Chapter 8

Gene Insertion and Deletion in Mosquitoes

Zach N. Adelman, Sanjay Basu and Kevin M. Myles
Fralin Life Science Institute and Department of Entomology, Virginia Tech, Blacksburg, VA, USA

RANDOM INSERTION OF GENETIC ELEMENTS

Transposons

Introduction. The poetical notion that the genome of an organism was a tranquil arrangement of cooperative genes working in harmony was forever destroyed with the fundamental discovery by Barbara McClintock of the existence of selfish genetic elements [1]. These transposable elements (TEs), or transposons, do not sit still and do not necessarily share the goals of the genome in which they reside, being capable of moving around the genome in an attempt to increase the frequency in which they are inherited. Class I transposons, descended from ancient retroviruses, generate an RNA intermediate, which is turned back into DNA in order to reintegrate into a new location in the genome. Class II transposons catalyze their own excision and reintegration; repair by the host cell machinery can restore the first copy resulting in a net gain in transposon number. Both types are highly mutagenic; the integration of a transposon into a new location in the genome can disrupt any gene or regulatory region present therein. Both transposon classes have been highly successful throughout evolutionary history and these sequences now make up a majority of the genome of many organisms, including humans. The mutagenic power of TEs is now widely recognized as a major driving force in evolution (reviewed in Refs [2,3]). As mentioned in Chapter 2, initial efforts at genetic control in mosquitoes utilized chromosomal translocations induced by irradiation. Following the successful transposon-based transformation of Drosophila melanogaster using a transposon known as the P element [4], a new era where much more targeted genome manipulation was possible was born. In a relatively short timeframe, the efforts of several groups to use P-element-based genetic transformation on malaria and dengue mosquitoes resulted in some rare recombination
events, but no true transposon-mediated integration \[5–7\]. Finally, the P element, so successful in *Drosophila*, was abandoned, and a search for new transposons to use in mosquito transformation experiments began. It was not until 10 years later that the barrier in mosquito transformation was broken with the successful transposon-based genetic modification of *Aedes aegypti* with the *Mos1* \[8\], *Hermes* \[9\], and *piggyBac* \[10\] elements, and the transformation of *Anopheles stephensi* \[11\] and *Anopheles gambiae* \[12\], with the transposons *minos* and *piggyBac*, respectively.

*How they work.* All current insect transposon-based transformation systems are based on class II DNA elements. These simple, selfish genetic elements consist of a pair of inverted repeats flanking a single open reading frame encoding the transposase protein. Following transcription and translation directed by the host cell machinery, the transpose binds selectively to its own inverted repeat sequences and catalyzes the excision and reintegration of the entire element (reviewed in Ref. \([2]\)). This mechanism, referred to as cut-and-paste transposition, results in the movement of the transposon from one genomic location to another. For insect transformation, this single genetic unit (transposon + inverted repeats) is split into a bipartite system, with the inverted repeats present in one plasmid DNA molecule, and the transpose protein (without inverted repeats) encoded separately (either on a separate plasmid or as *in vitro* transcribed RNA); the use of such transposons to modify mosquito genomes has been the subject of a number of excellent reviews \([13–15]\). Thus, any genetic segment of interest to the investigator can be placed in between the inverted repeats and subsequently be mobilized by the transpose from the plasmid and into the mosquito genome (Figure 8.1).

*Strengths.* Transposons have been the workhorse of the mosquito transformation field for more than 15 years now. They are easy to manipulate using standard molecular techniques and are consistent in their rates of transformation.

**FIGURE 8.1** Methods of germline transformation. Blue arrows indicate the L and R arms of the initial TE-based transposon. Green and red boxes indicate fluorescent marker genes, while yellow represents bacterial sequences such as antibiotic resistance genes or an origin of replication. White/orange arrowheads indicate *attP-attB* sites, while gray/black arrowheads indicate heterospecific *lox* sites. Dotted lines indicate crossing-over events during recombination.
transformation (typically ranging from 1% to 15%, depending on the mosquito/element combination). Transposon-based transformation allows the investigator to sample multiple locations of the genome due to the ubiquitous distribution of preferred target sites (typically the nucleotides TA or TTAA). This sampling of the genome can be used to identify new enhancer elements or genes [16], to find a genetic locus where insertion of the transposon yields the least effect on mosquito fitness [17], or to identify the most suitable location for expression of the transgene of interest [18–20]. Because the transposons used do not rely on any host factors, they can be easily adapted for use in new insect systems with confidence.

**Weaknesses.** Although a reliable means of inserting genes into the mosquito genome, transposons suffer from a number of weaknesses that have consequences for both studies on basic biology and applied strategies to develop mosquitoes to be used in genetic control programs (Figure 8.2). Certainly, the primary weakness is also their greatest strength—the ability to insert virtually anywhere in the genome. While a desired trait for forward genetic experiments, the lack of an ability to control where a transgene inserts severely complicates the development and characterization of transgenic strains for genetic control programs. The local chromatin structure or the presence of nearby enhancers/repressors can dramatically alter the expression pattern of the transgene, requiring investigators to characterize many insertion sites (typically 5–10) in order to find those that are most favorable [18–20]. Nearby TEs (or the inverted repeats of the transposon used itself) may induce epigenetic changes, resulting in substantial variation between individuals containing the same insertion [21]. Transcriptional initiation on the opposite strand by either recognized or cryptic promoters may induce the formation of dsRNA and result in silencing of the transgene [22].

**FIGURE 8.2** Position-based effects on transgene expression. A number of factors can influence the stability and/or expression of transgenes. Nearby enhancers/repressors (purple boxes) can reduce, increase, or change the spatial expression of the transgene. Nearby transposable elements (TEs) may induce heterochromatin formation and position effect variegation (PEV). Direct repeats may collapse following random DNA breaks, resulting in loss of the transgene. The presence of cryptic (or recognized) promoters may generate double-stranded RNA (dsRNA), resulting in gene silencing. The transgene may be linked to a deleterious mutation; selection for the transgene may then increase the frequency of this allele in the population, lowering its fitness.
Transgene insertions that occur nearby recessive deleterious mutations can suffer substantial fitness effects when made homozygous for the transgene [23,24]. Concerning long-term stability, the presence of short direct repeats sequences (common in mosquito genomes, particularly in Ae. aegypti) can result in recombination and loss of the entire transgene. Transposons are also limited in the cargo that they can carry, with maximum sizes of 10–12 kb. Early on, one of the biggest concerns regarding the use of transposons to develop transgenic mosquitoes was their ability to be remobilized by native elements in the mosquito genome. While such remobilization is indeed efficient in An. stephensi [16,25], in mosquito such as Ae. aegypti, even intentional remobilization has proven to be extraordinarily difficult [26–28], suggesting for at least this species that this is not a serious concern. However, the biological basis for this lack of remobilization is still unknown, and could certainly vary in the field. Thus, methods for permanently stabilizing transgenes have been developed, which remove one of the inverted terminal repeats [29] or both [30–32]. Finally, once favorable chromosomal integration sites are discovered and characterized, using transposon-based systems alone it is not possible to reuse the same integration site.

Docking-Site-Based Integration

Introduction. The large amount of work needed to characterize transposon-based integration sites in the mosquito genome for characteristics such as transgene expression and variation, fitness and stability has driven the implementation of site-specific, docking-site-based integration methods based on the bacteriophage ΦC31 recombinase [33]. First adapted to human cells [34] and then Drosophila [35], the ΦC31 system was first utilized in Ae. aegypti as a way to reuse favorable transposon insertions [36]; with subsequent success also in Aedes albopictus [37], An. stephensi [17], and An. gambiae [38,39].

How they work. In its native environment, the bacteriophage ΦC31 encodes a recombinase that recognizes a short DNA sequence in the phage genome (termed attP, for attachment-phage) and a short DNA sequence in the host bacterial genome (termed attB, for attachment-bacteria). Once both sites are bound, the recombinase catalyzes the integration of the phage genome into the bacterial genome; a process that is unidirectional. As these attachment sites are small and easy to manipulate, they were very readily converted into gene modification tools [34,35]. For typical experiments involving manipulating the genome of mosquitoes, the attP site is included in a normal transposon-based construct that integrates randomly into the mosquito genome. Once all of the insertion sites are characterized, the most favorable sites can be reused by encoding an attB site into a new target plasmid and introducing it into mosquito embryos while providing a source of recombinase (Figure 8.1).
**Strengths.** The ability to reuse well-characterized transposon-based insertion sites has made ΦC31-based recombination the current state-of-the-art practice for developing and testing genetically modified mosquito (GMM) strains for genetic control programs [40]. Reusing the same integration site allows a panel of related effector molecules to be evaluated together with the contributions of chromosomal position strictly controlled for. Overall rates of integration into attP target sites are similar to overall rates of transposon-based transposition but do not require the extensive downstream characterization of the site after the initial evaluation.

**Weaknesses.** Docking-site-based systems such as ΦC31 still depend on the random integration of a transposon and may also require additional steps to stabilize the transposon insertion. These insertions are thus restricted to a very small portion of the mosquito genome and are still ultimately descended from a single transformation event where loss of heterozygosity around the insertion site can depress fitness. While screening enough sites can identify insertions free of such effects, it remains to be seen whether the same will hold true under field conditions. ΦC31 integration results in the incorporation of an entire bacterial plasmid, including any antibiotic resistance genes, into the recipient genome. Removing this unwanted material thus requires additional downstream steps prior to inclusion of any modified strains in a genetic control program. Alternatively, including two attP target sites in the initial transposon allows the ability to perform recombinase-mediated cassette exchange [41], though this has not been reported in mosquitoes at the time of this writing (Figure 8.1).

**Other Recombinases**

Isolated from the phage P1, cre (causes recombination) recombinase is commonly used to catalyze recombination events in a wide variety of organisms [42–46], though this system is not commonly used in mosquitoes. Cre-mediated recombination occurs between two short loxP target sites, with each loxP site consisting of perfect 13-bp inverted repeats flanking an 8-bp asymmetric spacer. When two such loxP sites flank a gene of interest in cis, cre-mediated recombination results in the effective excision of the gene, leaving behind only a single loxP site. While cre recombinase has been shown to be highly efficient at catalyzing the excision of DNA sequences in Ae. aegypti [47], excision is so thermodynamically favored that transgene insertion events were never recovered [36]. To overcome this obstacle, heterospecific lox sites consisting of one wild-type loxP site and a second lox site containing one or more substitutions in the spacer region have been developed [48–50]. As cre-binding specificity is denoted entirely by the inverted repeats, these modified lox sequences bind cre normally and thus can recombine with themselves, but can no longer recombine with the wild-type loxP sequence. Such sites have been used to perform site-specific recombinase-
mediated cassette exchange in systems such as *Drosophila* [44], thus it is possible that this system could also be adapted to mosquitoes. The FLP/FRT recombination system derived from the baker’s yeast, *Saccharomyces cerevisiae*, is also heavily utilized for genomic manipulation in *D. melanogaster* (reviewed in Ref. [51]), and more recently, it has been shown to be effective in the silkworm [52]. Though this system was able to catalyze interplasmid recombination in *Ae. aegypti* embryos [53], it failed to produce any recombination events with the mosquito chromosome [47].

**SITE-SPECIFIC NUCLEASES**

**Homing Endonucleases**

*Introduction.* A new type of selfish DNA element that was first identified in yeast in the 1970s (reviewed in Refs [54,55]), homing endonucleases have been highly successful in invading eukaryotic, bacterial, archaeal, and even phage genomes despite their simple structure—a single open reading frame. Many (but not all) homing endonucleases identified to date recognize highly specific DNA target sequences found in self-splicing introns. Others, known as inteins, are translated along with a host protein and are posttranslationally processed in a manner that frees the homing endonuclease but leaves the host protein intact. Both strategies have the advantage of limiting detrimental effects to the host cell and its genome, enabling the spread of these selfish genes. Following the identification and biochemical characterization of many homing endonucleases in the 1980s and 1990s, researchers started to repurpose these simple elements for various genome manipulation applications. For example, the introduction of the target sequence recognized by the homing endonuclease I-SceI into the *Drosophila* genome enabled studies of DNA break repair and homologous recombination [56–59]. Homing endonucleases were relatively unknown to the vector biology community at this time, their adaptation beginning when Burt [60] suggested that the primitive form of genome invasion used by homing endonucleases may be well-suited for genetic strategies to control malaria or other vector-borne diseases. Indeed, several homing endonucleases have since been used effectively to introduce site-specific DNA breaks in *An. gambiae* [61,62] and to excise genes from *Ae. aegypti* [63,64]. Through biochemical redesign, the homing endonucleases I-CreI and I-AnI have been modified to recognize targets in the *An. gambiae* genome directly [65]. Transgenic expression of wild-type I-PpoI during spermatogenesis completely sterilized *An. gambiae* mosquitoes [62]; large cage trials using these transgenic sterile mosquitoes indicate effectivenes at crashing populations of this mosquito [66]. The expression of modified versions of I-PpoI in the male testes of *An. gambiae* has led to the development of sex-distorter strains that produce almost all males [39], while I-SceI was shown to successfully invade large cage populations [65].
How they work. The phrase “homing” refers to their ability as mature proteins to return to the site of their mRNA “birth” and introduce a double-stranded break on the homologous chromosome lacking the homing endonuclease gene. Unlike TEs and recombinases, homing endonucleases do not catalyze any reactions beyond DNA cleavage. This extraordinarily simplistic mode of action thus relies entirely on the cellular DNA repair machinery to generate a duplicate version of the homing endonuclease gene, by using the original gene as a repair template. Homing endonucleases can be categorized by their mode of catalyzing DNA cleavage into at least four completely independent families, and an extensive body of literature is available concerning the structural basis for DNA-binding specificity and double-stranded DNA break formation (reviewed in Ref. [54]). Of particular interest is that the target sequences for most homing endonucleases are 18 bp or longer, meaning that even in large eukaryotic genomes the probability of finding an endogenous site is very small. For example, the target sequence of I-SceI is 18 bp, and the corresponding chance of finding an exact match in a random sequence is $1 \times 10^{-10}$. Given the size of the An. gambiae (2.7 $\times$ 10^8 bp) and Ae. aegypti (1.3 $\times$ 10^9 bp) genomes, the corresponding chances of randomly finding an I-SceI target site are approximately 1:250 and 1:50, respectively.

Strengths. Homing endonuclease genes are relatively small, with dimeric nucleases such as I-CreI and I-PpoI both less than 500 bp; the larger monomeric nucleases such as I-SceI are still less than 1 kb. Thus, these genes and their resultant proteins are easy to manipulate, express, and purify. When integrated into a mosquito genome, their small size may also reduce the risks of acquiring deleterious mutations that can affect their function. Their extreme specificity can be expected to reduce or virtually eliminate the impacts of off-target cutting; the transgenic expression of I-SceI is well-tolerated by both Drosophila and An. gambiae [56,65]. Homing endonucleases are a mature technology that have been used to generate the first experimentally validated gene drive system in mosquitoes [65] as well as transgenic sex-distorter strains ready to be deployed in SIT-type programs [39].

Weaknesses. The extreme specificity of homing endonucleases due to extensive protein—DNA contacts at the target site also makes reengineering these molecules to recognize new target sequences extremely difficult. While such modifications are possible [39,67–69], the cost (primarily human capital) associated with modifying homing endonucleases to recognize new and diverse target sites is thus the primary barrier to their widespread use as gene editing and modification tools. While large-scale sequencing of new microbial genomes and comparative genomic analyses have accelerated the discovery of new homing endonuclease genes [70], many of these have reduced or completely lost catalytic activity, meaning that each new nuclease must still be verified experimentally.
Chimeric Nucleases (ZFNs and TALENs)

Instead of attempting to identify naturally occurring nucleases to cut a target DNA sequence using homing endonucleases, or modifying a homing endonuclease to recognize a related sequence, other research groups have attempted to create de novo targeted nucleases by fusing a custom DNA-binding domain to a nonspecific nuclease domain. Some of the most well-characterized DNA-binding domains at the time were considered to be the zinc-finger motifs found in a family of transcription factors. Fusions of several such zinc-finger motifs to the nuclease domain of the type IIS restriction enzyme FokI (isolated from the bacteria Flavobacterium okeanokoites) resulted in the first zinc-finger nucleases (ZFNs) [71]. As each zinc finger was responsible for mediating interactions with three nucleotides, increasing the number of fingers included in the chimeric protein served to increase specificity of the resultant nuclease. At the same time, cleavage by the FokI domain relied on dimerization of the nuclease domain, meaning two independent ZFNs were required—one to bind upstream of the target on one strand and one to bind downstream of the target on the opposing strand. ZFNs have since been used successfully to edit the genome of a number of organisms (reviewed in Refs [72,73]), including inducing gene correction in human cells [74]. It was not until 2013 that ZFNs were first used in mosquitoes, when DeGennaro et al. [75] generated null mutant alleles for the Ae. aegypti orco gene, a key component of the insect olfactory system. Orco-deficient mosquitoes were unable to differentiate human from animal subjects and could not be repelled by N,N-Diethyl-meta-toluamide (DEET), though in the presence of CO₂ orco-mutant mosquitoes were still attracted to vertebrates. Liesch et al. [76] used ZFNs to knockout a neuropeptide Y-like receptor in Ae. aegypti thought to be involved in shutting off host-seeking behavior after a female mosquito obtains a blood meal. Finally, ZFNs were used to knockout a critical component of the Ae. aegypti CO₂ receptor complex, rendering these mosquitoes unable to detect this critical host cue [77].

Following the decoding of how transcription-activator-like elements (TALEs) from Xanthomonas species bacteria bind specific DNA sequences in the nuclear genomes of the plants they infect [78,79], chimeric nucleases based on this platform (TALE nucleases, or TALENs) were rapidly developed as an alternative to ZFNs [80]. The DNA-binding region of each TALE was found to consist of a series of almost identical 34 amino acid repeats, the only changes being a two amino acid variable region—each unique combination specified binding to one of each of the four nucleotide bases A, C, G, and T. By simply stacking these repeats, it was possible to spell out the desired DNA-binding region, and efficient methods for building new TALENs were developed rapidly [81–85]. In mosquitoes, Aryan et al. [86] demonstrated that TALENs could induce gene editing rates of >30% in Ae. aegypti when targeting a gene involved in eye pigmentation. Smidler et al. [87] adopted a different
strategy, generating transgenic strains of *An. gambiae* each expressing one half of a TALEN pair targeting the *tep1* gene. No mutations were observed until the two strains were crossed with each other, at which point a fully active TALEN capable of heterodimerization and DNA cleavage was produced. Despite these successes, the cost and difficulty associated with generating new nucleases is likely to preclude their adoption as common laboratory reagents for mosquito genome modification considering the clustered regularly interspaced short palindromic repeat (CRISPR) revolution (see CRISPR/Cas9).

**Strengths.** At the time when each was developed, both ZFNs and TALENs were the leading technology for genome editing applications, and as described earlier, both platforms have been used successfully to edit the genomes of the most important vectors of malaria and dengue. In the laboratory, their expression can be easily controlled via plasmid-based expression, and when inherited directly [87], appear to be stable in the mosquito genome.

**Weaknesses.** Both ZFNs and TALENs are relatively difficult to engineer, particularly for smaller academic labs. ZFNs suffered from the fact that the rules that govern protein–DNA contacts are not straightforward and are heavily influenced by context. While several prediction platforms and validation schemes were developed for both ZFNs and TALENs [82–84,88,89], the cost and required expertise associated with these methods has prevented their widespread adoption. While TALENs possess a far simpler set of rules for DNA binding, these molecules suffer from the large and repetitive nature of the recombinant genes used to produce them. Both ZFNs and TALENs rely on the *FokI* nuclease, which in its wild-type form can form homodimers as well as heterodimers, resulting in cutting at off-target sites [90]. Efforts at structure-based redesign of the dimerization plane decreased such off-target activity [91–93] but also decreased on-target activity in *Ae. aegypti* [94]. Also, the repetitive nature of the ZF or TALE repeats may promote instability if these nucleases are inserted directly into the mosquito genome [95].

**CRISPR/Cas9**

**Introduction.** Found in the genome of some bacteria, CRISPRs represent a novel adaptive, heritable immune defense mechanism against foreign DNA [96,97]. These CRISPR loci contain short (30–40 nt) stretches of variable sequences, interspersed by direct repeats. Short RNA transcripts termed CRISPR RNAs bind one or more CRISPR-associated (Cas) proteins, in association with a short transactivating-RNA (trRNA) to direct the Cas endonuclease complex to a given target site. A minimalist system characterized from *Streptococcus pyogenes* involving a single Cas protein (Cas9) is sufficient to act as an RNA-directed dsDNA endonuclease in combination with the appropriate CRISPR short RNAs [97]. Due to the simplicity in designing new CRISPR RNAs, the CRISPR/Cas9 system has competed extremely well against other more established technologies and has been successfully
adapted to a range of organisms (reviewed in Ref. [98]). Indeed, both Basu et al. [94] and Kistler et al. [99] reported that the CRISPR/Cas9 system was highly efficient in *Ae. aegypti*, with editing rates as high as 90%. Substantially lower editing rates were reported by Dong et al. [100], though with only a limited number of guide RNAs.

*How it works.* In a natural system, CRISPRs are associated with trRNA that guides the Cas9 endonuclease to the complimentary sequence of the CRISPR, resulting in cleavage and indel events in the target DNA. Expression of a fusion RNA whereby the CRISPR RNA and transactivating RNA are represented by a single molecule reduces the system to a two component reprogrammable nuclease (one protein, one RNA) [97,101,102]. Following hybridization of the CRISPR RNA, the complimentary strand of DNA is cleaved by an HNH domain of Cas9, whereas the noncomplimentary strand is cleaved by a RuvC-like domain generating a blunt-ended double-stranded break 3 bp upstream from the so-called protospacer-adjacent motif (PAM, a trinucleotide sequence represented by the bases NGG for *S. pyogenes* Cas9). As with other site-specific engineered endonucleases, repair of the double-stranded DNA break is mediated by the host’s DNA damage response.

*Strengths.* The power of the CRISPR system is that developing new nucleases does not involve generating a new protein or a new complicated gene, but simply a new short RNA guide that can be synthesized cheaply using existing technology. This not only enables both high-throughput, large-scale editing screens [103], but also puts editing reagents in the hands of even small laboratories that could not afford homing endonuclease redesign, ZFNs or TALENs. Initial experiments in mosquitoes have used Cas9 endonuclease generated via *in vitro* transcribed mRNA or purified protein [94,99,100]. Other strategies such as transgenic driver lines should be readily achievable as well [104], though multiple groups were unable to obtain editing when using plasmid-expressed Cas9 in mosquitoes [99,100]. DNA plasmid constructs can also carry RNA Pol III-based U6 promoters to drive expression of the chimeric sgRNA, thereby facilitating ease-of-cloning and transfection/injection methods. The efficiency of the system in mosquitoes may be further increased by inclusion of multiple sgRNAs targeting the same or different genes within a single experiment, known as multiplexing, making it possible for multiple-gene knockout investigations [105].

*Weaknesses.* At present, the primary weaknesses associated with CRISPR/Cas9 center around off-target effects and specificity [106]. Several new analytical methods have been developed to identify such off-target effects [107,108], and it is clear that there is substantial variability in the potential for off-target cutting between different guide RNAs. In the context of mosquito genome modification, gene drive or GMMs, any such off-target effects could result in decreased fitness due to the introduction of indels or excessive chromosomal breakage. While no off-target effects were observed in initial experiments [94], these analyses were limited to a few
computationally predicted sites. Such prediction algorithms have recently been found to perform poorly compared to whole genome analysis methods [107,108]. Thus, the full range of off-target effects should be clarified experimentally prior to field-testing of active CRISPR RNAs. The current generation of CRISPR/Cas9 reagents is also restricted by the requirements of the PAM, any thus guide RNAs are restricted to the sequence N_{18}GG. As additional CRISPR reagents are developed with different PAM requirements [109], the number of targetable loci in a particular genome should increase further.

TARGETED INSERTION OF GENETIC ELEMENTS THROUGH HOMOLOGOUS RECOMBINATION

As noted above, the insertion of material into the mosquito genome via TEs is a random process. Without the ability to control the landing site, the types of manipulations that can be performed to generate useful phenotypes in mosquitoes are relatively limited. Genetic cargo typically consists of two or more expression cassettes (one marker, one or more effectors) that are expected to act relatively independent of the landing site. With the advent of reprogrammable site-specific nucleases, it becomes possible to generate double-stranded DNA breaks more strategically in the mosquito genome: if these breaks are repaired using a homology-driven process it becomes possible to introduce genetic material in a site-specific manner (Figure 8.3).

Liesch et al. [76] used a ZFN along with a donor vector containing 1.3/1.5 kbp of homologous sequence flanking the target site to integrate a fluorescent reporter construct into an *Ae. aegypti* neuropeptide Y-like receptor gene. Similarly, McMeniman et al. [77] used a ZFN targeting the *Ae. aegypti* Gr3 CO2 receptor along with a donor vector containing 0.8/0.8 kbp of homologous sequence to integrate a similar reporter into the *Ae. aegypti* genome. In the malaria mosquito, Bernardini et al. [110] used homologous recombination to insert a ΦC31 docking site onto the Y-chromosome. Taking advantage of a fortuitous transposon insertion, these researchers used the homing endonuclease I-SceI along with a donor vector containing 0.6/2.1 kbp of homologous sequence on each side of the target site. In all these cases, the efficiency of transgene insertion via homology-dependent repair was at least an order of magnitude lower than that of transposon-based systems. This appears to be due largely to the preference by mosquitoes for using end-joining repair to correct double-stranded DNA breaks, a preference shared by most higher eukaryotes. Manipulating the DNA damage response thus represents a critical bottleneck to targeting engineering of the mosquito genome [111]. Altering the ability of an insect to use end-joining factors such as Ligase 4 in *Drosophila* [112] or Ku70 in silkmoth [113] has been shown to increase rates of homology-dependent gene insertion. Basu et al. [94] adapted this concept to *Ae. aegypti*, where
silencing of the ku70 ortholog at the time of embryonic injection resulted in a significant boost to rates of homology-dependent repair. These experiments used homology arms of 1.7/2.2 and 1.1/1.7 kb to introduce an EGFP reporter following TALEN or CRISPR/Cas9-based DNA break induction. The ability to introduce genetic material at any location in the mosquito genome substantially increases the type of beneficial phenotypes that can be engineered into a mosquito population. For example, changes to native promoters or splice sites may alter the timing or expression pattern of genes to better respond to pathogen infection. Even the introduction of single base changes in key positions in the genome could have a dramatic beneficial effect, such as reverting insecticide-resistant mosquitoes back to a susceptible state.

CONTROLLING TRANSGENE EXPRESSION; THE CURRENT STATE OF THE MOSQUITO TOOLBOX

Promoters

Markers. Early mosquito transformation experiments used an eye pigmentation-based rescue to identify a successful transgene insertion. This
was based on the *Drosophila cinnabar* gene, which rescues an orthologous defect in a mosquito enzyme known as kyurenine hydroxylase [114]. Depending on the insertion site, eye pigmentation was restored from white to yellow, orange, red or purple depending on the expression of the transgene as dictated by the local chromosomal environment. Despite the simplicity of this system, which did not require fluorescence microscopy, this method had a number of drawbacks. The appropriate white-eyed mutation was only available for *Ae. aegypti*, and was difficult to rear due to substantial inbreeding which reduced both lifespan and the production of progeny. Also, many of the eye-color changes were only visible in adults, requiring massive numbers of progeny to be reared to adulthood. Fortunately, markers based on the expression of one or more fluorescent proteins quickly replaced eye pigment-based rescue. These markers work well in all mosquito species tested, with screening able to be performed in the early larval stages, substantially reducing the labor required. The synthetic eye-specific promoter 3×P3 based on the Pax-6 enhancer and the hsp70 core promoter is the most commonly used [115], along with more general promoters such as the *Drosophila* Act5C [19,116], the baculovirus IE-1 [12,117], and the *Ae. aegypti* polyubiquitin [118]. These promoters have the advantage of driving gene expression in a more robust fashion, enabling identification of transgenic individuals at earlier developmental stages and lower magnifications.

**Genes of interest.** Following the establishment of transgenic technology in mosquitoes, considerable effort was expended by the vector biology community to characterize promoter elements (for simplicity, the term “promoter” as used refers to both the actual promoter where transcription is initiated and any enhancer/repressor elements required for controlling proper expression) capable of driving the expression of transgenes in the most biologically relevant tissues. These efforts were highly successful, as for both dengue and malaria vectors, a catalog of validated promoters is available for driving gene expression in the mosquito gut, fat body, salivary glands, ovaries, and testes (Figure 8.4, Table 8.1). With the exception of the Vg promoter, which has been extensively dissected [119,120], little is known about the functional elements that contribute to the activity of each promoter. Identification of specific regulatory sequences that control tissue and temporal specificity would be valuable to the development of synthetic promoters with novel expression patterns (e.g., robust expression in both midgut and salivary glands). Some large-scale computational analyses have identified some conserved regulatory sequences [121,122]; the accuracy of such high-throughput approaches is likely to improve as RNA sequencing data helps refine gene models across the various mosquito genomes. Nevertheless, the current cadre of available promoters represents a sufficient entry point for most studies.
Identifying New Promoters, a Cautionary Tale

While many promoters have been developed for use in transgenic experiments, many others did not behave as expected when used to control transgenic constructs. While model organisms such as Drosophila have compact genomes that largely constrain the location of transcriptional regulatory regions such as enhancers/repressors to a few kilobases in most instances, this is not the case for the mosquito genomes, which have expanded substantially due to the proliferation of TEs and repetitive sequences. For example, the average intergenic region is $3 \times$ longer in An. gambiae ($\sim 17$ kb), and $10 \times$ longer in Ae. aegypti ($\sim 56$ kb) as compared to Drosophila ($\sim 6$ kb) [147]. Most promoter studies will also include the entire 5′UTR and first intron of the donor gene in order to include any functional elements present therein. In both Drosophila and Anopheles, the average intron size is manageable $\sim 1$ kb, while in Ae. aegypti introns on average are $5 \times$ longer ($\sim 5$ kb) [147]. It is not uncommon to find introns spanning 20–50 kb in mosquito genes, severely complicating efforts to include all possible functional elements in transgenic constructs. These increased distances are almost entirely due to the interspersed nature of repetitive sequences and/or TEs in and around protein-coding genes.

**FIGURE 8.4** Tissue-specific promoters for driving transgene expression in mosquitoes. Control elements have been characterized to allow transgene expression in a number of tissues in the mosquito body, including the eyes (green), salivary glands (yellow), midgut (red), fat body (gray), ovaries (blue), and flight muscle (purple). See Table 8.1 for a complete list of validated promoters.
<table>
<thead>
<tr>
<th>Promoter</th>
<th>Mosquito</th>
<th>Sex</th>
<th>Spatial Distribution</th>
<th>Temporal Distribution</th>
<th>Activity in Other Species</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mall</td>
<td>Ae. aegypti</td>
<td>♂</td>
<td>Proximal lateral lobe of salivary glands</td>
<td>Constitutive</td>
<td></td>
<td>[123]</td>
</tr>
<tr>
<td>Apy</td>
<td>Ae. aegypti</td>
<td>♂</td>
<td>Distal lateral and medial lobes of salivary glands</td>
<td>Constitutive</td>
<td></td>
<td>[123]</td>
</tr>
<tr>
<td>AeVg</td>
<td>Ae. aegypti</td>
<td>♂</td>
<td>Fat body</td>
<td>Blood meal induced</td>
<td>An. stephensi</td>
<td>[124]</td>
</tr>
<tr>
<td>AeCPA</td>
<td>Ae. aegypti</td>
<td>♂</td>
<td>Midgut</td>
<td>Blood meal induced</td>
<td>An. gambiae</td>
<td>[125,126]</td>
</tr>
<tr>
<td>AaVgR</td>
<td>Ae. aegypti</td>
<td>♂</td>
<td>Vitellogenin receptor</td>
<td>Ovary, blood-meal induced</td>
<td></td>
<td>[127]</td>
</tr>
<tr>
<td>β2-tub</td>
<td>Ae. aegypti</td>
<td>♂</td>
<td>Testes</td>
<td>Developmental</td>
<td></td>
<td>[128]</td>
</tr>
<tr>
<td>Nanos</td>
<td>Ae. aegypti</td>
<td>♂</td>
<td>Ovaries/embryos</td>
<td>Developmental</td>
<td></td>
<td>[129]</td>
</tr>
<tr>
<td>30K</td>
<td>Ae. aegypti</td>
<td>♂</td>
<td>Distal lateral lobes of salivary glands</td>
<td>Constitutive</td>
<td></td>
<td>[130]</td>
</tr>
<tr>
<td>Ae-Act4</td>
<td>Ae. aegypti</td>
<td>♂</td>
<td>Indirect flight muscles</td>
<td>Pupae</td>
<td>Ae. albopictus</td>
<td>[117]</td>
</tr>
<tr>
<td>AeUbL40</td>
<td>Ae. aegypti</td>
<td>♂</td>
<td>Ovary/early larval</td>
<td>Constitutive</td>
<td></td>
<td>[118]</td>
</tr>
<tr>
<td>AePUB</td>
<td>Ae. aegypti</td>
<td>♂</td>
<td>Whole body</td>
<td>Constitutive</td>
<td></td>
<td>[118]</td>
</tr>
<tr>
<td>AaHsp70Aa</td>
<td>Ae. aegypti</td>
<td>♂</td>
<td>Head, midgut, salivary gland, and ovary</td>
<td>Constitutive and induced</td>
<td></td>
<td>[131]</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Promoter</th>
<th>Mosquito</th>
<th>Sex</th>
<th>Spatial Distribution</th>
<th>Temporal Distribution</th>
<th>Activity in Other Species</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>AaHsp70Bb</td>
<td>Ae. aegypti</td>
<td>♀</td>
<td>Head, midgut, salivary gland, and ovary</td>
<td>Constitutive and induced</td>
<td></td>
<td>[131]</td>
</tr>
<tr>
<td>Aal-Act4</td>
<td>Ae. albopictus</td>
<td>♀</td>
<td>Indirect flight muscles</td>
<td>Pupae</td>
<td>Ae. aegypti</td>
<td>[18]</td>
</tr>
<tr>
<td>Hex1.2</td>
<td>Ae. atropalpus</td>
<td>♀</td>
<td>Larval fat body</td>
<td>Developmental</td>
<td>Ae. aegypti</td>
<td>[132]</td>
</tr>
<tr>
<td>exu</td>
<td>Ae. aegypti</td>
<td>♀</td>
<td>Ovary</td>
<td>Developmental</td>
<td></td>
<td>[133]</td>
</tr>
<tr>
<td>trunk</td>
<td>Ae. aegypti</td>
<td>♀</td>
<td>Ovary</td>
<td>Developmental</td>
<td></td>
<td>[133]</td>
</tr>
<tr>
<td>nitro</td>
<td>Ae. aegypti</td>
<td>♀</td>
<td>Ovary</td>
<td>Developmental</td>
<td></td>
<td>[133]</td>
</tr>
<tr>
<td>Agper1</td>
<td>An. gambiae</td>
<td>♀</td>
<td>Midgut</td>
<td>Constitutive</td>
<td>An. stephensi, Ae. fluviatilis</td>
<td>[134]</td>
</tr>
<tr>
<td>β2-tub</td>
<td>An. gambiae</td>
<td>♂</td>
<td>Testes</td>
<td>Constitutive</td>
<td>An. stephensi</td>
<td>[135]</td>
</tr>
<tr>
<td>AsVg1</td>
<td>An. stephensi</td>
<td>♀</td>
<td>Fat body</td>
<td>Blood meal induced</td>
<td></td>
<td>[136]</td>
</tr>
<tr>
<td>AsAAPP</td>
<td>An. stephensi</td>
<td>♀</td>
<td>Salivary glands</td>
<td>Constitutive (high) + Blood meal induced</td>
<td></td>
<td>[137]</td>
</tr>
<tr>
<td>AgVgt2</td>
<td>An. gambiae</td>
<td>♀</td>
<td>Fat body</td>
<td>Blood meal induced</td>
<td></td>
<td>[138]</td>
</tr>
<tr>
<td>AgApy</td>
<td>An. gambiae</td>
<td>♀</td>
<td>Salivary glands</td>
<td>Constitutive (low)</td>
<td></td>
<td>[139]</td>
</tr>
<tr>
<td>AgVasa</td>
<td>An. gambiae</td>
<td>♀, ♂</td>
<td>Gonads</td>
<td>Constitutive</td>
<td></td>
<td>[140]</td>
</tr>
<tr>
<td>Gene</td>
<td>Species</td>
<td>Sex</td>
<td>Tissue/Location</td>
<td>Expression Type</td>
<td>Target Species</td>
<td>Reference(s)</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------</td>
<td>------</td>
<td>----------------------------------</td>
<td>-----------------------</td>
<td>-------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>AgG12</td>
<td>An. gambiae</td>
<td>♀</td>
<td>Midgut</td>
<td>Blood meal induced</td>
<td>An. stephensi</td>
<td>[141]</td>
</tr>
<tr>
<td>Antryp1</td>
<td>An. gambiae</td>
<td>♀</td>
<td>Midgut</td>
<td>Nonresponsive to</td>
<td>An. stephensi</td>
<td>[141]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>blood meal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>α1-tub1b</td>
<td>An. gambiae</td>
<td>♀</td>
<td>Head, ventral nerve cord, and</td>
<td>Constitutive</td>
<td></td>
<td>[142]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>chordontal organs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AcCP</td>
<td>An. culicifacies A</td>
<td>♀</td>
<td>Midgut</td>
<td>Blood meal induced</td>
<td></td>
<td>[143]</td>
</tr>
<tr>
<td>Act88F</td>
<td>D. melanogaster</td>
<td>♀, ♂</td>
<td>Pupae and adult flight muscles</td>
<td>Developmental</td>
<td>Cx. quinquefasciatus Say</td>
<td>[144]</td>
</tr>
<tr>
<td></td>
<td>Synthetic</td>
<td>♀, ♂</td>
<td>Neural specific, eye, anal</td>
<td>Constitutive</td>
<td>Many insect species</td>
<td>[115]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>papillae</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 × P3</td>
<td>Synthetic</td>
<td>♀, ♂</td>
<td>Muscle</td>
<td>Constitutive</td>
<td>An. stephensi, Ae. aegypti</td>
<td>[135]</td>
</tr>
<tr>
<td>Actin5c</td>
<td>D. melanogaster</td>
<td>♀, ♂</td>
<td>Muscle</td>
<td>Constitutive</td>
<td>An. stephensi, Ae. aegypti, An. gambiae</td>
<td>[19,116,145]</td>
</tr>
<tr>
<td>IE-1</td>
<td>Baculovirus</td>
<td>♀, ♂</td>
<td>Whole body</td>
<td>Constitutive</td>
<td>An. stephensi, Ae. aegypti, An. gambiae</td>
<td>[50,146]</td>
</tr>
</tbody>
</table>
Practically speaking then, experimenters should proceed with caution when trying to identify additional promoters for driving transgene expression in mosquitoes. RNA-seq data, now abundantly available for both *An. gambiae* and *Ae. aegypti* can be used to identify multiple candidate genes with the desired expression pattern [133,141,148]. The candidates can be compared using parameters such as (i) distance to nearest gene (particularly if said gene/s have a different expression pattern); (ii) length of first intron (so 5′-UTR sequence up to the initiation codon can be included in test constructs); and (iii) presence/distribution of both unique and repetitive sequences upstream of transcription initiation. Ideally, a candidate test construct can be chosen for situations where functional elements are constrained to the genomic DNA fragment under investigation. For example, the *Ae. aegypti* 30 K [130] and Hsp70 [131] as well as the *An. gambiae* trpy1 [141] gene promoter fragments tested were constrained by the presence of an upstream gene in close proximity.

**Bipartite Tet Systems**

*Introduction.* Bipartite systems have been developed so that two independent events are required to control the expression of a transgene. As the name suggests, bipartite systems are comprised of two genetic components: a “driver” that expresses a specific exogenous transcription factor and a “responder” that employs a promoter that is exclusively activated only by that particular transcription factor.

*How it works.* The most common and well-established bipartite system in mosquitoes involves the activation of gene expression when the antibiotic tetracycline (or its analog, doxycycline) is withheld; this system is known as Tet-Off. Conversely, the opposing Tet-On system specifically activates gene expression in the presence of tetracycline (Tc). Both the Tet-Off and Tet-On systems take advantage of the naturally occurring negative-regulation of resistance-mediating genes by the tetracycline repressor (TetR) found in the bacteria *Escherichia coli*, specifically the regulatory components of the Tn10 tetracycline-resistance operon [149]. It was found that the TetR bound tetracycline more tightly than a previously used lacR complexed with IPTG of the lacR system. This allowed for employment of very low, nontoxic intracellular concentrations of tetracycline that still functioned efficiently. The TetR also shows very specific binding to the Tet operator (TetO) sequence, resulting in a reduction of leaky expression. Doxycycline also has better cell and tissue penetrative properties than IPTG, the inducer for the lacR system [150]. In the Tet-Off system, the TetR is fused with the activation domain of VP16, a protein encoded by herpes simplex virus involved in the triggering of the lytic phase of the viral infection from the latent, dormant state. The resultant chimeric tetracycline transactivation protein, referred to as tTA, binds to the TetO sequence, resulting in the recruitment of the transcriptional machinery and substantially increasing the expression of the gene under its control. When tetracycline is present in the diet of the mosquito, it binds...
to the tTa protein, thereby preventing DNA binding and gene activation. In the
Tet-On system, the reverse tetracycline transactivation protein can only recognize
and activate the TetO sequence in the presence of doxycycline. Initially used in
mosquitoes by Lycett et al. [151], the Tet-Off system has been utilized as a key
component of the transgenic strains used by Oxitec Limited for the genetic con-
trol of *Ae. aegypti*. Phuc et al. [19] demonstrated that the employment of the
Tet-Off system linked to lethal expression of tTAV (an optimized variant of tTA
sequence) that serves as both a transactivator and an effector can result in late-
acting lethality [19]. By killing mosquitoes at the fourth instar larval/pupal bound-
ary, this strategy retains density-dependent selection while preventing the emer-
gence of new adult mosquitoes. Functionally, this system uses Tc to prevent
tTAV from activating transcription from TetO that would otherwise consequently
express more tTAV in a positive feedback loop that would proceed uncontrolled
until the death of the mosquito (Figure 8.5A). While transgenic females must be

![Figure 8.5](image)

**FIGURE 8.5** Tetracycline-repressible control of transgene expression in OX531A and
OX3604C mosquitoes. While rearing in the factory, the presence of tetracycline (tet) or its ana-
log doxycycline binds to tTAV and prevents gene expression from TetO. Once released, the
tTAV is free to bind TetO, activating expression of either additional tTAV or the VP16 transacti-
vator, the accumulation of which is toxic to cells.

1. Any discussion of the transgenic constructs developed and used by Oxitec Limited for the
genetic control of dengue vectors must first and foremost acknowledge that this company has
voluntarily disclosed via publication all of the experimental details concerning how each mos-
quito strain was modified, though there is no legal requirement that they do so. This level of
transparency is in line with recent WHO recommendations [152] for testing genetically modified
mosquitoes and allows the scientific community and public at large the opportunity to evaluate
each transgenic mosquito strain on its own merits.
removed manually prior to field releases, the resultant strain (referred to as LA513 in Ref. [19], OX513A in Refs [153,154]) was the first transgenic mosquito used in an open-field release [153].

The transgene construct employed in another transgenic strain of *Ae. aegypti*, OX3604C, incorporates sex-specific splicing, under control of the *Ae-Act4* promoter [117] in combination with the tTAV/Tc system to control a female-specific flightless phenotype. In the presence or absence of Tc, the tTAV is not produced in males due to a combination of minimal promoter activity and an in-frame termination codon in the male splice form (Figure 8.5B). However, in females strong sex-specific expression from the promoter combined with correct splicing of the tTAV mRNA results in sufficient expression of the VP16 transactivator in the indirect flight muscles to render them nonfunctional [117]. OX3604C was successfully tested in large cage trials [155] but has not yet been used in open release. Similar flightless female transgenic strains have been developed and validated for the Asian tiger mosquito, *Ae. albopictus* [18] and the malaria vector *An. stephensi* [156]. Both of these strategies illustrate the potential for differential control of transgene expression and are particularly suited to cases where the expression of the transgene is expected to be detrimental to the health and/or survival of the mosquito.

**Strengths.** The nature of this approach allows for not only the ability to induce gene expression but also to limit its expression to specific temporal and spatial profiles. A distinct advantage of this approach allows for the scrutiny of genes whose expression may result in lethality, sterility, or a high fitness effects, as these effects will manifest only when both components are present. Another degree of control is afforded by systems that include an exogenous chemical component acting upon the expressed transgenic transactivation factor.

**Weaknesses.** As demonstrated by Phuc et al. [19], Fu et al. [117], and Labbe et al. [18] many transgenic lines must be generated and evaluated to obtain a transgenic strain that possesses the correct blend of restricted expression in the presence of Tc and sufficient activation upon removal of Tc to achieve the desired phenotypic switch. Once established, there is also the economic cost associated with treating large volumes of larval rearing water with Tc/doxycycline when transgenic mosquitoes are reared at large scale.

**Other Bipartite Expression Systems**

Other bipartite systems such as the yeast GAL4–UAS and the Neurospora Q systems may prove to be extremely valuable as a means to restrict transgene expression in candidate strains for genetic control programs. As opposed to the Tet-On/Off system, neither of these systems depends upon an exogenous chemical to allow for the repression/suppression of the switch. The promoter-dependent expression of the Gal4 yeast transactivator forms the driver line
component whereas the responder lines are generated to include a gene of interest under the transcriptional control of upstream activation sequences (UAS). When both components are present, Gal4 protein binds to the UAS permitting transcription of the candidate gene. Furthermore a natural suppressor of Gal4, Gal80, can provide an inhibitory effect allowing for more precise or complex investigations. This system has been successfully used in a number of different mosquito species including An. gambiae [157], An. stephensi [16], and Ae. aegypti [158,159]. The Q system involves genes identified from filamentous fungi Neurospora crassa whose qa gene cluster catabolizes quinic acid when glucose levels are limited. Essentially, quinic acid controls expression of the genes required for its own catabolism. The adapted system (similar in component nature to the Tet-On/Off system) utilizes the transcription factor, QF and its repressor, QS. In normal conditions, when glucose is abundant, QF is inhibited by QS thereby arresting expression of the qa gene cluster. However, when glucose is limited, the qa gene cluster expresses quinic acid that then binds QS in turn allowing QF to promote expression from the qa genes. Further, coupling together both the Gal4 and the Q system can attain even more precise control [160].

CONSIDERATIONS AND OUTLOOK

In the near term, first (transposon-based) and second (transposon + ΦC31) generation transgenic approaches will likely remain the dominant technologies used in genetic control strategies for both malaria and dengue. Indeed, all current and proposed genetically-modified mosquito trials involve these manipulations, and due to the amount of resources already invested, these will likely continue. However, with the advent of affordable site-specific gene editing of mosquito genomes, a range of manipulations not previously considered become manageable. For example, the entire need for promoter validation experiments as described above almost becomes unnecessary when a gene of interest can be recombined into the genome to be directly controlled by the native promoter. This can be done in a manner that is destructive to the original gene (replacement) or nondestructive (by recombining into an intron with a competing splice acceptor site). Site-specific recombination could also be used to insert or remove enhancer/repressor elements, thus changing the expression pattern of native mosquito genes in a beneficial manner without introducing any non-native DNA. Site-specific genome editing technologies provide new and powerful methods for stabilizing transgenes, through the generation of micro- or macroinversions [161], as well as new methods for performing gene drive [162,163]. Thus, while TE-based approaches will likely remain useful for the foreseeable future as a basic research tool, the precision offered by CRISPR-induced homologous recombination is unrivaled and may well-dominate the field of genetics-based control of diseases such as malaria and dengue in the coming years (Figure 8.6).
REFERENCES


FIGURE 8.6 Timeline of genome modification and manipulation tools in mosquitoes. Since the initial reports of cut-and-paste transposition in mosquitoes in the late 1990s, an increasing number of genetic tools have been validated including the integrase ΦC31, homing endonucleases (HEGs), zinc-finger nucleases (ZFNs), transcription-activator-like nucleases (TALENs), homologous recombination (HR), and CRISPR/Cas based gene editing.


Chapter 9

Gene Drive Strategies for Population Replacement

John M. Marshall1 and Omar S. Akbari2
1Division of Biostatistics, School of Public Health, University of California, Berkeley, CA, USA,
2Department of Entomology, University of California, Riverside, CA, USA

INTRODUCTION

After 3.8 billion years of research and development, Nature has provided inspiration for a plethora of human design problems. During the Renaissance, Leonardo da Vinci designed a flying machine inspired by the anatomy of birds. Today, Nature’s evolutionary solutions are informing the design of solar panels from photosynthesis, and digital displays using the light-refracting properties of butterfly wings. Nature’s intricate structures and processes may also help in the fight against mosquito-borne diseases. Gene drive—the process whereby natural mechanisms for spreading genes into populations are used to drive desirable genes into populations (e.g., genes conferring refractoriness to malaria or dengue fever in mosquitoes)—is another example of Nature’s processes being applied for the benefit of humanity. Gene drive systems may either spread from low initial frequencies or display threshold properties such that they are likely to spread if released above a certain frequency in the population and are otherwise likely to be eliminated.

Population replacement, in this context, refers to the process whereby a population of disease-transmitting mosquitoes is replaced with a population of disease-refractory ones. Several approaches are being explored to engineer mosquitoes unable to transmit human diseases, and there have been a number of notable successes. For example, Isaacs et al. have engineered Anopheles stephensi mosquitoes expressing single-chain antibodies that prevent Plasmodium falciparum malaria parasites from developing in the mosquito, thus preventing onward transmission of the parasite [1]. Gene drive systems are expected to be instrumental in spreading disease-refractory genes into wild mosquito populations, given the wide geographical areas that these species inhabit and the expectation that refractory genes will be associated with
at least modest fitness costs [2]. Gene drive systems are also being considered to implement population suppression strategies whereby genes conferring a fitness load or gender bias are instead driven into the vector population, thereby reducing disease transmission.

**Early Inspiration**

Initial suggestions for spreading desirable genes into insect pest populations date back to the early 1940s and involved the proposition of translocations [3,4] and transposable elements (TEs) [5], inspired from natural systems. Translocations are rearrangements of parts between nonhomologous chromosomes. If insects homozygous for a translocation are introduced into a population at high frequency, they are predicted to spread to fixation [6], and if the translocation is linked to a disease-refractory gene, it is predicted to consequently be driven into the population as well. Initial field trials with translocations were unsuccessful in demonstrating spread [7]; but this is likely a result of those translocations being generated using X-rays, which often induce high fitness costs.

The suggestion of using TEs to drive disease-refractory genes into mosquito populations was largely inspired by the observation that a TE known as the P element spread through most of the global *Drosophila melanogaster* population within the span of a few decades following natural acquisition from *Drosophila willistoni* [8]. TEs are able to spread through a population due to mechanisms that enable them to increase their copy number within a host genome and hence to be inherited more frequently in subsequent generations. As a result, they are able to spread into a population from very low initial frequencies even if they incur a fitness cost to their host [9]. It was hoped that the P element invasion of *Drosophila* could be repeated in disease-transmitting mosquito species using a TE attached to a disease-refractory gene; however, early laboratory work on TEs in mosquito vector species has failed to identify elements with high remobilization rates following integration into mosquito lines [10].

**Promising New Systems**

Two of the most promising gene drive systems at present also involve technologies inspired by Nature—the use of homing endonuclease genes (HEGs) observed to spread in fungi, plants, and bacteria [11], and a selfish genetic element known as Medea observed to spread in *Tribolium* beetles [12,13]. A synthetic Medea element has been developed in *Drosophila* that works by the hypothesis that Medea encodes both a maternally expressed toxin and a zygotically expressed antidote [14]. This combination results in the death of wild-type offspring of Medea-bearing mothers, thus favoring the Medea allele in subsequent generations and mimicking the behavior of the natural element.
in *Tribolium*. *Medea* was the first synthetic gene drive system to be developed and has a number of desirable design features; however, significant work is still ongoing to develop a *Medea* element in a mosquito disease vector.

Recently, there has been much excitement around HEGs as, while *Medea* was first engineered in *Drosophila*, a naturally occurring HEG has been shown to spread in a laboratory population of *Anopheles gambiae*, the main African malaria vector, containing an engineered target sequence for the HEG [15]. HEGs spread by expressing an endonuclease that creates a double-stranded break at specific target sequences lacking the HEG. Homologous DNA repair then copies the HEG to the cut chromosome, increasing its representation in subsequent generations. Similar to the aforementioned gene drive systems, HEGs are being considered to drive disease-refractory genes into mosquito populations; however, a number of additional strategies for their application are also being considered, which aim to suppress rather than replace mosquito populations [11], and progress has been made toward these ends as well [16].

**Design Criteria**

As the technology for developing gene drive systems for population replacement develops on a number of fronts, it is useful to consider design criteria for assessing the safety and efficacy of the various approaches. An excellent review by Braig and Yan [17] proposes several biological properties that an ideal gene drive system should or must have:

1. The gene drive system must be effective. That is, it must be strong enough to compensate for any loss in host fitness due to the presence of both itself and its transgenic load (manifest as a reduction in host fertility, life span, or competitiveness). It must be able to spread to very high frequency in a population on a timescale relevant to disease control (i.e., a few years) and must be unimpeded by wild-type vectors immigrating into the target area.

2. The gene drive system must be able to carry with it several large genes and associated regulatory elements. At the very least, a disease-refractory and marker gene will be needed along with regulatory elements; but multiple disease-refractory genes are preferable in order to slow the rate at which the pathogen evolves resistance to each of them.

3. Features should be included to minimize the rate at which linkage is lost between the drive system and disease-refractory genes, as even rare recombination events could be significant for wide-scale spread over a long time period.

4. It should be possible to use the gene drive system to introduce waves of refractory genes over time to counteract the effects of evolution.
of pathogen resistance, mutational inactivation of the refractory gene, or loss of linkage between the refractory gene and drive system.

5. The gene drive system should be easily adapted to multiple vector species. Human malaria, for instance, is transmitted by approximately 50 species of mosquitoes belonging to the genus *Anopheles*. In sub-Saharan Africa, the most important transmitters are *An. gambiae*, *Anopheles coluzzii*, *Anopheles arabiensis*, and *Anopheles funestus*, ideally all of which should be rendered refractory in a population replacement strategy.

Additional features of an ideal gene drive system were proposed by James to address ecological, epidemiological, and social issues, including safety [2]. Safety is a broad criterion that should be assessed through risk assessment in which potential hazards are identified along with their corresponding magnitudes and likelihoods. This provides a framework for managing the most significant risks and for the overall safety of the system to be scored. However, prior to a comprehensive risk assessment, a few general safety criteria for gene drive systems can be imagined.

6. The behavior of the gene drive system in the target species should be stable and predictable, thus minimizing the likelihood of unpredictable side effects in target species.

7. A mechanism should be available to prevent horizontal transfer of the gene drive system and/or refractory gene to nontarget species, thus minimizing the wider ecological impact of the release.

8. The gene drive system and refractory gene should not cause undesirable effects for human health, for instance, by selecting for increased virulence in the pathogen population. The gene drive system should also include a mechanism for removing the refractory gene from the population in the event of any adverse effect.

9. The gene drive system must be consistent with the social and regulatory requirements of the affected communities. For instance, public attitude surveys in Mali [18] highlight the importance of confined field trials prior to a wide-scale release, which could be achieved through the initial use of gene drive systems with high release thresholds followed by subsequent releases with more invasive systems.

10. The gene drive system should be cost-effective, as budgets for disease control are limited and a number of alternative interventions are available. The initial development of gene drive systems is expensive; but ongoing investment can be minimized by designing systems that are resilient to evolutionary degradation.

Cost-effectiveness is an important consideration, as it is not only relevant to the choice of gene drive system, but to whether gene drive should be used at all. In a recent modeling study, Okamoto et al. demonstrated the economic feasibility of releasing large numbers of insects carrying a dengue-refractory gene
without a gene drive system in order to reduce the dengue transmission potential of *Aedes aegypti* mosquitoes in Iquitos, Peru [19]. Wide-scale control of *Anopheles* malaria vectors in sub-Saharan Africa is less likely amenable to the mass release strategy; however, it is essential to assess this in terms of efficacy, safety, and cost-effectiveness prior to implementation.

In this chapter, we review a range of gene drive systems being considered to drive disease-refractory genes into mosquito vector populations. We divide gene drive systems into two broad categories: (i) those that spread by causing a double-stranded break at a specific target sequence and insert themselves at this location through DNA repair (e.g., HEGs) and (ii) those that use combinations of toxins and antidotes, active at different life stages, to favor their own inheritance (e.g., Medea). We also review modern approaches to developing translocations as forms of gene drive, which do not fit into either category. Systems using symbiotic or commensal microorganisms to mediate gene drive are covered in another chapter (e.g., *Wolbachia*). For each system, we review the biological mechanisms involved, the system’s current stage of development, and its alignment with the aforementioned design criteria.

**GENE DRIVE SYSTEMS THAT SPREAD VIA TARGET SITE CLEAVAGE AND REPAIR**

We begin by reviewing gene drive systems that manipulate inheritance in their favor by causing a double-stranded break at one or more specific target sites in the host’s genome and utilize the host’s homologous DNA repair mechanism to increase their genomic copy number. Gene drive systems of this type include TEs, HEGs, and a number of recently proposed HEG analogs, such as zinc-finger nucleases (ZFNs), transcription-activator-like effector nucleases (TALENs), and clustered, regularly interspaced, short palindromic repeats (CRISPRs).

**Transposable Elements**

TEs are genomic components capable of changing their position and sometimes replicating within a genome. Consequently, they show widespread prevalence throughout the genomes of many taxa, with various families of TEs accounting for ~90% of the Salamander genome, 50% of the *Ae. aegypti* genome, and 45% of the human genome. There are various classes of TEs, and those being considered for population replacement in mosquitoes belong to class 2. Class 2 elements contain both repeat sequences that mark their boundaries and their own transposase gene that catalyzes transposition. They move via a cut-and-paste mechanism [20], whereby transposition results in excision of the TE via two double-stranded breaks, leaving behind a gap where they have been excised. In some cases, this gap is filled by
homologous gap repair from a chromatid also having the TE. The excised TE is then inserted at another genomic location, resulting in their genomic copy number being increased by one. In a second replication mechanism, some TEs transpose during the S phase of the cell cycle. If a recently replicated element transposes to an unreplicated region of the genome, it will be replicated a second time, resulting in a net gain of one element in the genome.

**Current Status.** The widespread distribution of TEs in Nature together with observations of the rapid spread of the P element in *Drosophila* [8] inspired initial hopes that class 2 TEs could be inserted, along with disease-refractory and marker genes, into transgenic lines of *Ae. aegypti* (the main vector of dengue fever) and *Anopheles* vectors of malaria. Class 2 TEs lacking their transposase gene are often used as vectors for introducing novel genes into mosquitoes; hence, integration into mosquito lines is relatively straightforward. More problematic, however, has been the remobilization of TEs containing their own transposase gene once they have been integrated. An excellent review by O’Brochta et al. describes results from experiments in which four class 2 TEs—*Hermes*, *Mos1*, *Minos*, and *piggyback*—were used to create transgenic lines of *Ae. aegypti* [10]. In all cases, remobilization was shown to be highly inefficient. More recently, attempts were made to improve the post-integration mobility of *Hermes* in *Ae. aegypti* using an additional construct to express a transposase gene under the control of a testis-specific promoter [21]; however, remobilization was still only observed in less than 1% of the transgenic lines.

**Design Criteria.** The observed remobilization of natural TEs suggests that remobilization of introduced elements should also be possible; however, the regulation of TE mobility is complex, and it may require much experimentation to find TEs compatible with mosquito vectors. This work is likely not cost-effective, as TEs fail to satisfy most of the design criteria outlined earlier, and have been superseded by more recently proposed systems like HEGs and Medea. Of particular note, it is unlikely that TEs will be able to carry large inserts containing disease-refractory genes as transposition events are known to be imprecise and prone to DNA loss. Furthermore, a study on the Himar1 mariner element suggests that transposition rates decline substantially with increasing insert size [22], suggesting that elements which have lost their transgenic load will outspread those which have not [23]. Finally, the large numbers of target sites that TEs have undermine their predictability and stability in target species, and their wide species host range highlights the risk of horizontal gene transfer and spread in nontarget species.

**Homing Endonuclease Genes**

HEGs are highly efficient selfish genetic elements that spread by expressing an endonuclease that recognizes and cleaves a highly specific target sequence of 14–40 base pairs usually only present at a single site in the host
As the HEG is positioned directly opposite its target site, actually within its own recognition sequence, it induces a double-stranded break only in chromosomes lacking the HEG. The HEG is effectively copied to the target site, in a process referred to as “homing,” when the cell’s repair machinery uses the HEG-bearing chromosome as a template for homology-directed repair. When homing occurs in the germ line of the host organism, a HEG can be transmitted to progeny at a higher than Mendelian inheritance ratios, enabling its spread through a population (Figure 9.1A).

On the basis of observations of homing activity in a number of nonmetazoan organisms including yeast, fungi, algae, and plants, Burt proposed that HEGs could be used as a gene drive system for population replacement in mosquito disease vectors; however, he also proposed and favored their use as a population suppression system [11]. Burt proposed a suite of HEG-based

![Diagram of homing and X-shredder HEGs](image)

**FIGURE 9.1** Preferential inheritance of homing-based gene drive systems. (A) Left panel: A homing HEG (green) encodes an endonuclease that recognizes and cleaves a specific target sequence (red) on a wild-type chromosome (step 1). Once the target site is cleaved, the cell repairs the chromosomal break through homologous recombination using the HEG-bearing chromosome as a template (step 2). This two-step process results in the HEG effectively being copied to the wild-type chromosome in a process referred to as “homing,” thereby converting a HEG heterozygote into a HEG homozygote. Right panel: When a HEG-bearing male (green mosquito) is released into the wild and mates with a wild-type female (gray mosquito), the majority of their progeny inherit the HEG, and over time the HEG invades entire populations. (B) Left panel: For HEG-based population suppression, an X-shredder HEG is positioned on the Y chromosome (Y-linked X-shredder HEG). This HEG encodes an endonuclease that recognizes and cleaves chromosomal sequences that are repeated exclusively on the X-chromosome of the mosquito. When expressed during spermatogenesis, X-bearing spermatids are disrupted by the HEG, resulting in the majority of functional sperm being Y-bearing and containing the HEG. Right panel: When a Y-linked X-shredder HEG-bearing male (green mosquito) is released into the wild and mates with a wild-type female (gray mosquito), all resulting progenies are HEG-bearing males. Over time, this is predicted to induce an all-male population crash and potentially eventual extinction of the vector species.
strategies for genetic control of mosquito vectors—two involving population replacement and three involving population suppression:

1. First, the HEG could be linked to a disease-refractory gene and engineered to target a gene-sparse region of a chromosome (so as to reduce impacts on mosquito host fitness), thus carrying the disease-refractory gene with it as it spreads into the population.

2. In a related population replacement approach, the HEG could be engineered to target an endogenous gene involved in the development or transmission of the pathogen, thus reducing vector competence as it spreads [25]. This approach has the benefit that it does not involve an effector gene and hence is more resilient to evolutionary degradation; however, it does require a gene to be identified, the disruption of which would block pathogen transmission, and for a HEG to be engineered to target this, which is quite arduous.

3. In terms of population suppression, a HEG could be engineered that targets a native mosquito gene required in at least one copy for either mosquito survival or fertility. If a HEG of this type is active in the mosquito germ line, then it will increase in frequency in the population, inducing a genetic fitness load on the population as it spreads. This could lead to either population suppression or an eventual population crash.

4. An alternative to the homing-based applications of HEGs is to rely entirely on their target site cleavage activity. In the first of these approaches, known as the “autosomal X-shredder” strategy, a HEG can be designed to specifically cleave the X chromosome at multiple locations, effectively destroying it. If an X-shredder HEG is expressed during male meiosis, it will result in destruction of X-bearing male sperm. If females mate with males having the X-shredder, most viable sperm will be Y-bearing and hence most of the progeny will be male. This strategy will reduce the reproductive potential of the population; but it requires regular releases since the X-shredding gene is associated with a fitness cost and will only persist in the population for a few generations.

5. Finally, Burt proposed a “Y-linked X-shredder” strategy whereby, if the X-shredder HEG is located on the Y chromosome, then it will be driven into the population along with the transgenic Y chromosome as it induces an increasingly male gender bias. This approach would mimic naturally existing meiotic drive systems that bias sex ratios, although it could potentially induce a much larger gender bias than those observed in Nature [26–28], causing a cascade of male-only population crashes that could potentially lead to species extinction (Figure 9.1B).

Current Status. An encouraging result for homing-based HEG strategies has been the engineering of a naturally occurring HEG, I-SceI, which has been shown to cleave in Ae. aegypti [29] and spread in laboratory populations of both D. melanogaster and An. gambiae containing an engineered
target sequence for the HEG [15,30,31]. These results are encouraging because they show that, although HEGs have not been discovered in any metazoan species to date, there is nothing intrinsic about metazoan biology that prevents HEGs from homing. Furthermore, the fact that this was achieved in An. gambiae, the most important African malaria vector, is hopeful for its application to disease control. For the population replacement strategy to work in the wild, a HEG must be engineered or identified which has a target sequence in the wild mosquito genome. Engineering HEGs to recognize and cleave new target sequences has proven difficult thus far [32–34], and future research should focus on the development of novel approaches to circumvent these difficulties.

Population suppression strategies that rely solely on the target site cleave activity of HEGs have shown remarkable progress in recent years. A HEG originally discovered in the slime mold Physarum polycephalum, I-Ppo1 [35], was integrated into the An. gambiae genome and shown to recognize and cleave a conserved DNA sequence, repeated hundreds of times and located exclusively on the X chromosome cluster of ribosomal DNA genes in An. gambiae [36]. This cleavage activity is highly applicable to both the autosomal and Y-linked X-shredder strategies of HEG-driven population suppression and has also provided a novel genetic approach to the sterile insect technique for An. gambiae. The expression of I-Ppo1 during spermatogenesis in An. gambiae resulted in cleavage of the paternal X chromosome in differentiating spermatozoa, which was expected to result in a male bias among progeny. However, it turned out that the I-Ppo1 from mature sperm cells was carried over into the zygote, thus shredding the zygotic X chromosomes as well and rendering the transgenic males completely sterile [37]. It was later shown that transgenic mosquitoes engineered with I-Ppo1 could induce high levels on sterility in large cage populations, confirming the suitability of this technology for use in sterile insect population suppression programs [38]. This could be a useful first application of HEG technology in the wild given the self-limiting nature of sterile insect releases.

For X-shredder strategies to work, I-Ppo1 would need to be destabilized in order to minimize its carryover into the zygote by mature sperm. To this end, recent work by Galizi et al. has succeeded in expressing destabilized autosomal versions of I-Ppo1, which result in efficient shredding of the paternal X chromosome and are restricted to male meiosis [16]. Consequently, males carrying this construct are fully fertile and some insertions produce >95% male offspring bias. Males inheriting the autosomal I-Ppo1 gene also produce a male bias in their progeny, showing that the gender-biasing effect of autosomal X-shredders will remain in the population for several generations; however, continued releases would be required, as the X-shredder gene is not favored through inheritance when located on an autosome and is expected to be eliminated due to fitness costs. Nevertheless, for repeated releases, population suppression is expected, which would be
more efficient than the previously mentioned sterile male releases and would also be self-limiting, albeit over a longer period. Autosomal X-shredders could therefore be an appropriate second application of HEG technology.

The only remaining steps in order to realize the Y-linked X-shredder strategy are to dock the destabilized I-Ppo1 HEG onto the \textit{An. gambiae} Y chromosome and ensure that it is expressed during spermatogenesis. To this end, recent progress has been made in developing a Y chromosome docking line in \textit{An. gambiae} \cite{39}. Future work will focus on docking the HEG onto the Y chromosome and ensuring it can be expressed and function as anticipated.

\textit{Design Criteria.} HEG-based strategies for genetic control of vector-borne diseases are extremely promising given the remarkable progress made recently, most notably in the malaria vector \textit{An. gambiae}. HEGs are highly effective as a gene drive system, capable of spreading for low initial frequencies to high frequency on a short timescale. They are also relatively short sequences targeting very precise regions of the genome, suggesting both stability and a low rate of corruption due to evolutionary degradation. Species-specific regulatory sequences can be included to limit their horizontal transfer to nontarget species, and furthermore, a strategy has been proposed to reverse the spread of a deleterious HEG through the release of HEG-resistant alleles in the event of unforeseen consequences \cite{11}. Additionally, a wide range of HEG strategies are available displaying different levels of confinability, allowing them to be used at all stages of a phased release and to be tailored to the social and regulatory requirements of affected communities.

Target site cleavage strategies show more promise than those reliant on homing activity as they sidestep many of the abovementioned design criteria and are independent of disease-refractory genes. Target site mutagenesis and gap repair through nonhomologous end joining can both result in disruption of the HEG cleavage site, rendering certain individuals immune to the HEG and preventing the HEG from spreading through an entire population. For strategies in which a HEG disrupts a gene required for mosquito survival or fertility, HEG-resistant mutants will be favored in a population once they emerge. Furthermore, there is a possibility of losing the disease-refractory gene either through mutagenesis or during homology-directed repair—a concern that becomes more serious for larger inserts, and would render a population replacement strategy futile. The Y-linked X-shredder strategy is less vulnerable to target site mutagenesis as it targets so many loci on the X chromosome at once. It is, however, dependent on germ line gene expression on the \textit{An. gambiae} Y chromosome although this could potentially be achieved through the use of insulator sequences.

\textbf{TALENs and ZFNs}

TALENs and ZFNs have been proposed as alternative platforms for engineering homing-based gene drive systems \cite{40}—that is, systems that spread
by cleaving a specific target sequence and then using the cell’s repair machinery to copy themselves to the target site. The benefit of TALENs and ZFNs over HEGs is that they can be easily engineered to target desired DNA sequences due to the modular nature of their DNA-binding domains. TALENs are derived from naturally occurring proteins that are secreted by the pathogenic bacteria *Xanthomonas* spp. to alter gene expression in host plant cells [41,42]. These proteins contain arrays of highly conserved, repetitive DNA-binding domains, each recognizing only a single base pair, with specificity being determined by repeat-variable di-residues [43,44]. The relationship between these repeats and DNA recognition can be exploited to design TALENs that target virtually any desired DNA sequence. For ZFNs, DNA-binding specificity can be similarly manipulated, being determined by an array of finger modules that can be generated either by selection using large combinatorial libraries, or by rational design [45].

For both TALENs and ZFNs, DNA-binding modules can be combined with several types of domains, including transcriptional activators, nucleases, and recombinases, allowing for a comprehensive range of genetic modifications [46]. In terms of cleavage activity, a wide range of tailored recognition sequences can be cleaved efficiently as TALENs and ZFNs are fusion proteins consisting of a nonspecific fokI nuclease linked to a DNA-binding motif [47,48]. The TALEN or ZFN may then be copied to the cleaved target side by homology-directed repair, and hence used as a gene drive system for driving disease-refractory repair, and hence used as a gene drive system for driving disease-refractory genes into mosquito populations.

**Current Status and Design Criteria.** Both TALENs and ZFNs rely upon homing activity and thus, for the purposes of population replacement and control, are functionally similar to HEGs. Given this similarity, the range of replacement and suppression strategies outlined earlier is also applicable to these systems and many of the design issues are similar too. For example, TALENs and ZFNs are also expected to spread from low initial frequencies, species-specificity can be incorporated through the addition of regulatory elements, and a deleterious TALEN or ZFN can be removed from a population through the release of TALEN- or ZFN-resistant alleles. However, there are some important differences. In terms of cost-efficiency, both TALENs and ZFNs are easier to engineer to target specific DNA sequences, and consequently, they could be straightforwardly adapted to multiple vector species, which is particularly important for malaria control. However, concerns arise regarding their stability, as their repetitive nature makes them more prone to mutation and evolutionary degradation. Recent progress toward developing both TALEN- and ZFN-based gene drive systems in *D. melanogaster* have successfully demonstrated DNA-binding specificity, cleavage, and homing through homology-directed DNA repair; however, mutational inactivation led to a decline in effectiveness over just a short period of time [40]. Thus, if TALENs or ZFNs are to be useful as gene drive systems in the future, their stability issues must first be overcome.
Clustered, Regularly Interspaced, Short Palindromic Repeats

CRISPR is another promising system proposed, although not yet demonstrated, as an alternative platform for homing-based gene drive. The system is based on an adaptive immune process in bacteria whereby sequences derived from invading bacteriophages or plasmids are integrated into the bacterial CRISPR locus. This essentially provides bacterial cells with the ability to “remember” and protect themselves against previously encountered viral genomes and invasive, mobile genetic elements [49]. To perform nuclease activities, CRISPR systems use an array of CRISPR RNAs (crRNAs) derived from exogenous DNA targets (e.g., viral genomes), noncoding transactivating RNAs, and a cluster of CRISPR-associated (Cas) genes. Three types of CRISPR systems have been discovered, with type II CRISPR systems being best characterized. These consist of a Cas9 nuclease and a crRNA array encoding guide RNAs and auxiliary transactivating crRNAs to mediate target site cleavage [50]. As for the homing-based systems described earlier, if the double-stranded break is repaired by homology-directed repair, the CRISPR system may be copied to the cleaved target site and hence used as a gene drive system for population replacement similar to HEGs. If the target site cleavage activity is directed toward the X chromosome, then the population suppression strategies initially described for HEGs could also be realized.

Current Status. Recent encouragement for CRISPR-based gene drive has been provided by proof-of-principle studies showing that the type II CRISPR system from *Streptococcus pyogenes* can be modified to target endogenous genes in bacteria [51] and human cell lines [52,53]. It has subsequently been shown that CRISPR can be used to alter genes in a range of other species including insects such as *D. melanogaster* [54,55] and mosquitoes. Straightforwardly, utilizing this system in other organisms requires only two components—the Cas9 nuclease and guide RNAs [52,56]. DNA-binding specificity is determined by the first 20 nucleotides of the guide RNA as these designate the DNA target side that Cas9 will be guided to according to Watson–Crick DNA–RNA base pairing rules. The only restriction for the target site selection is that it must lie directly upstream of a protospacer adjacent motif sequence that matches the canonical form 5′-NGG. Aside from that, it is possible that the CRISPR system can be engineered to target and cleave essentially any genomic location, with subsequent homing and gene drive occurring via homology-directed repair, however this remains to be demonstrated.

Design Criteria. CRISPR-based gene drive has yet to be implemented; however, its mechanisms imply that the approach is achievable. In terms of design criteria, the system is very similar to TALENs and ZFNs—it is expected to spread from low initial frequencies, species-specificity can be incorporated through regulatory elements, and a deleterious CRISPR can be removed through release of CRISPR-resistant alleles. The system is active
in a range of species and target sites are even easier to engineer than for TALENs, suggesting the system would be easily adapted to multiple vector species. Another advantage of the CRISPR system is that it can be used to target multiple sequences in a single experiment [57], increasing its potential efficacy and decreasing the rate at which target site mutagenesis could slow its spread. A major concern, however, is that the CRISPR system itself may be degraded. The CRISPR system is quite large, consisting of promoters, the Cas9 gene, guide RNAs and, depending on the strategy being implemented, multiple disease-refractory genes and associated regulatory elements. A system this size is prone to mutation and errors introduced during homing, including potential loss of function of disease-refractory genes. These considerations may lead to population suppression strategies being favored for CRISPR-based drive systems; however, this would place selection pressure on mutant CRISPR alleles having lost their function and so the evolutionary stability of the CRISPR system will need to be explored and optimized if it is to provide a cost-effective alternative to the relatively stable yet difficult-to-engineer X-shredding HEGs.

TOXIN—ANTIDOTE GENE DRIVE SYSTEMS

We now move on to gene drive systems that use combinations of toxins and antidotes, active at different life stages, to favor their own inheritance [58]. Gene drive systems of this type include Medea, engineered forms of underdominance such as UD^MEL, self-limiting systems such as killer-rescue, and other toxin—antidote possibilities such as Semele, Medusa, and inverse Medea.

Medea

The story of Medea has origins in both Greek mythology and beetle biology. In Greek mythology, Medea was the wife of the hero Jason, to whom she had two children. Her marriage to Jason was hard-earned, transpiring only after she enabled him to plough a field with fire-breathing oxen, among other achievements; but despite this, he left her when the king of Corinth offered him his daughter. As a form of revenge, Medea killed their two children. From a biological perspective, such infanticide would make Medea an unfit mother; but if the trait is genetic and children that inherit it also have the ability to defend themselves, then mathematical models show that it actually has a selective advantage and, if present at modest levels in a population, is expected to become present among all individuals within a matter of generations [59,60]. This is simply because children who are able to defend themselves against a murderous parent are more fit than those who cannot.

The Greek analogy sounds bizarre; but genes displaying these properties do actually exist in Nature and have been discovered and characterized in
various regions of the world [12,61,62]. The first such element to be identified was in the flour beetle *Tribolium castaneum* [12] and was given the name Medea after both the character from Greek mythology, and as an acronym for “maternal-effect dominant embryonic arrest.” By crossing individuals from geographically isolated locations, it was found that Medea-bearing males gave rise to both wild-type and Medea-bearing offspring; but that Medea-bearing females only gave rise to Medea-bearing offspring. It appeared that Medea-bearing mothers were selectively killing non-Medea-bearing offspring; or alternatively that they were trying to kill all offspring and the Medea-bearing offspring were able to defend themselves.

The genetic factors involved in this behavior remain obscure; but the dynamics suggest a model in which Medea consists of two tightly linked genes—a maternally expressed toxin gene, the product of which causes all eggs to become unviable and a zygotically expressed antidote gene, the product of which rescues Medea-bearing eggs from the effects of the toxin [12,63]. In *Tribolium*, Medea dynamics are attributed to an insertion of a composite Tc1 transposon inserted between two genes both having maternal and zygotic components [13]. Remarkably, this system was reverse-engineered using entirely synthetic components in laboratory populations of *D. melanogaster* and was shown to rapidly drive population replacement [14,64]. These synthetic elements were constructed using two unique, tightly linked components—a maternal toxin consisting of maternally deposited microRNA designed to target an essential embryonic gene; and a zygotic antidote consisting of a tightly linked, zygotically expressed, microRNA-resistant version of the embryonic essential gene. The combination of these components results in the death of wild-type offspring of Medea-bearing mothers, thus favoring the Medea allele in subsequent generations and mimicking the behavior of the natural element in *Tribolium* (Figure 9.2A).

**Current Status.** Medea was the first synthetic gene drive system to be developed, in this case in *D. melanogaster* [14]. Given that the synthetic Medea elements were constructed using rationally designed synthetic components and well-understood, conserved molecular and genetic mechanisms, it should be possible to engineer Medea elements in a range of other insects including mosquitoes. The Medea drive strategy is particularly well-suited to driving disease-refractory genes into mosquito populations, and hence the development of several efficient refractory genes for each disease of interest is encouraged.

**Design Criteria.** In many ways, Medea is the ideal system for replacement of wild mosquito populations with disease-refractory varieties. Solutions are available for all of the design criteria outlined earlier, and Medea has an advantage over homing-based strategies for population replacement since it is stably integrated into the host chromosome, thus not affected by the substantial risk of loss during homology-directed repair. If introduced at modest population frequencies, Medea can spread and rapidly
FIGURE 9.2 Dynamics of toxin—antidote-based gene drive systems. (A) Medea elements distort the offspring ratio in their favor through the action of a maternally expressed toxin (MT) and a zygotically expressed antidote (ZA). This results in the death of wild-type offspring of heterozygous mothers and enables the Medea element to spread into a population from very low initial frequencies. Dynamics here are shown for a Medea element with no fitness cost, released at 10% in the population. Transgenic frequency refers to any individual carrying at least one copy of the element. (B) UD_MEL (maternal-effect lethal underdominance) is a toxin—antidote-based underdominant system consisting of two constructs, each of which possesses a maternally expressed toxin (MT1 and MT2) whose activity is manifest during progeny embryogenesis and a zygotic antidote (ZA1 and ZA2) capable of neutralizing the maternal toxin expressed by the opposite construct. This results in heterozygous females being sterile if mated to wild-type individuals, thus leading to the characteristic bistable dynamics of underdominant systems. Dynamics here are shown for UD_MEL constructs at independently assorting loci having no fitness costs. If released at a population frequency of 20%, the system spreads to fixation in the population; but if released at 15%, the system is eliminated. (C) Semele elements distort the offspring ratio in their favor through the action of a semen-based toxin (SBT) and a female-specific antidote (FA). This results in unviable crosses between transgenic males and wild-type females and favors transgenic individuals provided the Semele element is present at population frequencies exceeding ~36% (above this frequency, the selective advantage of the antidote exceeds the selective disadvantage of the toxin). Dynamics here are shown for a Semele element with no fitness cost. If released at a population frequency of 40%, the element spreads to fixation in the population; but if released at 30%, the system is eliminated. (D) Medusa is a two-construct, sex chromosome-linked drive system capable of inducing confineable and reversible population suppression. The system consists of four components—a maternally expressed, X-linked toxin (MT1) causes suppression of the female population and selects for the transgene-bearing Y since only transgenic male offspring have the corresponding Y-linked zygotically expressed antidote (ZA1). A zygotically expressed, Y-linked toxin (ZT2) and a zygotically expressed, X-linked antidote (ZA2) then selects for the transgene-bearing X when the transgene-bearing Y is present, creating a balanced lethal system. When present above a certain threshold frequency, Medusa spreads while creating a strong male gender bias leading to population suppression. Dynamics here are shown for Medusa constructs having no fitness costs. For two consecutive male-only releases at a population frequency of 50%, the population becomes entirely male as the system spreads to fixation in the population; but for two consecutive male-only releases at a population frequency of 40%, the system is eliminated.
replace a population, even in the presence of modest fitness costs [60]; however, Medea is unlikely to spread following a small-scale accidental release because its driving ability is low at low population frequencies [18].

Tight linkage between the toxin, antidote, and refractory genes by placing the toxin and refractory genes within an intron of the antidote gene can improve system stability and reduce the rate of loss of the refractory gene through recombination. However, in the event that the Medea element or refractory gene become unlinked, mutated, or rendered ineffective through parasite evolution, second-generation Medea elements can be generated that utilize toxin—antidote combinations distinct from those of the first-generation elements [14], making it possible to carry out multiple cycles of population replacement. This strategy can also be used to remove refractory genes from populations in the event of adverse effects. As the functional components of Medea are developed in mosquito species, it will become more cost-efficient to develop these elements and to adapt them to multiple vector species.

**Toxin—Antidote-Based Underdominance**

Underdominant systems display the property that heterozygotes, or their progeny, have lower fitness than either homozygote [65]. In the simplest case of a single biallelic locus for which matings between opposite homozygotes are sterile, whichever allele is more frequent in the population will tend to spread to fixation. Underdominant systems therefore display features similar to that of a bistable switch at the population level—if the system is present above a critical threshold frequency, it will tend to spread to fixation, while if it is present below the threshold, it will tend to be eliminated in favor of the alternative allele or chromosome. A variety of toxin—antidote systems have been proposed to achieve these underdominant dynamics and the critical threshold frequency depends on the system and fitness cost.

A range of underdominant systems is available in Nature, including chromosomal alternations such as inversions, translocations, and compound chromosomes [3,4]. We will return to translocations in the Translocation section; but will concentrate here on novel forms of underdominance that are in principle straightforward to engineer using combinations of toxins and antidotes. Toxin—antidote approaches to underdominance were originally proposed by Davis et al., who suggested an elegant system having two transgenic constructs, each of which possesses a gene whose expression induces lethality and a gene that suppresses the expression or activity of the gene inducing lethality carried by the other construct [66]. The constructs can either be inserted at the same locus on a pair of homologous chromosomes or at different loci on nonhomologous chromosomes. These systems display underdominant properties because individuals carrying neither or both constructs are viable; but a proportion of their offspring—those carrying just one of the