
gaaaatgaaa atgaaaatga agcaaaagta gaaggcggag gagagcaacc tgtgttccac 1080
tagtagcgaa tagtttagtc tagtttcgtc accaatcaac cttccaacca tcgttcaacc 1140
aatacctgag tcaacatcgt catcgttatc gtgccacaac tttattaaaa atgaaccttg 1200
tccgcgccac cgtagggtga tctaaggcga cctttcttac gggcgcgacc cacatgccat 1260
cgtcaccttc tccaatcaaa accaacagcc tgtaccgatg gtgtgcaatt gtgcgtgcgt 1320
gtgtgttatt agcaaaaaaa gagaaagagt cgacgagaga gagatagatc gagatcgaga 1380
gtacaaaaga gcagtagaaa tgttcgttgt ttgtttttcg taacacagtt gtttagccaa 1440
aatgggaatt tccaataatc ccgggggcgg ggaaatgcgg gaatactgcg tacacacata 1500
catcaatcaa aaagaaaaat ccttgcgcta catcactacc gtttgcgcgg tgctgatcta 1560
gagcagacca ctttccactc cactctacaa tcaatcaatc tgtgcagaag gtatggtaag 1620
acggcctttg agcgagtcac ggtcgccacc ataacgccgt ccgacgaggg ctgaatgcga 1680
actttgctaa tcgattttcc gctttctttt tatcccacct ccttttctct ccctctctct 1740
cttttgcact gccccttgta acccccaaaa aggtaaacga cacattaaga cctacgaagc 1800
gttggtgaag tcatcgctcg atccgaacag cgaccggctg acggaggacg acgacgagga 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
tgcgggtgcc agggcgtgcc cttgggctcc ccgggcgcgt actccacctc acccatgcga   60
tcgctccgga aagatacatt gatgagtttg gacaaaccac aactagaatg cagtgaaaaa  120
aatgctttat ttgtgaaatt tgtgatgcta ttgctttatt tgtaaccatt ataagctgca  180
ataaacaagt taacaacaac aattgcattc attttatgtt tcaggttcag ggggaggtgt  240
gggaggtttt ttaaagcaag taaaacctct acaaatgtgg tatggctgat tatgatctag  300
agtcgcggcc gctacaggaa caggtggtgg cggccctcgg tgcgctcgta ctgctccacg  360
atggtgtagt cctcgttgtg ggaggtgatg tccagcttgg agtccacgta gtagtagccg  420
ggcagctgca cgggcttctt ggccatgtag atggacttga actccaccag gtagtggccg  480
ccgtccttca gcttcagggc cttgtggatc tcgcccttca gcacgccgtc gcgggggtac  540
aggcgctcgg tggaggcctc ccagcccatg gtcttcttct gcattacggg gccgtcggag  600
gggaagttca cgccgatgaa cttcaccttg tagatgaagc agccgtcctg cagggaggag  660
tcttgggtca cggtcaccac gccgccgtcc tcgaagttca tcacgcgctc ccacttgaag  720
ccctcgggga aggacagctt cttgtagtcg gggatgtcgg cggggtgctt cacgtacacc  780
ttggagccgt actggaactg gggggacagg atgtcccagg cgaagggcag ggggccgccc  840
ttggtcacct tcagcttcac ggtgttgtgg ccctcgtagg ggcggccctc gccctcgccc  900
tcgatctcga actcgtggcc gttcacggtg ccctccatgc gcaccttgaa gcgcatgaac  960
tccttgatga cgttcttgga ggagcgcacc atggtggcga cctgtgggtc ccgggcccgc 1020
ggtaccgtcg actctagcgg taccccgatt gtttagcttg ttcagctgcg cttgtttatt 1080
tgcttagctt tcgcttagcg acgtgttcac tttgcttgtt tgaattgaat tgtcgctccg 1140
tagacgaagc gcctctattt atactccggc ggtcgagggt tcgaaatcga taagcttgga 1200
tcctaattga attagctcta attgaattag tctctaattg aattagatcc ccgggcgagc 1260
tcgaattaac cattgtggac cggtcagcgc tggcggtggg gacagctccg gctgtggctg 1320
ttcttgagag tcatcttcct gcggcacatc cctctcgtcg accagttcag tttgctgagc 1380
gtaagcctgc tgctgttcgt cctgcatcat cgggaccatt tgtacgggcc atccgccacc 1440
accaccatca ccaccgccgt ccatttctag gggcataccc atcagcatct ccgcgggcgc 1500
cattggcggt ggtgccaagg tgccattcgt ttgttgctga aagcaaaaga aagcaaatta 1560
gtgttgtttc tgctgcacac gatagttttc gtttcttgcc gctagacaca aacaacactg 1620
catctggagg gagaaatttg acgcctagct gtataactta cctcaaagtt attgtccatc 1680
gtggtataat ggacctaccg agcccggtta cactacacaa agcaagatta tgcgacaaaa 1740
tcacagcgaa aactagtaat tttcatctat cgaaagcggc cgagcagaga gttgtttggt 1800
attgcaactt gacattctgc tgtgggataa accgcgacgg gctaccatgg cgcacctgtc 1860
agatggctgt caaatttggc ccggtttgcg atatggagtg ggtgaaatta tatcccactc 1920
gctgatcgtg aaaatagaca cctgaaaaca ataattgttg tgttaatttt acattttgaa 1980
gaacagcaca agttttgctg acaatattta attacgtttc gttatcaacg gcacggaaag 2040
attatctcgc tgattatccc tctcgctctc tctgtctatc atgtcctggt cgttctcgcg 2100
tcaccccgga taatcgagag acgccatttt taatttgaac tactacaccg acaagcatgc 2160
cgtgagctct ttcaagttct tctgtccgac caaagaaaca gagaataccg cccggacagt 2220
gcccggagtg atcgatccat agaaaatcgc ccatcatgtg ccactgaagc gaaccggcgt 2280
agcttgttcc gaatttccaa gtgcttcccc gtaacatccg catataacaa gcagcccaac 2340
aacaaataca gcatcgagct cgagatggac tataaggacc acgacggaga ctacaaggat 2400
catgatattg attacaaaga cgatgacgat aagatggccc caaagaagaa gcggaaggtc 2460
ggtatccacg gagtcccagc agccgacaag aagtacagca tcggcctgga catcggcacc 2520
aactctgtgg gctgggccgt gatcaccgac gagtacaagg tgcccagcaa gaaattcaag 2580
gtgctgggca acaccgaccg gcacagcatc aagaagaacc tgatcggagc cctgctgttc 2640
gacagcggcg aaacagccga ggccacccgg ctgaagagaa ccgccagaag aagatacacc 2700
```

-continued

```
agacggaaga accggatctg ctatctgcaa gagatcttca gcaacgagat ggccaaggtg  2760
gacgacagct tcttccacag actggaagag tccttcctgg tggaagagga taagaagcac  2820
gagcggcacc ccatcttcgg caacatcgtg gacgaggtgg cctaccacga gaagtacccc  2880
accatctacc acctgagaaa gaaactggtg gacagcaccg acaaggccga cctgcggctg  2940
atctatctgg ccctggccca catgatcaag ttccggggcc acttcctgat cgagggcgac  3000
ctgaaccccg acaacagcga cgtggacaag ctgttcatcc agctggtgca gacctacaac  3060
cagctgttcg aggaaaaccc catcaacgcc agcggcgtgg acgccaaggc catcctgtct  3120
gccagactga gcaagagcag acggctggaa aatctgatcg cccagctgcc cggcgagaag  3180
aagaatggcc tgttcggaaa cctgattgcc ctgagcctgg gcctgacccc caacttcaag  3240
agcaacttcg acctggccga ggatgccaaa ctgcagctga gcaaggacac ctacgacgac  3300
gacctggaca acctgctggc ccagatcggc gaccagtacg ccgacctgtt tctggccgcc  3360
aagaacctgt ccgacgccat cctgctgagc gacatcctga gagtgaacac cgagatcacc  3420
aaggcccccc tgagcgcctc tatgatcaag agatacgacg agcaccacca ggacctgacc  3480
ctgctgaaag ctctcgtgcg gcagcagctg cctgagaagt acaaagagat tttcttcgac  3540
cagagcaaga acggctacgc cggctacatt gacggcggag ccagccagga agagttctac  3600
aagttcatca agcccatcct ggaaaagatg gacggcaccg aggaactgct cgtgaagctg  3660
aacagagagg acctgctgcg gaagcagcgg accttcgaca acggcagcat cccccaccag  3720
atccacctgg gagagctgca cgccattctg cggcggcagg aagattttta cccattcctg  3780
aaggacaacc gggaaaagat cgagaagatc ctgaccttcc gcatccccta ctacgtgggc  3840
cctctggcca ggggaaacag cagattcgcc tggatgacca gaaagagcga ggaaaccatc  3900
acccccctgga acttcgagga agtggtggac aagggcgctt ccgcccagag cttcatcgag  3960
cggatgacca acttcgataa gaacctgccc aacgagaagg tgctgcccaa gcacagcctg  4020
ctgtacgagt acttcaccgt gtataacgag ctgaccaaag tgaaatacgt gaccgaggga  4080
atgagaaagc ccgccttcct gagcggcgag cagaaaaagg ccatcgtgga cctgctgttc  4140
aagaccaacc ggaaagtgac cgtgaagcag ctgaaagagg actacttcaa gaaaatcgag  4200
tgcttcgact ccgtggaaat ctccggcgtg gaagatcggt tcaacgcctc cctgggcaca  4260
taccacgatc tgctgaaaat tatcaaggac aaggacttcc tggacaatga ggaaaacgag  4320
gacattctgg aagatatcgt gctgaccctg acactgtttg aggacagaga gatgatcgag  4380
gaacggctga aaacctatgc ccacctgttc gacgacaaag tgatgaagca gctgaagcgg  4440
cggagataca ccggctgggg caggctgagc cggaagctga tcaacggcat ccgggacaag  4500
cagtccggca agacaatcct ggatttcctg aagtccgacg gcttcgccaa cagaaacttc  4560
atgcagctga tccacgacga cagcctgacc tttaaagagg acatccagaa agcccaggtg  4620
tccggccagg gcgatagcct gcacgagcac attgccaatc tggccggcag ccccgccatt  4680
aagaagggca tcctgcagac agtgaaggtg gtggacgagc tcgtgaaagt gatgggccgg  4740
cacaagcccg agaacatcgt gatcgaaatg gccagagaga accagaccac ccagaaggga  4800
cagaagaaca gccgcgagag aatgaagcgg atcgaagagg gcatcaaaga gctgggcagc  4860
cagatcctga aagaacaccc cgtggaaaac acccagctgc agaacgagaa gctgtacctg  4920
tactacctgc agaatgggcg ggatatgtac gtggaccagg aactggacat caaccggctg  4980
tccgactacg atgtggacca tatcgtgcct cagagctttc tgaaggacga ctccatcgac  5040
aacaaggtgc tgaccagaag cgacaagaac cggggcaaga gcgacaacgt gccctccgaa  5100
gaggtcgtga agaagatgaa gaactactgg cggcagctgc tgaacgccaa gctgattacc  5160
cagagaaagt tcgacaatct gaccaaggcc gagagaggcg gcctgagcga actggataag  5220
gccggcttca tcaagagaca gctggtggaa acccggcaga tcacaaagca cgtggcacag  5280
atcctggact cccggatgaa cactaagtac gacgagaatg acaagctgat ccgggaagtg  5340
aaagtgatca ccctgaagtc caagctggtg tccgatttcc ggaaggattt ccagttttac  5400
aaagtgcgcg agatcaacaa ctaccaccac gcccacgacg cctacctgaa cgccgtcgtg  5460
ggaaccgccc tgatcaaaaa gtaccctaag ctggaaagcg agttcgtgta cggcgactac  5520
aaggtgtacg acgtgcggaa gatgatcgcc aagagcgagc aggaaatcgg caaggctacc  5580
gccaagtact tcttctacag caacatcatg aactttttca agaccgagat taccctggcc  5640
aacggcgaga tccggaagcg gcctctgatc gagacaaacg gcgaaaccgg ggagatcgtg  5700
tgggataagg gccgggattt tgccaccgtg cggaaagtgc tgagcatgcc ccaagtgaat  5760
atcgtgaaaa agaccgaggt gcagacaggc ggcttcagca aagagtctat cctgcccaag  5820
aggaacagcg ataagctgat cgccagaaag aaggactggg accctaagaa gtacggcggc  5880
ttcgacagcc ccaccgtggc ctattctgtg ctggtggtgg ccaaagtgga aaagggcaag  5940
tccaagaaac tgaagagtgt gaaagagctg ctggggatca ccatcatgga aagaagcagc  6000
ttcgagaaga atcccatcga ctttctggaa gccaagggct acaaagaagt gaaaaaggac  6060
ctgatcatca agctgcctaa gtactccctg ttcgagctgg aaaacggccg gaagagaatg  6120
ctggcctctg ccggcgaact gcagaaggga aacgaactgg ccctgccctc caaatatgtg  6180
aacttcctgt acctggccag ccactatgag aagctgaagg gctcccccga ggataatgag  6240
cagaaacagc tgtttgtgga acagcacaag cactacctgg acgagatcat cgagcagatc  6300
agcgagttct ccaagagagt gatcctggcc gacgctaatc tggacaaagt gctgtccgcc  6360
tacaacaagc accgggataa gcccatcaga gagcaggccg agaatatcat ccacctgttt  6420
accctgacca atctgggagc ccctgccgcc ttcaagtact ttgacaccac catcgaccgg  6480
aagaggtaca ccagcaccaa agaggtgctg gacgccaccc tgatccacca gagcatcacc  6540
ggcctgtacg agacacggat cgacctgtct cagctgggag gcgacaaaag gccggcggcc  6600
acgaaaaagg ccggccaggc aaaaaagaaa aagtaattaa ttaagaggac ggcgagaagt  6660
aatcatatgt ccgcattttg cgcaaaccag gcgcttagac aatttgcgcg taagcacatt  6720
cgaaatgtga aaagctgaaa gcagtggttt cgccagcccg agttcagcga aacggattcc  6780
ttccaagtgt ttgcattcct ggcggagtgt tcctcccaaa atgcactcac cctgcgtgca  6840
gtgccaaatc gtgagtttcc taattttttc atattgttta ttacctacca actaaagttg  6900
ttgttatata ttgcgtttta cgtacgacaa ataagttcgt attcagaaat atttgcgata  6960
agagagaact catttgcgat gaatctcatt gtatttagct aagtgccttg ataagtaagc  7020
ggaacagcag gaatatgaca ctccttggga aatacatgta agcgtctgta attagatata  7080
tatacacgca accaaatggt ccatggttga tttaagcact gcctgttgtc gaacattgct  7140
ataagcaaaa taaagaagca ttcattaatc taaaatttct tcaaagtgac ttcaatgatg  7200
atctctaggc tatagtgaaa gctgaaagct tatttgacaa tgcaagggaa agtgacgcac  7260
```

-continued

```
gtgcgtcgta tgggaccgcg cgcatctatt ctctcagcta attcccctaa tcattagtaa  7320
ttgacggcac gatttctgct tcttacttcc ttttactttg gagcttttca tcaataaaac  7380
cagtaccatg gccgtacgct caacggaaaa gcattcaaaa aaacccgcgt tcctcgtgtg  7440
atttgtgggt gagtggcgcc atctattaga gaatagctgt actacatctc gtggacgaag  7500
gggtcagaga agttgaaaga gagcttgatc gactgctatc caagctaggc gaggaaggga  7560
gatcgctaga gcaaaagaaa aaaaataagc aaatatcttt ttttataaca aatcgacgtt  7620
agcgaaatat gtttgaatcg atttaacggt tagaattccc tttggttcgt tcattatgcg  7680
aggcgcgcct ttgtatgcgt gcgcttgaag ggttgatcgg aaccttacaa cagttgtagc  7740
tatacggctg cgtgtggctt ctaacgttat ccatcgctag aagtgaaacg aatgtgcgta  7800
ggtatatata tgaaatggag ttgctctctg ctgtttaaca caggtcaagc gggttttaga  7860
gctagaaata gcaagttaaa ataaggctag tccgttatca acttgaaaaa gtggcaccga  7920
gtcggtgctt tttttttttg tatgcgtgcg cttgaagggt tgatcggaac cttacaacag  7980
ttgtagctat acggctgcgt gtggcttcta acgttatcca tcgctagaag tgaaacgaat  8040
gtgcgtaggt atatatatga aatggagttg ctctctgctg caataccacc cgtcagaggt  8100
tttagagcta gaaatagcaa gttaaaataa ggctagtccg ttatcaactt gaaaaagtgg  8160
caccgagtcg gtgctttttt ttacgcgtgg gtcccatggg tgaggtggag tacgcgcccg  8220
gggagcccaa gggcacgccc tggcacccgc a                                 8251
```

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

SEQ ID NO: 53          moltype = DNA  length = 44
FEATURE                Location/Qualifiers
misc_feature           1..44
                       note = multidsx?31L-R primer
source                 1..44
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 53
tgaacgattg gggtaccggt cttgacctgt gttaaacata aatg                   44

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

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

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

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

-continued

```
gagggtctca 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 ggtcaagcgg                                              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
ttatgtttaa cacaggtcaa gcggtggtca                                   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
cctttccatt catttatgtt taacacaggt caagcggtgg 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 tggtcaacga atactca                          37

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

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

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

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

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

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

SEQ ID NO: 72           moltype = DNA  length = 29
FEATURE                 Location/Qualifiers
misc_feature            1..29
                        note = intron 4 exon 5 boundary
source                  1..29
                        mol_type = other DNA
                        organism = unidentified
SEQUENCE: 72
gtttaacacc ggtggtcaac 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 tggtcaacga 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 ggtcataagc ggtggtcaac 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 ggtggtcaac 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 tggcgtggag ttgcgcaata 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 tggcgtggag ttgcgcaata 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 tggcgtggag ttgcgcaata 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 tggcgtggag ttgcgcaata ccacccgtca gagtggatga  120
```

-continued

```
taaactttc                                                          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
cctttccatt catttatgtt taacacaggt caagcggtgg tcaacgaata ctcacgattg  60
cataatctga acatgtttga tggcgtggag ttgcgcaata ccacccgtca gagtggatga  120
taaactttc                                                          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
cctttccatt catttatgtt taacacaggt caagcggtgg tcaacgaata ctcacgattg  60
cataatctga acatgtttga tggcgtggag ttgcgtaata ccacccgtca gagtggatga  120
taaactttc                                                          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
cctttccatt catttatgtt taacacaggt caagcggtgg tcaacgaata ctcacgattg  60
cataatctga acatgttcga tggcgtggag ttgcgcaata ccacccgtca gagtggatga  120
taaactttc                                                          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
cctttccatt catttatgtt caacacaggt caagcggtgg tcaacgaata ctcacgattg  60
cataatctga acatgttcga tggcgtggag ttgcgcaata ccacccgtca gagtggatga  120
taaactttc                                                          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
cctttccatt catttatgtt caacacaggt caaacggtgg tcaacgaata ctcacgattg  60
cataatctga acatgttcga tggcgtggag ttacgcaata ccacccgtca gagtggatga  120
taaactttc                                                          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
cctttccatt catttatgtt caacacaggt caaacggtgg tcaacgaata ctcacgattg  60
cataatctga acatgttcga tggcgtggag ttacgcaata ccacccgtca gagtggatga  120
taaacttt                                                           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 ctcaagattg  60
cataatctga acatgttcga tggcgtggag ttacgcaata ccacccgtca gagtggatga  120
taaactttc                                                          129

SEQ ID NO: 87           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: 87
cctttccatt catttatgtt caacacaggt caagcggtgg tcaacgaata ctcacgattg  60
cataatctga acatgttcga tggcgtggag ttacgcaata ccacccgtca gagtggatga  120
taaactttc                                                          129

SEQ ID NO: 88           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: 88
ccttaccatg catttatgtt caacacaggt caagcggtgg tcaacgaata ctcacgattg  60
cataatctga acatgttcga tggcgtggag ttacgcaaca ccacccgtca gagtggatga  120
taaactttc                                                          129

SEQ ID NO: 89           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: 89
cctttccatt catttatgtt caacacaggt caagcggtgg tcaacgaata ctcacgattg  60
cataatctga acatgttcga tggcgcggag ttgcgcaata ccacccgtca gagtggatga  120
taaactttc                                                          129

SEQ ID NO: 90           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: 90
cctttccatt catttatgtt caacacaggt caagcggtgg tcaacgaata ctcacgattg  60
cataatctga acatgttcga tggcgcggag ttgcgcaata ccacccgtca gagtggatga  120
taaactttc                                                          129

SEQ ID NO: 91           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: 91
cctttccatt catttatgct caacacaggt caggccgtgg tcaacgaata ctcacgattg  60
cacaatctga acatgttcga tggcgtggag ttgcgcaaca ccacccgtca gagtggatga  120
taaactttc                                                          129

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

-continued

```
                        organism = unidentified
SEQUENCE: 92
cctttccatt catttatgct caacacaggt caggccgtgg tcaacgaata ctcacgattg  60
cacaatctga acatgttcga tggcgtggag ttgcgcaaca ccacccgtca gagtggatga  120
taaactttc                                                          129

SEQ ID NO: 93           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: 93
ctttgccatt tatttatgcc caacacaggt caggccgtgg tcaacgaata ctcacgattg  60
cacaatctga acatgttcga tggcgtagag ttgcgcaacg ccacccgcca gagcggatga  120
taaacttcc                                                          129

SEQ ID NO: 94           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: 94
cctttccatt catttatgtt taacacaggt caagcagtgg tcaacgaata ttcacgattg  60
cataatctga acatgtttga tggcgtggag ttgcgcaata ccacccgtca gagtggatga  120
taaactttc                                                          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 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).

* * * * *