

Biocontrol

FIELD OF THE INVENTION

5 The present invention relates to an expression system capable of providing sterile, but competitive, sperm in arthropods, particularly insects, as well as uses thereof in methods of biocontrol (population control), quality control and sex selection of said arthropods.

10 INTRODUCTION

Insect pests of economic importance, originally native to certain parts of the world, are nowadays widely distributed through international trade and movement of people. Such pests develop large populations and generate damaging infestations of fruit and
15 vegetables worldwide. Potential control methods are wide ranging, including baited spray, direct insecticide spraying, biological control, Integrated pest management (IPM) approaches and the sterile insect technique (SIT) (Malacrida et al., 2007). Current control methods however, rely overwhelmingly on the use of chemical insecticides. Both direct and baited spray have the potential to cause a reduction in pollination due
20 to a decline in bees, and potential for animal or human intoxication. SIT on the other hand, is an environmentally friendly, species-specific method of pest control. It depends on the mass rearing, sterilisation and release of large numbers of sterile males who mate with wild females, causing a reduction in the wild population in the subsequent generation (Dyck et al., 2005; Knipling, 1955). If enough sterile males are
25 released for a sufficient time, the target population will collapse.

SIT relies on irradiation to sterilise the target pest species but this can have a negative impact on the released insects (Alphey, 2002; Alphey, 2007; Alphey *et al.*, 2007). Radiation affects all cells of the insect, not just the gametes, and so a degree of
30 damage to the released insect is unavoidable, with potentially negative effects on its performance (e.g. longevity or mating competitiveness). Radiation-sterilisation needs to be performed at a late developmental stage, limiting the options for release. Also, irradiation instruments are relatively large and costly (to obtain and to run), and tend to impose a degree of centralisation that may be undesirable for some programs. Finally,
35 the isotope-based irradiators which have been the mainstay of SIT programs to date

are becoming less favoured due to security concerns about the presence of substantial amounts of radioisotope in these instruments.

5 Alternatives to radiation have been tried in the past, particularly for mosquitoes. These include chemo-sterilisation and sterilisation by use of cytoplasmic incompatibility (CI, induced by *Wolbachia*), but each of these has its own disadvantages. Chemosterilants tend to be toxic or mutagenic compounds, leading to concerns over worker and environmental safety. *Wolbachia*-based systems depend on the lack of any equivalent *Wolbachia*-infected females in the wild, which may not be the case and also requires
10 that no such females are released; such stringent sex-separation may be difficult to achieve. This and other problems associated with current SIT programs could be overcome by the use of recombinant DNA methods (Morrison et al., 2010; Franz & Robinson, 2011).

15 A transgenic alternative to radiation-sterilisation has been suggested, termed Release of Insects carrying Dominant Lethals (RIDL: (Alphey, 2002; Alphey, 2007; Alphey and Andeasen, 2002; Alphey et al., 2010; Alphey et al., 2007; Alphey and Thomas, 1999; Thomas et al., 2000). In this system, insects are engineered to carry a dominant repressible lethal gene or genetic system. These are released into the wild; progeny of
20 matings between wild insects and RIDL insects that inherit a copy of the RIDL gene or construct will tend to die. The RIDL system may be designed to kill all progeny that inherit it, or only one sex. It may also be designed to kill the affected insects at a particular stage in development; this may have significant advantages in some species, e.g. some mosquitoes (Phuc et al., 2007). RIDL systems have been constructed in a
25 number of pest species (e.g. Fu et al., 2007; Gong et al., 2005; Phuc et al., 2007). Further information on the RIDL system may be found in WO 01/39599.

Our female lethal RIDL technology (female-specific RIDL, fsRIDL) is highly, even sometimes as much as 100%, effective in separating sexes and has been successfully
30 tested in laboratory, greenhouse and semi-field experiments. RIDL could be used with or without radiation to produce an effective product.

Despite the fact that this strategy provides a considerable advantage for SIT implementation on a number of pest insects, such as fruit flies; *Ceratitis capitata*,
35 *Bactrocera oleae* and *Anastrepha ludens*, Lepidoptera; *Pectinophora gossypiella* and

Plutella xylostella and mosquitoes; *Aedes aegypti* and *Aedes albopictus*, and can be employed on its own, irradiation might still be the method of sterilisation in certain markets. This is because, in the female-specific RIDL strains described to date, F1 males are fully viable and females are eliminated at a larval stage (i.e. after egg hatch).
5 Furthermore, regulatory pathways and public acceptance will be significantly eased by providing genetic sterilisation (or a trait that confers this). Genetic sterilisation in males would advantageously augment our current "female lethal" (fs-RIDL) strains.

Thus, there is a need in the art for an expression system that can provide a means of
10 genetic sterilisation in males akin to the effects of radiation in an SIT method, but without the associated reduction in fitness of the irradiated individuals.

Crisanti *et al.*, (Catteruccia *et al.*, 2009; Windbichler *et al.*, 2007; Windbichler *et al.*, 2008) have developed an expression system where an endonuclease (IppO-1, also
15 known as I-Ppol) is linked to the promoter from a constitutive structural gene, Beta-2 Tubulin. There a number of problems with this system, discussed herein, not least of which is that the experiments were largely unsuccessful in achieving their aims. However, it would also be particularly useful to be able to exert a degree of control on the timing of the genetic sterilising effect. This control is missing from the Crisanti
20 system.

For instance, one may want to allow breeding of individuals carrying an expression system in the lab, but also want the system to be activated or triggered when required, such as on or shortly before/after release. In other words, one may want to suppress
25 the effect of the system and/or induce it a particular point.

In short, it is desirable for an expression system of this kind to include a means for exerting control on the effect of the expression system. Such control is often referred to as "conditionality," such that a system including this control is a conditional system.
30 Conditional expression systems are known, for instance in insects, but not in the present context of genetic sterilisation. In any case, these conditional systems are not suitable for male germline expression. In fact, in order to harness existing conditional systems, such as the bipartite tet system, one cannot simply include them in a larger male germline expression system. Doing so, in the male germline, does not serve to
35 control expression of an effector (designed to achieve genetic sterilisation in said

germline). The reasons for are complex, but centre around the unusual conditions invoked by meiosis.

5 Despite this, we have surprisingly developed a suitable system for expression in a male germline. The system is capable of providing genetic sterilisation in the sense that expression of the system in the germline produces sperm that are not capable of forming a viable zygote. We have found that use of a conditional system such as tet can be harnessed, but that a significant overhaul of the entire expression system is required. It is extremely advantageous that the system that we have discovered
10 cleverly replicates the effects of radiation in SIT methods. Indeed, it allows for the production of male sterile insects without resorting to the use of debilitating radiation.

We have, therefore, shown that an arthropod expression system comprising a suitable effector harnessed to a conditional system via other regulatory regions can be made
15 capable of inducing conditional genetic sterility, sometimes referred to herein as "sperm lethality." However, it will be appreciated that the preferred intention is not to kill the sperm per se or even to prevent their production, but instead to produce sperm that cannot pass on their genetic information. However, the sperm are otherwise capable of competing with wildtype sperm or rendering a zygote inviable (i.e. prevent formation
20 of a viable zygote).

This solves the above problems by providing conditional, and preferably repressible, male sterility that works by allowing the production of sperm that are defective, in the sense of being unable to fertilise an egg to give a viable zygote or embryo (one
25 capable of developing to a fertile adult), but are still capable of entering or contacting an egg in such a way as to exclude other sperm.

Further technical background information may be found in GB2404382A, GB2355459A, JP2008067678A, WO2009/016627A, WO2008/134068A, WC Black *et al.* (Trends in Parasitology, 362-370, Vol 27, 2011), C Barreau *et al.* (Development, 1897-1902, Vol 135, 2008), G Fu *et al.* (Proc Natl Acad Sci USA, 4550-4554, Vol 107, 2010) and T Ant *et al.* (BMC Biology, 51, Vol 10, 2012).
30

SUMMARY OF THE INVENTION

Thus, in a first aspect, the present invention provides an arthropod male germline gene expression system suitable for conditional expression of an effector gene in an Arthropod male germline, the system comprising:

- a first expression unit comprising an effector gene and a promoter therefor operably linked thereto;
- a second expression unit comprising a coding sequence for a transcription factor and an upstream regulatory element operably linked thereto, the transcription factor being capable of acting upon the promoter in the first expression unit to drive expression of the effector gene, the upstream regulatory element including:
 - a promoter for the transcription factor; and
 - a 5' UTR adjacent a translation start site for the transcription factor coding sequence;

the upstream regulatory element driving sufficient expression of the transcription factor such that the transcription factor protein in turn drives transcription of the effector gene before meiosis.

The transcription factor is preferably a transcriptional activator, such as tTA, GAL4 or their variants. The effector is preferably an endonuclease, most preferably a 3-Zn finger nuclease. The promoter of the first expression unit is preferably a minimal promoter. The promoter of the upstream regulatory element in the second expression unit is most preferably from *topi*, *aly* or Beta-2 Tubulin (B2T) or homologues thereof.

The homologue is preferably that found in the target arthropod in which the system is to be expressed. This promoter is a male germline promoter, i.e. it is acted upon or activates transcription in the male germline. The 5' UTR in the upstream regulatory element of the second expression unit is preferably that from *hsp83*, preferably from *Medfly* or homologues of *hsp83*, particularly the homologue of *hsp83* found in the target arthropod (i.e. that in which the system is to be expressed). Alternatively, the 5' UTR may be that from B2T or homologues thereof, provided that said 5' UTR has been amended to remove or ameliorate the effects of transcriptional delay signals contained in the wild-type, especially if used in combination with the B2T promoter.

The first or second expression unit may also comprise an enhancer. Either or both of the first and second expression unit promoters, especially a minimal promoter, can be considered to further include an enhancer.

5 The expression units can be provided separately or together within the same construct. If separately, then the expression system can comprise separate constructs. The construct or constructs are preferably plasmids. The plasmids may comprise transposons. The transposons may in turn comprise transposable elements. Examples of transposons may include the piggyBac transposon.

10

The conditional nature of the expression of the effector gene is such that it can be controlled by a user or is otherwise influenced by outside factors. Such factors may be environment factors such as temperature (for instance in the case where the Gal4-UAS system is used), but are most preferably chemical entity, such as tetracycline or its analogues. Temperature can be controlled in the lab, and it could be envisaged that temperature changes during the course of the day or season may be harnessed. However, in some embodiments this is not preferred as one may wish to achieve a finer degree of control. In such instances, it is particularly preferred that the system is conditional in the sense of being inducible and most preferably repressible.

20

Inducible systems are known, for instance a GAL4-UAS set up may be employed where the transcription factor is GAL4 and the first expression comprises (outside of the effector gene coding sequence) the UAS region (CGG-N₁-CCG, where N can be any base) to which GAL4 binds or preferably an oligomer thereof. If the upstream regulatory element in the second expression unit comprises a suitable promoter and 5'UTR, then transcription of the Gal4 transcription factor may be induced by provision of a peptide or hormone for instance that acts on the promoter of the Gal4 transcription factor (directly or indirectly, i.e. causes transcription to be induced).

25

30

In another preferred example, the system may be an inducible system, where induction occurs by provision of a chemical entity, such a tetracycline or one of its analogues including doxycycline. In such a situation, the use of rtTA ("reverse tTA") as the transcription factor may be employed for instance, so that rtTA binds DNA only in the presence of tetracycline an analogue such as doxycycline. rtTA is described in WO

35

2001/059088 *inter alia*. In this case, provision of tetracycline (i.e. in the diet) or an

analogue such as doxycycline will allow the rtTA transcription factor in the present system to act on the first expression unit and hence induce expression of the present effector gene.

5 However, it is preferred that the system is repressible. A preferred example is that the transcription factor in the second expression unit is tTA or a variant thereof (tTAV, tTAV2, tTAV3 etc). These bind DNA unless tetracycline (Tc), or an appropriate analogue, is present. Tetracycline will block the DNA-binding of tTA, so there will be no interaction between tTA and an enhancer in the first expression unit and, therefore,
10 no transcription of the effector gene. Thus, as is well known (see for instance our RIDL publication referred to herein), tetracycline can be provided in the diet until such time as it is desired to de-repress (i.e. remove or relieve the repression of) expression of the effector gene. In the absence of tetracycline (or an analogue such as doxycycline), such as after field release or upon a switch of diet in the lab, the effector gene is
15 expressed.

Thus, in some embodiments it is preferred that the system is inducible, but in other embodiments, which are particularly preferred, the system is repressible.

20 The two expression units are preferably one of the two parts of a bipartite (conditional) expression system. Preferred examples include GAL4:UAS and the various tet systems. In the first case, the transcription factor of the second expression unit is preferably Gal4, whilst the first expression unit preferably comprises the UAS sequence for GAL-4 to bind. Suitable variants of GAL4, such as GAL4-VP16 are also envisaged.

25 In general, either or both of the expression units may comprise an enhancer. However, it is particularly preferred that the first expression unit comprises an enhancer. It is preferred that the transcription factor of the second expression unit is tTA or a variant (i.e. when the present expression system utilises the tet system to provide
30 conditionality). When this is the case, then the first expression unit preferably includes the tet operator (tetO). The tetO-mini promoter (tRE) element is particularly preferred. It provides together the promoter and enhancer elements of the upstream regulatory element of the second expression unit. The 426-bp TRE promoter contains seven tetO 18-mers fused to a mini-cytomegalo-virus (mini-CMV) promoter (see for instance M.
35 Ghosh *et al.*, Mol. Cell Biol. 2004, 24(23) 10193).

Although bipartite systems are preferred, where the first and second expression units are provided in separate constructs, the first expression unit and the second expression unit may also be provided on the same construct or plasmid. As such, the present system is preferably a plasmid or consists of two plasmids. Most preferably the two expression units are transformed as a single plasmid or a vector such that they are inserted at the same locus in the genome.

The arthropod is preferably an insect, as described further below. Male germline will be understood to include sperm, such that the system is capable of expressing the effector gene in sperm.

The effector confers or imparts paternal effect lethality under at least some circumstances, i.e. is, or is a part of a "paternal effect lethal" genetic system. In such a system, death of the offspring (taken here from the stage at which the sperm enters the egg, which in insects may not be simultaneous with membrane fusion) depends on the genotype of the father rather than of the zygote (or potential zygote). So, for example, in a dominant paternal effect lethal system, at least some of the wild type offspring from a heterozygous male mating a (homozygous) wild type female will be affected. Additional zygotic effects are possible of course, and anticipated within the invention. Several potential modes of action for such an effector are possible and envisioned. For example, in some embodiments, the effector is or comprises a nuclease. As such, sterility is achieved via what we term "paternal effect lethal." Here, the effector is expressed (to provide a functional nuclease protein) in the sperm. This may lead to DNA in sperm being affected such that the fertilised embryo has a reduced survival probability; indeed, this is a preferred example of a mechanism by which the paternal effect lethality may be achieved. It is also possible that the effector protein is passed into the egg where DNA cleavage can also occur. However, it is also envisaged that at least some effector transcript may also be passed into the egg, from the sperm, and translated in the egg. In both instances, the nuclease effector can thereby take its effect in the egg as well as the sperm.

The system is suitable for expression of the effector, but it will be understood that this may also be preferably referred to as 'capable of such expression' or 'adapted to express an effector in the male germline.'

The effector gene will be described further below, but it is preferably a reporter, such as a marker e.g. a fluorescent protein such as GFP, YFP and so forth. More preferably, however, the effector gene is a nuclease, of which suitable examples are provided
5 below. Most preferably it is both a nuclease and a reporter, for example a nuclease-fluorescent protein fusion (preferred examples of which include the well known Green Fluorescent Protein or Yellow Fluorescent Protein).

Where reference is made herein to a gene 'being' a reporter or a nuclease, for
10 instance, it will be appreciated the gene comprises DNA or RNA encoding a protein having that stated function.

The first expression unit comprises the effector gene. It also comprises a promoter therefor and which is operably linked thereto. The promoter is therefore suitable for
15 driving transcription of the effector gene. As mentioned above, the first expression unit may also comprise an enhancer, or at least a binding region or sequence for (i.e. recognised by) the transcription factor of the second expression unit.

The second expression unit comprises a coding sequence for the transcription factor.
20 It also comprises an upstream regulatory element. This in turn is operably linked to the coding sequence for the transcription factor so that it can drive transcription of the transcription factor.

The transcription factor is capable of acting upon the promoter in the first expression
25 unit to drive expression of the effector gene, although this may be via an enhancer of course, i.e. the transcription factor may not act directly on the promoter but via an enhancer instead. Other regulatory elements such as those for a 5' cap, a 5' UTR, 3' UTR and polyA tail are of course envisaged.

30 To drive transcription of the transcription factor, the upstream regulatory element of the second expression unit comprises a promoter and 5' UTR. Both of these need careful selection in order to provide sufficient expression of the effector. As such, the promoter of the second expression unit (i.e. in the upstream regulatory element) is preferably that from Beta-2 Tubulin (B2T). Alternatively, and more preferably, the promoter is
35 from topi or aly. The 5' UTR in the second expression unit (i.e. in the upstream

regulatory element) is preferably that from hsp83, for instance that from Medfly, but it may also be from B2T. If both the 5' UTR and promoter is from B2T, then one or other must be amended to such the translational delay signals removed or ameliorated. Reference to these genes includes their homologues of course.

5

The 5' UTR is defined herein as the sequence 5' adjacent (i.e. upstream) to the translation start site (i.e. minimally, the ATG start sit for translation). It has a median length of about 150 bases in eukaryotes and preferably extends up to the promoter, for instance around 50-500 bases upstream of the ATG start site.

10

Although the bulk of the framework in the second expression unit may be from B2T, albeit with the B2T ORF replaced with the transcription factor ORF, it will be appreciated that both the second expression unit's promoter and its 5' UTR cannot be those from B2T: at least one of them needs to be changed to provide sufficient accumulation of the effector transcript to occur before meiosis. If the B2T promoter is to be used, then the B2T 5' UTR must be the amended/ameliorated form described above, or it may be that from hsp83.

15

20

Alternatively, if the B2T 5' UTR is used, then the promoter must be another early-acting promoter. What is required is that the choice of the 5' UTR and the promoter for the second expression element must act together to allow for sufficient accumulation of the effector transcript to occur before meiosis.

25

Preferred examples of alternative promoters are the topi and the aly promoters. These may be used with a range of 5' UTRs. Examples of their sequences are provided below.

30

When reference is made to a particular genetic element such as a promoter, enhancer, 5' UTR or even an ORF being 'from' a certain a named gene, it will be appreciated that it does not actually mean that the element is removed from the reference gene, it simply means that this is the origin of the element. Another way to describe this would be 'derived from.' It is preferred that the species origin of the gene is the same as that of the target species. In other words, where it is desired to express the present effector in a Medfly, it is preferred that the elements are derived from the Medfly homologues of the mentioned gene. Failing that, however, the Drosophila versions are preferred.

35

In fact, it is preferred that the descending order of preference is:

- (most preferred) from the target species (in which the expression of the effector is envisaged);
- 5 - from the target genus (i.e. from another species within the same genus); and finally
- (least preferred) from the target family.

The reason is that the action of the preferred promoters at least may not be very well conserved across species. For instance, a *Drosophila* promoter may not work in
10 Medfly, so a promoter from the Medfly homologue of the same gene is preferred. However, as the coding sequence is well conserved, it is relatively simple to identify the beta-2-tubulin gene (for instance) in a given arthropod and hence identify by routine methods a suitable promoter fragment for the Medfly version.

15 A suitable promoter is normally identified within 1 to 2 kb upstream of the transcription start of the mRNA. Although this range is preferred in the present invention, some male germline promoters can be short and a 100-200 bp stretch is also preferred, either within the 1-2 kb window upstream of the transcription start of the mRNA, or even a 100-200 bp stretch upstream of the transcription start of the mRNA.

20 Similar considerations in terms of sequence conservation apply in respect of 5' UTRs, where primary sequence conservation is low. Thus, it is preferred that the 5' UTR is from the same species as the target arthropod, for instance following the above example, if the target is Medfly, then it is preferred that the 5' UTR is a 5' UTR from the
25 Medfly homologue of the same gene (identifiable as discussed above by reference to the more highly conserved ORF). One way to identify and define a 5' UTR is that it is attached (as RNA) to the 5' end of a sequence encoding the respective ORF, for example beta-2 tubulin.

30 In the case of both the 5' UTR and the promoter, it will be readily identifiable if the sequence used in the present expression is insufficient as there will be no expression of the transcription factor, which can be assayed in usual way or tested via the provision of a fusion protein linking the transcription factor ORF to a fluorescent protein.

Preferably, the promoter in the upstream regulatory element of the second expression unit is the *beta-2-tubulin* promoter. This is most preferably used in combination (in the upstream regulatory element of the second expression unit) with the hsp83 5' UTR described herein.

5

It is also preferred that the promoter in the upstream regulatory element is the *topi* promoter. To assist with identifying a homologue of *topi*, we hereby provide the *topi* ORF (see below). Further guidance is provided in Perzgasga et al (2004), for instance, which is hereby incorporated by reference and describes *topi* orthologues from
10 *Drosophila melanogaster*, *Drosophila pseudoobscura* and *Anopheles gambiae*, the latter being particularly preferred.

Alternatively, it is preferable that the promoter in the upstream regulatory element is the *aly* promoter. *Aly* represents a more recent gene duplication than *topi*, so is not
15 present as a male-germline-specific gene in as wide a range of species as *topi*. Nevertheless, where it is present this can be readily identified in the same way as for *topi* by reference to the conserved ORF. The *aly* ORF is provided below.

The transcription factor of the second expression unit is preferably tTA or a variant
20 thereof and the first expression unit comprises the tet operator (*tetO*). The amino acid sequences of tTA, tTAV, tTAV2 and tTAV3 is given below.

It is therefore preferable that the transcription factor of the second expression unit comprises polynucleotides encoding any of the amino acid sequences of tTA or its
25 variants, for instance those given above. As described above, it is also preferred that the transcription factor is GAL4 or a variant thereof and the first expression unit comprises the UAS site for GAL4.

For both tTA and Gal4, this preferably includes any sequences having (or encoding) at
30 least 70%, at least 90% or at least 95% amino acid sequence 'identity' or even the less stringent 'similarity' over at least 50 residues with one of said SEQ ID NOS, including the variants.

It will be understood that the placement of the transcription factor recognition sequence
35 (the DNA, for instance, sequence to which the transcription factor binds), must be

within the first expression unit and not within the ORF. For instance, with both tetO and UAS the insertion is within a few thousand bases upstream of the ATG start site, for instance. They are normally placed within a few hundred bases of the promoter, but can act out to a couple of kb.

5

In any case, it is preferred that the promoter is a minimal promoter. Together with the enhancer, one can refer to the {enhancer + minimal promoter} merely as a promoter for the sake of simplicity. Then the enhancer (transcriptional activator binding site, e.g. tetO or UAS) is part of the promoter by definition.

10

BRIEF DESCRIPTION OF THE FIGURES

Certain embodiments of the invention will now be described with reference to the following figures, wherein:

15

Figure 1 is a schematic drawing representing a design of the egg hatch rate assay;

Figure 2 shows percentage of OX4282-OX4104 male sterility on and off Tetracycline;

Figure 3 shows percentage of OX4282-OX4458 male sterility on and off Tetracycline;

Figure 4 shows percentage of OX4353 male sterility on and off Tetracycline;

20

Figure 5 shows percentage of OX4353 female sterility on and off Tetracycline;

Figure 6 shows repressible male-specific sterility in OX4718- σ 1 lines;

Figure 7 shows percentage of OX4705 olive fly male sterility on and off Tetracycline;

Figure 8 shows percentage of OX4705 female sterility on and off Tetracycline;

Figure 9 shows OX4466 strain hatch rate assay;

25

Figure 10 shows OX4467-E1 strain hatch rate assay;

Figure 11 shows hatch-rate assay of *Aedes aegypti* lines carrying both topi-tTAV and tetO-Dm-Protamine-FokI alleles;

Figure 12 shows hatch-rate assay of *Aedes aegypti* lines carrying both β 2-tubulin-tTAV and tetO-Ae-Protamine-FokI alleles;

30

Figure 13 shows OX4353 strains crossed to two leading RIDL female lethal lines (OX3864A and OX3647Q);

Figure 14 is a plasmid map of OX3866;

Figure 15 is a plasmid map of OX3867;

Figure 16 is a plasmid map of OX3671;

35

Figure 17 is a plasmid map of OX4112;

- Figure 18 is a plasmid map of OX4103;
Figure 19 is a plasmid map of OX4104;
Figure 20 is a plasmid map of OX3831;
Figure 21 is a plasmid map of OX4458;
5 Figure 22 is a plasmid map of OX4391;
Figure 23 is a plasmid map of OX4286;
Figure 24 is a plasmid map of OX3978;
Figure 25 is a plasmid map of OX4275;
Figure 26 is a plasmid map of OX4254;
10 Figure 27 is a plasmid map of OX4371.

DETAILED DESCRIPTION OF THE INVENTION

- 15 The terms 'coding sequence' and ORF may be used interchangeably herein.

As explained above, in a preferred embodiment, the sperm don't die (as the term 'sperm lethal[ity]' might imply to some). Instead, something in the sperm kills the zygote. When we use a nuclease this is probably genetic damage to the sperm DNA,
20 but it is also possible that RNA or protein carried by the sperm have their effect post-fertilisation (indeed, Burt/Crisanti have suggested this to explain the death of females as well as male embryos from their X-shredder in *An. gambiae*, see Windbichler et al 2008, *supra*).

- 25 The prior art teaches that one should use the B2T gene (promoter and ORF) and replace the ORF with a gene of interest. However, we have found that this does not lend itself to a conditional system of the type described herein. If the system is to be made conditional, we have found that it significant changes have to be made to this setup. One of the problems is that transcription shuts down once the cell enters
30 meiosis. Although translation can still occur in the post-meiotic cell(s), further transcription cannot (post-meiotic transcription of a few exceptional genes has recently been described by Barreau et al (2008), but the general point is valid). Thus, if it is desired to express a protein of interest (from said gene of interest), such as an effector protein, then there must be sufficient expression of the effector gene (to provide a
35 corresponding effector RNA) before meiosis. By "effector RNA" it is meant RNA, e.g.

mRNA, coding for (which when translated and optionally post-translationally modified, cleaved if a fusion protein, and/or folded) provides an amino acid sequence that functions as the effector protein.

5 In the context of the present invention, "sufficient" provision of a transcript or a protein relates to both the quantity and timing of said transcript or protein. Thus, when applied to the effector gene, it means that the present system ensures that at least the transcript for the effector protein is generated before meiosis (timing) and that enough of said transcript is thereby provided to achieve the desired protein function. It will be
10 appreciated that this includes all interim events in protein expression, such as optional processing of the transcript, translation of the transcript into an amino acid sequence having a primary structure and optional modification thereof, and provision of secondary, tertiary and even quaternary structure (say for instance in the case of a dimer).

15

However, when a conditional system is employed, the prior art systems cannot provide sufficient functional effector protein to be effective. This is primarily thought to be because the such conditional systems require not only transcription and translation of the effector protein, but also transcription and translation of a control factor protein that
20 acts as a transcription factor on (the regulatory units of) the effector protein. There is simply not time for the additional requirements of a full cycle of transcription, translation and protein function, followed by further transcription, all before meiosis. In other words if, as is the case here, the conditional system employs a transcription factor, then that transcription factor needs to be transcribed and translated itself to provide a
25 functional transcription factor protein. That transcription factor protein must then in turn act upon the regulatory elements of the effector gene (encoding the effector protein), such that there is sufficient accumulation of effector transcript in each cell *before* meiosis to in time allow translation of the effector transcript in each cell *following* meiosis.

30

What we have found is that it is not merely sufficient to simply replace the ORF in the prior art systems with an ORF coding for a transcription factor. Instead, the B2T promoter region of the art *also* has to be either amended or replaced entirely so that the system can produce sufficient effector transcript before meiosis (to then have a
35 functional (translated) protein post-meiosis). Thus, in our new conditional system, new

upstream regulatory elements (such as a promoter and/or 5'UTR) are required to drive expression of the transcription factor ORF. The new upstream regulatory elements must act early enough in spermatogenesis to generate sufficient transcription factor functional protein to in turn act upon regulatory elements controlling expression of the effector and thereby generate sufficient effector transcript before meiosis occurs and transcription shuts down. It is a particular advantage of the present invention that it is able to achieve this in the context of a conditional system.

The effector protein (i.e. the functional protein encoded by the effector gene) most preferably has a deleterious effect, post-meiotically, on the ability of the sperm to fertilise an egg to produce a viable zygote.

The system is preferably dominant, in the sense that the sterility it causes in the sperm is dominant. This is where the concept of paternal-effect lethality comes in, see further below. The genetic sterility caused by a preferred effector such as a nuclease is preferably dominant *in the male*, so that all sperm from a heterozygous male are affected. It will be appreciated that this is not essential, since we expect to release homozygous males, but preferable. This is in contrast to dominant *in the zygote*, as is the case in conventional RIDL. In these prior art systems, one copy (inherited with the sperm) is enough to kill the zygote, but of course eggs fertilized by sperm (from a heterozygous male) that carry a wild-type allele are not affected.

The arthropod is preferably an insect and suitable examples of both insect and non-insect arthropods are provided herein. However, it is particularly preferred that the arthropod is a mosquito, particularly of a species able to transmit malaria or dengue, or an agricultural pest, such as a fruit fly.

The arthropod in which the present system is expressed is a male arthropod and expression occurs in male germline cells thereof, particularly the gonads, i.e. the testes, where spermatogenesis occurs. In addition, or alternatively, expression may occur in the sperm themselves. Preferably, the expression occurs in the majority of said cells i.e. at least 50% of said cells, but it can be much higher, for instance at least 80%, at least 90%, at least 95% or most preferably 99-100% of said cells.

The expression system is, therefore, capable of, adapted to or suitable for expression of a gene in the male germline of an arthropod. By "gene" here it is meant principally a protein. In other words, the function of this expression system is to express a protein in the male germline (germline cells) of an arthropod. The timing of this expression is crucial and is discussed further elsewhere, but it must be sufficient to render sperm unable to fertilise an egg to give a viable zygote. Ideally, the expression system itself is, or is comprised within, a plasmid, transposon or other transposable genetic element capable of expression in the male germline of the arthropod, i.e. by transformation. The expression system is, therefore, preferably a polynucleotide expression system, preferably DNA, RNA or a mixture of the two.

The expression from the system is preferably conditional. Although it may be inducible, it is particularly preferred that the conditionality is repressible such that expression only occurs in the absence of a repressor. In an inducible system, expression will only occur in the presence of an inducer. A particularly preferred repressible system for inclusion in the present expression system is the *tet* (tetracycline) system or the GAL4/UAS system, both described further herein.

It will be appreciated that the promoter is capable of expression in the male germline cells of an arthropod, i.e. is capable of initiating transcription therein. It will also be appreciated that the promoter is operably linked to the regulatory elements and coding sequence of the present system.

It may be that the system comprises a single coding sequence or more than one coding sequence. The one or more coding sequences may be joined as a fusion or provided separately. Suitable examples of further coding sequences are markers or reporters such as the fluorescent proteins (e.g. GFP, EFP, YFP, etc), but further effectors are also preferred to provide greater specificity.

Aside from the specifically described upstream regulatory elements of the second expression unit, the present system also may include further regulatory elements as appropriate. These facilitate, i.e. enable, expression in the male germline cells of an arthropod. Furthermore, as described below, the further regulatory elements facilitate pre-meiotic processing and translation of RNA. Thus, the further regulatory elements are not part of the promoter but preferably include the 5'UTR in the first expression unit

and/or a 3'UTR in both the first and second expression units. Thus, these further regulatory elements can be considered as untranslated sequences, as is the case with the specified 5' UTR of the second expression unit. They all preferably include at least a substantial part of a 5'UTR or 3'UTR, but most preferably a complete 5'UTR or a complete 3'UTR. Most preferably, at least a substantial part of the 5'UTR and a 3'UTR are provided (and preferably both are complete). The regulatory elements thus may include all or part of a 5'UTR, including the various regulatory sequences, associated or found within the 5'UTR that enable translation of an RNA sequence, for instance one that has recently been transcribed from DNA.

All the promoters described herein should preferably include features such as transcription factor and RNA polymerase binding sites. These are promoter elements (typically) not UTR elements (though there can be an overlap). A 5' UTR generally has some translation start signals, possibly also RNA stability, localisation, translation control and intron sequences (though not necessarily).

The same follows for other further regulatory elements that may include a 5' cap and/or a polyA tail. In other words, ideally all of these elements are present in at least a substantial part, i.e. that sufficient to provide their necessary regulatory function. The skilled person would be able to assess whether or not sufficient expression of the coding sequences was provided in the presence or absence of all or part of these regulatory elements, as the fundamental requirement is the downstream functionality of the sperm. The skilled person is able to assess whether the sperm are able to compete, which is advantageous. He is also able to readily identify and whether the sperm produce viable zygotes, which is not advantageous. If the sperm do not compete, and/or if they do produce viable zygotes, then the expression levels of the effector will need revision. In the present invention, the sperm (by which is meant transformed or GM (genetically modified, i.e. carrying the transgene/the present expression system)) are preferably able to compete with wildtype sperm but do not produce viable zygotes. Thus, if one of these is not achieved, especially if viable zygotes are produced, then the system needs revision.

However, what must be borne in mind is that the processing of the RNA of the effector ORF/coding sequence into said effector transcript occurs pre-meiotically and accumulates in sufficient levels such that the later translated and functional protein has

the desired effect on the ability of the sperm to fertilise an egg to produce a viable zygote. This is discussed in further detail elsewhere.

5 Where the effector encodes the preferred nuclease, for instance, this is a deleterious effect. This occurs at any point (although obviously after the effector transcript has been processed and translated and so forth into a functional protein. The basic requirement is that there must be sufficient transcript of the effector before meiosis as transcription shuts down at that point. However, it is envisioned that the effector protein may be functional post-mitotically (but certainly post-meiotically). T

10

Preferably, all the regulatory elements are homologous to each other, i.e. derived from the same gene. It is particularly preferred that, in order to fine-tune expression levels, that the regulatory elements are heterologous to each other, such that the, for instance, 3'UTR may be derived from a different gene to the chosen 5'UTR. It is preferable that 15 the 5' cap and polyA tail are homologous to the promoter, i.e. from the same gene as the promoter.

20

As the effector gene is one that is not normally expressed in germline cells of a male arthropod, it will also be appreciated that the regulatory elements are heterologous to the gene (i.e. heterologous to the coding sequence). In all of this, it will be appreciated that heterologous refers to the origin of the genetic element such as the promoter, 5'UTR (or other regulatory element) or coding sequence, such that they may come from different genes from the same organism; from conserved genes from different organisms; or even different (unrelated) genes from different organisms. In other 25 words, said elements can be *from* (in the sense of *derived from*, although modification is envisaged) a range of different genes within a single organism or a range of different genes from different organisms, with the proviso that they are sufficient to provide the necessary expression levels in the male germline of the arthropod, i.e. in the gonads or sperm thereof. In particular, the coding sequence may not necessarily be from an 30 arthropod at all whilst for example the regulatory elements may also be preferably derived from bacteria or viruses if this assists with the appropriate degree and timing of protein expression from the coding sequence, particularly in respect of pre-meiotic translation.

The coding sequence encodes an effector protein. As mentioned above, the coding sequence is preferably heterologous to the other elements in the expression system such as the promoter and/or the regulatory elements. Whilst the coding sequence in the expression system will be a polynucleotide, it is capable of being transcribed into a suitable messenger RNA (mRNA) sufficient to be then translated into a functional protein. Alternative splices of the RNA are envisaged, regulated by intronic (splice control) sequences in co-operation with a spliceosome. The advantage of providing or controlling alternative splices is that it adds an additional level of regulation to the protein expression. Further guidance on this is provided in our earlier publication WO 2007/091099.

The timing and location of the expression of the effector protein are critical to this invention and are discussed further below. However, it will be appreciated that the function of the effector is to preferably have a deleterious effect on the ability of the sperm to fertilise an egg to produce a viable zygote.

As described herein, it is important that the female feels as though she has been properly fertilised otherwise she will seek a further mate or, indeed, there may already be other wildtype sperm with which the present sperm needs to compete. However, the sperm will be unable to compete if its general functions are significantly impaired, such as its ability to "swim".

The preferred effector has the deleterious effect on the ability of the sperm to fertilise an egg to produce a viable zygote. Preferably, it induces or directly causes DNA damage. Particular preferred examples of this are nucleases, particularly endonucleases. Further examples of these are provided below.

The promoters drive, i.e. are capable of or adapted to initiate, transcription of the effector coding sequence, as well as the regulatory elements surrounding it, into RNA. The timing of this is crucial and highlighted elsewhere. Translation of the effector can occur before or after meiosis.

The effector protein has a deleterious effect, especially after meiosis (post-meiotically). The sperm are created by spermatogenesis from male germline cells carrying the present expression system. Said sperm therefore preferably carry or comprise at least

one copy of the effector protein, but preferably significantly more than one copy. This deleterious effect reduces the ability of the sperm to fertilise an egg. Thus, the present system acts well before the induction of lethality in embryo-specific systems as disclosed in Horn and Schetelig (Horn, C., Wimmer, A.E. 2003, Schetelig, M.F.,
5 Handler, M.A. 2012). It is strongly preferred that the effector allows the sperm to have sufficient motility such that they are able to compete with normal (i.e. wildtype or untransformed) sperm in the rush to reach the egg. Again it is noteworthy that it is not essential that the sperm carry a copy of the DNA (gene encoding the effector). Indeed, this is preferred and is a key difference from zygotically active RIDL systems, as the
10 present system can:

- ensure that there is no egg hatch;
- be used in combination with genetic sexing;
- provides conditionality; and
- pure-breeding, zygotic lethality and resistance.

15 Generally, insemination is in the female genital tract. Males transfer sperm during mating which the female stores. Mature eggs passing down the female reproductive tract are exposed to this sperm and fertilised. For some species (Medfly, *Aedes aegypti*), females typically mate only once and use the stored sperm for all their eggs.
20 Some other insects, e.g. some moths, mate rather more often.

However, it is particularly preferred that the effector causes sufficient DNA damage that the formation of a viable zygote is impossible, for instance because the haploid genetic information provided by the sperm is damaged. For instance, it is preferred that the
25 sperm's DNA has a double-strand breakage therein. This preferred embodiment results from the use of endonucleases, examples of which are provided herein. Thus, the present invention prevents the formation of viable zygote in the first place, rather than expressing a further protein in the zygote once formed. This is an important distinction over the prior art, such as the RIDL technique, where functional zygotes are
30 formed and then an effector is expressed to kill the zygote. In the present invention, no viable zygote is ever formed.

Thus, it is preferred that the zygote may never advance beyond the single cell stage. Early insect embryos, for instance, divide their nuclei without cell division, then
35 cellularise, so go from 1 cell to >1000 in one step. However, it is also preferred that the

embryo merely fails to hatch as a larva (which is a desirable feature in the field). A minimum requirement is, therefore, preferably that the individual fails to develop into a viable adult.

5 Thus, it can be seen that the present invention provides, as a solution to the problems with the prior art, a delicate balancing act of pre-meiotic accumulation of transcript of/for a heterologous effector protein, preferably an endonuclease, before meiosis occurs in the spermatogenic process, such that, after meiosis, the resulting sperm carry copies of the translated effector sequence which can then take effect in the
10 sperm. This effect takes place before fertilisation, thus effectively rendering the sperm sterile, such that the sperm are sterile, active sperm. This is the "sperm lethal" effect described herein.

There does not seem to be a strong checkpoint for DNA integrity at meiosis. Therefore
15 pre-meiotic damage to DNA is generally tolerated. Indeed, this is probably what radiation does: radiation leads to variable sperm head size, whilst head size is proportional to DNA content. This variable DNA content may therefore derive from uneven segregation of DNA at meiosis due to DNA damage cause by radiation in pre-meiotic cells. However, there is a strong checkpoint in mitosis (normally), so DNA
20 damage in germline stem cells (GSCs) or spermatogonial cells is likely to prevent development of functional (in the sense of ability to fertilise an egg) sperm.

According to a further aspect of the invention, there is provided a method of expressing the effector protein in a gonad or sperm. Preferably, the method comprises
25 transforming the gonad with the present expression system. This aspect also relates to a method of transformation.

Also provided is a method of population control comprising expressing said protein via the present expression system in the gonads of a male arthropod. The preferred
30 arthropods are described herein.

The invention also provides a method of resistance management. In principle for any zygotically active RIDL system, including an embryo-active one, there is a possibility that something in the zygote's genome (e.g. genetic material inherited from the mother)
35 could prevent or reduce the intended lethal effect. In respect of RIDL this would be an

inherited resistance factor. To illustrate, a resistance factor might be something that reduces the level of expression of the lethal effector, or reduces the sensitivity of the target of the lethal effector to that effector. However, for the present invention, when an endonuclease is used, it is hard to envisage how that could happen. The damage is already done – using a nuclease as an example, the DNA of the sperm has already been damaged and it is hard to envision how this might be corrected in the egg, thus making the damage permanent.

With respect to resistance management, one could envision a situation where if we were using female-specific RIDL and saw resistance arise in the field. We could add a system according to the present invention, keep the fsRIDL for sex separation and the present invention for sterility. Alternatively we could simply switch to the present system, however sex-separation has distinct benefits for SIT in some species.

This resistance management is not completely comprehensive – behavioural resistance is still possible (if females can discriminate between fertile and sterile males, there will be strong selection for those that preferentially mate fertile males, thereby avoiding exposure to sperm affected by the invention. Such behavioural resistance has been observed in radiation-based SIT programs, but only very rarely (we believe that there is only know of one good example). This argument is closely analogous to one that proponents of radiation-sterilisation have used, arguing for the superiority of radiation damage in the sperm to a zygotic (RIDL) lethal for the reasons outlined above.

In the present methods, it is preferred that the conditional system, preferably the *tet* system described herein, is used to provide a further degree of control. This allows, for instance, breeding under laboratory conditions, i.e. in the presence of a repressor, such as tetracycline. Upon release, or removal of the tetracycline, the repression of the system is removed and the effector protein is thereby expressed.

Male sterility is useful in the present context in both agriculture, to prevent egg hatch for instance, as well as in disease control, for instance in control of disease vectors such as mosquitoes (those responsible for transmission of malaria and dengue fever, for instance). Thus, the present invention also provides a method of biocontainment comprising expression of the system in a population or release of males (carrying the

system) into the field. With a repressible system, which is preferred, the strain becomes dependent on the repressor and cannot establish in the wild.

5 The invention also provides a method of quality control, for instance, by including a reporter such as a fluorescent protein, preferably Green Fluorescent Protein (GFP) or any of the other coloured fluorescent proteins known in the art. This may be the effector protein per se, acting as a transformation marker. Other examples of fluorescent proteins used as transformation markers include DsRed, DsRed2 and AmCyan.

10

Separate transformation markers may also be used, including those described here. Transcription of these transformation markers may be under the control of a separate promoter to that of the first or second expression unit. Examples of such promoters include muscle actin promoter, 3xP3, hrIE and hr5IE1.

15

However, more preferably, the fluorescent protein can be linked to the effector protein in the present system, so that this reporter protein and expression thereof will allow one to assess the degree of inclusion of a transgene or other effector into the population. This has at least some of the following advantages:

- 20 (1) like any such marker it identifies the presence of the transgene, so one can follow inheritance. The more tightly the marker is linked to the trait of interest e.g. the lethal system, the less likely it is that mutations occur which inactivate one but not the other. In practice, though, if they (the marker and transgene) are on the same inserted DNA segment this is extremely unlikely in any case;
- 25 (2) if linked in the sense of fused, a marker shows expression of the effector protein. This would allow one to look at actual expression. For example, in a preferred instance of tet-repressible expression of a nuclease, fusion of the nuclease to a fluorescent reporter would allow one to check that insects to be released were expressing the nuclease. Presence of the fluorescent marker would tell you that (i)
- 30 the male has at least one copy of the transgene; (ii) that the expression system is functioning correctly in the sense of giving expression of effector in the absence of the repressor; (iii) that the insects are expressing the nuclease-FP fusion (and therefore were not, for example, inadvertently reared in the presence of the repressor); [(iv) assuming male-specific expression, are male]; (v) by inference they
- 35 are indeed sterile. In quality control (QC) terms, this gives you much more certainty

over 'they're sterile' than merely knowing that the male possesses a copy of the transgene, which is what you get from using a linked (genetically linked, e.g. adjacent gene) fluorescent marker;

- 5 (3) with a higher-powered microscope you can see when the expression comes on and where the protein is localised within the cell. This is a helpful development tool and also for monitoring consistency, for instance in ongoing QC to establish whether the system does today what it did yesterday/last year, but also in the context of sperm-to-sperm consistency of expression.; and
- (4) a further advantage is in respect of fluorescent sperm, discussed below.

10

There is a clear functional connection for a nuclease to cleave DNA, so if it is not in the nucleus it is unlikely to have the desired DNA-cleaving effect. As such, it is therefore also preferred that a nuclear localisation signal is provided to ensure that the nuclease is localised to the nucleus.

15

In a further aspect, a method of quality control is hereby provided, comprising inducing or de-repressing expression of the present expression system in a target group of individuals and determining whether those individuals meet expected criteria such as size, number, developmental stage or localisation. For instance, if the system includes means to express a reporter such as a fluorescent protein, either as the effector or as part of a fusion protein for instance, then the individuals where expression from the system has been induced or de-repressed will become visible under suitable wavelengths of light.

20

25

It is preferred that the present system includes at least one spacer. Such spacers can advantageously be positioned between any of the present elements of the system. For instance, a spacer may be provided between the promoter and the regulatory elements and/or between the regulatory elements and the coding sequence, to thereby provide a "buffer" between these elements to ensure proper functionality thereof. As such, the spacer has no function in gene expression other than to separate these elements although it may optionally include a number of restriction sites, if this is deemed to be useful. Ideally, it should not include any transcription binding factor sites, etc as these might interfere with expression of the effector.

30

It is also preferred that the effector may be in the form of a fusion sequence or protein, such that, for instance, a nuclease is fused to a marker such that transcription and translation of the effector also leads to transcription and translation of the marker. This has the advantage of showing exposure of a sperm to a nuclease, the presence of the fluorescent protein being indicative that the nuclease has been expressed. The fluorescent proteins may be viewed under fluorescence microscopy using excitation filters suitable for the particular fluorescent protein. Examples of these are our strains LA4466 or LA4467 (LA4466 = PB-hr5IE1-DsRed-Aeprot-tGFP-EcoRI and LA4467 = PB-hr5IE1-DsRed-Aeprot-tGFP-FokICD). It should be noted that earlier strains were labelled "LA" after the inventor, Luke Alphey, but since then we have adopted the prefix "OX" after the Applicant, Oxitec. As such, prefixes LA and OX may be used interchangeably for an individual strain. LA4466/OX4466 and LA4467/OX4467 are both examples of a nuclease function successfully fused to fluorescent expression.

It is also envisaged that the present system and methods can be used to produce fluorescent sperm. For instance, a reporter such as those mentioned above could be linked to the promoter or, indeed, under a separate promoter, such as *tetO* promoter enhancer system if the effector is tTA or any of its variants. Fluorescent sperm would be advantageous for visual separation of sperm or gonads, particularly in methods of dissection or sex selection. In particular, it infers the ability to determine with which male individual a female has mated, which is useful in the context of a field release program. Such a method might include, providing (e.g. trapping) wild females; dissecting them; looking for stored sperm and see whether such sperm carry the present system, i.e. are fluorescent. This will very quickly tell you whether a female:

- (i) is unmated (has not yet mated);
- (ii) mated with a wild type male (as it shows non- fluorescent sperm);
- (iii) mated with a transgenic male carrying the present system (which would show fluorescent sperm; or
- (iv) mated both types of male (shown by the presence of fluorescent and non-fluorescent sperm).

Since 'who are the females mating?' is a key question in assessing and managing an SIT-type program, this is a useful advantage. There are several papers proposing this, e.g. the Malacrida et al 2007 paper already cited, but not in the context we provide here.

An example of this concept is construct (OX3878), which was designed to fluorescently mark sperm heads, in particular in *Aedes aegypti*. This construct utilised an *Aedes aegypti* tGFP-tagged protamine and its regulatory sequences to express a fluorescent fusion protein in a sperm-specific manner. Strong green fluorescence was detected both in whole dissected testis and in the isolated sperm from OX3878 male mosquitoes. This shows that expression of fluorescent proteins can be driven by the present system to produce fluorescent sperm and, furthermore, that the fluorescence can be usefully detected.

10

Thus, in a further aspect, the invention provides a method of determining the mating status of a female arthropod, comprising use of a system according to the present invention in a transgenic male (i.e. a released) population, wherein said system comprises a marker such as a fluorescent reporter protein; and where sperm is present, assaying for the presence of said marker in a female; the presence of the marker being indicative that the female has mated with a transgenic male carrying the system. The presence of sperm that does not carry the marker is indicative that the female mated with a wild-type (not a transgenic) individual. Where sperm is not present it is indicative that the female has not yet, or recently, mated. In some embodiments, the method includes releasing the transgenic males carrying the present system and allowing them to mate with females (i.e. wildtype females).

20

It will be appreciated that in the present invention the expression of the system must be invoked or allowed to occur. In other words, the conditions required to induce expression of the effector (and if separate, an optional marker such as a reporter) are adhered to. In the case of a repressible tet system, for instance, this means removal of tetracycline from the diet of the transgenic males.

25

The overarching aim of the present invention is to provide male sterility. However, what we have also shown is that it is possible to provide sterile sperm, which are still capable of competing with wildtype sperm. This is advantageous as it leads to a greater degree of population control because if the wildtype sperm is easily capable of outcompeting the transformed sperm, and females likely mate both types of male, then there will be only a marginal reduction in the population at the next generation. Even if

30

females typically only mate once, in such a situation there may be strong selection for increased remating.

An ubiquitin fusion protein may also advantageously be included in the present system.

5 This has the advantages of giving expression of both proteins as a fusion, i.e. single polypeptide, but (co-translational) cleavage into two separate proteins. This is useful if one of the proteins does not tolerate fusions (tTA, for example, tends not to function with N-terminal fusions). You still have co-expression in a compact form, however you lose the ability to use the tag (e.g. fluorescent protein, FP) to determine the sub-cellular
10 location of the fusion protein as it is does not stay fused to it.

It is worth reiterating that the present promoter is an “early acting” germline promoter, thus providing the necessary levels of transcription before meiosis. The promoter is defined further below, but again it is worth bearing in mind that the promoter should
15 preferably not act any earlier than the *topi* promoter after mitosis and, especially, not in the stem cells (at least where the effector does not damage such cell types, i.e. where the effector is not a nuclease).

The promoter should have a germline effect and it is preferable that expression of the
20 system is conditional. Ideally, spermatogenesis should be substantially completed before any negative effects of the expression of the effector are seen. It is preferred that there is no discernable effect on sperm function until after egg entry. Whilst DNA damage could perhaps be seen as ‘a negative effect,’ one can view DNA in a sperm merely as “cargo” as there is no transcription in the sperm. Any DNA damage caused
25 by the effector must be sufficient to prevent the production of viable progeny.

Thus, the present invention preferably provides conditional germline specificity (in terms of expression).

30 In a preferred example, the ‘framework’ for the second expression unit in particular is based on the B2T wild-type gene from Dm or the target arthropod. The coding sequence is replaced with that for the transcription factor and the promoter and/or 5’ UTR sequences are also changed. Indeed, even a heterologous enhancer may be inserted or a heterologous 3’ UTR used. Clearly, if all these changes were made, there

would be little left of the original sequence, so it will be appreciated that this is one way to build up the second expression unit.

5 In respect of the regulatory elements, particularly the promoter and/or 5'UTR of the upstream regulatory element in the second expression unit, it is important that there is no delayed translational effect for the present transcription factor. One way to achieve this is to use the regulatory elements from a gene known to transcribe and translate at a sufficient, preferably strong, level before meiosis. Suitable examples would include chaperone genes, preferably the HSP family of genes, in particular hsp83. In another
10 preferred embodiment, the 3'UTR may be derived from a virus, such as SV40.

In a particularly preferred embodiment, the second expression unit of the present system comprises a promoter from Beta 2 tubulin (B2T) combined with an amended B2T 5' UTR or a 5'UTR from hsp83. Optionally, a 3'UTR from SV40 may be used.
15 Either or both of the promoter and the 5' UTR may be from *topi*. *topi* refers to the *Drosophila* gene *matotopetli*. However, the present invention includes functional homologues and paralogues from other species. These can be identified by reference to the conserved ORF as described above. In the case of a 5'UTR from *topi*, the promoter may also be from *topi*, although it is envisaged that it could be from any other
20 of the promoters disclosed herein, for instance B2T. Again, when the promoter from *topi* is used and/or the 5'UTR from *topi* is used, the 3'UTR is preferably also from *topi*, as are the remaining regulatory elements such as the 5' cap and the polyA tail. The reason for this is that *topi* has an "early" expression pattern in spermatogenesis, such that it is able to drive suitable transcription and translation after mitotic divisions but
25 prior to meiosis.

In the case of B2T, whilst the promoter is useful, we have found that there does appear to be a delay signal involved with the 5'UTR from B2T, hampering early translation. It was for this reason that the 5'UTR of B2T can be replaced by the 5'UTR from, for
30 instance, a chaperone such as hsp83. Thus, it will be seen that whilst in some instances the promoter and regulatory elements are homologous to each other, or at least preferably from conserved homologues from different species, in some instances it may be necessary to use promoters and regulatory elements that are heterologous to each other in order to fine-tune the expression patterns acquired for the present
35 invention. As seen in the Examples, we provide an example of each scenario.

In another aspect, the present invention also provides an arthropod, preferred examples of which are provided herein, transformed with the present system or by the present methods. In other words, the invention also provides a transformant or a
5 genetically modified arthropod, preferably as further defined herein. It will be appreciated that said arthropod is a male, preferably whose gonads carry the present system, such that expression of the effector occurs during spermatogenesis.

10 It is an advantage of the present invention that the promoter and the regulatory elements act together in synergy to provide the desired expression pattern.

As mentioned above, the promoter is preferably from a testis-specific gene or at least one sufficient to provide "early" expression during spermatogenesis. Alternatives include more constitutive promoters, such as structural promoters, for instance, the
15 tubulin family, particularly the beta tubulins and most preferably the beta 2 tubulin promoter, and homologues thereof. When this is used, it is necessary to use upstream regulatory element that does not have the translational delay signals seen with at least some instances of beta 2 tubulin's upstream regulatory element. An advantage to using the B2T promoter is that the B2T gene coding sequence is highly conserved and
20 it and a suitable promoter fragment can be readily identified and isolated from a given arthropod species by a skilled person.

An example of the B2T promoter sequence is given below.

25 An example of the ameliorated B2T promoter sequence is given below.

An example of a 5'UTR from B2T is given below. If B2T promoters from other species are used in the present invention, then a skilled person will be readily able to identify the 5'UTR based on its conserved nature from the above SEQ ID NO. They will then
30 be able to replace it with another 5'UTR. Prepared examples include the 5'UTR from chaperones, particularly the hsp family, particularly hsp83. A suitable example, the 5'UTR from hsp83 is given below.

A suitable example of a 3'UTR that has been used successfully in embodiments of the
35 present invention is that from SV40. This 3'UTR is given below.

The *topi* coding sequence is largely conserved between mosquitoes such as *Aedes aegypti* and Medfly (*C.capitata*). As for B2T, one can clone the coding region of *topi* (or part of it) by sequence similarity (lots of methods including molecular and sequence-based ones), then go 5' to the transcription start and take a chunk of DNA 5' to that as your promoter. How much is not always immediately clear; conservatively go 5' until you reach the next transcribed region and take all that, but in practice male germline promoters tend to be pretty short (a few hundred bases), so 1kb 5' of transcription start should be enough, 2-5 if you are feeling cautious. Trial-and-error testing here is also obvious: hook promoter up to FP, make transgenics, look at expression pattern.

topi is useful because it has early expression and is linked to spermatogenesis. It is also advantageous because it is a relatively "compact" system, i.e. consists of relatively few polynucleotides. Again, it is testes-specific and, indeed, it is expressed earlier than B2T. The expression compared to a B2T promoter is perhaps a little weaker, but this may be advantageous in some respects if the levels of expression need to be ameliorated. *Topi* is an example of a transcription factor and so promoters and/or regulatory elements from other transcription factors that express in the testes and are preferably testes-specific (i.e. expressed only in the testes) are hereby preferred.

We have found that a stronger overall sterilisation effect was seen in crosses where nuclease expression was driven by *Topi* promoter, compared to B2-tubulin, particularly in *Aedes aegypti*. Nevertheless, significant male sterility was observed in both cases, rendering both *topi* and the altered form of B2-tubulin suitable promoters for the "paternal lethality effect" in mosquitoes, especially *Aedes aegypti*.

Genes whose product (e.g. encoded protein) is required only at or after meiosis are likely to be translated only shortly before, or after, meiosis, even if transcribed earlier. In contrast, transcription factors needed to drive the expression of such genes must be expressed (transcribed and translated) early enough for their protein product to accumulate sufficiently to drive adequate expression of target genes prior to the cessation of transcription before the meiotic divisions. Therefore, where it is desired to express a transcriptional activator such as tTA in the male germline, the regulatory elements of a male germline transcription factor may be suitable with minimal modification.

Suitable endonucleases are described in greater detail below. However, preferred embodiments include zinc-finger endonucleases as seen for instance in LA4104. Other alternatives include IppO1, also referred to as I-Ppol, as used by Crisanti *et al* (Catteruccia et al., 2009; Windbichler et al., 2011; Windbichler et al., 2007; Windbichler et al., 2008). This has certain advantages such as it has a very long recognition sequence, which is correspondingly rare in random sequence. However, it does not have high specificity relative to some restriction enzymes, for example, in that it will tolerate (i.e. still cut) sequences with a degree of divergence from the canonical recognition sequences. Windbichler et al 2007 show growth arrest in *An. gambiae* tissue culture cells, which is reasonable evidence (as they also conclude) that expression would be toxic, but they don't show it directly.

A different result is from Windbichler et al 2008, where expression of I-Ppol – which they think should cut only the X chromosome in *An. gambiae* as its target site in the rDNA seems only to be on the X chromosome in this species - gave completely sterile males rather than their expectation (no viable daughters due to damage to the paternally-derived X chromosome, but viable sons). Their proposed explanation, for which they provide some supporting data, was that the I-Ppol itself is transmitted in the sperm to the fertilised egg, where it cuts the maternally-derived X chromosome as well (they seem to assume as protein, but could potentially equally be as mRNA). This is more an issue of perdurance, which relates to protein (or mRNA) stability.

An alternative preferred endonuclease is the *Fok*-1 protamine fusion endonuclease. Further preferred alternatives include the *EcoRI* protamine fusion endonuclease. Protamine is a DNA binding protein and has a generally a very low sequence specificity. This is combined with Fok-1, a type IIS cleavage domain. This cleavage domain must dimerise in order to cleave its target. This is useful, because the site need to be close enough together in order to give rise to non-linear concentration effects. An effector that acts as a monomer is expected to have its effect (here, DNA cleavage) in proportion to its concentration. For many applications one would prefer a non-linear dose-response curve, so that the effect is near zero up to a certain point, but then increases to full effectiveness relatively quickly above that point, the limit of this being a binary 'threshold' effect. A nuclease such as protamine-FokI is predicted to have a degree of this non-linearity. Protamine binds DNA but has little or no sequence

specificity. Therefore at low concentration (e.g. molecules per nucleus) the protamine-FokI proteins will tend to be scattered randomly around the chromatin, rarely being in sufficiently close proximity/orientation to dimerise and cut a chromosome.

5 However, as the concentration increases the probability of such proximity greatly increases, leading to a non-linear relationship between concentration and cutting. This facilitates the selection of a promoter (and specific transgene insertion), as the system is relatively inert even with low-but non-zero levels of off-target (basal) expression, while still having the desired effect at higher expression (in the intended expression domain, de-repressed in the case of a repressible expression system). A similar effect
10 can be achieved where the effector must dimerise (or form a larger complex, e.g. tetramer) prior to binding to DNA. Where a more linear effect is desired, this may readily be accomplished within the method of the invention, by using a nuclease domain that does not need to dimerise, or where the necessary subunits are provided
15 in a single polypeptide (e.g. two copies of the FokI domain rather than one). Additional manipulation of the system can be achieved by using nucleases of greater or lesser sequence specificity, as the available protein molecules will be 'focused' by the specificity and affinity of the DNA binding domain to a larger or smaller number of sites, leading to a greater or lesser degree of concentration at those sites.

20 It is preferred that the Protamine gene (or protein coding sequence) is obtained from the same species as that of the target species. It is preferable that the Protamine gene is derived from *D. melanogaster*. It is also preferable that the Protamine gene is derived from *Aedes aegypti*.

25 The examples relating to the constitutively expressed constructs OX4466 and OX4467 were designed to investigate the functionality of the effector proteins Aeprotamin-EcoR1 and Aeprotamin-FokI, respectively. These examples show that both these effector proteins should certainly work when used in a sperm lethal system, i.e. under
30 the control of the early acting promoter (second expression unit). In fact, we have also gone on to show this for the Aeprotamin-FokI binary construct. Having proved the functionality for both effector proteins, we developed and injected the Aeprotamin-FokI binary construct (see the example "OX4282-OX4627 Topi-tTAV-driven expression of tetO-Ae-Protamine-FokI-CD"), which confirmed the positive outcome of the previous
35 (constitutively expressed) system. It is, therefore, entirely reasonable to expect that an

EcoR1 system will also work in the Sperm Lethal setting. Accordingly, the effector gene is preferably Aeprotamin-EcoR1 or Aeprotamin-Fok1. These are most preferably for use in mosquitoes and in particular *Aedes*, especially *Aedes aegypti*.

5 Other type II endonucleases include Eco32I, BfiI, and MboII, for instance. These endonucleases are homodimeric. In general, dimeric endonucleases are advantageous. They are dimeric because they only cleave DNA when dimerised. When they are suitably dimerised, they lead to double stranded DNA breaks.

10 Other endonucleases may include HEG's (Homing Endonucleases). These HEG's can be monomers or dimers but generally have low specificity (in the sense that they don't require perfect matches to their (very long) recognition sequences, but they certainly don't just cut random sequence). Other alternatives include restriction endonucleases from bacteria. These also have low specificity.

15 Accordingly, the skilled person can choose the level of specificity required. It will be appreciated that low specificity endonucleases will break or damage the DNA on a greater number of occasions than high specificity endonucleases, for a given recognition sequence. HEGs have very long recognition sequences that would occur
20 by chance at a frequency often less than one per genome. So despite their imperfect specificity for that sequence they may not cut at all (e.g. I-SceI in Windbichler *et al.*, 2011 only cuts the engineered site, not the rest of the genome - at least not at a high enough frequency to cause evident trouble in their experiment).

25 Thus, a further level of fine-tuning is possible by appropriate selection of endonucleases as the effector. The nuclease effector fusion protein has been found to be fully functional in three different diptera species tested so far, namely *C. capitata*, *B. oleae* and *Aedes aegypti*. These species are particularly preferred as are other species in the same genus.

30 It is particularly preferred that the arthropod (the host organism in which the present system is expressed) is an insect, preferably a tephritid. In particular, it is preferred that the insect is from the Order Diptera, especially higher Diptera and particularly that it is a tephritid fruit fly, preferably Medfly (*Ceratitis capitata*), preferably Mexfly
35 (*Anastrepha ludens*), preferably Oriental fruit fly (*Bactrocera dorsalis*), Olive fruit fly

(*Bactrocera oleae*), Melon fly (*Bactrocera cucurbitae*), Natal fruit fly (*Ceratitis rosa*), Cherry fruit fly (*Rhagoletis cerasi*), Queensland fruit fly (*Bactrocera tyroni*), Peach fruit fly (*Bactrocera zonata*) Caribbean fruit fly (*Anastrepha suspensa*) or West Indian fruit fly (*Anastrepha obliqua*). It is also particularly preferred that the host organism is a mosquito, preferably from the genera *Stegomyia*, *Aedes*, *Anopheles* or *Culex*. Particularly preferred are *Stegomyia aegyptae*, also known as *Aedes aegypti*, *Stegomyia albopicta* (also known as *Aedes albopictus*), *Anopheles stephensi*, *Anopheles albimanus* and *Anopheles gambiae*.

Within Diptera, another preferred group is Calliphoridae, particularly the New World screwworm (*Cochliomyia hominivorax*), Old World screwworm (*Chrysomya bezziana*) and Australian sheep blowfly (*Lucilia cuprina*). Lepidoptera and Coleoptera are also preferred, especially moths, including codling moth (*Cydia pomonella*), and the silk worm (*Bombyx mori*), the pink bollworm (*Pectinophora gossypiella*), the diamondback moth (*Plutella xylostella*), the Gypsy moth (*Lymantria dispar*), the Navel Orange Worm (*Amyelois transitella*), the Peach Twig Borer (*Anarsia lineatella*) and the rice stem borer (*Tryporyza incertulas*), also the noctuid moths, especially Heliothinae. Among Coleoptera, Japanese beetle (*Popilla japonica*), White-fringed beetle (*Graphognatus* spp.), Boll weevil (*Anthonomous grandis*), corn root worm (*Diabrotica* spp) and Colorado potato beetle (*Leptinotarsa decemlineata*) are particularly preferred.

However, as we have shown in the Examples, the present system and methods can be implemented across Diptera, including higher and lower Diptera. Higher Diptera are therefore preferred. Lower Diptera are also preferred.

In some embodiments, it is preferred that the insect is not a Drosophilid. Thus, in some embodiments, expression in Drosophilids, especially *Drosophila melanogaster*, is excluded.

It is preferred that the expression of the effector protein leads to a phenotypic consequence in the organism, namely sterility. It is particularly preferred that the functional protein can be associated with visible markers (including fluorescence).

Where reference to a particular nucleotide or protein sequence is made, it will be understood that this includes reference to any mutant or variant thereof, having

substantially equivalent biological activity thereto. Preferably, the mutant or variant has at least 85%, preferably at least 90%, preferably at least 95%, preferably at least 99%, preferably at least 99.9%, and most preferably at least 99.99% sequence identity with the reference sequences.

5

We anticipate that:

a) the present sterility is dominant, so that all sperm of a heterozygous male would be affected (not necessarily all defective, as the system might be selective for some type of sperm, or less than 100% effective against those sperm that should be affected);

10

and

b) the mechanism is not dependent on inheritance by the zygote of a dominant lethal gene via these sperm.

15

The present system is therefore not a RIDL system, as both a) and b) distinguish from RIDL.

20

One potential advantage of radiation had been that it is difficult to see what heritable or genetic change in the wild target population could overcome the sterilising effect of radiation. Thus it is difficult to imagine resistance arising in the wild population to this aspect of radiation-based SIT (other types of resistance, such as behavioural resistance e.g. assortative mating (Dhillon et al., 2005), might nonetheless arise). In contrast, genes or alleles conferring resistance to the killing or incapacitating action of RIDL genes might conceivably arise. Similarly, though the precise biochemical basis of CI is not known, embryos infected with suitable strains of *Wolbachia* are able to reverse CI-induced male sterility, therefore a biochemical/genetic resistance to CI-induced sterility seems conceivable.

25

30

The present invention overcomes these difficulties. The sterilising mechanisms are designed to act prior to zygotic gene expression, so that the possibility of heritable resistance is severely restricted. For example, consider the use of a nuclease expressed in spermatogenesis. Under restrictive conditions (for fertility, permissive for expression of the nuclease), this will tend to damage the DNA in much the same way as does radiation, e.g. inducing double-strand breaks. Therefore it is equally difficult to imagine how a wild target population could develop resistance to such sterility or paternal-effect lethality. It may be that the artificial mass-rearing population could

35

develop resistance, however use of a suitable conditional expression system, so that little damage is done to the sperm (or other cells or tissues) under permissive (for fertility) conditions, will minimise the selective pressure for such alleles. Though similar in this respect to radiation, the method of the present invention has several advantages. The sterilising effect is restricted to the gonads or gametes, and so is less likely to reduce the performance of the insects. By selection of a suitable repressible expression system, sterility may be achieved simply by removal of the repressor. In this version, the system also provides biocontainment, in that insects released into the wild (deliberately or inadvertently) are sterile, or some or all of their progeny are sterile.

We intend to combine the female lethal (or flightless) RIDL technology with the present "sperm lethal" invention to develop insect products suitable for implementation in an SIT program. We show in the examples that the present "sperm lethal" system is viable in combination with RIDL technology. Of course there will be also species where sex separation is not desirable, and others where it is desirable but can be achieved by other means (e.g. physical size-based sex separation, as is suitable for *Aedes aegypti*). Thus, preferably, the present expression system can also be combined with existing female lethal (or flightless) technology to develop further arthropod, and particularly insect, products suitable for implementation in an SIT-type program. Male-only release is considered desirable for SIT for several species, for example where adult females are actually or potentially harmful even if sterile (e.g. mosquitoes, by biting, or Medfly, by oviposition into unripe fruit), also for Medfly co-released sterile females appear to 'distract' sterile males from seeking out wild females (Rendón et al., 2004).

Suitable sex separation systems include separation based on natural sexual dimorphism (e.g. size for *Aedes aegypti*, pupation time for tsetse fly); induced sexual dimorphism (e.g. Catteruccia et al., 2005). Alternatively, one sex may be eliminated by use of a lethal genetic sexing system; such systems have been constructed by classical genetics (e.g. Franz, 2005; Klassen and Curtis, 2005) and also by recombinant DNA methods, including embodiments of RIDL (Fu et al., 2007; Thomas et al., 2000). Both conditional expression systems may use the same condition, so that control of each of the two phenotypes is co-ordinately controlled. So, for example, if each system were repressed by tetracycline (or suitable analogues thereof, such as chlortetracycline), rearing in the presence of a suitable concentration of tetracycline would allow the strain to be grown to adequate numbers. In the last generation before

release, rearing without tetracycline would lead to derepression of both the female-specific lethal system and the proposed sperm lethal system. Therefore, females would die and the remaining males would be sterile.

5 A degree of cross-talk between the two tet-repressible systems may be anticipated, such that both effectors will be expressed by both systems. Expression of the sperm lethal effector in females is unlikely to have any undesirable consequences, since it is intended that these females die. Cross-talk the other way, i.e. expression of the female-lethal effector in spermatogenesis may be more problematic, depending on the nature
10 of the female-lethal effector and its effect on spermatogenesis. Such cross-talk, if deemed undesirable, may readily be avoided restricting expression of the female-lethal effector to females, for example by use of sex-specific alternative splicing to regulate the production of functional effector (Fu et al., 2007).

15 The combined technologies can be achieved in two ways:
- the proposed invention can be inserted separately into the insect's genome by transposon mediated transformation. Thorough analysis of the strains generated will lead to candidate strains with the desired features. These can be back crossed to lead female lethal (preferably fs-RIDL) strains to generate double homozygotes for the two
20 insertions.

In a particularly preferred embodiment, the two systems may be combined on a single DNA construct, with a conditional female-specific lethal construct. This construct would therefore provide both genetic sexing and male sterility.

25 Spermatogenesis is a highly specialized process of cellular differentiation resulting in the formation of functional spermatozoa for successful reproduction. In principle, the process of spermatogenesis is well conserved in all sexually proliferating organisms although the size and shape of the mature sperm vary considerably among different
30 species. Many details are comparable between mammals and *Drosophila* making the fly a very good model system to study fertility defects. *Drosophila* germ cells, like those of mammals, are set aside early in embryonic development and migrate through the primordium of the hindgut into the interior of the embryo where they join the somatic parts of the embryonic gonads (Zhao and Garbers, 2002). At the end of the third larval
35 instar and the onset of pupariation, the first germ cells have entered meiosis.

Spermatogenesis is a continuous process during adult life and, thus, the adult testes contain all stages from stem cells to mature sperm.

In general, for male (insect) germline cells in spermatogenesis, transcription shuts down shortly before meiosis (Fuller, 1993). There may be a few genes transcribed later (Barreau et al., 2008), but these are exceptional. In general, genes whose products are required at or after meiosis are nonetheless transcribed before meiosis; the mRNA is stored and translated later, i.e. when required, e.g. post meiotically. The mRNA is then typically degraded after translation. Many genes follow this pattern; examples included β 2-tubulin and protamines, also the meiotic regulator twine. Of these, twine protein is required for entry into meiosis, β 2-tubulin is required for the meiotic spindle and post-meiotically for the flagellum; protamines are required strictly post-meiotically. Thus, suitable control of timing can be afforded.

In principle it should be possible to disrupt sperm production by affecting (e.g. killing) somatic cells in the testis (e.g. cyst cells) that are required for spermatogenesis, or by affecting hormone production etc in the male or (slightly more promising) expressing or abrogating factors in the seminal fluid. However, such methods would at almost certainly produce, at best, males with no sperm, or obviously aberrant sperm. For this reason, we have sought to focus on male-germline expression.

Germ line promoter sequences

Therefore, a single-gene expression construct can be readily made which will express the gene of interest ("effector") in the male germline. Many genes have been identified in *Drosophila* which express during spermatogenesis and many of these are specific to the germline, or to the male germline. Such expression data are readily obtainable directly from the literature; there are also large-scale projects (Chintapalli et al., 2007) and <http://www.flyatlas.org/>) which have performed expression analysis on many of the genes in the *Drosophila* genome (e.g. FlyAtlas (Chintapalli et al., 2007) and <http://www.flyatlas.org/>, *Drosophila* testis expression database (www.fly-ted.org) and data collated on FlyBase e.g. <http://flybase.bio.indiana.edu/>). An expression construct can therefore be made as follows:

(1) identify a suitable testis-expressed gene, e.g. β 2-tubulin, which encodes a male-germline-specific isoform of β -tubulin (Nielsen et al., 2001) (a recent report (Jattani et al., 2009) indicates that the expression and function of β 2-tubulin may not be

strictly confined to the male germline in *Drosophila*, nonetheless a promoter fragment from *An. gambiae* β 2-tubulin was used to express a potent nuclease, known to cut sequences in the *An. gambiae* genome, without obvious effects other than in the male germline and in embryos fertilised by sperm from such males);

- 5 (2) isolate a wild type copy of the gene;
- (3) replace the coding region of the gene with the effector gene; and
- 4) make transgenic insects carrying this gene.

Male-germline-specific expression of fluorescent protein reporters has been achieved
10 by this method in *Drosophila* using a range of genes involved in spermatogenesis such as protamine, *don juan* and *sneaky*, for example (Raja and Renkawitz-Pohl, 2005; Santel et al., 1997; Wilson et al., 2006). β 2-tubulin is a well-conserved gene, so an extension to this method, in step 2 isolating the appropriate gene from the species of interest (or a related species, such as *Anopheles gambiae* for use in *Anopheles*
15 *stephensi* (Catteruccia et al., 2005)) will allow use in an arbitrarily selected arthropod species. This approach has indeed been successfully used for male-germline-specific expression of a fluorescent protein reporter in several species including the Mediterranean fruit fly (Malacrida et al., 2007), *Aedes aegypti* (Maynard-Smith et al., 2007) and *Anopheles gambiae* (Catteruccia et al., 2005), all of which used promoter
20 fragments from the β 2-tubulin of the target species, except for Catteruccia et al. who used a promoter fragment from *Anopheles gambiae* β 2-tubulin to drive spermatogenesis-specific expression of a fluorescent protein in *Anopheles stephensi*.

“Sperm lethality” will be used in conjunction with a conditional expression system to
25 ensure transition through generations under permissive conditions. This condition can be temperature, for example working through a temperature-sensitive protein effector. However, temperature based conditional systems tend to be “leaky”, for example. The TSL sexing medfly strain (Casares, 2002), a system based on the presence/absence of an externally applied molecule (e.g. RU486/mifepristone (Osterwalder et al., 2001))
30 could be an alternative. However, it is not suitable for a bipartite expression system (Brand et al., 1994; Brand and Perrimon, 1993; Fussenegger, 2001; Fussenegger et al., 2000; Gossen and Bujard, 1992; Gossen and Bujard, 2002; Victorinová and Wimmer, 2007).

On the other hand, the Tet-off or Tet-on system is the scheme of choice (Gossen and Bujard, 1992; Gossen and Bujard, 2002). This system depends on the expression of a transcription factor which conditionally drives expression of a suitable gene (the effector). In the tet systems, binding of the synthetic transcription factor tTA (or its variants, including rtTA) is affected by the concentration of tetracycline and/or related compounds such as chlortetracycline or doxycycline. In the tet-off expression system, binding of tTA to tetO is inhibited by tetracycline, therefore expression of the effector is inhibited by tetracycline, hence "Tet-off". A similar expression system regulated by streptogramin is also known (Fussenegger et al., 2000). However, expression of tTA under the control of β 2-tubulin, for example, will not lead to significant expression of the effector, even in the absence of tetracycline. This is because expression of tTA will be substantially post-meiotic (more precisely, substantially after the shut-down of transcription prior to meiosis), and therefore too late to drive expression of the effector, even in the absence of tetracycline.

The key is to understand the timing of gene expression in spermatogenesis in arthropods and in particular in insects. There is essentially no gene expression after initiation of meiosis. Therefore, to use a bipartite expression system such as tet-off (or tet-on), transcription of the sequence-specific transcription factor (tTA) has to be sufficiently far in advance of meiosis to allow accumulation of tTA mRNA, translation of tTA protein and tTA-dependent transcription of the effector all before meiosis. Most genes expressed during spermatogenesis are transcribed in spermatocytes (i.e. before meiosis) but then translated later, when the protein product is actually required. For most proteins involved in sperm differentiation or function (and therefore most sperm proteins), this means translation after meiosis. Therefore, suitable expression for our purposes means early transcription AND early translation of the tTA. This is not widely recognised and also not easy to achieve. Thus, we were the first to identify that there was a problem here.

A suitable 3'UTR, can be constructed using heterologous sequence, rather than sequence derived from the testis-expressed gene, also all or part of the coding region of the testis-expressed gene may be retained, for example. Male-germline-specific expression of fluorescent protein reporters has been achieved by this method in *Drosophila* using a range of genes involved in spermatogenesis such as protamine, *don juan* and *sneaky*, for example (Raja and Renkawitz-Pohl, 2005; Santel et al., 1997;

Wilson et al., 2006). Similarly, 5' UTR can be substituted with a heterologous sequence to avoid translation after meiosis. The five prime untranslated region (5' UTR), can contain sequences for controlling gene expression by way of regulatory elements for example. sequences that promote or inhibit translation initiation or introns within 5' UTRs have been linked to regulation of gene expression and mRNA export (Cenic et al., 2011). It begins at the transcription start site and ends one nucleotide (nt) before the start codon (usually AUG) of the coding region.

In order to obtain adequate expression of tTAV and respectively of the effector gene, in other words in order to obtain "sperm lethality," we have substituted (in the upstream regulatory element of the second expression unit) the 5' UTR, and in some cases the 3' UTR, of the β 2-tubulin gene with the corresponding sequence(s) from hsp83 (Theodoraki & Mintzas, 2006) and SV40 (Simian vacuolating virus 40 or *Simian virus 40*) (Cheng et al., 2009), respectively, see the discussion herein and the present Examples.

In other words, the preferred 5' UTR is from hsp83. This is preferably from Medfly as that is what we used in specific examples, but homologues from other species are also preferred. The preferred 3' UTR is from SV40, but could also be from hsp83 (species as above for the 5' UTR). We have found the choice of the 5' UTR to be more important however than that of the 3' UTR.

Earlier expression of a transactivator, exemplified here as tTA may be also achieved by any of the following methods. One example is by identifying promoters acting earlier in spermatogenesis. This means promoters acting such that a transcription factor, e.g. tTA, expressed under the control of this promoter, *will* be able to drive the expression of a tTA-responsive effector construct. We prefer this not to be a promoter substantially active in the germline stem cells, as such genes (e.g. *vasa*) are expressed in immature stages; this will tend to lead to expression of the effector (in induced or de-repressed conditions) to be expressed very early in development, possibly leading to damage to or loss of the germline and hence the production of no or fewer gametes. If the expression system were induced (or derepressed) later in development, the time course of spermatogenesis, i.e. the time between stem cell division and the production of mature sperm (several days in *Drosophila*), would lead to a slow response (e.g.

sterility) to the induction (or derepression); this effect is compounded by the ability of males to store sperm.

We therefore prefer to use promoter elements from genes expressed in spermatogonia or primary spermatocytes. As discussed above, many genes expressed in primary spermatocytes are involved in later functions and are translated post-meiotically, therefore we prefer genes (and promoters from genes) with pre-meiotic functions. A particularly preferred class of genes are those encoding transcription factors, or components of transcriptional complexes, which are involved in the expression of spermatogenesis genes, e.g. later-acting genes. A key reason for preferring such genes is that they are, almost by definition, expressed (at the protein level) early enough in spermatogenesis to be able to drive the pre-meiotic expression of other genes (of course negative regulatory elements of such complexes are also preferred, it is the timing of expression rather than the specific function of the translated protein that is important).

Several classes of such genes are known. However, as discussed below, not all are suitable and some are preferred over others. Classes include the *can*-class of meiotic arrest genes. This class comprises *can*, *mia*, *sa*, *nht* and *rye*; these encode testis specific paralogues of broadly expressed TATA-binding protein associated factors (TAFs) (Hiller et al., 2004). Another class is the *aly*- class of meiotic arrest genes. These include *aly*, *comr*, *achi/vis*, *topi* and *tomb* which encode components of a complex of sequence specific DNA binding proteins and associated factors paralogous to a broadly expressed transcriptional regulatory complex known as dREAM or Myb-MuvB (Beall et al., 2007; Jiang et al., 2007; Jiang and White-Cooper, 2003; White-Cooper et al., 2000). Most of these genes appear to have arisen by duplication of conserved 'somatic' genes, e.g. ancestral TAFs or Myb-MuvB components that functioned in both soma and germline, followed by specialisation of these versions to a germline- or male-germline-specific expression and role. Several of these duplications appear to have occurred relatively recently in evolution, meaning that the germline-specific paralogs exist in a variable, and in some cases quite narrow, phylogenetic/taxonomic range. We prefer to use the promoter(s) of gene(s) conserved from *Drosophila* to the species of interest, and particularly prefer promoters of genes conserved across a wide phylogenetic/taxonomic range; this conservation both makes

the identification of homologues easier, and makes the specific gene more generally useful.

Such (conserved early acting) genes can be identified in one of two ways:

5

1. **By homology with genes known, e.g. from *Drosophila*, to have suitable patterns and levels of expression.**

10 This is achieved by homology searches in sequence databases. However, some spermatogenesis genes are rapidly evolving, and specific homologues may not be readily identifiable across a wide phylogenetic range. For example, *bag-of-marbles* (*bam*) meets the criteria above of early expression – *bam* is expressed in spermatogonia (and also in oogenesis) and early function – it is involved in regulating the number of mitotic divisions of spermatogonia (Gonczy et al., 1997; Kawase et al., 15 2004). Furthermore, a fragment of the *bam* promoter has been used successfully in a bipartite expression system (GAL4/UAS) in spermatogenesis (and also oogenesis) in *Drosophila* (Jiang et al., 2007). However, *bam* is recognisable only in Drosophilids. Another gene with a similar expression pattern, whose gene product is thought to interact with that of *bam*, is *benign gonial cell neoplasm* (*bgcn*). Unfortunately, this 20 gene is also not well conserved across insects; both genes show signs of positive selection and rapid evolution (Bauer DuMont et al., 2007).

Difficulties also arise with several members of the *can*-class and *aly*-class meiotic arrest genes. Many of them appear to have derived from a duplication of an ancestral 25 gene presumably used in both somatic and germline cells. The germline versions tend to be rapidly evolving, so that the somatic paralogue is more easily identified than the germline paralogue. Furthermore, the series of duplications leading to these somatic/germline gene pairs in *Drosophila* seems to post-date the divergence of Diptera from other orders of insects, so germline-specific paralogs of most of these 30 genes do not exist outside Diptera, and several of them only in higher Diptera.

However, exceptions do exist; these are readily identified by sequence comparisons with insects from other orders (e.g. *Apis mellifera*, *Tribolium castaneum*, *Bombyx mori* – the list of insects for which a significant fraction of the genome has been sequenced 35 is rapidly increasing and readily accessed by a person skilled in the art). One example of such a gene is *matotopetli* (*topi*), an *aly*-class gene which encodes a putative

sequence-specific DNA binding protein identified in a two-hybrid screen as binding to another *aly*-class protein (Comr) and is therefore presumably a component of a testis-specific Myb-MuvB complex. Topi homologues were identified in *Drosophila* and *Anopheles gambiae* (Perezgasga et al., 2004); by sequence similarity searches of public sequence databases we have also identified homologues in several other orders of insects including Coleoptera (*Tribolium castaneum*, NW_001092862.1), Hymenoptera (*Apis mellifera*, NW_001253507.1). Topi is therefore an example of a well-conserved gene with a suitable expression pattern (based on (i) quantitative rtPCR data on FlyAtlas and elsewhere; (ii) developmental expression pattern shown by RNA *in situ* hybridisation to be in primary spermatocytes, especially early primary spermatocytes (Perezgasga et al., 2004); (iii) predicted function as part of a male germline-specific transcription complex (Beall et al., 2007; Perezgasga et al., 2004); (iv) actual function, by phenotypic analysis of mutants, affects expression of a number of other spermatogenesis genes (Perezgasga et al., 2004)).

Genes with early acting promoters can also be identified by:

2. Homolog identification of the selected gene(s) from the target species

Alignments of relevant sequences from species of greater or lesser phylogenetic distance from the species of interest will identify regions likely to be conserved in gene of interest in the target species. Preferably, to enrich for the gene of interest and to avoid problems with introns, cDNA from RNA extracted from males will be used, most preferably from dissected testes. From a cloned fragment of the transcribed region of the gene of interest, a suitable promoter fragment may be obtained.

A candidate for a suitable promoter fragment comprises at least 50 nucleotide bases of genomic DNA 5' to the transcription start, and, preferably, the transcription start itself and at least 1 nucleotides of transcribed region. This will comprise the promoter fragment and an open reading frame for the reporter. It may also contain 5' and 3' UTR sequences, either from the same gene as the promoter fragment, or another region. Those from the same gene are likely to function in spermatogenesis, if the gene has been correctly identified as active in spermatogenesis, however they may also contain signals, e.g. for delayed translation, that would be undesirable for expression of a protein, such as tTA, that must function pre-meiotically. Regulatory sequences from the same gene as the promoter fragment are therefore preferred for

genes whose function is known or thought (e.g. by homology) to be pre-meiotic (e.g. *topi*), but heterologous regulatory sequences are preferred for genes whose function is known or thought to be meiotic or post-meiotic.

- 5 For the purposes of our work we have isolated and tested *topi* promoter sequences from *C. capitata* and *A. aegypti* with various effector proteins, the present Examples.

Protein effectors to give a late sperm phenotype

10 'Late' here means after the point to which sperm function is required, e.g. transfer to the female or entry into the egg. The objective is to produce sperm that are transferred to the female and will induce refractoriness to remating in the female, and will do well in sperm competition if the female does remate. Sperm competition occurs, or can occur, when a female mates more than one male; in this circumstance the question arises of which males sire what proportion of the female's embryos. It is likely that males
15 transferring no sperm will do badly in this instance. Because of the likelihood or possibility of evolutionary responses to the use of males with no sperm, or sperm incapable of being transferred to a female, or incapable of competing with other sperm after transfer to a female, it is desirable for biological control purposes to engineer sterile males that do produce sperm, that these sperm are capable of being transferred
20 to a female and are capable of competing with other sperm after transfer to a female.

Paternal-Effect Lethals

Our approach, which is our favoured method to achieve "sterility" is to this is to
25 construct paternal-effect lethals, whereby the sperm enter an egg but no viable zygote (capable of developing to a fertile adult) is formed. Preferably, this effect is generated by males with a single copy of the paternal-effect lethal, though the use of multiple copies is also envisioned. The release into the environment of males homozygous for a conditional paternal-effect lethal is particularly preferred, as such a strain is
30 substantially stable during rearing. Paternal-effect lethals (Pals) of the invention characteristically affect all or most of the sperm produced by a male carrying at least one copy of the Pal. In particular, the zygote may be adversely affected, e.g. killed, sterilised or its development (e.g. sexual phenotype(s)) changed, based on the genotype of the parent. This is in contrast to RIDL, wherein the effect on the zygote is
35 based on the genotype of the embryo. Thus, progeny of mating between a wild type female and a male heterozygous for a RIDL construct at a single locus will typically

give ~50% normal progeny and ~50% which inherit the RIDL construct and may be affected by it. In contrast, all of the progeny of a male heterozygous for a Pal may be affected.

- 5 Several types of protein-based effector are envisioned. Protein or RNA effectors may be transmitted with the sperm to affect the egg or developing zygote. These may be on the surface of the sperm (e.g. as *Drosophila* sex peptide, SP), or produced and stored within it. Sperm, and cells in other stages in spermatogenesis, are remarkably resistant to some types of protein, for example pro-apoptotic proteins, presumably
- 10 because the apoptotic pathway has been co-opted for other purposes in spermatogenesis (Arama et al., 2003; Cagan, 2003).

It is preferred that the effector have a direct biochemical effect on the sperm, rather than merely using the sperm as a vehicle via which to enter the egg (and then to have

15 an effect there). This is due to considerations of potential resistance. In an SIT-type program, the released males have been produced in an artificial rearing facility, so they and their genotype is to some extent under control, or at least variations can potentially be identified and eliminated more easily than they could be from a wild population of the same species. A biochemical effect acting in the egg (or developing zygote) may

20 potentially be altered or mitigated by changes in that egg; if these are heritable then there is at least the basis for potential heritable resistance to a toxin producing this biochemical effect. In contrast, it is difficult to see how the maternal or zygotic genotype could compensate for at least some types of damage that has already been done to the sperm before it enters the egg (or before it enters the female). One

25 example of such damage is damage to the genetic information contained by the sperm, other examples, though perhaps with less certainty about the ability of the maternal/zygotic genome to compensate, include loss or damage to essential components of early processes such as sperm membrane breakdown or sperm nucleus decondensation, for example, or centrosome function.

30 One preferred class of paternal effect lethals are nucleases. Though sperm contribute several things to the zygote, one key one is genetic information. If this genetic information is damaged to the extent that some or (preferably) substantially all of the zygotes are non-viable, then this forms the basis for a suitable form of sterility through

35 paternal-effect lethality. Radiation-sterilisation, as used for example in conventional

5 Sterile Insect Technique, is an example of conditional (in this case inducible by irradiation) paternal-effect lethality; sterilisation with chemosterilants, e.g. thiotepa, is another. Each of these approaches work by damaging the DNA in the sperm, thus degrading the genetic information that it carries to the point that many zygotes die (Robinson, 2005). Chemosterilants such as thiotepa and bisazir, for example, may leave toxic residues and are generally regarded as unacceptable for widespread use.

10 Suitable expression of a suitable nuclease in the sperm, or during spermatogenesis, will have the effect of damaging the genetic information of the gamete, or resulting gamete, without similarly damaging the genetic information of somatic cells of the male. This will reduce the degree of incapacitation of the male associated with the sterilisation process. In contrast, the use of radiation or chemosterilants typically exposes all cells approximately equally to the sterilising agent.

15 A suitable nuclease is one that cuts DNA, preferably generating double-stranded breaks. DNA-modifying enzymes which affect the storage or interpretation of genetic information, or which lead to indirect cutting or modification of the DNA, e.g. by cellular DNA repair machinery, are also here classed as nucleases; it will be understood that references to 'cutting' DNA, for example, therefore include 'modifying' DNA, as appropriate.

20 Suitable expression is conditional expression, such that under restrictive conditions (for fertility, obviously these are permissive conditions for expression of the nuclease) at least 30%, preferably more than 50% and most preferably >90% of the sperm from a male with at least one copy of the nuclease-comprising paternal-effect lethal system (a 'PAL male') are incapable of fertilising an egg to form a viable zygote capable of surviving to give a fertile adult under typical rearing conditions, e.g. as found in the wild, or laboratory conditions approximating those found in the wild. Conversely, under permissive conditions (for fertility), equivalent PAL males should give significantly more viable zygotes than under restrictive conditions; preferably at least 50% of the zygotes should survive.

35 Preferably, expression of the nuclease in males is substantially restricted to the germline. While this may mean that expression is substantially restricted to the male germline, this is not essential. For some applications, elimination of females is

desirable (e.g. genetic sexing in SIT (Alphey, 2007; Franz, 2005)). Under circumstances where females are to be eliminated anyway, expression of the nuclease in females may not be undesirable. Indeed it may be desirable, if used to eliminate, or assist in the elimination of, females. This would be an example of potential 'dual use' – using the nuclease both as a PAL and as (zygotic) dominant lethal to kill some or all of the females. Many nucleases are known. Relevant classes include restriction endonucleases and zinc finger nucleases and transcription activator-like effector (TALE) nucleases (Hockemeyer et al., 2011; Mahfouz et al., 2011; Miller et al., 2011) and homing endonucleases. Different classes and different members within a class, show different levels of sequence specificity at or near the site at which they cut. In principle, an enzyme with a very high degree of sequence specificity will cut at a relatively small number of sites per genome, perhaps only one. This is not preferred, as it allows the possibility of resistance by mutation or variation of the target site. Where a specific sequence is repeated many times, so that there are multiple sites in the genome despite a relatively high degree of specificity for a relatively long recognition sequence, this is more acceptable. However, we prefer that the nuclease has multiple target sites per genome. This means recognising a relatively long sequence that is nonetheless present in multiple copies in the genome or a relatively short recognition sequence or a significant degree of redundancy within the target sequence, of incomplete specificity for the nominal target sequence, or indeed a nuclease with little or no sequence specificity at all. I-Ppol is of this type, having a long recognition sequence but it's in a highly conserved section of rDNA, which is present in multiple copies per genome. I-Ppol is further discussed elsewhere.

Zinc finger nucleases (ZFNs) have been described, wherein each zinc finger provides sequence-specific binding to a short nucleotide sequence, e.g. 3 nucleotides. Higher affinity and greater sequence specificity can therefore be provided by combining multiple such zinc fingers into a single protein. If this is combined with a nuclease, e.g. the nuclease domain of the restriction endonuclease FokI, an artificial sequence-specific nuclease can be constructed, with arbitrary sequence specificity (Kim et al., 1996). Such synthetic zinc finger nucleases have been developed for gene therapy purposes, for example (Urnov et al., 2005). Considerable effort has gone into improving their specificity, to reduce cutting to a single site in the genome, e.g. (Miller et al., 2007). In an example in *Drosophila*, zinc-finger nucleases have been produced which, when expressed in transgenic flies, specifically damage the target locus, e.g.

the *yellow* and later also *rosy* and *bw* loci (Beumer et al., 2006; Bibikova et al., 2002). An unwanted side-effect was observed, which was that high level expression of some of the ZFNs used was toxic. It was further shown that this toxicity depended on the nuclease, rather than the DNA binding activity, of the toxic ZFNs (Beumer et al., 2006).
5 Though undesirable for gene targeting, this broader specificity is attractive for our purpose of generating damage at several or many sites.

The use of homing endonucleases (HEGs) is generally not preferred, as they tend to recognise relatively long (15-40 bp, though often accepting some mismatches to the
10 nominal target sequence) nucleotide sequences which therefore occur rarely, if at all, in any given insect genome. The minimum number of acceptable recognition/cutting sites is one per diploid genome; one per haploid genome is preferred and recognising/cutting multiple sites per haploid genome is particularly preferred. An example of a HEG likely to be inappropriate is I-SceI. Originally isolated from yeast
15 (*Saccharomyces cerevisiae*) mitochondria, I-SceI has been used in *Drosophila* as part of a system allowing or promoting targeted homologous recombination (Gong and Golic, 2003; Rong and Golic, 2000; 2001; Rong et al., 2002). One attractive feature of I-SceI for that homologous recombination system was precisely that I-SceI does *not* readily cut any endogenous sequences in the *Drosophila melanogaster* genome, and
20 will therefore tend to cut specifically at engineered sites inserted on transgenes. Though some other insect genomes may by chance have one or more sites recognised by I-SceI, this enzyme is not generally preferred for use in the present invention. One example of a homing endonuclease that cuts multiple sites per genome is Ppol. Though this has a rather specific, long recognition site, it corresponds to a highly
25 conserved sequence in an rDNA gene. Since multiple copies of this rDNA gene are present in all eukaryotic genomes, multiple target sites are available. There are potential mechanisms, such as gene conversion, whereby a 'resistant', i.e. resistant to Ppol-mediated cutting, mutation of such a gene could spread through the rDNA sequences, however there are sufficient copies of this sequence that the risk of
30 resistant genomes, where all or almost all of these sites have become resistant, seems rather low. I-SceI and Ppol have both been shown to function in mosquito (*Anopheles gambiae*) tissue culture cells (Windbichler et al., 2007); as one might expect expression of Ppol was deleterious to the cells, leading to cell proliferation arrest, whereas I-SceI was relatively innocuous. Homing endonucleases are a class of selfish DNA element
35 which can spread through populations of the fungi wherein they have been identified.

No homing endonucleases are known in animals, but it has been proposed that their unusual properties (cutting a homologous non-HEG-bearing chromosome and being copied into it by the cell's DNA repair machinery (Burt and Trivers, 2006)) could be used to construct a 'gene drive' system (Burt, 2003); proof-of-principle has been accomplished using I-SceI and an artificial target site (Windbichler et al., 2011). A gene drive system is a system for spreading genes through a target, e.g. wild, population where the gene to be spread does not confer a simple selective advantage (e.g. Alphey et al., 2002).

In addition to the 'conventional' use of HEGs as gene drive systems, which resembles their natural spreading mechanism and involves copying the HEG into cut DNA, it has also been proposed that a HEG located on the Y chromosome that specifically cuts the X chromosome might lead to males from whom the only viable gametes are Y-bearing (Deredec et al., 2008).

HEGs typically recognise a target site of 15-40bp. ZFNs recognise approximately 3bp per zinc finger. In the typical configuration, ZFNs need to dimerise to cut DNA; this means an effective recognition sequence of 18bp. A specific 18nt sequence will typically occur once in 4^{18} nucleotides, which is approx 7×10^{10} nucleotides or 7000Mb. For comparison, insect genomes are typically a few hundred Mbp, and the human genome about 3000Mbp, so such sequences will typically occur 0-1 times per genome. Of course this is a simplified calculation, and ignores the effects of GC content, repetitive DNA, etc, but the general point still holds, that the specificity of such long recognition sequences is desirable for purposes that need specificity, such as gene therapy and gene targeting, but molecules with a shorter recognition sequence are generally preferred for the purpose of cutting the genome two or more times.

One such class of enzymes are restriction endonucleases. These typically have recognition sites of 4-10bp, which will typically cut a eukaryotic genome many times (4^{10} is approximately 1×10^6). Some restriction endonucleases are sensitive to the methylation state of the substrate DNA; enzymes which are insensitive to methylation, or which cut DNA modified in a way characteristic of the target genome, are preferred. For example, *Drosophila melanogaster* genomic DNA is substantially unmethylated, therefore an enzyme such as *DpnI*, which cuts only adenomethylated DNA, is not preferred. In contrast, to cut the same sequence (GATC), alternative enzymes such as

DpnII, MboI or Sau3AI would be preferred. FokI, which is not methylation sensitive and for which the nuclease domain is known to function in a range of cell types as well as in vitro, is particularly preferred.

- 5 The nuclease need not have any substantial sequence specificity. Another preferred class of nuclease is one in which a nuclease domain, for example from FokI, is combined with a DNA binding domain, said DNA binding domain having little or no sequence specificity (though a general preference for some types of DNA over others, for example for GC-rich or AT-rich regions or sequences, is acceptable). Particularly preferred examples of this class of effector are protamine-nuclease and histone-nuclease fusions. Some of the advantages of protamine-nuclease fusions are described above. Two types of nuclease domain are envisioned. Firstly, one which requires to bind at least one other protein ('dimerise'; it will be understood that a requirement to form larger complexes, e.g. trimers, tetramers, etc are included in this term), either with itself (homodimer) or with at least one different protein (heterodimer) in order to cut DNA; secondly one which does not need to so dimerise.

The key difference between these two types is their concentration/function relationships. An enzyme that does not need to dimerise will, broadly, function in proportion to its concentration. Therefore very low concentrations will produce some DNA damage or modification; higher concentrations will produce more. This is advantageous when the conditional expression system produces relatively low amounts of effector. The time course of spermatogenesis is likely to give the effector at least some hours to act, and probably rather longer. In contrast, an enzyme that does need to dimerise will typically have a non-linear dose-response function. Particularly for an enzyme that can bind at many sites in the genome, at low concentration it is unlikely that two enzyme molecules will meet in such a way as to be able to cut. At higher concentrations, it is much more than proportionately likely (typically increasing as the square of concentration for homodimers, cube for homotrimers, etc). This may be advantageous where, for example, the conditional expression system is somewhat leaky, producing a low, non-zero level of effector in at least some cells other than the intended cells (e.g. other than in the male germline, or, as another example, in male germline cells where intended expression is in spermatogonia or later stages of spermatogenesis only). Then this low 'basal'

expression should be relatively harmless, or at least show a more-than-linear difference in its activity in the non-target cell relative to its activity in the target cell(s).

5 Another relevant type of leaky activity relates to the conditionality of the conditional expression system – it is unlikely that the supposedly ‘off’ condition will in fact give zero expression, therefore an additional system, such as this, which exaggerates the phenotypic difference between the ‘on’ and ‘off’ states is desirable. Note that these arguments and potential benefits apply to other classes of effector. Within nuclease effectors, they apply to zinc finger nucleases as well as to protamine-nucleases, etc. In particular, the FokI nuclease domain needs to dimerise to cut DNA. Normally this is by homodimerisation. However, variants are known which alter this. The two nuclease domains can be included in the same protein, connected by a suitable linker, so that dimerisation is not required (i.e. intramolecular, rather than intermolecular dimerisation is sufficient for). Also, mutants of FokI are known where heterodimerisation is required (Miller et al., 2007). Note that it is not required that each of the two dimerisation domains have nuclease activity. This permits another method of achieving greater specificity. The two (or more) proteins which need to heterodimerise may be expressed under separate control (e.g. transcriptional control, but also translational or degradation control, for example). Only cells expressing both domains will have significant activity. Therefore, leaky expression of one protein in a cell that does not also suffer leaky expression of the other protein will have little or no adverse effect.

10

15

20

EXAMPLES

Example 1 : *Germ line specific promoters*

5

In order to test the suitability of an isolated promoter region, a fluorescent protein can be used as a reporter, and particularly a fusion between a tetracycline-responsive transcriptional activator (tTA) and a fluorescent protein. This allows both direct visualisation of protein expression, and also further testing of the suitability of the timing and level of expression directed by the candidate suitable promoter fragment for the specific purpose of use as a part of a conditional bipartite expression system functional in spermatogenesis. Expression of the reporter from such insertions illustrates the activity of the candidate suitable promoter fragment. Assessment of several such insertion lines allows a determination as to the likely suitability of the promoter fragment. Several insertion lines should be examined due to 'position effect'; a well-known phenomenon whereby the expression of an inserted gene is influenced by the chromatin context into which it is inserted (Wilson et al., 1990).

The actual ability of the candidate suitable promoter fragment to function as part of a conditional expression system in the male germline can be tested as above but including a suitable expression system, or relevant component thereof, e.g. encoding tTA or a related sequence. Use of a tTA-fluorescent protein (tTA-FP) fusion in the reporter test above allows the same insertion line, or set of insertion lines, to be used for both tests. Crossing to a suitable line, e.g. tRE-X, where X may be an effector or reporter, under conditions permissive for expression from the expression system, will allow a determination of whether the expression system (promoter-tTA, tRE-X) is functional in spermatogenesis and, e.g. by fluorescence microscopy on dissected testes, where, when and how much X and/or tTA-FP is expressed. Similar experiments under conditions repressive for expression will allow a determination as to whether the expression system is conditional.

If expression studies indicate that the candidate suitable promoter fragment is *not* in fact suitable, one can take a different promoter fragment from the same gene. Such a revised fragment is typically longer, as the original fragment may have omitted some important regulatory elements, however in the case of expression being specific but too

late in development, e.g. post-meiotic, even though the gene from which the promoter fragment was derived is thought (e.g. based on RNA *in situ* data) to be transcribed sufficiently early, it may be desirable to eliminate some elements, especially UTR elements, to remove signals for translational delay. It may also be desirable to test longer and shorter fragments to identify regions of the promoter necessary for or influencing gene expression levels and temporal and tissue specificity; this approach has been widely used to analyse promoters and can be achieved simply by iterations of the above procedure with variants of the original candidate promoter fragment.

RNA in situ hybridisation experiments demonstrated that that the Medfly *β2-tubulin* transcription is at an early-to-late stage in Medfly spermatogenesis. It is assumed that the Medfly *β2-tubulin* transcript starts at a similar stage as *Dmβ2-tubulin*, which is before meiosis and is synthesised from the late third instar larval stage, before there is significant meiosis in the developing testes. Sperm marking systems using this gene promoter have been developed for the mosquitoes *Anopheles stephensi* (Catteruccia, Benton et al. 2005) and *Ae. aegypti* (Smith, Walter et al. 2007), for Medfly (Scolari, Schetelig et al. 2008) during the period of our research towards sperm lethality and later the Caribbean fruit fly, *Anastrepha suspensa* (Zimowska, Nirmala et al., 2009). According to this, we have amplified the *β2-tubulin* promoter from the medfly genome using the following primers:

Forward: CTCCCGTGCGATATCCTAGGCCCCCATGTTACAAGGCTG (SEQ ID NO: 53)

Reverse: AGCCATTTTGGTTAATTGAAATCCCTAAAATAAATGTAATTCATTTTTCG (SEQ ID NO: 54)

Construct OX3671 (Figure 16) contains the Medfly *β2-tubulin* promoter (the 1556bp promoter fragment was amplified from Medfly genomic DNA) fused to the DsRed2 sequence. Most of the *Ccβ2-tubulin* 3'-UTR sequence was omitted and replaced by a commonly used 3'-UTR - SV40 –known to express in a variety of species. Expression of the transgene should result in sperm fluorescing red under the appropriate excitation filter. DsRed2 expression was clearly visible in the abdomen of transgenic males in all three lines obtained, with two areas of more intense fluorescence, seemingly the testes (data not shown).

To further assess whether the DsRed2 marker expresses in sperm, sexually mature transformed male's testis were dissected. Results demonstrate that mature sperm exhibits strong DsRed2 expression compared with sperm from a wild type male. In alignment with B2-tubulin expression pattern, early spermatocytes do not exhibit fluorescence, which starts to be clearly visible in the spermatid stages of spermatogenesis.

In order to test whether we would obtain similar patterns of fluorescence expression using the bipartite "tet-off system", two fluorescent reporter constructs were built; OX3866 (Figure 14) (*tetO*-mini promoter-DsRed2-mls) and OX3867 (Figure 15) (*tetO*-mini promoter-protamine fused with DsRed2). The *tetO*-mini promoter element is known as tRE and is preferred. In OX3866 (Figure 14) the DsRed2 reporter is fused to a membrane localisation signal (mls) which is predicted to lead to any expressed fluorescent protein being membrane-bound or -associated after expression, whereas in OX3867 (Figure 15) the DsRed2 is fused to a *Drosophila* protamine. Both reporter genes were tested with promoters driving expression of tTAV in somatic cells and found to be functional.

Following the finding that the $\beta 2$ -tubulin promoter drives expression of DsRed2 in sperm, albeit too late for our purposes, construct OX3831 (Figure 20) was designed and modified from OX3671 (Figure 16), with Medfly $\beta 2$ -tubulin promoter driving tTAV-fluorescence fusion gene. In this construct, the *Drosophila aly* 5' UTR region replaced that of $\beta 2$ -tubulin promoter under the speculation that tTAV protein would then express earlier during spermatogenesis. Since the TurboGFP sequence was fused with the tTAV, sperm of OX3831 (Figure 20) transgenic males should exhibit green fluorescence under the appropriate excitation filter. To test this, adult males from seven transgenic lines were dissected and observed under a fluorescence microscope. Although fluorescence expression was not readily visible in the testes of non-dissected males, in dissected testes of OX3831 (Figure 20) adult males fluorescence expression was visible in the spermatid bundles. Some lines expressed stronger fluorescence than others but all spermatids in OX3831 (Figure 20) strains displayed fluorescence testis-specific sperm marker expression.

OX3831-heterozygous flies were crossed with OX3866- and OX3867-heterozygous flies (Figure 14 and 15 respectively) (*tetO*-DsRed2 strains). More than 20 adult male

progeny that had been raised on non-tetracycline food (i.e. permissive conditions for tetO-DsRed2 expression) in different crosses and at different times were dissected and observed under a high magnification fluorescence microscope. Only one male showed strong DsRed2 expression, which was visible in several spermatid bundles. In other males, no DsRed2 expression in any spermatid bundles was detected. A possible reason for only one male demonstrating strong DsRed2 expression is late transcription and translation of tTAV-fluorescence, driven by the *Ccβ2-tubulin* promoter. This late expression may mean that there is insufficient time for tTAV to bind to the *tetO* sequence and further induce enough DsRed2 transcripts before meiosis. According to our results, only by chance, some of the spermatocytes in some males have enough DsRed2 transcripts accumulated before meiosis and DsRed2 fluorescence can be detected in spermatid bundles. Reverse transcriptase PCR analysis on isolated testes and carcasses of OX3831 (Figure 20) males, using tTAV specific primers, demonstrated the testis-specificity of the 5' UTR-*aly Ccβ2-tubulin* promoter.

It was apparent at this stage that an even earlier expression of tTAV was necessary in order to allow for adequate expression of the reporter gene in the male germ line. For this reason, we developed and tested construct OX4282 which contained the 5' UTR of the *Hsp83* gene. *Hsp83* is expressed strongly both in germline and somatic cells of the Mediterranean fruit fly *Ceratitis capitata* and is not considered to contain any delayed-translation signals. In construct OX4282, the reporter gene; TurboGFP, was not fused to the tTAV sequence but placed adjacent to the tetO sequence. Furthermore, the tetO-reporter and promoter-tTAV components were combined in a single plasmid in an attempt to assess the *Hsp83* 5' UTR at a more immediate fashion. tTAV should be expressed in the male germline of transformed individuals and, in the absence of tetracycline, by binding to *tetO* should induce expression of the adjacent TurboGFP marker gene. In other words, male testes of those strains carrying this construct should show TurboGFP expression when reared off tetracycline, and expression should be repressed by tetracycline. Testes dissection of adult males reared in the absence of tetracycline, revealed the presence of strong green (turboGFP) fluorescence in three out of 6 lines tested.

The fact that TurboGFP expression was not observed in the other lines is possibly due to positional effects of the transgene insertion. In the lines exhibiting TurboGFP

fluorescence off tetracycline, expression was totally repressed by tetracycline (data not shown).

5 OX4282 males were crossed with OX3867 (tetO-protamine-DsRed2) (Figure 15) females in each cage in order to further examine the ability of the promoter to drive adequate expression of the reporter gene; in this case DsRed2 using the “tet-off” bipartite conditional system.

10 TurboGFP expression can be detected at different stages of spermatogenesis (in both elongated spermatid and spermatocytes, although red fluorescence was only detected in elongated spermatids but not in early spermatocytes (Figure 3). TurboGFP expression was detected in both the sperm head (where nucleus is located) and tail of the spermatids.

15 The above results indicate that there is a slight delay in the expression of the reporter gene (DsRed2 in this experiment) compared to tTAV expression; as this is estimated by the amount of green fluorescence observed. Considering that red fluorescence was apparent in later stages of spermatogenesis (elongated spermatids), it is likely that the reporter is transcribed before meiosis but translated after meiosis.

20

Dmtopi is a testis-specific gene which encodes a testis-specific Zn-finger protein that physically interacts with *Comr* (Perezgasga, Jiang et al. 2004). *Dmtopi* is not required for the nuclear localisation of *Aly* or *Comr*, but is required for their accumulation on chromatin. In *Drosophila*, although all genes that depend on *aly* or *comr* for expression
25 also depend on *achi/vis* and/or *topi*, there are a few genes, whose transcription depends on *achi/vis* and *topi* but not on *aly* or *comr*. (Perezgasga, Jiang et al. 2004). Perhaps more significantly, many of the *aly*-class genes seem to have arisen from gene duplications. Following duplication, one of the pair has assumed a somatic role and one a germline role. This has two limitations in respect of using these genes as a
30 source of male germline promoters in a wide range of insects.

Firstly, the two genes may be quite similar, making it relatively difficult to identify the germline-specific version. Secondly, and much more significantly, many of these duplications appear to be quite recent. Most insects may therefore have the ancestral
35 version, which is a single gene performing both germline and somatic functions, and

hence no separate germline gene and promoter. *topi* is unusual in this respect in that it has no obvious somatic alternative, and it also seems more ancient than most of the aly-class duplications. It is therefore likely to provide a potential germline-specific promoter in a much wider range of insects than the other aly-class genes. One caveat for this is that the sequence conservation of *topi* is clear, but the functional conservation, and in particular the expression pattern, of *topi* in other insects is generally unknown. However, this reservation applies also to other male-germline genes. For these various reasons, *topi* was chosen for further investigation as a source of a male-germline promoter.

Before developing *topi*-based constructs in Medfly, the conservation of expression pattern of *topi* was first tested in the mosquito *Ae. aegypti*; availability of genome sequence meant that the *topi* homologue and putative promoter fragment could rapidly be isolated. RISH (RNA In Situ Hybridisation) results showed that *Ae. aegypti topi* (*Aetopi*) gene transcription starts from the primary spermatocyte stage; an expression profile that resembles that of *Dmtopi* in *Drosophila* testes (Perezgasga, Jiang et al. 2004). An *Ae. aegypti* transgenic strain (OX4286, Figure 23), with tTA expression driven by a 1233-bp sequence, including 1168 bp of *Aetopi* promoter and 65 bp *Aetopi* 5'-UTR), was developed. After crossing 3 independent OX4286 (Figure 23) lines with a tetO-reporter strain (OX3978, tetOAehsp70 mini promoter-Amcyan, Figure 24), clear Amcyan expression was seen in testes and spermatogenesis cells in all three crosses rearing on non-TET diet (data not shown). Expression was tetracycline-repressible.

After the *Ae. aegypti topi* promoter was shown to induce testis-specific expression of tTAV and, consequently, expression of a reporter gene, the expression pattern of *Ceratitis capitata topi* was analysed. Using the nucleotide sequence of this gene in Medfly, primers were designed to verify that *Cctopi* is a testis-specific gene in Medfly. Ten pairs of testes dissected from ten wild-type males were used to extract RNA (for dissection details see section 2.6.4.3). RNA was extracted from the remaining carcasses to provide non-testes controls. Reverse transcriptase analysis using the primers TopitestF1' and 'TopitestR' were used for these PCRs. The results confirmed that *Cctopi* expression is testis-specific. RISH also demonstrated that *Cctopi* is transcribed in Medfly testes (data not shown). This gene promoter was therefore chosen for use in new constructs for developing testes-specific expression in transgenic strains.

TopitestF1 primer: GTAAC TCCCGTTCCTGAGACAACA (SEQ ID NO:55)

TopitestR primer: CGATATGGAGTGGGTGAAACCTCA (SEQ ID NO: 56)

- 5 Construct OX4275 (Figure 25) comprises a putative promoter fragment (1178 bp) from *Cctopi* driving DsRed2 (SV40 3'UTR). Dissected testes from adult males of all strains obtained with this construct did not reveal any signs of DsRed2 expression (data not shown). To further test the *Cctopi* promoter, a construct with the same promoter sequence driving tTAV expression was developed (OX4254, Figure 26). OX4254-
- 10 heterozygous flies were crossed with OX3867 (tetO-DsRed2) (Figure 15) heterozygous flies (tetO-DsRed2 strains) of the opposite sex and the adult male progeny of these crosses were dissected and assessed for red fluorescence. No DsRed2 expression in any spermatid bundles was detected in those male testes.
- 15 As the short length of the putative *Cctopi* promoter in OX4275 (Figure 25) and OX4254 (Figure 26) may have resulted in a lack of apparent function, a new construct - OX4371 (Figure 27) – was made based on the following modifications. This construct contains a 1708-bp sequence from the putative *Cctopi* coding region, over 530bp more than in previous constructs including a 484bp possible coding sequence and a 55bp intron in
- 20 the retained portion of the coding region. As tTAV generally does not function after fusion with another protein at its N-terminus, a 228bp ubiquitin gene sequence (sequence based on *Drosophila* ubiquitin but optimised for expression in a range of insects and synthesised by Geneart Ltd) was used between the *Cctopi* coding region and tTAV to separate the two gene products post-translationally via cleavage by the
- 25 ubiquitin protease (Varshavsky 2005). This construct contains both the newly designed *Cctopi* promoter driving tTAV and a tetO-Dmhsp70 mini promoter-TurboGFP reporter. Therefore, expression of tTAV should be detected by fluorescence in testes of males reared without tetracycline in their larval diet. Weak TurboGFP expression in spermatogenesis cells was detected in dissected testes from OX4371 (Figure 27) male
- 30 adults reared off tetracycline in one of the two lines tested. No TurboGFP expression was detected in dissected testes of any males reared on tetracycline, indicating repressible expression.
- 35 Concurrent to the work with OX4371 (Figure 27), OX4391 (Figure 22) strains were generated and analysed.

OX4391 (Figure 22) differs from OX4371 (Figure 27) in one aspect: the SV40 3' UTR of tTAV in OX4371 (Figure 27) was replaced by the *Cctopi* endogenous 3' UTR in OX4391 (Figure 22). As mentioned before, 3'-UTRs can influence the fate of a particular mRNA, for example transcript stability or level of translation (Mazumder, Seshadri et al. 2003). Considering that the 3' UTR has been shown to play an important role in mRNA processing, we hypothesised that the endogenous 3' UTR from *Cctopi* may confer the gene's desired expression patterns to our tTAV transgene. Reverse transcription PCR was used to amplify the *Cctopi* 3'-UTR. Testes from 2-3-day-old male adults reared off tetracycline from 6 OX4391 (Figure 22) strains were dissected. 3 strains showed strong TurboGFP expression in male germ line cells. Weak fluorescence was detected in the other three.

To further test the newly designed *topi* promoter sequence, we set up crosses with OX4391 (Figure 22) males (reared on non-tetracycline diet) with wild type females and assessed the presence of fluorescence in dissected spermathecae. Examination of sperm stored in spermathecae under a fluorescence microscope demonstrated that TurboGFP was indeed detectable.

The above results indicate that our isolated *topi* promoter sequence (as described above) is expressed in the male germ line and adequately drives expression of tetO reporter genes in testes. Crosses with flies comprising tetO-nuclease transgenes are discussed later.

Example 2 Effector proteins

The objective of this work is to produce sperm that are transferred to the female and will lead to a low or indeed none of the embryos -which the female would otherwise have produced- surviving. Same sperm should induce adequate refractoriness to remating in the female, while will do well in sperm competition if the female does remate. Our approach to this is to construct paternal-effect lethals, whereby the sperm can enter an egg but no viable zygote (capable of developing to a fertile adult) is formed. Ideally, this effect is generated by males with a single copy of the paternal-effect lethal, though the use of multiple copies is also envisioned. It is also preferred that the effector have a direct biochemical effect on the sperm, rather than merely

using the sperm as a vehicle via which to enter the egg (and then to have an effect there). This is due to considerations of potential resistance. Nucleases are a preferable option for the purposes of this invention. Theoretically, if the genetic information carried by the sperm is damaged to the extent that some or (preferably) substantially all of the zygotes are non-viable, then this forms the basis for a suitable form of sterility through paternal-effect lethality. Different classes of nucleases have been explored in an attempt to induce sperm specific damage. All nucleases were tested as part of the "tet-off" bipartite conditional system; that is linked to a *tetO* sequence. This approach allows for the assessment of various effector proteins in combination with different germline specific promoter sequences, without the burden of creating new transformant strains. Furthermore, it provides a more realistic situation in terms of future application.

Zinc finger nucleases (ZFNs) have been described, wherein each zinc finger provides sequence-specific binding to a short nucleotide sequence, e.g. 3 nucleotides. Higher affinity and greater sequence specificity can be provided by combining multiple such zinc fingers into a single protein. If this is combined with a nuclease, e.g. the nuclease domain of the restriction endonuclease FokI, an artificial sequence-specific nuclease can be constructed, with arbitrary sequence specificity. We have tested the hypothesis that ZFNs can produce DNA breaks in elongated spermatids by crossing these lines to various male germ line specific promoters. Two (OX4103) (Figure 18) and three (OX4104, Figure 19) Zinc finger nuclease constructs were designed and tested. Previous assessment, by crossing to promoter-tTAV lines expressing in somatic tissues, indicated that the 3-Zn finger nuclease exhibits a greater "lethal" effect and the strongest of these lines was therefore used for the purposes of this work. This nuclease is therefore preferred.

Homing endonucleases are a type of restriction enzymes typically encoded by introns. They act on the cellular DNA of the cells that synthesize them and they tend to recognise relatively long (15-40 bp, though often accepting some mismatches to the nominal target sequence) nucleotide sequences which therefore occur rarely, if at all, in any given insect genome. The minimum number of acceptable recognition/cutting sites is one per diploid genome; one per haploid genome is preferred and recognising/cutting multiple sites per haploid genome is particularly preferred. One example of a homing endonuclease that cuts multiple sites per genome is IPpol.

Though this has a rather specific, long recognition site, it corresponds to a highly conserved sequence in an rDNA gene. Since multiple copies of this rDNA gene are present in all eukaryotic genomes, multiple target sites are available. A tetO-IPPO1 (OX4112, Figure 17) construct was designed and tested.

5

Restriction endonucleases have recognition sites of 4-10bp, which will typically cut a eukaryotic genome many times. Such nucleases have no substantial sequence specificity. FokI, which is not methylation sensitive and for which the nuclease domain is known to function in a range of cell types as well as in vitro, is particularly preferred for our aim. FokI nuclease domain combined with a DNA binding domain of little or no sequence specificity is of particular interest. Preferred examples of this class of effector are protamine-nuclease fusions. An advantage of the protamine-nuclease fusion is the need for dimerization (or polymerisation) either with itself (homodimer) or with at least one different protein (heterodimer) in order to cut DNA. An enzyme that does need to dimerise will typically have a non-linear dose-response function. Particularly for an enzyme that can bind at many sites in the genome, at low concentration it is unlikely that two enzyme molecules will meet in such a way as to be able to cut. This may be advantageous where, the conditional expression system is leaky –either through the promoter or through the conditional system itself, producing a low, non-zero level of effector in at least some cells other than the intended cells (e.g. other than in the male germline). A tetO-protamine-Fok1 construct (OX4458, Figure 21) has been designed and tested.

10

15

20

25

30

35

Assessment of male sterility in the bipartite system mentioned before is referred to as “Egg hatch rate assays or experiments”. A schematic representation of the design is illustrated in Figure 1. Promoter lines drive expression of tTAV in the male germline, preferably with little or no expression elsewhere, while effector lines are linked to the *tetO* sequence, expression of which is correlated to tTAV, both temporally and spatially. Transformation markers in promoter and effector lines utilize different fluorescent genes for accurate evaluation of the results, i.e. promoter lines fluoresce green, while effector lines fluoresce red under appropriate excitation filters. Effector (E) and Promoter (P) lines are crossed without tetracycline, i.e. permissive (for expression) conditions. Eggs from these crosses are collected and divided into either permissive conditions (without tetracycline) or repressive conditions (with tetracycline) and reared accordingly. Pupae are screened for expression of line-specific fluorescent markers

and double heterozygotes – with both markers – collected. These are checked for equal male to female ratio after eclosion and males crossed to wild type females – again either on- or off-tet. Double-heterozygote females to wild type males crosses are used to test whether any observed effects are sex specific. Wild type control is included as the reference point for egg hatch rates. Eggs from the crosses are collected and counted. Three days later counting is repeated and numbers of un-hatched eggs assessed.

Figure 1. : Design of the egg hatch rate assay.

Results of a number of these crosses are shown below. All crosses presented here have followed the basic design described above.

OX4282 (*PB-HrIE-AmCyan-SV40-TurboGFP-teto14-ccTubulin-hsp83-tTAV-SV40*) x **OX4104** (*PB-YAFN-hsp83-tetO21-Hr5-IE1-Red*) (Figure 19)

15

Figure 2. OX4282-OX4104 male sterility on and off Tetracycline. Three independent, autosomal insertion lines of OX4282 transposon carrying tetracycline repressible transactivator (tTAV) driven by the β 2-tubulin promoter from *Ceratitis capitata* gene (I, L, G) were crossed to line OX4014 carrying a tetO-3Zn-finger effector. Progeny of these crosses was reared and bred either on a diet with (100 μ g/ml; tet+) or without tetracycline (tet-). Males carrying both driver and effector alleles were crossed to the wild type females and the hatching rates of eggs obtained from these crosses were calculated (percentage of laid eggs that hatched). Two different egg collections of 100-150 eggs each were used for each cross. Wild type males crossed to wild type females in the presence or absence of tetracycline were used as controls. Crosses where highly significant male sterility was observed (chi-squared test, * represents $P < 0.01$, ** represents $P < 0.001$) are marked with asterisks.

Three OX4282 lines were crossed to a 3Zn-finger line (OX4104, Figure 19) which has previously shown promising results when crossed to two generic promoters (Hsp83 and OP). No adverse effects were recorded for flies containing both plasmids, indicating that basal expression of the effector in somatic tissue was not high enough to show an adverse effect. Only 14% egg hatching was observed from line L and 27% from line I. There was no significant reduction in egg hatching from line G. Both lines L and I contain a single copy of the transgene in contrast to line G, which contains two

35

(or more) copies. Results clearly indicate that the 5' Hsp83 UTR-Cctubulin-SV40 3' UTR promoter drives adequate expression of tTAV in the male germ line to induce sperm sterility, as indicated by the reduction of hatching larvae from the resulting progeny of these males. Additionally, no reduction in female fertility was observed indicating that the promoter acts in a male-germline specific fashion (data not shown). However, the fact that there was no significant reduction in the progeny of one OX4282 line points toward positional effects influencing the phenotype.

OX4282 (*PB-HrIE-AmCyan-SV40-TurboGFP-teto14-ccTubulin-hsp83-tTAV-SV40*) x
OX4112 (*PB-IPPO-1-hsp83-tetO21-Hr5-IE1-Red*)

The same OX4282 lines were crossed to an I-Ppo1 line, which had demonstrated lethality when crossed to two promoters (Hsp83 and Opie2) active in somatic cells. Up to 50% sterility was observed compared to wild type and tetracycline controls, using this effector (data not shown). Results demonstrate that the sterility seen in this assay is the result of the activation of nuclease expression by tTAV and subsequent sperm DNA cleavage. Sterility levels indicate not enough penetrance. The fact that the same line exhibited a strong lethal effect when it was crossed to generic promoter lines may suggest that not enough molecules of this protein were produced in the male germline to cause a desirable effect. It is probable that the protein needs to exceed a certain threshold for a strong effect to occur which also increases the possibility of positional effects of the transgene influencing performance. This, together with the reduced numbers of pupae containing both transgenes that were observed during this assay, make I-Ppo1 a less appropriate, and therefore less preferred, candidate for developing sperm lethality.

OX4282 (*PB-HrIE-AmCyan-SV40-TurboGFP-teto14-ccTubulin-hsp83-tTAV-SV40*) x
OX4458 (*PB-AttP-Hr5-IE1-DsRed2-SV40-teto21-Dmprotamine-nuclease*)

Figure 3. OX4282-OX4458 male sterility on and off Tetracycline - Repressible male-specific sterility in transgenic Medflies.

Four independent, autosomal insertion lines of OX4458 (Figure 21) transposon carrying a tetO-Protamine-FokI effector (B1, D1, D2,F2) were crossed to the OX4282L driver line carrying the tetracycline repressible transactivator (tTAV) driven by a promoter from *Ceratitis capitata* β 2-tubulin gene. Progeny of these crosses was reared and bred either on a diet with (100 μ g/ml; tet+) or without tetracycline (tet-). Males carrying both driver and effector alleles were crossed to the wild type females and the hatching rates of eggs obtained from these crosses were calculated (percentage of laid eggs that hatched). Four different egg collections of 200-500 eggs each were used for each cross. Wild type and OX4282L male crosses with wild type females in the presence or absence of tetracycline were used as controls. Crosses where highly significant male sterility was observed (chi-squared test, $P < 0.0001$) are marked with asterisks.

OX4458 (Figure 21) construct contains a single FokI cleavage domain fused to a *Drosophila* protamine under the transcriptional control of the tetO operator in a single ended *piggyBac*-derived vector with hr5-IE1-DsRed2 as a transformation marker. The OX4458 (Figure 21) construct should not exert any effect on its own. Expression of the effector fusion protein occurs when OX4458 (Figure 21) line is crossed to suitable tTAV expressing line, in double heterozygotic progeny, possessing both alleles, and in permissive conditions (without tTAV repressor – tetracycline).

Four lines with single, autosomal transgene insertion were crossed to the tTAV expressing line – OX4282L which exhibited higher promoter activity in previous experiments (see above). In all four double-heterozygote combinations a severe reduction was observed in the hatch rate of eggs from females mating transgenic males. The sex-specificity of the observed effect was confirmed using transgenic females carrying both transgenes crossed to wild type males (data not shown).

According to the results shown above, the combination of the “altered” tubulin promoter and the *Drosophila* protamine-FokI effector, seem to produce a male germ line specific lethal effect with minimal (or absent) adverse effects on the general fitness of the males in any other way than that tested. Additionally, females do not seem to be significantly affected by the expression of the transgenes, supporting the male germ line specificity

of the sequences used. The next step would be to design a construct that will contain both promoter and effector sequences in a single plasmid (OX4353). Such a construct would provide a more realistic situation of the transposon to be used for the development of a paternal-effect lethal system.

5

OX4353 (PB-HrIE-AmCyan-SV40-teto14-Dmprotamine-nuclease-ccTubulin-hsp83-tTAVnew-SV40)

4 lines were considered to be single insertion events apart from line F which was thought to have two insertions, based on the inheritance pattern of the transformation marker. All lines were tested by crosses to wild type in the presence and absence of tetracycline as described before. Females from the same lines were also crossed to wild type males and their fertility was assessed in a similar way.

15 **Figure 4. Percentage of OX4353 male sterility on and off Tetracycline.**

Five independent, autosomal insertion lines of OX4353 transposon carrying tetO-Protamine-FokI effector and tetracycline repressible transactivator (tTAV) driven by a β 2-tubulin promoter from *Ceratitis capitata* were generated in Medflies (A, B, C, D, F). Progeny of these lines was reared and bred either on a diet with (100 μ g/ml; tet+) or without tetracycline (tet-). Males were crossed to the wild type females and the hatching rates of eggs obtained from these crosses were calculated (percentage of laid eggs that hatched). Two egg collections of 100-150 eggs were used for each cross. Wild type male crossed to wild type females in the presence or absence of tetracycline were used as controls. Crosses where highly significant male sterility was observed (chi-squared test, ** represents $P < 0.001$, *** represents $P < 0.0001$) are marked with asterisks.

Figure 5. Percentage of OX4353 female sterility on and off Tetracycline. Five independent, autosomal insertion lines of OX4353 transposon carrying tetO-Protamine-FokI effector and tetracycline repressible transactivator (tTAV) driven by the β 2-tubulin promoter from *Ceratitis capitata* were generated in Medflies (A, B, C, D, F). Progeny of these lines was reared and bred either on a diet with (100 μ g/ml; tet+) or without tetracycline (tet-). Females were crossed to wild type males and the hatching rates of eggs obtained from these crosses were calculated (percentage of laid eggs that hatched). Two egg collections of 100-150 eggs from each cross were used. Wild type

males crossed to wild type females in the presence or absence of tetracycline were used as controls. No significant female sterility was observed (chi-squared test, $P > 0.05$).

5 Results indicate that heterozygous OX4353 males were up to 100% sterile in the absence of tetracycline (permissive conditions) in the diet. However males of some lines were sub-fertile even in the presence of tetracycline and females from most lines showed a slight reduction of fertility in the absence of tetracycline. From the above results, we can conclude that expression of the transgene is operative in all 5 lines
10 tested; in terms of inducing sperm lethality, however the fine tuning of achieving complete (or very close) sperm lethality with minimal effects on the general fitness of the insects that contain that transgene, seems to be greatly influenced by position effects (i.e. the location of the inserted sequences in the insect's genome and the regulatory elements in proximity).

15

The work described here shows that a 1030bp Medfly $\beta 2$ -tubulin promoter fragment is enough to drive tTAV expression and further drive effector gene expression in spermatogenesis under the control of tetracycline. The OX4353 strain proved to be a functional, tetracycline-repressible sperm-lethal strain. Additionally, one can also
20 conclude that the *Dm*protamine retains at least one of its key properties when expressed in Medfly; that is binding to sperm DNA. The protamine sequence is not well conserved, thus the positive outcome was uncertain prior to these experiments. These results indicate promise for further use of this promoter/effector combination in relation to the bipartite "tet-off" system, with a view for Medfly population control.

25

OX4718 (PB4 Hrie1-AmC-MexMActPro-DsR-tetO21-Prota-mCh-FokI-CcBTubPro-tTAV2-Hrie1-ZsG)

30

OX4718 is an example of a single construct carrying both promoter and effector components, on Medfly. This plasmid was injected into pre-blastoderm *Ceratitis capitata* embryos. Pupae expressing both red and green fluorescent protein (4-ended pb construct containing different coloured pb ends (Dafaala et al., 2006) – signifying insertion of complete transposon – were found among G1 progeny in two Go crosses: V (1 pupa) and σ (47 pupae). The single pupa from OX4718-V line did not survive to
35 adulthood. Pupae from OX4718- σ line displayed two clearly distinct phenotypes:

stronger (30 pupae) and weaker (17 pupae) and were named $\sigma 1$ and $\sigma 2$ respectively. These were propagated, mostly as single male or female crosses to wild type and at this stage are considered two independent insertion events. OX4718- $\sigma 1(b)$ and OX4718- $\sigma 2(b)$ lines were discontinued at G2 stage. Phenotypic analysis of G2 pupae indicated multiple insertions in each of these lines. Further analysis of lines at the G3 stage confirmed that both $\sigma 1(a)$ and $\sigma 1(c)$ lines are single, autosomal insertions. OX4718- $\sigma 2(a)$ carries a single insertion on X chromosome – as suggested by the alternating absence or presence of the transgene in males of different generations and has not been further analysed for the purposes of this study.

OX4718- $\sigma 1(a)$ and OX4718- $\sigma 1(c)$ lines were reared and bred either on a diet with (100 $\mu\text{g/ml}$; tet+) or without tetracycline (tet-) and were crossed to wild type. Wild type to wild type and OX4718 female crosses with wild type males in the presence or absence of tetracycline were used as controls. Fresh – not older than 24 hours – eggs from these crosses were collected on day 4 after setting up cages. Three collections per cross / cage were performed. Total number of eggs was compared with the number of eggs that had failed to hatch after four days. Hatching rates were calculated as the mean percentage of laid eggs that hatched; these data showed significant sterility of OX4718- $\sigma 1$ males in the absence of tetracycline. Results are shown on **Figure 6**. OX4718- $\sigma 1(a)$ and X4718- $\sigma 1(c)$ lines reared in the presence or the absence of tetracycline were crossed to the wild type females and the hatching rates of eggs obtained from these crosses were calculated (percentage of laid eggs that hatched). Wild type to wild type and OX4718 female crosses with wild type males were used as controls. Crosses where highly significant male sterility was observed (chi-squared test, $P < 0.0001$) are marked with asterisks.

The above results demonstrate successful development of conditional male-specific sterility in *Ceratitis capitata* in a format suitable for field use; i.e. individual promoter and effector molecules were assembled in a single construct.

Figures 7 and 8 Percentage of OX4705 male and female sterility on and off Tetracycline in Olive Fly.

OX4705 (PBMexMActPro-DsR-tetO21-Prota-mCh-FokI-CcBTubPro-tTAV2)

This provides another example of a single construct containing both promoter and effector components, in this case in olive fly.

- 5 Based on the encouraging results obtained in Medfly (*Ceratitis capitata*), similar plasmids were developed for a relative Tephritid; *Bactrocera oleae*, commonly referred to as olive fly. Plasmid OX4705 incorporates the altered form of B₂ tubulin driving expression of tTAV in the male germline and subsequently that of tetO and DroProtamine-FokI fusion effector. The plasmid also incorporates a novel Mexfly
10 (*Anastrepha ludens*) muscle actin promoter which drives the expression of DsRed2 fluorescent protein as a transformation marker.

Micro-injection of plasmid OX4705 on olive fly pre-blastoderm embryos generated 10 independent insertion events. All strains exhibited a strong fluorescent phenotype
15 (fluorescent marker expression) under a microscope with the appropriate excitation filter and were single insertions according to Mendelian laws of inheritance. Nine insertions were autosomal while one was on the X chromosome. All strains were tested for male and female sterility in the presence and absence of tetracycline from the larval medium. Wild type crosses provided additional control. Results are shown on Figures 7
20 and 8.

Figure 7 shows the percentage of OX4705 olive fly male sterility on and off tetracycline. 9 independent, autosomal insertion lines of OX4605 transposon carrying tetO-Protamine-FokI effector and tetracycline repressible transactivator (tTAV) driven by an
25 altered form of β 2-tubulin promoter from *Ceratitis capitata* were generated in olive fly (A, A1, A2, A3, B, B1, F, F1, P). Progeny of these lines was reared and bred either on a diet with (100 μ g/ml; tet+) or without tetracycline (tet-). Males were crossed to wild type females and the hatching rates of eggs obtained from these crosses were calculated (percentage of laid eggs that hatched). Three egg collections of 100-150
30 eggs were used for each cross to provide statistical significance. Wild type males crossed to wild type females in the presence or absence of tetracycline were used as controls.

Figure 8 shows the percentage of OX4705 olive fly female sterility on and off
35 tetracycline. 9 independent, autosomal insertion of OX4705 transposon carrying tetO-

Protamine-FokI effector and tetracycline repressible transactivator (tTAV) driven by the β 2-tubulin promoter from *Ceratitis capitata* were generated in olive fly (A, A1, A2, A3, B, B1, F, F1, P). Progeny of these lines was reared and bred either on a diet with (100 μ g/ml; tet+) or without tetracycline (tet-). Females were crossed to wild type males and the hatching rates of eggs obtained from these crosses were calculated (percentage of laid eggs that hatched). Two egg collections of 100-150 eggs from each cross were used. Wild type males crossed to wild type females in the presence or absence of tetracycline were used as controls. No significant female sterility was observed.

Data strongly indicates formidable OX4705 olive fly male sterility when males are reared in the presence of tetracycline in the larval diet, with no reduction in male fertility in the absence of tetracycline from the larval diet, as compared to wild type controls (Figure 7). Female fertility (Figure 8) and general fitness of males were unaffected suggesting a male germline specific expression of the "paternal lethal cassette".

15

Figure 9 Percentage of OX4466 male and female sterility on and off Tetracycline

Following the success of paternally transmitted lethal effect in *Ceratitis capitata*, and *Bactrocera oleae*, the focus was shifted to another Diptera of economic importance; *Aedes aegypti*. The examples below provide evidence of significant male sterility on this species utilising similar promoter and effector sequences as in *C. capitata*. Slight alterations occurred in some constructs utilising genomic sequences of this organism to achieve maximum results.

20

OX4466 (PB-hr5IE1-DsRed-Aepro-tGFP-EcoRI)

25

This also provides an example of another effector; Aepro-EcoRI. OX4466 was injected in pre-blastoderm *Aedes aegypti* embryos. The components of this construct are constitutively expressed rather than inducible by the tet-off system. 4 independent insertion events were generated. Males and females from each strain were backcrossed to wild type mosquitoes of the opposite sex. Wild type insects were used as control. Females were allowed to oviposit on wet filter papers for 24 hours. Survival of progeny was assessed. Results are shown in Figure 9. The chart shows the percentage of collected embryos that hatched from each test cross. The inside table presents actual recorded numbers. The experiment demonstrates a significant reduction in the number of hatched eggs when these were fathered by OX4466 males;

30

35

reduction in embryo viability was less apparent in line D, probably due to positional effects. Results confirm the expected sex-specificity of embryo lethality; embryos from transgenic females are equally viable to these of the wild type control cross.

- 5 The *Aedes*-protamine-nuclease effector DNA sequence (*EcoRI*) of construct OX4466 was fused to a fluorescent gene (*turboGFP*) so that expression of the nuclease should co-exist with expression of the fluorescent protein in the sperm. Fluorescent microscopy of sperm isolated from dissected testis in a number of OX4466 males, showed strong GFP co-localization with nucleus / sperm heads. This is an example of
10 nuclease function fused to fluorescent expression.

Figure 10 Percentage of OX4467 male and female sterility on and off Tetracycline

OX4467 (PB-hr5IE1-DsRed-Aeprot-tGFP-FokICD)

15

OX4467 is identical to plasmid OX4466, the only difference being that the nuclease effector is Fok1 rather than EcoRI.

Injection of this plasmid to pre-blastoderm *Aedes aegypti* embryos resulted in 3
20 independent insertion events. Two of these events were lost in generation G1 through transgenic G0 males; indicating very strong expression of the paternal lethal effect. The remaining strain was analysed as previously. Results are shown in Figure 10. The chart shows the percentage of collected embryos that hatched from each test cross. The inside table presents actual recorded numbers. The experiment demonstrates a
25 significant reduction in the number of hatched eggs when these were fathered by OX4467 males. Results confirm the expected sex-specificity of embryo lethality; embryos from transgenic females are equally viable to these of the wild type control crosses.

- 30 As in OX4466 plasmid, the *Aedes*-protamine-nuclease effector DNA sequence (*Fok1*) was fused to a fluorescent gene (*turboGFP*) so that expression of the nuclease should co-exist with expression of the fluorescent protein in the sperm. Fluorescent microscopy of sperm isolated from dissected testis in a number of OX4467-E1 males, showed strong GFP co-localization with nucleus / sperm heads. This is another
35 example of nuclease function fused to fluorescent expression.

OX4286 (PB-AeTopi-tTAV-K10-3xP3-DsR)

OX4286 (Figure 23) is a piggyBac transposon based construct that contains tTAV driven by *Aedes aegypti* derived topi promoter and topi 5'UTR. OX4286 (Figure 23) employs 3xP3 driven DsRed as a transformation marker. 3xP3 is an artificial, eye-specific promoter, responsive to the evolutionary conserved Pax-6 transcription factor and is active during embryonic, larval and pupal stages. To test the activity of the Topi promoter, a reporter line, OX3979-Ae, was used. The OX3979-Ae contains an AmCyan coding sequence under the control of tetO operator, integrated into a genomic docking site using the phage C31 system. It expresses hr5IE1 driven DsRed as the transformation marker. Double heterozygotes carrying both Topi-tTAV and tetO-AmCyan, generated by crossing together OX4628B and OX3979 lines, showed clear expression of AmCyan in *Aedes aegypti* testis from later larval stages.

OX4635 (PB-HrIE-AmCyan-SV40-Aebeta2tubulin-hsp83-tTAV-SV40)

To test the suitability of the B2 tubulin promoter for use in a conditional male sterility system in *Aedes aegypti* mosquitoes, OX4635, a piggyBac -based construct was built. Smith et al., in their 2007 paper described cloning and characterization of the *Aedes aegypti* B2-tubulin promoter and defined its 959 bp fragment as sufficient for driving DsRed expression in mosquito testis – in stage and tissue-specific manner similar to endogenous promoter. This represented successful direct expression of a reporter gene and was similar in terms of design to our previously tested constitutive male sterile system. To adapt the promoter for use in our conditional expression system we decided to remove as far as possible the transcribed sequences of B2-tubulin, reasoning that these were likely to mediate the translational delay typical of B2-tubulin which would be highly undesirable for the proposed bipartite expression system. The 5'UTR, along with first 36 bp of ORF present in the sequence used by Smith et al (2007), was removed and replaced with an hsp83 minimal promoter from Mediterranean fruit fly. This altered promoter was employed to drive tTAV expression in construct OX4635. OX4635 contains hr5IE1 – driven AmCyan as the transformation marker.

Conditional male sterility in *Aedes aegypti***Figure 11 OX4282-OX4627 Topi-tTAV-driven expression of tetO-Ae-Protamine-FokI-CD (OX4286-B x OX4458)**

5

OX4627 (PB-AeProt-FokI-sv40-polyA-hrIE1-DsRed)

For the second part of the conditional sterility system; the nuclease effector protein, construct OX4627 was built in a similar fashion as OX4458 (Figure 21) (used with success in *C. capitata*); the main difference being that the protamine sequence utilised
10 originated from *Aedes aegypti* and not *D. melanogaster* for optimal results.

Strain OX4282B was crossed to four different OX4627 strains. Double-heterozygous males were crossed to wild type females and resulting embryos were scored for hatching rates. Wild type, OX4627 alone, and female controls were included.
15 Significant (up to 100%) reduction of embryonic hatch rate was observed in all 4 samples of males carrying both alleles reared on a diet without tetracycline. Results are shown in Figure 11. Progeny of crosses between OX4286B and OX4627 lines was reared either on a diet with (tet+) or without tetracycline (tet-). Males carrying both driver and effector alleles were crossed to the wild type females and the hatching rates
20 of eggs obtained from these crosses were calculated (percentage of laid eggs that hatched). Wild type males' crosses with wild type females were used as controls. Crosses where highly significant male sterility was observed (chi-squared test) are marked with asterisks.

Figure 12 OX4635-OX4627 β 2-tub-tTAV-driven expression of tetO-Ae-Protamine-FokI-CD (OX4635 x OX4627)

Two OX4635 strains were crossed to four different OX4627 strains. Double-heterozygous males were crossed to wild type females and resulting embryos scored for hatch rate. A wild type cross was used as control. Significant reduction of embryonic
30 hatch rate was observed in some heterozygous males (for each allele) carrying both alleles reared on a diet without tetracycline. The fact that not all samples demonstrated the same rate of sperm sterility is believed to be due to positional effects of the various transgene insertions. Results are shown in Figure 12. Progeny of crosses between OX4635 and OX4627 lines was reared either on a diet with (tet+) or without
35 tetracycline (tet-). Males carrying both driver and effector alleles were crossed to the

wild type females and the hatching rates of eggs obtained from these crosses were calculated (percentage of laid eggs that hatched). Wild type males' crosses with wild type females were used as controls. Crosses where highly significant male sterility was observed (chi-squared test) are marked with asterisks.

5

A stronger overall sterilisation effect was seen in crosses where nuclease expression was driven by Topi promoter, compared to B2-tubulin, in *Aedes aegypti*. Nevertheless, significant male sterility was observed in both cases, rendering both topi and the altered form of B2-tubulin suitable promoters for the "paternal lethality effect" in the mosquito species *Aedes aegypti*. The stronger effect exerted by Topi promoter could be however, related to more than one insertion of Topi-tTAV allele. According to a phenotype segregation data, there are several Topi-tTAV insertions in OX4286B line – some of them being linked to the male sex determination locus. *Aedes aegypti* lack sexually dimorphic chromosomes, instead sex is determined by the presence or absence of male sex-determination locus (M).

10

15

The nuclease effector fusion protein; protamine-Fok1, is fully functional in three different diptera species tested so far, namely *C. capitata*, *B. oleae* and *Aedes aegypti*.

20

Male sterile and female lethal strains crossed together

In order to examine how the sperm lethal technology may interact with the female lethal RIDL technology and the possible effects on the general performance of a final product containing both transgenes, an experiment was set up where Medfly females of OX4353 (two lines were selected B and F) were crossed to males of a Medfly female lethal line (OX3864A and OX3647Q). The use of RIDL technology was first described in WO 01/39599. Individuals containing both insertions were selected according to their fluorescent phenotypes and sterility and female lethality assays were in place. In the absence of tetracycline, the progeny of these crosses should be male only and sterile. Male fertility and female lethality were assessed in the presence and absence of tetracycline (100ng/μl). Results confirmed that no female progeny was produced in the absence of tetracycline in any of the crosses, whereas a normal 50:50 male to female ratio was obtained when larvae were grown on food containing tetracycline. Males containing both insertions were back crossed to wild type females and male sterility was assessed as in previous crosses. Results are presented in **Figure 13**. Data

25

30

35

indicates that the male sterility of the OX4353 strains tested remained unaffected by the presence of the female lethal positive feedback or indeed it has been reduced slightly further although this can be attributed to stochastic variation. Importantly so, these crosses strongly suggest that the presence of female and sperm lethal plasmids can co-exist in a single organism without any adverse effects on performance of either insertion.

Role of the 5' and 3' UTR in the Cc β 2-tubulin promoter in sperm-lethal strains

The Cc β 2-tubulin promoter as described elsewhere in literature (Catteruccia et al., 2005; Smith et al., 2007; Scolari et al., 2008; Nirmala et al., 2009) did not drive DsRed2 expression in all sperm in OX3671 (**Figure 16**) strains, possibly related to the late transcription or translation of the Cc β 2-tubulin gene. Late expression may result in insufficient tTAV protein binding to tetO repeats for activation of the function of either a reporter gene or a lethal gene in the transgenic strain. Constructs with a modified 5'UTR -OX3831, OX4282 and OX4353 – showed higher levels of fluorescence or lethality in sperm, which illustrates the importance of the 5'UTR in gene expression in spermatogenesis.

In *D. melanogaster* most genes are expressed before meiosis and the products are stored in the developing germline cells, only a few genes are transcribed after meiosis (White-Cooper 2009). The timing of transcription and translation, and the stability of RNA and protein products, are vital for developing sperm-lethal transgenic Medfly strains. The importance of this may vary from one type of effector to another. Indeed Windbichler et al speculate that part of their problem was that the nuclease was too stable, and persisted into the embryo (Windbichler et al. 2008). On the other hand, for sperm labelling with a fluorescent protein, it is clearly essential that the fluorescent protein survives into the mature sperm. Unmodified Cc β 2-tubulin is expressed in the male germline, as confirmed by using promoters from this gene to drive a reporter gene expression in spermatocyte cells. However, this Cc β 2-tubulin promoter in its normal configuration, i.e. combined with 5'UTR sequences derived from the same gene, may drive translation of tTAV very late, and just before meiosis. In which case, there is not enough time for the tTAV protein to bind to tetO sequences and further induce sufficient expression of the adjacent reporter gene, when crossed to a tetO-reporter strain. With the unmodified Cc β 2-tubulin promoter, only one in 20 testes showed

reporter (DsRed2) expression in a few spermatid bundles. On the other hand, 5' UTR replacement resulted in greater fluorescence expression in early spermatocytes and spermatids of dissected male testis in strains OX3831 (**Figure 20**) and OX4282. The 3'UTR contains sequences that regulate translation efficiency and mRNA stability, therefore replacing the 3' UTR of the tubulin gene with a sequence that is known to function in a wide range of species; one possibly eliminates the regulatory elements responsible for late translation of the tubulin gene.

OX4371 (*PB-HrIE-AmCyan-SV40-TurboGFP-teto14-Topi-ubi-tTAV2-SV40*)

OX4112 (*PB-IPPO-1-hsp83-tetO21-Hr5-IE1-Red*)

By inheritance pattern, OX4371 (**Figure 27**) lines appeared to have a single copy of the transgene. When crossed to line OX4112 (**Figure 17**), which contains the effector IPpo1, there was a slight reduction in the number of individuals containing both plasmids (as compared to wild type pupae of the same cross). However, there was no significant reduction in the number of hatched eggs in most lines with the exception of line E which exhibited a 40% reduction in male fertility.

OX4371 (*PB-HrIE-AmCyan-SV40-TurboGFP-teto14-Topi-ubi-tTAV2-SV40*)

OX4104 (*PB-YAFN-hsp83-tetO21-Hr5-IE1-Red*) (**Figure 19**).

When OX4371 (**Figure 27**) lines were crossed to the 3-Zn finger line, flies containing both plasmids were totally viable and healthy. There was no reduction in male fertility of line F; however there was a 45% and 40% reduction in lines B and E respectively.

OX4391 (*PB-HrIE-AmCyan-SV40-TurboGFP-teto14-Topi-ubi-tTAV2-Topi3'UTR*)

OX4104 (*PB-YAFN-hsp83-tetO21-Hr5-IE1-Red*) (**Figure 19**)

Lines OX4391G, C and D were estimated to have a single insertion while lines B and H seemed to have two copies of the transgene each. When OX4391 (**Figure 22**) lines were crossed to the tetO-3zn-finger line (OX4104, **Figure 19**), there did not seem to be any adverse effect on the viability of the individuals containing both plasmids. This indicates that there was very low (if any) basal expression of the transgenes in somatic tissue. Out of all 5 lines analysed, B had a 70% reduction in the ability of sperm to produce viable progeny, while line H was only 40% fertile. There was no significant

difference in egg viability on and off T in the rest of the lines. Results strongly indicate that the presence of two copies in this particular combination significantly decreases the fertility of the males tested.

- 5 **OX4391** (*PB-HrIE-AmCyan-SV40-TurboGFP-teto14-Topi-ubi-tTAV2-Topi3'UTR*)
 OX4112 (*PB-IPPO-1-hsp83-tetO21-Hr5-IE1-Red*)

10 When the same OX4391 (**Figure 22**) lines were crossed to the IPPO1 line, the number of progeny containing both plasmids in the absence of T was reduced compared to progeny containing either of the two plasmids or neither. Egg viability was assessed as described in similar experiments. There was no significant reduction in the number of larvae hatched when compared to the wild type control in lines B, C, F and G. There was a 50% embryo viability reduction in line H.

15 Both *Aetopi* and *Cctopi* promoters have shown ability to direct testis-specific expression of a reporter gene and tTAV, which subsequently induced testis-specific expression of a reporter gene or a lethal gene. Specifically, a 1233-bp sequence, including 1168 bp of *Aetopi* promoter and 65 bp *Aetopi* 5'-UTR, was shown to drive testes-specific expression of the tTAV-DsRed fusion protein in *Ae. aegypti* (data not shown), and tTAV expression can further induce testes-specific AmCyan reporter gene expression. A 1178 bp putative *Cctopi* promoter did not show evidence of testes-specific activity in Medfly. A larger (1708 bp) *Cctopi* fragment sequence was found to be sufficient to drive testes-specific tTAV expression.

25 From an applied genetic engineering perspective, the *Cctopi* putative promoter fragment used was validated as a new male germline-specific promoter in Medfly, with earlier expression than the previously characterised *β -2-tubulin* promoter.

REFERENCES

- Alphey, L. (2007). "Engineering insects for the Sterile Insect Technique", in M. Vreysen, A. Robinson, and J. Hendrichs, (eds.), *Area-wide control of insect pests: from research to field implementation*. Dordrecht, The Netherlands, Springer, pp. 51-60.
- Alphey, L., Beard, B., Billingsley, P., Coetzee, M., Crisanti, A., Curtis, C. F., Eggleston, P., Godfray, C., Hemingway, J., Jacobs-Lorena, M., James, A., Kafatos, F., Mukwaya, L., Paton, M., Powell, J., Schneider, W., Scott, T., Sine, B., Sinden, R., Sinkins, S., Spielman, A., Touré, Y., and Collins, F. (2002). "Malaria control with genetically modified vectors." *Science*, 298, pp. 119-121.
- Arama, E., Agapita, J., and Steller, H. (2003). "Caspase activity and a specific cytochrome C are required for sperm differentiation in *Drosophila*." *Developmental Cell*, 4, pp. 687-697.
- Barreau, C., Benson, E., Gudmannsdottir, E., Newton, F., and White-Cooper, H. (2008). "Post-meiotic transcription in *Drosophila* testes." *Development*, 135, pp. 1897-1902.
- Bauer DuMont, V., Flores, H., Wright, M., and Aquadro, C. (2007). "Recurrent positive selection at Bgcn, a key determinant of germ line differentiation, does not appear to be driven by simple coevolution with its partner protein Bam." *Mol. Biol. Evol.*, 24(1), pp. 182-191.
- Beall, E., Lewis, P., Bell, M., Rocha, M., Jones, D., and Botchan, M. (2007). "Discovery of tMAC: a *Drosophila* testis-specific meiotic arrest complex paralogous to Myb-Muv B." *Genes Dev.*, 21, pp. 904-919.
- Beumer, K., Bhattacharyya, G., Bibikova, M., Trautman, and Carroll, D. (2006). "Efficient gene targeting in *Drosophila* with Zinc-finger nucleases." *Genetics*, 172, pp. 2391-2403.
- Bibikova, M., Golic, M., Golic, K., and Carroll, D. (2002). "Targeted chromosomal cleavage and mutagenesis in *Drosophila* using Zinc-finger nucleases." *Genetics*, 161, pp. 1169-1175.
- Brand, A., Manoukian, A., and Perrimon, N. (1994). "Ectopic expression in *Drosophila*." *Meth. Cell Biol.*, 44, pp. 635-654.
- Brand, A., and Perrimon, N. (1993). "Targeted gene expression as a means of altering cell fates and generating dominant phenotypes." *Development*, 118, pp. 401-415.
- Burt, A. (2003). "Site-specific selfish genes as tools for the control and genetic engineering of natural populations." *Proc. Biol. Sci.*, 270, pp. 921-928.
- Burt, A., and Trivers, R. (2006). *Genes in conflict: The biology of selfish genetic elements*, Cambridge, Belknap Press, Harvard University Press.
- Cagan, R. (2003). "Spermatogenesis: Borrowing the apoptotic machinery." *Current Biology*, 13, pp. R600-R602.
- Catteruccia, F., Benton, J., and Crisanti, A. (2005). "An *Anopheles* transgenic sexing strain for vector control." *Nature Biotechnology*, 23(11), pp. 1414-1417.
- Catteruccia, F., Crisanti, A., and Wimmer, E. (2009). "Transgenic technologies to induce sterility." *Malaria Journal*, 8(Suppl 2), pp. S7.
- Chintapalli, V., Wang, J., and Dow, J. (2007). "Using FlyAtlas to identify better *Drosophila* models of human disease." *Nature Genetics*, 39, pp. 715-720.
- Deredec, A., Burt, A., and Godfray, H. (2008). "Population genetics of using homing endonuclease genes in vector and pest management." *Genetics*, 179, pp. 2013-2026.
- Dhillon, M., Singh, R., Naresh, J., and Sharma, H. (2005). "The melon fruit fly, *Bactrocera cucurbitae*: a review of its biology and management." *Journal of Insect Science*, 5, pp. 40.

- Dyck, V. A., Hendrichs, J., and Robinson, A. S. (2005). "Sterile Insect Technique: Principles and practice in Area-Wide Integrated Pest Management". City: Springer: The Netherlands, pp. 801.
- 5 Franz, G. (2005). "Genetic sexing strains in mediterranean fruit fly, an example for other species amenable to large-scale rearing for the sterile insect technique", in V. A. Dyck, J. Hendrichs, and A. S. Robinson, (eds.), *Sterile Insect Technique. Principles and practice in area-wide integrated pest management*. The Netherlands, Springer, pp. 427-451.
- 10 Fu, G., Condon, K. C., Epton, M. J., Gong, P., Jin, L., Condon, G. C., Morrison, N. I., Dafa'alla, T. H., and Alphey, L. (2007). "Female-specific insect lethality engineered using alternative splicing." *Nature Biotechnology*, 25(3), pp. 353-357.
- 15 Fuller, M. (1993). "Spermatogenesis", in M. Bate and A. Martinez Arias, (eds.), *The development of Drosophila melanogaster*. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press, pp. 71-147.
- Fussenegger, M. (2001). "The impact of mammalian gene regulation concepts on functional genomic research, metabolic engineering, and advanced gene therapies." *Biotechnol. Prog.*, 17, pp. 1-51.
- 20 Fussenegger, M., Morris, R., von Stockar, B., Fux, C., Timann, M., Thompson, C., and Bailey, J. (2000). "Novel Streptogramin-based gene regulation systems for mammalian cells." *Nat. Biotech.*, 18, pp. 1203-8.
- Gonczy, P., Matunis, E., and DiNardo, S. (1997). "*bag-of-marbles* and *benign gonial cell neoplasm* act in the germline to restrict proliferation during *Drosophila* spermatogenesis." *Development*, 124, pp. 4361-4371.
- 25 Gong, P., Epton, M., Fu, G., Scaife, S., Hiscox, A., Condon, K., Condon, G., Morrison, N., Kelly, D., Dafa'alla, T., Coleman, P., and Alphey, L. (2005). "A dominant lethal genetic system for autocidal control of the Mediterranean fruitfly." *Nat. Biotech.*, 23, pp. 453-456.
- 30 Gong, W., and Golic, K. (2003). "Ends-out, or replacement, gene targeting in *Drosophila*." *Proc. Nat. Acad. Sci. (USA)*, 100, pp. 2556-2561.
- Gossen, M., and Bujard, H. (1992). "Tight control of gene expression in mammalian cells by tetracycline- responsive promoters." *Proc Natl Acad Sci U S A*, 89(12), pp. 5547-51.
- 35 Gossen, M., and Bujard, H. (2002). "Studying gene function in eukaryotes by conditional gene inactivation." *Annu. Rev. Genet.*, 36, pp. 153-173.
- Hiller, M., Chen, X., Pringle, M., Suchorolski, M., Sancak, Y., Viswanathan, S., Bolival, B., Lin, T. Y., Marino, S., and Fuller, M. (2004). "Testis-specific TAF homologs collaborate to control a tissue-specific transcription program." *Development*, 131, pp. 5297-5308.
- 40 Hockemeyer, D., Wang, H., Kiani, S., Lai, C. S., Gao, Q., Cassady, J. P., Cost, G. J., Zhang, L., Santiago, Y., Miller, J. C., Zeitler, B., Cherone, J. M., Meng, X., Hinkley, S. J., Rebar, E. J., Gregory, P. D., Urnov, F. D., and Jaenisch, R. (2011). "Genetic engineering of human pluripotent cells using TALE nucleases." *Nat Biotech*, 29(8), pp. 731-734.
- 45 Horn, C., Wimmer, A.E., (2003) "A transgene-based, embryo-specific lethality system for insect pest management". *Nature Biotechnology*, 1, pp. 64-70
- Jattani, R., Patel, U., Kerman, B., and Myat, M. M. (2009). "Deficiency screen identifies a novel role for beta 2 tubulin in salivary gland and myoblast migration in the *Drosophila* embryo." *Developmental Dynamics*, 238(4), pp. 853-863.
- 50 Jiang, J., Benson, E., Bausek, N., Doggett, K., and White-Cooper, H. (2007). "Tombola, a tesmin/TSO1 family protein, regulates transcriptional activation in the *Drosophila* male germline and physically interacts with Always early." *Development*, 134(1549-1559).

- Jiang, J., and White-Cooper, H. (2003). "Transcriptional activation in *Drosophila* spermatogenesis involves the mutually dependent function of *aly* and a novel meiotic arrest gene *cookie monster*." *Development*, 130, pp. 563-573.
- 5 Kawase, E., Wong, M., Ding, B., and Xie, T. (2004). "Gbb/Bmp signaling is essential for maintaining germline stem cells and for repressing *bam* transcription in the *Drosophila* testis." *Development*, 131, pp. 1365-1375.
- Kim, Y.-G., Cha, J., and Chandrasegaran. (1996). "Hybrid restriction enzymes: Zinc finger fusions to *FokI* cleavage domain." *Proc. Nat. Acad. Sci. (USA)*, 93, pp. 1156-1160.
- 10 Klassen, W., and Curtis, C. F. (2005). "History of the sterile insect technique", in V. A. Dyck, J. Hendrichs, and A. S. Robinson, (eds.), *Sterile Insect Technique. Principles and practice in area-wide integrated pest management*. The Netherlands, Springer, pp. 3-36.
- Knippling, E. F. (1955). "Possibilities of insect control or eradication through the use of sexually sterile males." *J Econ Entomol*, 48, pp. 459 - 469.
- 15 Mahfouz, M. M., Li, L., Shamimuzzaman, M., Wibowo, A., Fang, X., and Zhu, J.-K. (2011). "De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks." *Proceedings of the National Academy of Sciences*.
- 20 Malacrida, A., Scolari, F., Schetelig, M., Bertin, S., Gasperi, G., and Wimmer, E. (2007). "A transgenic sperm marking system in the medfly, as a tool for pest control strategies and sperm use analysis." *Entomological Research*, 37, pp. (Suppl. 1): A56.
- Maynard-Smith, L., Chen, L.-C., Banaszynski, L., Ooi, A., and Wandless, T. (2007). "A directed approach for engineering conditional protein stability using biologically silent small molecules." *Journal of Biological Chemistry*, 282(34), pp. 24866-24872.
- 25 Miller, J., Holmes, M., Wang, J., Guschin, D., Lee, Y.-L., Rupniewski, I., Beausejour, C., Waite, A., Wang, N., Kim, K., Gregory, P., Pabo, C., and Rebar, E. (2007). "An improved zinc-finger nuclease architecture for highly specific genome editing." *Nature Biotechnology*, 25(7), pp. 778-785.
- 30 Miller, J. C., Tan, S., Qiao, G., Barlow, K. A., Wang, J., Xia, D. F., Meng, X., Paschon, D. E., Leung, E., Hinkley, S. J., Dulay, G. P., Hua, K. L., Ankoudinova, I., Cost, G. J., Urnov, F. D., Zhang, H. S., Holmes, M. C., Zhang, L., Gregory, P. D., and Rebar, E. J. (2011). "A TALE nuclease architecture for efficient genome editing." *Nat Biotech*, 29(2), pp. 143-148.
- 35 Nielsen, M., Turner, F., Hutchens, J., and Raff, E. (2001). "Axoneme-specific β -tubulin specialization: a conserved C-terminal motif specifies the central pair." *Current Biology*, 11, pp. 529-533.
- 40 Osterwalder, T., Yoon, K., White, B., and Keshishian, H. (2001). "A conditional tissue-specific transgene expression system using inducible GAL4." *Proc. Nat. Acad. Sci. (USA)*, 98(22), pp. 12596-12601.
- Perezgasga, L., Jiang, J., Bolival, B., Hiller, M., Benson, E., Fuller, M., and White-Cooper, H. (2004). "Regulation of transcription of meiotic cell cycle and terminal differentiation genes by the testis-specific Zn-finger protein *matotopetli*." *Development*, 131, pp. 1691-1702.
- 45 Phuc, H. K., Andreasen, M. H., Burton, R. S., Vass, C., Epton, M. J., Pape, G., Fu, G., Condon, K. C., Scaife, S., Donnelly, C. A., Coleman, P. G., White-Cooper, H., and Alphey, L. (2007). "Late-acting dominant lethal genetic systems and mosquito control." *BMC Biology*, 5, pp. 11.
- 50 Raja, S., and Renkawitz-Pohl, R. (2005). "Replacement by *Drosophila melanogaster* protamines and Mst77F of histones during chromosome condensation in late

- spermatids and role of Sesame in the removal of these proteins from the male pronucleus." *Molecular and Cellular Biology*, 25(14), pp. 6165-6177.
- 5 Rendón, P., McInnis, D., Lance, D., and Stewart, J. (2004). "Medfly (Diptera:Tephritidae) genetic sexing: large-scale field comparison of males-only and bisexual sterile fly releases in Guatemala." *Journal of Economic Entomology*, 97(5), pp. 1547-1553.
- Robinson, A. S. (2005). "Genetic basis of the sterile insect technique", in V. A. Dyck, J. Hendrichs, and A. S. Robinson, (eds.), *Sterile Insect Technique. Principles and practice in area-wide integrated pest management*. The Netherlands, Springer, 10 pp. 95-114.
- Rong, Y., and Golic, K. (2000). "Gene targeting by homologous recombination in *Drosophila*." *Science*, 288(5473), pp. 2013-8.
- Rong, Y., and Golic, K. (2001). "A targeted gene knockout in *Drosophila*." *Genetics*, 157, pp. 1307-1312.
- 15 Rong, Y., Titen, S., Xie, H., Golic, M., Bastiani, M., Bandyopadhyay, P., Olivera, B., Brodsky, M., Rubin, G., and Golic, K. (2002). "Targeted mutagenesis by homologous recombination in *Drosophila melanogaster*." *Genes Dev.*, 16, pp. 1568-1581.
- Santel, A., Winhauer, T., Blümer, N., and Renkawitz-Pohl, R. (1997). "The *Drosophila don juan (dj)* gene encodes a novel sperm specific protein component characterized by an unusual domain of a repetitive amino acid motif." *Mechanisms of Development*, 64, pp. 19-30.
- 20 Schetelig, M.F., Handler, M.A. (2012) "strategy for enhanced transgenic strain development for embryonic conditional lethality in *Anastrepha suspensa*". PNAS, 109 (24), pp. 9348-9353
- 25 Thomas, D. D., Donnelly, C. A., Wood, R. J., and Alphey, L. S. (2000). "Insect population control using a dominant, repressible, lethal genetic system." *Science*, 287(5462), pp. 2474-2476.
- Urnov, F., Miller, J., Lee, Y.-L., Beausejour, C., Rock, J., Augustus, S., Jamieson, A., Porteus, M., Gregory, P., and Holmes, M. (2005). "Highly efficient endogenous human gene correction using designed zinc-finger nucleases." *Nature*, 435, pp. 646-651.
- Victorinová, I., and Wimmer, E. (2007). "Comparative analysis of binary expression systems for directed gene expression in transgenic insects." *Insect Biochem. Mol. Biol.*, 37, pp. 246-254.
- 35 White-Cooper, H., Leroy, D., MacQueen, A., and Fuller, M. (2000). "Transcription of meiotic cell cycle and terminal differentiation genes depends on a conserved chromatin associated protein, whose nuclear localisation is regulated." *Development*, 127, pp. 5463-5473.
- 40 Wilson, C., Bellen, H. J., and Gehring, W. J. (1990). "Position effects on eukaryotic gene expression." *Annu Rev Cell Biol*, 6, pp. 679-714.
- Wilson, K., Fitch, K., Bafus, B., and Wakimoto, B. (2006). "Sperm plasma membrane breakdown during *Drosophila* fertilization requires *sneaky*, an acrosomal membrane protein." *Development*, 133(24), pp. 4871-4879.
- 45 Windbichler, N., Menichelli, M., Papathanos, P. A., Thyme, S. B., Li, H., Ulge, U. Y., Hovde, B. T., Baker, D., Monnat, R. J., Burt, A., and Crisanti, A. (2011). "A synthetic homing endonuclease-based gene drive system in the human malaria mosquito." *Nature*, 473(7346), pp. 212-215.
- Windbichler, N., Papathanos, P., Catteruccia, F., Ranson, H., Burt, A., and Crisanti, A. 50 (2007). "Homing endonuclease mediated gene targeting in *Anopheles gambiae* cells and embryos." *Nucl Acids Res*, 35, pp. 5922 - 5933.
- Windbichler, N., Papathanos, P. A., and Crisanti, A. (2008). "Targeting the X chromosome during spermatogenesis induces Y chromosome transmission

ratio distortion and early dominant embryo lethality in *Anopheles gambiae*." *PLoS Genet*, 4(12), pp. e1000291.

5

SEQUENCES

The following relate to SEQ ID NOs: 1-56 provided hereafter.

10

SEQ ID NOS: 1-5 Full beta-2 tubulin sequences - different insects

SEQ ID NO: 1

>gi|167822003|gb|EU386342.1| Ceratitis capitata beta-2 tubulin mRNA, complete cds

SEQ ID NO: 2

15

>gi|158742|gb|M20420.1|DROTUBB2A D.melanogaster beta-2 tubulin mRNA, complete cds

SEQ ID NO: 3

>gi|111035017|gb|DQ833526.1| Aedes aegypti beta-2 tubulin (B2t) gene, complete cds

20

SEQ ID NO: 4

>gi|219815271|gb|EU938673.1| Bactrocera dorsalis beta-2 tubulin gene, complete cds

SEQ ID NO: 5

25

>gi|219815267|gb|EU938671.1| Anastrepha suspensa beta-2 tubulin gene, complete cds

SEQ ID NOS: 6-10: beta-2 tubulin 5' UTR sequences - different insects

SEQ ID NO: 6

30

>gi|167822003|gb|EU386342.1| Ceratitis capitata beta-2 tubulin 5'UTR

SEQ ID NO: 7

>gi|158742|gb|M20420.1|DROTUBB2A D.melanogaster beta-2 tubulin 5' UTR

35

SEQ ID NO: 8

>gi|219815271|gb|EU938673.1| Bactrocera dorsalis beta-2 tubulin 5' UTR

SEQ ID NO: 9

>gi|111035017|gb|DQ833526.1| Aedes aegypti beta-2 tubulin (B2t) 5' UTR

SEQ ID NO: 10

5 >gi|219815267|gb|EU938671.1| Anastrepha suspensa beta-2 tubulin 5' UTR

SEQ ID NOS: 11-15 : beta-2 tubulin 3' UTR sequences - different insects

SEQ ID NO: 11

>gi|167822003|gb|EU386342.1| Ceratitis capitata beta-2 tubulin 3' UTR

10

SEQ ID NO: 12

>gi|158742|gb|M20420.1|DROTUBB2A D.melanogaster beta-2 tubulin mRNA 3' UTR

SEQ ID NO: 13

15 >gi|111035017|gb|DQ833526.1| Aedes aegypti beta-2 tubulin (B2t) gene 3' UTR

SEQ ID NO: 14

>gi|219815271|gb|EU938673.1| Bactrocera dorsalis beta-2 tubulin gene 3' UTR

20 SEQ ID NO: 15

>gi|219815267|gb|EU938671.1| Anastrepha suspensa beta-2 tubulin gene 3' UTR

tTAV and variants

SEQ ID NO. 16: Open reading frame of tTAV

25 SEQ ID NO. 17: Protein sequence of tTAV

SEQ ID NO. 18: Open reading frame of tTAV2

SEQ ID NO. 19: Protein sequence of tTAV2

30 SEQ ID NO. 20: Open reading frame of tTAV3

SEQ ID NO. 21: Protein sequence of tTAV3

SEQ ID NO: 22 - 5' UTR Beta-2 tubulin from *D. melanogaster*

SEQ ID NO: 23-24 - b2 tubulin promoter sequence of drosophila melanogaster (SEQ ID NO 24) aligned with the mRNA sequence of Dm b2-tubulin (5'UTR is included in the alignment) (SEQ ID NO 23)

5 SEQ ID NO: 25 - ORF of Hsp83 Medfly

SEQ ID NO: 26 - 5' UTR as given by genebank

SEQ ID NO: 27 - Hsp83 5' UTR from 4353

10

SEQ ID NO: 28 - Amino acid sequence of Hsp83 medfly

SEQ ID NO: 29 – Dm Aly 5'UTR from OX3831

15 SEQ ID NO: 30 – Dm Aly promoter

SEQ ID NO: 31 – TETR

SEQ ID NO: 32 - VP16

20 TETR and VP16 sequences combined make the tTAV sequence.

SEQ ID NO: 33 – DmTopi cDNA

SEQ ID NO: 34 – Dm promoter sequence as found on fly base

25

SEQ ID NO: 35 - # OX4254 Cc topi promoter

SEQ ID NO: 36 - # OX4275 Cc topi promoter

30 SEQ ID NO: 37 - # OX4371 Cc topi promoter

SEQ ID NO: 38-40 – Alignment of LA4254 (SEQ ID NO 38) cf LA4275 (SEQ ID NO 39) cf LA4371 (SEQ ID NO 40)

35 SEQ ID NO: 41 – Aedes topi promoter from 4286

SEQ ID NO: 42 – Ae topi 5'UTR from 4286

SEQ ID NO: 43 – Topi *Drosophila melanogaster* ORF (coding region)

5

SEQ ID NO: 44 - Aly ORF from *Drosophila melanogaster*(coding region)

SEQ ID NO: 45 – LA 3671 B2 tubulin promoter and 5'UTR

10

SEQ ID NO: 46 – LA 4353 B2 tubulin promoter

SEQ ID NO: 47 – LA 4353 hsp83 5' UTR

SEQ ID NO: 48 - B2 tubulin 5'UTR

15

SEQ ID NO: 49 - AeProtamine

SEQ ID NO: 50 - SG4

20

SEQ ID NO: 51 – Fok1 cleavage domain

SEQ ID NO: 52 – Aeprotamine-SG4-Fok1

SEQ ID NO: 53 – Medfly B2 tubulin promoter forward primer

25

SEQ ID NO: 54 – Medfly B2 tubulin promoter reverse primer

SEQ ID NO:55 – TopitestF1 primer

30

SEQ ID NO: 56 – Topitest R primer

SEQ ID NO: 23-24

b2 tubulin promoter sequence of drosophila melanogaster aligned with the mRNA sequence of Dm b2-tubulin (5'UTR is included in the alignment)

35

	extended dro	TCCTTTATTGAGATTAAACGGTCAAATCAATAGATAAAAAGAAAAC TTATTACATATTTAAA -----
5	extended dro	GAATGATGAAATTTTTAAAATTCATTGTATCATATGTTATTTCGGCCACTGTAACCGAAAT -----
	extended dro	CAACCATTTTTGGCGGATGCTGTGTGTTTGTGTTTGCTGACAACTATCGATTTTGTCTAGAC -----
10	extended dro	GCAGCATCTTTAACTGAACGAAAAAGGCGCGTGGTGCAAAATATATTAATTGATTATAGA -----
15	extended dro	TCGTAGTGATTATATTTGAGACTATATGATGAAGCGACAGAATGTCCGTACCCTTTCCCT -----
	extended dro	GGTGGTATGCACTTTTACCTATCTTTTAATTGGAGCCGCCGTGTTTCGATTCCCTGGAGTC -----
20	extended dro	ACCAACGGAGGCCAAAAGATGGGAATTCCTACAGAGTGAGAAGCTTGTGATTATTAAAC -----
	extended dro	CTAATTTCTTAGTAATGAATTTATTTAATCAATTGTAGCCGTTAAGAACAAC TTTGTTAG -----
25	extended dro	AAAGTACAATGTGACTGACGAGGATTTCCGTGTGATGGAAATCGTCATCATTGAAAATAA -----
	extended dro	GCCCCACAAGCCGGACCTCAGTGGAATTCGCTGGAGCTTTCTATTTTCAGCACGGTTGT -----
30	extended dro	ACTGGCAATGATAGGTAAATTAATTATCTATTAAATATGATTATTGAATAGATTATAAT -----
35	extended dro	TCTGTTGTAAC TTTCTTTAGGATATGGTCATTCTACGCCAGTTACAATTCGGGAAAAAGC -----
	extended dro	ATTTTGTATGGCTATGCTATGGTAAGTGAAC TTACAATCCCAATTTCCAGTCTTCTAAA -----
40	extended dro	GATATTCCTTATTAGGTGGGCATCCCGCTGGGTCTGGTGATGTTCCAGTCTATCGGAGA -----
45	extended dro	ACGTCTGAATAAGTTTGCATCCGTGATAATAAGGCGGGCAAAGAGAGCCAGTGGAGCTCG -----
	extended dro	CTGTACGGATGCCACCGAAATGAATCTCATGTTGGCCACCGGAATGCTCTCCTCCATAAT -----
	extended	AATCACC ACTGGAGCAGCAGTCTTTTCCCGATACGAGGGTTGGAGCTACTTCGATAGCTT

	dro	-----
	extended	CTACTATTGTTTTGTACCTTGACGACAATTGGTTTCGGCGATTATGTGGCATTGCAGAA
5	dro	-----
	extended	CGACCAAGCTCTAACTAATAAGCCTGGCTATGTGGCGCTGAGCTTGGTCTTCATCCTATT
	dro	-----
10	extended	CGGCTTGGCCGTGGTGGCCGCCAGTATCAATCTATTGGTGCTCCGATTTCATGACCATGTG
	dro	-----
	extended	AGTCCATGTTCTATTGCAGGAAATATCTTATTTAATGGATTTTAAATCACAGGCAAGCAG
	dro	-----
15	extended	AGGATGCCAAGAGAGATGAGCAGGATGCTCAGAACTTGGCTGGAATGCCAGCCGGTGA
	dro	-----
	extended	CCTTCGATGATGAGTCCAGTACAATATGCACGGCAAGCTGCTGGAGAACAACCTACACAA
20	dro	-----
	extended	CGGAGAACGATGAGACCGCTCCCTGTGTTCCCTGCACCTGCATGGGTGGCACCAGGTGCC
	dro	-----
25	extended	TGAATCATGAGCAGTTCGTGGACCCGGACTTTCAGCCTACCGACATTATCGAGAGCACCT
	dro	-----
	extended	TGTGCCTGAAGCGAGCCTCCGTCTGATATCCGTACAGCCAGCTGTGGGACTCCTCATTGT
	dro	-----
30	extended	AGGAGCCAGAGCCAATGGATCACCAAATCGTAGTTACAATCCTGTAGAGAACCATCCGCC
	dro	-----
	extended	GCCAAAATTTGGTTGTTAGACAAACCTTCCTCCCTACGTAGATTTTAAACCAGGATGGG
35	dro	-----
	extended	TCATAATACATATAAGTTTGGAGAGCAAGGTTAATAGTCTTTAAAAGGCAGTTTTTGCTT
	dro	-----
40	extended	AAGAAATAATCGACCCATCCCATTTATACACCCATATAAACATTTACAAAGGAGTAAAATC
	dro	-----
	extended	CAGGACATCCATGTCAATATCAATCGTATCATCTGGTCGGTAGCCTTGAATCCTCTATT
45	dro	-----
	extended	GCTTCCAAGGCACCGCCAAATCCATCCCATCTCGAATTTTAGCCGTATATTGTTTATCT
	dro	-----
	extended	ATGTAAGTACTATTAAAGTTTGTGCTCAAAACGGAGAACTGAGTTTTCTGAAATCGGGGT

dro -----

5 extended GTGTGAAATGTGTCGAAGTCGGAATCGTAGTAGCCTATTTGTGAACATTTCGGTGTAGTA
dro -----GGAAATCGTAGTAGCCTATTTGTGAACATTTCGGTGTAGTA

10 extended ATCCAAGCCAGGTTTCAGTTCACCTCAGTATCAGCTAGCACGTACACGACTAAAATCTAAA
dro ATCCAAGCCAGGTTTCAGTTCACCTCAGTATCAGCTAGCACGTACACGACTAAAATCTAAA

15 extended CCTGAAAAATTATACGTTTAAATATTTCAGTCTTTTGCCGATTTTGCCCCACTCAGACTG
dro CCTGAAAAATTATACGTTTAAATATTTCAGTCTTTTGCCGATTTTGCCCCACTCAGACTG

20 extended TTTTAAAAGCTCGATTTTTTTTTGTACCATTTTTTCGGTGTGAAAAAGGGGCCCTAACTT
dro TTTTAAAAGCTCGATTTTTTTTTGTACCATTTTTTCGGTGTGAAAAAGGGGCCCTAACTT
***** ** ****

25 extended TACTATCAAAATGCGTGAAATGTACACATTAGGCCGGTCAATGCGGTAACCAGATCGG
dro TACTATCAAAATGCGTGAAATGTACACATTAGGCCGGTCAATGCGGTAACCAGATCGG

30 extended TGGTAAATTCTGGGAGGTAATCTCGGATGAGCACTGTATAGATGCGACCGGAACGTACTA
dro TGGTAAATTCTGGGAGGTAATCTCGGATGAGCACTGTATAGATGCGACCGGAACGTACTA

35 extended CGGCGATAGTGATCTCCAGCTGGAGCGCATCAATGTATACTACAATGAAGCCACCGGTGC
dro CGGCGATAGTGATCTCCAGCTGGAGCGCATCAATGTATACTACAATGAAGCCACCGGTGC

40 extended CAAGTATGTGCCACGCGCAATTCTCGTGGACCTGGAGCCCGGCACCATGGATTTCGGTTTCG
dro CAAGTATGTGCCACGCGCAATTCTCGTGGACCTGGAGCCCGGCACCATGGATTTCGGTTTCG

45 extended TTCTGGCGCCTTTGGCCAGATCTTCCGGCCGGACAATTTGTGTTTGCCAATCGGGAGC
dro TTCTGGCGCCTTTGGCCAGATCTTCCGGCCGGACAATTTGTGTTTGCCAATCGGGAGC

extended AGGCAACAAC TGGGCCAAGGGTCATTACACCGAGGGTGCTGAAC TGGTGGATTCCGTCTT
dro AGGCAACAAC TGGGCCAAGGGTCATTACACCGAGGGTGCTGAAC TGGTGGATTCCGTCTT

extended GGATGTGGTGCGAAAGGAGTCCGAGGGATGCGATTGCCTTCAGGTAAGTTTGGGGGTTT
dro GGATGTGGTGCGAAAGGAGTCCGAGGGATGCGATTGCCTTCAGG-----

extended GGGAAATTTATCTGAAAAAGTTTACCCTACTTTTCTCCAACAGGGCTTCAGCTGACCCAC
dro -----GCTTCAGTCGACCCAC

5	extended	TCGCTGGGTGGCGGCACTGGCTCCGGCATGGGAACCCTGCTGATCTCGAAGATCCGCGAG
	dro	TCGCTGGGTGGCGGCACTGGCTCCGGCATGGGAACCCTGCTGATCTCGAAGATCCGCGAG *****
10	extended	GAGTACCCGGACCGCATCATGAACACCTTCTCGGTGGTGCCCTCGCCCAAGGTGTCCGAT
	dro	GAGTACCCGGACCGCATCATGAACACCTTCTCGGTGGTGCCCTCGCCCAAGGTGTCCGAT *****
15	extended	ACGGTGGTGGAGCCCTACAATGCCACCCTGAGTGTGCATCAGCTGGTGGAGAACACCGAT
	dro	ACGGTGGTGGAGCCCTACAATGCCACCCTGAGTGTGCATCAGCTGGTGGAGAACACCGAT *****
20	extended	GAGACGTACTGCATCGACAACGAGGCGTTGTATGACATCTGCTTCCGCACACTGAAGCTG
	dro	GAGACGTACTGCATCGACAACGAGGCGTTGTATGACATCTGCTTCCGCACACTGAAGCTG *****
25	extended	ACCACGCCCACCTACGGTGACCTGAACCATCTGGTTTCGGCCACCATGTCTGGTGTGACG
	dro	ACCACGCCCACCTACGGTGACCTGAACCATCTGGTTTCGGCCACCATGTCTGGTGTGACG *****
30	extended	ACCTGCCTGCGCTTCCCTGGCCAGCTGAACGCTGATCTTCGCAAGCTGGCCGTGAACATG
	dro	ACCTGCCTGCGCTTCCCTGGCCAGCTGAACGCTGATCTTCGCAAGCTGGCCGTGAACATG *****
35	extended	GTACCCTTCCCCGGCTGCACCTTCTTCATGCCCGGATTGCGACCGCTCACCTCGCGAGGA
	dro	GTACCCTTCCCCGGCTGCACCTTCTTCATGCCCGGATTGCGACCGCTCACCTCGCGAGGA *****
40	extended	TCGCAACAATACCGGGCCCTTACCGTTCCGGAGCTGACCCAGCAGATGTTTCGATGCCAAG
	dro	TCGCAACAATACCGGGCCCTTACCGTTCCGGAGCTGACCCAGCAGATGTTTCGATGCCAAG *****
45	extended	AACATGATGGCTGCGTGCGATCCCGACATGGTCGCTATCTGACCGTCGCCGCCATCTTC
	dro	AACATGATGGCTGCGTGCGATCCCGACATGGTCGCTATCTGACCGTCGCCGCCATCTTC *****
50	extended	CGTGGCCGCATGTCCATGAAGGAGGTGGACGAGCAGATGCTCAACATTCAGAACAAAGAAC
	dro	CGTGGCCGCATGTCCATGAAGGAGGTGGACGAGCAGATGCTCAACATTCAGAACAAAGAAC *****
55	extended	AGCAGCTTCTTCGTGGAATGGATCCCGAATAACTGCAAGACAGCGGTGTGCGATATTCGG
	dro	AGCAGCTTCTTCGTGGAATGGATCCCGAATAACTGCAAGACAGCGGTGTGCGATATTCGG *****
60	extended	CCCAGAGGTCTCAAGATGTCGGCCACCTTCATTGGCAACTCCACCGCCATTTCAGGAGCTA
	dro	CCCAGAGGTCTCAAGATGTCGGCCACCTTCATTGGCAACTCCACCGCCATTTCAGGAGCTA *****

	extended	TTCAAACGGGTTTCGGAGCAGTTCACCGCCATGTTCCGAAGGAAGGCCTTCTTGCAATTGG
	dro	TTCAAACGGGTTTCGGAGCAGTTCACCGCCATGTTCCGAAGGAAGGCCTTCTTGCAATTGG *****
5	extended	TACACCGGCGAGGGAATGGACGAAATGGAATTCACAGAGGCCGAGAGCAACATGAACGAC
	dro	TACACCGGCGAGGGAATGGACGAAATGGAATTCACAGAGGCCGAGAGCAACATGAACGAC *****
10	extended	TTGGTTTCTGAATATCAGCAGTACCAGGAGGCGACTGCCGATGAGGAGGGCGAATTCGAT
	dro	TTGGTTTCTGAATATCAACAGTACCAGGAGGCGACTGCCGATGAGGAGGGCGAATTCGAT *****
15	extended	GAAGACGAAGAGGGTGGCGGCGATGAATAATAGGATTAACCTCCCACTCAAGATCACACA
	dro	GAAGACGAAGAGGGTGGCGGCGATGAATAATAGGATTAACCTCCCACTCAAGATCACACA *****
20	extended	TGAACACCAAAACAGGCTAGCAGGGGAACCCATTAGGAAGGCACAACACATTGGATCTTT
	dro	TGAACACCAAAACAGGCTAGCAGGGGAACCCATTAGGAAGGCACAATACATTGGATCTTT *****
25	extended	GGGCCTTAGCATATTGTGCTTCGAGGCCCGTCGGTTGTACATATTTCTATATGGATTCT
	dro	GGGCCTTAGCATATTGTGCTTCGAGGCCCGTCGGTTGTACATATTTCTATATGGATTCT *****
30	extended	GCTCATGTTGCAATTGCTGTGATTTCTGGGTTACGAATAAATGTTGATTTATAAGCAGAC
	dro	GCTCATGTTGCAATTGCTGTGATTTCTGGGTTACGAATAAATGTTGATTTATAAGCAGAC *****
35	extended	AAGATTACCAACAGCATTTTGCATATTTTATACCGTTCAAAGGCATTGCATAAACCTA
	dro	AAGATTACCAACAGCATTTTGCATATTTT----- *****

SEQ ID NO: 38-40 – Alignment of LA4254 (SEQ ID NO 38) cf LA4275 (SEQ ID NO 39) cf LA4371 (SEQ ID NO 40):

40	4254	GCCGTTCAAGTCAAATGTGATATTCACAACTATTGAGCAGAGAATTCCATTAATGTACATA 60
	4275	GCCGTTCAAGTCAAATGTGATATTCACAACTATTGAGCAGAGAATTCCATTAATGTACATA 60
	4371	-----
45	4254	TGATTTTGAATTGCTGCAACAAAAAATATTTAAATGGTTTAGCAAGGTTAATTAAGTGTA 120
	4275	TGATTTTGAATTGCTGCAACAAAAAATATTTAAATGGTTTAGCAAGGTTAATTAAGTGTA 120
	4371	-----
50	4254	AATGACAGATTTTTTTTACATACACCACTTCGCCCTGTAGCTAGTTGCGAGTTTACTT 180
	4275	AATGACAGATTTTTTTTACATACACCACTTCGCCCTGTAGCTAGTTGCGAGTTTACTT 180
	4371	-----
	4254	CAGTTTCTATCTAATTCGTTTGAATCCATATGGCAGAATTACAGTGTAATGGACGCTCTC 240
	4275	CAGTTTCTATCTAATTCGTTTGAATCCATATGGCAGAATTACAGTGTAATGGACGCTCTC 240
	4371	-----CTAATTCGTTTGAATCCATATGGCAGAATTACAGTGTAATGGACGCTCTC 50

5	4254	TTACTTTTTAGGCTTAAAAACACATTAAAGATCAATTAATTTTAAGGAATAAGCAAAT 300
	4275	TTACTTTTTAGGCTTAAAAACACATTAAAGATCAATTAATTTTAAGGAATAAGCAAAT 300
	4371	TTACTTTTTAGGCTTAAAAACACATTAAAGATCAATTAATTTTAAGGAATAAGCAAAT 110

10	4254	AAAATTACTCCGGCGTTTCAGATATTGGAAATATAGAATAATGTAACATTTAAAATAAGGC 360
	4275	AAAATTACTCCGGCGTTTCAGATATTGGAAATATAGAATAATGTAACATTTAAAATAAGGC 360
	4371	AAAATTACTCCGGCGTTTCAGATATTGGAAATATAGAATAATGTAACATTTAAAATAAGGC 170

15	4254	CTAATATTTATCAATTATCAAGACATATGTATATACATGATTATGCAAAAGGTATTCAT 420
	4275	CTAATATTTATCAATTATCAAGACATATGTATATACATGATTATGCAAAAGGTATTCAT 420
	4371	CTAATATTTATCAATTATCAAGACATATGTATATACATGATTATGCAAAAGGTATTCAT 230

20	4254	TTTTAATAATGCAGGGAAAACTACAGCTAAACAACAACGTAATCAATTCCTACTTGGTA 480
	4275	TTTTAATAATGCAGGGAAAACTACAGCTAAACAACAACGTAATCAATTCCTACTTGGTA 480
	4371	TTTTAATAATGCAGGGAAAACTACAGCTAAACAACAACGTAATCAATTCCTACTTGGTA 290

25	4254	TTTCTTCGTTCCCTTTAACATTTTTTCATAACAGTAGGTTTTCAATATTTAGATGTAA 540
	4275	TTTCTTCGTTCCCTTTAACATTTTTTCATAACAGTAGGTTTTCAATATTTAGATGTAA 540
	4371	TTTCTTCGTTCCCTTTAACATTTTTTCATAACAGTAGGTTTTCAATATTTAGATGTAA 350

30	4254	ATGAAAAATGTACGGTTTCCGTGGCAAGCTTAACCTGCCATTCTTCTGAACAATTTAATC 600
	4275	ATGAAAAATGTACGGTTTCCGTGGCAAGCTTAACCTGCCATTCTTCTGAACAATTTAATC 600
	4371	ATGAAAAATGTACGGTTTCCGTGGCAAGCTTAACCTGCCATTCTTCTGAACAATTTAATC 410

35	4254	TAATAATTTTTCATTATCTAAGGCGTCAATTTAAATGGCAAAGTATTAATATTCTTGATG 660
	4275	TAATAATTTTTCATTATCTAAGGCGTCAATTTAAATGGCAAAGTATTAATATTCTTGATG 660
	4371	TAATAATTTTTCATTATCTAAGGCGTCAATTTAAATGGCAAAGTATTAATATTCTTGATG 470

40	4254	GTTGCCTAAATTTTAGAAATAAACTGAATGCTATTAACCTAAGGAAGTTGAGGTAAAAG 720
	4275	GTTGCCTAAATTTTAGAAATAAACTGAATGCTATTAACCTAAGGAAGTTGAGGTAAAAG 720
	4371	GTTGCCTAAATTTTAGAAATAAACTGAATGCTATTAACCTAAGGAAGTTGAGGTAAAAG 530

45	4254	TTTTGTTTAAATTCACATATGTTGGAATATCGTCATCAAAAATAAATGTGTCCTGTAAT 780
	4275	TTTTGTTTAAATTCACATATGTTGGAATATCGTCATCAAAAATAAATGTGTCCTGTAAT 780
	4371	TTTTGTTTAAATTCACATATGTTGGAATATCGTCATCAAAAATAAATGTGTCCTGTAAT 590

50	4254	TAATATGTTTATCGTTTAGTTTTAAATTTAAATTAATTTAAGTTAACTGTAATGGGTGT 840
	4275	TAATATGTTTATCGTTTAGTTTTAAATTTAAATTAATTTAAGTTAACTGTAATGGGTGT 840
	4371	TAATATGTTTATCGTTTAGTTTTAAATTTAAATTAATTTAAGTTAACTGTAATGGGTGT 650

55	4254	ACTCAATCGTTGGATTAGAAATTGAAAGCGGAGGCAAATATAATTTTTCGGTGTTGGGTA 900
	4275	ACTCAATCGTTGGATTAGAAATTGAAAGCGGAGGCAAATATAATTTTTCGGTGTTGGGTA 900
	4371	ACTCAATCGTTGGATTAGAAATTGAAAGCGGAGGCAAATATAATTTTTCGGTGTTGGGTA 710

60	4254	AGTGTTACAATTGCAACAGTTTTAAATTAGAATAATTAATATATGAAAATGCATTAAA 960
	4275	AGTGTTACAATTGCAACAGTTTTAAATTAGAATAATTAATATATGAAAATGCATTAAA 960
	4371	AGTGTTACAATTGCAACAGTTTTAAATTAGAATAATTAATATATGAAAATGCATTAAA 770

65	4254	ATCAAAAATATCCATGATTAAATCATATTTAAATGTAGAAATTAATAACACTAAAATAT 1020
	4275	ATCAAAAATATCCATGATTAAATCATATTTAAATGTAGAAATTAATAACACTAAAATAT 1020
	4371	ATCAAAAATATCCATGATTAAATCATATTTAAATGTAGAAATTAATAACACTAAAATAT 830

	4254	TTTGGTAAATTAAGACACTATCAAAAACTCGAAAAAGTAGGCTAGCTTCTATGTCAA 1080
	4275	TTTGGTAAATTAAGACACTATCAAAAACTCGAAAAAGTAGGCTAGCTTCTATGTCAA 1080

4371 TTTGGTAAATTAAGACACTATCAAAAACTCGAAAAAGTAGGCTAGCTTCTATGTCAA 890

5 4254 GGCGCCATTTTTAAAGAACAATAGATCTAGAAATACTGCAGAGTCCGCAAAATTTTGAA 1140
4275 GGCGCCATTTTTAAAGAACAATAGATCTAGAAATACTGCAGAGTCCGCAAAATTTTGAA 1140
4371 GGCGCCATTTTTAAAGAACAATAGATCTAGAAATACTGCAGAGTCCGCAAAATTTTGAA 950

10 4254 TTTATTTTATAAATATAAACTAAATTAATCCACTAG----- 1178
4275 TTTATTTTATAAATATAAACTAAATTAATCCACTAG----- 1178
4371 TTTATTTTATAAATATAAACTAAATTAATCCACTAGTATGAATACAAGTGAAGAAGAA 1010

15 4254 -----
4275 -----
4371 TTTTCGAATTCGATGCGTGGCTATCCGAGCAACTGTTGCTCMATTAAGAATTTAAT 1070

20 4254 -----
4275 -----
4371 TCAGATTATAGAAAAAGTCGGTTGGTGATGCATCGACMACATTTGTATTTCTTCCGGT 1130

25 4254 -----
4275 -----
4371 AGTCTCAGCTGTTGCTGAAGGAGAACCTCAGACTTAACAAAATCAGACTTGAAAAC 1190

30 4254 -----
4275 -----
4371 TACGAGCCTGTTTCAAATTATCTACCACTAATATATCTTCTTCGATCTGAACGAT 1250

35 4254 -----
4275 -----
4371 GGGACAGTTCATTAACCTCATTTGTAACCCGTTCTGAGACAACTAATGGTAAAT 1370

40 4254 -----
4275 -----
4371 GAGACAGTGAAACAAACGGCTGAATCATCTTTGATGTAAACAGAAGAGGAATTAAAGCTT 1430

45 4254 -----
4275 -----
4371 TTGAAATTTTGGAGTCACAGCCAATACTAATCAGTTTGGTGTATTGTATATCGGAT 1490

50 4254 -----
4275 -----
4371 ATACTTAATAATATCTAATTTCTTGATCTTTTCAGACACAAAATCATATGTTCAAACGA 1550

55 4254 -----
4275 -----
4371 AATGTCTTCCAAACACATATTTGTGATTATGACGAACACCACAATCTAATTGCTCCACC 1670

60 4254 -----
4275 -----
4371 AATAACATCAACACCTCTCAGCAAACCAATAAAGAAGAACCATTAACAGTAGAACC 1730

65 4254 -----
4275 -----
4371 TGCATGTATTCGTTTATTGCGTGAAAATCAAATTCGAATCCGACGACCTAGCTCG 1785

CLAIMS:

- 1) An arthropod male germline gene expression system suitable for conditional
5 expression of an effector gene in an Arthropod male germline, the system comprising:
- a first expression unit comprising an effector gene and a promoter therefor operably linked thereto;
 - a second expression unit comprising a coding sequence for a transcription factor and an upstream regulatory element operably linked thereto, the
10 transcription factor being capable of acting upon the promoter in the first expression unit to drive expression of the effector gene, the upstream regulatory element including:
 - a promoter for the transcription factor; and
 - a 5' UTR adjacent a start site for the transcription factor coding
15 sequence;
- the upstream regulatory element driving sufficient expression of the transcription factor such that the transcription factor protein in turn drives transcription of the effector gene before meiosis.
- 20 2) A gene expression system according to claim 1, wherein the transcription factor is a transcriptional activator, such as tTA, GAL4 or their variants.
- 3) A gene expression system according to any preceding claim, wherein the effector is endonuclease, most preferably a 3-Zn finger nuclease.
- 25 4) A gene expression system according to any preceding claim, wherein the promoter of the first expression unit is a minimal promoter.
- 5) A gene expression system according to any preceding claim, wherein the promoter
30 of the upstream regulatory element in the second expression unit is most from topi, aly or Beta-2 Tubulin (B2T) or homologues thereof.
- 6) A gene expression system according to any preceding claim, wherein the 5' UTR in the upstream regulatory element of the second expression unit is that from hsp83,

preferably from Medfly or homologues of hsp83, particularly the homologue of hsp83 found in the target arthropod.

5 7) A gene expression system according to any preceding claim, wherein the Gal4-UAS system is used, such that the transcription factor is GAL4.

8) A gene expression system according to any of claims 1-6, wherein the system is an inducible system, where induction occurs by provision of a chemical entity, such a tetracycline or one of its analogues including doxycycline.

10

9) A gene expression system according to claim 8, wherein the transcription factor in the second expression unit is tTA or a variant thereof (tTAV, tTAV2, tTAV3).

15

10) A gene expression system according to any preceding claim, wherein the transcription factor of the second expression unit is tTA or a variant and the first expression unit includes the tet operator (tetO).

20

11) A gene expression system according to any preceding claim, wherein the arthropod is an insect, preferably a Dipteran, including a Tephritid, such as a Medfly or Olive fly.

12) A gene expression system according to any preceding claim, wherein the effector is a nuclease and confers or imparts paternal effect lethality.

25

13) A method of expressing an effector protein in a gonad or sperm, comprising transforming the gonad with the expression system according to any preceding claim.

14) A method of population control comprising expressing the effector protein of the expression system according to any of claims 1-12 in the gonads of a male arthropod.

30

15) A method of resistance management comprising use of the expression system according to any of claims 1-12.

16) A method of quality control, comprising including a reporter, such as a fluorescent protein as, or in addition to, the effector protein in the expression system according to

any of claims 1-12, such that individuals where expression from the system has been induced or de-repressed become visible under suitable wavelengths of light.

- 5 17) A method of determining the mating status of a female arthropod, comprising use of the expression system according to any of claims 1-12 in a transgenic male population, wherein said system comprises a marker such as a fluorescent reporter protein; and where sperm is present, assaying for the presence of said marker in a female; the presence of the marker being indicative that the female has mated with a transgenic male carrying the system.