Imperial College London

PhD Project:

A functional study of the Y chromosome in the malaria vector *Anopheles gambiae*

Federica Bernardini

CID: 00693630

London, February 2015

Imperial College of Science, Technology and Medicine

Division of Cell and Molecular Biology

Supervisors: Professor Andrea Crisanti

Dr Nikolai Windbichler

Progress Review Panel: Professor Austin Burt

Professor George Christophides

I hereby declare that this project was entirely my own work and that any additional sources of information have been duly cited.

Abstract

Despite its function in sex determination and its role in driving genome evolution, the Y chromosome remains poorly understood in most species. Y chromosomes are gene-poor, repeat-rich and largely heterochromatic and therefore represent a difficult target for genetic engineering. The Y chromosome of the human malaria vector *Anopheles gambiae* appears to be involved in sex determination although very little is known about both its structure and function. The work described in this thesis led to the characterization of an *Anopheles gambiae* transgenic strain obtained by transposase-mediated integration of a transgene construct onto the Y chromosome. Using meganuclease-induced homologous repair a site-specific recombination signal was introduced onto the Y chromosome and the resulting docking line, named YAttP, was proven to allow secondary integration. To demonstrate its utility, the activity of a germline-specific promoter when located on the Y chromosome was studied. *Anopheles arabiensis* is another important vector of human malaria. Since active insertions onto the Y chromosome are extremely rare, a scheme based on crossing and selection was used to overcome F1 hybrid male sterility and introgress the modified *Anopheles gambiae* Y chromosome in the *Anopheles arabiensis* genetic background. Fertility of Y-introgressed males, tested after up to 10 backcross generations, was comparable to fertility of wild-type *Anopheles arabiensis* males. The molecular manipulation of the Y chromosome in *Anopheles gambiae*, opens up a number of ways to explore one of the most fascinating of evolution's upshots and its introgression in the *Anopheles arabiensis* genetic background may answer important questions on the similarity and differences in Y chromosome biology of closely related species. The Y-linked fluorescent transgenes allow automated sex separation of these important vector species, providing the means to generate large single-sex populations. Furthermore, the possibility of introducing genes of interest

specifically onto the Y chromosome makes these strains a valuable tool for vector control strategies.

Acknowledgments

The first person that I want to thank is Andrea, my boss. Without him these words and everything that has informed them throughout these years in London wouldn't have existed. I would love to mention all the amazing people I have been lucky enough to meet one by one; I am in debt to all of you. The need to be concise allows me to mention only a few. The first time I met Niki he drew me something on a piece of paper and said, "this is going to be your PhD project". That drawing (I still keep it) turned out to be my first scientific publication, for this and for his time I want to say thanks. A massive thanks goes to Tony for all his guidance, scientific and non, his trusting and his blues. Thanks to Miriam for all the talking and thanks to Roberto for all the listening, you guys made this a wonderful journey. Carla, Vale, Drew, Chryssa, Alekos and Lucy thanks for taking care of me, it really made me feel loved. My acknowledgment goes to all members of Crisanti lab, present and past, because they contributed in one way or another to the achievements of this project. Thanks to Austin and Karen, an inspiring example of tenacity and enthusiasm. The support of some people, even if not directly involved in the scientific development of this project, has been vital. Alessandra, thanks for keeping me at home, no matter where I am. People from Kin2000, you dressed my life with joy and glitter, I am grateful for this. Last but not least, Robert Nesta Marley, thanks for the positive vibrations.

Un immenso grazie va alla mia famiglia per essere paziente e in particolare ai miei nonni di cui piu' sento la mancanza.

Publications

Site-specific genetic engineering of the *Anopheles gambiae* Y chromosome

F. Bernardini, R. Galizi, M. Menichelli, P. A. Papathanos, V. Dritsou, E. Marois, A. Crisanti and N. Windbichler

PNAS 2014

A synthetic sex ratio distortion system for the control of the human malaria mosquito

R. Galizi, L. A. Doyle, M. Menichelli, F. Bernardini, A. Deredec, A. Burt, B. L. Stoddard, N. Windbichler and A. Crisanti

Nature Communication 2014

These papers are attached at the end of this thesis.

Some figures generated for these manuscripts have been used within this thesis.

Table of Contents

1 Introduction 1

1.1 Malaria 1

1.2 The Parasite 1

1.3 The vector 4

1.3.1 The *Anopheles gambiae* complex 6

1.3.2 *A. gambiae s.s.* and *A. arabiensis*, major vectors of human malaria 9

1.4 Malaria control strategies 11

1.4.1 Parasite control 11

1.4.2 Vaccines 12

1.4.3 Vector control 14

1.4.3.1 Environmental management 14

1.4.3.2 Insecticides 15

1.4.3.3 Sterile insect technique 16

1.4.3.4 Genetic modification of mosquitoes 17

1.4.3.5 Release of Insect carrying a Dominant Lethal 19

1.4.3.6 Gene drive systems for insect disease vectors 21

1.4.3.7 Homing endonuclease genes 22

1.4.3.8 Sex ratio distortion by disrupting sex chromosomes segregation 23

1.4.3.9 Sex ratio distortion by targeting Anopheles sex determination 25

1.5 The *Anopheles* Y chromosome 27

2 Aim of this study 29

3 Materials and Methods 30

3.1 Mosquitoes Rearing 30
3.2 Isolation of Y-specific sequences 30
3.3 Bacterial Cultures 31
3.4 Generation of I-SceI RNA 31
3.5 Generation of Transformation Vectors 31
3.6 Polymerase Chain Reaction (PCR) 32
3.7 Preparative Restriction Enzyme Digests 32
3.8 Agarose gel electrophoresis 33
3.9 Purification of DNA fragments from agarose gel 33
3.10 Dephosphorylation of vectors before ligation 33
3.11 Ligations 34
3.12 Transformation of TOP10 Chemically Competent cells 34
3.13 Colony Screening by PCR 34
3.14 Mini-preparation of plasmid DNA 35
3.15 DNA sequencing 35
3.16 Midi-preparation of plasmid DNA 35
3.17 Large-scale 'Maxi-Prep' preparation of plasmid DNA 36
3.18 Spectrophotometry of DNA 36
3.19 Embryo Microinjections 36
3.20 Confocal Microscopy 37
3.21 Fluorescence in situ hybridization (FISH) 37
3.22 Genetic crosses and fertility assays 38

4 Results 39

4.1 Engineering of the *A. gambiae* Y chromosome 39

4.1.1 The transgenic line T4 carries a transgene that segregates with the Y chromosome.....40

4.1.2 The Y-linked transgene is stable and can serve as a marker for automated sexing 44

4.1.3 Characterization of the genomic neighbourhood of the T4 transgene.....45

4.1.4 Identification of a Y-specific polymorphism 55

4.1.5 Generation of the Y-linked φC31 docking line YAttP.....59

4.1.6 Secondary integration of a germline-specific marker gene by site-specific recombination 64

4.1.7 The *vasa* regulatory region drives GFP expression during early male spermatogenesis from the Y chromosome 67

4.2 Introgression of a modified *A. gambiae* Y chromosome in the *A. arabiensis* genetic background.....69

4.2.1 Introgression attempt of YAttP and YI-PpoI chromosomes through F1 hybrid females generated from *A. gambiae* females mated with *A. arabiensis* males 72

4.2.2 Introgression of *YAttP* and *YI-PpoI* chromosomes through F1 hybrid females generated from *A. arabiensis* females mated with *A. gambiae* males.....81

4.2.2.1 Advanced introgression of the A. gambiae YI-PpoI chromosome.....87

4.2.2.2 Single backcrosses of YI-PpoI-introgressed males in cups.....88

4.2.2.3 Establishment of the YI-PpoI-introgressed line 90

4.2.2.4 Introgression of the YAttP chromosome.....91

4.2.2.5 Y-introgressed males backcrosses to A. gambiae females 91

5 Discussion 94

5.1 Engineering of the *A. gambiae* Y chromosome 94

5.2 Introgression of a modified *A. gambiae* chromosome in the *A. arabiensis* genetic background 100

6 Conclusions 109

7 Future plans 110

8 References 111

9 Appendix 119

9.1 Overview of transgenic lines generated 119

9.2 Single backcrosses of *YI-PpoI*-introgressed males in cups 120

9.3 List of primers 123

9.4 Scientific publications 126

Index of Figures

Figure 1: Malaria Parasite life cycle. 3

Figure 2: Mosquito life cycle. 5

Figure 3: Geographic distribution of members of the *Anopheles gambiae* complex in Africa. 8

Figure 4: Site-specific recombination mediated by the φC31 system. 19

Figure 5: Release of Insect carrying a Dominant Lethal (RIDL). 20

Figure 6: Mechanism of action of a Homing endonuclese gene (HEG). 23

Figure 7: Mechanism of action of the endonuclease I-PpoI. 25

Figure 8: Generation of the transgenic line T4. 42

Figure 9: Fluorescent pattern of T4 larvae. 43

Figure 10: Fluorescence in situ hybridization (FISH) of mitotic chromosomes in T4 line. 43

Figure 11: Flow cytometry analysis and automated sex separation of the T4 strain. 45

Figure 12: Identification of a Y chromosome insert carrying BAC and insert size determination. 47

Figure 13: PCR and RT-PCR analysis of the Y-specific ORF BigORF1. 51

Figure 14: PCR and RT-PCR analysis of the Y-specific ORF BigORF1. 52

Figure 15: RT-PCR analysis of the Y-specific ORFs ORF1 and ORF10. 53

Figure 16: PCR and RT-PCR analysis of the Y-specific ORFs ORF1 and ORF10. 54

Figure 17: PCR analysis of the non-Y-specific ORFe. 55

Figure 18: Alignment of the non-Y-specific ORFe from Contig 12 with the *A. gambiae* X chromosome. 56

Figure 19: PCR analysis of the non-Y-specific ORFe. 57

Figure 20: RT-PCR analysis of the non-Y-specific ORFe. 58

Figure 21: Generation of the transgenic line YAttp. 61

Figure 22: PCR analysis of the transgenic line YAttP. 63

Figure 23: Fluorescent patter of YAttP larvae. 63

Figure 24: Generation of the transgenic line YVasG. 65

Figure 25: Fluorescent pattern in YVasG larvae. 66

Figure 26: Confocal analysis of GFP expression in dissected testes of transgenic YVasG adult males. 68

Figure 27: *A. gambiae* and *A. arabiensis* karyotype. 71

Figure 28: *A. gambiae* females crossed to *A. arabiensis* males. 73

Figure 29: *F1 hybrid* females crossed to *A. gambiae transgenic* males. 74

Figure 30: *Class 1* males crossed to *A. arabiensis* females. 75

Figure 31: *Class 2* males crossed to *A. arabiensis* females (2nd backcross). *Class 2* males crossed to *class 1* females (cross 1). *Class 2* males crossed to *class 2* females (cross 2). 76

Figure 32: *Class 1 males crossed to class 2* females. 78

Figure 33: *Class 3 males crossed* to *A. arabiensis* females. 79

Figure 34: *Class 4 males* crossed to *A. arabiensis* females. 80

Figure 35: *A. arabiensis* females crossed to *A. gambiae* males. 82

Figure 36: *F1 hybrid** females crossed to *A. gambiae transgenic* males. 83

Figure 37: *Class 1** males crossed to *A. arabiensis* females. 84

Figure 38: *Class 2** males crossed to *A. arabiensis* females (2nd backcross). *Class 2** males crossed to *class 1** females (cross 1). *Class 2** males crossed to *class 2** females (cross 2). 85

Figure 39: Deposition rate in single backcrosses in cups. 89

Figure 40: Hatching rate of single backcrosses in cups. 89

Figure 41: PCR analysis of *YI-PpoI*-introgressed line. 90

Figure 42: Y-introgressed males crossed to *A. gambiae* females. 92

Figure 43: *Class 5* males crossed to *A gambiae* females. 93

Figure 44: Association between different classes of males and fertility/sterility phenotype. 102

Figure 45: Association between different classes of males and fertility/sterility phenotype. 105

Index of Tables

Table 1: PCR analysis of selected open reading frames (ORFs) 50

Table 2: Embryo microinjections and fluorescent phenotypes observed in the outcross progeny.....62

Table 3: Embryo microinjections and fluorescent phenotypes observed in the outcross progeny.....66

Table 4: Hatching rate recovered from crosses described in Figure 30. 77

Table 5: Hatching rate recovered from crosses described in Figure 32 and 33. 80

Table 6: Hatching rate recovered from crosses described in Figure 37. 86

Table 7: Hatching rate recovered from backcrosses of *YI-PpoI*-introgressed males to *A. arabiensis* females in cage. 87

1 Introduction

1.1 Malaria

In 2700 B.C. ancient Chinese medical writings described for the first time the symptoms of what would later be named malaria, from the Italian for "bad air" (*mal'aria*). Malaria is an infectious disease presently considered endemic in 104 countries and territories with populations living in Sub-Saharan Africa having the highest risk of infection. Every year, 300-500 million people are infected by malaria and more than 1 million people die as a consequence of *Plasmodium* parasite infections (WHO, 2012).

1.2 The Parasite

Malaria is caused by members of the genus *Plasmodium. Plasmodium* species are Apicomplexa and exhibit a heteroxenous life cycle involving a vertebrate host and an arthropod vector. Five different species of *Plasmodium* infect humans: *Plasmodium falciparum, Plasmodium vivax, Plasmodium malariae, Plasmodium ovale* and *Plasmodium knowlesi* (Sherman, 1998; Singh and Daneshvar, 2013). *Plasmodium falciparum* is responsible for the most severe cases of human malaria. During the blood meal the mosquito ingests the male and female parasite gametocytes, micro- and macrogametocytes respectively, circulating in the bloodstream of an infected person. Ingestion of gametocytes induces gametogenesis which involves a process known as exflagellation that leads to the formation of microgametes, flagellated forms that fertilize the macrogamete to form the zygote. The zygote develops into a motile elongated ookinete which penetrates the gut epithelial cells where it matures into an oocyst. The oocyst undergoes multiple divisions that result in the production of immature sporozoites. Upon rupture of the mature oocyst the sporozoites are

released into the hemocoel, from this cavity they migrate to invade the salivary glands. In the salivary glands the sporozoites continue their maturation until inoculation into a new host when a blood meal is taken by the mosquito. Once in the human host the sporozoites migrate to the liver and invade the hepatocytes where they develop into exoerythrocytic schizonts. During this stage the parasite undergoes an asexual replicative process to form merozoites. Following the hepatocytes rupture, merozoites are released into the circulatory system and invade the erythrocytes where they undergo a throphic period followed by an asexual replication, this is known as the erythrocytic schizont stage. Newly formed merozoites are released upon ruptures of the erythrocytes. The simultaneous burst of the infected erythrocytes and the release of antigens and waste products are responsible for the intermittent fever paroxysms associated with malaria. As an alternative to schizogony some of the parasites undergo a sexual cycle and differentiate into either micro- or macrogametocytes. Gametocytes are ingested by the mosquito during the blood meal (Figure 1).

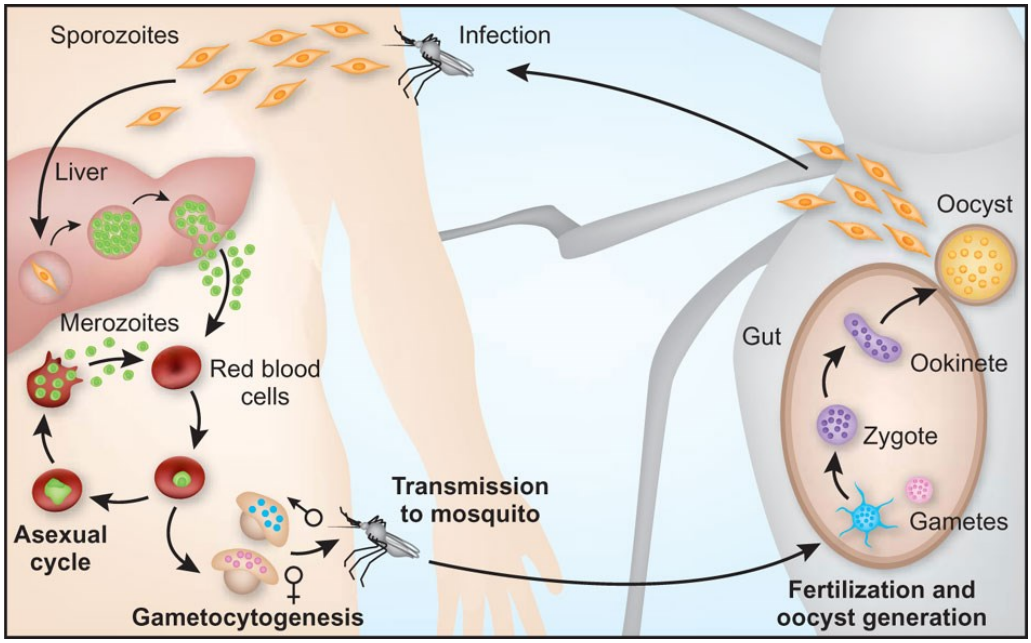


Figure 1: Malaria Parasite life cycle.

Parasite gametocytes are ingested by the mosquito during its blood meal on an infected person. In the mosquito gut the gametocytes mature in gametes, this is followed by fertilization e formation of the zygote. The zygote develops in a motile, elongate oockinete that penetrates the gut epithelial cells where matures in an oocyst. Multiple divisions of the oocyst result in the formation of sporozoites that migrate through the hemocoel in the salivary glands. Sporozoites are injected in the human bloodstream during the mosquito blood meal. Sporozoites invade the liver and, through an asexual replicative process, merozoites are generated. Following the hepatocytes rupture, merozoites are released in the circulatory system and invade the erythrocytes. New merozoytes are generated and released upon rupture of the blood cells. Alternatively some parasites undergo a sexual cycle and differentiate in gametocytes that are ingested by the mosquito during its blood meal (Pasvol, 2010).

1.3 The vector

The exclusive vector of human malaria is represented by mosquitoes belonging to the genus *Anopheles* (Davidson, 1964a). Mosquitoes undergo separate and distinct stages during their life cycle: egg, larva, pupa, and adult (Figure 2). The eggs, laid by the females in stagnant water, hatch into larvae that live in the water and come to the surface to breathe. Larvae grow until they are able to change into pupae that undergo a resting, non-feeding stage. These three stages can last from 5 to 14 days, depending on the species and the environmental conditions i.e. food and temperature. When development is complete, the pupal skin splits and the mosquito emerges as an adult. A few days after emerging from the pupal shell, adult mosquitoes mate. The mating, in most species, takes place around dusk and requires the males to form swarms into which the females fly to mate. Adults mosquitoes feed on the nectar of plants, only females need a blood meal for the full development of the eggs. After mating and blood feeding, female mosquitoes rest for a few days until they are ready to lay the eggs. While in captivity adult mosquitoes can live up to a month, in the wild the lifespan it is estimated to be around 2 weeks (Clements, 1992).

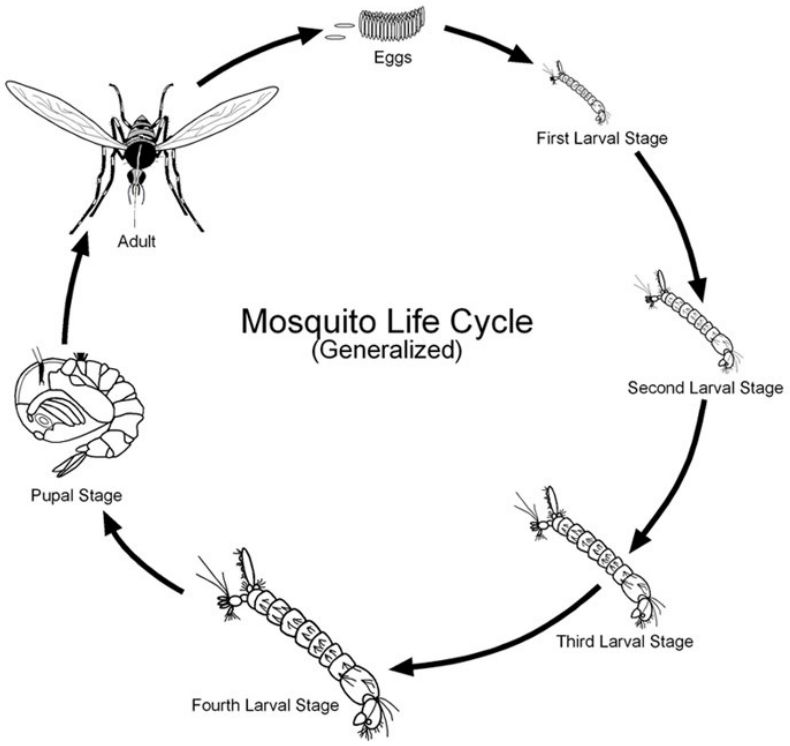


Figure 2: Mosquito life cycle.

After mating and blood meal eggs are laid by female mosquitoes in stagnant water. Eggs hatch into larvae that grow and develop into pupae. When development is complete the pupal skin splits and the mosquito emerges as an adult (Catherine A. Hill, 2008).

1.3.1 The *Anopheles gambiae* complex

The genus *Anopheles* includes 465 formally recognized species of which approximately 70 transmit human malaria parasites. *Anopheles gambiae* is a complex that includes eight sibling species that represent the most widespread and potent vectors of malaria in Sub-Saharan Africa: *A. gambiae sensu stricto* (*A. gambiae s.s.*), *A. arabiensis*, *A. bwambae*, *A. melas*, *A. merus*, *A. quadriannulatus*, *A. amharicus*, *A. comorensis* (White, 1985; Hunt *et al.*, 1998; Harbach, 2012; Coluzzi *et al.*, 2002; Coetzee *et al.*, 2013). The species in the complex were identified based on morphological features (Brunhes J, 1997), fixed differences in chromosomal inversions (Davidson G, 1973; Coluzzi *et al.*, 1979) and demonstration of F1 hybrid sterility in crosses between populations (Coetzee *et al.*, 2013; Davidson G, 1973; Davidson, 1956; Davidson, 1964a, Davidson, 1964b). Species of the complex differ in their ecology and geographic distribution (Figure 3). *A. merus* and *A. melas* are salt-water breeders and so their distribution is restricted to sites along the East and West African coasts respectively. The vectorial efficiency of these species in terms of malaria transmission is considered to be low, adult females bite human beings generally in absence of alternative hosts. A third salt-water species, *A. bwambae*, spreads in hot springs in Semliki Forest National Park in Eastern Uganda. *A. quadriannulatus* and *A. amharicus*, distributed in South-Eastern Africa and *A. amharicus*, in Ethiopia (Hunt *et al.*, 1998; Coetzee *et al.*, 2013), show a marked zoophily and therefore are not considered to be involved in the transmission of human malaria. *A. comorensis* occurs in the island of Grande Comore in the Indian Ocean and was described as a distinct species on the basis of morphological characters. The two remaining fresh-water species, *A. gambiae s.s.* and *A. arabiensis*, have the widest geographic distribution and show a marked anthropophily, these features make these two species the most important vectors of human malaria (Gillies M, 1968; Coetzee *et al.*, 2000). Populations of *A. gambiae*

s.s. are thought to be undergoing speciation (Hunt *et al.*, 1998; Harbach, 2012; Coetzee *et al.*, 2013). A number of subpopulations have been characterized based on chromosomal polymorphisms in the form of paracentric inversions (five chromosomal forms identified) or based on a single base substitution at the 28S rDNA sequence (two molecular forms identified). Recently, of the two molecular forms, M and S, the M form has been elevated to the species status and renamed *A. coluzzii* (Coetzee *et al.,* 2013). It has been shown that inversions occur in a non-random manner along the chromosomes arising more often on the right arm of the chromosome 2 (2R) (Pombi *et al.*, 2008). This observation, together with the inversions frequency throughout the geographical distribution of the species, suggest that these chromosomal polymorphisms might be maintained by selection that allows different species to survive and exploit a variety of habitats (Toure *et al.*, 1998; Coluzzi *et al.*, 2002; Lee *et al.*, 2009). Genetic divergence, enhanced by reduced recombination associated with inversions, will progressively increase between populations leading eventually to reproductive isolation. Nevertheless, latest studies on *A. gambiae s.s.* molecular forms suggested that 3 small regions located near the centromere of the X, 2L and 3L chromosomes contain genes directly involved in reproductive isolation, based on this observation these regions have been referred as islands of speciation (Wu, 2001; Wu and Ting, 2004), gene flow between the two populations is thought to be responsible for low levels of divergence over the rest of the genome (97%). An alternative model, incidental islands of divergence, suggests that divergence between the two molecular forms is extended to a large portion of the genome and that gene flow between these two forms is really low (Turner *et al.*, 2005; Hahn *et al.*, 2012). The validation of a speciation model requires a deep analysis of the genetic structure of the mosquito populations and a robust method to analyse the frequency of hybrid individuals in order to estimate the gene flow between members of the complex.

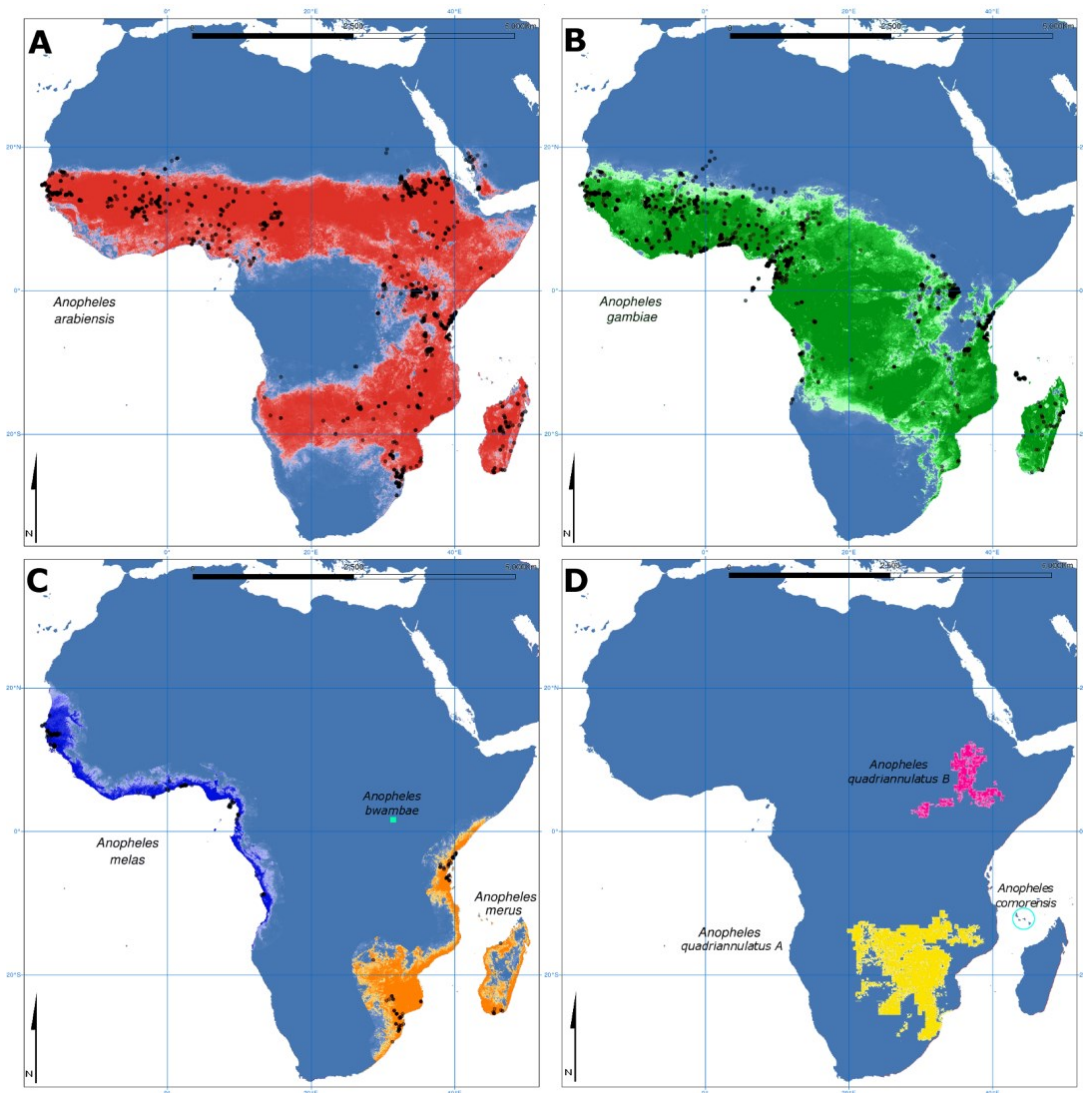


Figure 3: Geographic distribution of members of the *Anopheles gambiae* complex in Africa.

Distribution in Africa of the *A. gambiae* complex is shown. A: *A. arabiensis* (red); B: *A. gambiae s.s.* (green); C: *A. melas* (blue), *A. merus* (orange) and *A. bwambae* (cyan); D: *A. quadriannulatus* (yellow), *A. amharicus* (magenta) and *A. comorensis* (cyan circle) (Lee, 2013).

1.3.2 *A. gambiae s.s.* and *A. arabiensis*, major vectors of human malaria

A. gambiae s.s. (from here after referred to as *A. gambiae*) and *A. arabiensis* are species of primary medical importance as they are the most anthropophilic members of the *A. gambiae* complex and have the widest distribution. They show different distributional perimeters and asynchronous seasonal prevalence due to divergent responses to climatic factors. *A. gambiae* predominates in zones of forest and humid savannah whereas *A. arabiensis* prevails in arid savannahs and steppes, including those of the South-Western part of the Arabian Peninsula. In the sympatric areas, occupied by both the species, change in seasonal prevalence are observed showing an increase in the relative frequency of *A. arabiensis* during the dry season. *A. gambiae* and *A. arabiensis* are strictly fresh-water breeders and they share a similar larval ecology. This species, morphologically indistinguishable, differ by multiple fixed chromosomal rearrangements on the X chromosome but share three inversions on chromosome 2. These arrangements, classified in relation to the standard polytene sequence found in *A. quadrianilatus*, are known as 2La, fixed in *A. arabiensis* and polymorphic in *A. gambiae*, and 2Rb and -c, polymorphic in both species (Coluzzi *et al.*, 1979). In areas where the distribution of *A. gambiae* and *A. arabiensis* overlaps, hybrids are present at really low frequency (0.02-0.76%) (Toure *et al.*, 1998; Temu *et al.*, 1997; Mawejje *et al.*, 2013). It has been suggested that pre-zygotic isolation mechanisms occur in the wild. Cues used for species recognition are still unclear but they might involve differences in the wing beat frequencies in the two species (Gibson *et al.*, 2010). Post-zygotic isolation mechanisms are also known. F1 hybrids obey Haldane's rule (Haldane, 1922; Schilthuizen *et al.*, 2011) with males, the heterogametic sex, being sterile and females, the homogametic sex, being fertile. Inviability has also been reported in hybrids between *A. gambiae* and *A. arabiensis*, Slotman *et al.* suggested that this is likely due to recessive factors on the X chromosome of *A. gambiae* incompatible with at

least one factor on each autosome in *A. arabiensis* (Slotman *et al.*, 2004). The same authors suggested that the X chromosome has a disproportionately large effect on male hybrid sterility (Slotman *et al.*, 2005). Low frequency of hybrids between the two species has been recently reported by a study conducted on samples collected in Eastern Uganda. The authors of this study showed that most of the samples were hybrids generations beyond F1 highlighting the role of gene flow between species in leading to adaptive introgression and the implication of this for vector control (Weetman *et al.*, 2014).

1.4 Malaria control strategies

1.4.1 Parasite control

The targeting of the human stages of the malaria parasite by using chemotherapeutic drugs has been proven to be quite efficient for the reduction of the disease, especially in many parts of the Western World. Antimalarials such as quinine, made from a plant found in the rainforests in South America, have a long history of use, dating back to the 17th century. After the 1940s, other drugsbased on quinine derivates, such as chloroquine, with fewer unpleasant side effects, have been used as antimalarials for prevention and treatment. Introductions in the latter years of the 20th century included proguanil, sulfadoxine-pyrimethamine (SP), amodiaquine, mefloquine and atovaquone. However, these drugs exert a dramatic selection pressure on the parasites that results in the development of resistance (Peters, 1987). In the late 1950's, after about 5 years of mass application in the Global Malaria Eradication Programme, resistance has been reported for chloroquine in South East Asia, it quickly spread to other areas in Asia and then invaded Africa in the following three decades. SP has been implemented for just a year before resistance emerged. Similar fortune was reported for mefloquine, atovaquone and progunil from South East Asia, where these drugs were more commonly employed (Breman *et al.*, 2004; Petersen *et al.*, 2011). Drugs derived from artemisinin, a Chinese herb, are the most effective treatments for malaria at present. Since this class of drugs has a short *in vivo* life, they can be effectively used in combination with other long acting drugs such as lumefantrine, amodiaquine, mefloquine, piperaquine, pyronaridine or, in certain areas, SP. These therapies, comparatively expensive, are known as artemisinin-based combination therapies (ACTs). Yet resistance against artemisinin has been reported in South East Asia and more recently in Western Cambodia, Thailand, Vietnam,

Eastern Myanmar and Northern Cambodia, and there are signs of resistance emerging in Central Myanmar, Southern Laos and North-Eastern Cambodia (Walker *et al.*, 2000; Uhlemann *et al.*, 2005; Jambou *et al.*, 2005). History suggests that drugs employment and development is undoubtedly vital for controlling the impact of malaria on human health, yet it is unlikely to be a long-term solution to the problem.

1.4.2 Vaccines

Malaria vaccines are an area of excessive research. However, despite many years of effort, there has no effective vaccine been introduced into clinical practice, yet. As the use of an effective vaccine remains an elusive goal for every national malaria control programme, a great number of leads are emerging, with three vaccine strategies (Matuschewski, 2006):

1) Pre-erythrocytic, liver stage vaccines.

2) Erythrocytic, blood stage vaccines.

3) Transmission blocking, mosquito stage vaccines.

The observation that low numbers of sporozoites invade the liver, due to the natural bottleneck caused by epidermal injection, has encouraged the targeting of liver stage sporozoites as an appealing approach. Pre-erythrocytic vaccines are designed to prevent the entry of sporozoites into hepatocytes or destroy infected hepatocytes, by evoking protective immune responses that do not normally develop in natural infection, thereby preventing clinical disease. This approach was initially applied in animal models of malaria by vaccination with radiation-attenuated sporozoites (Herrington *et al.*, 1991; Hoffman *et al.*, 2002), the recent generation of genetically engineered, live attenuated sporozoites by a defined single gene knock-out, is considered to be safer for further application methods (Mueller *et al.*, 2005;

Van Dijk *et al.*, 2005). A more cost-effective prospect, compared to that of a whole organism vaccine, is the development of a potential subunit vaccine. The most developed and promising subunit vaccine candidate, targeting the liver stages of *Plasmodium* development, is RTS,S. This candidate vaccine aims at protection achieved by immunization with the circumsporozoite (CS) protein (Casares *et al.*, 2010) and is currently being evaluated in a phase III trial in Africa (Crawley *et al.*, 2010). Vaccines against the pathogenic asexual blood stage of *Plasmodium* parasites are designed with the aim of preventing clinical disease by destroying the merozoites. Since protective semi-immunity to malaria is acquired after repeated *Plasmodium* infections, it may be possible to mimic and stimulate the acquisition of naturally acquired immunity by a vaccine (Crompton *et al.*, 2010). However, relatively few blood stage antigens are in clinical development as vaccines, so far (Crompton *et al.*, 2010). These include erythrocyte-binding antigen–175 (EBA-175) (El Sahly *et al.*, 2010), serine-repeat antigen 5 (SERA5) (Horii *et al.*, 2010), apical membrane antigen 1 (AMA1) (Sagara *et al.*, 2009), glutamate-rich protein (GLURP) (Hermsen *et al.*, 2007) and merozoite surface protein 1 (MSP1) (Ogutu *et al.*, 2009), MSP2 (Genton *et al.*, 2003) and MSP3 (Druilhe *et al.*, 2005), all of which are highly expressed on the surface of the merozoite. Unfortunately, phase II trials of the most advanced blood stage candidates, AMA1 and MSP1, did not represent efficacy in African children. Extensive *P. falciparum* polymorphism variation in parasite surface proteins, means that generation of an efficient blood stage vaccine against this species is much more challenging. Transmission blocking vaccines target the intravector stages of *Plasmodium* development. The immune system could be induced to develop antibodies able to prevent the gametocytes from producing multiple sporozoites in the mosquito gut wall. Candidate antigens have been identified and approved for clinical trials (DeWeerdt, 2012). However, the application of these vaccines, used alone, could, at best, lead to the interruption of malaria

transmission only in areas with low transmission rate (Stowers and Carter, 2001). If malaria is to be eradicated a broader range of small molecule therapeutics, able to target various stages in *Plasmodium*'s life cycle, are required. Thus, interest in the development of a multi-antigen, multistage vaccine, is now increasing (Butler, 2009). This approach will possibly offer better protection than do single-antigen vaccines, and avoid vaccine failure caused by genetic diversity in parasite populations (Saul and Fay, 2007).

1.4.3 Vector control

The life cycle of *Plasmodium* is strictly linked to the mosquitoes, as such, a strategy employed to impair the vectorial capacity of mosquitoes would break the chain of transmission as effectively as a vaccine or antimalarial drug would. Discussed below are a number of different strategies currently employed or envisaged for vector control.

1.4.3.1 Environmental management

Environmental management for vector control aims to prevent or minimise vector propagation and reduce man-vector-pathogen contact. The distribution of vector-borne diseases is strictly linked to the ecological requirements of the local vector species. In the contest of malaria, the aquatic environment represents a critical factor for the mosquito life cycle. Therefore, environmental management for malaria control frequently aims at introducing changes in the local hydrology or in water-use practices (Randell *et al.*, 2010). This can be achieved by environmental modification or environmental manipulation. Modification implies permanent changes such as landscaping, drainage, land reclamation and filling. Manipulation refers to activities that reduce larval breeding sites through temporary changes

such as removing aquatic weeds from irrigation and drainage canals. The realization of such procedures, especially for environmental modification, requires significant capital investment. Environmental management for vector control is not intended to replace other control strategies but it provides a basis for other methods, such as chemical control, to build on in a complementary fashion, while reducing the environmental costs and resistance risks incurred, for example, by excessive use of insecticides.

1.4.3.2 Insecticides

The use of insecticides represents an essential component in the control of vector borne diseases (Townson *et al.*, 2005). Indoor residual spraying (IRS) and in particular pyrethroids and the most well-known dichlorodiphenyltrichloroethane (DDT), have been used since the 1940s in a number of malaria infected areas resulting in considerable reductions in disease morbidity and mortality (WHO, 1995). IRS combined to the subsequently developed insecticide-impregnated bed nets (ITNs) and long-lasting insecticidal nets (LLINs) represent the most prominent malaria preventive measure for large-scale deployment in highly endemic areas. It has been estimated that 5% of the population at risk is protected by indoor residual spraying (IRS) (Phillips, 2001) and due to their efficacy, acceptance and low cost, between 2000 and 2012, the proportion of households owning at least one bed net in Sub-Saharan Africa increased from 3% to 53% (WHO, 2012). Despite the historical optimism about the use of insecticides, ITNs and LLINs, there has been increasing debate concerning the potential for environmental damage and resistance in the target vectors (Curtis *et al.*, 2003; Maxwell *et al.*, 2002). Resistance mechanisms develop predominantly though both evolution of metabolic insecticide detoxification and target site mutation (Martinez-Torres *et al.*, 1998). Mosquito

resistance to public health insecticides has been identified in 64 countries around the world, affecting all major vector species and all classes of insecticides (WHO, 2012). The problems associated with the use of insecticides, ITNs and LLINs has led to encouragement in developing alternative methods of malaria control.

1.4.3.3 Sterile insect technique

In 1955 Dr. R.C. Bushland and Dr. E.F. Knipling laid the foundations for a new strategy for vector control, the sterile insect technique (SIT) (Knipling, 1955). Unlike insecticides, SIT has the advantage to be an environment-friendly species-specific vector control strategy. SIT involves mass-releases of sterile males, commonly generated by low doses of radiations, over infested areas. These males, sexually active but genetically sterile, compete with native males for mating with females leading to a decline in female's reproductive potential. Sequential releases of large number of sterile males (10 to 100 times the number of native males) cause a progressive increase of the ratio of sterile to normal insects, this leads to the decrease of the native insect population and eventually to its extinction. SIT has been successfully employed for the eradication of the New World screwworm fly, *Cochliomyia homnivorax*, from Libya, the Southern States of the USA, Mexico and all of Central America (Krafsur *et al.*, 1987; Lindquist *et al.*, 1992). The same approach led to the drastic reduction of populations of Mediterranean fruit fly, an economically important fruit pest, in large areas of Central America (Curtis, 2005). Few suppression and/or elimination programmes have been attempted also on mosquito populations, an example is given by the elimination of an isolated *A. albimanus* population in El Salvador (Lofgren *et al.*, 1974; Benedict and Robinson, 2003). Nevertheless, the lack of a reliable high-throughput sexing methods, necessary to guarantee the male-

exclusive releases, the loss of male fitness that arises from the sterilization methods, and the immigration of mated females into release areas, raise a number of technological problems that affect the success of this technique in large scale programmes.

1.4.3.4 Genetic modification of mosquitoes

The medical importance of vectors of human pathogens has led to the development of new approaches to investigate their gene structure and function. The use of genetically modified mosquitoes (GMM) for malaria control has gained worldwide recognition in the past 20 years. Transposable elements such us *piggyBac*, *Hermes* and *Mariner* can be used to deliver genes of interest in a genome host (Calos and Miller, 1980; Munoz-Lopez and Garcia-Perez, 2010). This is possible by microinjection-based germline transformation technique (Lobo *et al.*, 2006). Co-injection of embryos with a source of *transposase* in trans, such as an helper plasmid or RNA, enables the transposone to move into the host genome. The identification of individuals carrying the transgene is made easier by the incorporation of fluorescent markers, such as green, red and cyan fluorescent proteins (GFP, RFP and CFP), into the transgenic construct (Berghammer *et al.*, 1999). The availability of molecular tools for site-specific integration such as the phage φC31 system has led to the possibility of target transgenes on specific chromosomal locations. The φC31 system requires co-injection of embryos with a plasmid carrying a gene of interest and a source of the phage φC31 *integrase* in trans. This enzyme catalyses the recombination reaction between attachment sites called *AttP* and *AttB*. One of these sites is inserted into a known chromosomal location while the other is on the plasmid carrying the gene of interest. If recombination between the attachment sites occurs, the gene of interest will be inserted onto a specific location of the genome and hybrids sites,

called *attL* and *attR*, will be generated (Thorpe and Smith, 1998) (Figure 4). Since the generation of the first transgenic mosquito (Miller *et al.*, 1987), this technique has been largely improved and, with the availability of genome sequences (Holt *et al.*, 2002), it has enormously increased the knowledge about the biology of anopheline mosquitoes and thus offered the possibility of intervention to impair their capacity as vectors of malaria parasite (Catteruccia *et al.*, 2000; Jasinskiene *et al.*, 1998; Kokoza *et al.*, 2001; Coates *et al.*, 1998; Perera *et al.*, 2002; Grossman *et al.*, 2001; Allen *et al.*, 2001; Handler, 2001). Vector control strategies based on GMM can be grouped in two broad categories: population suppression or population replacement. Release of transgenic mosquitoes for population replacement aims to convert a wild population into one whose phenotype would be impaired in parasite transmission. A fundamental step toward the application of such strategies has been the understanding of the biological interactions between the mosquito and the parasite. This has resulted in the identification of key immunity genes and immunity pathways that affect levels of parasitemia (Christophides *et al.*, 2004; Mendes *et al.*, 2008; Michel and Kafatos, 2005; Osta *et al.*, 2004; Waterhouse *et al.*, 2007; Ghosh *et al.*, 2009; Ghosh *et al.*, 2009; Marrelli *et al.*, 2007). In contrast to population replacement, GM strategies focusing on population suppression aim to use transgenic mosquitoes to reduce or eradicate the wild population by negatively affecting its fitness. In this scenario transgenic technologies have been proven to benefit existing technique such as the previously mentioned sterile insect technique. The lack of a reliable high-throughput sexing method required for the success of this technology, could be overcome by the generation of mosquito strains that express fluorescent markers, such as GFP and RFP, under the control of sex specific promoters (Catteruccia *et al.*, 2005) or alternatively these markers could be inserted on to the sex chromosomes.

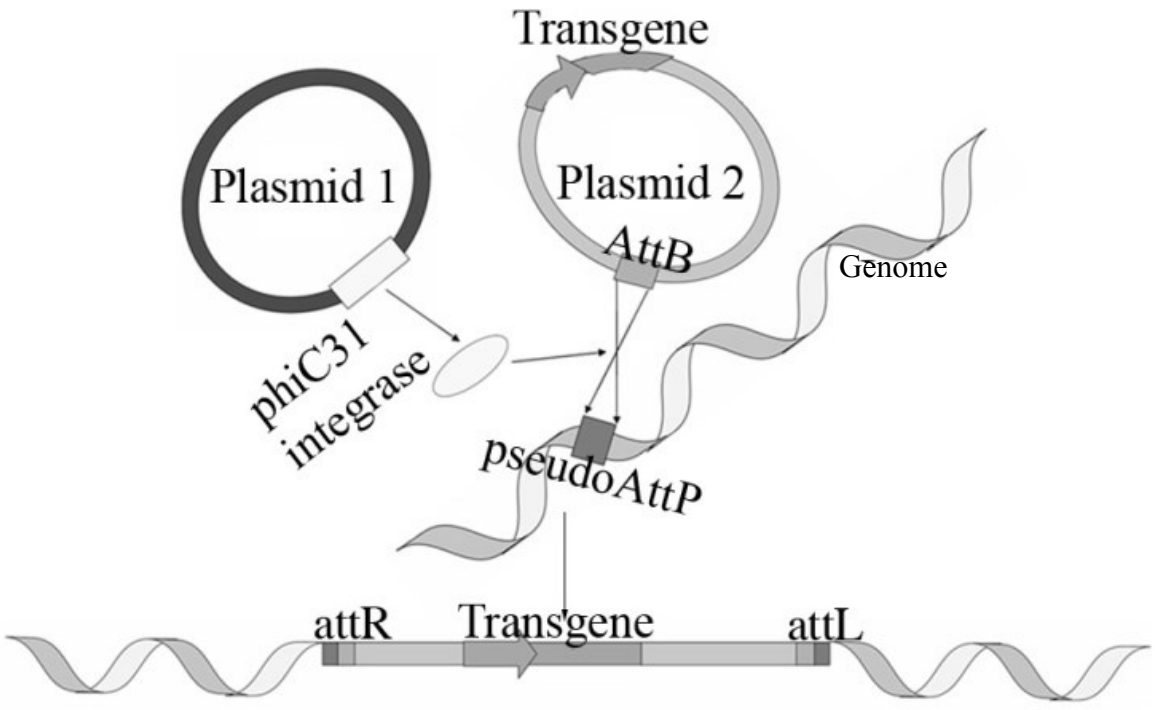


Figure 4: Site-specific recombination mediated by the φC31 system.

When mosquito embryos are co-injected with a plasmid carrying a transgene and an *AttB* site (Plasmid 2) and a source of the phage C31 integrase (Plasmid 1), specific integration of the transgenic construct in a chromosomal region where an *AttP* site is inserted, occurs and hybrid sites called *attL* and *attR* are generated (Francisco Martin, 2011).

1.4.3.5 Release of Insect carrying a Dominant Lethal

An alternative strategy for vector control would be to select against females by expressing lethal genes in a female-specific manner. Based on this approach a modified version of sterile insect technique has been developed, the release of insect carrying a dominant lethal (RIDL) (Thomas *et al.*, 2000) (Figure 5). The lethal gene has to be expressed only in the final generation of mosquitoes that is released such us their rearing in the insectary will not be affected. In order to conditionally express the lethal gene the tetracycline-repressible expression system (Tet Off system) can be used (Gossen and Bujard, 1992). This is a two component system that requires the expression of the *transactivator* (tTA) under the control of a promoter of choice and an effector gene under the regulatory sequences (*tetO*) recognized by the transactivator. The presence of tetracycline in rearing conditions impedes

the binding of the transactivator to the *tetO* sequence so the effector gene will not be expressed. The absence of tetracycline in the wild allow the lethal gene to be expressed thus RIDL can be applied to suppress a target population. Despite the potential for population suppression offered by sterile insect technique and RIDL several fundamental problems still persist. Both the techniques require in fact mosquitoes mass rearing and repeated releases (Benedict and Robinson, 2003). The effectiveness and success of techniques that aim population replacement of eradication rely on their "invasive" character, in fact the lethal gene or any other genetic factor employed for vector control needs to increase in frequency within the population after the release. This invasive character is called gene drive.

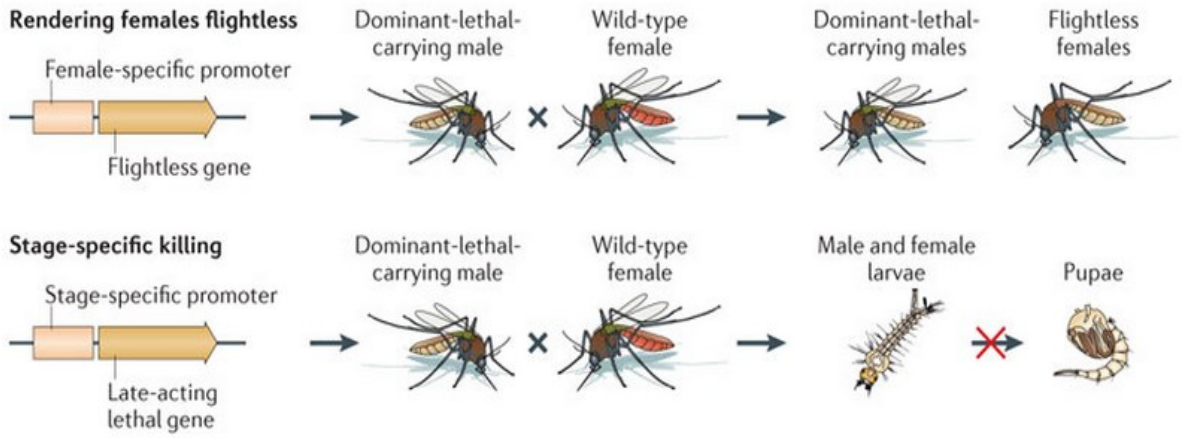


Figure 5: Release of Insect carrying a Dominant Lethal (RIDL).

Two examples of application of RIDL are given. In the first scenario, males carrying a female-acting transgene that results in a loss of flying ability are released in the open field. These males mate with wild-type females, and the resulting female offspring are flightless and, hence, unable to mate or find human hosts. In the second scenario, males carrying a transgene that causes late-acting lethality are released in the open field. These males mate with wild-type females, and the resulting offspring die as pupae (shown) or adults (McGraw and O'Neill, 2013).

1.4.3.6 Gene drive systems for insect disease vectors

Selfish genetic elements respond to the necessity of factors with an invasive character for gene drive in a target population. These elements show a frequency of inheritance greater than the Mendelian ratio and are known to spread within a population even when their presence does not confer any advantage for the host organism (Hurst and Werren, 2001). The selfish elements known so far belong to different categories such as transposable elements, meiotic drive genes, homing endonuclease genes and *Wolbachia* (Sinkins and Gould, 2006). For these elements to be good candidates for gene drive systems they need to respond to some criteria (Braig, 2002; James, 2005). One feature of great importance for the drive mechanism is its ability to spread within the population in a time scale that suits the vector control strategy applied (Boëte and Koella, 2003). Another important feature is the persistence of the drive mechanism within the population, the linkage between the drive mechanism and the effector gene to be driven has to be as resistant as possible to potential loss. The possibility to use a combination of effector genes might overcome the problematic that could arise due to the development of mechanisms of resistance. Given that malaria mosquito vectors belong to a complex and that within this complex there are at list two major vectors, a drive system that can be used to drive effector genes in sibling vector species would be of great interest. Nevertheless, the drive mechanism needs to be specie-specific to avoid any effect on non-target species and prevent any ecological case. Ideally, in case of an unanticipated negative effect of the drive system applied, the possibility of removal of the effector gene from the population might result in an advantage. Once candidate selfish genes have been identified their efficacy needs to be evaluated in small field trial populations before releases in the wild.

1.4.3.7 Homing endonuclease genes

Homing endonuclease genes (HEGs) are selfish genes that encode highly specific enzymes that recognize and cleave a long target sequence (20–30 bp) (Chevalier and Stoddard, 2001; Stoddard, 2005). The recognition site is characteristic of chromosomes that lack the HEG, in fact the HEG itself is inserted in the middle of its own recognition sequence, and so chromosomes carrying the HEG are protected from the endonuclease recognition and cleavage. In a scenario where the HEG is in heterozygosis with the wild type allele, the endonuclease will cut the intact recognition site present in the chromosome lacking the HEG. The broken chromosome will be repaired by the cell's recombinational repair machinery and the intact homologue chromosome, harbouring the HEG, will be used as a template. After the repair, both chromosomes will contain a copy of the HEG, the heterozygote will be converted into a homozygote (Figure 6). This mechanism of action of HEGs is called homing and results in dominant transmission and inheritance of the HEG (super Mendelian ratio). Naturally occurring HEGs do not compromise gene function when integrated in the genome because their recognition site is associated to intronic sequences. The understanding of the mechanism by which HEGs spread opened up new possibilities for vector control strategies that aim at population suppression or population refractoriness. A HEG can be engineered in a way that the endonuclease recognizes a sequence in the middle of an essential target gene. The target gene is chosen so that the presence of the HEG in heterozygosis is not deleterious for the organism. The HEG is driven by a germline-specific promoter so that the heterozygous zygote develops normally, but during meiosis there is a super Mendelian transmission of the HEG in the gametes. If such a construct is introduced at low frequency into a population, then initially it will appear mostly in the heterozygous state, and so it will show transmission-ratio

distortion (TRD) but few harmful effects. It will therefore increase in frequency, until it reaches an equilibrium frequency at which the harmful effects balance the TRD (Burt, 2003).

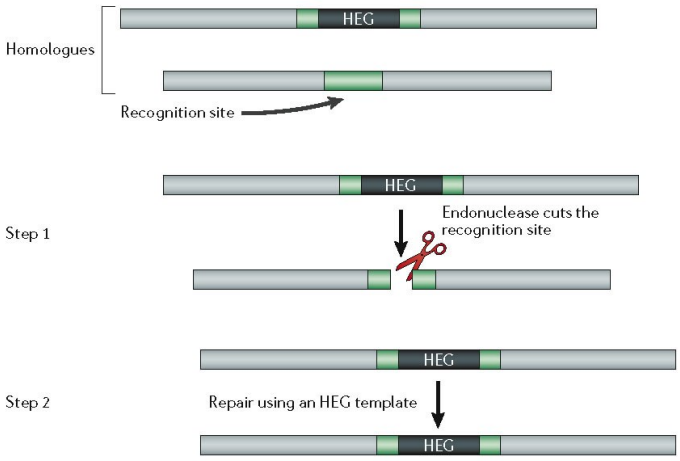


Figure 6: Mechanism of action of a Homing endonuclese gene (HEG).

The HEG in heterozygosis produces an endonuclease that cleaves a specific recognition sequence in a homologous location of the genome. This gap is then repaired by the cell's repair machinery that uses the homologous chromosome, where the HEG is inserted, as a template leading to another copy of the HEG and therefore, to homozygosis (Sinkins and Gould, 2006).

1.4.3.8 Sex ratio distortion by disrupting sex chromosomes segregation

It has been suggested that population suppression or elimination could be achieved by biasing the sex ratio towards males (Hamilton, 1967). In *Aedes* and *Culex* mosquitoes naturally occurring Y chromosome meiotic drivers have been identified which cause the X chromosome breaks during meiosis I of spermatogenesis (Hickey and Craig, 1966) (Sweeny and Barr, 1978). The cause of these mechanisms is not clear but the identification of these naturally occurring "selfish" sex chromosomes in mosquitoes encouraged the idea that the expression of Y-linked HEGs targeting the X chromosome during meiotic stages of spermatogenesis would determine a selective "shredding" of the X chromosome. If this happens male mosquitoes will lack sperms bearing the X chromosome and carry sperms bearing only Y chromosome. These males will give only sons when mated with wild-type females and all the sons will inherit the

sex distorter. In this scenario and considering the X-chromosome bearing sperms are unable to fertilize embryos, the synthetic Y chromosome will quickly spread into the population leading to its suppression by elimination of females (Burt, 2003). Based on these assumptions proposals for the generation of synthetic sex-ratio distorters have been suggested and modelled (Burt, 2003; Deredec *et al.*, 2008). I-PpoI is a homing endonuclease that has high specificity for a conserved sequence within the ribosomal rDNA repeats located in a single cluster on the mosquito's X chromosome (Flick *et al.*, 1998; Windbichler *et al.*, 2007). If this endonuclease is expressed during spermatogenesis in transgenic mosquitoes the paternal X chromosome is expected to be cut and therefore only Y chromosome bearing sperms are produced (Figure 7). A first attempt to assess the feasibility and validity of this approach was made in 2008 by Windbichler *et al.* (Windbichler *et al.*, 2008). The expression of *I-PpoI* during spermatogenesis in transgenic mosquitoes succeed in cutting the paternal X chromosomes, however, it also resulted in complete male sterility due to the persistence of stable I-PpoI protein in sperm cells that leads to the cleavage of the maternal X chromosome in the zygote. While this first experiment failed in inducing a sex ration distortion through a mosquito population, it gave proof of principle for the potential of such technique for vector control. In a later experiment, a number of destabilized I-PpoI were generated with the aim of diminishing the *in vivo* half-life of the endonuclease and thereby restrict its activity to male meiosis and prevent the mortality of the later developing zygotes. Of all the transgenic males carrying different grade of destabilized I-PpoI, some variants gave a significant sex ratio bias ranging from 70.2 to 97.4%. Together with a high level of male biased sex distortion a few strains showed a fertility rate comparable to the wild-type control males. Population cage studies showed that transgenic males carrying the sex distorter led to the suppression of a wild-type *A. gambiae* population within six generations. The transgenic strains object of the

above works harbour the endonuclease *I-PpoI*, driven by a germline-specific promoter, on an autosomal chromosome (Galizi *et al.*, 2014). For the efficacy of a vector control strategy based on releases of such transgenic males, the possibility to insert the endonuclease directly on the Y chromosome would represent a powerful tool which in theory could lead to the elimination of the target population even after releasing a very low number of transgenic males.

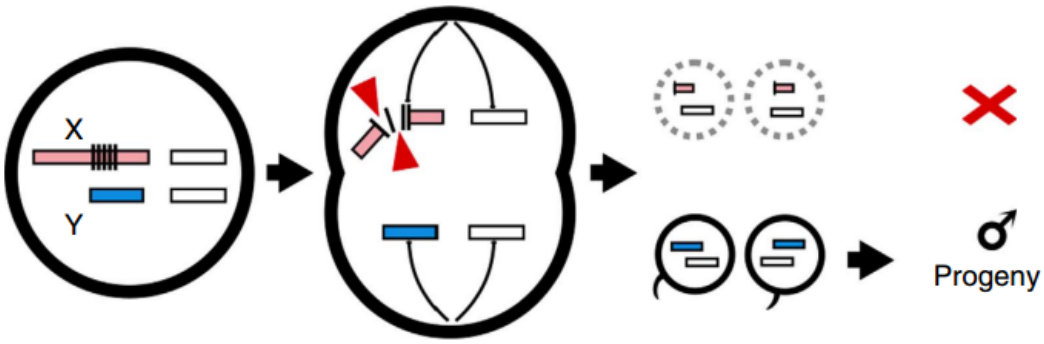


Figure 7: Mechanism of action of the endonuclease I-PpoI.

During spermatogenesis the endonuclease I-PpoI (red triangle) recognizes and cleaves a conserved sequence (black bars) within the ribosomal rDNA repeats located exclusively on the X chromosome. Due to the cuts of the paternal X chromosomes only Y chromosome bearing sperms are produced resulting in a male-biased progeny (Galizi *et al.*, 2014).

1.4.3.9 Sex ratio distortion by targeting Anopheles sex determination

Invasive sex ratio distortion systems could also be generated by targeting HEGs against genes that control the mosquito sex determination pathway (Deredec *et al.*, 2008). In this scenario the target gene, expressed only in females, would have a key role in the activation of female-specific differentiation cascades. Therefore, its knock down would result in masculinisation of females. The molecular components involved in the sex determination pathway have been studied in several species and with great detail in *Drosophila melanogaster* (Baker and Belote, 1983). In *Drosophila* sex determination is determined during the early stages of

embryogenesis and achieved by a balance of female determinants on the X chromosome and male determinants on the autosomes. The X:A ratio is the primary signal that irreversibly sets the state of activity of *sex lethal* (*Sxl*) gene sometime around the blastoderm stage. If *Sxl* is activated it enables the pathway that will lead to female-specific gene expression through its RNA-binding activity. Only in the presence of SXL protein the primary transformer (*tra*) transcript is spliced to produce an mRNA-encoding active TRA protein. TRA is a RNA-binding protein that produces female-specific splicing of the doublesex (dsx) pre-mRNA. The mRNA produced by this splicing pattern encodes a DSX-F protein, a transcription factor that globally represses male-specific gene expression. In the absence of active SXL protein, the splicing pattern of *tra* primary transcript produces an mRNA that does not encode functional TRA protein. In the absence of active TRA protein, splicing of the dsx primary transcript leads to the production of a DSX-M transcription factor that represses female-specific gene expression (Salz and Erickson, 2010). In other organisms sex determination is based on the X/Y system. In this model a trans-acting male determining gene is expressed from the male sex chromosome during early embryogenesis, and functions to inhibit factors that destine the sex determination pathway to the default female form (Wilhelm *et al.*, 2007). In *Anopheles* the *doublesex* gene has been characterized but no upstream regulator has been identified (Scali *et al.*, 2005). Despite being the key regulator gene in *Drosophila*, *Sxl* does not appear to play a key discriminatory role in the control of sex determination outside the drosophilids. Thus, the *Drosophila* sex determination cascade, while offering a large amount of insight, is not likely to be a good paradigm for *Anopheles* when looking at genes that act upstream of transformer.

1.5 The *Anopheles* Y chromosome

A. gambiae has two pairs of submetacentric autosomes, termed 2 and 3, and a pair of heteromorphic sex chromosomes X and Y, XX in females and XY in males (Clements, 1992). The autosomes are divided into two "arms" at the centromere. The longer arm is referred as the right arm and the shorter as the left arm. Extensive non-pairing regions exist between the X and the degenerate Y chromosome and evidence points to a factor located on the Y chromosome which primarily determines the sex in *Anopheles*. The first evidence in Anopheline mosquitoes, which links the primary sex determining signal to the presence of the Y chromosome, came from observations that showed complete sex linkage of recessive eye colour genes (Mason, 1967). In 1979 Baker *et al.* found that triploid *A. culifacies* individuals with 3 X chromosomes where phenotypically females whereas XXY individuals showed a male phenotype (Baker and Sakai, 1979). However these mosquitoes also had a triploid set of autosomes, so no definite conclusion could be drawn nor could a *Drosophila* like system be ruled out entirely. Chromosome translocation experiments in several *Anopheles* species, including *A. albimanus* (Rabbani MG, 1972; Rabbani and Kitzmiller, 1975), *A. gambiae* (Krafsur, 1972; Curtis, 1976) and *A. culicifacies* (Richard H. Baker, 1978), provide the best evidence for the role of the Y chromosome. However due to its heterochromatic state, the Y does not polytenize and does not appear on polytene translocation maps but can be seen only in crude mitotic chromosome spreads. Hence, the arguments used in these translocation experiments were often circular in a way that male inheritance of translocations was used as evidence pointing to the involvement of the Y chromosome which could not be observed directly. Current models suggest that the evolutionary differentiation of Y chromosomes begins with the acquisition of a male determining factor on a proto-Y chromosome (Charlesworth, 1996; Muller, 1932). This is followed by a progressive suppression of recombination between the

still largely homomorphic proto-sex chromosomes, a process attributed to the acquisition of sexually antagonistic mutations, beneficial to the heterogametic sex but detrimental to the homogametic sex (Lahn *et al.*, 2001; Fisher, 1931; Rice, 1984). The lack of recombination, together with the male-limited transmission, leads to the degeneration of the Y chromosome which involves accumulation of deleterious mutations, spread of transposable elements and silencing of all or most of the genes present on the proto-Y (Steinemann and Steinemann, 2000; Burgoyne, 1998; Castillo *et al.*, 2010). As a result Y chromosomes of many species appear to be strongly heterochromatic and harbour only few genes often involved in male fertility (Rice, 1994; Bachtrog, 2003; Carvalho *et al.*, 2000; Carvalho *et al.*, 2001; Steinemann and Steinemann, 1998; Maggert and Golic, 2002). The accumulation of repetitive sequences, many of which are also present on other chromosomes, hampers the assembly of Y chromosome contigs following shotgun sequencing. Indeed, despite the completion of the *A. gambiae* genome (Holt *et al.*, 2002), and the knowledge that the primary signal is likely to be associated with the inheritance of the Y (Baker and Sakai, 1979; Mason, 1967), no assembly of the *Anopheles* Y chromosome has been achieved. At present public databases host only a few hundreds kilobases of *Anopheles gambiae* sequences attributed to the Y, a chromosome which is estimated to comprise 10% of the genome and to be at least 20 megabases in size. None of these Y-specific scaffolds have been physically mapped, as the Y chromosome does not polytenize. Recently a number of candidate genes have been identified on the Y chromosome of anopheline mosquitoes (Hall *et al.*, 2013; Criscione *et al.*, 2013). As interfering with male fertility is an essential part of vector control strategies such as the sterile insect technique, the identification of such genes is of particular interest to mosquito biologists.

2 Aim of this study

The research project described in this thesis has been developed with the aim to generate biological tools for the application of vector control strategies based on genetically modified mosquitoes. The subject of this research focuses upon *Anopheles gambiae* and *Anopheles arabiensis*, the two mosquito species considered to be the major vectors of human malaria. The generation of a transgenic mosquito strain that allows male-exclusive traits is expected to benefit existing vector control strategies such as those requiring mass-releases of sterile males. Moreover, such a transgenic strain will pave the way for the application of new vector control technologies such as those based on endonuclease genes. These assumptions led to the attempt of engineering the *Anopheles gambiae* Y chromosome as an approach to generate a transgenic docking line that allows the insertion of genes of interest specifically onto this chromosome and therefore establish male-exclusive genetic traits. A different procedure based on introgression was used to generate a Y docking strain for the control of *Anopheles arabiensis* populations. The success of this project is intended to complement existing techniques in the fight against human malaria.

3 Materials and Methods

3.1 Mosquitoes Rearing

Wild-type *A. gambiae* (G3), *Anopheles arabiensis* (Dongola) and the transgenic mosquito lines T4, YAttP, YVasG and YI-PpoI were reared under standard condition at 28 °C and 80% relative humidity with access to fish food as larvae and 5% (wt/vol) glucose solution as adults. For eggs production, young adult mosquitoes (3–5 d after emergence) were allowed to mate for at least 6 days and then fed on mice. Three days later, an egg bowl containing rearing water (dH2O supplemented with 0.1% pure salt) was placed in the cage. One to two days after hatching, larvae were placed into rearing water containing trays. The protocols and procedures used in the study were approved by the Animal Ethics Committee of Imperial College in compliance with UK Home Office regulations.

3.2 Isolation of Y-specific sequences

High molecular weight DNA was extracted from transgenic T4 larvae (final larval stage). A BAC library was constructed in vector pIndigoBAC5-HindIII from Epicentre by using the standard *HindIII* cloning site. The library was pooled in 1 × 96-well plate with each well containing approximately 125 independent primary clones with an average insert of 120 kb. Primers *actin5rev* and *exfp3sEq* were used to screen the clones for the presence of a Y-specific insert. Size determination of the insert was performed by *NotI* digestion and Pulsed field gel electrophoresis (PFGE) gel separation. Single Molecule Real Time (SMRT) technology was chosen to sequence the positive BAC clone and PacBio SMRT Portal assembler was used for sequence quality filtering and assembly. The contigs generated were uploaded to GenBank (accession nos. KJ608148–KJ608160).

3.3 Bacterial Cultures

One Shot® TOP10 *Escherichia coli* strains (Invitrogen), involved in the cloning steps, were grown in Laura-Bertani (LB) medium (10g/L tryptone, 5g/L yeast extract, 10g/L NaCl) or on LB agar plates (LB + 15g/L agar). Appropriate selective antibiotics were added to select for propagation of the plasmid vectors.

3.4 Generation of I-SceI RNA

The mMESSAGEmachine kit (Ambion) was used to obtain a 5'capped mRNA (I-SceI RNA) coding for the homing endonuclease I-SceI. The *I-SceI* gene was amplified from the target plasmid pP[v+,70I-SceI] by using primers *ForI-SceI*, containing a T7 RNA polymerase promoter, and *RevI-SceI*.

3.5 Generation of Transformation Vectors

The Phusion Site-Directed Mutagenesis Kit was used to generate the 3xP3[AttP]RFP plasmid. For this cloning strategy primers *F1* and *R2*, containing one-half *AttP* site each at the 5', were used. The plasmid attBCFP-VasaGFP was generated by cloning of a blunt synthetic fragment, attB-fragment, containing enhanced cyan fluorescent protein (eCFP) preceded by an *AttB* site and followed by the SV40 terminator, into the *StuI* site of pfVasIntGFP carrying the VasGFP (Papathanos *et al.*, 2009).

3.6 Polymerase Chain Reaction (PCR)

PCR reactions were performed using an Applied Biosystems Veriti PCR thermal cycler. Typically, reactions were catalysed by Phusion® High-Fidelity PCR Master Mix (New England Biolabs) containing 0.04 U/µl Phusion® DNA Polymerase, 2x Phusion® HF Buffer and 400 µM of each dNT. This reaction was carried out in a final volume of 20µl and 0.5 µM of the forward and reverse primers and 20-100ng of template DNA were added. Each round of PCR synthesis involves three steps: denaturation (98ºC for 30 seconds), annealing (for primers > 20nt the annealing was performed for 10–30 seconds at a Tm +3°C of the lower Tm primer. For primers ≤ 20 nt, an annealing temperature equal to the Tm of the lower Tm primer was used) and elongation (72ºC, 15-30 seconds per kb of template). This three-step 'PCR-cycle' was repeated 25 times. Termination is the final step, where elongation is carried on at 72ºC for 10 minutes. PCR products were purified using the QIAquick PCR Purification Kit (Quiagen) as per the manufacturer's instructions.

3.7 Preparative Restriction Enzyme Digests

Restriction enzymes (purchased from Fermentas and New England Biolabs) were used following the conditions specified by the suppliers. Typically, 1-2µg of DNA was digested in 1-2 unit of enzyme, to a final volume of 20 µl including 2 µl of 10x restriction buffer. The reaction was incubated at the optimum temperature for the specific enzyme (normally 37ºC) for 2 hours. If any of the restriction products were required for subsequent cloning, the restriction digest was loaded on a 1% agarose gel to isolate and purify any band of interest.

3.8 Agarose gel electrophoresis

For size separation, DNA was run on a 0.8%-1.0% agarose gel [w/v] containing ethidium bromide (0.5 µg/mL) in 0.5 x TAE buffer (20 mM Tris-acetate, 0.5 mM EDTA [pH 8.0], pH 8.3) at 90-120 V. DNA bands were examined under UV light at 312 nm on a Herolab UVT-20M transilluminator. 2.5 µL Hyperladder I™ (Bioline) was included to estimate the size and quantity of the fragments.

3.9 Purification of DNA fragments from agarose gel

The bands of interest were excised from the agarose gels and cleaned using the QIAquick Gel Extraction Kit (Quiagen) as per the manufacturers' instructions.

3.10 Dephosphorylation of vectors before ligation

Removal of 5' phosphate groups from DNA was performed using Antarctic Phosphatase (New England Biolabs) to avoid self-ligation and re-circularization of linearized plasmids. This enzyme is active on 5' overhangs, 5' recessed and blunt ends. A reaction containing 1µg of linearized DNA, 1x Antarctic Phosphatase Reaction Buffer and Phosphatase enzyme (0.01 u/µl) was incubated at 37ºC for 30 minutes in a final volume of 25 µl. The reaction was stopped by adding 0.8 µl of EDTA (0.5M) and subsequently purified using the QIAquick PCR Purification Kit (Quiagen) as per manufacturer's instructions.

3.11 Ligations

Ligations were either performed with T4 DNA ligase (Roche) or with Takara Ligase (Takara) following manufacturer's instructions. These enzymes catalyse the fusion of two strands of DNA between the 5'-phosphate and the 3'-hydroxyl groups of adjacent nucleotides. Multiple reactions were performed with differing molar ratios of insert to vector to ensure appropriate molar conditions for efficient ligation. Typically, ligation reactions were set up in a final volume of 10µl. Normally, the concentrations used were 50ng of vector DNA and an equal to 3-fold concentration of insert. The reactions were incubated at 16ºC overnight.

3.12 Transformation of TOP10 Chemically Competent cells

DNA constructs were transformed using One Shot® TOP10 Chemically Competent Cells (Invitrogen), by adding 50 µl of cells to 5 µl of ligation reaction in a 1.5 ml microcentrifuge tube. The tubes were flicked gently to mix the contents and placed on ice for 30 minutes. Heat-shock was carried out at 42ºC in a water bath for 30 seconds. The tubes were immediately placed on ice for 2 minute and 250 µl of pre-warmed S.O.C. Medium (supplied with the kit) were added to the mixture and placed in a shaking incubator at 37ºC for 1 hour. Cells were spread on agar plates supplemented with a suitable antibiotic and incubated at 37ºC overnight.

3.13 Colony Screening by PCR

The following day, individual colonies from plates were added into PCR tubes using pipette tips then streaked onto a fresh replicate agar plate (with the suitable antibiotic) using a

numbered template and incubated at 37ºC overnight. The reactions were performed in a final volume of 10µl using 0.5 units/reaction Taq DNA Polymerase (Quiagen), 2 µl dNTP mix (10 mM each), 0.5 µM of the forward and reverse primer and 1x PCR Buffer. The PCR cycling program was repeated 35 times: denaturation (94ºC for 5 minutes), annealing (temperature approximately 5°C below Tm of primers for 30 seconds), elongation (72ºC, 60 seconds per kb of template). DNA samples were analysed by gel electrophoresis and visualized using UV transilluminator. Usually, 2-4 positive clones were inoculated from plates into 5 ml of LB medium containing antibiotic, and incubated overnight at 37 ºC with moderate shaking.

3.14 Mini-preparation of plasmid DNA

Plasmid DNA was extracted from bacteria using QIAprep Miniprep Kits (Quiagen) as per the manufacturer's instructions.

3.15 DNA sequencing

1 µg of DNA, in a final volume of 15 µl, was sent to Beckman Coulter Genomics (Takeley, UK) and sequentially analysed using Vector NTI software.

3.16 Midi-preparation of plasmid DNA

Midi-preps were prepared using PureLink™ HiPure Plasmid Midiprep Kit (Invitrogen) as per the manufacturer's instructions. A colony was grown in 25 ml of LB and the appropriate amount of antibiotic and incubated overnight with shaking at 37°C.

3.17 Large-scale 'Maxi-Prep' preparation of plasmid DNA

Large-scale purification (100µg-1µg) of plasmid DNA was achieved using the QIAGEN Plasmid Maxi Kit (QIAGEN), following the manufacture's instructions. The protocol essentially follows the alkaline lysis method but the scale of culture is increased. Endofree Maxi-preps were used when preparing plasmids for *Anopheles* embryo injections.

3.18 Spectrophotometry of DNA

Determination of DNA concentration and purity was performed using a Nanodrop Spectrophotometer by loading 1.5µl of sample with absorbance readings taken at wavelengths of 260nm and 280nm, simultaneously. At 260nm, absorbance was used to calculate the concentration of the ds DNA, where an OD of 1 unit corresponds to ~50µg/ml. The ratio between the reading at 260nm and 280nm (OD_{260}:OD_{280}) presented the purity, where high quality preparations of DNA have an OD_{260}:OD_{280} ratio of ~1.8.

3.19 Embryo Microinjections

Embryos were injected using a Femtojet Express injector and a Narishige 202ND micromanipulator mounted on an inverted microscope (Nikon TE-DH100W) with a mixture of 0.2 µg/µL 3xP3[AttP]RFP plasmid or 0.8 µg/µL in vitro-transcribed I-SceI RNA, and 0.4 µg/µL pVas2-I-SceI helper plasmid for the generation of the transgenic mosquito line YAttP. The hatched larvae were screened for transient expression of the RFP marker, and the 0.18% positive was grown up and crossed to wild-type mosquitoes. The progeny of these crosses was analysed for red fluorescence. Individual larvae showing expression of the selectable marker were then separated and the adults that emerged were crossed individually with wild-type

mosquitoes to obtain transgenic lines. Embryos were injected with a mixture of 0.2 µg/µL attBCFP-VasaGFP plasmid and 0.5 µg/µL Integrase helper plasmid to generate the transgenic mosquito line YVasG. The hatched larvae were grown up, and the surviving males were crossed to wild-type mosquitoes. The progeny of these crosses was analysed for eCFP fluorescence. Individual larvae showing expression of the selectable marker were then separated and the adults that emerged were crossed individually with wild-type mosquitoes to obtain transgenic lines. Transgenic mosquitoes at different developmental stages were analysed on a Nikon inverted microscope (Eclipse TE200) at a wavelength of 488 nm to detect GFP expression, 563 nm to detect RFP expression and 442 nm to detect eCFP expression.

3.20 Confocal Microscopy

Dissected testes from transgenic lines YVasG and Vas2GFP were fixed in methanol-free 4% formaldehyde (Pierce) in PBS for 30 min and washed three times for 15 min in 0.1% Tween-20 PBS. Testes were then transferred on fresh slides containing Vectashield mounting medium with DAPI (Vectorlabs) with coverslips. Testes images were taken by using a Zeiss LSM 510 Laser scanning confocal microscope and a 20× objective in two cannels (DAPI and GFP) sequentially.

3.21 Fluorescence in situ hybridization (FISH)

FISH of the sexual chromosomes, X and Y, in the transgenic line T4 was performed according to a previously established protocol (Timoshevskiy *et al.*, 2012). The X chromosome was marked by using as a probe an intergenic spacer region of rDNA generated by PCR using primers *UN* and *GA* and labelled with the cyanine dye Cy5. The Y chromosome was marked by

using as a probe the p-HomeT plasmid Cy3-labeled by nick translation (Roche Nick Translation Kit). Chromosomes where counterstained with DAPI.

3.22 Genetic crosses and fertility assays

Crosses for the introgression experiments were set up in BugDorm-1 cages with size 30x30x30 cm. For any cross a number of 100 female and 100 male mosquitoes was used, the number of males varied through the experiment depending of the progeny recovered from the previous generation. To assay fertility of the Y-introgressed males single crosses in cups were set up. Y-introgressed males were singularly introduced into a cup together with one *A. arabiensis* female (*introgressed cups*). Parallel the same number of cups were set up for *A. arabiensis* males and females as a control (*control cups*). After 4 days mating and blood feeding of females, eggs were collected from every cup and the hatching rate (number of larvae/number of eggs) relative to every backcross was calculated.

4 Results

4.1 Engineering of the *A. gambiae* Y chromosome

The use of genetic modified mosquitoes has been proven to benefit the understanding of molecular and biological mechanisms in the contest of malaria research. The previously discussed sterile insect technique, based on mass-releases of sterile males, has been efficiently applied for the eradication of insect pest, however the lack of a reliable high-throughput sexing methods, necessary to guarantee the male-exclusive releases, and the loss of male fitness that arises from the sterilization methods, represent a major problem that affect the success of this technique in large scale programmes. The generation of a transgenic docking strain that allows expression of gene of interest inserted onto the Y chromosome is expected to enable automated sexing methods based on fluorescent markers or, alternatively, genetic sexing traits such as insecticide resistance markers could be introduced to the Y chromosome providing the means for large releases of only male mosquitoes. Together with the improvement of existing technique for vector control, the possibility to generate male-exclusive traits would allow the development of new techniques such as those based on endonucleases genes.

4.1.1 The transgenic line T4 carries a transgene that segregates with the Y chromosome

Transposase-mediated germline transformation is a technique routinely used in laboratories where generation of genetic modified mosquitoes is required. Although hundreds of transgenic *Anopheles* strains have been generated since the inception of this technique, so far no strain harbouring a construct inserted into the Y chromosome has been described. Due to its heterochromatic nature, the Y chromosome was thought to be refractory to the random integration catalysed by transposase or, alternatively, marker genes commonly used to identify transformation events undergo complete silencing when located on the Y chromosome. In this paragraph the generation of a Y-linked transgenic line obtained by transposase-mediated germline transformation is described. Wild-type *A. gambiae* embryos were injected with the plasmid pHome-T together with in-vitro transcribed piggyBac transposase helper RNA (Windbichler *et al.*, 2011). The pHome-T plasmid contains *piggyBac* inverted repeats for transposase-mediated random integration as well as fluorescent markers, the GFP driven by the neuronal *3xP3* promoter and the RFP under the control of the *Drosophila Actin5C* promoter. In addition, the construct contains the 18 base pair I-SceI endonuclease recognition site, located within the GFP open reading frame (Figure 8). This site is not present in the *Anopheles* genome and can be used to specifically cut chromosomes carrying the integrated plasmid exclusively at this position. A number of G0 mosquitoes was obtained expressing the fluorescent markers transiently and they were separately crossed to wild-type mosquitoes. One male G0 founder gave rise to 42 transgenic offspring all of which emerged as males. When 5 of these transgenic males were crossed to wild-type females in single mating experiments, again, all transgenic larvae were emerging as male adults (206 out of 421) whereas the non-transgenic larvae (215 out of 421) emerged as females. These results

indicated that the transgene construct had potentially integrated within the Y chromosome (Figure 8). Expression of both fluorescent marker genes in this transgenic line, termed T4, was clearly visible however the expression of RFP, driven by the *Actin5C* promoter, was patchy and visible mainly in the periphery of the larval gut. The pattern of neuronal GFP fluorescence appeared to be normal resembling that of autosomal insertions (Figure 9). To confirm the Y-linkage of transgene in the T4 strain, fluorescence in situ hybridization (FISH) was performed. In situ hybridization experiments are routinely conducted on polytene chromosomes from ovarian nurse cells. Due to his heterochromatic state, the Y chromosome does not polytenize but can be seen only in crude mitotic chromosome spreads. It has been shown that larval imaginal discs in anopheline mosquitoes are particularly suitable for obtaining high-quality mitotic chromosome preparations (Timoshevskiy *et al.*, 2012). Therefore, FISH was performed using the pHome-T plasmid as a probe against mitotic chromosome spreads, generated from male larval imaginal discs, of line T4. As a control the multicopy rDNA locus on the X chromosome was used. A single weak signal was obtained from the smaller sex chromosome (Figure 10) supporting the notion that the transformation construct had integrated into the Y chromosome.

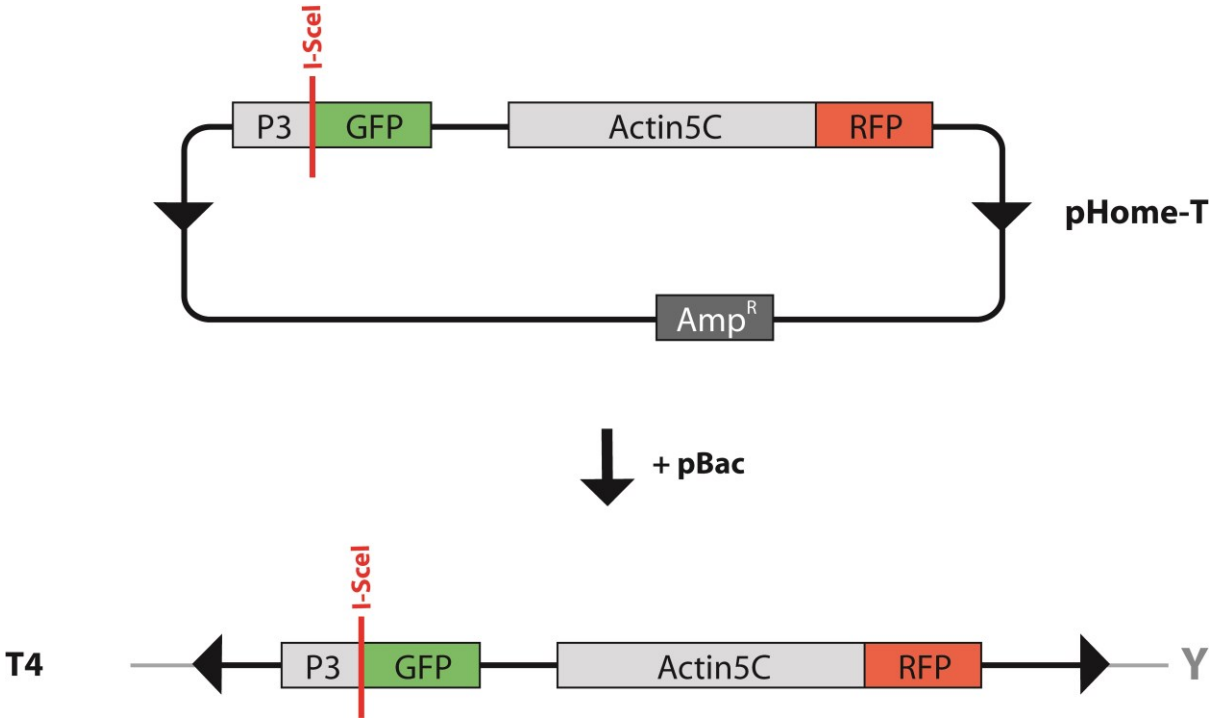


Figure 8: Generation of the transgenic line T4.

Wild-type *A. gambiae* embryos were injected with the plasmid pHome-T together with in-vitro transcribed *piggyBac* transposase helper RNA. The pHome-T plasmid contains *piggyBac* inverted repeats (black triangle) for transposase-mediated random integration as well as the *3xP3*-GFP and *Actin5C*-RFP fluorescent markers. In addition, the construct contains the 18 base pair I-SceI endonuclease recognition site, located within the GFP open reading frame. Integration of the transgenic construct onto the Y chromosome is shown.

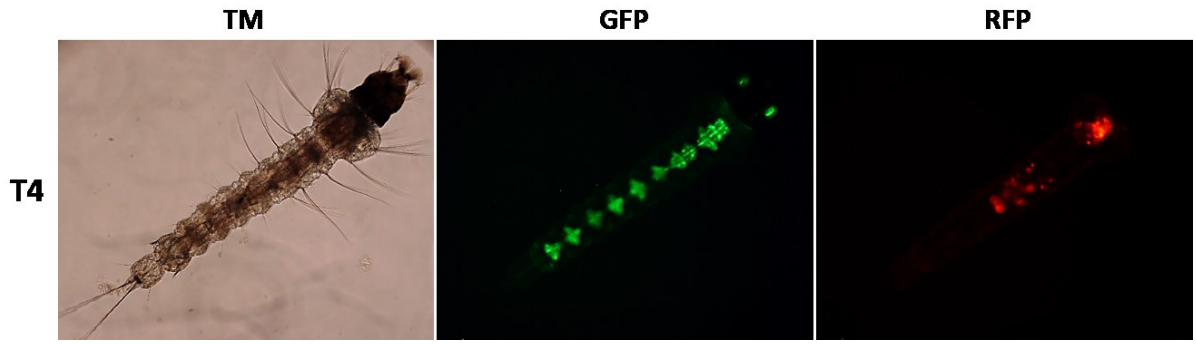


Figure 9: Fluorescent pattern of T4 larvae.

Expression of green (GFP) and red (RFP) fluorescent markers in the tissues of second instar larvae shown as transmitted-light image (TM) in *Left*.

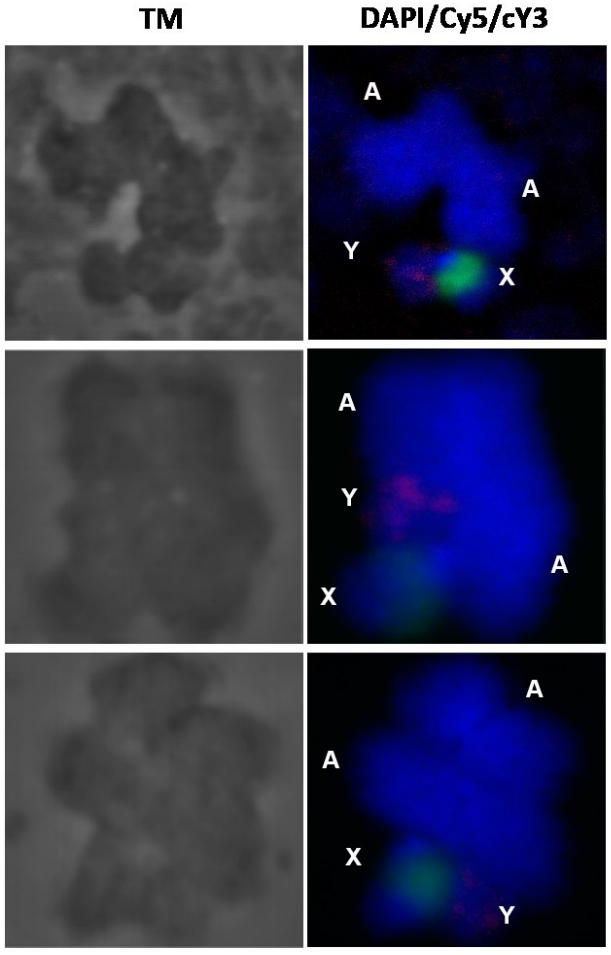


Figure 10: Fluorescence in situ hybridization (FISH) of mitotic chromosomes in T4 line.

FISH with probes designed against the X-linked rDNA (labeled with Cy5, green) and the pHome-T construct (labeled with Cy3, red) hybridized to the transgenic line T4. Mitotic chromosome slides were prepared from imaginal discs of fourth instar larvae. X and Y indicate the sex chromosomes and A indicates autosomes. Chromosome preparation is shown as transmitted-light image (TM) in *Left*.

4.1.2 The Y-linked transgene is stable and can serve as a marker for automated sexing

The Y linkage of the transgene in T4 line made this strain a good candidate for high-throughput sexing methods, required for the efficiency of vector control strategies such us sterile insect technique. Recombination between the *A. gambiae* X and Y chromosomes has not been reported to occur; however, evidence is derived mainly from genetic experiments by using chromosomal translocations (Krzywinski *et al.*, 2004). The Y-linked transgene allows to test experimentally whether exchange between these chromosomes occurs at significant rate. A total of 16,750 T4 larvae were sorted in the complex object parametric analyser and sorter (COPAS) particle sorter according to their distinct intensities of green fluorescence, because the expression of GFP is expected to occur in male larvae only (Figure 11a). The GFP-negative cloud corresponding to putative female larvae (6,415 individuals) was entered for a second run in the COPAS to check for the presence of contaminating GFP+ larvae. No GFP-expressing male larvae were detected (Figure 11b). The GFP-positive and -negative larvae were separately reared to adulthood. All GFP positive adults emerged as males, whereas all GFP negatives were found to be females, thus indicating the utility of line T4 for automated sex separation. This COPAS sorting experiment was repeated multiple times at intervals of several months, encompassing at least 12 generations. Sorting was always perfectly accurate, indicating that the transgene is stable. No putative X/Y recombination events were detected in the T4 line, indicating that recombination was either very rare or that the T4 insertion site is located in a region of the Y chromosome that does not readily recombine with the X. Thus, automated sorting of the T4 line represents a fast and reliable method to generate pure populations of either transgenic T4 males or non-transgenic, virgin females.

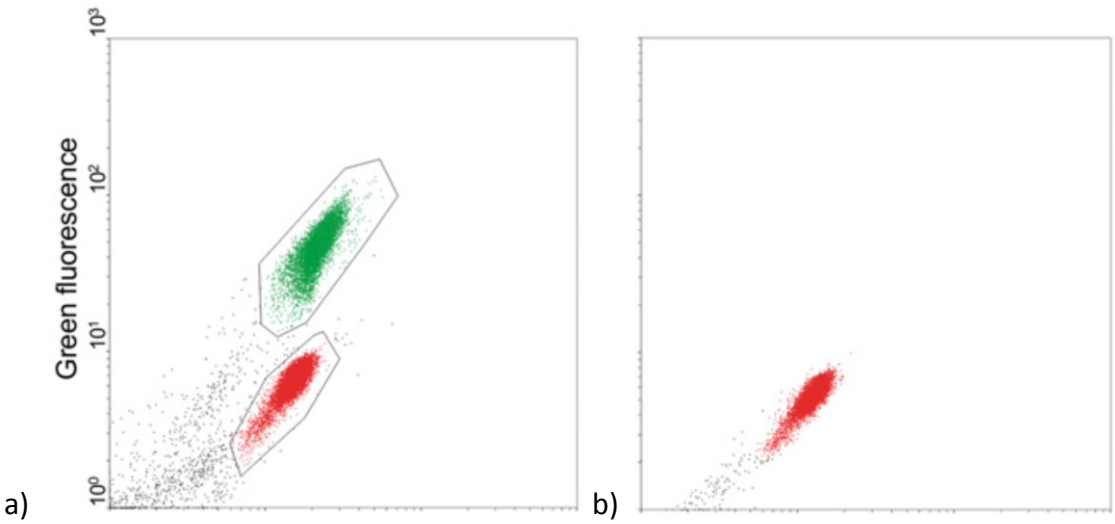


Figure 11: Flow cytometry analysis and automated sex separation of the T4 strain.

a) A total of 16,750 larvae of the T4 line were analysed according to their level of green and red fluorescence (green and red clouds) and sorted via he gates indicated (black lines).
b) Purified larvae (6,415 individuals) gated in a) as having only background fluorescence were subjected to a second sorting run in the COPAS to check for the absence of contaminating GFP positive larvae.

4.1.3 Characterization of the genomic neighbourhood of the T4 transgene

Expression of fluorescent maker genes in two different tissues of T4 males suggested that the transgene construct had inserted within a region of chromatin that allows transcriptional activity. Such a euchromatic region is of interest because it might host male-specific genes. Therefore, the genomic neighbourhood of the T4 transgene was explored. Inverse PCR is a technique routinely used for the determination of insert locations. Known sequences of the insert can be used to design primers that amplify a small region of the genomic DNA flanking the insert. The amplified product can then be sequenced and compared with DNA databases to locate the genomic sequence where the insertion occurred. Such a strategy, if applied to the Y-linked T4 transgene, offers the possibility to identify, by an *in silico* analysis of the sequence derived from the inverse PCR product and the *A. gambiae* genome database, scaffolds containing Y sequences. Therefore, inverse PCR was performed using primers (*alt3R2*

and *alt5F2*) binding to the *piggyBac* inverted repeats of the integrated T4 construct. A genomic sequence was obtained that aligned to AAAB01003622.1, a scaffold not assigned to a known chromosome but that was previously characterized as a scaffold containing fragments originating exclusively from male libraries (Krzywinski *et al.*, 2004). This scaffold has a length of only 1.7 kb. With the aim to extend the knowledge of the T4 transgene flanking regions, a bacterial artificial chromosomes (BAC) library from genomic DNA of T4 males was generated. This approach allows the generation of BACs containing large inserts, 150-350 Kb, of genomic DNA. The exclusive presence of the T4 transgene on the Y chromosome can be used to isolate BACs harbouring Y chromosome inserts and therefore to provide Y sequences. L4 larvae (final larvae stage) were screened for the presence of the fluorescent markers and high molecular weight (HMW) DNA was extracted. The BAC vector used to generate the library was pIndigoBAC5-HindIII and the cloning site was *HindIII*. The library was pooled in 1x96-well plate with each well containing about 125 independent primary clones with an average insert of 110 Kb. A specific PCR primer pair (*Actin5rev* and *exfp3seq*) was made to amplify a single specific region of the transgene. *Actin5rev* and *exfp3seq* primers amplified a DNA fragment 588 bp long when the genomic DNA extracted from T4 males was analysed (Figure 12a). A Y chromosome insert carrying clone, named BAC 4327-1 9C, was selected (Figure 12b). Size determination of the insert was performed by *NotI* digestion and pulsed-field gel electrophoresis (PFGE) gel separation, suitable for large DNA molecules. Two bands, corresponding to two Y chromosome inserts, ~80kb and ~15kb in size, were identified (Figure 12c).

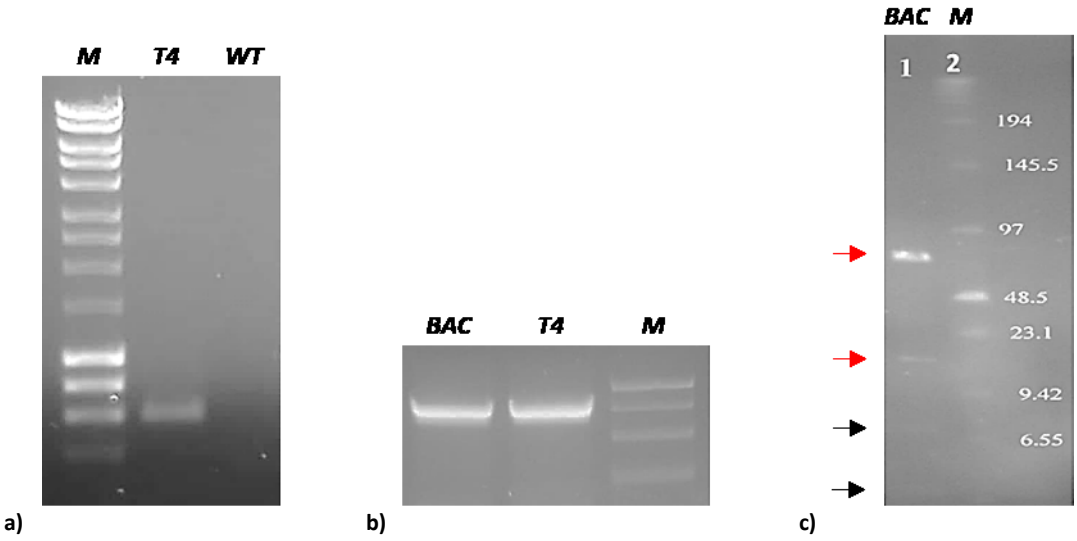


Figure 12: Identification of a Y chromosome insert carrying BAC and insert size determination.

a) Lane 1: hyperladder I (M). Lane 2: genomic DNA from T4 male (T4). Lane 3: genomic DNA from wild-type *A. gambiae* male (WT). **b)** Lane 1: BAC 4327-1 9C (BAC). Lane 2: genomic DNA from T4 male (T4): Lane 3: Marker (200, 400, 600, 800 bp) (M). Samples in a) and b) were tested with a pair of primers that amplify a region (588 bp) of the T4 transgenic construct. **c)** Lane 1: *NotI* digestion and PFGE gel separation of BAC 4327-1 9C (BAC). Lane 2: Marker (size shown in figure) (M). Red arrows indicate the Y chromosome inserts. Black arrows indicate the BAC 4327-1 9C bone.

Single Molecule Real Time (SMRT™) technology was chosen to sequence the BAC 4327-1 9C, this approach offers the possibility to obtain a high number of long reads (1000-3000bp) suited to analyse complex and repetitive sequences such those potentially present on the Y chromosome. The BAC sequencing provided a total number of 150292 unfiltered reads. PacBio SMRT Portal for sequence quality filtering and assembly was used to exclude the adapter dimers, required for the previous steps of this technique, and short fragments (11-100bp). Firstly, one contig, 23171 bp long was obtained. Therefore, a second assembly attempt was done in order to include a significant amount of sequence information that escaped the first assembly attempt. For the re-assembly, filtering-stringency was lowered. This strategy provided a number of 13 contigs (contig 1-13) that are in line with the expected size of the insert in BAC 4327-1 9C. An *in silico* analysis of the contigs was performed using the basic local alignment search tool (BLAST) algorithm. The sequence of all contigs was found to

consist, in a large part, of short sequence repeats and sequence fragments matching known transposable elements. In order to identify any putative Y-specific gene, a cut-off of 100 amino acids as a minimum length was used to select putative open reading frames (ORFs) within these contigs. A number of 39 ORFs were selected. PCR performed on genomic DNA showed that 3 ORFs out the 39 were amplified from wild-type or T4 males but not from wild-type females (Table 1). To assess any expression of the Y-specific ORFs identified, mRNA was extracted from adult mosquitoes (T4 males, wild-type males and wild-type females), wild-type testes and embryos of 0, 2, 4, 6 and 8 hours. RNA extraction was followed by first strand cDNA synthesis and reverse transcriptase (RT)-PCR analysis. None of the Y-specific ORFs showed expression in any of the samples analysed (Figure 13, Figure 14, Figure 15 and Figure 16).

Contig	ORF designator	PCR amplification from genomic DNA			Comment
		wt males	wt females	T4 males	
1	ORFA (324bp)				
	ORFB (309bp)				
2	ORFC (384bp)				Putative transposase
	ORF2 (378bp)	✓	✓	✓	
3	ORF10 (1110bp)	✓	✓	✓	Pol polyprotein
4	BigORF2 (654bp)				
	BigORF1 (495bp)	✓		✓	
	ORF1 (219bp)	✓		✓	
	ORF10 (408bp)	✓		✓	
	ORF12 (468bp)				
5	ORF12 (435bp)	✓	✓	✓	Putative polyprotein
	BigORF2 (432bp)				
	ORF17 (399bp)	✓	✓	✓	
6	ORF4 (336bp)	✓	✓	✓	
	ORF10 (354bp)	✓	✓	✓	RNA-dependent DNA polymerase
	ORF11 (303bp)	✓	✓	✓	RNA-dependent DNA polymerase
	ORF12 (486bp)				
7	ORF3 (390bp)	✓	✓	✓	Transposon T1-2
	ORF6 (315bp)				
8	ORF20 (528bp)				Putative transposase
	ORF21 (414bp)				

9	ORF2 (480bp)	✓	✓	✓	
10	ORF7 (384bp)	✓	✓	✓	Putative reverse transcriptase
	ORF18 (384bp)				
11	ORF20 (438bp)	✓	✓	✓	
	ORF21 (426bp)	✓	✓	✓	
	ORF22 (354bp)				
	ORF23 (315bp)				
	ORF24 (528bp)	✓	✓	✓	
12	ORFa (561bp)				
	BigORF1 (477bp)	✓	✓	✓	
	ORFh (705bp)	✓	✓	✓	
	ORFe (477bp)	✓	✓	✓	Putative reverse transcriptase
	ORFf (492bp)	✓	✓	✓	Putative reverse transcriptase
	ORFg (465bp)	✓	✓	✓	
	ORFI (339bp)				
	ORFL (441bp)	✓	✓	✓	
	ORFM (306bp)				
	ORFb (366bp)	✓	✓	✓	Putative reverse transcriptase

Table 1: PCR analysis of selected open reading frames (ORFs)

A number of 39 putative ORFs were selected and tested for Y-linkage. Primers used for the PCR analysis are listed in the appendix.

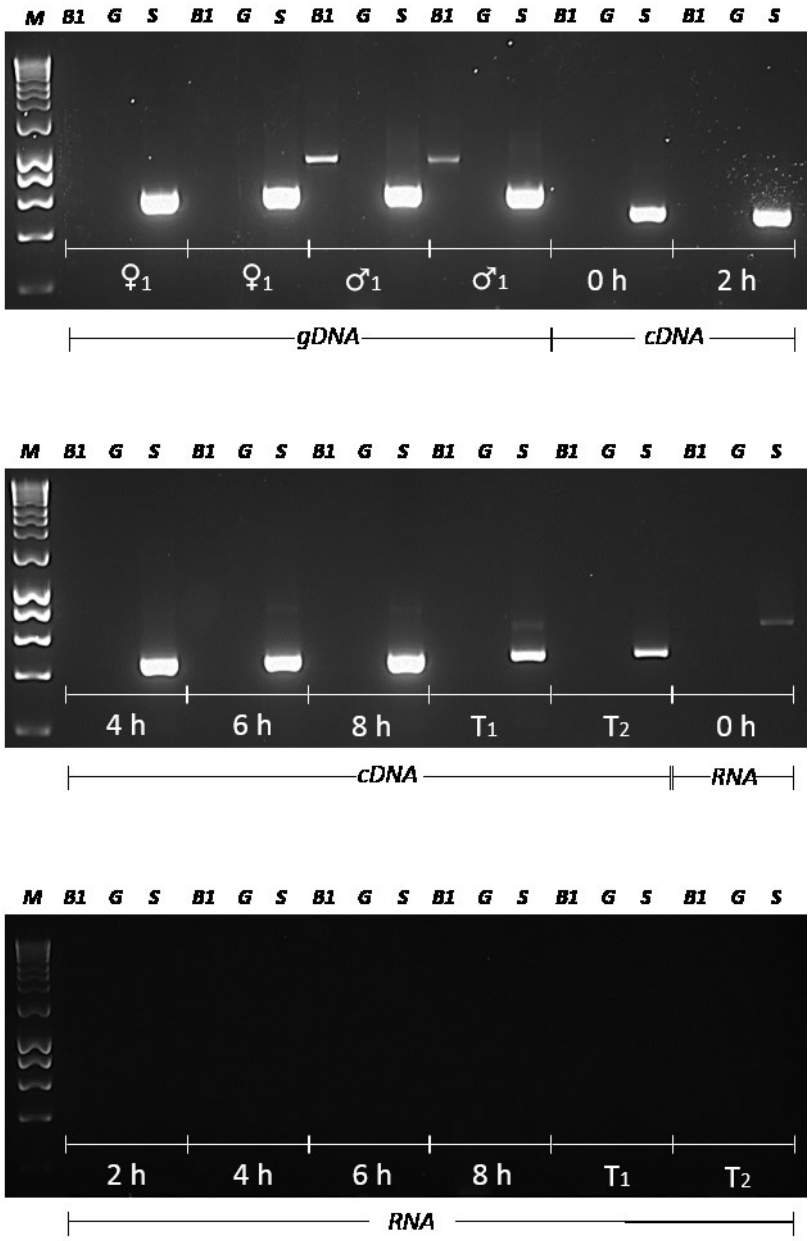


Figure 13: PCR and RT-PCR analysis of the Y-specific ORF BigORF1.

PCR analysis was performed on genomic DNA (gDNA) of wild-type adult female and male mosquitoes (♀, ♂). RT-PCR analysis was performed on cDNA of wild-type embryos 0, 2, 4, 6 and 8 hours old (0, 2, 4, 6 and 8 h) and on wild-type testes (T1 and T2). RNA was tested to exclude DNA contamination in the samples. In figure M indicates hyperladder I. Targets of the PCR analysis are indicated as B1 (BigORF1) (~1000 bp), G (GFP) (460 bp) and S (S7 ribosomal protein gene) (580 bp).

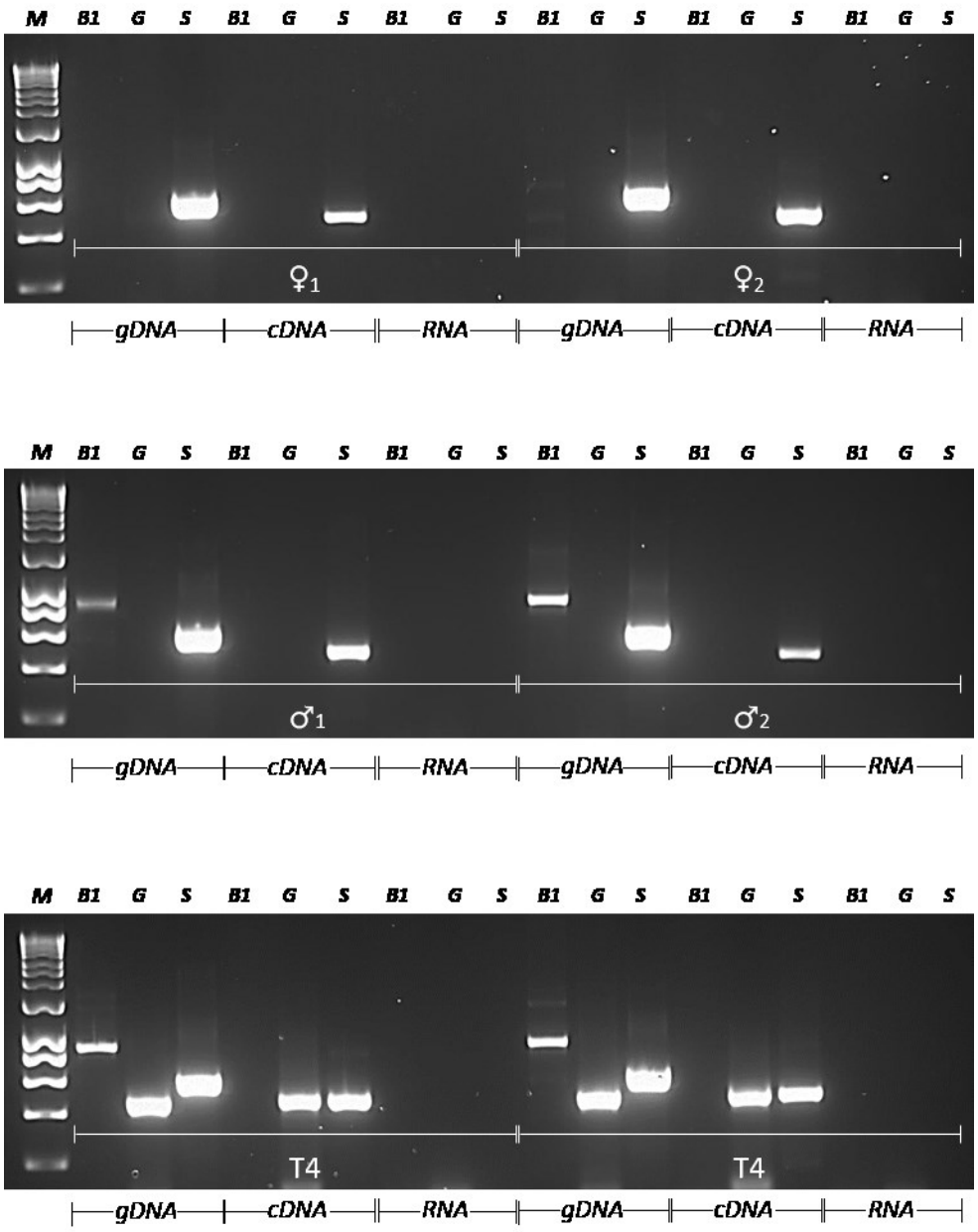


Figure 14: PCR and RT-PCR analysis of the Y-specific ORF BigORF1.

PCR and RT-PCR analysis were performed on genomic DNA (gDNA) and cDNA of wild-type adult female and male mosquitoes and adult T4 males (♀, ♂, T4). RNA was tested to exclude DNA contamination in the samples. In figure M indicates hyperladder I. Targets of the PCR analysis are indicated as B1 (BigORF1) (~1000 bp), G (GFP) (460 bp) and S (S7 ribosomal protein gene) (580 bp).

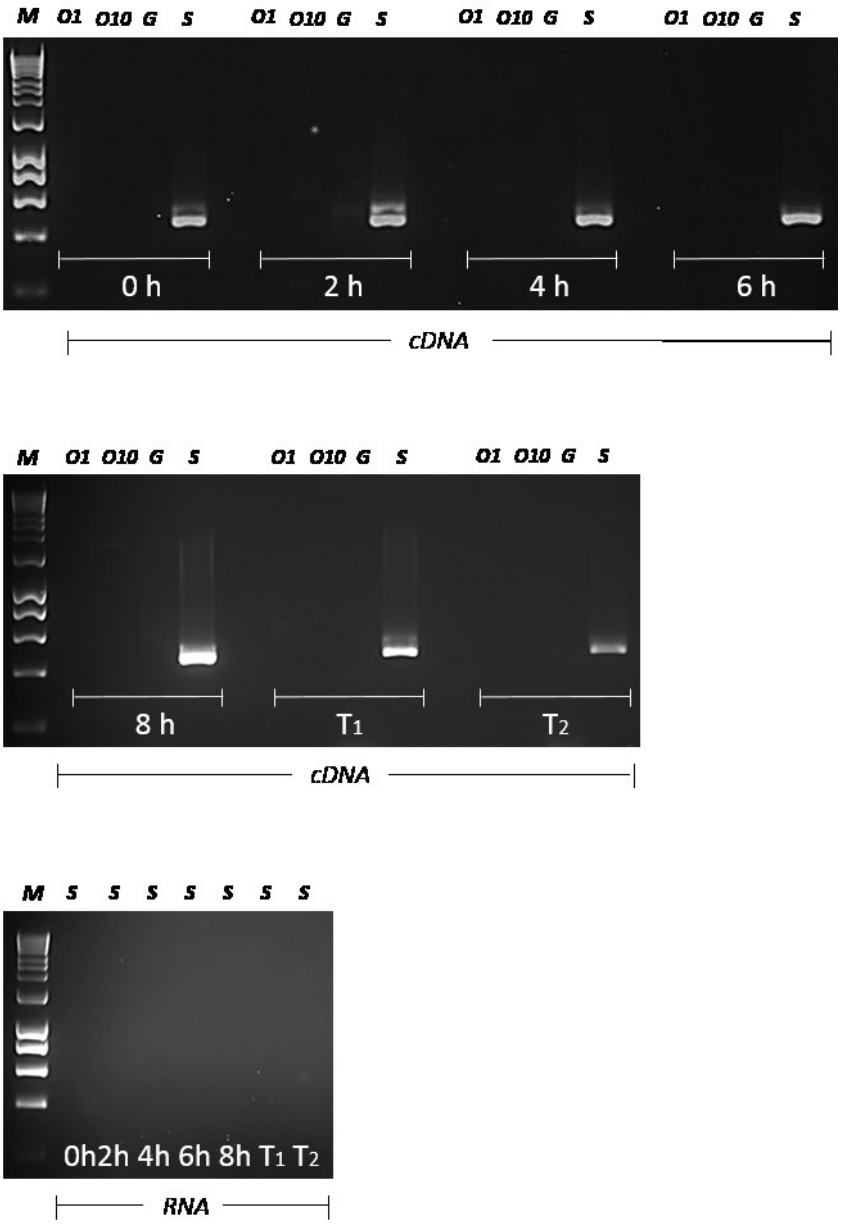


Figure 15: RT-PCR analysis of the Y-specific ORFs ORF1 and ORF10.

RT-PCR analysis was performed on cDNA of wild-type embryos 0, 2, 4, 6 and 8 hours old (0, 2, 4, 6 and 8 h) and on wild-type testes (T1 and T2). RNA was tested to exclude DNA contamination in the samples. In figure M indicates hyperladder I. Targets of the RT-PCR analysis are indicated as O1 (ORF1) (219 bp), O10 (ORF 10) (408 bp), G (GFP) (460 bp) and S (S7 ribosomal protein gene) (580 bp).

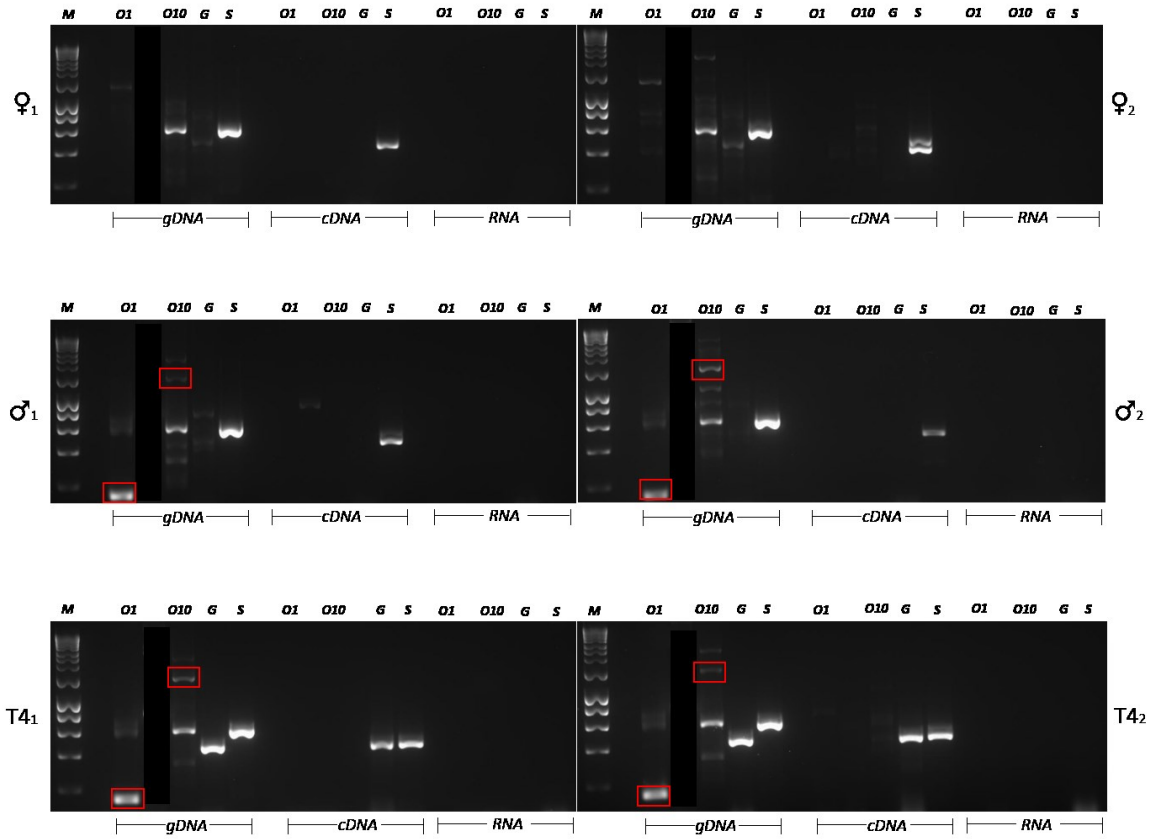


Figure 16: PCR and RT-PCR analysis of the Y-specific ORFs ORF1 and ORF10.

PCR and RT-PCR analysis were performed on genomic DNA (gDNA) of wild-type adult female and male mosquitoes and adult T4 males (♀, ♂, T4). RNA was tested to exclude DNA contamination in the samples. In figure M indicates hyperladder I. Targets of the PCR analysis are indicated as O1 (ORF1) (219 bp), O10 (ORF 10) (408 bp), G (GFP) (460 bp) and S (S7 ribosomal protein gene) (580 bp). In figure the red boxes facilitate the visualization of the Y-specific ORFs.

4.1.4 Identification of a Y-specific polymorphism

The analysis of the selected ORFs showed some interesting outcomes for the non-Y-specific ORFe on contig 12. Interestingly, a noticeable shift of the band was present only in wild-type females when PCRs were performed with genomic DNA (Figure 17). *In silico* analysis of ORFe, using BLAST algorithm, led to the identification of a homologues sequence on the *A. gambiae* X chromosome in a region that lacks of an annotated transcript. An extra string of 24 nucleotides is present on the X-linked ORFe (Figure 18). A primer named *Yspec2* was designed in order to specifically amplify ORFe from the Y chromosome. *Yspec2* spans a sequence that in the X-linked ORFe is interrupted by the 24 nucleotide long gap (Figure 18). PCRs were performed with *Yspec2* and *ORFe reverse* primers using genomic DNA from T4 males, wild-type males and wild-type females. As shown in Figure 19, only when male genomic DNA was used for the PCR reactions, strong bands of the expected size (374 bp) were obtained whereas really faint bands were obtained, along with non-specific bands from female genomic DNA.

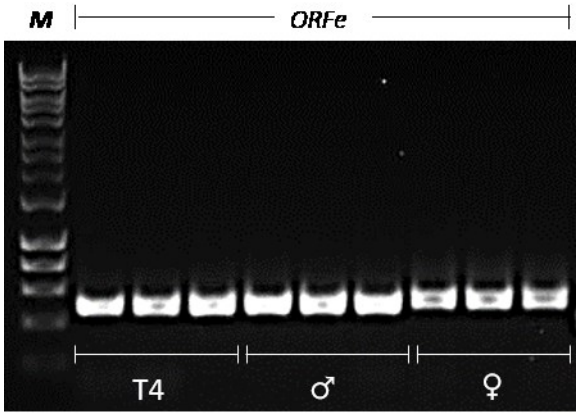


Figure 17: PCR analysis of the non-Y-specific ORFe.

PCR analysis was performed on genomic DNA of T4 males and wild-type adult male and female mosquitoes (T4, ♂, ♀). Lane 1: hyperladder I (M). A pair of primers that amplify ORFe (477 bp) was used (see table 1). A shift of the bands due to an extra string of 24 bp is shown in females.

Figure 18: Alignment of the non-Y-specific ORFe from Contig 12 with the *A. gambiae* X chromosome.

The BLAST tool available on Vectorbase (https://www.**vectorbase**.org/) was used to align ORFe from Contig 12 (query) with the *A. gambiae* reference genome. A region of homology was found in the X chromosome (sbjct). A 24 nucleotide gap was detected at position 120 in the query sequence. *ORFe forward* and *reverse* primers amplify ORFe from the X and Y chromosome. *Yspec2* and *ORFe reverse* primers amplify ORFe specifically from the Y chromosome.

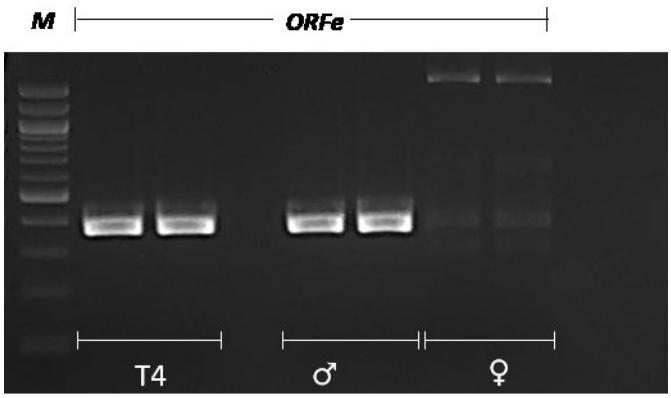


Figure 19: PCR analysis of the non-Y-specific ORFe.

When PCR analysis is performed with *Yspec2* and *ORFe reverse* primers (Figure 18), a bright band (374 bp) is obtained from T4 and wild-type males (T4, ♂) genomic DNA. A faint band together with non-specific bands were obtained when genomic DNA from wild-type female (♀) was used. In figure M indicates hyperladder I.

Gene expression was also investigated for ORFe. For this analysis *ORFe forward* and *reverse* primers were used (Figure 18). Expression was detected in adult males (wild-type and T4) and embryos of 3 and 9 hours (Figure 20). Moreover, one out of two testes samples analysed gave an amplicon.

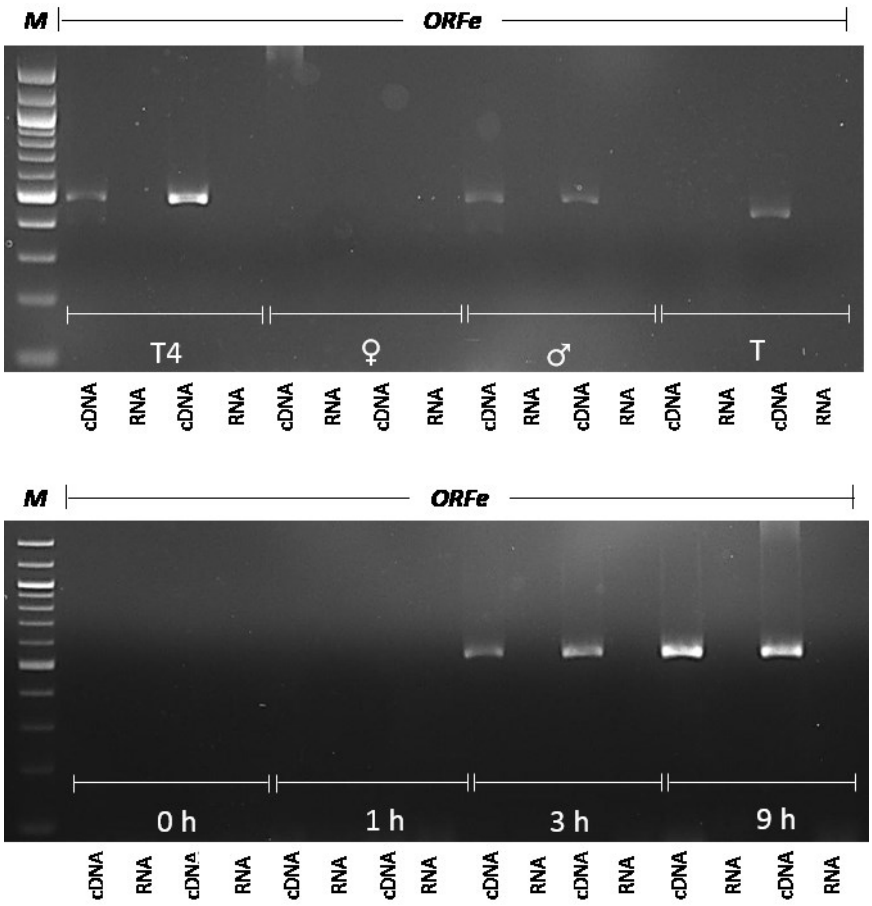


Figure 20: RT-PCR analysis of the non-Y-specific ORFe.

RT-PCR analysis was performed on cDNA of T4 males and wild-type adult female and male mosquitoes (T4, ♀, ♂), wild-type embryos 0, 1, 3 and 9 hours old (0, 1, 3, 9 h) and on wild-type testes (T). RNA was tested to exclude DNA contamination in the samples. A pair of primers that amplify a region of ORFe was used (Table 1). In figure M indicates hyperladder I.

4.1.5 Generation of the Y-linked φC31 docking line YAttP

The fact that the T4 locus was transcriptionally active made it an exciting target for gene engineering. As the ability to introduce transgenes to the Y chromosome would allow the generation of male-exclusive genetic traits, which is of great interest for vector control, a "knock-in" strategy was designed in order to introduce a site-specific recombination signal onto the Y chromosome. The approach relied on the well characterized endonuclease I-SceI and the presence of an I-SceI site within the GFP ORF of the T4 transgene. It was previously shown that I-SceI cleavage can trigger homologous repair between homologous chromosomes and from plasmid repair templates (Windbichler *et al.*, 2011). Therefore, the plasmid 3xP3[AttP]RFP was generated by site-directed mutagenesis of the pHome-T construct (Figure 21). The plasmid 3xP3[AttP]RFP carries the RFP marker under the control of the *3xP3* promoter. The plasmid was designed to contain, between the promoter and the RFP coding sequence, an *AttP* recombination site that replaces both the GFP coding sequence as well as the *Actin5C* promoter which are present in pHome-T. The 3xP3[AttP]RFP plasmid retains regions of homology to the T4 locus of 0.6 kb and 2.1 kb, 5' and 3' of the *AttP* site respectively. The *AttP* recombination signal is recognized by the phage φC31 integrase and would allow for the subsequent site-specific integration of constructs carrying a corresponding *AttB* signal (Thorpe and Smith, 1998; Meredith *et al.*, 2011). If I-SceI induced DNA double strand breaks followed by homologous repair from 3xP3[AttP]RFP in the germline progenitor cells it could be observed as male larvae switching fluorescent reporters i.e. showing red but no longer any green fluorescence. This approach allows for the identification of transgenics and, since the *AttP* site is located between the *3xP3* promoter and RFP coding sequence, it would also allow the identification of subsequent φC31 integrations by the loss or conversion of red fluorescence. The plasmid 3xP3[AttP]RFP was co-injected into embryos of transgenic line T4

together with a source of I-SceI represented either by in-vitro transcribed I-SceI mRNA or a helper plasmid in which the *I-SceI* coding sequence is placed under the control of the *A. gambiae vasa* promoter (Papathanos *et al.*, 2009). Table 2 shows the outcome of these experiments. A total of 11 3xP3-RFP positive progeny out of 6160 larvae screened (0.18%) was observed when using helper plasmid. No RFP positive larvae were obtained from the progeny of males injected with I-SceI RNA. Genomic PCR and sequencing confirmed that cassette replacement had occurred in this new transgenic line termed YAttP (Figure 22). In Figure 23 the fluorescent pattern of the YAttP strain is shown.

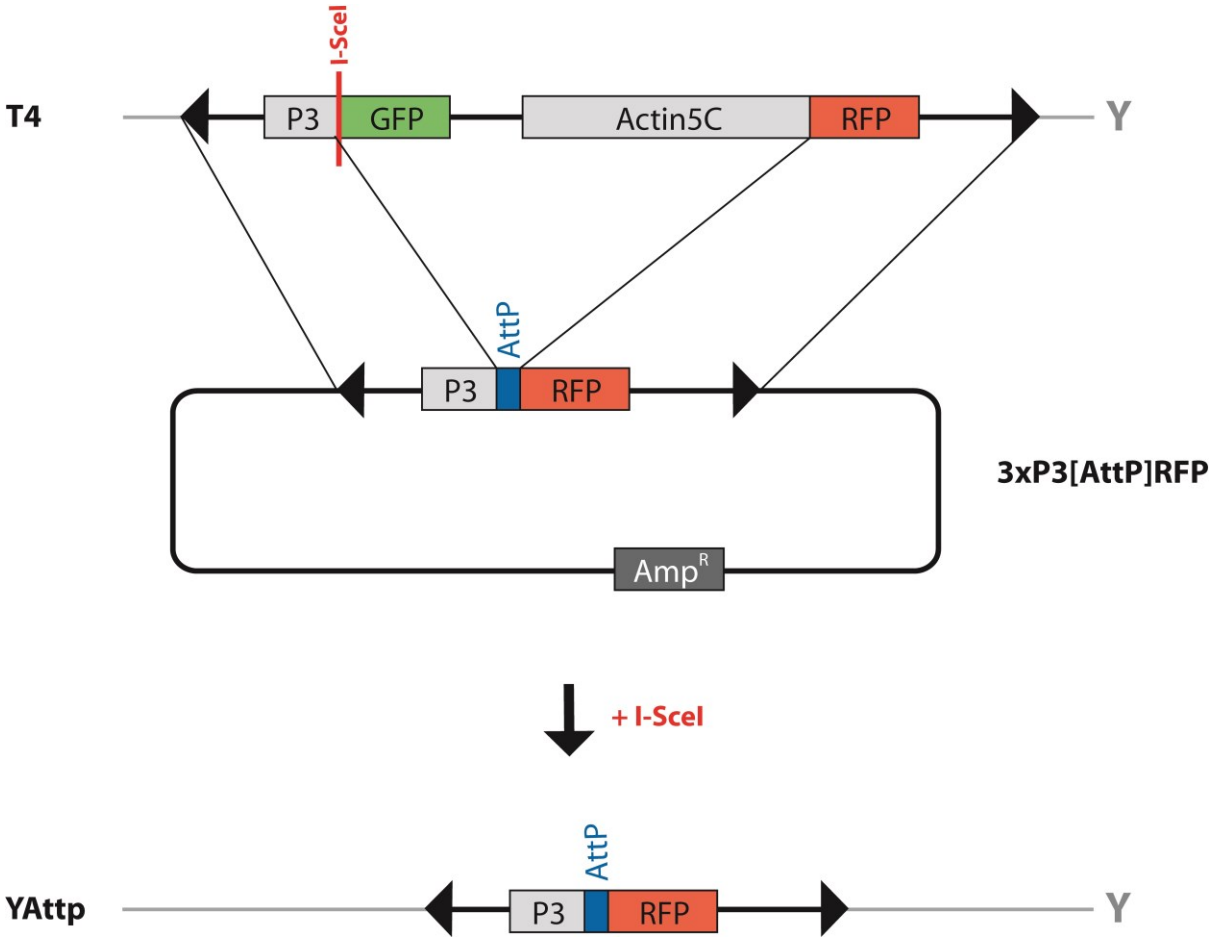


Figure 21: Generation of the transgenic line YAttp.

T4 embryos were injected with the plasmid 3Xp3[AttP]RFP together with an I-SceI donor. The 3Xp3[AttP]RFP plasmid carries the RFP marker under the control of the *3xP3* promoter. The plasmid was designed to contain, between the promoter and the RFP coding sequence, an *AttP* recombination site. Homologous recombination, induced by the endonuclease I-SceI, leads to the replacement of both the GFP coding sequence as well as the *Actin5C* promoter, present in the T4 transgene, with the *AttP* site. The resultant engineered Y chromosome in line YAttP is shown.

Source of I-SceI	Embryos injected	Transient males outcrossed	Larvae screened	Fluorescent phenotypes observed	
RNA	~2000	6	964	non-fluorescent	473 (♀)
				Actin5C-RFP/3xP3-GFP	461 (♂)
				Actin5C-RFP	30 (♂)
				3xP3-RFP	**0**
Plasmid	~4000	15	6160	non-fluorescent	3113 (♀)
				Actin5C-RFP/3xP3-GFP	2820 (♂)
				Actin5C-RFP	216 (♂)
				3xP3-RFP	**11** (♂)

Table 2: Embryo microinjections and fluorescent phenotypes observed in the outcross progeny.

T4 males injected with 3xP3[AttP]RFP plasmid and either in-vitro transcribed I-SceI RNA or the pVas2-I-SceI helper plasmid were outcrossed to wild-type females. The resulting progeny was screened for fluorescence. Phenotypes indicating successful gene conversion or integration are shown in bold.

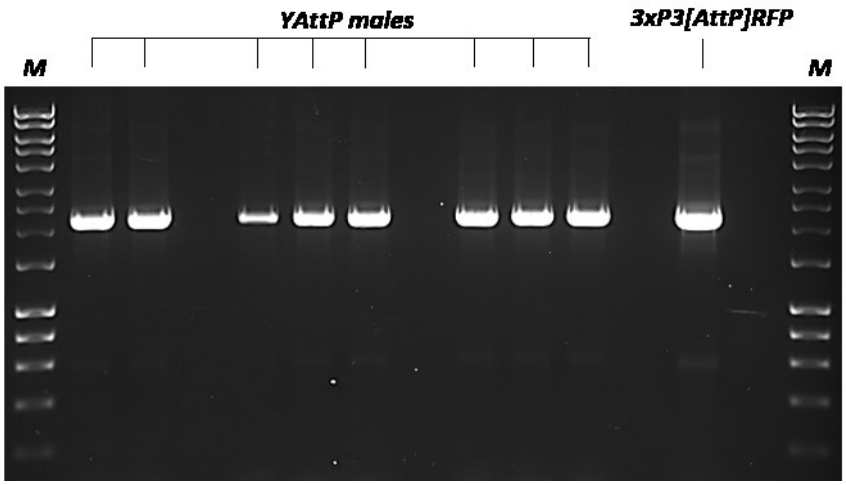


Figure 22: PCR analysis of the transgenic line YAttP.

PCR analysis was performed on genomic DNA of 8 YAttP males to confirm that cassette replacement in the T4 transgene occurred. 3xP3[AttP]RFP plasmid was used as positive control. In figure M indicates hyperladder I. A pair of primers that amply a region of 3xP3[AttP]RFP was used.

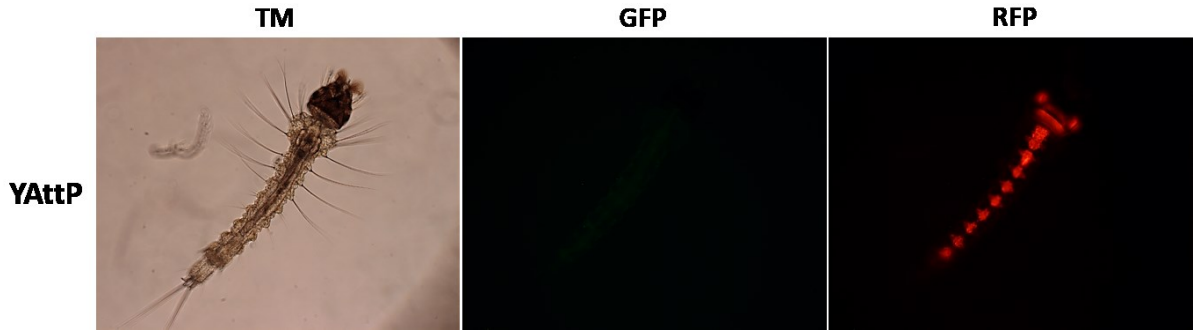


Figure 23: Fluorescent patter of YAttP larvae.

Expression of green (GFP) and red (RFP) fluorescent markers in the tissues of second instar larvae shown as transmitted-light image (TM) in *Left*.

4.1.6 Secondary integration of a germline-specific marker gene by site-specific recombination

The success of vector control strategies such as those based on homing endonuclease genes require the availability of germline-specific regulatory sequences suitable to drive gene expression during meiosis (Burt, 2003). It was previously shown that the regulatory regions of the *Anopheles* gene *vasa* drive expression of transgenes in the male and female germline throughout larval and adult gonad development (Papathanos *et al.*, 2009). A male-specific version of the *vasa* promoter was generated by using a shortened 5' untranslated region (UTR). However, this approach not only reduced overall activity of the promoter, but also led to the loss of expression in the male germline stem cells (GSCs) of the testes. An alternative strategy in order to express genes, driven by the *vasa* promoter in a male specific manner, would be to insert such constructs onto the Y chromosome. In this scenario, despite the *vasa* promoter drives expression of genes in both the male and female germline, it will only be present in males, carrying the X and the Y sexual chromosomes, and it will be absent in females, carrying only X chromosomes. To study the activity of germline-specific regulatory elements on the Y and to demonstrate the utility of line YAttP as a tool for the site-specific genetic engineering of the Y chromosome the attBCFP-VasGFP plasmid was generated (Figure 24). In this vector an *AttB* site was placed upstream of a promoterless eCFP (enhanced cyan fluorescent protein) coding sequence. In addition, it carries a cassette in which GFP is placed under the control of the *vasa* promoter. Recombination between *AttP* and *AttB* sites and integration of the plasmid was designed to replace the RFP coding sequnce with the CFP ORF and could be detected by screening for the shift to blue fluorescence in the progeny. Table 3 shows the outcome of experiments in which line YAttP was injected with attBCFP-VasGFP as well as a helper plasmid expressing the φC31 *integrase* from a *vasa* promoter sequence. A

total of 111 embryos were injected and 5 male survivors hatched. Since the plasmid attBCFP-VasGFP does not allow the detection of transient fluorescence in G0 individuals all male survivors were outcrossed to wild-type females. Out of 791 progeny screened 6 CFP positive larvae (0.76%) were obtained. Fluorescent pattern of the strain YVasG is shown in Figure 25.

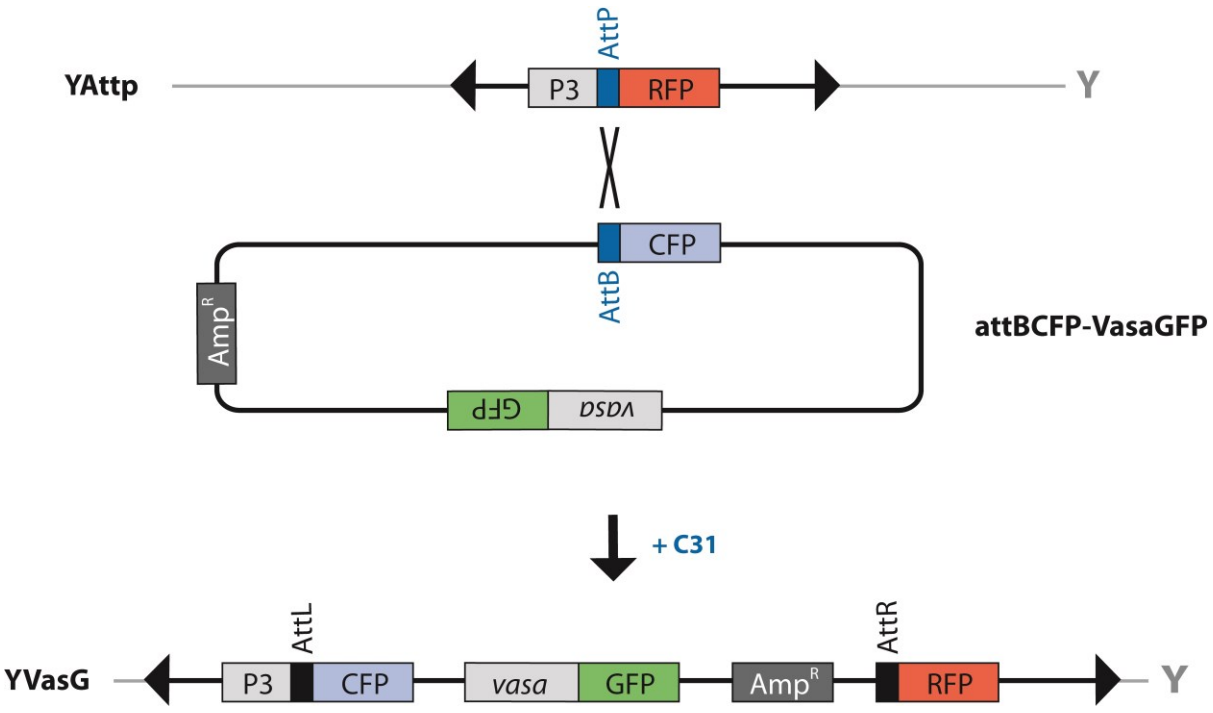


Figure 24: Generation of the transgenic line YVasG.

YAttP embryos were injected with the plasmid attBCFP-VasaGFP together with a φC31 integrase donor. In attBCFP-VasaGFP plasmid an *AttB* site was placed upstream of a promoterless eCFP coding sequence. In addition, it carries a cassette in which GFP is placed under the control of the *vasa* promoter. The integrase catalyses the recombination reaction between the attachment sites (*AttB* and *AttP*) bringing to the site-specific integration of the attBCFP-VasaGFP construct onto the Y chromosome.

Source of Integrase	Embryos injected	Male survivors outcrossed	Larvae screened	Fluorescent phenotypes observed	
Plasmid	111	5	791	non-fluorescent	390 (♀)
				3xP3-RFP	395 (♂)
				3xP3-CFP/Vas-GFP	**6** (♂)

Table 3: Embryo microinjections and fluorescent phenotypes observed in the outcross progeny.

YAttP males injected with attBCFP-VasaGFP plasmid and the integrase helper plasmid were outcrossed to wild-type females. The resulting progeny was screened for fluorescence. Phenotypes indicating successful gene conversion or integration are shown in bold.

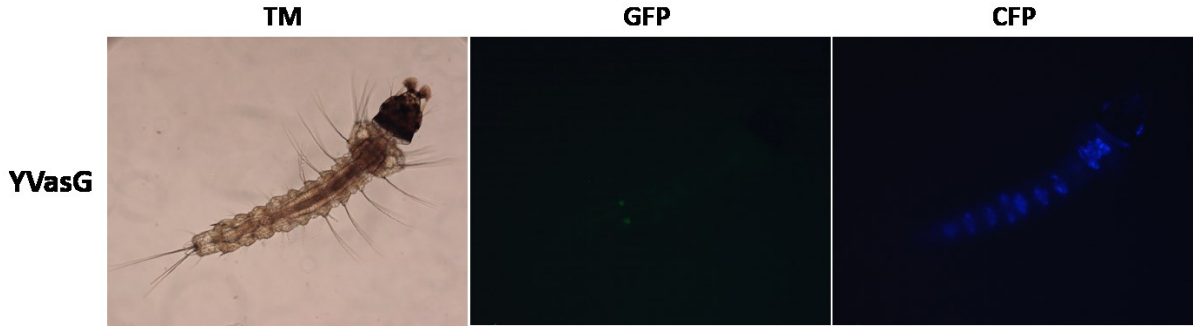


Figure 25: Fluorescent pattern in YVasG larvae.

Expression of green (GFP) and red (RFP) fluorescent markers in the tissues of second instar larvae shown as transmitted-light image (TM) in *Left*.

4.1.7 The *vasa* regulatory region drives GFP expression during early male spermatogenesis from the Y chromosome

All transgenic G2 larvae of strain YVasG showed a gonad-specific GFP signal (Figure 25) from the L1 larval stage onwards and developed into adult males as expected. Testes from male adults of the YVasG strain were dissected and analysed using confocal microscopy (Figure 26). GFP expression was detectable in all stages of spermatogenesis including the germline stem cells (GSCs) in the apical tip and developing spermatocytes but not in supporting somatic cells. DAPI staining of the hub region demonstrated that cells in the anterior tip of the testis were expressing GFP thus indicating that the regulatory regions of *vasa* were active in GSCs when located on the Y chromosome. The pattern of GFP expression is identical to testes from line Vas2GFP that expresses *vasa* driven GFP from position 2465559 of chromosome 3L. This suggests that the YAttP strain, in combination with the previously characterized non-sex-specific *vasa* promoter, allows the male-exclusive expression of transgenes starting at the earliest stages of spermatogenesis.

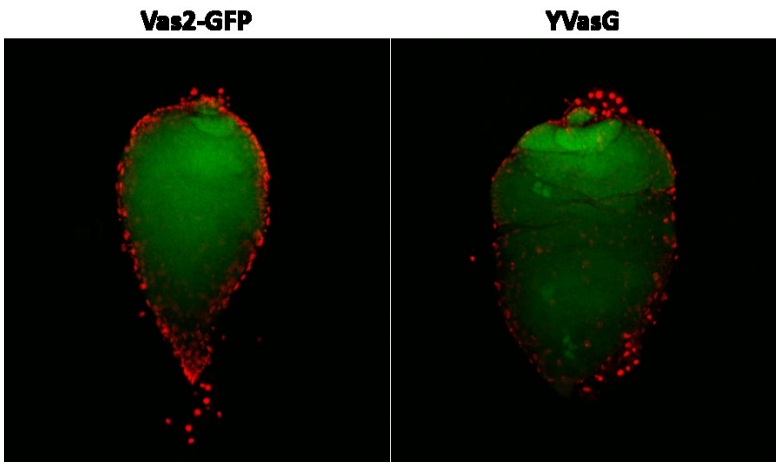


Figure 26: Confocal analysis of GFP expression in dissected testes of transgenic YVasG adult males.

In testes from YVasG adult males GFP expression (green) is detectable in all stages of spermatogenesis. The pattern of GFP expression is identical to testes from line Vas2GFP that expresses *vasa* driven GFP from position 2465559 of chromosome 3L. DAPI staining (red) of the hub region demonstrated that cells in the anterior tip of the testis were expressing GFP thus indicating that the regulatory regions of *vasa* were active in germline stem cells when located on the Y chromosome.

4.2 Introgression of a modified *A. gambiae* Y chromosome in the *A. arabiensis* genetic background

In the previews chapter the engineering of the *A. gambiae* Y chromosome has been described. The transgenic strain generated, YAttP, represents a valuable tool for the application of vector control strategies that target *A. gambiae* populations. *A. arabiensis* is considered a major vector of human malaria. An attempt to engineer the Y chromosome in this species, as previously described for *A. gambiae*, would be of great interest for vector control strategies that aim to eradicate malaria. Since active insertions onto the Y chromosome are extremely rare, a scheme based on crossing and selection was used to overcome F1 hybrid males sterility, predicted by Haldane's rule (Haldane, 1922), and introgress the modified *A. gambiae* Y chromosome in the *A. arabiensis* genetic background. The introgression of two modified *A. gambiae* Y chromosomes was attempted:

- The YAttP chromosome, generated as part of this PhD project and previously described (Bernardini *et al.*, 2014).
- The YI-PpoI chromosome. Despite active insertion of transgenic constructs onto the Y chromosome have been described as extremely rare, the strain YI-PpoI was generated by transposase-mediated random integration during a different experimental project in our laboratory (unpublished data). In this strain, named YI-PpoI, the Y-linked construct carries the RFP marker under the control of the neuronal *3xP3* promoter and the endonuclease *I-PpoI* driven by the male spermatogenesis-specific *b2 tubulin* promoter (Windbichler *et al.,* 2008; Catteruccia *et al.*, 2005) . As described in the introduction, the endonuclease I-PpoI has high specificity for a conserved sequence within the ribosomal rDNA repeats located in a single cluster on

the X chromosome. Expression of *I-PpoI* during spermatogenesis is expected to cause a selective "shredding" of the X chromosome leading to the production of only Y chromosome bearing sperms. In the transgenic strain YI-PpoI, despite the RFP phenotype is clearly visible, the endonuclease *I-PpoI* is not expressed and therefore no sex distortion is achieved in the progeny. Inactivation of the *I-PpoI* is due to meiotic sex chromosomes inactivation (MSCI) a mechanism responsible for the silencing of chromosomes that fail to pair with their homologues partners during meiosis (Vibranovski, 2014). The introgression of the *A. gambiae* YI-PpoI chromosome was attempted in the *A. arabiensis* genetic background to assess any activation of the Y-linked endonuclease *I-PpoI* gene when present in a hybrid genetic scenario.

Four different mosquito strains were involved in the experiments described in this chapter (see appendix):

- *A. gambiae* (G3) male and female mosquitoes
- *A. arabiensis* (Dongola) male and female mosquitoes
- *A. gambiae* YAttP male mosquitoes
- *A. gambiae* YI-PpoI male mosquitoes

A. gambiae and the *A. arabiensis* share the same number of chromosomes with two pairs of sub metacentric autosomes and a pair of heteromorphic sex chromosomes X and Y, XX in females and XY in males. A schematic representation of these mosquitoes' karyotype is used in the present chapter to describe the genetic crosses involved in the Y chromosome introgression experiment.

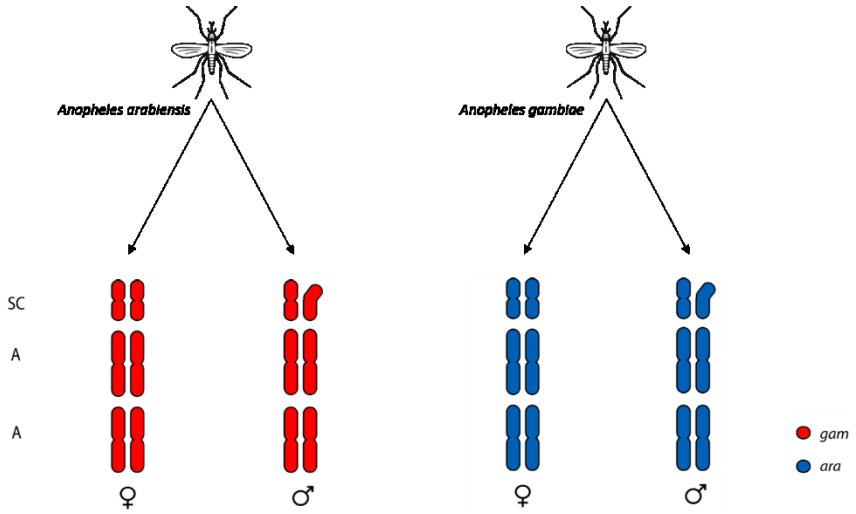


Figure 27: *A. gambiae* and *A. arabiensis* karyotype.

Karyotype in *Anopheles gambiae* and the *Anopheles arabiensis* comprises two pairs of submetacentric autosomes, termed 2 and 3, and a pair of heteromorphic sex chromosomes X and Y, XX in females and XY in males. Paste legend from previews pictures. SC: sexual chromosomes (XX in females, XY in males). A: autosomes. gam: *A. gambiae* genome. ara: *A. arabiensis* genome. No sex ratio distortion was observed in the progeny.

4.2.1 Introgression attempt of YAttP and YI-PpoI chromosomes through F1 hybrid females generated from *A. gambiae* females mated with *A. arabiensis* males

Introgression of the modified *A. gambiae* Y chromosomes, YAttP and YI-PpoI, was attempted in the *A. arabiensis* genetic background. According to Haldane's rule (Haldane, 1922), F1 hybrid males sterility is known to occur in all intercrosses in the *A. gambiae* complex (Coluzzi *et al.*, 1979). The strategy used for the introgression experiment here described involved the initial cross of *A. gambiae* females to *A. arabiensis* males to generate a large number of *F1 hybrids* (Figure 28). When *F1 hybrid* males were crossed either to *A. arabiensis* or *A. gambiae* females, sterility was observed (eggs collected after females blood feeding did not hatch) confirming that Haldane's rule holds true for these species. In order to overcome *F1 hybrid* males sterility, the introgression experiment was carried out through *F1 hybrid* fertile females. Crosses were set up in two separate cages, one containing *F1 hybrid* females and *A. gambiae* YAttP males, the other containing *F1 hybrid* females and *A. gambiae* YI-PpoI males. To simplify the description of the experiments I refer to the YAttP and YI-PpoI males as *transgenic* males (Figure 29). In the transgenic strains the RFP contained in the Y-linked constructs allows the selection of males at the very early larval stage by screening for *3xP3* RFP phenotype. From both the cages a large number of progeny was recovered (>2000), I refer to these progeny as *class 1* males and females. For the aim of the experiment *class 1* males were selected and backcrossed to *A. arabiensis* females (Figure 30). A large number of progeny was recovered (>2000) and males, named *class 2* males, were kept to set up the second backcross to *A. arabiensis* females. Fertility of these males was also tested in crosses with females exhibiting a different genetic background, cross 1 and cross 2 (Figure 31). After mating and blood feeding of the females, eggs were collected from all the experimental cages (six), a total number of

4052 eggs from the YAttP cages and 9384 eggs from the YI-PpoI cages. After several days none of these eggs hatched suggesting sterility associated to *class 2* males (Table 4).

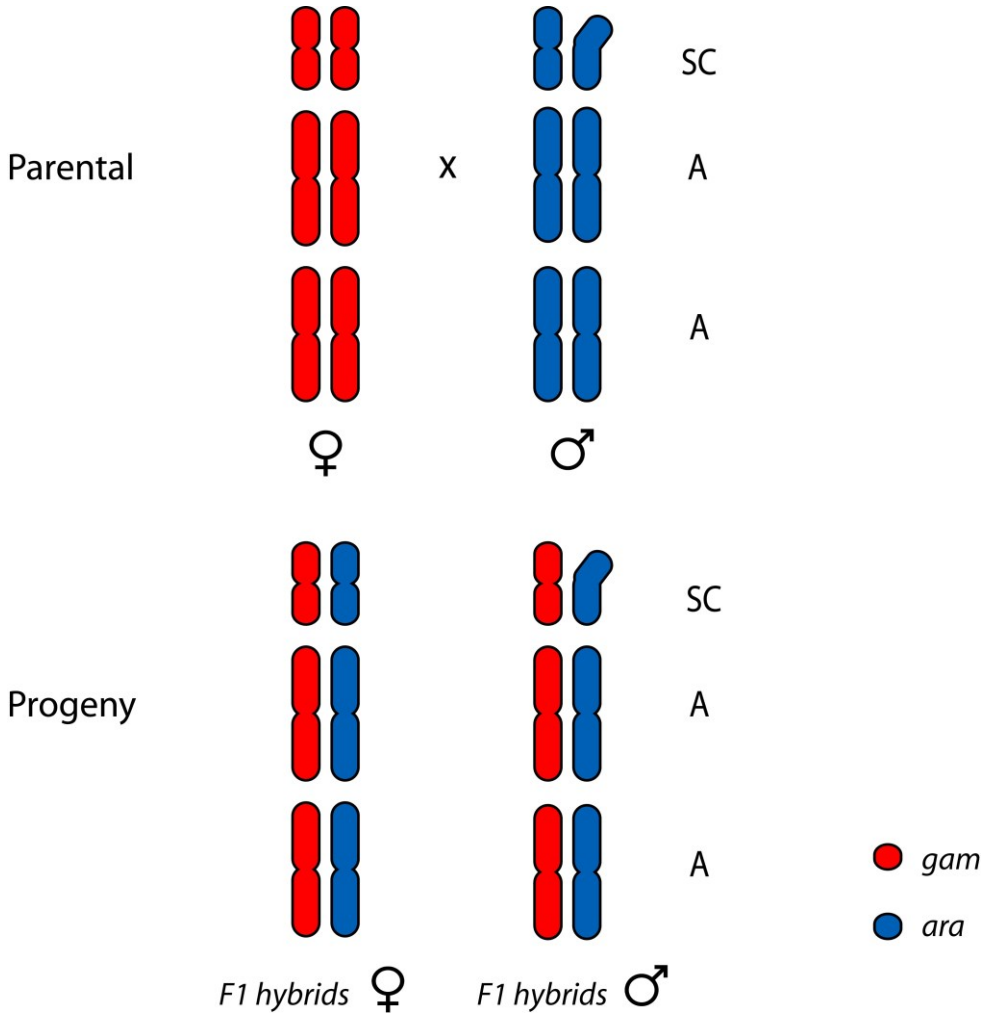


Figure 28: *A. gambiae* females crossed to *A. arabiensis* males.

100 *A. gambiae* females were crossed to 100 *A. arabiensis* males to generate *F1 hybrids*. *F1 hybrid* males, the heterogametic sex, exhibit sterility, while *F1 hybrid* females, the homogametic sex, are fully fertile. SC: sexual chromosomes (XX in females, XY in males). A: autosomes. gam: *A. gambiae* genome. ara: *A. arabiensis* genome. No sex ratio distortion was observed in the progeny.

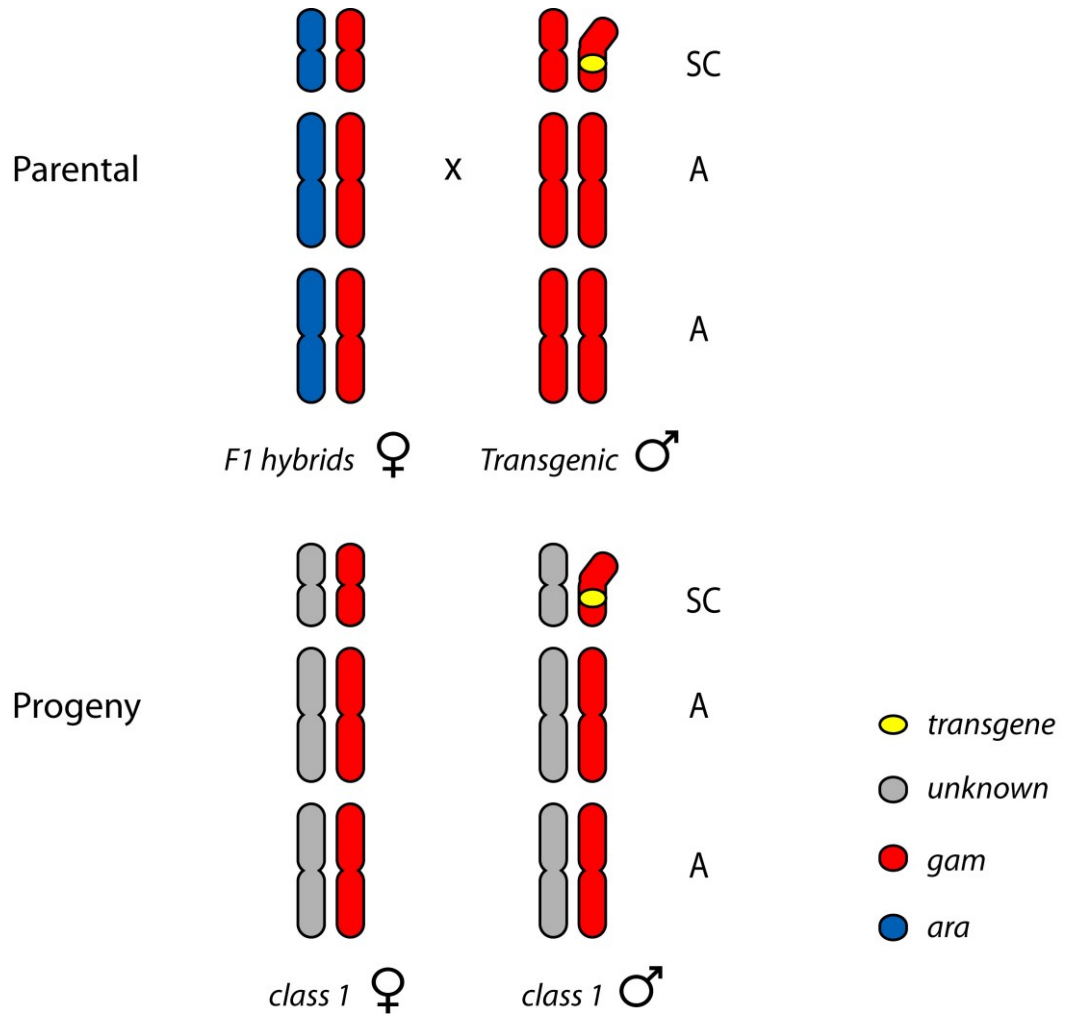


Figure 29: *F1 hybrid* females crossed to *A. gambiae transgenic* males.

100 *F1 hybrid* females were crossed to 100 *A. gambiae transgenic* males. The transgene is inserted onto the Y chromosome. In the progeny recovered, *class 1* males and females, the paternal genetic contribution has an *A. gambiae* composition whereas, the maternal genetic contribution is the result of meiotic recombination between hybrid chromosomes, therefore the exact genetic contribution of the two species cannot be predicted. SC: sexual chromosomes (XX in females, XY in males). **A:** autosomes. gam: *A. gambiae* genome. ara: *A. arabiensis* genome. unknown: relative amount of *A. gambiae* and *A. arabiensis* genome unknown. transgene: construct inserted onto the Y chromosome. No sex ratio distortion was observed in the progeny.

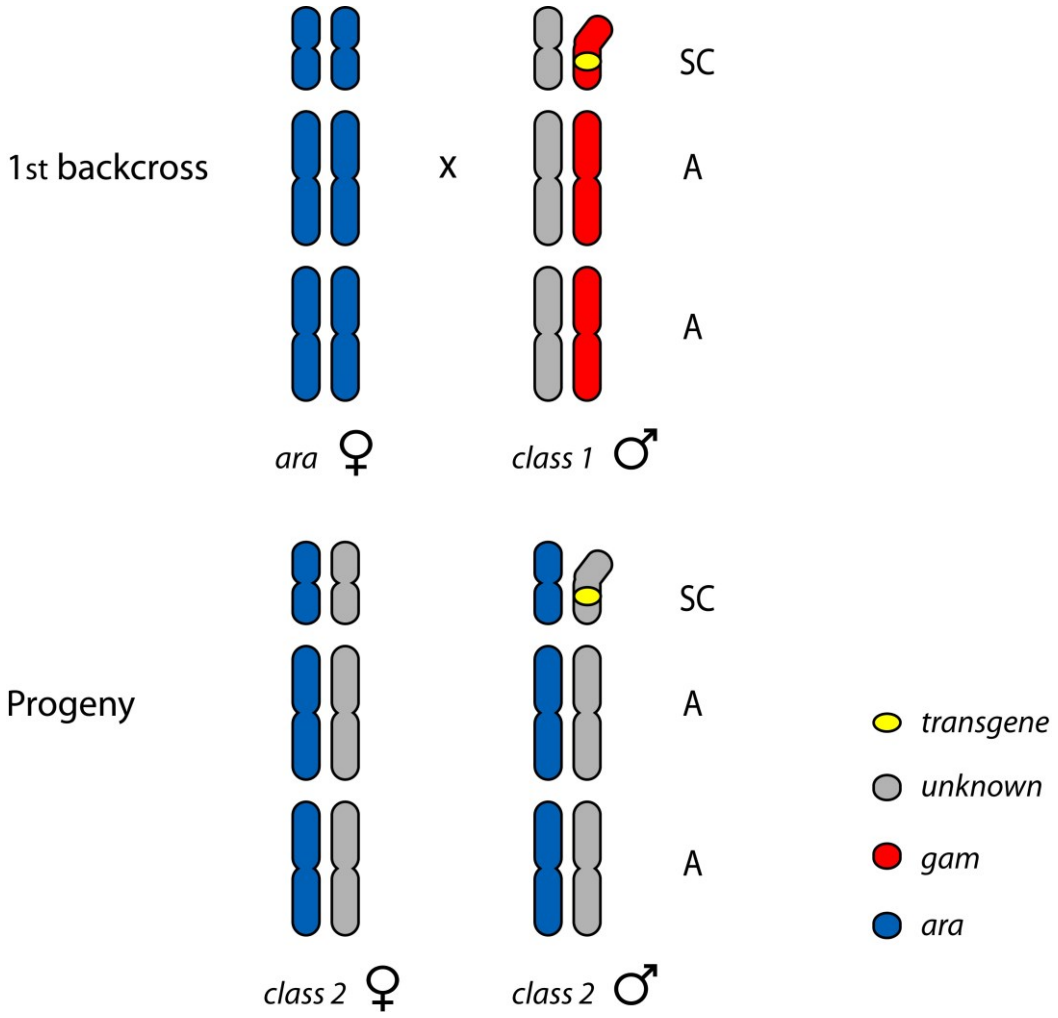


Figure 30: *Class 1* males crossed to *A. arabiensis* females.

100 *class1* males were backcrossed to 100 *A. arabiensis* females. A large number of progeny, identified as *class2* females and males, were recovered from the cage. SC: sexual chromosomes (XX in females, XY in males). A: autosomes. gam: *A. gambiae* genome. ara: *A. arabiensis* genome. unknown: relative amount of *A. gambiae* and *A. arabiensis* genome unknown. transgene: construct inserted onto the Y. No sex ratio distortion was observed in the progeny.

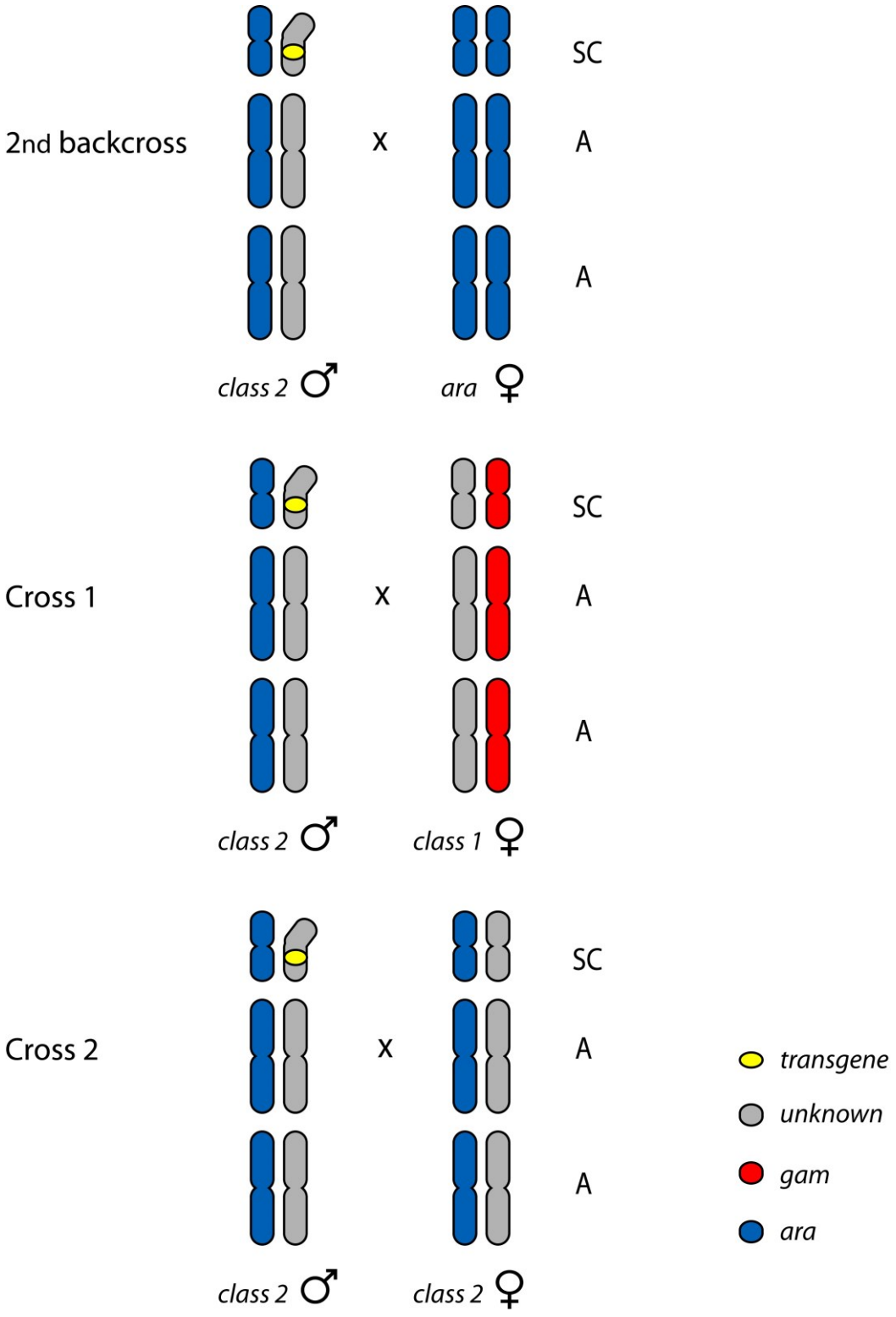


Figure 31: *Class 2* males crossed to *A. arabiensis* females (2nd backcross). *Class 2* males crossed to *class 1* females (cross 1). *Class 2* males crossed to *class 2* females (cross 2).

Crosses of *class 2* males with females exhibiting a different genetic background are shown. SC: sexual chromosomes (XX in females, XY in males). A: autosomes. gam: *A. gambiae* genome. ara: *A. arabiensis* genome. unknown: relative amount of *A. gambiae* and *A. arabiensis* genome unknown. transgene: construct inserted onto the Y chromosome. No sex ratio distortion was observed in the progeny.

a)

YAttP	N. of eggs recovered	N. of larvae hatched
2nd backcross	*408*	*0*
Cross 1	*300*	*0*
Cross 2	*3344*	*0*
Total	*4052*	*0*

b)

YI-PpoI	N. of eggs recovered	N. of larvae hatched
2nd backcross	*827*	*0*
Cross 1	*3347*	*0*
Cross 2	*5210*	*0*
Total	*9384*	*0*

Table 4: Hatching rate recovered from crosses described in Figure 30.

a) Data collected from the cages where the introgression of the YAttP chromosome was attempted.
b) Data collected from the cages where the introgression of the YI-PpoI chromosome was attempted.

No larvae were recovered from the eggs collected.

In order to overcome the sterility barrier encountered in the first batch of experiments new crosses were set up. C*lass 1* males (YAttP and YI-PpoI) were crossed in two separate cages to *class 2* females (sibling of YAttP-Y and YI-PpoI *class 2* males respectively) (Figure 32). A large number of progeny was recovered from the crosses. Male progeny, *class 3* males, underwent the first backcross to *A. arabiensis* females (Figure 33). A number of 251 out of 1450 larvae were recovered for the *YAttP* cage, while 8 out of 987 larvae were collected from the *YI-PpoI* cage (Table 5). Male progeny, named *class 4* males, was selected to undergo the second backcross to *A. arabiensis* females (Figure 34). A number of 170 and 5100 eggs were recovered respectively from the YAttP cage and the YI-PpoI cage. After several days none of these eggs hatched in larvae (Table 5). The experiments described in this paragraph led to the occurrence

of male sterility and therefore to the interruption of the attempt to introgress the *A. gambiae* Y chromosome into the *A. arabiensis* genetic background. These results highlight the important role of post-zygotic mechanisms in limiting genetic flow between different species. However, a slightly different approach was undertaken in an attempt to achieve the Y chromosome introgression. Experiments in relation to this attempt are described in the next paragraph.

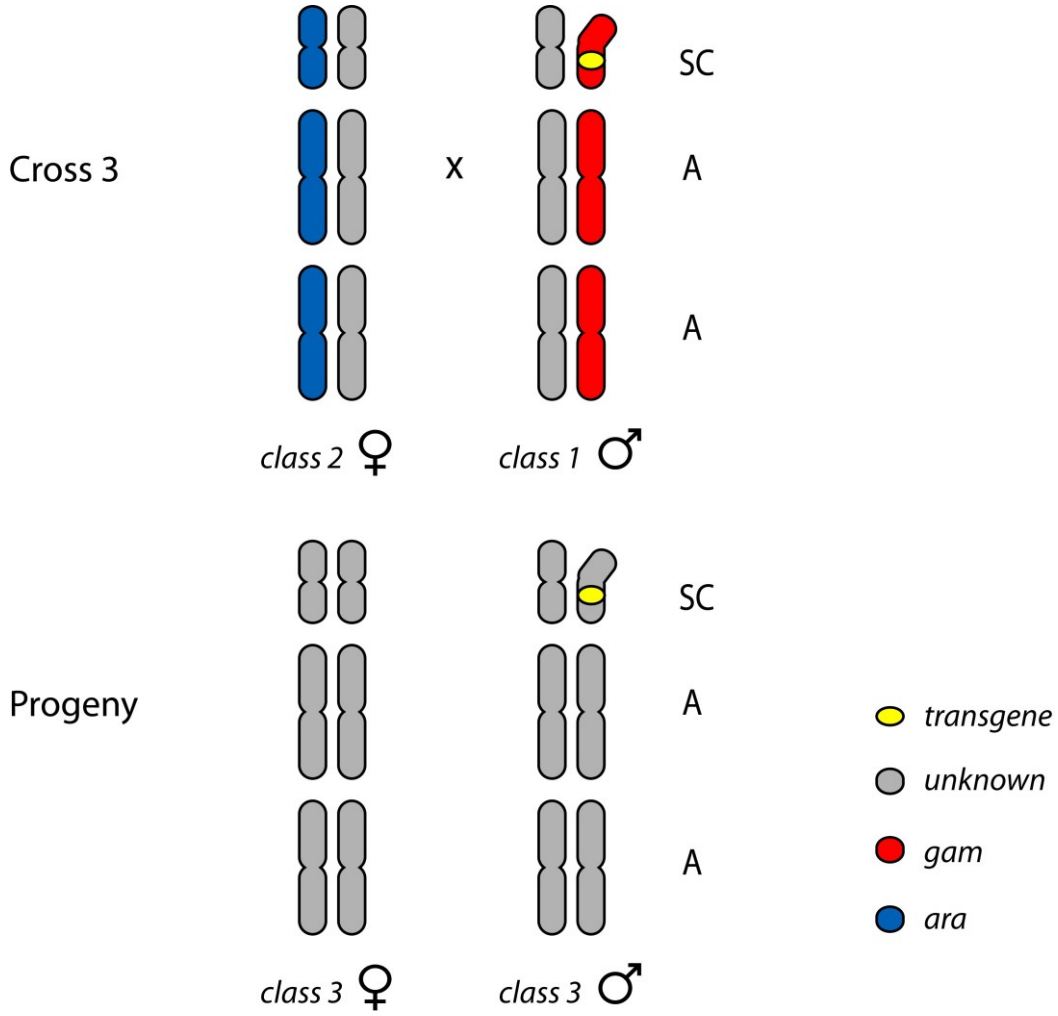


Figure 32: *Class 1 males crossed to class 2* females.

100 *class1* males were crossed to 100 *class 2* females. A large number of progeny, identified as *class 3* females and males, were recovered from the cage. SC: sexual chromosomes (XX in females, XY in males). A: autosomes. gam: *A. gambiae* genome. ara: *A. arabiensis* genome. unknown: relative amount of *A. gambiae* and *A. arabiensis* genome unknown. transgene: construct inserted onto the Y chromosome. No sex ratio distortion was observed in the progeny.

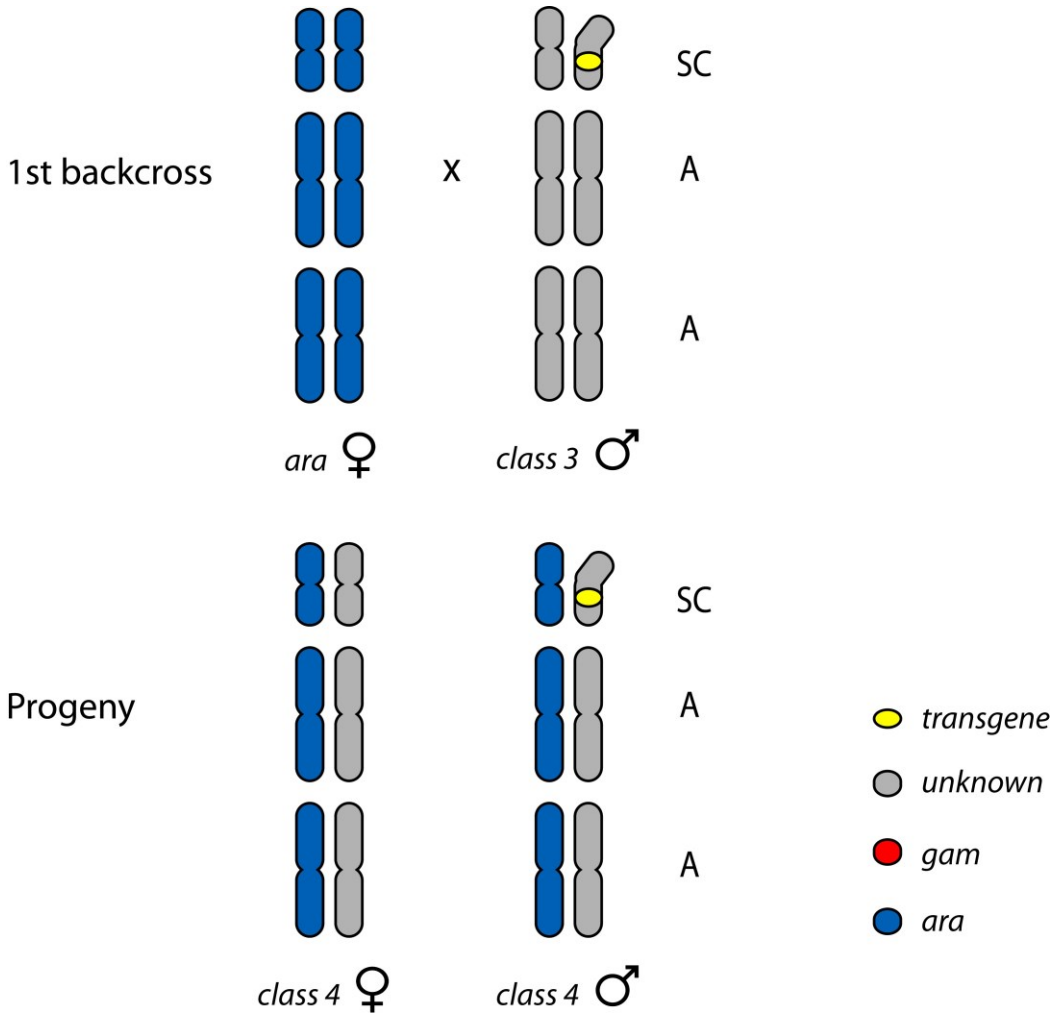


Figure 33: *Class 3 males crossed* to *A. arabiensis* females.

100 *class 3* males were crossed to 100 *A. arabiensis* females. Progeny, identified as *class 4* females and males, were recovered from the cage. SC: sexual chromosomes (XX in females, XY in males). A: autosomes. gam: *A. gambiae* genome. ara: *A. arabiensis* genome. unknown: relative amount of *A. gambiae* and *A. arabiensis* genome unknown. transgene: construct inserted onto the Y chromosome. No sex ratio distortion was observed in the progeny.

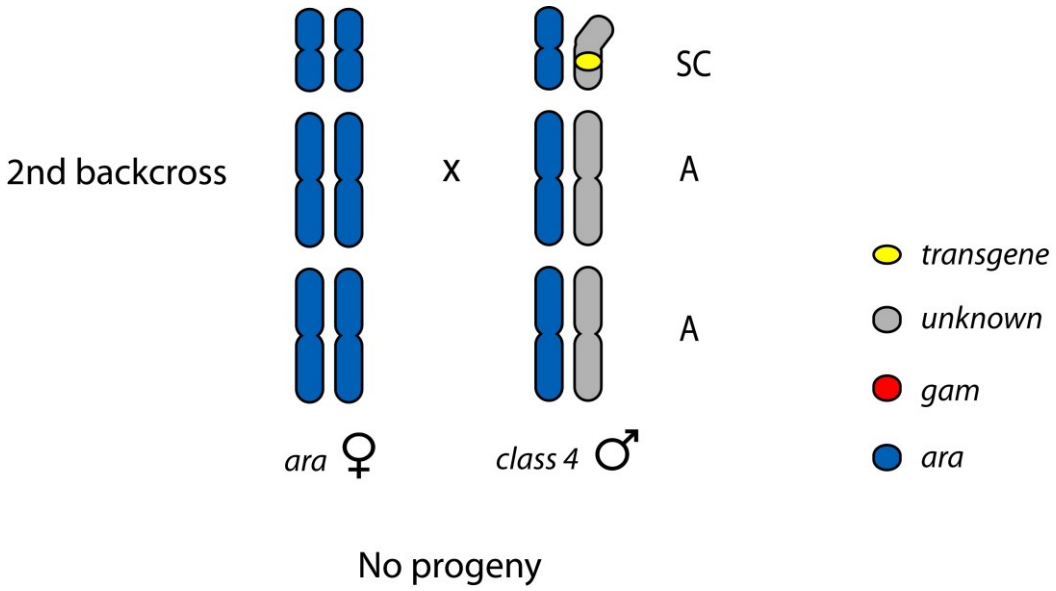


Figure 34: *Class 4 males* crossed to *A. arabiensis* females.

C*lass 4* males were crossed to 100 *A. arabiensis* females. None of the collected eggs hatched in larvae. SC: sexual chromosomes (XX in females, XY in males). A: autosomes. gam: *A. gambiae* genome. ara: *A. arabiensis* genome. unknown: relative amount of *A. gambiae* and *A. arabiensis* genome unknown. transgene: construct inserted onto the Y chromosome. No sex ratio distortion was observed in the progeny.

a)

YAttP	N. of eggs recovered	N. of larvae hatched
1st backcross	*1450*	*251*
2st backcross	*170*	*0*

b)

YI-PpoI	N. of eggs recovered	N. of larvae hatched
1st backcross	*987*	*8*
2nd backcross	*5100*	*0*

Table 5: Hatching rate recovered from crosses described in Figure 32 and 33.

a) Data collected from the cages where the introgression of the *YAttP* was attempted.
b) Data collected from the cages where the introgression of the *YI-PpoI* was attempted.

Eggs collected from the second backcross to *A. arabiensis* females did not hatch.
In figure arrows indicate the source of males for each backcross.

4.2.2 Introgression of *YAttP* and *YI-PpoI* chromosomes through F1 hybrid females generated from *A. arabiensis* females mated with *A. gambiae* males

The experiments described in this session followed the same strategy applied for the first introgression attempt with the exception that sex of the parental was inverted. *A. arabiensis* females were crossed to *A. gambiae* males to generate *F1 hybrids** (Figure 35). As previously assessed and according to Haldane's rule (Haldane, 1922), *F1 hybrid** males showed sterility when crossed either to *A. gambiae* or *A. arabiensis* females (the eggs collected after females blood feeding did not hatch). The introgression experiment was carried out through the *F1 hybrid** fertile females. Crosses were set up in two separate cages, one containing *F1 hybrids** females and *A. gambiae* YAttP males, the other containing *F1 hybrids** females and *A. gambiae* YI-PpoI males. To simplify the description of the experiment I refer to the YAttP-Y and YI-PpoI males as *transgenic* males (Figure 36). A large number of progeny was recovered from both cages (>2000), I refer to these progeny as *class 1** males and females. *Class 1** males were selected and backcrossed to *A. arabiensis* females (Figure 37). The male progeny recovered, *class 2** males, underwent the second backcross to *A. arabiensis* females. In addiction these males were crossed to different classes of females, cross 1 and cross 2 (Figure 38). After mating and blood feeding of the females, eggs were collected from all the experimental cages (six). Of the YAttP cages, second backcross and cross 2 gave a number of 400 and 2638 eggs respectively that did not hatch, while 34 out of 6992 eggs collected form cross 1 hatched in larvae (Table 6). Of the YI-PpoI cages, cross 2 gave 1420 eggs that did not hatch, while 7 out of 2095 and 157 out of 2795 eggs, from second backcross and cross 1 respectively, hatched in larvae (Table 6). Males were selected from the hatched larvae and used to carry on the introgression experiment. To simply the description of the data, the introgression of the two

modified Y chromosomes (YAttP-Y and YI-PpoI) will be from this point onwards treated separately.

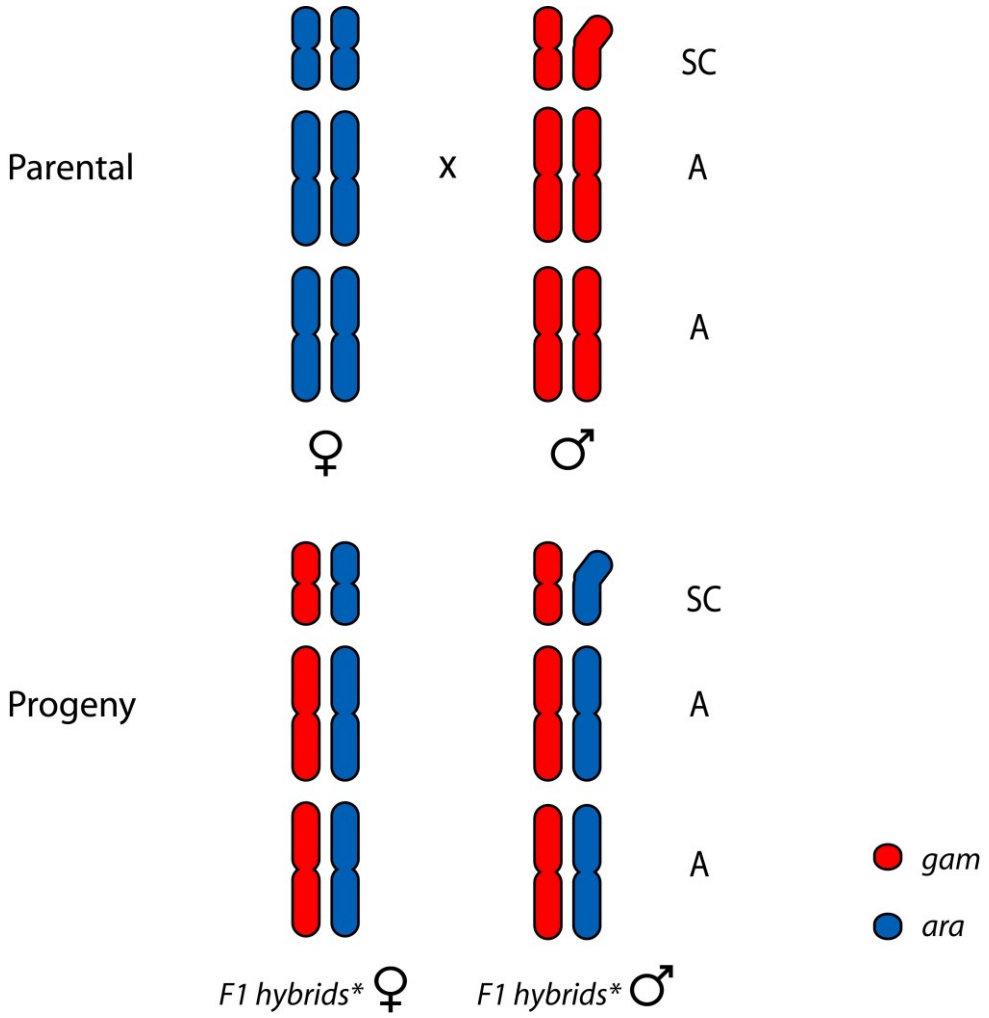


Figure 35: *A. arabiensis* females crossed to *A. gambiae* males.

100 *A. arabiensis* females were crossed to 100 *A. gambiae* males to generate *F1 hybrids**. *F1 hybrid** males, the heterogametic sex, exhibit sterility, while *F1 hybrid** females, the homogametic sex, are fully fertile. SC: sexual chromosomes (XX in females, XY in males). A: autosomes. gam: *A. gambiae* genome. ara: *A. arabiensis* genome. No sex ratio distortion was observed in the progeny.

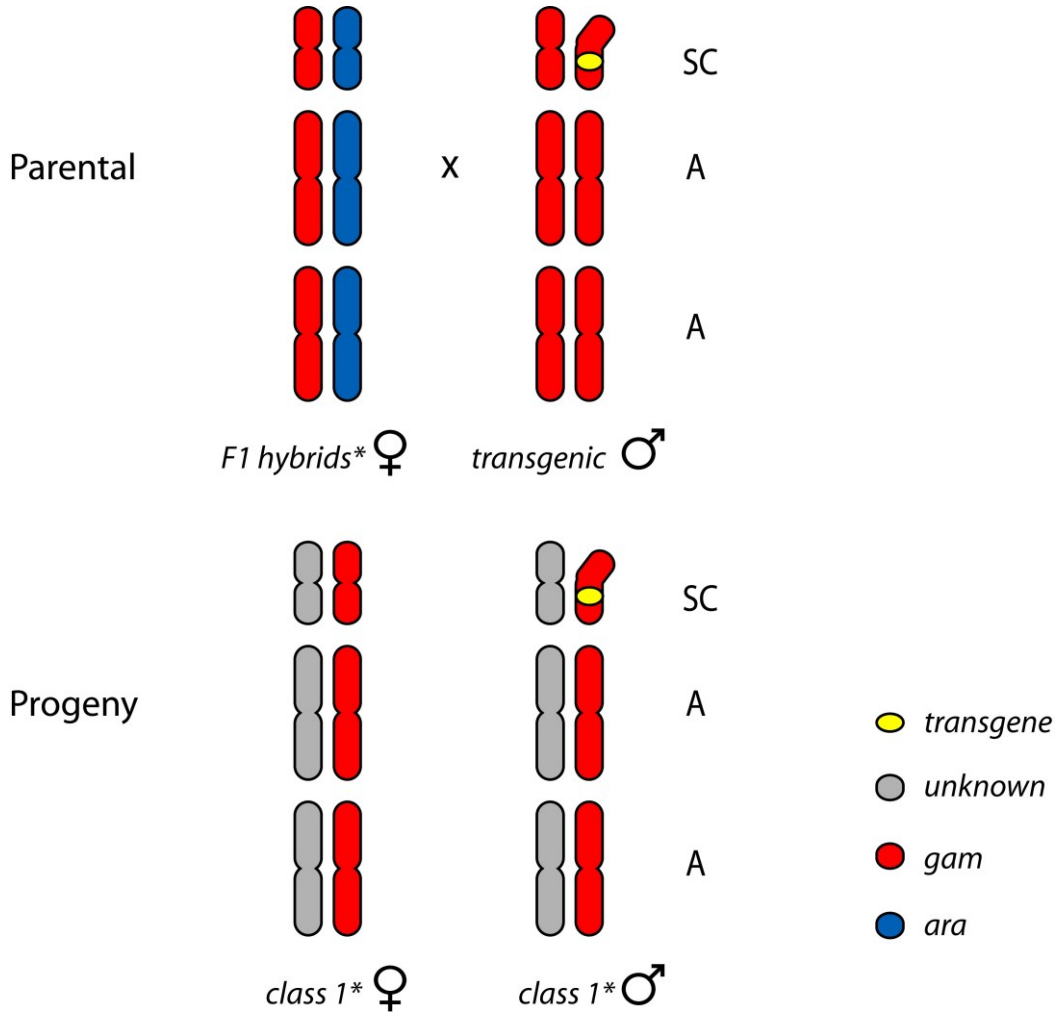


Figure 36: *F1 hybrid females crossed to *A. gambiae transgenic* males.**

100 *F1 hybrid** females were crossed to 100 *A. gambiae transgenic* males. The transgene is inserted onto the Y chromosome. In the progeny recovered, *class 1** males and females, the paternal genetic contribution has an *A. gambiae* composition whereas, the maternal genetic contribution is the result of meiotic recombination between hybrid chromosomes, therefore the exact genetic contribution of the two species cannot be predicted. SC: sexual chromosomes (XX in females, XY in males). A: autosomes. gam: *A. gambiae* genome. ara: *A. arabiensis* genome. unknown: relative amount of *A. gambiae* and *A. arabiensis* genome unknown. transgene: construct inserted onto the Y chromosome. No sex ratio distortion was observed in the progeny.

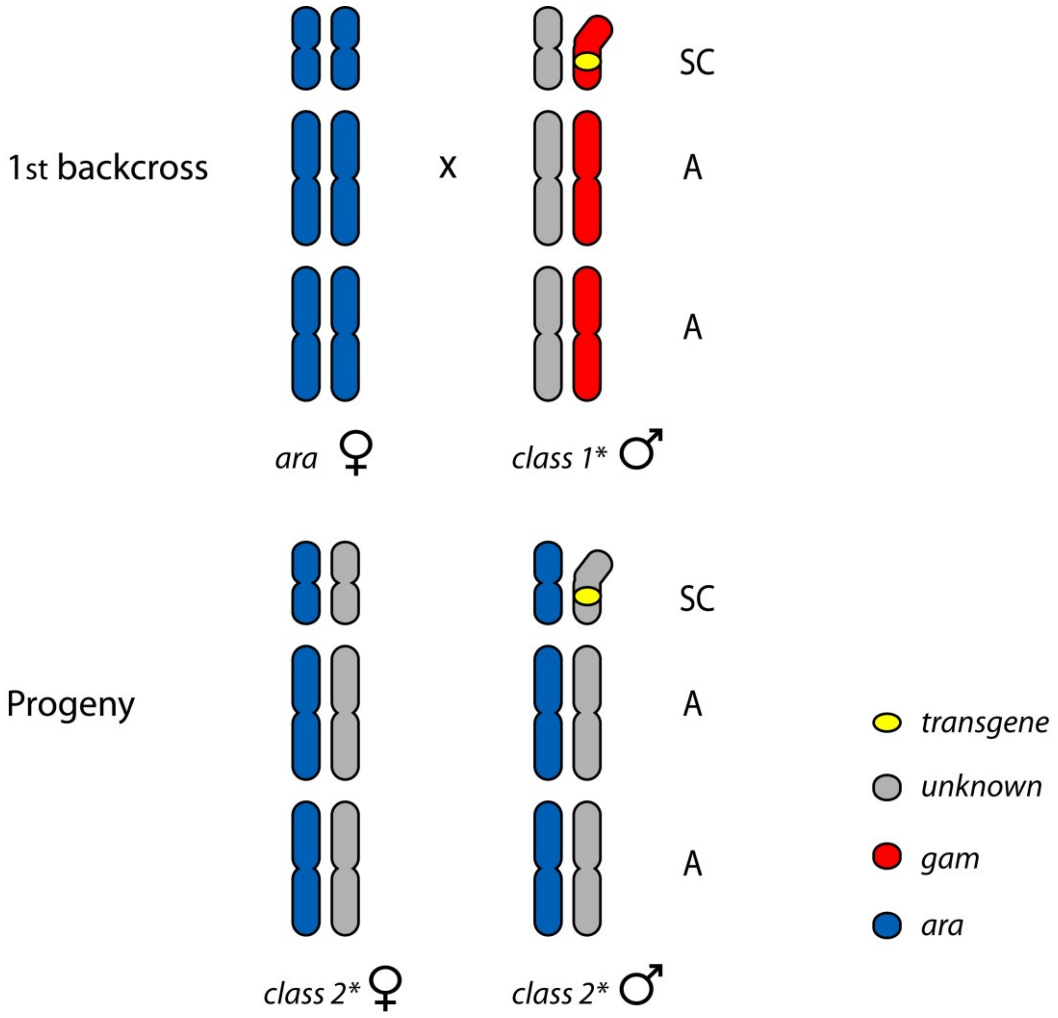


Figure 37: *Class 1 males crossed to *A. arabiensis* females.**

100 *class1** males were backcrossed to 100 *A. arabiensis* females. A large number of progeny, identified as *class2** females and males, were recovered from the cage. SC: sexual chromosomes (XX in females, XY in males). A: autosomes. gam: *A. gambiae* genome. ara: *A. arabiensis* genome. unknown: relative amount of *A. gambiae* and *A. arabiensis* genome unknown. transgene: construct inserted onto the Y chromosome. No sex ratio distortion was observed in the progeny.

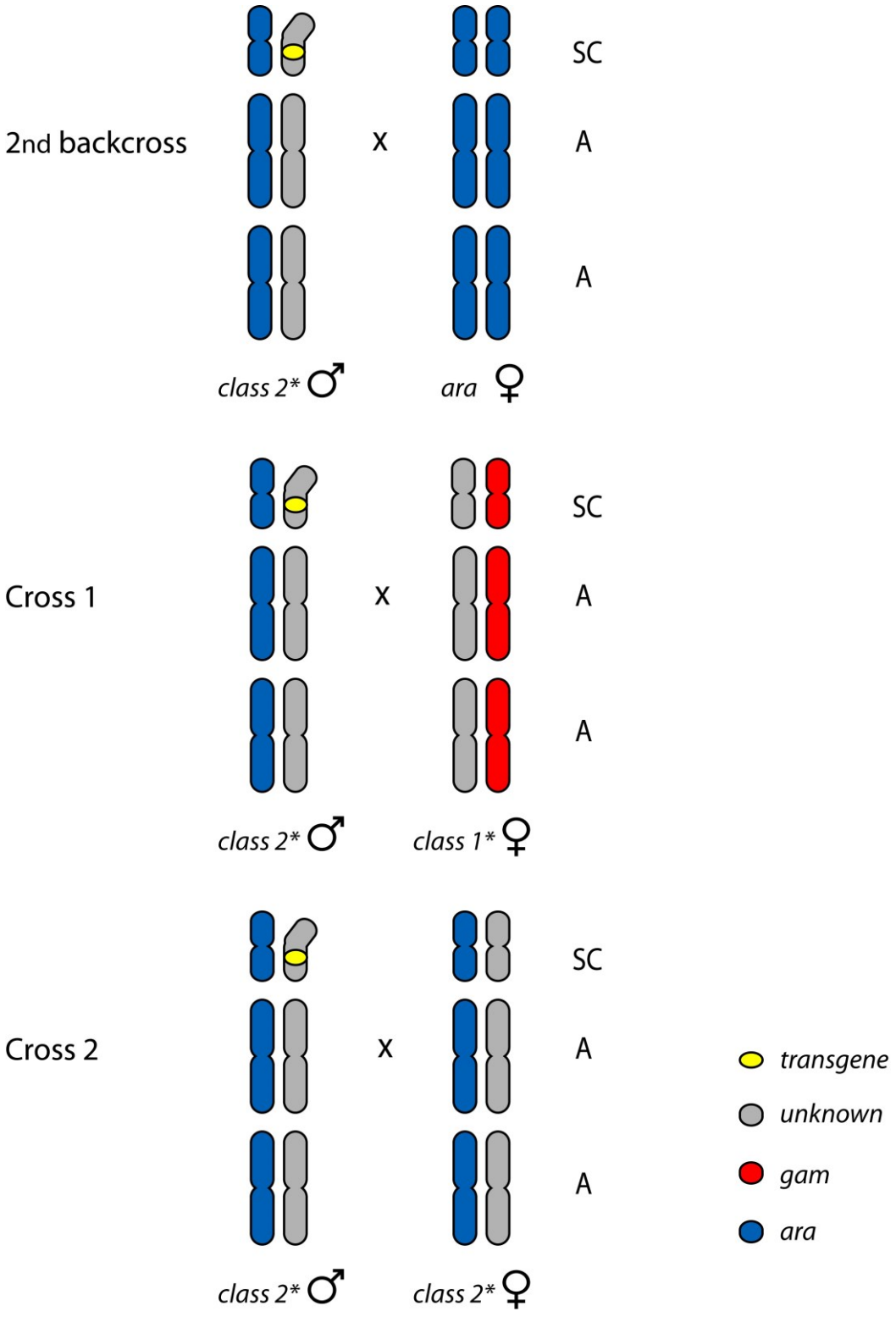


Figure 38: *Class 2 males crossed to *A. arabiensis* females (2nd backcross). *Class 2** males crossed to *class 1** females (cross 1). *Class 2** males crossed to *class 2** females (cross 2).**

Crosses of *class 2** males with females exhibiting a different genetic background are shown. SC: sexual chromosomes (XX in females, XY in males). A: autosomes. gam: *A. gambiae* genome. ara: *A. arabiensis* genome. unknown: relative amount of *A. gambiae* and *A. arabiensis* genome unknown. transgene: construct inserted onto the Y chromosome. No sex ratio distortion was observed in the progeny.

a)

YAttP	N. of eggs recovered	N. of larvae hatched
2nd backcross	*400*	*0*
Cross 1	*6992*	*34*
Cross 2	*2638*	*0*
Total	*10030*	*34*

b)

YI-PpoI	N. of eggs recovered	N. of larvae hatched
2nd backcross	*2095*	*7*
Cross 1	*2795*	*15*
Cross 2	*1420*	*0*
Total	*6310*	*22*

Table 6: Hatching rate recovered from crosses described in Figure 37.

a) Data collected from the cages where the introgression of the YAttP chromosome was attempted.
b) Data collected from the cages where the introgression of the YI-PpoI chromosome was attempted.

Male progeny selected from the hatched eggs was used to carry on the introgression.

4.2.2.1 *Advanced introgression of the A. gambiae YI-PpoI chromosome*

Progeny recovered from YI-PpoI cross 1 and second backcross (Table 6) underwent a process of selection of males and backcross to *A. arabiensis* females. The same strategy was applied to progeny of further generations. While the introgression in males derived from cross 1 failed after a few generations due to emergence of sterility (data not shown), the introgression in males derived from the second backcross has been carried on for many generations (up to 20 so far). At any generation, after mating and blood feeding of females, eggs were collected from the experimental cage and the hatching rate was measured. Data collected for 11 generations of backcrossing in cage of YI-PpoI-introgressed males to *A. arabiensis* females are shown in Table 7. The hatching rate values showed the occurrence of bottlenecks through the generations analysed. However, further backcrossing in cage, led eventually to the overcoming of the bottlenecks observed in earlier generations (data not shown) and a large number of progeny was recovered from the backcrosses in cage.

YI-PpoI	N. of eggs recovered	N. of larvae hatched
1st backcross	*>2000*	*>2000*
2nd backcross	*2095*	*7*
3rd backcross	*124*	*97*
4th backcross	*640*	*376*
5th backcross	*150*	*125*
6th backcross	*500*	*9*
7th backcross	-	*120*
8th backcross	*400*	*22*
9th backcross	*200*	*120*
10th backcross	*300*	*41*
11th backcross	*1200*	*980*

Table 7: Hatching rate recovered from backcrosses of *YI-PpoI*-introgressed males to *A. arabiensis* females in cage.

In figure arrows indicate the source of males for each backcross.

4.2.2.2 Single backcrosses of YI-PpoI-introgressed males in cups

When the hatching rate is measured from eggs collected from multiple crosses in cage, paternity and maternity cannot be assigned to the eggs. To assess fertility, after 11 generations of introgression in cage, single backcrosses of *YI-PpoI*-introgressed males to *A. arabiensis* females were set up in cups. From the 11th backcross in cage 980 larvae out of 1200 eggs hatched (Table 7). A number of 50 males were randomly selected. Every male was introduced into a cup together with one *A. arabiensis* female (*introgressed cups*). Parallel the same number of cups were set up for wild-type *A. arabiensis* males and females as a control (*control cups*). After 4 days mating and blood feeding of females, eggs were collected from every cup and the hatching rate relative to every single backcross was calculated. Progeny from the *introgressed cups* with higher hatching rate was kept, the males were selected and used to set up further backcrosses to *A. arabiensis* females in cups. The strategy of selection of higher hatching rate and backcross to *A. arabiensis* females in cups was applied to a total of 7 generations. Parallel *control cups* were set up. Despite single crosses in cups represent a good strategy for tracking the parental of progeny, mosquitoes do not mate efficiently in the experimental circumstances described (mating in mosquitoes require formation of swarms). Therefore, only from a few number of experimental cups (*introgressed* and *controls cups*) progeny was recovered (Figure 39). The hatching rate data relative to the single backcrosses in cups are shown in the appendix, a summary of these data is given in Figure 40.

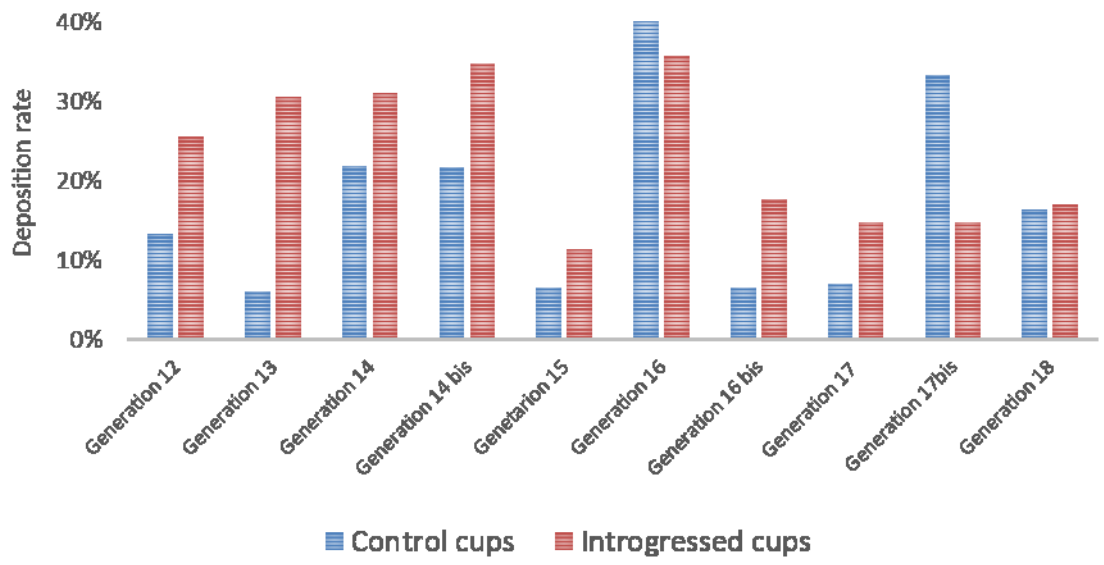


Figure 39: Deposition rate in single backcrosses in cups.

To assess fertility of *YI-PpoI*-introgressed males, single backcrosses to *A. arabiensis* females were set up in cups. The graph shows the percentage of cups from which eggs were recovered after mating and blood feeding of females. Red for *introgressed cups* (1 *YI-PpoI*-introgressed male x 1 *A. arabiensis* female) and blue for *control cups* (1 *A. arabiensis* male x 1 *A. arabiensis* female).

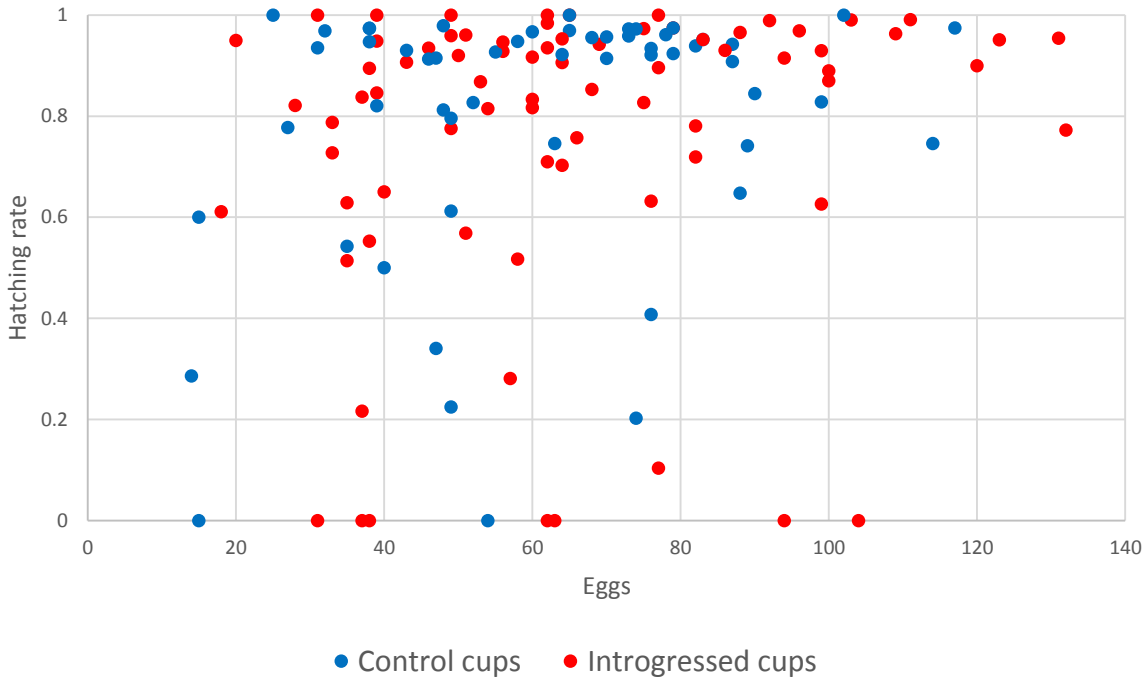


Figure 40: Hatching rate of single backcrosses in cups.

The hatching rate relative to 7 generations of single backcrosses in cups is showed. Red for *introgressed cups* (1 *YI-PpoI*-introgressed male x 1 *A. arabiensis* female) and blue for *control cups* (1 *A. arabiensis* male x 1 *A. arabiensis* female). A detailed description of the data relative to any generation analysed can be found in the appendix.

4.2.2.3 Establishment of the YI-PpoI-introgressed line

A number of 50 *YI-PpoI*-introgressed males, selected for higher hatching rate through the single backcrosses in cups, were crossed in cage with *A. arabiensis* females. After mating and blood feeding of females eggs were collected from the cage. The hatched progeny was grown and the emerging mosquitoes, males and females, were allowed to mate. In order to establish a mosquito line and ensure its maintenance, this process has been repeated for many generations (up to 6 so far). The exhibition of *3xP3*-RFP phenotype and PCR analysis of genomic DNA from *YI-PpoI*-introgressed males confirm the presence of the transgenic *A. gambiae* Y chromosome in the YI-PpoI introgressed line (Figure 41).

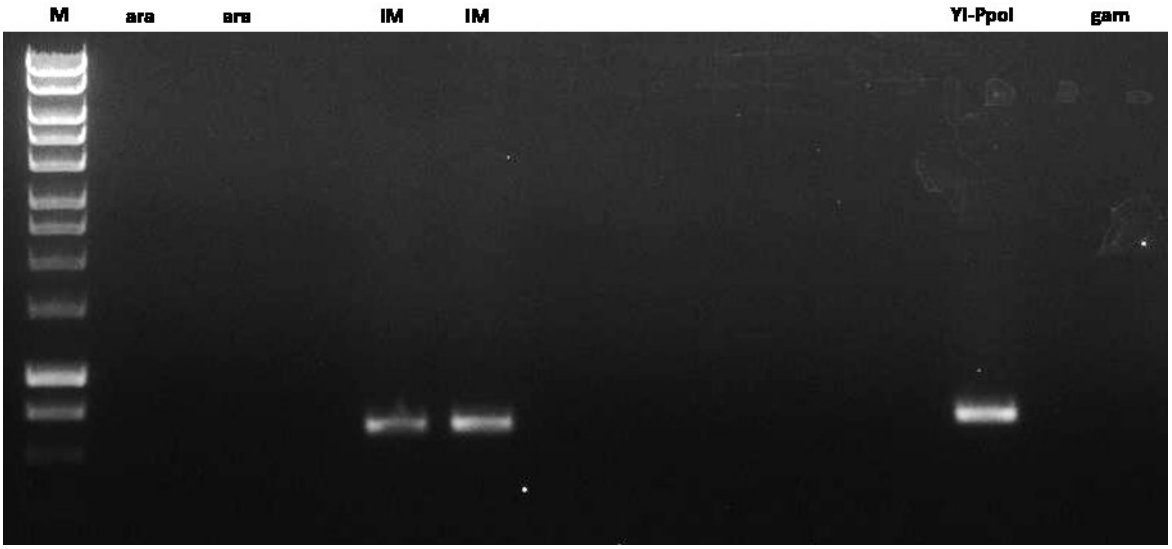


Figure 41: PCR analysis of *YI-PpoI*-introgressed line.

Lane 1: hyperladder I (M). Lane 2 and 3: genomic DNA from wild-type *A. arabiensis* males (ara). Lane 5 and 6: genomic DNA from *YI-PpoI*-introgressed males (IM). Lane 12: genomic DNA from the *A. gambiae* transgenic strain YI-PpoI (YI-PpoI) . Lane 13: genomic DNA from wild-type *A. gambiae* males (gam).

4.2.2.4 Introgression of the YAttP chromosome

The strategy applied for the introgression on the *A. gambiae* YI-PpoI chromosome was used to introgress the *A. gambiae* YAttP chromosome in the *A. arabiensis* background. No detailed eggs hatching and laying data were recorded through the experiment, however, the bottlenecks recorded at the early stage of the introgression, similarly to those observed for the YI-PpoI chromosome, were overcome and the introgression has been carried out for many generations in cage (up to 20 so far).

4.2.2.5 Y-introgressed males backcrosses to A. gambiae females

In Y-introgressed males the initial presence of *A. gambiae* DNA in the genetic background (*class 1* males) has been progressively substituted by *A. arabiensis* genome. This was possible through the setup of many generations of backcrossing of males to *A. arabiensis* females. Exhibition of *3xP3*-RFP phenotype in Y-introgressed males confirms the *A. gambiae* origin of the Y chromosome or, at least, the presence of the *A. gambiae* region of the Y chromosome harbouring the transgenic construct. Fertility of Y-introgressed males when crossed to *A. arabiensis* females was proven to be comparable to fertility of wild-type *A. arabiensis* males (Figure 40). The fertility of these males when crossed to *A. gambiae* females was also tested. A number of 50 introgressed males, YAttP and YI-PpoI, were randomly chosen and crossed to 50 *A. gambiae* females in two separate cages. After mating and blood feeding of females a large number of progeny was recovered from each cage (Figure 42). Males, named *class 5* males, were selected and crossed to *A. gambiae* females. None of the eggs collected after mating and blood feeding of females hatched in larvae suggesting sterility associated to *class 5* males (Figure 43).

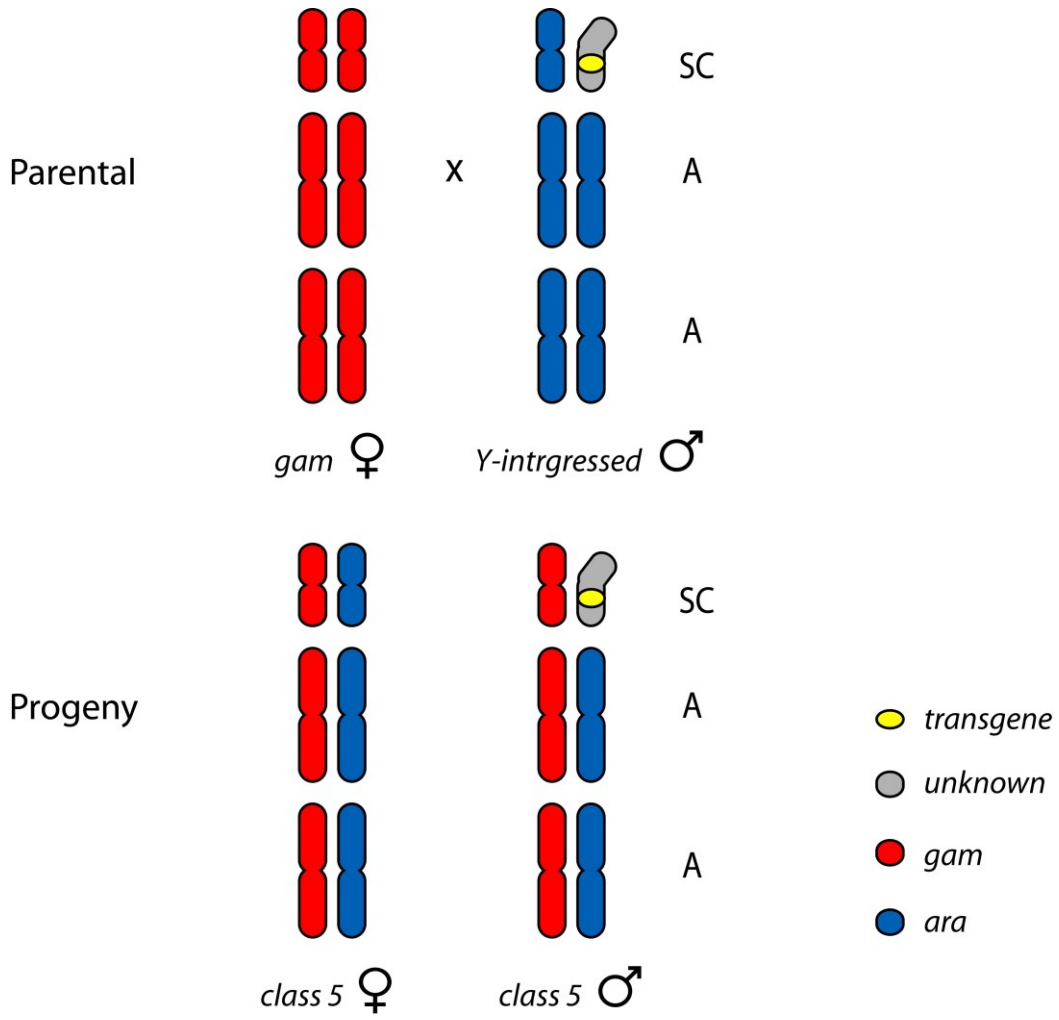


Figure 42: Y-introgressed males crossed to *A. gambiae* females.

Y-introgressed males (YAttP and YI-PpoI) where crossed to *A. gambiae* females. A large number of progeny, identified as *class 5* females and males, were recovered. SC: sexual chromosomes (XX in females, XY in males). A: autosomes. gam: *A. gambiae* genome. ara: *A. arabiensis* genome. transgene: construct inserted onto the Y chromosome. No sex ratio distortion was observed in the progeny.

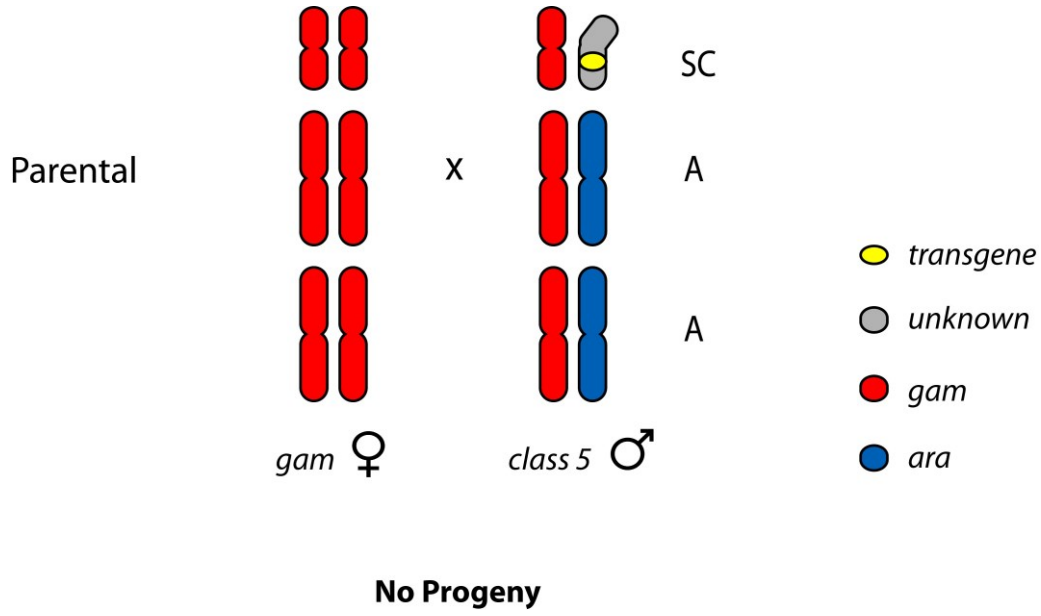


Figure 43: *Class 5* males crossed to *A gambiae* females.

Eggs recovered from the cages were *class 5* males, carrying either the *A. gambiae YAttP* or *YI-PpoI*, were crossed to *A. gambiae* female, did not hatched in larvae. SC: sexual chromosomes (XX in females, XY in males). A: autosomes. gam: *A. gambiae* genome. ara: *A. arabiensis* genome. transgene: construct inserted onto the Y chromosome. No sex ratio distortion was observed in the progeny.

5 Discussion

5.1 Engineering of the *A. gambiae* Y chromosome

Techniques such as transposase-mediated transformation have become routine in the study of the human malaria vector *Anopheles gambiae*. Although hundreds of transgenic *Anopheles* strains have been generated, none have been described harbouring a construct inserted onto the Y chromosome, despite it representing an estimated 10% of the genome. Y chromosomes are gene-poor, repeat-rich and largely heterochromatic. In higher eukaryotes, gene expression and the accessibility of DNA is determined, to a large extent, by chromatin structure. These facts suggested that, because of its heterochromatic nature, the Y chromosome was refractory to the random integration catalysed by transposases or, that marker genes, commonly used to identify transformation events, would undergo complete silencing when located on the Y chromosome. The study carried out during this PhD project was developed following the identification of a Y-linked transgenic line obtained in our laboratory by transposase-mediated random integration. This strain, named T4, is the first reported in *A. gambiae* with a transgene inserted onto the Y chromosome. Fluorescent maker genes (GFP and RFP) are expressed in two different tissues of T4 males suggesting that the transgene construct had inserted within a region of the chromosome permissive to transcriptional activity. As the Y chromosome is largely heterochromatic, such a euchromatic region represented a good candidate for the search of putative male-specific genes. In a number of organisms Y chromosome genes have been found to be essential for male fertility or sex determination (Miller, 2004). Several Diptera species, including *A. gambiae*, are believed to possess a Y-linked primary sex determination signal (Mason, 1967; Krafsur, 1972, Curtis, 1976; Baker and Sakai, 1979; Rabbani MG, 1972; Rabbani and Kitzmiller, 1975). Although subordinate genes like

transformer and *doublesex* have been identified and studied in a number species, the nature of this primary signal remains elusive. The identification of such a factor is expected to have direct applications to mosquito vector control. It could be the basis for novel genetic sexing technologies as is required for sterile insect release or, in combination with gene manipulation technologies and genetic driving mechanisms, to distort a population's sex-ratio towards males (Deredec *et al.*, 2008). In order to explore the surrounding region of the T4 transgene, a bacterial artificial chromosome (BAC) harbouring Y sequences was isolated from a library generated from the T4 strain. The sequencing of this BAC provided around 110 kb of sequences belonging to the Y, therefore enhancing the knowledge of this mostly unexplored chromosome. A number of putative ORFs were identified from the Y sequences. The majority of these putative ORFs, analysed by genomic PCR, were also found to be present in females, indicating that these sequences are likely to be additionally present on chromosomes other than the Y. In agreement with the notion that the Y is a repeat-rich chromosome, many of these ORFs matched known repetitive elements and transposons. A putative ORF, named ORFe, appeared to be present in both, X and Y, sexual chromosomes with an extra string of 24 nucleotides present exclusively on the X chromosome. ORFe showed expression in early embryos and only in male adult mosquitoes. Expression of this ORF was recovered also in adult testes. This finding makes ORFe an interesting target for studies that aim to identify regulatory regions that drive gene expression during spermatogenesis as this is an important requirement for the application of vector control strategies such as those based on homing endonuclease genes. A number of Y-specific PCR markers have been identified that may aid in characterizing anopheline population history and geographic structure, which is critical for understanding genetic structure and speciation and for the effective implementation of malaria control strategies (Hammer and Zegura, 1996; Hurles and Jobling, 2001). The Y-linked

transgene in the T4 line has been proven to be stable, making this strain ideally suited for automated fluorescence sexing. Although the use of COPAS automated fluorescent larval sorting is limited to research purposes, it has been suggested that future iterations of this technology could be scaled up to develop robust protocols for the mass production of sorted insects. This property may become useful in applications requiring large numbers of pure males, such as sterile insect technique (Condon *et al.*, 2007), and in experiments that require genetic crosses to a large number of virgin wild-type females, readily obtained here from the T4 line purified at the early larval stage. The fact that transgenic males give rise to male progeny that all inherit the transgene represents an additional advantage of these strains because, unlike with autosomal markers for sexing (Marois *et al.*, 2012), no segregation of the sexing marker occurs so that a pure-breeding stock can be maintained. This feature allows a sexing trait to be combined with other autosomally encoded traits, for example male sterility. The fact that the T4 locus was transcriptionally active made it an exciting target for genome engineering. A site-specific recombination signal (*AttP*) was introduced onto the Y chromosome using meganuclease-induced homologous repair. This result demonstrated that sequence-specific chromosomal breaks can be repaired by homologous recombination with a circular synthetic repair template, and therefore, that *A. gambiae* is amenable to knock-in genome engineering procedures. The availability of novel programmable endonucleases based on the transcription activator-like effector nuclease (TALEN) or clustered regularly interspaced short palindromic repeats (CRISPR/Cas) architectures as well as a growing number of sequences uniquely present on the Y chromosome such as those described in this study, suggest that this powerful approach could be extended to introduce foreign sequences (such as docking sites or desired mutations) also into other regions of the Y or, indeed, into any locus of interest on other chromosomes. Although homologous recombination was inefficient

(occurring in 0.18% of the progeny), it opened up the possibility for the more efficient site-specific integration method. Secondary integration mediated by the ϕC31 integrase was achieved and occurred at a rate at least 4× higher. The generation of the transgenic strain YVasG, demonstrates the ability to specifically insert transgenes into the Y chromosome, this allows for the generation of male-exclusive traits by using non-sex-specific components and to express tagged or mutant forms of Y-linked genes. Alternatively to automated sexing based on fluorescent markers, genetic sexing traits such as insecticide resistance markers can now be easily introduced to the Y chromosome providing the means for large releases of only male mosquitoes. Moreover, testes of YVasG males show expression of GFP, driven by the *vasa* promoter, in all stages of spermatogenesis including the germline stem cells (GSCs) in the apical tip and developing spermatocytes but not in supporting somatic cells. This finding suggests that the strain YAttP, in combination with the previously characterized non-sex-specific *vasa* promoter, allows the male-exclusive expression of transgenes starting at the earliest stages of spermatogenesis thus, encouraging the development of vector control strategies such those based on homing endonuclease genes. It has been suggested that population suppression or elimination could be achieved by biasing the sex ratio towards males (Hamilton, 1967). As part of a different experimental project I contributed to during my PhD, a synthetic male sex ratio distortion system, based on the activity of the endonuclease I-PpoI, was generated in our laboratory. This system operates by destroying the X chromosome during male spermatogenesis and creates progeny consisting almost exclusively of males. Population cage studies showed that transgenic males carrying the sex distorter led to the suppression of a wild-type *A. gambiae* population within six generations. These transgenic strains harboured a sex distorter, driven by the male spermatogenesis-specific *b2 tubulin* promoter, on an autosomal chromosome (Galizi *et al.*, 2014). When transgenes are

expressed from autosomal locations they are transmitted to only half of the male progeny. In its current form, this technology would require continuous mass releases to achieve population suppression. In contrast, if the distorter trait could be successfully linked to the Y chromosome, this chromosome would be converted into a selfish sex chromosome that, in theory, could spread through and eliminate a natural vector population even when seeded by a very low number of transgenic males (Hamilton, 1967). The development of such technique is made challenging by the observation that sex chromosomes are inactivated during meiosis. This phenomenon is known as meiotic sex chromosomes inactivation (MSCI) and determines the silencing of chromosomes that fail to pair with their homologues partners during meiosis thus, protecting against aneuploidy in subsequent generations (Vibranovski, 2014). Despite active integration onto the Y chromosome are really rare, during a different experimental project (unpublished data), a new transgenic Y line was generated in our laboratory by transposase-mediate random integration. In this strain, named YI-PpoI, the fluorescent marker, driven by the neuronal promoter *3xP3*, is clearly expressed while the *I-PpoI*, previously described, driven by the *b2 tubulin* promoter, is silenced, therefore no sex distortion is observed in the progeny. Contrary to the *vasa* promoter, the regulatory regions of the *b2 tubulin* gene drive gene expression during late stages of spermatogenesis in *A. stephensi* and *A. gambiae* (Windbichler et al., 2008; Catteruccia et al., 2005). No expression of *b2 tubulin*-GFP in YI-PpoI males suggests that the *b2 tubulin* promoter activation is hampered by MSCI. Detection of *vasa*-GFP in YVasG males is likely due the earlier activation of this promoter and the presence of GFP in later stages of spermatogenesis might be due to persistence of stable GFP protein. The potential for the application of vector control strategy such us the synthetic sex ratio distortion system here described, encourages studies that aim to characterize biological mechanisms such us spermatogenesis and MSCI. The identification of genes

expressed during different stages of spermatogenesis could lead to the isolation of regulatory sequences that escape silencing due to MSCI. Alternatively, the increasing availability of Y-specific sequences, together with the availability of programmable endonucleases such as TALEN or CRISPR/Cas could be used to induce insertion of docking sites in regions of the Y chromosome that might not be subjected to gene silencing during meiosis. The *A. gambiae* YAttP line described in this work is expected to improve the efficiency of existing vector control strategies such as those requiring mass-releases of sterile males. Furthermore, this strain paves the way for the development and the application of new vector control technologies such as those based on endonuclease genes. Therefore, the transgenic line YAttP represents a promising tool in the fight against malaria.

5.2 Introgression of a modified *A. gambiae* chromosome in the *A. arabiensis* genetic background

Within the *Anopheles* complex there are eight sibling species that transmit human malaria with *A. gambiae* and *A. arabiensis* being the most anthropophilic and wide spread members. The efficiency of vector control strategies that aim to eradicate malaria rely on the possibility of their application on sibling vector species. The value of the *A. gambiae* YAttP strain for the improvement of existing vector control strategies and its potential for the development of new approaches, has been previously discussed. Therefore, the generation of a transgenic *A. arabiensis* strain with an engineered Y chromosome as that described in *A. gambiae* would be of great interest (Bernardini *et al.*, 2014). However, due to the heterochromatic nature of the Y chromosome this approach is challenging. An alternative strategy would have been to introgress the engineered *A. gambiae* Y chromosome into the *A. arabiensis* genetic background. The success of this approach is made difficult by the observation that pre- and post-zygotic isolation mechanisms occur between these species (Gibson *et al.*, 2010; Haldane, 1922; Slotman *et al.*, 2004). Pre-zygotic isolation mechanisms can be easily overcome by using experimental conditions that favour interspecific mating i.e. set up of interspecific crosses in cage. Post-zygotic isolation mechanisms represent the critical step for the progress of the introgression. In fact, hybrids between these two species obey Haldane's rule: "When in the offspring of two different animal races one sex is absent, rare, or sterile, that sex is the heterozygous (heterogametic) sex". The rule holds true also for other groups such as flies and mammals, where males represent the heterogametic sex, and also birds and butterfly, where the heterogametic sex is represented by females (Schilthuizen *et al.*, 2011). Based on these observations, heterogametic sex chromosomes seem to be the critical feature. Despite many models have been proposed to explain the genetic basis of Haldane's rule, its origin it is still

unclear. A scheme based on crossing and selection was used to overcome F1 hybrid males sterility and introgress the *A. gambiae* YAttP chromosome in the *A. arabiensis* genetic background. Moreover, the *A. gambiae* YI-PpoI strain has been previously described in this chapter. In this strain the fluorescent marker, driven by the neuronal promoter *3xP3*, is clearly expressed while the sex distorter *I-PpoI*, driven by the spermatogenesis-specific promoter *b2 tubulin*, is silenced due to MSCI (unpublished data) (Vibranovski, 2014). If the introgression of the *A. gambiae* Y chromosome in a sibling genetic background generated an alteration of the gene expression profiling in hybrid individuals, this could lead, in theory, to the activation of genes otherwise silenced such as the *I-PpoI* in the YI-PpoI strain. In order to assess any reactivation of the sex distorter, parallel to the introgression of the *A. gambiae* YAttP chromosome, the introgression of the *A. gambiae* YI-PpoI chromosome, in the *Anopheles arabiensis* genetic background, was attempted. The introgression was carried out through crosses of F1 hybrid fertile females to Y-transgenic *A. gambiae* males, selection of male progeny and backcrossing with *A. arabiensis* females. This strategy, applied generation after generation, aimed to diminish the amount of *A. gambiae* genome in the hybrid individuals with the exception of the Y chromosome whose presence is indicated by the expression of the fluorescent marker. As previously discussed for the T4 line no putative recombination has been observed between the *A. gambiae* X and Y chromosomes. This is in accord with previous works that showed no evidence of recombination between heteromorphic X and Y chromosomes (Bachtrog, 2003; Rice, 1994). During the introgression experiment here described all progeny was screened for the presence of the red fluorescent marker at any generation. Larvae expressing the fluorescent marker always developed in males, while fluorescent negative larvae always developed in females. These observations encouraged the assumption that the Y chromosome progressively selected though the introgression is an

unaltered or mostly unaltered *A. gambiae* transgenic Y chromosome. However, pseudoautosomal regions (PARs) described as homologous sequences localized at the tips of the X and Y chromosomes, have been shown to recombine in a number of groups (Helena Mangs and Morris, 2007). Therefore, recombination between the sexual chromosomes, X and Y, cannot be entirely excluded with the exception of the Y chromosome region where the transgene is inserted (Y-introgressed males express the fluorescent marker). The experimental approach used in this work underlined differences in fertility of different classes of males:

- Wild-type *A. gambiae* and *A. arabiensis* and the transgenic *A. gambiae* males (YAttP and YI-PpoI) exhibited full fertility (Figure 44).
- F1 hybrid males, characterized by a haploid set of chromosomes from each species, showed full sterility as predicted by Haldane's rule (Figure 44) (Haldane, 1922).

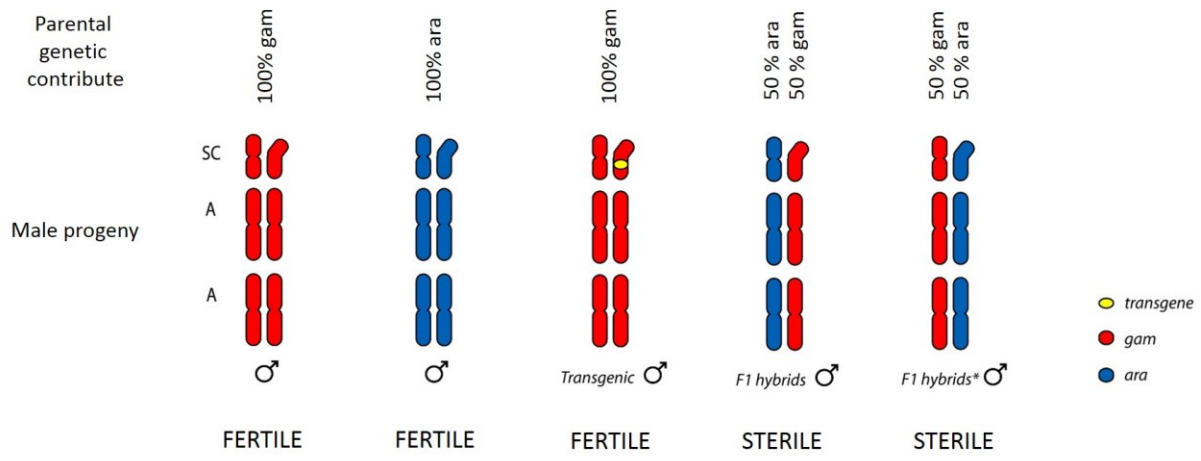


Figure 44: Association between different classes of males and fertility/sterility phenotype.

SC: sexual chromosomes (XX in females, XY in males). A: autosomes. gam: *A. gambiae* genome. ara: *A. arabiensis* genome. transgene: construct inserted onto the Y chromosome. No sex ratio distortion was observed in the progeny.

- In class 1 males the paternal genetic contribution has an *A. gambiae* composition. The maternal genetic contribution is the result of hybrid meiosis. Since the recombination rate of interspecific chromosomes during meiosis cannot be predicted, the maternal genetic contribution in class 1 males has been indicated, through the description of this project, as unknown (Figure 45). However, when class 1 males were backcrossed to *A. arabiensis* females a large number of progeny was recovered suggesting that fertility of these males was not strongly affected. Despite the maternal genetic contribution in class 1 males cannot be predicted, exhibition of fertility, exclude an entire *A. arabiensis* genetic contribution, this scenario would in fact be shared with F1 hybrid males that are fully sterile. Therefore, in class 1 males the genetic background is expected to exceed in the *A. gambiae* genetic composition. It has been suggested that hybrid phenomena such as inviability and sterility are due to incompatibility between interspecific genes (Slotman *et al.*, 2004). In class 1 males these incompatibilities, fully manifested in F1 hybrids might have been overcome by loss of *A. arabiensis* "*A. gambiae*-incompatible" regions of DNA.
- In *class 2* males the maternal genetic contribution has an *A. arabiensis* composition while the paternal contribution is the result of hybrid meiosis and its exact composition cannot be predicted (Figure 45). Interestingly, contrary to class 1 males, class 2 males, exhibited high sterility, only 56 out of 29776 eggs (0.002%), recovered from the crosses of these males, hatched in larvae. The genetic composition of class 2 males is unique due to the presence of a full *A. arabiensis* X chromosome with the putative unaltered transgenic *A. gambiae* Y. This set of sexual chromosomes resembles that present in F1 hybrids which, similarly to class 2 males, express full sterility. This observation, suggested prematurely, a correlation between the presence

of interspecific sexual chromosomes in hybrid individuals and sterile phenotype. This was in accord with the X-Y model for the explanation of Haldane's rule. This model suggests that the interaction between the X chromosome from one species and the Y chromosome from the other species, which only occur in the heterogametic sex, it is responsible for the inviability or sterility observed (Coyne, 1985). However, this X-Y model is not well supported in those species where the males lack the Y chromosome i.e. species in which females are XX and males are XO, the model would in fact predict male and female hybrids be equally affected while in grasshoppers, for example, male hybrids are more frequently inviable or sterile than female hybrids (Wu and Davis, 1993). High sterility of class 2 males provided further support for the role of interspecific gene incompatibility in causing sterility. Slotman *et al.* have previously suggested that in hybrids between *A. gambiae* and *A. arabiensis* sterility effects are due to incompatibility between interspecific X and autosomes (Slotman *et al.*, 2004). In the simple scenario where sterility is due to incompatibility of the X chromosome with a locus on each interspecific autosome, only 6.25 % of class 2 males are expected to be fertile i.e. lacking both the *A. gambiae* autosomal factors incompatible with the full *A. arabiensis* X chromosome. The bottleneck recovered during the introgression experiment, following the crosses of class 2 males (0.002% hatching rate), supports the findings of Slotman *et al.* and suggests a more complex scenario than the one here hypnotized for the explanation of hybrid sterility i.e. more incompatibility factors might be involved.

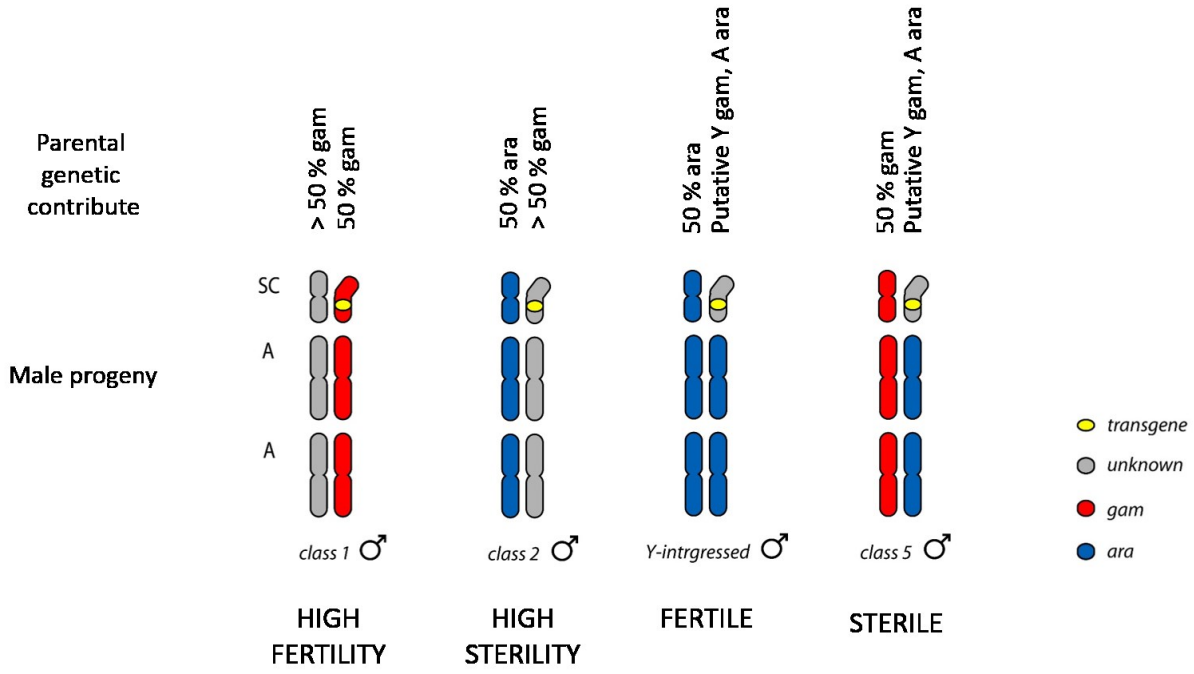


Figure 45: Association between different classes of males and fertility/sterility phenotype.

SC: sexual chromosomes (XX in females, XY in males). A: autosomes. gam: *A. gambiae* genome. ara: *A. arabiensis* genome. transgene: construct inserted onto the Y chromosome. No sex ratio distortion was observed in the progeny.

Male larvae that got through the bottleneck observed were selected and the introgression continued. The hatching rate recovered from cages through 10 generations of backcrossing showed values between 0.018 and 0.83%. Degrees of fertility in different generation hybrids has been reported (Oka *et al.*, 2004). This phenomenon, known as hybrid breakdown, is due the variety in combination of interspecific genes into the hybrid genetic backgrounds. Nevertheless, further backcrossing to *A. arabiensis* females succeeded in the generation of Y-introgressed males whose fertility appeared to be comparable to *A. arabiensis* wild-type males. If fertility was hampered in earlier generation hybrids by incompatibility between interspecific genes, it is reasonable to assume that, after many generations of backcrossing, the dilution of the *A. gambiae* genome in the hybrids, due the progressive contribution of *A.*

arabiensis females, determined a degrease and eventually the overcoming of those incompatibilities.

- In Y-introgressed males the Y chromosome, based on the assumptions previously made, is expected to be an unaltered/almost unaltered transgenic *A. gambiae* Y, the X chromosome has an *A. arabiensis* origin and the autosomes, after up to 20 generations of backcrossing to *A. arabiensis* females, are expected to have this species composition (Figure 45). The successful introgression in the *A. arabiensis* genetic background of the transgenic *A. gambiae* Y chromosome, and the observation that it might be an unaltered *A. gambiae* Y, suggests that the X-Y model for the explanation of Haldane's rule, previously considered, is likely not to be applied on these mosquitoes species i.e. interspecific sexual chromosomes are not responsible for sterility observed. Moreover, crosses of the Y-introgressed males to *A. gambiae* females generate class 5 males that exhibit sterility.

- In class 5 males the maternal genetic contribution has an *A. gambiae* composition. Based on the assumptions made, the paternal genetic contribution is expected to have an *A. gambiae* composition for the Y chromosome and an *A. arabiensis* composition for the autosomes (Figure 45). The genetic background of class 5 resembles that in F1 hybrids with the exception of the sexual chromosomes, co-specific in class 5 males and interspecific in F1 hybrids. Interestingly both these classes of males are sterile. This result highlights a minor role of interspecific sexual chromosomes in determining hybrids sterility. Conversely, the hypothesis that sterility is due to incompatibilities between interspecific X and autosomes is supported together with a minor role of the Y chromosome.

In order to validate the speculations here discussed, the genetic background of the Y-introgressed males needs to be explored. The assumption that in these males the Y chromosome has an *A. gambiae* composition has to be assessed and the *A. arabiensis* composition of the autosomes confirmed. Deep (50x) DNA sequencing of a number of Y-introgressed males has been arranged. The data recovered from the DNA sequencing are currently under analysis but preliminary results suggest that the hypothesis that the Y chromosome in Y-introgressed males is an unaltered transgenic *A. gambiae* Y seems to be supported. The outcome of these data promise to be helpful for the understating of the complex genetic associated with hybrids. The peculiar genetic background of Y-introgressed males and their fertile phenotype can shade light on the genetic basis of Haldane's rule in anopheline mosquitoes, made previously difficult by the lack of a deep genetic analysis of hybrid individuals due to either inviability or sterility. Moreover, the parallel RNA sequencing of Y-introgressed and wild-type A. *arabiensis* males, that is next step for the progress of this project, is expected to answer important questions on the similarity and differences in Y chromosome biology of closely related species. The presence of really low frequency of hybrids between *A. gambiae* and *A. arabiensis* in areas were these species are sympatric suggests the presence of pre-zygotic isolation mechanisms (Toure *et al.*, 1998; Temu *et al.*, 1997; Mawejje *et al.*, 2013). Cues used for species recognition are still unclear but they might involve differences in the wing beat frequencies in the two species (Gibson *et al.*, 2010). Furthermore, it has been suggested that the Y chromosome controls mating behaviour in *Anopheles* mosquitoes (Fraccaro *et al.*, 1977). A role of the Y chromosome in determining pre-zygotic isolation mechanisms might emerge from behavioural studies of the Y-introgression males and their comparison with wild-type *A. arabiensis* and *A. gambiae* males. *A. gambiae* and *A. arabiensis* represent the major vectors of human malaria. The work

described in this thesis has led to the generation of an *A. gambiae* transgenic line, YattP, which is expected to be a promising tool in the fight against malaria. The generation of the *YAttP*-introgressed males provide a candidate strain for the application of the vector control strategies discussed for *A. gambiae* also on the sibling species *A. arabiensis*.

6 Conclusions

Due to the widest geographic distribution and marked anthropophily, *A. gambiae* and *A. arabiensis* are considered the major vectors of human malaria. The study described in this thesis has led to the generation of an *A. gambiae* transgenic line, YattP, which is expected to be a promising tool for the improvement of existing vector control techniques and for the implementation of new technologies against the disease. Furthermore, the *YAttP*-introgressed strain, generated by genetic introgression, is a good candidate for the application of such strategies also on the sibling species *A. arabiensis*.

7 Future plans

The YattP and YattP-introgressed strains described within this thesis are valuable tools for both the investigation and development of strategies for vector control. However, the potential use of these biological strains for the suppression of target populations, depends critically upon parameters including, but not limited to, fitness, competitiveness and fertility. These parameters must be studied further in the context of the strains, with a focus upon their use for the generation of a transgenic sex-distorter strain. Moreover, the data generated by the introgression experiment has provided the basis for experiments aimed at better understanding the biology of the Y chromosome, and the mechanisms of post-zygotic isolation between closely related species.

8 References

ALLEN, M. L., O'BROCHTA, D. A., ATKINSON, P. W. & LEVESQUE, C. S. 2001. Stable, germ-line transformation of Culex quinquefasciatus (Diptera: Culicidae). *J Med Entomol,* 38**,** 701-10.

BACHTROG, D. 2003. Adaptation shapes patterns of genome evolution on sexual and asexual chromosomes in Drosophila. *Nat Genet,* 34**,** 215-9.

BAKER, B. S. & BELOTE, J. M. 1983. Sex Determination and Dosage Compensation in Drosophila Melanogaster. *Annual Review of Genetics,* 17**,** 345-393.

BAKER, R. H. & SAKAI, R. K. 1979. Triploids and male determination in the mosquito, Anopheles culicifacies. *J Hered,* 70**,** 345-6.

BENEDICT, M. Q. & ROBINSON, A. S. 2003. The first releases of transgenic mosquitoes: an argument for the sterile insect technique. *Trends in Parasitology,* 19**,** 349-355.

BERGHAMMER, A. J., KLINGLER, M. & WIMMER, E. A. 1999. A universal marker for transgenic insects. *Nature,* 402**,** 370-1.

BERNARDINI, F., GALIZI, R., MENICHELLI, M., PAPATHANOS, P.-A., DRITSOU, V., MAROIS, E., CRISANTI, A. & WINDBICHLER, N. 2014. Site-specific genetic engineering of the Anopheles gambiae Y chromosome. *Proceedings of the National Academy of Sciences,* 111**,** 7600-7605.

BOËTE, C. & KOELLA, J. C. 2003. Evolutionary ideas about genetically manipulated mosquitoes and malaria control. *Trends in Parasitology,* 19**,** 32-38.

BRAIG, H. R. Y., G. 2002. Genetically Engineered Organisms: Assessing Environmental and Human Health Effects. *CRC Press, Boca Raton, Florida*.

BREMAN, J. G., ALILIO, M. S. & MILLS, A. 2004. Conquering the intolerable burden of malaria: what's new, what's needed: a summary. *Am J Trop Med Hyg,* 71**,** 1-15.

BRUNHES J, L. G., GEOFFROY B. 1997. frotropical Anopheles.1. Description of new species and modifications of taxonomical status (Diptera:Culicidae). *Ann Soc Entomol Fr.*

BURGOYNE, P. S. 1998. The mammalian Y chromosome: a new perspective. *Bioessays,* 20**,** 363-6.

BURT, A. 2003. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proceedings of the Royal Society B: Biological Sciences,* 270**,** 921-928.

BUTLER, D. 2009. Initiative targets malaria eradication. *Nature,* 462**,** 19.

CALOS, M. P. & MILLER, J. H. 1980. Transposable elements. *Cell,* 20**,** 579-595.

CARVALHO, A. B., DOBO, B. A., VIBRANOVSKI, M. D. & CLARK, A. G. 2001. Identification of five new genes on the Y chromosome of Drosophila melanogaster. *Proc Natl Acad Sci U S A,* 98**,** 13225-30.

CARVALHO, A. B., LAZZARO, B. P. & CLARK, A. G. 2000. Y chromosomal fertility factors kl-2 and kl-3 of Drosophila melanogaster encode dynein heavy chain polypeptides. *Proc Natl Acad Sci U S A,* 97**,** 13239-44.

CASARES, S., BRUMEANU, T. D. & RICHIE, T. L. 2010. The RTS,S malaria vaccine. *Vaccine,* 28**,** 4880-94.

CASTILLO, E. R., BIDAU, C. J. & MARTI, D. A. 2010. Neo-sex chromosome diversity in Neotropical melanopline grasshoppers (Melanoplinae, Acrididae). *Genetica,* 138**,** 775-86.

CATHERINE A. HILL, J. F. M. A. S. C. 2008. Purdue Entomology.

CATTERUCCIA, F., BENTON, J. P. & CRISANTI, A. 2005. An Anopheles transgenic sexing strain for vector control. *Nat Biotechnol,* 23**,** 1414-7.

CATTERUCCIA, F., NOLAN, T., LOUKERIS, T. G., BLASS, C., SAVAKIS, C., KAFATOS, F. C. & CRISANTI, A. 2000. Stable germline transformation of the malaria mosquito Anopheles stephensi. *Nature,* 405**,** 959-62.

CHARLESWORTH, B. 1996. The evolution of chromosomal sex determination and dosage compensation. *Current Biology,* 6**,** 149-162.

CHEVALIER, B. S. & STODDARD, B. L. 2001. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Research,* 29**,** 3757-3774.

CHRISTOPHIDES, G. K., VLACHOU, D. & KAFATOS, F. C. 2004. Comparative and functional genomics of the innate immune system in the malaria vector Anopheles gambiae. *Immunol Rev,* 198**,** 127-48.

CLEMENTS, A. N. 1992. *The Biology of Mosquitoes: Development, nutrition, and reproduction*, Chapman & Hall.

COATES, C. J., JASINSKIENE, N., MIYASHIRO, L. & JAMES, A. A. 1998. Mariner transposition and transformation of the yellow fever mosquito, Aedes aegypti. *Proceedings of the National Academy of Sciences,* 95**,** 3748-3751.

COETZEE, M., CRAIG, M. & LE SUEUR, D. 2000. Distribution of African Malaria Mosquitoes Belonging to the Anopheles gambiae Complex. *Parasitology Today,* 16**,** 74-77.

COETZEE, M., HUNT, R. H., WILKERSON, R., DELLA TORRE, A., COULIBALY, M. B. & BESANSKY, N. J. 2013. Anopheles coluzzii and Anopheles amharicus , new members of the Anopheles gambiae complex. *2013,* 3619**,** 29.

COLUZZI, M., SABATINI, A., DELLA TORRE, A., DI DECO, M. A. & PETRARCA, V. 2002. A polytene chromosome analysis of the Anopheles gambiae species complex. *Science,* 298**,** 1415-8.

COLUZZI, M., SABATINI, A., PETRARCA, V. & DI DECO, M. A. 1979. Chromosomal differentiation and adaptation to human environments in the Anopheles gambiae complex. *Trans R Soc Trop Med Hyg,* 73**,** 483-97.

CONDON, K. C., CONDON, G. C., DAFA'ALLA, T. H., FU, G., PHILLIPS, C. E., JIN, L., GONG, P. & ALPHEY, L. 2007. Genetic sexing through the use of Y-linked transgenes. *Insect Biochem Mol Biol,* 37**,** 1168-76.

COYNE, J. A. 1985. The genetic basis of Haldane's rule. *Nature,* 314**,** 736-8.

CRAWLEY, J., CHU, C., MTOVE, G. & NOSTEN, F. 2010. Malaria in children. *The Lancet,* 375**,** 1468-1481.

CRISCIONE, F., QI, Y., SAUNDERS, R., HALL, B. & TU, Z. 2013. A unique Y gene in the Asian malaria mosquito Anopheles stephensi encodes a small lysine-rich protein and is transcribed at the onset of embryonic development. *Insect Mol Biol,* 22**,** 433-41.

CROMPTON, P. D., PIERCE, S. K. & MILLER, L. H. 2010. Advances and challenges in malaria vaccine development. *J Clin Invest,* 120**,** 4168-78.

CURTIS, C. F., J. AKIYAMA AND G. DAVIDSON 1976. A genetic sexing system in Anopheles gambiae species A. *Mosquito News*.

CURTIS, C. F., JANA-KARA, B. & MAXWELL, C. A. 2003. Insecticide treated nets: impact on vector populations and relevance of initial intensity of transmission and pyrethroid resistance. *J Vector Borne Dis,* 40**,** 1-8.

CURTIS, W. K. A. C. F. 2005. Sterile Insect Technique *Springer Netherlands***,** chap. 1 pp. 3-36.

DAVIDSON, G. 1956. Insecticide Resistance in Anopheles gambiae Giles: a Case of Simple Mendelian Inheritance. *Nature,* 178**,** 863-864.

DAVIDSON, G. 1964a. Anopheles gambiae, a complex of species. *Bulletin of the World Health Organization,* 31**,** 625-634.

DAVIDSON, G. 1964b. THE FIVE MATING-TYPES IN THE ANOPHELES GAMBIAE COMPLEX. *Riv Malariol,* 43**,** 167-83.

DAVIDSON G, H. R. 1973. The crossing and chromosome characteristics of a new, sixth species in the Anopheles gambiae complex. . *Parassitologia*.

DEREDEC, A., BURT, A. & GODFRAY, H. C. 2008. The population genetics of using homing endonuclease genes in vector and pest management. *Genetics,* 179**,** 2013-26.

DEWEERDT, S. 2012. Vaccines: The take-home lesson. *Nature,* 484**,** S24-S25.

DRUILHE, P., SPERTINI, F., SOESOE, D., CORRADIN, G., MEJIA, P., SINGH, S., AUDRAN, R., BOUZIDI, A., OEUVRAY, C. & ROUSSILHON, C. 2005. A Malaria Vaccine That Elicits in Humans Antibodies Able to Kill <italic>Plasmodium</italic> <italic>falciparum</italic>. *PLoS Med,* 2**,** e344.

EL SAHLY, H. M., PATEL, S. M., ATMAR, R. L., LANFORD, T. A., DUBE, T., THOMPSON, D., SIM, B. K., LONG, C. & KEITEL, W. A. 2010. Safety and immunogenicity of a recombinant nonglycosylated

erythrocyte binding antigen 175 Region II malaria vaccine in healthy adults living in an area where malaria is not endemic. *Clin Vaccine Immunol,* 17**,** 1552-9.
FISHER, R. 1931. The evolution of dominance. *Biol Rev Amb Philos Soc*.
FLICK, K. E., JURICA, M. S., MONNAT, R. J., JR. & STODDARD, B. L. 1998. DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature,* 394**,** 96-101.
FRACCARO, M., TIEPOLO, L., LAUDANI, U., MARCHI, A. & JAYAKAR, S. D. 1977. Y chromosome controls mating behaviour on Anopheles mosquitoes. *Nature,* 265**,** 326-328.
FRANCISCO MARTIN, K. B., MARIEN COBO, PILAR MUNOZ, PER ANDERSON AND MIGUEL G. TOSCANO 2011. New Vectors for Stable Gene Modification. chapter 1.
GALIZI, R., DOYLE, L. A., MENICHELLI, M., BERNARDINI, F., DEREDEC, A., BURT, A., STODDARD, B. L., WINDBICHLER, N. & CRISANTI, A. 2014. A synthetic sex ratio distortion system for the control of the human malaria mosquito. *Nat Commun,* 5.
GENTON, B., AL-YAMAN, F., BETUELA, I., ANDERS, R. F., SAUL, A., BAEA, K., MELLOMBO, M., TARAIKA, J., BROWN, G. V., PYE, D., IRVING, D. O., FELGER, I., BECK, H. P., SMITH, T. A. & ALPERS, M. P. 2003. Safety and immunogenicity of a three-component blood-stage malaria vaccine (MSP1, MSP2, RESA) against Plasmodium falciparum in Papua New Guinean children. *Vaccine,* 22**,** 30-41.
GHOSH, A. K., DEVENPORT, M., JETHWANEY, D., KALUME, D. E., PANDEY, A., ANDERSON, V. E., SULTAN, A. A., KUMAR, N. & JACOBS-LORENA, M. 2009. Malaria parasite invasion of the mosquito salivary gland requires interaction between the Plasmodium TRAP and the Anopheles saglin proteins. *PLoS Pathog,* 5**,** e1000265.
GIBSON, G., WARREN, B. & RUSSELL, I. J. 2010. Humming in tune: sex and species recognition by mosquitoes on the wing. *J Assoc Res Otolaryngol,* 11**,** 527-40.
GILLIES M, D. M. B. 1968. The Anophelinae of Africa south of Sahara.
GOSSEN, M. & BUJARD, H. 1992. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A,* 89**,** 5547-51.
GROSSMAN, G. L., RAFFERTY, C. S., CLAYTON, J. R., STEVENS, T. K., MUKABAYIRE, O. & BENEDICT, M. Q. 2001. Germline transformation of the malaria vector, Anopheles gambiae, with the piggyBac transposable element. *Insect Molecular Biology,* 10**,** 597-604.
HAHN, M. W., WHITE, B. J., MUIR, C. D. & BESANSKY, N. J. 2012. *No evidence for biased co-transmission of speciation islands in Anopheles gambiae*.
HALDANE, J. B. S. 1922. Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics,* 12**,** 101-109.
HALL, A. B., QI, Y., TIMOSHEVSKIY, V., SHARAKHOVA, M. V., SHARAKHOV, I. V. & TU, Z. 2013. Six novel Y chromosome genes in Anopheles mosquitoes discovered by independently sequencing males and females. *BMC Genomics,* 14**,** 273.
HAMILTON, W. D. 1967. Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science,* 156**,** 477-88.
HAMMER, M. F. & ZEGURA, S. L. 1996. The role of the Y chromosome in human evolutionary studies. *Evolutionary Anthropology: Issues, News, and Reviews,* 5**,** 116-134.
HANDLER, A. M. 2001. A current perspective on insect gene transformation. *Insect Biochemistry and Molecular Biology,* 31**,** 111-128.
HARBACH, R. 2012. Mosquito Taxonomic Inventory.
HELENA MANGS, A. & MORRIS, B. J. 2007. The Human Pseudoautosomal Region (PAR): Origin, Function and Future. *Current Genomics,* 8**,** 129-136.
HERMSEN, C. C., VERHAGE, D. F., TELGT, D. S., TEELEN, K., BOUSEMA, J. T., ROESTENBERG, M., BOLAD, A., BERZINS, K., CORRADIN, G., LEROY, O., THEISEN, M. & SAUERWEIN, R. W. 2007. Glutamate-rich protein (GLURP) induces antibodies that inhibit in vitro growth of Plasmodium falciparum in a phase 1 malaria vaccine trial. *Vaccine,* 25**,** 2930-40.
HERRINGTON, D., DAVIS, J., NARDIN, E., BEIER, M., CORTESE, J., EDDY, H., LOSONSKY, G., HOLLINGDALE, M., SZTEIN, M., LEVINE, M. & ET AL. 1991. Successful immunization of humans with irradiated

malaria sporozoites: humoral and cellular responses of the protected individuals. *Am J Trop Med Hyg,* 45**,** 539-47.

HICKEY, W. A. & CRAIG, G. B. 1966. Genetic Distortion of Sex Ratio in a Mosquito, AEDES AEGYPTI. *Genetics,* 53**,** 1177-1196.

HOFFMAN, S. L., GOH, L. M., LUKE, T. C., SCHNEIDER, I., LE, T. P., DOOLAN, D. L., SACCI, J., DE LA VEGA, P., DOWLER, M., PAUL, C., GORDON, D. M., STOUTE, J. A., CHURCH, L. W., SEDEGAH, M., HEPPNER, D. G., BALLOU, W. R. & RICHIE, T. L. 2002. Protection of humans against malaria by immunization with radiation-attenuated Plasmodium falciparum sporozoites. *J Infect Dis,* 185**,** 1155-64.

HOLT, R. A., SUBRAMANIAN, G. M., HALPERN, A., SUTTON, G. G., CHARLAB, R., NUSSKERN, D. R., WINCKER, P., CLARK, A. G., RIBEIRO, J. M., WIDES, R., SALZBERG, S. L., LOFTUS, B., YANDELL, M., MAJOROS, W. H., RUSCH, D. B., LAI, Z., KRAFT, C. L., ABRIL, J. F., ANTHOUARD, V., ARENSBURGER, P., ATKINSON, P. W., BADEN, H., DE BERARDINIS, V., BALDWIN, D., BENES, V., BIEDLER, J., BLASS, C., BOLANOS, R., BOSCUS, D., BARNSTEAD, M., CAI, S., CENTER, A., CHATURVERDI, K., CHRISTOPHIDES, G. K., CHRYSTAL, M. A., CLAMP, M., CRAVCHIK, A., CURWEN, V., DANA, A., DELCHER, A., DEW, I., EVANS, C. A., FLANIGAN, M., GRUNDSCHOBER-FREIMOSER, A., FRIEDLI, L., GU, Z., GUAN, P., GUIGO, R., HILLENMEYER, M. E., HLADUN, S. L., HOGAN, J. R., HONG, Y. S., HOOVER, J., JAILLON, O., KE, Z., KODIRA, C., KOKOZA, E., KOUTSOS, A., LETUNIC, I., LEVITSKY, A., LIANG, Y., LIN, J. J., LOBO, N. F., LOPEZ, J. R., MALEK, J. A., MCINTOSH, T. C., MEISTER, S., MILLER, J., MOBARRY, C., MONGIN, E., MURPHY, S. D., O'BROCHTA, D. A., PFANNKOCH, C., QI, R., REGIER, M. A., REMINGTON, K., SHAO, H., SHARAKHOVA, M. V., SITTER, C. D., SHETTY, J., SMITH, T. J., STRONG, R., SUN, J., THOMASOVA, D., TON, L. Q., TOPALIS, P., TU, Z., UNGER, M. F., WALENZ, B., WANG, A., WANG, J., WANG, M., WANG, X., WOODFORD, K. J., WORTMAN, J. R., WU, M., YAO, A., ZDOBNOV, E. M., ZHANG, H., ZHAO, Q., et al. 2002. The genome sequence of the malaria mosquito Anopheles gambiae. *Science,* 298**,** 129-49.

HORII, T., SHIRAI, H., JIE, L., ISHII, K. J., PALACPAC, N. Q., TOUGAN, T., HATO, M., OHTA, N., BOBOGARE, A., ARAKAKI, N., MATSUMOTO, Y., NAMAZUE, J., ISHIKAWA, T., UEDA, S. & TAKAHASHI, M. 2010. Evidences of protection against blood-stage infection of Plasmodium falciparum by the novel protein vaccine SE36. *Parasitol Int,* 59**,** 380-6.

HUNT, R. H., COETZEE, M. & FETTENE, M. 1998. The Anopheles gambiae complex: a new species from Ethiopia. *Trans R Soc Trop Med Hyg,* 92**,** 231-5.

HURLES, M. E. & JOBLING, M. A. 2001. Haploid chromosomes in molecular ecology: lessons from the human Y. *Mol Ecol,* 10**,** 1599-613.

HURST, G. D. D. & WERREN, J. H. 2001. The role of selfish genetic elements in eukaryotic evolution. *Nat Rev Genet,* 2**,** 597-606.

JAMBOU, R., LEGRAND, E., NIANG, M., KHIM, N., LIM, P., VOLNEY, B., EKALA, M. T., BOUCHIER, C., ESTERRE, P., FANDEUR, T. & MERCEREAU-PUIJALON, O. 2005. Resistance of Plasmodium falciparum field isolates to in-vitro artemether and point mutations of the SERCA-type PfATPase6. *Lancet,* 366**,** 1960-3.

JAMES, A. A. 2005. Gene drive systems in mosquitoes: rules of the road. *Trends in Parasitology,* 21**,** 64-67.

JASINSKIENE, N., COATES, C. J., BENEDICT, M. Q., CORNEL, A. J., RAFFERTY, C. S., JAMES, A. A. & COLLINS, F. H. 1998. Stable transformation of the yellow fever mosquito, Aedes aegypti, with the Hermes element from the housefly. *Proceedings of the National Academy of Sciences,* 95**,** 3743-3747.

KNIPLING, E. 1955. Possibilities of insect control or eradication through the use of sexually sterile males. *J. Econ. Entomol*.

KOKOZA, V., AHMED, A., WIMMER, E. A. & RAIKHEL, A. S. 2001. Efficient transformation of the yellow fever mosquito Aedes aegypti using the piggyBac transposable element vector pBac[3xP3-EGFP afm]. *Insect Biochemistry and Molecular Biology,* 31**,** 1137-1143.

KRAFSUR, E. S. 1972. Production of reciprocal translocations in Anopheles gambiae species A. *Trans R Soc Trop Med Hyg,* 66**,** 22-3.

KRAFSUR, E. S., WHITTEN, C. J. & NOVY, J. E. 1987. Screwworm eradication in North and Central America. *Parasitol Today,* 3**,** 131-7.

KRZYWINSKI, J., NUSSKERN, D. R., KERN, M. K. & BESANSKY, N. J. 2004. Isolation and characterization of Y chromosome sequences from the African malaria mosquito Anopheles gambiae. *Genetics,* 166**,** 1291-1302.

LAHN, B. T., PEARSON, N. M. & JEGALIAN, K. 2001. The human Y chromosome, in the light of evolution. *Nat Rev Genet,* 2**,** 207-16.

LEE, G. C. L. A. Y. 2013. The Distribution of Genetic Polymorphism and Patterns of Reproductive Isolation Among Natural Populations

LEE, Y., CORNEL, A., MENESES, C., FOFANA, A., ANDRIANARIVO, A., MCABEE, R., FONDJO, E., TRAORE, S. & LANZARO, G. 2009. Ecological and genetic relationships of the Forest-M form among chromosomal and molecular forms of the malaria vector Anopheles gambiae sensu stricto. *Malaria Journal,* 8**,** 75.

LINDQUIST, D. A., ABUSOWA, M. & HALL, M. J. 1992. The New World screwworm fly in Libya: a review of its introduction and eradication. *Med Vet Entomol,* 6**,** 2-8.

LOBO, N. F., CLAYTON, J. R., FRASER, M. J., KAFATOS, F. C. & COLLINS, F. H. 2006. High efficiency germ-line transformation of mosquitoes. *Nat. Protocols,* 1**,** 1312-1317.

LOFGREN, C. S., DAME, D. A., BREELAND, S. G., WEIDHAAS, D. E., JEFFERY, G., KAISER, R., FORD, H. R., BOSTON, M. D. & BALDWIN, K. F. 1974. Release of chemosterilized males for the control of Anopheles albimanus in El Salvador. 3. Field methods and population control. *Am J Trop Med Hyg,* 23**,** 288-97.

MAGGERT, K. A. & GOLIC, K. G. 2002. The Y chromosome of Drosophila melanogaster exhibits chromosome-wide imprinting. *Genetics,* 162**,** 1245-58.

MAROIS, E., SCALI, C., SOICHOT, J., KAPPLER, C., LEVASHINA, E. & CATTERUCCIA, F. 2012. High-throughput sorting of mosquito larvae for laboratory studies and for future vector control interventions. *Malaria Journal,* 11**,** 302.

MARRELLI, M. T., LI, C., RASGON, J. L. & JACOBS-LORENA, M. 2007. Transgenic malaria-resistant mosquitoes have a fitness advantage when feeding on Plasmodium-infected blood. *Proceedings of the National Academy of Sciences,* 104**,** 5580-5583.

MARTINEZ-TORRES, D., CHANDRE, F., WILLIAMSON, M. S., DARRIET, F., BERGE, J. B., DEVONSHIRE, A. L., GUILLET, P., PASTEUR, N. & PAURON, D. 1998. Molecular characterization of pyrethroid knockdown resistance (kdr) in the major malaria vector Anopheles gambiae s.s. *Insect Mol Biol,* 7**,** 179-84.

MASON, G. F. 1967. Genetic studies on mutations in species A and B of the Anopheles gambiae complex. *Genet Res,* 10**,** 205-17.

MATUSCHEWSKI, K. 2006. Vaccine development against malaria. *Curr Opin Immunol,* 18**,** 449-57.

MAWEJJE, H. D., WILDING, C. S., RIPPON, E. J., HUGHES, A., WEETMAN, D. & DONNELLY, M. J. 2013. Insecticide resistance monitoring of field-collected Anopheles gambiae s.l. populations from Jinja, eastern Uganda, identifies high levels of pyrethroid resistance. *Med Vet Entomol,* 27**,** 276-83.

MAXWELL, C. A., MSUYA, E., SUDI, M., NJUNWA, K. J., CARNEIRO, I. A. & CURTIS, C. F. 2002. Effect of community-wide use of insecticide-treated nets for 3-4 years on malarial morbidity in Tanzania. *Trop Med Int Health,* 7**,** 1003-8.

MCGRAW, E. A. & O'NEILL, S. L. 2013. Beyond insecticides: new thinking on an ancient problem. *Nat Rev Micro,* 11**,** 181-193.

MENDES, A. M., SCHLEGELMILCH, T., COHUET, A., AWONO-AMBENE, P., DE IORIO, M., FONTENILLE, D., MORLAIS, I., CHRISTOPHIDES, G. K., KAFATOS, F. C. & VLACHOU, D. 2008. Conserved mosquito/parasite interactions affect development of Plasmodium falciparum in Africa. *PLoS Pathog,* 4**,** e1000069.

MEREDITH, J. M., BASU, S., NIMMO, D. D., LARGET-THIERY, I., WARR, E. L., UNDERHILL, A., MCARTHUR, C. C., CARTER, V., HURD, H., BOURGOUIN, C. & EGGLESTON, P. 2011. Site-specific integration and expression of an anti-malarial gene in transgenic Anopheles gambiae significantly reduces Plasmodium infections. *PLoS One,* 6**,** e14587.

MICHEL, K. & KAFATOS, F. C. 2005. Mosquito immunity against Plasmodium. *Insect Biochem Mol Biol,* 35**,** 677-89.

MILLER, D. 2004. Sex determination: insights from the human and animal models suggest that the mammalian Y chromosome is uniquely specialised for the male's benefit. *The Journal of Men's Health & Gender,* 1**,** 170-181.

MILLER, L., SAKAI, R., ROMANS, P., GWADZ, R., KANTOFF, P. & COON, H. 1987. Stable integration and expression of a bacterial gene in the mosquito Anopheles gambiae. *Science,* 237**,** 779-781.

MUELLER, A. K., CAMARGO, N., KAISER, K., ANDORFER, C., FREVERT, U., MATUSCHEWSKI, K. & KAPPE, S. H. 2005. Plasmodium liver stage developmental arrest by depletion of a protein at the parasite-host interface. *Proc Natl Acad Sci U S A,* 102**,** 3022-7.

MULLER, H. J. 1932. Some Genetic Aspects of Sex. *The American Naturalist,* 66**,** 118-138.

MUNOZ-LOPEZ, M. & GARCIA-PEREZ, J. L. 2010. DNA transposons: nature and applications in genomics. *Curr Genomics,* 11**,** 115-28.

OGUTU, B. R., APOLLO, O. J., MCKINNEY, D., OKOTH, W., SIANGLA, J., DUBOVSKY, F., TUCKER, K., WAITUMBI, J. N., DIGGS, C., WITTES, J., MALKIN, E., LEACH, A., SOISSON, L. A., MILMAN, J. B., OTIENO, L., HOLLAND, C. A., POLHEMUS, M., REMICH, S. A., OCKENHOUSE, C. F., COHEN, J., BALLOU, W. R., MARTIN, S. K., ANGOV, E., STEWART, V. A., LYON, J. A., HEPPNER, D. G., JR., WITHERS, M. R. & FOR THE, M. S. P. M. V. W. G. 2009. Blood Stage Malaria Vaccine Eliciting High Antigen-Specific Antibody Concentrations Confers No Protection to Young Children in Western Kenya. *PLoS ONE,* 4**,** e4708.

OKA, A., MITA, A., SAKURAI-YAMATANI, N., YAMAMOTO, H., TAKAGI, N., TAKANO-SHIMIZU, T., TOSHIMORI, K., MORIWAKI, K. & SHIROISHI, T. 2004. Hybrid breakdown caused by substitution of the X chromosome between two mouse subspecies. *Genetics,* 166**,** 913-924.

OSTA, M. A., CHRISTOPHIDES, G. K., VLACHOU, D. & KAFATOS, F. C. 2004. Innate immunity in the malaria vector Anopheles gambiae: comparative and functional genomics. *J Exp Biol,* 207**,** 2551-63.

PAPATHANOS, P., WINDBICHLER, N., MENICHELLI, M., BURT, A. & CRISANTI, A. 2009. The vasa regulatory region mediates germline expression and maternal transmission of proteins in the malaria mosquito Anopheles gambiae: a versatile tool for genetic control strategies. *BMC Molecular Biology,* 10**,** 65.

PASVOL, G. 2010. Protective hemoglobinopathies and Plasmodium falciparum transmission. *Nat Genet,* 42**,** 284-285.

PERERA, O. P., HARRELL II, R. A. & HANDLER, A. M. 2002. Germ-line transformation of the South American malaria vector, Anopheles albimanus, with a piggyBac/EGFP transposon vector is routine and highly efficient. *Insect Molecular Biology,* 11**,** 291-297.

PETERS, W. 1987. Chemotherapy and drug resistance in malaria. *New York: Academic Press*.

PETERSEN, I., EASTMAN, R. & LANZER, M. 2011. Drug-resistant malaria: molecular mechanisms and implications for public health. *FEBS Lett,* 585**,** 1551-62.

PHILLIPS, R. S. 2001. Current Status of Malaria and Potential for Control. *Clinical Microbiology Reviews,* 14**,** 208-226.

POMBI, M., CAPUTO, B., SIMARD, F., DI DECO, M., COLUZZI, M., TORRE, A., COSTANTINI, C., BESANSKY, N. & PETRARCA, V. 2008. Chromosomal plasticity and evolutionary potential in the malaria vector Anopheles gambiae sensu stricto: insights from three decades of rare paracentric inversions. *BMC Evolutionary Biology,* 8**,** 309.

RABBANI, M. G. & KITZMILLER, J. B. 1975. Studies on x-ray induced chromosomal translocations in Anopheles albimanus. II. Laboratory evaluation of sexual competitiveness of translocation males. *Am J Trop Med Hyg,* 24**,** 1027-30.

RABBANI MG, K. J. 1972. Chromosomal translocations in Anopheles albimanus Wiedemann. *Mosquito News,* 32.

RANDELL, H. F., DICKINSON, K. L., SHAYO, E. H., MBOERA, L. E. & KRAMER, R. A. 2010. Environmental management for malaria control: knowledge and practices in Mvomero, Tanzania. *Ecohealth,* 7**,** 507-16.

RICE, W. R. 1984. Sex Chromosomes and the Evolution of Sexual Dimorphism. *Evolution,* 38**,** 735-742.

RICE, W. R. 1994. Degeneration of a nonrecombining chromosome. *Science,* 263**,** 230-2.

RICHARD H. BAKER, R. K. S., UMAIMA T. SAIFUDDIN AND ALMAS PERVEEN 1978. Induced chromosomal aberrations in Anopheles culicifacies. *Mosquito News*.

SAGARA, I., DICKO, A., ELLIS, R. D., FAY, M. P., DIAWARA, S. I., ASSADOU, M. H., SISSOKO, M. S., KONE, M., DIALLO, A. I., SAYE, R., GUINDO, M. A., KANTE, O., NIAMBELE, M. B., MIURA, K., MULLEN, G. E., PIERCE, M., MARTIN, L. B., DOLO, A., DIALLO, D. A., DOUMBO, O. K., MILLER, L. H. & SAUL, A. 2009. A randomized controlled phase 2 trial of the blood stage AMA1-C1/Alhydrogel malaria vaccine in children in Mali. *Vaccine,* 27**,** 3090-8.

SALZ, H. K. & ERICKSON, J. W. 2010. Sex determination in Drosophila: The view from the top. *Fly (Austin),* 4**,** 60-70.

SAUL, A. & FAY, M. P. 2007. Human Immunity and the Design of Multi-Component, Single Target Vaccines. *PLoS ONE,* 2**,** e850.

SCALI, C., CATTERUCCIA, F., LI, Q. & CRISANTI, A. 2005. Identification of sex-specific transcripts of the Anopheles gambiae doublesex gene. *The Journal of experimental biology,* 208**,** 3701-3709.

SCHILTHUIZEN, M., GIESBERS, M. C. W. G. & BEUKEBOOM, L. W. 2011. Haldane's rule in the 21st century. *Heredity,* 107**,** 95-102.

SHERMAN, I. W. 1998. *Malaria: Parasite Biology, Pathogenesis, and Protection*, ASM Press.

SINGH, B. & DANESHVAR, C. 2013. Human infections and detection of Plasmodium knowlesi. *Clin Microbiol Rev,* 26**,** 165-84.

SINKINS, S. P. & GOULD, F. 2006. Gene drive systems for insect disease vectors. *Nat Rev Genet,* 7**,** 427-435.

SLOTMAN, M., DELLA TORRE, A. & POWELL, J. R. 2004. The genetics of inviability and male sterility in hybrids between Anopheles gambiae and An. arabiensis. *Genetics,* 167**,** 275-87.

SLOTMAN, M., TORRE, A. D. & POWELL, J. R. 2005. Female Sterility in Hybrids between Anopheles gambiae and A. arabiensis, and the Causes of Haldane's Rule. *Evolution,* 59**,** 1016-1026.

STEINEMANN, M. & STEINEMANN, S. 1998. Enigma of Y chromosome degeneration: Neo-Y and Neo-X chromosomes of Drosophila miranda a model for sex chromosome evolution. *Genetica,* 102-103**,** 409-420.

STEINEMANN, M. & STEINEMANN, S. 2000. Common mechanisms of Y chromosome evolution. *Genetica,* 109**,** 105-11.

STODDARD, B. L. 2005. Homing endonuclease structure and function. *Q Rev Biophys,* 38**,** 49-95.

STOWERS, A. & CARTER, R. 2001. Current developments in malaria transmission-blocking vaccines. *Expert Opin Biol Ther,* 1**,** 619-28.

SWEENY, T. L. & BARR, A. R. 1978. Sex Ratio Distortion Caused by Meiotic Drive in a Mosquito, Culex pipiens L. *Genetics,* 88**,** 427-446.

TEMU, E. A., HUNT, R. H., COETZEE, M., MINJAS, J. N. & SHIFF, C. J. 1997. Detection of hybrids in natural populations of the Anopheles gambiae complex by the rDNA-based, PCR method. *Ann Trop Med Parasitol,* 91**,** 963-5.

THOMAS, D. D., DONNELLY, C. A., WOOD, R. J. & ALPHEY, L. S. 2000. Insect Population Control Using a Dominant, Repressible, Lethal Genetic System. *Science,* 287**,** 2474-2476.

THORPE, H. M. & SMITH, M. C. M. 1998. In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proceedings of the National Academy of Sciences,* 95**,** 5505-5510.

TIMOSHEVSKIY, V. A., SHARMA, A., SHARAKHOV, I. V. & SHARAKHOVA, M. V. 2012. Fluorescent in situ hybridization on mitotic chromosomes of mosquitoes. *J Vis Exp***,** e4215.

TOURE, Y. T., PETRARCA, V., TRAORE, S. F., COULIBALY, A., MAIGA, H. M., SANKARE, O., SOW, M., DI DECO, M. A. & COLUZZI, M. 1998. The distribution and inversion polymorphism of chromosomally recognized taxa of the Anopheles gambiae complex in Mali, West Africa. *Parassitologia,* 40**,** 477-511.

TOWNSON, H., NATHAN, M. B., ZAIM, M., GUILLET, P., MANGA, L., BOS, R. & KINDHAUSER, M. 2005. Exploiting the potential of vector control for disease prevention. *Bull World Health Organ,* 83**,** 942-7.

TURNER, T. L., HAHN, M. W. & NUZHDIN, S. V. 2005. Genomic islands of speciation in Anopheles gambiae. *PLoS Biol,* 3**,** e285.

UHLEMANN, A. C., CAMERON, A., ECKSTEIN-LUDWIG, U., FISCHBARG, J., ISEROVICH, P., ZUNIGA, F. A., EAST, M., LEE, A., BRADY, L., HAYNES, R. K. & KRISHNA, S. 2005. A single amino acid residue can determine the sensitivity of SERCAs to artemisinins. *Nat Struct Mol Biol,* 12**,** 628-9.

VAN DIJK, M. R., DOURADINHA, B., FRANKE-FAYARD, B., HEUSSLER, V., VAN DOOREN, M. W., VAN SCHAIJK, B., VAN GEMERT, G. J., SAUERWEIN, R. W., MOTA, M. M., WATERS, A. P. & JANSE, C. J. 2005. Genetically attenuated, P36p-deficient malarial sporozoites induce protective immunity and apoptosis of infected liver cells. *Proc Natl Acad Sci U S A,* 102**,** 12194-9.

VIBRANOVSKI, M. D. 2014. Meiotic sex chromosome inactivation in Drosophila. *J Genomics,* 2**,** 104-17.

WALKER, D. J., PITSCH, J. L., PENG, M. M., ROBINSON, B. L., PETERS, W., BHISUTTHIBHAN, J. & MESHNICK, S. R. 2000. Mechanisms of artemisinin resistance in the rodent malaria pathogen Plasmodium yoelii. *Antimicrob Agents Chemother,* 44**,** 344-7.

WATERHOUSE, R. M., KRIVENTSEVA, E. V., MEISTER, S., XI, Z., ALVAREZ, K. S., BARTHOLOMAY, L. C., BARILLAS-MURY, C., BIAN, G., BLANDIN, S., CHRISTENSEN, B. M., DONG, Y., JIANG, H., KANOST, M. R., KOUTSOS, A. C., LEVASHINA, E. A., LI, J., LIGOXYGAKIS, P., MACCALLUM, R. M., MAYHEW, G. F., MENDES, A., MICHEL, K., OSTA, M. A., PASKEWITZ, S., SHIN, S. W., VLACHOU, D., WANG, L., WEI, W., ZHENG, L., ZOU, Z., SEVERSON, D. W., RAIKHEL, A. S., KAFATOS, F. C., DIMOPOULOS, G., ZDOBNOV, E. M. & CHRISTOPHIDES, G. K. 2007. Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science,* 316**,** 1738-43.

WEETMAN, D., STEEN, K., RIPPON, E., MAWEJJE, H., DONNELLY, M. & WILDING, C. 2014. Contemporary gene flow between wild An. gambiae s.s. and An. arabiensis. *Parasites & Vectors,* 7**,** 345.

WHITE, G. B. 1985. Anopheles bwambae sp.n., a malaria vector in the Semliki Valley, Uganda, and its relationships with other sibling species of the An.gambiae complex (Diptera: Culicidae). *Systematic Entomology,* 10**,** 501-522.

WHO 1995. Vector control for malaria and other mosquito-borne diseases.

WHO 2012. World Malaria Report 2012. WHO; World Health Organization.

WILHELM, D., PALMER, S. & KOOPMAN, P. 2007. Sex determination and gonadal development in mammals. *Physiol Rev,* 87**,** 1-28.

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. & CRISANTI, A. 2011. A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature,* 473**,** 212-215.

WINDBICHLER, N., PAPATHANOS, P. A., CATTERUCCIA, F., RANSON, H., BURT, A. & CRISANTI, A. 2007. Homing endonuclease mediated gene targeting in Anopheles gambiae cells and embryos. *Nucleic Acids Research,* 35**,** 5922-5933.

WINDBICHLER, N., PAPATHANOS, P. A. & 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**,** e1000291.

WU, C.-I. 2001. The genic view of the process of speciation. *Journal of Evolutionary Biology,* 14**,** 851-865.

WU, C.-I. & TING, C.-T. 2004. Genes and speciation. *Nat Rev Genet,* 5**,** 114-122.

WU, C. I. & DAVIS, A. W. 1993. Evolution of postmating reproductive isolation: the composite nature of Haldane's rule and its genetic bases. *Am Nat,* 142**,** 187-212.

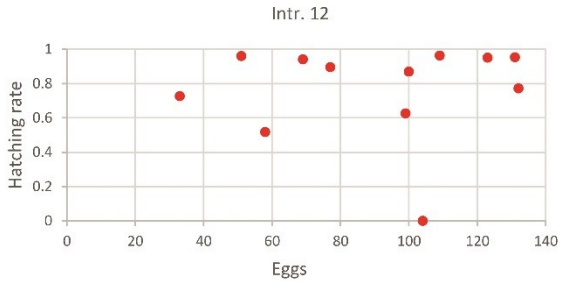
9 Appendix

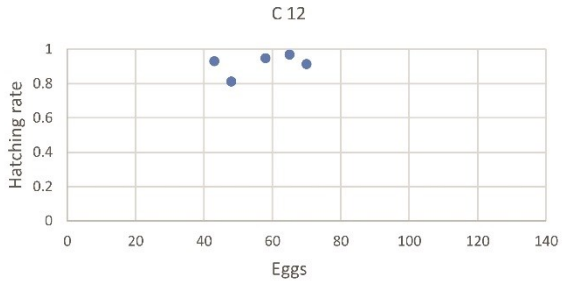
9.1 Overview of transgenic lines generated

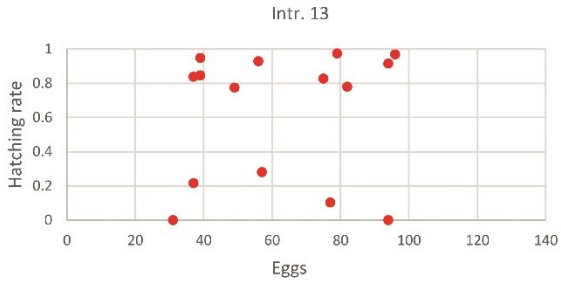
- T4 line: the Y-linked construct contains 3xP3-GFP and Actin5C-RFP fluorescent markers. In addition, the construct contains the 18-bp I-SceI endonuclease recognition site, located within the GFP ORF.
- YAttP line: the Y-linked construct carries the RFP marker under the control of the *3xP3* promoter. An *AttP* recombination site is placed between the promoter and the RFP coding sequence.
- YVasG line: the Y-linked construct contains an *AttB* site placed upstream of a promoterless eCFP coding sequence. In addition, it carries a cassette in which GFP is placed under the control of the germline promoter *vasa*.
- YI-PpoI line: the Y-linked construct expresses the hendonuclease *I-PpoI* driven by the male spermatogenesis-specific *b2 tubulin* promoter. The YI-PpoI line was generated in our lab as part of a different experimental project (unpublished data).

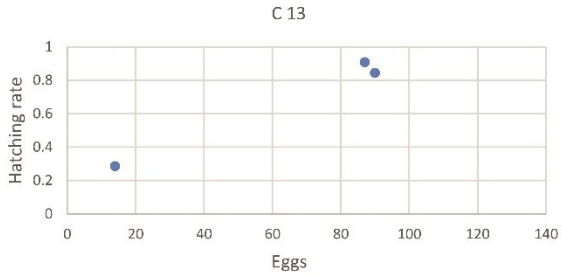
9.2 Single backcrosses of *YI-PpoI*-introgressed males in cups

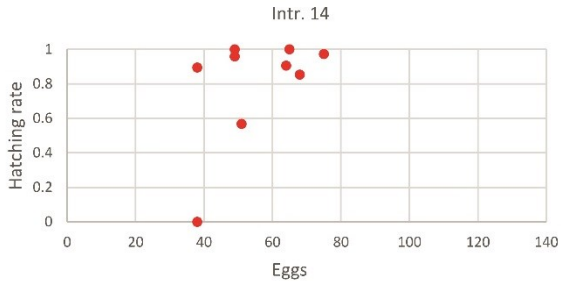
The graphs show the hatching rate data measured for single backcrosses of *YI-PpoI*-introgressed males to *A. arabiensis* females in cups. These data span 7 generations of backcrossing. On the left side, the hatching rate relative to the *introgressed cups* is indicated by red dots (Intr. 12-18). On the right, the hatching rate relative to the *control cups* is indicated by blue dots (C 12-18).

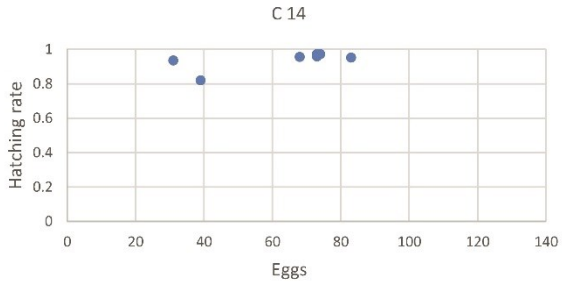


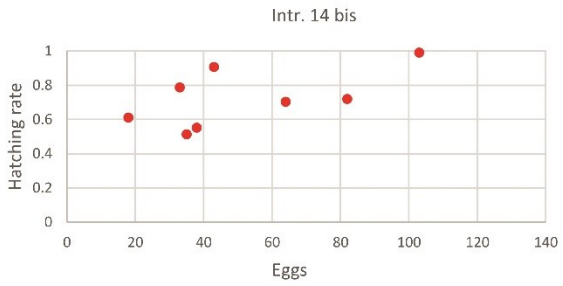


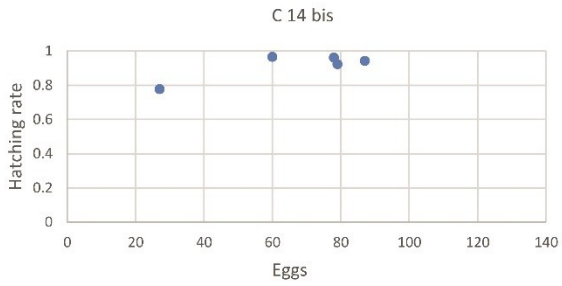


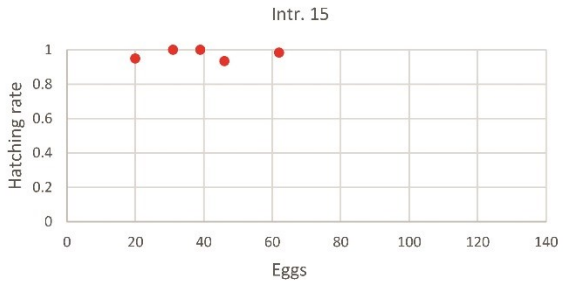


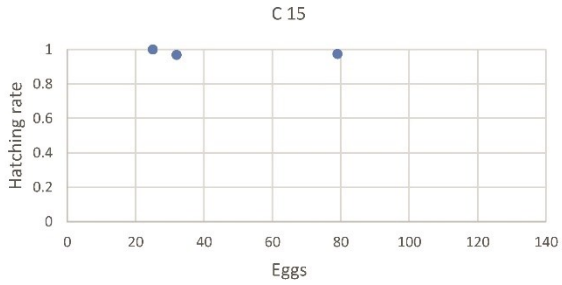


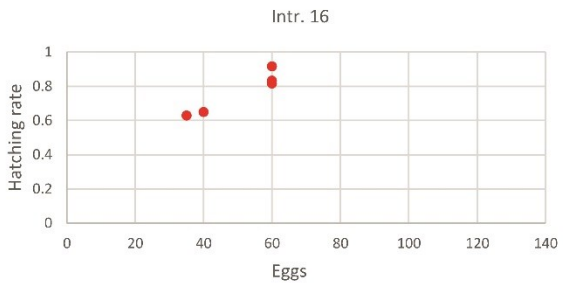


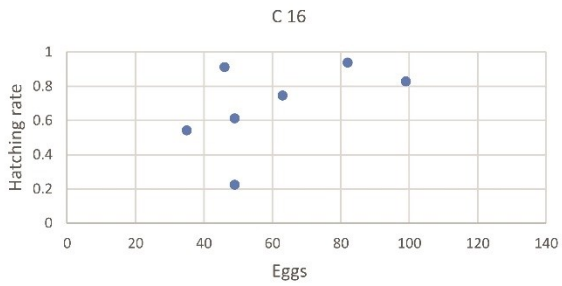

Intr. 14 bis
Hatching rate
Eggs
0
0.2
0.4
0.6
0.8
1
0
20
40
60
80
100
120
140

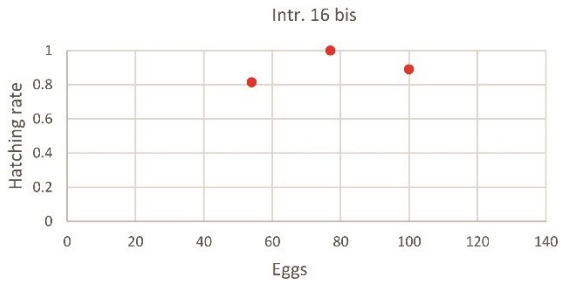

C 14 bis
Hatching rate
Eggs
0
0.2
0.4
0.6
0.8
1
0
20
40
60
80
100
120
140

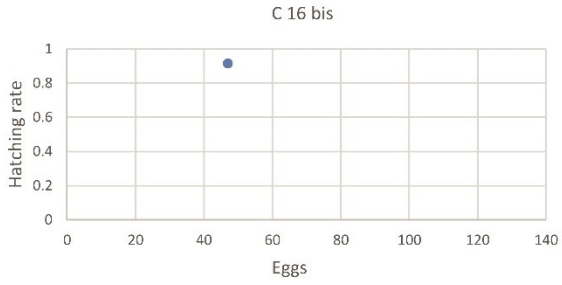

Intr. 15
Hatching rate
Eggs
0
0.2
0.4
0.6
0.8
1
0
20
40
60
80
100
120
140

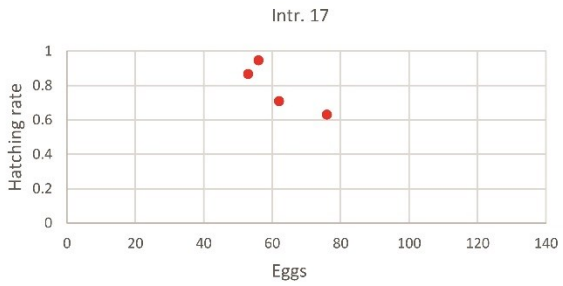

C 15
Hatching rate
Eggs
0
0.2
0.4
0.6
0.8
1
0
20
40
60
80
100
120
140

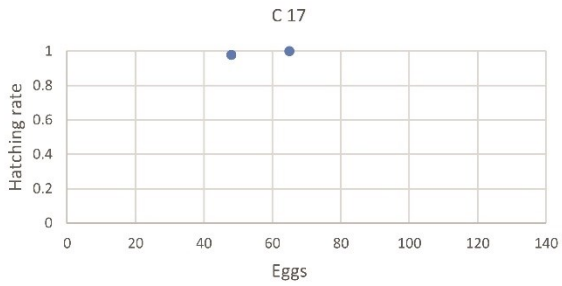

Intr. 16
Hatching rate
Eggs
0
0.2
0.4
0.6
0.8
1
0
20
40
60
80
100
120
140

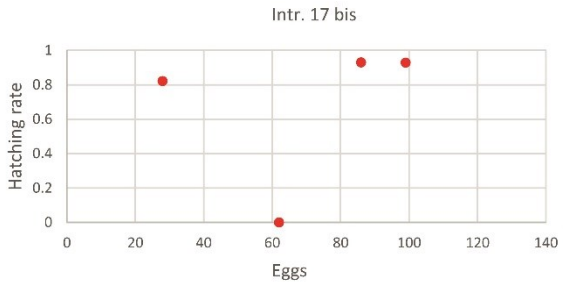

C 16
Hatching rate
Eggs
0
0.2
0.4
0.6
0.8
1
0
20
40
60
80
100
120
140

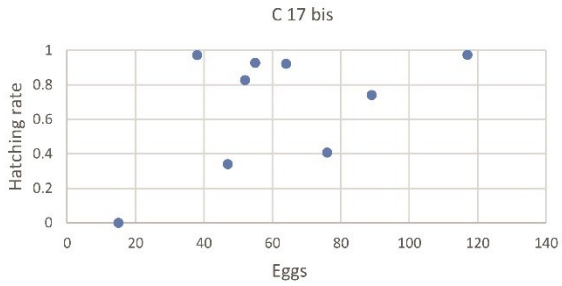

Intr. 16 bis
Hatching rate
Eggs
0
0.2
0.4
0.6
0.8
1
0
20
40
60
80
100
120
140

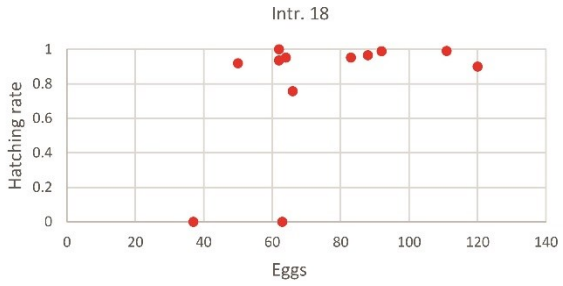

C 16 bis
Hatching rate
Eggs
0
0.2
0.4
0.6
0.8
1
0
20
40
60
80
100
120
140

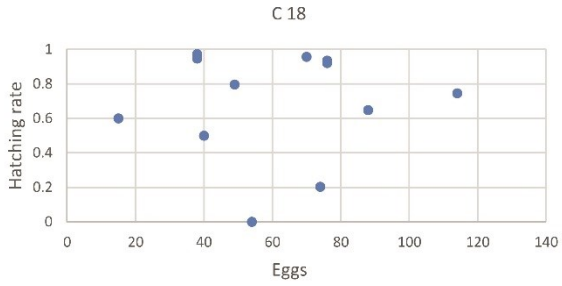

Intr. 17
Hatching rate
Eggs
0
0.2
0.4
0.6
0.8
1
0
20
40
60
80
100
120
140


C 17
Hatching rate
Eggs
0
0.2
0.4
0.6
0.8
1
0
20
40
60
80
100
120
140


Intr. 17 bis
Hatching rate
Eggs
0
0.2
0.4
0.6
0.8
1
0
20
40
60
80
100
120
140


C 17 bis
Hatching rate
Eggs
0
0.2
0.4
0.6
0.8
1
0
20
40
60
80
100
120
140


Intr. 18
Hatching rate
Eggs
0
0.2
0.4
0.6
0.8
1
0
20
40
60
80
100
120
140


C 18
Hatching rate
Eggs
0
0.2
0.4
0.6
0.8
1
0
20
40
60
80
100
120
140

9.3 List of primers

PRIMERS	SEQUENCE (5'-3')
ACTIN5REV	tggaaatgagaagtaaggtgcatctgca
EXFP3SEQ	atccgccacaacatcgaggacg
FORI-SCEI	cgcgtaatacgactcactatagggcgaattgggtaccgggccccccctcgaggtcgacggtatcgataagcttgatatcat-gggatcatcatcagacgacgaagctacggccgatgcgca
REVI-SCEI	ttttttttttttttttttttttatttcaggaaagtttcggaggagatagtgttcggcagtttgtaca
F1	ctttgagttctctcagttgggggcacgcaacatggtgcgctcctccaagaacg
R2	gttaccccagttggggcgtagggtctagtcgactctagcggtaccccgattg
UN	gtgtgccccttcctcgatgt
GA	ctggtttggtcggcacgttt
YSPEC2	acgcgacacagtacgcga
ALT3R2	agtataccatcttagctggttcg
ALT5F2	cgatgagcatatcctctctgc
ORFA (324BP) CONTIG 1	Tgcgcactgcccgct
	gcggcgattaccggtgc
ORFB (309BP) CONTIG 2	ccatagccaacctcgccaact
	atttggtttcgggcggtttatgt
ORFC (384BP) CONTIG 2	ctgcccttgtgaacctaggacatg
	ggttacgatggatattcaggtgcg
ORF2 (378BP) CONTIG 2	tctttgatgtggttggtgatctgtacg
	atcctcaacaagggaaacgagcac
ORF10 (1110BP) CONTIG 3	gacgttaagtcggcatacttgtatgga
	cagcttggcaaggattgctca
BIGORF2 (654BP) CONTIG 4	atggcccgttccctcttaggc
	tggattaggtcggtgttgatctctgag
BIGORF1 (495BP) CONTIG 4	actgcgttcggcaaaagacg
	atgcatcaggcttcactcgatg
BIGORF2 (654BP) CONTIG 4	atggcccgttccctcttaggc
	tggattaggtcggtgttgatctctgag
BIGORF1 (495BP) CONTIG 4	actgcgttcggcaaaagacg
	atgcatcaggcttcactcgatg
ORF1 (219BP) CONTIG 4	cagatctcttcccaggataccaatcatg
	cgtccacttcgtggagctgag
ORF10 (408BP) CONTIG 4	atcggtacaattcacatcaaacatcgat
	taggtgaatgtggcgtcagctg
ORF12 (468BP) CONTIG 4	catagcttccctcctcgaactcatc
	ccatggagggttgcgtaatgagta
ORF12 (435BP) CONTIG 5	ataactgccatttgggacaaccgt
	ttacatttccgtaagctcggtcgtctac
BIGORF2 (432BP) CONTIG 5	gtgccgccggacgg
	actcgttcttttgggcagccatt
ORF17 (399BP) CONTIG 5	gaagcacaaaaagcacccgac
	gctttcgctatttgtttctcattcga
ORF4 (336BP) CONTIG 6	ttcaagtgtacatattcgtttaaatggatgaa
	atggcttatggcatgttagaatcggt

ORF10 (354BP)	ttacgatgaatttactgccgtataggtatggt
CONTIG 6	atgtctatgtgtttatttttgacagatgaatgc
ORF11 (303BP)	gaacattttactggattaaaatttaaatggtaagaccttaac
CONTIG 6	gattccatcaaataccttttagaaaatgattcaaga
ORF12 (486BP)	atgttctactgtgagaacctgtgctgtacg
CONTIG 6	agtcatggcatcaagagcatatacatagttc
ORF3 (390BP)	tggtgtccatgttggctgattg
CONTIG 7	tgctaatactgcaatcgccgactg
ORF6 (315BP)	atggaattaacttgcgccagctg
CONTIG 7	gaaacctgaccaccgcgagt
ORF20 (528BP)	gcgctgagaatgctcagtgc
CONTIG 8	ccgcgcaactgttaccattacg
ORF21 (414BP)	gacttatccagccctttgcattgtc
CONTIG 8	gactcaatcactagcaaacaaacacatgc
ORF2 (480BP)	aaccggcagctttctcccag
CONTIG 9	tgcgaagcgtacgagagacca
ORF7 (384BP)	attcagccacactgtcgttgatgt
CONTIG 10	ccagtcaagctctactcgatgaactaagc
ORF18 (384BP)	aggatcgtgttgatgactcggtagatagt
CONTIG 10	atgatcaagacactaaggaagcgtcca
ORF20 (438BP)	actgcaagcgtacagccacg
CONTIG 11	tgtgttctattattctccattgaccatcgt
ORF21 (426BP)	cgcaagacgtccggtaaacg
CONTIG 11	aatggtactgcaccgaccgc
ORF22 (354BP)	cgtgcatcgcgagtggag
CONTIG 11	tctacaatcgctccgaggattacagatac
ORF23 (315BP)	agcgacgctgggttagaccg
CONTIG 11	attcgtggcgccgtgct
ORF24 (528BP)	gtgtcgaagtgcctgtgacca
CONTIG 11	cagacgctgacagcgtctcttc
ORFA (561BP)	tcaagagtggtccgcgacct
CONTIG 12	gccgtggccatattgtatcgtg
BIGORF1 (477BP)	acgtctgtgcggctatgacgag
CONTIG 12	catgttgatccgtctcgaccaac
ORFH (705BP)	Tatctgccacgcgccacat
CONTIG 12	gggagtacgaagcgaatcgaagac
ORFE (477BP)	gtacgcaacacaacaggatgcag
CONTIG 12	atccgtatagccacaagcagcg
ORFF (492BP)	gagcacatccggcctactgc
CONTIG 12	gcgtacgtatctacacgaccagctg
ORFG (465BP)	ggttacgcccgtagcaagaacg
CONTIG 12	gcaatccgccgtacgtgc
ORFI (339BP)	tgtccgtctccggaatccaattac
CONTIG 12	ggtcgagactgtaggacgaatacatcg
ORFL (441BP)	Tcgaggccagcgcgatac
CONTIG 12	gcatcgtgatgcatgttccgt
ORFM (306BP)	gacctacctgtccagagggacatattagc
CONTIG 12	caagcgtcgattatgacaggttatgc

ORFB (366BP) CONTIG 12	atggctgctcccaagatgaacg ggcggacttcgacgattctg

9.4 Scientific publications

Attached are the scientific publications related to the body of the research described within this thesis.

Site-specific genetic engineering of the *Anopheles gambiae* Y chromosome

Federica Bernardini[a], Roberto Galizi[a], Miriam Menichelli[a], Philippos-Aris Papathanos[b], Vicky Dritsou[b], Eric Marois[c], Andrea Crisanti[a,b], and Nikolai Windbichler[a,1]

[a]Department of Life Sciences, South Kensington Campus, Imperial College London, London SW7 2AZ, United Kingdom; [b]Dipartimento di Medicina Sperimentale Via Gambuli, Centro di Genomica Funzionale, University of Perugia, 06132 Perugia, Italy; and [c]Institut de Biologie Moléculaire et Cellulaire, Institut National de la Santé et de la Recherche Médicale U963, Centre National de la Recherche Scientifique, Unité Propre de Recherche 9022, 67084 Strasbourg, France

Edited by Anthony A. James, University of California, Irvine, CA, and approved April 17, 2014 (received for review March 18, 2014)

Despite its function in sex determination and its role in driving genome evolution, the Y chromosome remains poorly understood in most species. Y chromosomes are gene-poor, repeat-rich and largely heterochromatic and therefore represent a difficult target for genetic engineering. The Y chromosome of the human malaria vector *Anopheles gambiae* appears to be involved in sex determination although very little is known about both its structure and function. Here, we characterize a transgenic strain of this mosquito species, obtained by transposon-mediated integration of a transgene construct onto the Y chromosome. Using meganuclease-induced homologous repair we introduce a site-specific recombination signal onto the Y chromosome and show that the resulting docking line can be used for secondary integration. To demonstrate its utility, we study the activity of a germ-line–specific promoter when located on the Y chromosome. We also show that Y-linked fluorescent transgenes allow automated sex separation of this important vector species, providing the means to generate large single-sex populations. Our findings will aid studies of sex chromosome function and enable the development of male-exclusive genetic traits for vector control.

biotechnology | SIT

Mosquito species of the *Anopheles gambiae* complex represent the principal vectors of human malaria, and they pose an enormous burden on global health and economies. Every year, 300–500 million people are infected by malaria and more than 1 million people die as a consequence of *Plasmodium* parasite infections (1). The malaria mosquito *A. gambiae* has two pairs of autosomes, termed 2 and 3, and a pair of heteromorphic sex chromosomes X and Y, XX in females and XY in males (2). Extensive nonpairing regions exist between the X and the degenerate Y chromosome. and evidence points to a factor located on the Y chromosome that primarily determines the sex in *Anopheles* (3). Current models suggest that the evolutionary differentiation of Y chromosomes begins with the acquisition of a male determining factor on a proto-Y chromosome (4, 5). This event is followed by a progressive suppression of recombination between the still largely homomorphic proto-sex chromosomes, a process attributed to the acquisition of sexually antagonistic mutations, which are beneficial to the heterogametic sex but detrimental to the homogametic sex (6–8). The lack of recombination, together with the male-limited transmission, leads to the degeneration of the Y chromosome, which involves accumulation of deleterious mutations, spread of transposable elements, and silencing of all or most of the genes present on the proto-Y (9–11). As a result, Y chromosomes of many species appear to be strongly heterochromatic and harbor only few genes often involved in male fertility (12–17). The accumulation of repetitive sequences, many of which are also present on other chromosomes, hampers the assembly of Y chromosome contigs following shotgun sequencing. Indeed, despite the completion of the *A. gambiae* genome project (18), and the knowledge that the primary signal is likely to be associated with the inheritance of the Y (3, 19), no assembly of the *Anopheles* Y chromosome has been achieved. At present, public databases host only a few hundred kilobases of *A. gambiae* sequences attributed to the Y, a chromosome that is estimated to comprise 10% of the genome and to be at least 20 Mb in size. None of these Y-specific scaffolds have been physically mapped, because the Y chromosome does not polytenize. The exploration of the Y chromosome will improve our understanding of the evolutionary forces involved in driving chromosome evolution and may enable the manipulation of the molecular pathways that control sex determination and sexual differentiation in mosquitoes. In a number of organisms, Y chromosome genes have been found to be essential for male fertility or sex determination. Recently, a number of excellent candidate genes potentially involved in these processes have been identified on the Y chromosome of anopheline mosquitoes (20, 21). Because interfering with male fertility is an essential part of vector control strategies such as the sterile insect technique, the identification of such genes is of particular interest to mosquito biologists. In this paper, we demonstrate the targeted molecular manipulation of the Y chromosome in *A. gambiae*, thus opening up a number of ways to explore one of the most fascinating of evolution's upshots and to harness this genetic tool for vector control.

Significance

Interfering with sex determination and male fertility are potentially powerful approaches for the genetic control of the human malaria vector *Anopheles gambiae*. Despite this fact, the male-specific Y chromosome of this mosquito has remained largely unexplored, because of its repetitive, heterochromatic structure. Little is known about its ability to support gene transcription in different tissues and during gametogenesis, yet this information is crucial for understanding the function of this chromosome. We show, using a combination of knock-in and site-specific genetic engineering steps, how transgenes can be specifically introduced onto the Y chromosome. The Y-linked strains we have created provide the means to generate large single-sex populations and to establish male-exclusive genetic traits for the control of this important vector species.

Author contributions: N.W. designed research; F.B., R.G., M.M., and E.M. performed research; P.-A.P. and N.W. contributed new reagents/analytic tools; F.B., P.-A.P., V.D., E.M., and N.W. analyzed data; and A.C. and N.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession nos. KJ608148–KJ608160).

[1]To whom correspondence should be addressed. E-mail: n.windbichler@imperial.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1404996111/-/DCSupplemental.

Results

The Transgenic Line T4 Carries a Transgene That Segregates with the Y Chromosome. We injected wild-type *A. gambiae* embryos with the plasmid pHome-T together with in vitro-transcribed piggyBac transposase helper RNA (*Materials and Methods*) (22). The pHome-T plasmid contains piggyBac inverted repeats for transposase-mediated random integration and the 3xP3-GFP and Actin5C-RFP fluorescent markers. In addition, the construct contains the 18-bp I-SceI endonuclease recognition site, located within the GFP ORF. This site is not present in the *Anopheles* genome and can be used to specifically cut chromosomes carrying the integrated plasmid exclusively at this position. We obtained a number of G_0 mosquitoes expressing the fluorescent markers transiently, and they were separately crossed to wild-type mosquitoes. One male G_0 founder gave rise to 42 transgenic offspring, all of which emerged as males. When five of these transgenic males were crossed to wild-type females in single mating experiments, we again observed that all transgenic larvae were emerging as male adults (206 of 421), whereas the non-transgenic larvae (215 of 421) emerged as females. These results indicated that the transgene construct (Fig. 1) had potentially integrated within the Y chromosome. Expression of both fluorescent marker genes in this transgenic line, termed T4, was clearly visible (Fig. 2); however, the expression of RFP driven by the *Drosophila* Actin5C promoter was patchy and visible mainly in the periphery of the larval gut. The pattern of neuronal GFP fluorescence appeared to be normal, resembling that of autosomal insertions. We performed fluorescence in situ hybridization (FISH)

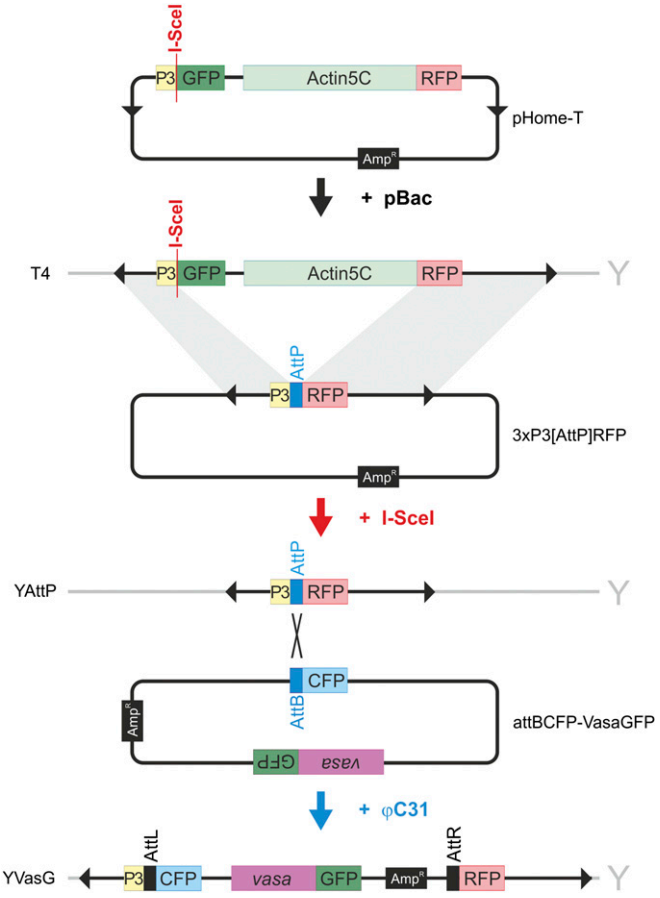


Fig. 1. Site-specific genetic engineering of the Y chromosome. Overview of constructs and transgenic lines generated in this study and the stepwise approach taken to modify the Y chromosome.

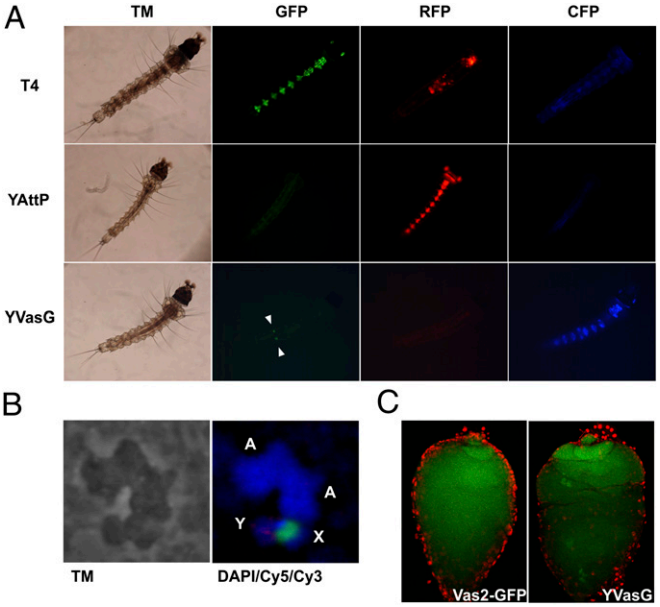


Fig. 2. Phenotype of Y-linked strains. (*A*) Expression of green (GFP), red (RFP), and blue (CFP) fluorescent markers in the tissues of L2 larvae shown as transmitted-light image (TM) in *Left*. The white arrowheads indicate GFP expression in the developing larval gonads. (*B*) FISH with probes designed against the X-linked rDNA (labeled with Cy5) and the pHome-T construct (labeled with Cy3) hybridized to the *A. gambiae* transgenic line. Mitotic chromosome slide preparations were prepared from imaginal discs of fourth instar larvae. X and Y indicate the sex chromosomes and A indicates autosomes. (*C*) Confocal analysis of GFP expression in dissected testes of transgenic adult males.

by using the pHome-T plasmid as a probe against mitotic chromosome spreads generated from male larval imaginal discs of line T4. As a control, the multicopy ribosomal DNA (rDNA) locus on the X chromosome was used. In all cases, a single weak signal was obtained from the smaller sex chromosome (Fig. 2*B* and Fig. S1), supporting the notion that the transformation construct had integrated into the Y chromosome.

Characterization of the Genomic Neighborhood of the T4 Transgene. Expression of the fluorescent reporter transgenes suggested that the construct had inserted within a region of chromatin allowing transcriptional activity. We performed genomic mapping by inverse PCR using primers binding to the PiggyBac-inverted repeats to characterize the genomic region flanking the integrated construct. We obtained genomic sequences that aligned to AAAB01003622.1, a scaffold not assigned to a known chromosome but that was characterized as a scaffold containing fragments originating exclusively from male libraries (23). This scaffold has a length of only 1.7 kb, and we therefore generated a BAC library from genomic DNA of T4 males and used transgene-specific primers to identify a single positive clone of approximately 110 kb. This clone was sequenced and assembled into a total of 13 contigs. The sequence of all contigs was found to consist, in a large part, of short sequence repeats and sequence fragments matching known transposable elements. Using a cutoff of 100 aa as a minimum length, we selected 39 putative ORFs within these contigs to test for Y linkage. PCR performed on genomic DNA showed that 3 ORFs of the 39 were amplified from wild-type or T4 males but not from females (Table S1). We next used RNAseq data from adult males and females to identify novel transcribed regions (NTRs) mapping to the 13 contigs (Fig. S2). Of 20 identified NTRs, 12 show a male-specific expression pattern similar to recently identified Y-linked genes used as a control. We performed similarity searches with NTRs supported by putative

ORFs, and although some may present previously unidentified Y linked genes, others match to known Y-linked repetitive elements (Dataset S1).

The Y-Linked Transgene Is Stable and Can Serve as a Marker for Automated Sexing. Recombination between the *A. gambiae* X and Y chromosomes has not been reported to occur; however, evidence is derived mainly from genetic experiments by using chromosomal translocations (23). The Y linkage of the transgene in line T4 allowed us to test experimentally whether exchange between these chromosomes occurs at a significant rate. These experiments also allowed us to test whether Y linkage of the T4 transgene was stable and whether recombination or remobilization of the transgene could break its linkage with the Y chromosome. A total of 16,750 T4 larvae were sorted in the Complex Object Parametric Analyzer and Sorter (COPAS) particle sorter according to their distinct intensities of green fluorescence, because the expression of GFP is expected to occur in male larvae only (Fig. 3*A*). The GFP-negative cloud corresponding to putative female larvae (6,415 individuals) was entered for a second run in the COPAS to check for the presence of contaminating GFP^+ larvae. No GFP-expressing male larvae were detected. The GFP-positive and -negative larvae were separately reared to adulthood. All GFP-positive adults emerged as males, whereas all GFP negatives were found to be females, thus indicating the utility of line T4 for automated sex separation. This COPAS sorting experiment was repeated multiple times at intervals of several months, encompassing at least 12 generations. Sorting was always perfectly accurate, indicating that the transgene is stable. No putative X/Y recombination events were detected in the T4 line, and we concluded that recombination was either very rare or that the T4 insertion site is located in a region of the Y chromosome that does not readily recombine with the X. Thus, automated sorting of the T4 line represents a fast and reliable method to generate pure populations of either transgenic T4 males or nontransgenic, virgin females.

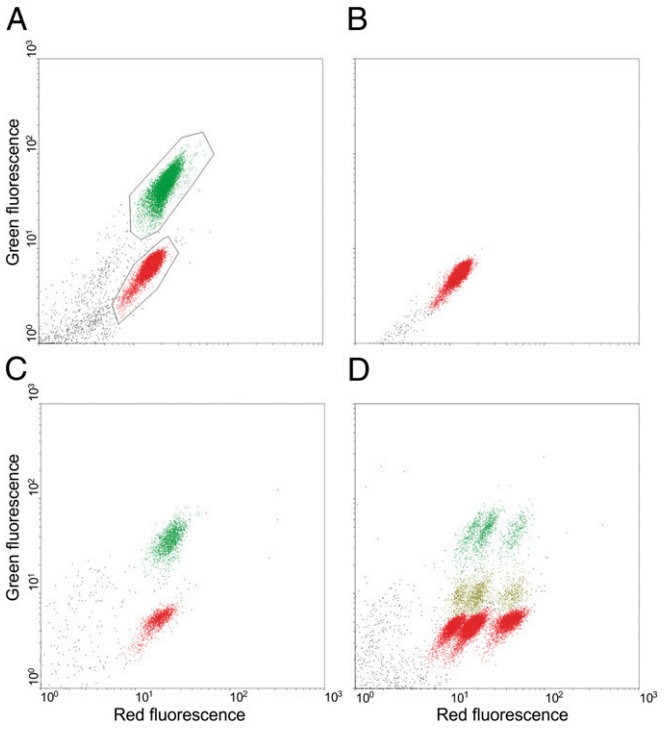


Fig. 3. Flow cytometry analysis and automated sex separation of the T4 strain. (*A*) A total of 16,750 larvae of the T4 line were analyzed according to their level of green and red fluorescence (green and red clouds) and sorted via the gates indicated (black lines). (*B*) Purified larvae (6,415 individuals) gated in *A* as having only background fluorescence were subjected to a second sorting run in the COPAS to check for the absence of contaminating GFP positive larvae. (*C*) Analysis of the fluorescence profile of ~2,500 F_1 larvae carrying the T4 transgene and expressing the I-SceI nuclease from which 1,246 GFP-positive larvae (upper compact cloud) and 902 GFP-negative larvae (lower compact cloud) were COPAS purified. (*D*) Analysis of approximately 9,100 F_2 larvae from an intercross of the F_1. Three classes of red fluorescent larvae are seen along the *x* axis, indicating normal segregation of the red-marked I-SceI transgene. Three classes of green fluorescent larvae are also seen along the *y* axis: high green fluorescence (6.3% of larvae), low green fluorescence (6.8% of larvae), and GFP-negative larvae (87%).

The Y Chromosome Is Accessible to Modification in the Male Germ Line. The presence of the unique I-SceI site within the integrated transformation construct opened the possibility to specifically modify the Y chromosome, if it was accessible to modification in the male germ line. To test for this hypothesis, we crossed T4 males to females of the transgenic line VFS1. This line transcribes the I-SceI homing endonuclease in the germ-line cells under the control of the germ line-specific *vasa* promoter and it also carries a 3xP3-RFP marker gene. As expected, no indication of I-SceI activity was found in the F_1 generation as COPAS sorted GFP-positive larvae and GFP-negative larvae gave rise to male and female adults (>500 individuals were analyzed for each sex), respectively (Fig. 3*C*). Transgenic male and female F_1 individuals were intercrossed, and the F_2 was analyzed by fluorescent sorting. Three classes of red fluorescent larvae are seen along the *x* axis, indicating normal segregation of the RFP-marked I-SceI transgene (Fig. 3*D*). In addition, we observed varying levels of green fluorescence in F_2 larvae on the *y* axis: High GFP fluorescence (seen in 6.3% of all larvae) corresponds to male larvae in which the GFP transgene had not been affected by I-SceI activity, or in which the I-SceI–caused DNA break was repaired in a way that preserved GFP integrity. Lower GFP fluorescence (6.8% of larvae) indicates male larvae in which I-SceI caused a DNA break within GFP followed by imprecise DNA repair that lowered GFP fluorescence. GFP-negative individuals (87%) are expected to include all female larvae, and male larvae in which I-SceI caused a mutation that completely abolished GFP activity. Of 1,134 adult mosquitoes arising from these GFP-negative larvae, 697 were females and 437 (38.5%) were males, indicating a high level of I-SceI activity. Judging by the proportion of male larvae that carried a detectable modification in their Y chromosome, the minimal rate of I-SceI activity can be estimated to be approximately 80%. However, this number is likely to be an underestimate, because some Y chromosomes may have been cut and repaired precisely, or repaired without causing a loss of GFP fluorescence. We concluded that the transgene construct on the Y chromosome was accessible and could be efficiently modified by I-SceI.

Generation of the Y-Linked φC31 Docking Line YAttP. We have shown that I-SceI cleavage can trigger homologous repair between homologous chromosomes and from plasmid repair templates (22). We designed such a "knock-in" strategy to place an AttP-specific integration site onto the Y chromosome. We generated the plasmid 3xP3[AttP]RFP (Fig. 1) by site-directed mutagenesis of the pHome-T construct. It carries the RFP (monomeric DsRed) marker under the control of the 3xP3 promoter. The plasmid was designed to contain, between the promoter and the RFP coding sequence, an AttP recombination site that replaces both the GFP coding sequence and the Actin5C promoter, which are present in pHome-T. The plasmid retains regions of homology to the T4 locus of 0.6 kb and 2.1 kb, 5′ and 3′ of the AttP site, respectively. The AttP recombination signal is recognized by the phage φC31 integrase and would allow for the subsequent site-specific integration of constructs carrying a corresponding AttB signal (24, 25). We hypothesized that I-SceI induced DNA double-strand breaks followed by homologous repair from 3xP3[AttP]RFP in the germ-line progenitor cells could

be observed as male larvae switching fluorescent reporters, i.e., showing red but no longer any green fluorescence. This approach allows for the identification of transgenics, and, because the AttP site is located between the 3xP3 promoter and DsRed sequences, would also allow the identification of subsequent φC31 integrations by the loss or conversion of red fluorescence. We coinjected 3xP3 [AttP]RFP into embryos of transgenic line T4 together with a source of I-SceI. We used either in vitro-transcribed I-SceI mRNA or a helper plasmid in which the I-SceI coding sequence is placed under the control of the *A. gambiae vasa* promoter. Table 1 shows the outcome of these experiments. We observed 11 3xP3-RFP–positive progeny of 6,160 larvae screened (0.18%) when using helper plasmid. No RFP-positive larvae were obtained from the progeny of males injected with I-SceI RNA. Genomic PCR and sequencing confirmed that cassette replacement had occurred in this transgenic line now termed YAttP (Fig. 2). This result shows that induced sequence-specific chromosomal breaks can be repaired by homologous recombination with a circular synthetic repair template and, therefore, that *A. gambiae* is amenable to knock-in gene engineering procedures.

Secondary Integration of a Germ Line-Specific Marker Gene by Site-Specific Recombination. To study the activity of germ line-specific regulatory elements on the Y and to demonstrate the utility of line YAttP as a tool for the site-specific genetic engineering of the Y chromosome, we generated the construct attBCFP-VasGFP (Fig. 1). In this vector, an AttB site was placed upstream of a promoterless eCFP coding sequence. In addition, it carries a cassette in which GFP is placed under the control of the *vasa* promoter (26). Recombination between AttP and AttB sites and integration of the plasmid was designed to replace the RFP CDS with the CFP ORF and could be detected by screening for the shift to blue fluorescence in the progeny. Table 1 shows the outcome of experiments in which we injected line YAttP with attBCFP-VasGFP and a helper plasmid expressing the φC31 integrase from a *vasa* promoter sequence. We injected a total of 111 embryos from which 5 male survivors hatched. Because the plasmid attBCFP-VasGFP does not allow the detection of transient fluorescence in G_0 individuals, all male survivors were outcrossed to wild-type females. Of 791 progeny screened, we obtained 6 CFP-positive larvae (0.76%). We confirmed the identity of the newly generated Y-linked strain YVasG by genomic PCR and sequencing.

The *vasa* Regulatory Region Drives GFP Expression During Early Male Spermatogenesis from the Y Chromosome. All transgenic G_2 larvae of strain YVasG showed a gonad-specific GFP signal (Fig. 2*A*) from the L1 larval stage onwards and developed into adult males as expected. We dissected testes from male adults of the YVasG strain and analyzed GFP fluorescence in the gonads by using confocal microscopy (Fig. 2*C*). GFP expression was detectable in all stages of spermatogenesis including the germ-line stem cells (GSCs) in the apical tip and developing spermatocytes but not in supporting somatic cells. DAPI staining of the hub region demonstrated that cells in the anterior tip of the testis were expressing GFP, thus indicating that the regulatory regions of *vasa* were active in GSCs when located on the Y chromosome. The pattern of GFP expression is identical to testes from line Vas2GFP that expresses *vasa*-driven GFP from position 2465559 of chromosome 3L. This finding suggests that strain YAttP in combination with the previously characterized non–sex-specific *vasa* promoter allows the male-exclusive expression of transgenes starting at the earliest stages of spermatogenesis.

Discussion

Although hundreds of transgenic *Anopheles* strains have been generated since the inception of transposase-mediated germ-line transformation, so far no strain harboring a construct inserted into the Y chromosome has been described, despite this chromosome representing an estimated 10% of the genome. In higher eukaryotes, gene expression and the accessibility of DNA is determined, to a large extent, by chromatin structure. These facts suggested that, because of its heterochromatic nature, the Y chromosome was refractory to the random integration catalyzed by transposases or, alternatively, that marker genes commonly used to identify transformation events would undergo complete silencing when located on the Y chromosome. Here, we characterized a Y-linked transgenic line obtained by random integration. Expression of fluorescent maker genes in two different tissues suggested that the transgene construct had inserted within a region of chromatin, allowing transcriptional activity. Such a euchromatic region is of interest because it might host male-specific genes. We identified a number of Y-specific PCR markers that may aid in characterizing *anopheline* population history and geographic structure, which is critical for understanding genetic structure and speciation and for the effective implementation of malaria control strategies (27, 28). The majority of selected putative ORFs analyzed by genomic PCR were also amplified from females, indicating that these sequences are likely to be present also on other chromosomes. We also identified a number of transcribed regions in the neighborhood of the integrated construct. Male-specific NTRs could represent putative Y-linked genes and lend themselves for future functional

Table 1. Embryo microinjections and fluorescent phenotypes observed in the outcross progeny

Source	Embryos injected	Transient males/male survivors outcrossed	Larvae screened	Fluorescent phenotypes observed	
I-SceI					
RNA	~2,000	6	964	Nonfluorescent	473 (♀)
				Actin5C-RFP/3xP3-GFP	461 (♂)
				Actin5C-RFP	30 (♂)
				3xP3-RFP	**0**
Plasmid	~4,000	15	6,160	Nonfluorescent	3,113 (♀)
				Actin5C-RFP/3xP3-GFP	2,820 (♂)
				Actin5C-RFP	216 (♂)
				3xP3-RFP	**11** (♂)
Integrase					
Plasmid	111	5	791	Nonfluorescent	390 (♀)
				3xP3-RFP	395 (♂)
				3xP3-CFP/Vas-GFP	**6** (♂)

T4 males injected with 3xP3[AttP]RFP plasmid and either in-vitro transcribed I-SceI RNA or the pVas2-I-SceI helper plasmid were outcrossed to wild-type females. The resulting progeny was screened for fluorescence. YAttP males injected with attBCFP-VasaGFP plasmid and the integrase helper plasmid were outcrossed to wild-type females. The resulting progeny was screened for fluorescence. Phenotypes indicating successful gene conversion or integration are shown in bold.

analysis. Many of the ORFs selected for genomic PCR and those underlying NTRs largely match known repetitive elements and transposons. Interestingly, putative paralogues of a number of the identified transcribed regions map to scaffold AAAB01008885.1 close to the centromere of the X chromosome (Dataset S1).

The fact that the T4 locus was transcriptionally active made it an exciting target for gene engineering. We used line T4 in combination with a plasmid template to stimulate homologous recombination and introduce a site-specific recombination signal onto the Y chromosome. Our approach relied on the well characterized endonuclease I-SceI and the presence of an I-SceI site within the GFP ORF of the transgene. The availability of novel programmable endonucleases based on the transcription activator-like effector nuclease (TALEN) or clustered regularly interspaced short palindromic repeats (CRISPR/Cas) architectures as well as a growing number of sequences uniquely present on the Y chromosome such as those described in this study suggests that this powerful approach could be extended to introduce foreign sequences (such as docking sites or desired mutations) also into other regions of the Y or, indeed, into any locus of interest on other chromosomes. Although homologous recombination was inefficient (occurring in 0.18% of the progeny), it helped to successfully introduce the AttP recombination site, paving the way for the more efficient site-specific integration method. We found that secondary integration mediated by the φC31 integrase occurred at least at a 4× higher rate.

Although the use of COPAS automated fluorescent larval sorting is limited to research, it has been suggested that future iterations of this technology could be scaled up to develop robust protocols for the mass production of sorted insects. The strains carrying Y-linked fluorescent markers we have developed are, as we show, ideally suited for such automated fluorescence sexing. This property may become useful in applications requiring large numbers of pure males, such as SIT (29), and in experiments that require genetic crosses to a large number of virgin wild-type females, readily obtained here from the T4 line purified at the early larval stage. The fact that transgenic males give rise to male progeny that all inherit the transgene represents an additional advantage of these strains because, unlike with autosomal markers for sexing (30), no segregation of the sexing marker occurs so that a pure-breeding stock can be maintained. This feature allows combining the sexing trait with other autosomally encoded traits, for example male sterility. Alternatively, genetic sexing traits such as insecticide resistance markers can now also be easily introduced to the Y chromosome. The ability to introduce transgenes to the Y chromosome also allows the generation of male-exclusive genetic traits. We have shown that the regulatory regions of the *Anopheles* gene *vasa* drive expression of transgenes in the male and female germ line throughout larval and adult gonad development (26). We then attempted to generate a male-specific version of the *vasa* promoter by using a shortened 5′ UTR. However, this approach not only reduced overall activity of the promoter, but also lead to the loss of expression in the GSCs of the testes. As demonstrated here, the ability to express transgenes from the Y chromosome allows for the generation of male-exclusive traits by using non–sex-specific components and to expressed tagged or mutant forms of Y-linked genes.

We have recently generated a synthetic male sex ratio distortion system that operates by destroying the X chromosome during male spermatogenesis and creates progeny consisting almost exclusively of males (Galizi et al., under review). The induction of extreme male-biased sex ratios could be a powerful strategy to supress or eliminate pest populations. However, transgenes when expressed from autosomal locations are transmitted to only half of the male progeny. In its current form, this technology would require continuous mass releases to achieve population suppression. In contrast, if the distorter trait could be successfully linked to the Y chromosome, this chromosome would be converted into a selfish sex chromosome that, in theory, could spread through and eliminate a natural vector population even when seeded by a very low number of transgenic males (31). Alternative uses of Y-linked meiotic drive have also been described (32). The YattP line described here paves the way to establishing such vector control strategies and, thus, represents a promising tool in the fight against malaria.

Materials and Methods

Mosquito Rearing. The wild-type *A. gambiae* strain (G_3) and the transgenic mosquito line YAttP and YVasG were reared under standard condition at 28 °C and 80% relative humidity with access to fish food as larvae and 5% (wt/vol) glucose solution as adults. For egg production, young adult mosquitoes (3–5 d after emergence) were allowed to mate for at least 6 d and then fed on mice. Three days later, an egg bowl containing rearing water (dH_2O supplemented with 0.1% pure salt) was placed in the cage. One to two days after hatching, the larvae (L1 stage) were placed into rearing water containing trays. The protocols and procedures used in the study were approved by the Animal Ethics Committee of Imperial College in compliance with UK Home Office regulations.

Analysis of Y-Specific BAC Clone. High molecular weight DNA was extracted from transgenic T4 individuals. A BAC library was constructed in vector pIndigoBAC5-HindIII from Epicentre by using the standard HindIII cloning site. The library was pooled in 1 × 96-well plate with each well containing approximately 125 independent primary clones with an average insert of 120 kb. Primers actin5rev (5′TGGAAATGAGAAGTAAGGTGCATCTGCA3′) and exfp3sEq. (5′ATCCGCCACAACATCGAGGACG3′) were used to screen the clones for the presence of a Y-specific insert. Size determination of the insert was performed by NotI digestion and PFGE gel separation. Single Molecule Real Time (SMRT) technology was chosen to sequence one positive BAC clone, and PacBio SMRT Portal assembler was used for sequence quality filtering and assembly. The contigs were uploaded to GenBank (accession nos. KJ608148–KJ608160).

RNAseq Analysis. Illumina paired-end 101-bp reads derived from libraries of whole adult males and females (unfed) were aligned against the assembled 13 contigs by using STAR (33). To control for reads that map to repetitive regions, only reads that mapped uniquely within all 13 contigs were considered. We used Cufflinks (34) to predict transcripts, and the aligned reads were used to generate updated consensus sequences of the contigs, which generally increased ORF size. Sequences encompassing transcribed regions that displayed male-specific expression were BLAT searched (BLAST-like alignment tool searched; http://genome.cshlp.org/content/12/4/656) against the *A. gambiae* PEST assembly (AgamP3) to identify whether these may already be present in the genome assembly, or to identify putative paralogues on other chromosomes. As a control, we mapped male and female reads against sequences of two recently described Y-linked genes, gYG1 and gYG2 (20), and the α-tubulin (AGAP001219) gene expressed ubiquitously in both sexes.

Generation of I-SceI RNA. The mMESSAGEmachine kit (Ambion) was used to obtain a 5′capped mRNA (I-SceI RNA) coding for the homing endonuclease I-SceI. The *I-SceI* gene was amplified from the target plasmid pP[v+,70I-SceI] by using primers ForI-SceI 5′CGCGTAATACGACTCACTATAGGGCGAATTGGGTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCATGGGATCATCATCAGACGACGAAGCTACGGCCGATGCGCA3′ containing a T7 RNA polymerase promoter and RevI-SceI (5′TTTTTTTTTTTTTTTTTTTTTTTATTTCAGGAAAGTTTCGGAGGAGATAGTGTTCGGCAGTTTGTACA3′).

Generation of Transformation Vectors. The Phusion Site-Directed Mutagenesis Kit was used to generate the 3xP3[AttP]RFP plasmid. For this cloning strategy, we used primers, F1 (5′-CTTTGAGTTCTCTCAGTTGGGGGCACGCAACATGGTGCGCTCCTCCAAGAACG-3′) and R2 (5′-GTTACCCCAGTTGGGGCGTAGGGTCTAGTCGACTCTAGCGGTACCCCGATTG-3′), containing one-half attP site each at the 5′. The plasmid attBCFP-VasaGFP was generated by cloning of a blunt synthetic fragment, attB-fragment, containing eCFP preceded by an attB site and followed by the SV40 terminator, into the StuI site of pfVasIntGFP carrying the VasGFP (26).

Embryo Microinjections. Embryos were injected by using a Femtojet Express injector and a Narishige 202ND micromanipulator mounted on an inverted microscope (Nikon TE-DH100W) with a mixture of 0.2 μg/μL 3xP3[AttP]RFP plasmid or 0.8 μg/μL in vitro-transcribed I-SceI RNA, and 0.4 μg/μL pVas2-I-SceI

helper plasmid for the generation of the transgenic mosquito line YAttP. The hatched larvae were screened for transient expression of the DsRed marker, and the 0.18% positive was grown up and crossed to wild-type mosquitoes. The progeny of these crosses was analyzed for DsRed fluorescence. Individual larvae showing expression of the selectable marker were then separated, and the adults that emerged were crossed individually with wild-type mosquitoes to obtain transgenic lines. Embryos were injected with a mixture of 0.2 μg/μL attBCFP-VasaGFP plasmid and 0.5 μg/μL Integrase helper plasmid to generate the transgenic mosquito line YVasG. The hatched larvae were grown up, and the surviving males were crossed to wild-type mosquitoes. The progeny of these crosses was analyzed for eCFP fluorescence. Individual larvae showing expression of the selectable marker were then separated, and the adults that emerged were crossed individually with wild-type mosquitoes to obtain transgenic lines. Transgenic mosquitoes at different developmental stages were analyzed on a Nikon inverted microscope (Eclipse TE200) at a wavelength of 488 nm to detect eGFP expression (Filter 535/20 nm emission, 505 nm dichroic) and 563 nm (Filter 630/30 nm emission, 595 nm dichroic) to detect DsRed expression.

COPAS-Assisted Analysis and Sorting of Larval Populations. COPAS sorting was performed as described in Marois et al. (30). Briefly, newly hatched larvae were transferred to the reservoir of the large-particles flow cytometry COPAS SELECT instrument (Union Biometrica) equipped with a multiline argon laser (488, 514 nm) and a diode laser (670 nm), analyzed, and sorted with the Biosort5281 software with a 488-nm filter and the following acquisition parameters: Green PMT 500, Red PMT 600, Delay 8; Width 6, pure mode selection with superdrops. The flow rate was kept below 15 detected objects per second. Larvae identified as male and female were dispensed sequentially in separate Petri dishes. The LMD files generated by the COPAS software were imported in the WinMDI freeware for data analysis and figure preparation.

Confocal Microscopy. Dissected testes from transgenic lines YVasG and Vas2GFP were fixed in methanol-free 4% formaldehyde (Pierce) in PBS for 30 min and washed three times for 15 min in 0.1% Tween-20 PBS. Testes were then transferred on fresh slides containing Vectashield mounting medium with DAPI (Vectorlabs) with coverslips. Testes images were taken by using a Zeiss LSM 510 Laser scanning confocal microscope and a 20× objective in two cannels (DAPI and GFP) sequentially.

FISH. FISH of the sexual chromosomes, X and Y, in the transgenic line T4 was performed according to a previously established protocol (35). The X chromosome was marked by using as a probe an intergenic spacer region of rDNA generated by PCR using primers UN (GTGTGCCCCTTCCTCGATGT) and GA (CTGGTTTGGTCGGCACGTTT) and labeled with the cyanine dye Cy5. The Y chromosome was marked by using as a probe the p-Home T plasmid Cy3-labeled by nick translation (Roche Nick Translation Kit). Chromosomes where counterstained with DAPI.

ACKNOWLEDGMENTS. We thank Cecilia Cheung and Stephen M. Rothery from the Facility for Imaging by Light Microscopy, Tony Nolan, and Maria Sharakhova. This work is funded by a grant from the Foundation for the National Institutes of Health through the Vector-Based Control of Transmission: Discovery Research program of the Grand Challenges in Global Health initiative of the Bill & Melinda Gates Foundation and the European Research Council under European Union's Seventh Framework Programme ERC Grant 335724 (to N.W.).

1. World Health Organization (2012) World Malaria Report 2012. (World Health Organization, Geneva).
2. Clements AN (1992) *The Biology of Mosquitoes: Development, Nutrition, and Reproduction* (Chapman & Hall, London).
3. Baker RH, Sakai RK (1979) Triploids and male determination in the mosquito, Anopheles culicifacies. *J Hered* 70(5):345–346.
4. Charlesworth B (1996) The evolution of chromosomal sex determination and dosage compensation. *Curr Biol* 6(2):149–162.
5. Muller HJ (1932) Some genetic aspects of sex. *Am Nat* 66(703):118–138.
6. Lahn BT, Pearson NM, Jegalian K (2001) The human Y chromosome, in the light of evolution. *Nat Rev Genet* 2(3):207–216.
7. Fisher RA (1931) The evolution of dominance. *Biol Rev Camb Philos Soc* 6(4):345–368.
8. Rice WR (1984) Sex chromosomes and the evolution of sexual dimorphism. *Evolution* 38(4):735–742.
9. Steinemann M, Steinemann S (2000) Common mechanisms of Y chromosome evolution. *Genetica* 109(1-2):105–111.
10. Burgoyne PS (1998) The mammalian Y chromosome: A new perspective. *Bioessays* 20(5):363–366.
11. Castillo ER, Bidau CJ, Martí DA (2010) Neo-sex chromosome diversity in Neotropical melanopline grasshoppers (Melanoplinae, Acrididae). *Genetica* 138(7):775–786.
12. Rice WR (1994) Degeneration of a nonrecombining chromosome. *Science* 263(5144): 230–232.
13. Bachtrog D (2003) Adaptation shapes patterns of genome evolution on sexual and asexual chromosomes in Drosophila. *Nat Genet* 34(2):215–219.
14. Carvalho AB, Lazzaro BP, Clark AG (2000) Y chromosomal fertility factors kl-2 and kl-3 of Drosophila melanogaster encode dynein heavy chain polypeptides. *Proc Natl Acad Sci USA* 97(24):13239–13244.
15. Carvalho AB, Dobo BA, Vibranovski MD, Clark AG (2001) Identification of five new genes on the Y chromosome of Drosophila melanogaster. *Proc Natl Acad Sci USA* 98(23):13225–13230.
16. Steinemann M, Steinemann S (1998) Enigma of Y chromosome degeneration: Neo-Y and neo-X chromosomes of Drosophila miranda a model for sex chromosome evolution. *Genetica* 102–103(1-6):409–420.
17. Maggert KA, Golic KG (2002) The Y chromosome of Drosophila melanogaster exhibits chromosome-wide imprinting. *Genetics* 162(3):1245–1258.
18. Holt RA, et al. (2002) The genome sequence of the malaria mosquito Anopheles gambiae. *Science* 298(5591):129–149.
19. Mason GF (1967) Genetic studies on mutations in species A and B of the Anopheles gambiae complex. *Genet Res* 10(3):205–217.
20. Hall AB, et al. (2013) Six novel Y chromosome genes in Anopheles mosquitoes discovered by independently sequencing males and females. *BMC Genomics* 14:273.
21. Criscione F, Qi Y, Saunders R, Hall B, Tu Z (2013) A unique Y gene in the Asian malaria mosquito Anopheles stephensi encodes a small lysine-rich protein and is transcribed at the onset of embryonic development. *Insect Mol Biol* 22(4):433–441.
22. Windbichler N, et al. (2011) A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature* 473(7346):212–215.
23. Krzywinski J, Nusskern DR, Kern MK, Besansky NJ (2004) Isolation and characterization of Y chromosome sequences from the African malaria mosquito Anopheles gambiae. *Genetics* 166(3):1291–1302.
24. Thorpe HM, Smith MC (1998) In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc Natl Acad Sci USA* 95(10):5505–5510.
25. Meredith JM, et al. (2011) Site-specific integration and expression of an anti-malarial gene in transgenic Anopheles gambiae significantly reduces Plasmodium infections. *PLoS ONE* 6(1):e14587.
26. Papathanos PA, Windbichler N, Menichelli M, Burt A, Crisanti A (2009) The vasa regulatory region mediates germline expression and maternal transmission of proteins in the malaria mosquito Anopheles gambiae: A versatile tool for genetic control strategies. *BMC Mol Biol* 10:65.
27. Hammer MF, Zegura SL (1996) The role of the Y chromosome in human evolutionary studies. Evolutionary Anthropology: Issues. *News Rev (Melb)* 5(4):116–134.
28. Hurles ME, Jobling MA (2001) Haploid chromosomes in molecular ecology: Lessons from the human Y. *Mol Ecol* 10(7):1599–1613.
29. Condon KC, et al. (2007) Genetic sexing through the use of Y-linked transgenes. *Insect Biochem Mol Biol* 37(11):1168–1176.
30. Marois E, et al. (2012) High-throughput sorting of mosquito larvae for laboratory studies and for future vector control interventions. *Malar J* 11(1):302.
31. Hamilton WD (1967) Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science* 156(3774): 477–488.
32. Huang Y, Magori K, Lloyd AL, Gould F (2007) Introducing desirable transgenes into insect populations using Y-linked meiotic drive - a theoretical assessment. *Evolution* 61(4):717–726.
33. Dobin A, et al. (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1): 15–21.
34. Trapnell C, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28(5):511–515.
35. Timoshevskiy VA, Sharma A, Sharakhov IV, Sharakhova MV (2012) Fluorescent in situ hybridization on mitotic chromosomes of mosquitoes. *J Vis Exp* (67):e4215.

nature COMMUNICATIONS

ARTICLE

Received 12 Mar 2014 | Accepted 28 Apr 2014 | Published 10 Jun 2014

DOI: 10.1038/ncomms4977 OPEN

A synthetic sex ratio distortion system for the control of the human malaria mosquito

Roberto Galizi[1,2], Lindsey A. Doyle[3], Miriam Menichelli[1], Federica Bernardini[1], Anne Deredec[1], Austin Burt[1], Barry L. Stoddard[3], Nikolai Windbichler[1,*] & Andrea Crisanti[1,2,*]

It has been theorized that inducing extreme reproductive sex ratios could be a method to suppress or eliminate pest populations. Limited knowledge about the genetic makeup and mode of action of naturally occurring sex distorters and the prevalence of co-evolving suppressors has hampered their use for control. Here we generate a synthetic sex distortion system by exploiting the specificity of the homing endonuclease I-PpoI, which is able to selectively cleave ribosomal gene sequences of the malaria vector *Anopheles gambiae* that are located exclusively on the mosquito's X chromosome. We combine structure-based protein engineering and molecular genetics to restrict the activity of the potentially toxic endonuclease to spermatogenesis. Shredding of the paternal X chromosome prevents it from being transmitted to the next generation, resulting in fully fertile mosquito strains that produce >95% male offspring. We demonstrate that distorter male mosquitoes can efficiently suppress caged wild-type mosquito populations, providing the foundation for a new class of genetic vector control strategies.

[1] Department of Life Sciences, Imperial College London, South Kensington Campus, London SW7 2AZ, UK. [2] Centro di Genomica Funzionale, University of Perugia, Dipartimento di Medicina Sperimentale Via Gambuli, Edificio D, 3° Piano, 06132 Perugia, Italy. [3] Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to N.W. (email: n.windbichler@imperial.ac.uk) or to A.C. (email: arcs@imperial.ac.uk).

In most pest species, males contribute few material resources to the next generation, and it is the number and productivity of females that determines future population size. The induction of extreme male-biased reproductive sex ratios has thus been suggested as an attractive method to suppress or eliminate pest populations[1,2]. For this reason, the identification of naturally occurring sex ratio distorters found in two medically important genera of mosquitoes, *Aedes* and *Culex*, has generated substantial interest[3,4]. However, attempts to assess the suitability of such natural distorter strains for controlling caged mosquito populations revealed that the degree of distortion was insufficient to achieve population suppression[5]. Resistance in female mosquitoes was often swiftly selected for, indicating the prevalence of natural resistance alleles. Although a detailed molecular understanding of these distorters has remained elusive, cytological observations indicated that a bias towards male gamete production in *Aedes* is associated with preferential breakage of the X chromosome during male meiosis, thus suggesting one possible mechanism of action[6]. We therefore hypothesized that the selective induction of DNA double-stranded breaks in the X chromosome during male meiosis might be a potential strategy to distort the sex ratio of the progeny towards male mosquitoes, if the normal production of Y-bearing gametes was unaffected (Fig. 1a).

Anopheles gambiae, the main vector of human malaria represents an ideal experimental model to test this hypothesis because of the availability of the highly specific homing endonuclease I-PpoI[7], which cuts a conserved sequence within the ribosomal rDNA repeats located in a single cluster on the mosquito's X chromosome[8]. The expression of wild-type I-PpoI during spermatogenesis in transgenic mosquitoes causes cleavage of the paternal X chromosome but also results in complete male sterility because the protein's stability and persistence in mature sperm cells leads to subsequent cleavage of the maternal X chromosome in the zygote[9]. However, these preliminary experiments also indicated that the majority of developmentally arrested embryos had been fertilized by Y-bearing spermatozoa, providing a rationale for attempting to uncouple chromosomal transmission distortion from male sterility[9].

In this study, we systematically destabilize the I-PpoI endonuclease with the objective to reduce its *in vivo* half-life and thereby restrict its activity to male meiosis, which occurs earlier than the formation of the zygote after female fertilization. Our objective is population suppression via the development of transgenic male mosquitoes that are not sterile and produce mainly Y chromosome-bearing sperm (and hence will produce only male progeny when mated with wild-type female mosquitoes). Male transgenic progeny fathered by these transgenic male mosquitoes also produce Y-biased sperm so that the effect can be self-sustaining.

Results

Generation of destabilized I-PpoI variants. We have generated a series of I-PpoI variants that displays a range of thermal unfolding temperatures spanning about 20 degrees, and a corresponding range of half-lives that differ by as much as 30-fold. To accomplish this, we used the crystal structure of the enzyme bound to its DNA target site to guide the necessary alterations of the protein's sequence and structure[7]. We selected residues for site-directed mutagenesis that were involved either (i) in the coordination of structural zinc ions in the protein core, (ii) in hydrophobic packing in the protein core or (iii) in subunit interactions within the protein dimer interface (Fig. 1b). Of the 11 residues that were subjected to point mutagenesis, alteration of each of the eight individual residues involved in zinc coordination resulted in a dramatic reduction in expression and recovery of recombinant protein, presumably because of disrupted folding. In contrast, mutation of two residues in the hydrophobic core (H106 and L111) and one residue in the dimer interface (W124) yielded soluble proteins that were readily purified using the same protocol used for the wild-type protein (as did simultaneous mutation of L111 and W124 to alanine and leucine, respectively). The thermal denaturation temperatures ('Tm') of the wild-type and mutant I-PpoI proteins were then determined by conducting temperature-dependent protein denaturation experiments using circular dichroism (CD) spectroscopy on purified protein samples (Supplementary Fig. 1a). Under these experimental conditions the Tm of the wild-type protein was determined to be 54.4 °C (± 0.2 °C). Mutations of the residues in the hydrophobic core and/or dimer interface all reduced the thermostability of the corresponding recombinant proteins with Tm values ranging from 49.4 °C (L111A; $\Delta Tm = -5°$) to 35.1 °C (L111A/W124L; $\Delta Tm = -19.3°$) while maintaining cooperative unfolding transitions for each endonuclease variant.

***In vitro* characterization of I-PpoI variants**. We next performed kinetic unfolding experiments in which purified enzymes were placed at constant temperatures (30 or 37 °C) while monitoring the gradual unfolding of the protein by measuring the CD signal (sensitive to a protein's secondary structure) at 206 nm for 4 h (Supplementary Fig. 1b,c). The thermal half-life of each construct at 37° ranged from 73.5 h (wild-type enzyme) to 2 h (H106A; the L111A/W124L double mutant displayed a half-life of ~4 min). At 30 °C, none of the constructs exhibited significant thermal unfolding over 4 h (except for the double mutant, which displayed a half-life of ~35–40 min). The specific activity of the wild-type and mutant enzymes was then determined using freshly prepared enzyme under assay conditions (a DNA digest performed at 22 °C for 5 min) where unfolding of each enzyme variant would not contribute to reduced catalytic activity (Supplementary Fig. 1d,e). With the exception of the double mutant, all of the constructs fell within a threefold activity range (from a value of ~7 pmol min^{-1} μg^{-1} for the wild-type enzyme to 2.6 pmol min^{-1} μg^{-1} for the H106A construct). The properties of all mutants are summarized in Table 1.

Generation and characterization of transgenic lines. We then designed germline transformation constructs to express each of the I-PpoI structural variants showing reduced stability in transgenic mosquitoes using the male spermatogenesis-specific β2 tubulin promoter[10] (Fig. 2a). The constructs were designed to express both I-PpoI and enhanced green fluorescent protein (eGFP) as in frame fusion proteins or, using a 2A ribosomal stuttering signal, as distinct protein chains. We generated at least two independent transgenic lines for each I-PpoI variant. We determined the genomic location (Supplementary Table 1) and I-PpoI mRNA expression levels (Supplementary Fig. 2) in all transgenic lines that showed a single chromosomal integration of the transgene. Constructs that had integrated on the X chromosome failed to show significant levels of I-PpoI expression in the testes, in agreement with the notion that this chromosome undergoes general transcriptional silencing during male meiosis[11]. The mRNA abundance of strains carrying autosomally integrated constructs ranged from twofold higher to 35-fold lower compared with mosquitoes expressing wild-type I-PpoI, thus offering the possibility to assess different combinations of expression levels and thermostability. We analysed protein levels of recombinant I-PpoI in testes with

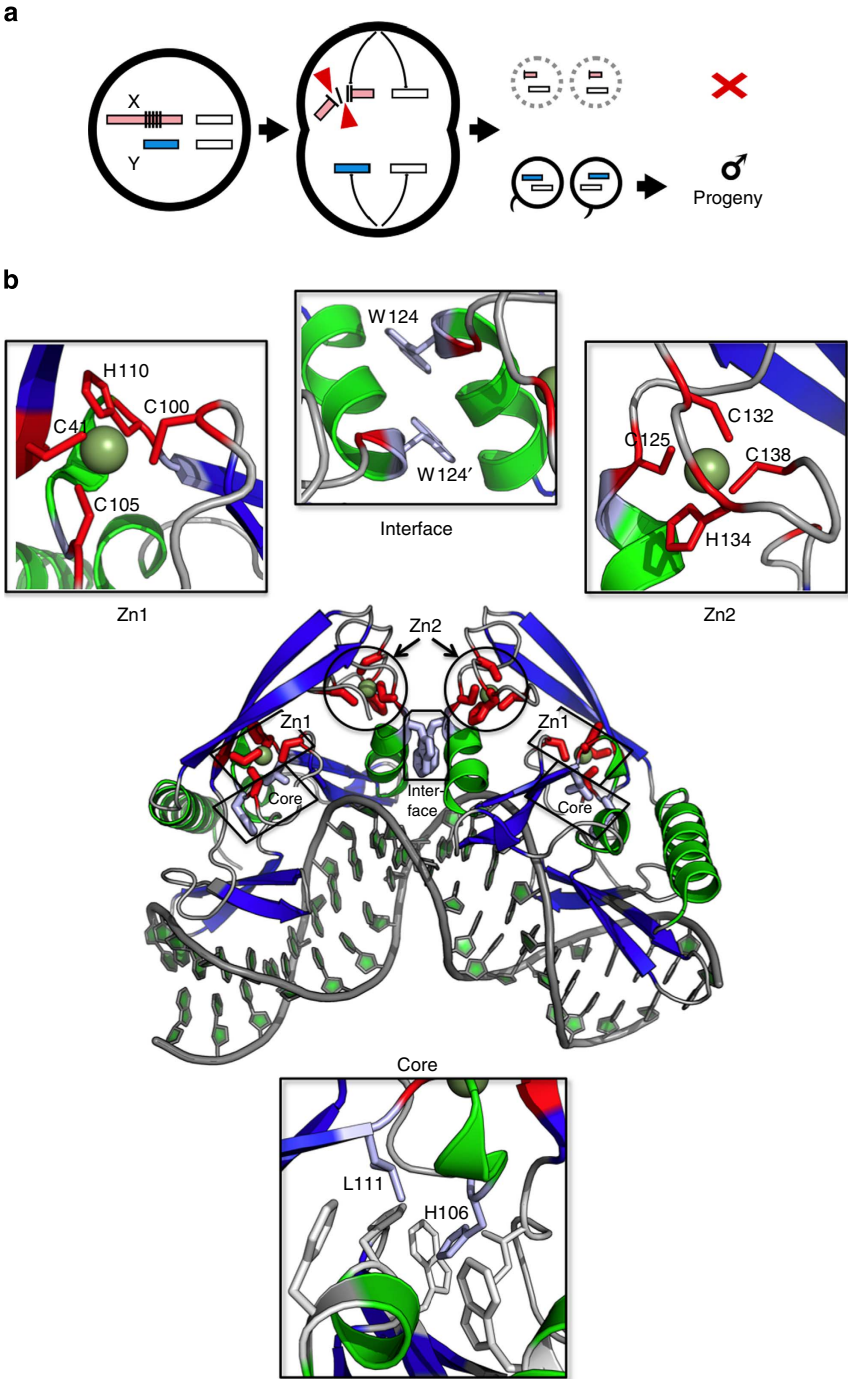


Figure 1 | Generation of a synthetic sex distortion system. (**a**) Proposed model for the distortion of the reproductive sex ratio towards males based on meiotic X shredding. A multicopy target sequence (black bars) on the X chromosome (but not present on the Y chromosome) is targeted by an endonuclease (red triangle) during spermatogenesis. Shredding of the X chromosome favours the unaffected Y-bearing sperm and results in the production of a male-biased progeny. (**b**) Structure of I-PpoI a homodimeric, 'His-Cys box' homing endonuclease that recognizes and cleaves a highly conserved rDNA sequence. Boxes: a total of 11 residues were mutated in this study. Eight of those positions correspond to residues that coordinate two zinc ions in each subunit ('Zn1' and 'Zn2'). One position contributes two symmetry-related tryptophan side chains (W124 and W124′) that form a significant fraction of the dimer interface; the final two positions (L111 and H106) are found in small hydrophobic core regions in each subunit.

western blot experiments using, for each variant, the strain with the highest level of mRNA expression. We observed that a reduced *in vitro* thermostability of the I-PpoI mutants translated into significantly reduced protein levels of the eGFP::I-PpoI protein variants *in vivo* consistent with lower eGFP fluorescence signal observed in the testes of these lines (Supplementary Fig. 3).

Destabilized I-PpoI distorts the sex ratio towards males. To investigate whether the reduction of I-PpoI stability resulted in a dissociation of gamete sex ratio distortion from embryonic toxicity, we determined the larval hatching rate and the number of eggs laid as measures of fertility as well as the adult sex ratio in crosses of transgenic male mosquitoes carrying I-PpoI variants to wild-type female ones (Supplementary Table 2). As a control,

transgenic female mosquitoes of each strain were crossed to wild-type male ones. Male mosquitoes of the strain expressing wild-type I-PpoI were sterile as previously observed for the strains ^{gfp}I-PpoI-1 and ^{gfp}I-PpoI-2 in which I-PpoI was expressed as an eGFP fusion protein[9]. No effect on fertility or the sex ratio was observed in all strains expressing the I-PpoI double-mutant L111A/W124L, which displays the lowest *in vitro* folded stability (Fig. 2b,c) or in strains carrying X-linked transgenes (Supplementary Fig. 4). Significant male-biased sex ratios ranging from 70.2 to 97.4% were observed in the progeny of male mosquitoes carrying the remaining I-PpoI variants (Fig. 2b). Some of the strains showed a significant reduction in the number of hatching progeny. Strains carrying I-PpoI variants H106A and W124L suffered a lower impact on their hatching rates than did those carrying the more stable variants L111A and W124A. With the possible exception of variant H106A egg laying rates were generally unaffected and showed little correlation with mRNA expression levels or the I-PpoI variants tested (Supplementary Fig. 5).

Notably, strains ^{gfp}111A-2, ^{gfp}124L-2 and ^{gfp}124L-3 combined a high level of male-biased sex distortion (95.0%, 95.5% and 97.4%, respectively) with a fertility rate that did not differ significantly from the controls. Interestingly, these strains also showed the lowest mRNA levels within their respective sets. We found that, while low rearing temperatures could cause a modest reduction in fertility in ^{gfp}111A-2 male mosquitoes, the sex distortion phenotype remained unaffected at both low (24 °C) and high (32 °C) temperatures (Supplementary Table 3) in all three strains. In addition, we found that the sex distortion phenotype was stably inherited from male mosquitoes to their transgenic sons. For four subsequent generations, transgenic male mosquitoes of ^{gfp}111A-2 fathers showed comparable levels of fertility and male-biased sex ratios in their offsprings (Supplementary Table 4). Finally, we investigated the phenotype of homozygous distorter males since such males would transmit the distorter to all their male progeny resulting in a stronger suppressive effect on target populations. Whereas the addition of an extra copy of the transgene construct in strains ^{gfp}111A-2 and ^{gfp}124L-3 caused a reduction in the hatching rate, it had no effect on the fertility of ^{gfp}124L-2, which showed hemizygous mRNA levels less than half of ^{gfp}124L-3 (Supplementary Table 5).

Table 1 | Properties of wild-type and mutant I-PpoI variants determined *in vitro*.

I-PpoI variant	Tm (°C)	Half-life (h)	Specific activity (pmol min^{-1}mg^{-1})
Wild-type	54.4	73.5	7.02 (±1.1, $N=3$)
L111A	49.4	16.9	6.1 (±0.6, $N=3$)
W124A	47.2	29.6	3.4 (±0.4, $N=3$)
W124L	45.2	9.4	2.8 (±0.8, $N=3$)
H106A	42.6	1.96	2.6 (±0.4, $N=3$)
L111A+W124L	35.1	0.067 (4 min)	0.6 (±0.1, $N=3$)

Analysis of individuals with modified X chromosomes. We analysed at the molecular and phenotypic levels the small fraction of female offspring of transgenic male mosquitoes and found molecular evidence for both misrepair and copy number variation of the ribosomal gene cluster in these individuals (Supplementary Fig. 6a,b). We also observed a significant female-biased sex ratio in the progeny of female survivors crossed to wild-type males suggesting that loss of viability occurs in individuals that had inherited only a damaged X chromosome (Supplementary Table 6). We also investigated whether the X chromosomes of daughters from transgenic males were intrinsically resistant to I-PpoI activity. For this purpose, we utilised daughters of ^{gfp}111A-2 males, carrying paternal 'modified' X chromosomes that had been exposed to I-PpoI activity. Females that had also inherited the transgene were crossed to wild-type males. The transgenic male progeny, having a 50% chance of carrying the

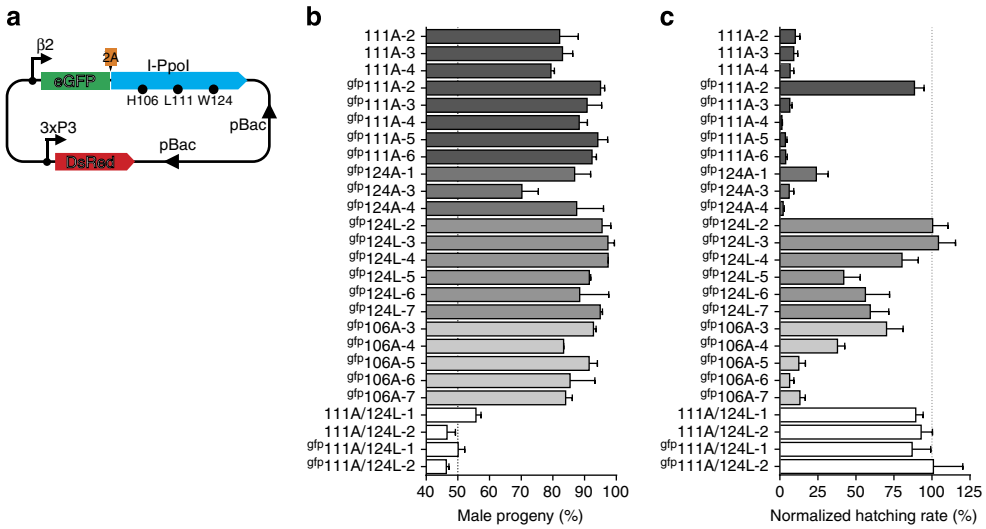


Figure 2 | Engineered I-PpoI variants distort the sex ratio towards males. (**a**) Germline transformation construct showing mutated I-PpoI residues. (pBac) piggyBac inverted repeats; (3xP3) Pax promoter driving the DsRed marker gene (DsRed); I-PpoI and GFP reading frames under the control of β2 tubulin promoter (β2); (2A) self-cleaving peptide used in some constructs. (**b**) The adult sex ratio (error bars represent the s.e.m.) of the progeny of hemizygous transgenic males and wild-type females of all transgenic lines with single copy autosomal integrations. Experiments were performed in at least two independent generations. The total numbers of sexed individuals are summarized in Supplementary Table 2. (**c**) The larval hatching rate (error bars represent s.e.m.) of the progeny of hemizygous transgenic males and wild-type females of all transgenic lines with single copy autosomal integrations normalized against the hatching rate of hemizygous transgenic females and wild-type males of the same lines. Experiments were performed in at least two independent generations. The total number of females analysed is summarized in Supplementary Table 2.

potentially resistant X chromosome, was then crossed to wild-type females and the sex ratio of individual crosses was recorded. This analysis showed a male bias in the progeny of all males that did not differ significantly from the ^{gfp}111A-2 stock (Supplementary Fig. 6d). These experiments suggested that X chromosomes of female survivors are damaged as a result of I-PpoI activity yet remaining susceptible to further cleavage. A stochastic process most likely accounts for the fact that some damaged X chromosomes allow the constitution of functional gametes, resulting in female progeny.

Population cage experiments. Although the robustness of these traits under variable natural conditions remains to be studied, our findings have important implications for the development of genetic vector control measures. To demonstrate this we performed a set of population cage studies to assess whether the I-PpoI sex distorter strains could suppress wild-type *A. gambiae* caged populations. We observed in five independent cage experiments that the release of hemizygous ^{gfp}111A-2 males at an over-flooding ratio of 3X was effective in suppressing caged wild-type populations (achieving elimination within six generations in four out of five cages). As expected, the release of sterile 'control' males expressing wild-type I-PpoI at the same ratio had no measurable effect on three control populations (Fig. 3) under these experimental conditions. We observed hatching rates averaging below 20% in the terminal generations of these caged populations, indicating that under field conditions the accumulation of X chromosome damage would significantly contribute to the demise of target populations.

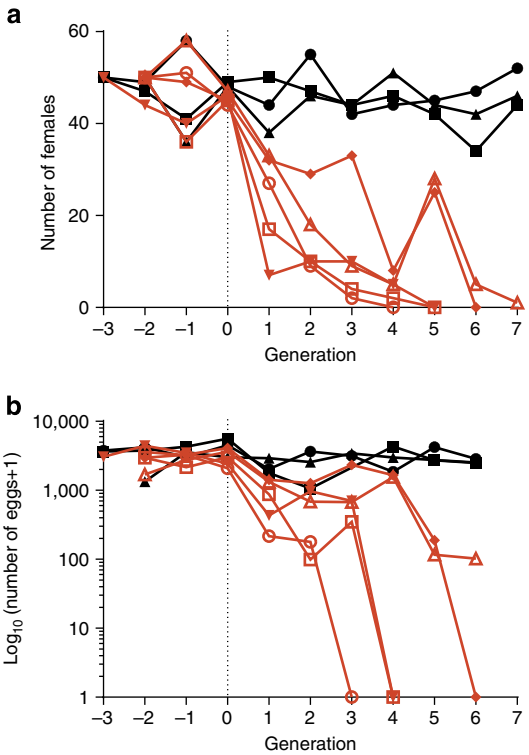


Figure 3 | Transgenic ^{gfp}111A-2 males suppress wild-type mosquito caged populations. Into caged wild-type mosquito populations, initially composed of 50 male and 50 female individuals, hemizygous ^{gfp}111A-2 (red lines) or hemizygous ^{gfp}I-PpoI-2 control males (black lines) were released starting from generation 0 (dotted line). The release rate of 3X was set relative to the initial number of males in the target population. Shown is the number of adult females in each cage population (**a**) as well as the number of eggs laid by each of the caged population in that generation (**b**).

Discussion

As with any other forms of pest control, it is important to consider possible pathways for the evolution of resistance. The universal conservation of the I-PpoI recognition sequence[12] within the multicopy essential 28S rDNA locus of animals makes the rapid selection of target site resistance unlikely. Other rDNA sequences or alternative X-specific sequence repeats could also be targeted—for example, by recently described programmable DNA endonucleases[13,14]. Our findings thus suggest a more universal approach towards achieving sex distortion.

The introduction of a sex distorter favouring the production of males into a pest population has been predicted to be a highly effective control strategy[3]. Depending on the use of one or multiple autosomal distorter loci the single releases of such males may be at least two orders of magnitude more effective than sterile males and repeated releases 2–70 times more effective than equivalent releases of female-killing alleles[2]. Using a discrete-generation model specifically parameterized with the properties of our transgenic strains, we find the release of single-locus homozygous I-PpoI males to have a greater suppressive effect than equivalent releases of males causing sterility or female-specific lethality and can effectively reduce vector populations at low release ratios (Supplementary Fig. 7). The scaling up of COPAS automated fluorescent larval sorting technology[10,15] or the use of conditional expression systems will facilitate the sustained mass releases of these autosomal strains in the field. In addition, the insertion of the I-PpoI transgene on the male sex chromosome may generate a driving Y chromosome that, in theory, could eliminate a natural vector population even when seeded with a very low number of transgenic males[1,16].

Methods

Engineering of destabilized I-PpoI variants. Eleven individual residues were chosen for mutation based upon their location in any of three separate regions of the protein: coordination of structural zinc ions within the protein fold (C41, C100, C105, H110, C125, C132, H134 and C138), formation of the dimer interface (W124) or packing within the hydrophobic core (H106 and L111). Each residue was individually converted into an alanine. Residue W124 was more extensively altered to phenylalanine, leucine or alanine (which correspond to serial reductions in the size of the larger indole side chain) because that residue plays an important role in the packing of the enzyme dimer interface (Fig. 1b). All point mutants of I-PpoI and the wild-type were expressed using a bacterial expression plasmid construct. The vector contains the I-PpoI reading frame (amplified from pI3-941 (ref. 17)) cloned between the *Bam*HI and *Nde*I sites within the pET11c[18] plasmid. The 15 point mutations were incorporated into the I-PpoI scaffold, first individually and then in pairwise combinations as described in the text, using a QuikChange II XL Site-Directed Mutagenesis Kit (Stratagene) following the manufacturer's protocol. The identity of the point mutants were verified using DNA sequencing.

Purification of I-PpoI variants. Protein expression and purification was performed as previously described[18]. Briefly, each expression construct was transformed into *Escherichia coli* BL21-CodonPlus (DE3)-competent cells (Promega Biotech). A 5-ml LB starter culture was inoculated with a single colony and grown at 37 °C for 3–4 h h, and then added to 1 l of LB media pre-warmed to 37 °C and grown until OD_{600} of ~0.8. Isopropyl β-D-1-thiogalactopyranoside and zinc acetate were then simultaneously added to concentrations of 0.5 and 1.5 mM, respectively, and the culture was moved to 16 °C and allowed to grow for an additional 16 h. Cells were harvested via centrifugation at 4 °C. Pellets were resuspended in 50 ml of 50 mM Tris pH 8, 100 mM NaCl, 5% Glycerol, 1 mM $MgCl_2$, 5 mM dithiothreitol and 0.5 mM phenylmethyl sulphonyl fluoride. Cells were lysed via sonication and centrifuged to remove debris. The supernatant was diluted twofold with 50 mM Tris pH 8 and then applied to a 5-ml HiTrap Heparin HP (GE Healthcare) affinity column equilibrated with 50 mM Tris pH 8. A salt gradient was then run from 0 to 0.6 M NaCl under constant buffering conditions. Using this protocol, wild-type I-PpoI eluted at ~0.4 M NaCl. Those point mutants that were soluble eluted within the bounds of the same salt gradient. The proteins were used immediately in subsequent assays described below.

Determination of denaturation temperatures and thermal half-lives. Purified I-PpoI variants were diluted to concentrations between 10 and 20 μM and exchanged via overnight dialysis into 10 mM potassium phosphate buffer at pH 8. CD data were collected on a Jasco J-815 spectrometer in a 0.1-cm path length

cuvette. Initial UV CD wavelength scans (250 –190 nm) were collected for each construct at 20 °C and then at 90 °C; the wavelength where the change in CD signal strength was greatest (206 nm) was then chosen for monitoring during a variable temperature scan. Temperature melts spanning a range from 4 to 95 °C were measured using 2° increments with temperature regulated by a Peltier device. Sample temperature was allowed to equilibrate for 30 s before measurement. Thermal denaturation half-points (T_m values) were determined by curve-fitting using the SpectraManager software supplied by the manufacturer. In subsequent experiments conducted for each construct, enzyme samples at 11.5 μM protein concentration, under the buffer conditions listed above, were set to a single temperature value (30 or 37 °C), and the loss of CD ellipticity signal (monitored at 206 nm wavelength) was followed for 4 h, with measurements taken every 30 s. The average of the first six values (collected between 0 and 3 min) and of the last six values (collected between 237 and 240 min) were used as values corresponding to the beginning and end points of each experiment (that is, $t = 0$ and $t = 4$ h). The protein was then completely unfolded at 95° and an unfolded baseline signal was monitored for 30 min and also averaged. This value was subtracted from the $t = 0$ and $t = 4$ h ellipticity values described above. These values were then converted to fraction unfolded protein, and each construct's thermal half-life was determined by assuming (i) that unfolding followed a first-order exponential decay (which agreed well with all unfolding curves over the time course of the experiment) and (ii) that the 95 °C average baseline represents 100% unfolded protein. Using this approach, the thermal half-life ('$t_{1/2}$') was calculated according to

$$F = e^{-(t/t(1/2))\times\ln(2)} \rightarrow \ln(F) = (-t/t_{1/2})\times\ln(2) \rightarrow t_{1/2} = -t\times(\ln(2)/\ln(F)) \quad (1)$$

where F = fraction folded and t = time.

Determination of specific activity. Protein activity cleavage assays were performed using a substrate corresponding to a 937-bp polymerase chain reaction (PCR) product that was generated using the pBend3-Ppo plasmid[19]. To create that plasmid, oligonucleotides containing the I-PpoI target site (5′-CTAGATGACTCT CTTAAGGTAGCCAA-3′ and its complement 5′-TCGATTGGCTACCTTAAGA GAGTCAT-3′) were annealed and then ligated into pBend3 (ref. 20) between the *Sal*I and *Xba*I restriction sites. The DNA substrate contains the I-PpoI target site flanked by 576 and 361 bp, respectively. I-PpoI constructs were diluted to concentrations ranging from 50 to 1,000 nM enzyme in 10 mM Tris pH 7.5, 50 mM NaCl, 0.1 mM EDTA. Samples were then added in a 1:10 v/v ratio into a standard cleavage buffer (2.5 mM CAPS/CHES pH 10, 5 mM NaCl, 0.2 mM dithiothreitol, 1 mM $MgCl_2$, 0.01 mg ml^{-1} bovine serum albumin and 20 nM DNA substrate), corresponding to final enzyme concentrations of 5–100 nM (adjusted for each construct to allow for production of an easily measurable amount of product during the reaction, without completely digesting the substrate). The digests proceeded at room temperature (23 ± 0.5°) for 5 min and were then quenched with 2 × stop buffer (2% SDS, 0.1 M EDTA, 20% glycerol and 0.2% bromophenol blue). The reaction products were separated and visualized using agarose gel electophoresis (2% TAE) and the resulting bands were quantitated using ImageJ.

Generation of germline transformation constructs. eGFP::I-PpoI was amplified using the common primer GFP-PpoI-F (5′-GAGCTGTACAAGTCCGGACTC AGA-3′) and unique the chimeric primers:

Ppo111A-SanDI (5′-CGTTGGGACCCGGGCACCAGTTTCTGCCTTTGTTG TCGTCTAGCGACTCCCAGCATGCGTGCAAGGGATTGTGGCATCG-3′,

Ppo124A-SanDI (5′-CGTTGGGACCCGGGCATGCGTTTCTGCCTTTGTTG TCGTCTAGCGACTCCCAGCACAAGTGCAAGGGATTGTGGCATCG-3′), Ppo124L-SanDI (5′-CGTTGGGACCCGGGCATAAGTTTCTGCCTTTGTTGTC GTCTAGCGACTCCCAGCACAAGTGCAAGGGATTGTGGCATCG-3′) and Ppo111A/124L-SanDI (5′-CGTTGGGACCCGGGCATAAGTTTCTGCCTTTGTT GTCGTCTAGCGACTCCCAGCATGCGTGCAAGGGATTGTGGCATCG-3′) from pBac[3xP3-DsRed]β2-eGFP::I-PpoI (ref. 9). Each 0.45-kb product was digested with *Bgl*II and *San*DI and ligated back into pBac[3xP3-DsRed]β2-eGFP::PpoI to generate the final transformation constructs.

In order to generate pBac[3xP3-DsRed]β2-eGFP(F2A)I-PpoI, we used primers β2-AfeI-F (5′-CCCAGCGCTTAGAGAGCAACGCTCGTGC-3′) and F2A-NLS-R (5′-CAGCTTCAGCAGGTCGAAGTTCAGCGTCTGCTTCACCGGGGCGCGCT TGGCGCGCTTGTACAGCTCGTCC-3′) as well as primers F2A-NLS-F (5′-CTG AACTTCGACCTGCTGAAGCTGGCCGGCGACGTGGAGTCGAACCCGGGCC CTAAGAAGAAGAGGAAGGTGA-3′, which encode the F2A peptide sequence[21] and Ppo124W-SanDI (5′-CCGTTGGGACCCGGGCACCAGTTTCTGCCTTTGT TGTCGTCTAG-3′) to amplify two fragments from pBac[3xP3-DsRed]β2-eGFP::I-PpoI that were used as template for a second PCR performed with the primer pair β2-AfeI-F and Ppo124W-SanDI. The final 1.3-kb *Afe*I and *San*DI digested was ligated into pBac[3xP3-DsRed]β2-eGFP::I-PpoI. In order to obtain the constructs pBac[3xP3-DsRed]β2-eGFP(F2A)I-PpoI-111A and pBac[3xP3-DsRed]β2-eGFP (F2A)I-PpoI-111A/124L, a fragment of 1.9 kb was digested from the respective eGFP::PpoI plasmids and moved into pBac[3xP3-DsRed]β2-eGFP(F2A)I-PpoI to replace the original eGFP::I-PpoI sequence. Each construct contains piggyBac inverted repeats and DsRed under the control of the 3xP3 promoter as a germline transformation marker.

Generation of transgenic mosquito lines. *A. gambiae sensu stricto* embryos (strain G3, from here on referred to as wild-type) were injected using a Femtojet Express injector and a Narishige 202ND micromanipulator mounted on an inverted microscope (Nikon TE-DH100W) with a mixture of 0.2 μg μl^{-1} of the selected plasmid and 0.4 μg μl^{-1} of helper plasmid containing a *vasa*-driven piggyBac transposase (Eric Marois, unpublished). An average of 375 embryos was injected for each plasmid with a 17.5% of survivors obtained. The hatched larvae were screened for transient expression of the DsRed marker and positives (∼54%) crossed to wild-type mosquitoes. The progeny was analysed for DsRed fluorescence. DsRed positives were crossed individually with wild-type mosquitoes to obtain transgenic lines. Transgenic lines originated from a single integration event were selected using inverse PCR and Mendelian inheritance of the transgene scored in the progeny (Supplementary Table 1). Transgenic mosquitoes were analysed on a Nikon inverted microscope (Eclipse TE200) at a wavelength of 488 nm to detect eGFP expression (Filter 535/20 nm emission, 505 nm dichroic) and 563 nm to detect DsRed expression (Filter 630/30 nm emission, 595 nm dichroic). Transgenic lines were named after the I-PpoI variant (the ^{gfp} superscript indicates eGFP fusion) they express followed by an identifier digit. All transgenic lines were maintained in a way so that in each generation transgenic females were back-crossed back to G3 wild-type males.

Inverse PCR. For each line, 500 ng of genomic DNA were digested with 10 units of Sau3AI or HinP1I as separate reactions; 5 μl of each digestion was re-ligated with T4 DNA ligase (Takara) in a final volume of 20 μl, of which 5 μl were subjected to PCR. The piggyBac flanking regions were initially amplified with primers 5F1 (5′-GACGCATGATTATCTTTTACGTGAC-3′) and 5R1 (5′-TGACACTTAC CGCATTGACA-3′) for 5′ junctions; or 3F1 (5′-CAACATGACTGTTTTTAAAG TACAAA-3′) and 3R1 (5′-GTCAGAAACAACTTTGGCACATAT-3′) for 3′ junctions, followed by a second inner PCR reaction using primers 5F2 (5′-GCGAT GACGAGCTTGTTGGTG-3′) and 5R2 (5′-TCCAAGCGGCGAATGAGATG-3′) or 3F2 (5′-CCTCGATATACAGACCGATAAAAC-3′) and 3R2 (5′-TGCATTTGC CTTTCGCCTTAT-3′) on 5 μl of the first reaction. Insertion sites were sequenced using primers pB-5SEQ (5′-CGCGCTATTTAGAAAGAGAGA-3′) for 5′ junctions and pB-3SEQ (5′-CGATAAAACACATGCGTCAATT-3′) for 3′ junctions.

qRT–PCR and western blot analysis. Mosquito abdomens were dissected and 10 pairs of testes from each transgenic line were pooled to constitute a biological replicate for total RNA and protein extraction using TRI reagent (Ambion). RNA was reverse-transcribed using Superscript II (Invitrogen) after TURBO DNA-free (Ambion) treatment following the manufacturer's instructions. Quantitative real-time–PCRs (qRT–PCR) analyses were performed on cDNA using the Fast SYBR-Green master mix on a StepOnePlus system (Applied Biosystems). Ribosomal protein *Rpl19* gene was used for normalization. At least two independent biological replicates from independent crosses were subjected to duplicate technical assays. We used primers RPL19Fwd (5′-CCAACTCGCGACAAAACATTC-3′), RPL19Rev (5′-ACCGGCTTCTTGATGATCAGA-3′), eGFP-F (5′-CGGCGTGCAGTGCT TCA-3′) and eGFP-R (5′-CGGCGCGGGTCTTGT-3′). For the western blot analysis, proteins were separated under denaturing conditions (SDS–PAGE) on 4–15% precast gels (Criterion TGX, Bio-Rad) and subsequently transferred to Hybond ECL membranes (Amersham). Immunoblotting was performed with α-GFP Rabbit mAb and α-Tubulin Rabbit mAb (Cell Signalling Technology) at a 1:1,000 dilution and a 1:1,000 α-rabbit IgG HRP-linked secondary antibody (Cell Signalling Technology) using the Pierce SuperSignal West Femto Substrate (Thermo Scientific). Chemiluminescence signals were recorded using a Bio-Rad ChemiDoc XRS + system and quantified using the Bio-Rad Image Lab software.

rDNA analysis. Genomic DNA was extracted from females using the Wizard Genomic DNA Purification Kit (Promega). Primers rgDNA-F4 (5′-GTAAGGGA AGTCGGCAAACTGGATCCG-3′) and rgDNA-R4 (5′-GAGGAAGCCCTCCCAC CTATGCTGC-3′) were used to amplify the I-PpoI target site within the ribosomal 28S gene using the Phusion HF polymerase (Thermo Scientific). One-fifth of each PCR product was loaded in a 1.5% agarose gel and the remainder purified (Qiagen PCR purification kit) and then digested with 100 units of I-PpoI for 3 h at 37 °C. The reaction was then run in a 1.5% agarose gel after heat inactivation at 65 °C for 15 min. qRT–PCRs were performed using *Rpl19* as internal control for normalization and three independent technical replicates for each sample. Primers rgDNA18S-F (5′-CCGTCCAGACTCGCGTCCGTC-3′), rgDNA18S-R (5′-GGGC ACCATGAAGCTGGGTCG-3′) and rgDNA5.8S-F (5′-GGCAGGGGATCACTC GGCTCA-3′), rgDNA5.8S-R (5′-CGTCCGATGCGCCATATGCGTT-3′) were used.

Crosses and fertility assays. To assay fertility and the adult sex ratio, 10 transgenic males were crossed to 10 wild-type females and as a control 10 transgenic females were crossed to 10 wild-type males (Fig. 2 and Supplementary Table 2). In all crosses, mosquitoes were allowed to mate for 5 and 3 days after the blood meal gravid females were placed individually in oviposition cups. The number of eggs laid as well as the number of larvae hatching were counted. The larvae were then pooled and reared to adulthood. The total number of adult male and female mosquitoes from each cross were subsequently counted. Each transgenic line was assayed for at least two consecutive generations. To determine

whether the sex distortion phenotype was stably inherited, 10 ^{gfp}111A-2 males were crossed to 10 wild-type females (Supplementary Table 4). Their transgenic sons were backcrossed in the same way for four consecutive generations. The numbers of eggs, larvae hatched as well as the adult sex ratio were determined as described above. To generate homozygote individuals for the analysis of the effect of transgene copy number, 15 transgenic females and 15 transgenic males were crossed and from the offspring approximately 400 L1 larvae were sorted using a 'COPAS Mosquito' large particle flow cytometer (Union Biometrica) using a 488-nm solid-state laser and emission filters for the detection of GFP filter (510 nm, band width of 23 nm) and DsRed filter (610 nm, band width of 20 nm). Homozygote and heterozygote larvae were selected and separated according to the intensity of DsRed fluorescence and 20 individuals of each type were crossed to wild-type females (Supplementary Table 5). The numbers of eggs laid and the larval hatching rate were measured as described. Females were allowed to oviposit individually and the progeny of each female was reared to adulthood separately to determine the sex ratio. To study temperature dependence, three different sets of experiments were performed with ^{gfp}111A-2, ^{gfp}124L-2 and wild-type strains reared at 24 °C, the standard 28 and 32 °C (80% humidity) from larval stage L1 (Supplementary Table 3). ^{gfp}111A-2 and ^{gfp}124L-2 males were crossed to wild-type females and transgenic females were crossed to wild-type males as a control. Mosquitoes were mated for 5 days and females were allowed to lay eggs 3 days after a blood meal at standard temperature of 28 °C. The egg laying and hatching rate as well as the adult sex ratio were determined from the progeny of individual females as described. To test for the occurrence of potentially cleavage-resistance X-chromosomes 10 ^{gfp}111A-2 females, originating from crosses between ^{gfp}111A-2 males with wild-type females, were crossed to 10 wild-type males, in two independent experimental replicates. Twenty transgenic males that originated from these crosses (termed ^{gfp}111A-2^{mod}) were individually crossed to wild-type females. As positive and negative controls, 20 ^{gfp}111A-2 stock males or 20 ^{gfp}111A-2 stock females were individually crossed to 20 wild-type females or males, respectively (Supplementary Fig. 6d). To analyse the female progeny of transgenic males, we crossed 10 ^{gfp}111A-2 males with 10 wild-type females and then crossed, in four batches, 10–15 transgenic female survivors (termed ^{gfp}111A-2^{mod}) with 10–15 wild-type males (Supplementary Table 6). As a control, an equal number of stock ^{gfp}111A-2 females were crossed to wild-type males. The egg laying and hatching rate as well as the adult sex ratio were determined from the progeny of individual females as described.

Population cage experiments. We initially established eight populations of 50 male and 50 female wild-type individuals in population cages (BugDorm-1, Megaview). Each generation mosquitoes were allowed to mate for 5–7 days and then all females were bloodfed on mice. After 3 days, females were allowed to oviposit and the eggs were washed with 2% household bleach (in water) followed by water on a round filter paper (Whatman, 125 mm). High-resolution images of these filter papers were taken and used to automatically count all eggs using the *Anopheles* Eggcounter software[22]. To approximate a natural population growth rate of approximately nine females produced per female we applied a mortality rate of 50% at the egg stage. This was performed by cutting the filter paper containing eggs in half or by counting and selecting eggs manually when there were fewer than ~200 eggs. The remaining eggs were hatched and a maximum of 110 L1 larvae selected at random and moved to a new tray to be reared to the pupal stage at which all individuals were manually sexed and males and females were allowed to emerge in separate holding cages. After 48 h males and females as well as 150 release males (during the intervention phase) were introduced into a new cage to establish the next generation. The rationale for this approach is as follows: we established in the initial generations before releases that females laid an average of 2,815 eggs per cage or ~56.3 eggs per per female under these conditions. Using COPAS sorting we counted all larvae and determined a hatching rate of 68.4% under these conditions. Around 90% of larvae reared under these conditions reached the adult stage. Hemizygous release males (either ^{gfp}111A-2 males or ^{gfp}I-PpoI-2 males for sterile control releases) were reared in separate trays under identical conditions alongside the main populations but were COPAS-sorted at the L1 stage for the presence of the fluorescent transgene.

Discrete generations model of mosquito dynamics. The dynamics of the mosquito population is described by a discrete-generation model. Each adult female produces a number f of progeny per generation. This progeny survives to adulthood with a probability P_L that declines with the density of larvae (L): $P_L = \theta_L \times F(L)$, where θ_L is the probability of surviving through the larval stage at very low density and $F(L)$ characterizes the density-dependent survival arising from the competition among larvae and following a Beverton–Holt equation $F(L) = \frac{\alpha}{\alpha + L}$.

The recurrence for the number of female adults (A_t) is therefore:

$$A_t = \frac{1}{2} f A_{t-1} \theta_L F(f \times A_{t-1}) \quad (2)$$

Parameterisation: $f = 130.1$, $\theta_L = 0.076$ and $\alpha = 500{,}000$. In the absence of control, equation (2) set for equilibrium $\left(A^* = \frac{1}{2} f A^* \theta_L F(f A^*)\right)$ provides the equilibrium density value: $A^* = \frac{\alpha}{f}\left(\frac{f\theta_L}{2} - 1\right) \approx 15{,}157$. We investigate the equilibrium density of vectors as a function of the number of males released per generation for each control strategy. Solutions were derived either analytically (for SIT and RIDL) or numerically (for autosomal X-shredder) allowing to compare the thresholds of minimal release size to eliminate the target mosquito population for the different strategies.

Modelled control strategies. Releases of intervention males into the population result in a genetic and phenotypic diversity that is incorporated into the model by distinguishing groups of individuals with identical genotypes and thus identical life-history traits. In both models, females are assumed to mate once in their life, on the first day they emerge, with a male chosen at random from the entire adult population. We characterize adult females according to their genotype and that of their mate: A_{ij} is a female of genotype i and fertility f_i that mated a male of genotype j, $\{i,j\} \in \{0,1,2\}^2$. Each of these matings gives rise to a specific set genotypes in the progeny. Individuals carrying autosomal sex ratio distorters are indexed by whether they have 0, 1 or 2 autosomal copies of the distorter transgene, which is inherited in a Mendelian manner. At male meiosis, these alleles induce shredding of the X chromosome and distort to m_x the proportion of gametes that inherit the Y chromosome. We consider two types of releases: males being either heterozygous or homozygous carriers of the distorter transgene. Males carrying Y-linked sex ratio distorters are indexed by whether they carry the distorter-bearing Y or not. The construct induces X shredding and is transmitted along the Y to a proportion m_x of sperm. By mating with wild females sterile males (SIT strategy) prevent the generation of viable progeny and reduce the recruitment of the next generation. Individuals carrying a dominant late-acting lethal gene (RIDL strategy) are indexed by whether they have 0, 1 or 2 RIDL alleles. These dominant alleles induce the death of mosquitoes (either in both sexes 'bisexual form' or in females only 'female-specific form') at the end of the larval stage.

References

1. Hamilton, W. D. Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science* **156,** 477–488 (1967).
2. Schliekelman, P., Ellner, S. & Gould, F. Pest control by genetic manipulation of sex ratio. *J. Econ. Entomol.* **98,** 18–34 (2005).
3. Hickey, W. A. & Craig, Jr G. B. Genetic distortion of sex ratio in a mosquito, *Aedes aegypti*. *Genetics* **53,** 1177–1196 (1966).
4. Sweeny, T. L. & Barr, A. R. Sex ratio distortion caused by meiotic drive in a mosquito, *Culex pipiens* L. *Genetics* **88,** 427–446 (1978).
5. Robinson, A. S. Sex-ratio manipulation in relation to insect pest control. *Annu. Rev. Genet.* **17,** 191–214 (1983).
6. Newton, M. E., Wood, R. J. & Southern, D. I. A cytogenetic analysis of meiotic drive in the mosquito, *Aedes aegypti* (L.). *Genetica* **46,** 297–318 (1976).
7. Flick, K. E., Jurica, M. S., Monnat, Jr R. J. & Stoddard, B. L. DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature* **394,** 96–101 (1998).
8. Windbichler, N. *et al.* Homing endonuclease mediated gene targeting in *Anopheles gambiae* cells and embryos. *Nucleic Acids Res.* **35,** 5922–5933 (2007).
9. Windbichler, N., Papathanos, P. A. & Crisanti, A. Targeting the X chromosome during spermatogenesis induces Y chromosome transmission ratio distortion and early dominant embryo lethality in *Anopheles gambiae*. *PLoS Genet.* **4,** e1000291 (2008).
10. Catteruccia, F., Benton, J. P. & Crisanti, A. An *Anopheles* transgenic sexing strain for vector control. *Nat. Biotechnol.* **23,** 1414–1417 (2005).
11. Magnusson, K. *et al.* Demasculinization of the *Anopheles gambiae* X chromosome. *BMC Evol. Biol.* **12,** 69 (2012).
12. Nomura, N., Nomura, Y., Sussman, D., Klein, D. & Stoddard, B. L. Recognition of a common rDNA target site in archaea and eukarya by analogous LAGLIDADG and His-Cys box homing endonucleases. *Nucleic Acids Res.* **36,** 6988–6998 (2008).
13. Joung, J. K. & Sander, J. D. TALENs: a widely applicable technology for targeted genome editing. *Nat. Rev. Mol. Cell Biol.* **14,** 49–55 (2013).
14. Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. *Nat. Methods* **10,** 957–963 (2013).
15. Marois, E. *et al.* High-throughput sorting of mosquito larvae for laboratory studies and for future vector control interventions. *Malar. J.* **11,** 302 (2012).
16. Deredec, A., Burt, A. & Godfray, H. C. The population genetics of using homing endonuclease genes in vector and pest management. *Genetics* **179,** 2013–2026 (2008).
17. Muscarella, D. E., Ellison, E. L., Ruoff, B. M. & Vogt, V. M. Characterization of I-Ppo, an intron-encoded endonuclease that mediates homing of a group I intron in the ribosomal DNA of *Physarum polycephalum*. *Mol. Cell Biol.* **10,** 3386–3396 (1990).
18. Flick, K. E. *et al.* Crystallization and preliminary X-ray studies of I-PpoI: a nuclear, intron-encoded homing endonuclease from *Physarum polycephalum*. *Protein Sci.* **6,** 2677–2680 (1997).
19. Eastberg, J. H., Eklund, J., Monnat, Jr R. & Stoddard, B. L. Mutability of an HNH nuclease imidazole general base and exchange of a deprotonation mechanism. *Biochemistry* **46,** 7215–7225 (2007).
20. Zwieb, C. & Adhya, S. Improved plasmid vectors for the analysis of protein-induced DNA bending. *Methods Mol. Biol.* **30,** 281–294 (1994).

21. Fang, J. *et al.* Stable antibody expression at therapeutic levels using the 2A peptide. *Nat. Biotechnol.* **23,** 584–590 (2005).
22. Mollahosseini, A. *et al.* A user-friendly software to easily count *Anopheles* egg batches. *Parasit. Vectors* **5,** 122 (2012).

Acknowledgements

We thank Eric Marois, Silke Fuchs, Tony Nolan, Ann Hall, Andrew Hammond and Shani Jayawardhan. This study was funded by a grant from the Foundation for the National Institutes of Health through the Vector-Based Control of Transmission: Discovery Research (VCTR) programme of the Grand Challenges in Global Health initiative of the Bill & Melinda Gates Foundation. This study was funded by the European Research Council under the European Union's Seventh Framework Programme ERC grant no. 335724 awarded to N.W. Funded by the European Commission and Regione Umbria under grant agreement I-MOVE-PCOFUND-GA-2010-267332.

Author contributions

R.G., L.A.D., M.M. and F.B. performed the experiments. R.G., B.L.S., N.W. and A.C. designed the experiments. R.G., A.D., A.B., B.L.S. and N.W. performed analyses. B.L.S., N.W. and A.C. wrote the paper.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Galizi, R. *et al.* A synthetic sex ratio distortion system for the control of the human malaria mosquito. *Nat. Commun.* 5:3977 doi: 10.1038/ncomms4977 (2014).