12 Modelling Threshold-Dependent Gene Drives: a Case Study Using Engineered Underdominance

Matthew P. Edgington and Luke Alphey*
The Pirbright Institute, Woking, Surrey, UK

12.1 Introduction to Threshold-Dependent Gene Drives

Gene drives have been proposed as valuable tools in the fight against a range of globally important issues, including vectors of disease, invasive species and agricultural pests. These approaches are classified primarily based on their persistence and/or invasiveness. Here we consider persistent (i.e. self-sustaining) and low-invasiveness (i.e. threshold-dependent) approaches using engineered underdominance as a case study.

Gene drive is a phenomenon whereby a particular gene (or suite of genes) can bias inheritance in its (their) own favour, thus allowing it to increase in frequency over successive generations, even when deleterious to carrier individuals (Sinkins and Gould, 2006; Alphey, 2014, 2020; Champer *et al*., 2016; NASEM, 2016; Leftwich *et al*., 2018;) (see Raban and Akbari, Chapter 8; Champer, Chapter 9, this volume). This may occur by a range of natural or synthetic mechanisms, most of which act in one of two ways: (i) conversion of progeny individuals into other genotypes; or (ii) reducing the fitness (or killing) of progeny individuals of certain genotypes.

The precise configuration of gene drive components can lead to systems with a wide range of different behavioural characteristics, and these are often used to classify the various gene drive systems. Perhaps the most common classifications are based on their intended purpose (usually population modification or suppression), invasiveness (ability to spread into non-target populations) and persistence (whether they remain in the population or diminish over time).

Owing mainly to coverage in popular media outlets, the term gene drive is often associated with only widely known systems such as some CRISPR-based homing approaches. These are expected to have relatively straightforward behaviour in that they are highly invasive (spreading from extremely small releases and so likely also to spread to all populations linked by any degree of gene flow), highly persistent (at least in absence of resistance) and able to be used flexibly for either population suppression or modification (see Bottino-Rojas and James, Chapter 11, this volume). As discussed previously (James, 2005; NASEM, 2016; Harvey-Samuel *et al*., 2019; Long *et al*., 2020; Lanzaro *et al*., 2021), the first gene

*Corresponding author, email: luke.alphey@pirbright.ac.uk

 Transgenic Insects (eds M.Q. Benedict and M.J. Scott)
DOI: 10.1079/9781800621176.0012

drive trials are likely to be conducted in remote areas such as highly isolated islands to limit the probability of spill-over into non-target populations. It is debatable whether geographical containment of this type is adequate for non-localized (sometimes also referred to as 'global') gene drive approaches, as it would be difficult to guarantee perfect confinement within the trial area. Another potential drawback is that any gene drive releases would presumably need to gain regulatory approval from all affected countries, which for global systems could be argued to be all countries in which the target species (and any capable of forming fertile hybrids) are present (see Beech *et al.*, Chapter 25, this volume). For the first gene drive trial releases, such widespread regulatory approval required for a global system would seem challenging to obtain, at least for widely distributed species. Localized gene drives may need approval only in the target territory and may be particularly suitable where homogenous modification of every population of the species is not desired.

Though not as widely discussed in the media as non-localized drives, several designs for localized drives have been proposed and subject to considerable analysis. Here we focus on two-locus engineered underdominance (UD) – an example of threshold-dependent gene drive that should be persistent, reversible and spatially restricted. This would appear to answer much of the concern around non-localized gene drive approaches, since the system is unlikely to spread significantly beyond any initial trial site and can be reversed easily in the event of unintended consequences – features likely to prove desirable to regulatory bodies and other stakeholders in the context of initial proof-of-concept gene drive field trials.

In this chapter, we outline a range of modelling approaches that have been used to demonstrate the key characteristics of this approach, including threshold introduction frequencies, reversibility, spatial limitation and robustness to mutation/resistance. We then go on to discuss alternative configurations based on the use of sex-specific components and their effect on introduction thresholds. We conclude with a discussion on the cycle of information between mathematical models and experimental data along with a range of areas for future modelling that will be important in providing information on the anticipated effects of these systems when released into target populations.

12.2 Two-Locus Engineered Underdominance

Underdominance, also known as negative heterosis, is a natural phenomenon and the opposite of the better-known overdominance (positive heterosis or hybrid vigour). Thus, underdominance occurs where hybrids are of lower fitness than either of two different true-breeding parental strains; for practical purposes, one of these parental strains is wild-type, the other is the underdominance-based (UD) gene drive strain. In a single locus scenario, modelling of such selection against hybrids has been shown to allow for the eventual fixation of one allele, with the other being eliminated (Wilson and Turelli, 1986; Altrock *et al.*, 2010, 2011). More recently, transgenes displaying these properties have been developed and tested (Reeves *et al.*, 2014; Maselko *et al.*, 2020; Buchman *et al.*, 2021). The UD concept can also be expanded beyond a single locus. UD gene drives can potentially be developed using transgenic constructs containing toxin and antidote components. The particular configuration considered here is two-locus UD as originally proposed by Davis *et al.* (2001). This approach requires two distinct transgenic constructs to be inserted at independently segregating (unlinked) genomic loci, each of which comprises a lethal effector (toxin) and a suppressor (antidote) for the toxin of the other transgenic construct (Fig. 12.1). One (or optionally both) of these transgenic constructs will also contain a genetic cargo aimed at producing a desirable phenotype within the target population, for example a reduced ability to transmit a given pathogen (e.g. Franz *et al.*, 2006, 2014; Buchman *et al.*, 2019) (see Franz, Chapter 22, this volume). This combination of transgenic components gives an underdominance-like

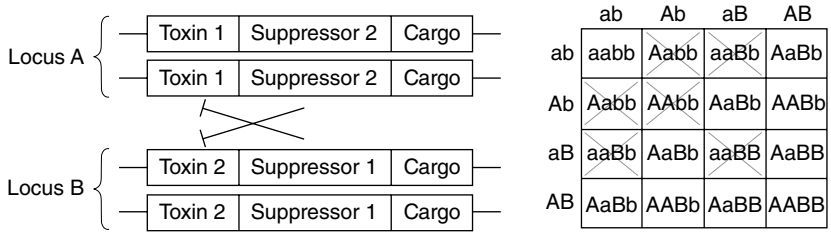

Fig. 12.1. A schematic diagram illustrating the workings of a two-locus engineered underdominance gene drive system. (Left) This gene drive design requires the introduction of two distinct transgenic constructs at independently segregating genomic loci. Each transgenic construct comprises a toxin, an element suppressing the toxin of the other transgenic construct and a desirable genetic cargo (which may be included within one or both constructs). **(Right)** A Punnett square demonstrating the creation of an underdominance-like effect. Here haplotypes outside the square represent gametes from a viable parent (one maternal and one paternal). The resulting offspring genotypes are listed inside the square and it is here that lethality manifests in individuals carrying just one of the transgenic constructs (grey crosses), creating a selective pressure for individuals to carry either both transgenic constructs or neither.

effect by creating a negative selection (via a lethal effect) on individuals carrying just one of the transgenic constructs (since they contain a toxin but not the antidote from the other transgenic construct). This results in a positive selection pressure for individuals to carry either both or neither of the transgenic constructs (Fig. 12.1). The precise strength of this selection pressure is dependent on several factors, including the degree of lethality conferred on affected genotypes and the fitness cost caused by the presence of the gene drive constructs, for example imperfect suppression of the lethals by the suppressors, or insertional effects of the transgene on nearby genes.

12.3 Mathematical Modelling Approaches

Building gene drives in the laboratory is an inherently time-consuming and expensive activity. Mathematical modelling, on the other hand, can be conducted relatively quickly and, by comparison, inexpensively and allows systematic exploration of parameter space far more readily than empirical methods. It is therefore extremely beneficial to model any proposed gene drive approach in advance of (or at least concurrently with) the laboratory development of transgenic components, and this has indeed been widespread practice. Models can be used to determine essential performance targets that must be met for engineered gene drives to achieve their intended function, particularly within laboratory-based experiments or field-based trial releases. The structure and complexity of such gene drive models can vary dramatically, with each having their own respective benefits and limitations. It is here that experienced mathematical modellers are key in determining the most appropriate model to use in any given scenario while understanding and being able to communicate to non-modellers how a given model structure may influence modelling outcomes. It seems intuitive that mathematical modelling has the potential to save vast amounts of experimental time, effort and money where gene drive designs and engineered components are not fit for purpose. Perhaps less obvious is the requirement for models, and model-based conclusions, to be used (or at least scrutinized) by experienced modellers who understand the impact of model assumptions, complexity and limitations.

A variety of mathematical modelling approaches have been used to provide insight into the predicted performance of UD gene drives; however, much of this work has focused on deterministic population genetics

models. These can provide insight into the basic function and utility of UD systems. Such models typically consider an idealized scenario consisting of an infinitely large population (avoiding stochastic effects and the need for integer numbers of individuals) that is isolated (closed from any migration) and panmictic (randomly mating). For simplicity, it is also commonly assumed that females mate only once and produce a 1:1 (female to male) sex ratio in their offspring. Attention is typically restricted to the case whereby no resistance mechanisms can emerge and where each component of the introduced transgenes is immutable and perfectly linked (i.e., toxin, antidote and cargo genes are unable to separate from one another). Finally, it is commonly assumed that generations of offspring in modelled populations are synchronous (i.e., non-overlapping), which may not always be realistic but can apply to laboratory caged populations or wild populations that are synchronized by climatic factors (e.g., wet and dry seasons). This allows for the use of simple recurrence relations (i.e., difference equations) to model the population genetics resulting from the release of such a gene drive. This typical set of simplifying assumptions is also adopted in the example below.

Much of the modelling of UD gene drives uses genotype-based population genetics models. In the case of a two-locus approach such as UD, this results in a total of nine possible genotypes – homozygous, heterozygous or wild-type for the transgene at each of two loci – and therefore a set of nine difference equations. However, since UD is based entirely on Mendelian inheritance and lethality to certain offspring genotypes, the offspring in each generation are directly related to the proportions of each allele present in the parental generation rather than the precise parental genotypes. This allows the consideration of a simpler model that requires the tracking of only the four haplotype frequencies (ab, Ab, aB and AB, where a/b represent wild-type alleles and A/B their transgenic counterparts). This results in a set of four difference equations that may be solved recursively in a manner similar to that originally presented by Davis *et al*. (2001). Magori and Gould (2006) considered a similar model structure but additionally allowed for multiple insertions of each transgene, though this is not included here. This model is of the form:

$$ab_{t+1} = \frac{\left[\left(ab_t^{\,2}\right) + \left(\frac{1}{2}\varepsilon^2 ab_t AB_t\right) + \left(\frac{1}{2}\varepsilon^2 aB_t Ab_t\right)\right]}{\bar{\Omega}} = \frac{f_1}{\bar{\Omega}}, \quad \text{(Equation 1)}$$

$$aB_{t+1} = \frac{\left[\left(\frac{1}{2}\varepsilon^2 ab_t AB_t\right) + \left(\frac{1}{2}\varepsilon^2 aB_t Ab_t\right) + \left(\varepsilon^3 aB_t AB_t\right)\right]}{\bar{\Omega}} = \frac{f_2}{\bar{\Omega}}, \quad \text{(Equation 2)}$$

$$Ab_{t+1} = \frac{\left[\left(\frac{1}{2}\varepsilon^2 ab_t AB_t\right) + \left(\frac{1}{2}\varepsilon^2 aB_t Ab_t\right) + \left(\varepsilon^3 Ab_t AB_t\right)\right]}{\bar{\Omega}} = \frac{f_3}{\bar{\Omega}}, \quad \text{(Equation 3)}$$

$$AB_{t+1} = \frac{\left[\left(\frac{1}{2}\varepsilon^2 ab_t AB_t\right) + \left(\frac{1}{2}\varepsilon^2 aB_t Ab_t\right) + \left(\varepsilon^3 aB_t AB_t\right) + \left(\varepsilon^3 Ab_t AB_t\right) + \left(\varepsilon^4 AB_t^{\,2}\right)\right]}{\bar{\Omega}} = \frac{f_4}{\bar{\Omega}}, \quad \text{(Equation 4)}$$

where

$$\bar{\Omega} = f_1 + f_2 + f_3 + f_4, \quad (5)$$

is the sum of numerators in (Equation 1)–(Equation 4), ε denotes the fitness (relative to wild-type) of an individual carrying a transgenic construct and t denotes the generation from which the next allele frequency is computed. For simplicity, here we assume that each transgenic construct (A and B) confers the same degree of fitness cost on carrier individuals and that these are applied multiplicatively where individuals carry more than one transgenic construct (up to a maximum of four, where the relative fitness would be given by ε^4). Note that the parameter ε can take any value in the range from zero (completely non-viable) to one (equally as fit as wild-type) and that the consideration of multiplicative relative fitness ensures that the overall value for any genotype also remains in the range from zero to one. Like much of the modelling literature, here we assume that toxins are fully penetrant (i.e., no viable offspring result) and similar for antidotes (i.e., a single antidote copy provides full rescue).

The model presented here provides one of the simplest possible models of UD gene drive and is useful for determining various base-level characteristics of the system. As with all models, this is based on a range of simplifying assumptions (described above), each of which is likely to have its own implications. There is a wide range of other modelling approaches that can be (and have been) used to capture the anticipated effects of relaxing one or more of these model assumptions. Several of these are briefly discussed in the following sections, focusing primarily on results rather than extensive modelling detail. Modelling of UD systems has spanned a range of model structures, including difference equations, ordinary differential equations (ODEs), delay differential equations (DDEs), partial differential equations (PDEs) and stochastic models, each of which provides insights into different aspects of anticipated UD behaviour.

12.4 Introduction Thresholds

Underdominance acting at a single locus has been shown to produce bistable dynamics: either homozygotic state can be stable, depending on the initial frequencies and relative fitness of each type, as shown for underdominant alleles (Wilson and Turelli, 1986; Altrock *et al*., 2010, 2011) and chromosome translocations (Curtis, 1968). UD gene drives seek to capture a similar effect synthetically, via the introduction of toxin and antidote elements. While the threshold dependence of UD has been demonstrated using numerical simulation under a range of release sizes, to our knowledge a full equilibrium analysis has yet to be conducted. This can be achieved either analytically or computationally and here we focus on the latter, using the numerical continuation software package XPPAUT (Ermentrout, 2002), producing results shown in Fig. 12.2.

These results show two distinct regimes of behaviour separated by a particular relative fitness parameter $\varepsilon^* \approx 0.725$ (as these are applied multiplicatively, this gives UD double homozygotes a relative fitness of just $\sim$0.276). In the region $\varepsilon < \varepsilon^*$ there are two possible equilibrium states, with either the wild-type (stable) or gene drive (unstable) alleles at fixation. This can be interpreted as a scenario in which the gene drive is unable to establish itself, no matter how many gene drive-carrying individuals are introduced. The unstable equilibrium whereby gene drive alleles are at fixation is not biologically feasible, since it would imply there were no wild-type individuals present at the time of release – rendering the release of a gene drive unnecessary. The more interesting region ($\varepsilon > \varepsilon^*$) displays four possible equilibrium states. The first is the unstable (and not biologically feasible) equilibrium state with gene drive alleles at fixation. The three remaining equilibria together constitute a bistable system (i.e. two stable equilibria separated by an unstable equilibrium). Focusing on the gene drive allele (Fig. 12.2c) below the unstable equilibrium, the elimination of the gene drive is the only stable equilibrium – representing negative selection against the gene drive when introduced at a sub-threshold frequency. Above the unstable equilibrium, the gene drive moves towards a stable equilibrium with high gene drive frequency – representing positive

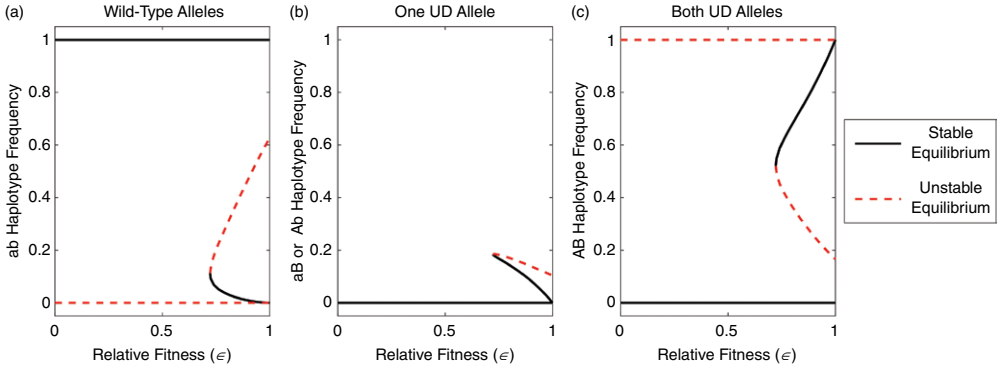

Fig. 12.2. A bifurcation diagram showing the possible equilibria of a two-locus engineered underdominance gene drive and their associated stability properties. Here stable equilibria are shown by solid black lines while unstable equilibria are shown by red dashed lines, with panel (a) showing ab (i.e. fully wild-type), (b) Ab or aB (i.e. a single UD allele) and (c) AB (i.e. both UD alleles) haplotype frequencies. These diagrams show the bistable nature of this gene drive approach. When introduced above a certain threshold, the gene drive system increases in frequency, towards the stable equilibrium with non-zero AB haplotype frequency. When introduced below this threshold, the system decreases in frequency, heading towards the stable equilibrium with a zero AB haplotype frequency. These diagrams also demonstrate the existence of a maximum tolerable fitness cost for two-locus engineered underdominance gene drive systems of around 28% per construct. Bifurcation analysis was conducted using XPPAUT continuation software (Ermentrout, 2002) and results were plotted using MATLAB (R2020b, The MathWorks Inc., Natick, Massachusetts).

selection when introducing the gene drive above the threshold frequency. Theoretically it is possible that the system would attain the unstable equilibrium state and remain there; however practically this is exceedingly unlikely, due to the many and varied stochastic effects present in the real world.

A feature evident in the results of Fig. 12.2 is that the equilibrium state for a UD system does not necessarily comprise only gene drive homozygotes. Where there are no fitness costs associated with the gene drive, the system can reach fixation (i.e. 100% gene drive homozygotes). However, where fitness costs are non-zero, wild-type alleles are expected to be present at a frequency that increases with the fitness costs of the system (Fig. 12.2b).

While useful in displaying the bistable (i.e. threshold-dependent) nature of a UD gene drive, Fig. 12.2 is not necessarily of direct use when planning a gene drive release. This is due to the combination of AB and Ab/aB haplotypes present at the unstable equilibrium (i.e. the introduction threshold). In practice it would be more convenient to know a single gene drive frequency above which the system must be introduced for it to increase in frequency within the target population. Fortunately, this can be calculated by summing to obtain the overall gene drive allele frequency for each point on the unstable equilibrium line. This results in a single threshold gene drive allele frequency (as shown in Fig. 12.3) that must be exceeded through any combination of heterozygote or homozygote individuals. Note that these results align with the pattern observed in several previous studies (e.g. Magori and Gould, 2006; Edgington and Alphey, 2017, 2018; Dhole *et al.*, 2018, 2020; Leftwich *et al.*, 2018), though precise thresholds may differ due to assumptions on the application of fitness costs to individuals carrying multiple transgenic constructs and the choice of presentation method.

12.5 Relaxing Model Assumptions

As discussed above, gene drive models are based on a range of simplifying assumptions,

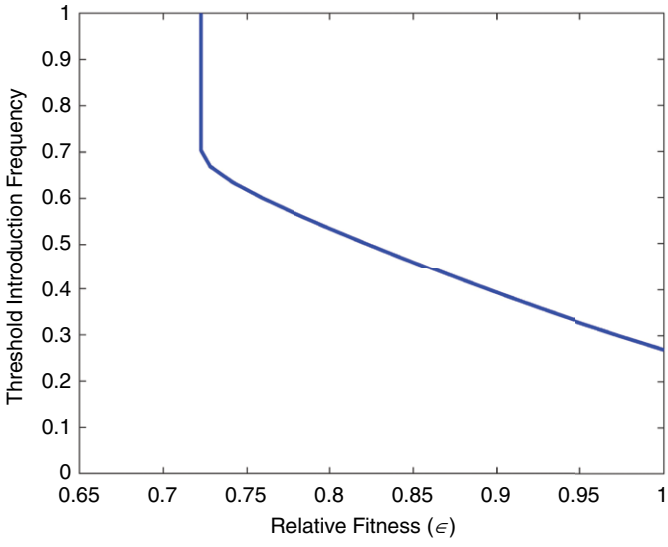

Fig. 12.3. Threshold introduction frequencies required for an engineered underdominance system to spread in a target population over a range of relative fitness parameters. Note that this figure was generated by summing gene drive allele frequencies from results in Fig. 12.2, but mirror those previously shown in Edgington and Alphey (2017, 2018), using alternative mathematical models – although some slight differences are observed due to assumptions on how fitness costs should be applied to individuals carrying multiple transgenic constructs.

each of which has its own implications for the outcomes and applicability of models to different scenarios. The model outlined in section 12.3 represents one of the simplest useful representations of a UD gene drive system and, as shown in section 12.4, allows key characteristics of this gene drive design to be elucidated. Of course, this model structure can be altered to allow for the relaxation of any of the model assumptions outlined above, thereby allowing their implications to be explored. In the following sections we discuss studies exploring the relaxation of three such model assumptions: (1) the presence of resistance formation and mutation of transgenic constructs; (2) the reversal of UD gene drives; and (3) the incorporation of spatial effects.

12.5.1 Resistance formation and mutation

A common set of simplifying assumptions for gene drive modelling studies is that elements within a single transgenic construct are perfectly linked (i.e., unable to separate), do not undergo mutation (i.e., lose function of transgenic components) and that no other resistance mechanisms emerge. Edgington and Alphey (2019) relaxed this assumption by modelling a scenario whereby transgenic constructs accumulate loss-of-function mutations at a constant rate. To our knowledge, rates of mutation in insects likely to be targeted by gene drives are not well studied and could vary considerably, depending on the molecular biology of the gene drive system. Thus, mutation rates (per gene) are assumed to fall within the range 10^{-4}–10^{-8} that should span rates relevant to a range of target insect species. This parameter range is based on measured mutation rates in *Drosophila melanogaster* (estimated as being of the order 10^{-9} per nucleotide per generation) (Haag-Liautard *et al.*, 2007; Keightley *et al.*, 2014); the size of gene drive constructs in previous studies (~1–10 kb) (Windbichler *et al.*, 2011; Reeves *et al.*, 2014; Champer *et al.*, 2017) and an

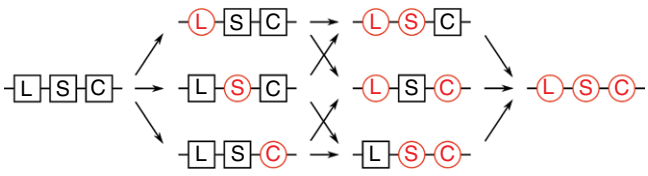

Fig. 12.4. Transgenic constructs are assumed to mutate at a rate of *m* per gene. This is assumed to be low enough that multiple mutations per generation may be neglected. For example, the initial transgenic construct (say, A) mutates at a rate of 3*m*, producing mutations in the lethal (giving A_L), suppressor (A_S) and the cargo (A_C) gene each at a rate of *m*. Then, transgenic constructs possessing one mutated gene (e.g. A_L) mutate at a rate of 2*m* (giving A_{LS} and A_{LC} each at rate *m*). Transgenic constructs with two mutated genes (e.g. A_{LS}) then mutate at rate *m*, producing constructs with all three genes mutated (i.e. A_{LSC}). Here, non-mutated genes are represented by black squares whereas genes with loss-of-function mutations are shown in red circles. (Figure originally published in Edgington and Alphey, 2019.)

assumption that ~1–10% of nucleotides in transgenic constructs are essential for gene function. Note that these mutation rates are assumed to be low enough that multiple mutations (i.e., in more than one gene) within a single generation may be neglected, leading to the pattern of mutation shown in Fig. 12.4.

The original study describing this scheme of mutation (Edgington and Alphey, 2019) considered a genotype-based formulation, resulting in a set of 2025 genotypes, of which 819 were non-viable. This could be reduced to a haplotype-based formulation with 81 haplotypes (81 difference equations), making the model simpler and faster to formulate, code and simulate.

This study found that UD displays an increase in frequency that is almost completely unaffected by such mutation where mutated transgenic constructs conferred a greater fitness cost than their non-mutated counterparts – such cases will not be discussed further. Loss-of-function mutations are therefore only of concern where mutated transgenic constructs are of higher fitness than non-mutated versions. Here, the introduced UD system would initially increase in frequency if introduced above the required threshold. Over time each type of mutated transgenic construct will begin to accumulate, with the rate of accumulation and maximum frequencies varying depending on which loss-of-function mutations are present. Results in Edgington and Alphey (2019) show that, for a range of mutation rates and fitness costs, it is transgenic constructs with a single loss-of-function mutation in either the lethal or cargo gene that achieve the greatest maximal frequencies (Fig. 12.5). Constructs with loss-of-function in two genes achieve lower frequencies and are dominated by those where the antidote (suppressor) gene is unaffected. Combined, the mutated transgenic constructs reach high overall frequencies, with a concurrent decrease in the frequency of non-mutated constructs. Since UD approaches typically achieve an equilibrium in which wild-type alleles remain present (see Fig. 12.2), these begin to replace the mutated transgenes due to their relative fitness advantage, eventually returning the population to a fully wild-type state (Fig. 12.5). It is yet to be studied in depth whether the stable co-existence of mutated and non-mutated transgenic constructs is possible, but it was not observed under any parameter set or initial condition considered in Edgington and Alphey (2019).

For different fitness costs and mutation rates, the observed dynamics remained broadly similar to those in Fig. 12.5, though the timescales and maximal frequencies of each mutation vary. Higher mutation rates reduce the period over which the UD system persists at high frequency – essentially the period in which the gene drive would maintain efficacy. Even though UD systems can be eliminated by mutations, they are predicted to remain at high frequency for hundreds of generations – likely long enough for the system to have produced its desired effect. As an example, *Aedes aegypti* mosquitoes (vectors of dengue, Zika, yellow fever and chikungunya viruses) undergo approximately

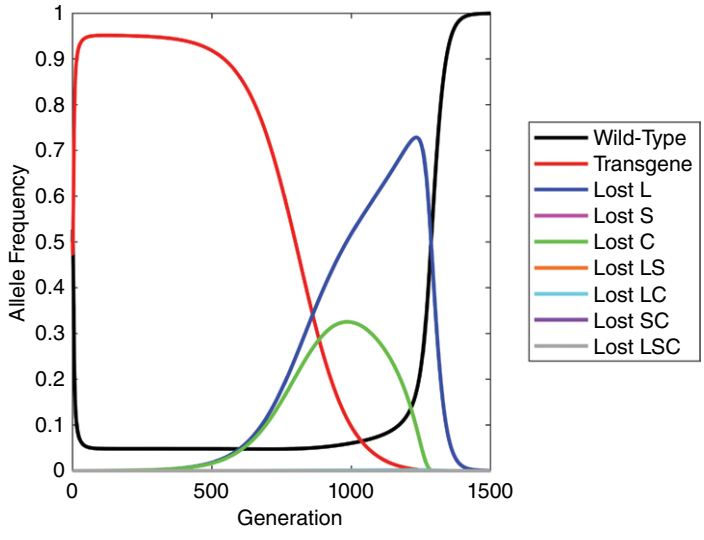

Fig. 12.5. An example numerical simulation showing the effects of mutation within transgenic constructs of an engineered underdominance gene drive system. Results here are presented for a 1:1 (introduced to wild) introduction of non-mutated double homozygote (AABB) individuals into a wild-type (aabb) population. Note that several alleles only reach very low maximum frequencies and thus they appear to overlie one another along the horizontal axis. Results are shown for an engineered underdominance system with 5% fitness cost per non-mutated construct, a 4% fitness cost per mutated construct and a mutation rate of $m = 10^{-6}$.

ten generations per year, meaning that the introduced UD system should persist at high frequency for at least ten years, even in the face of accumulating loss-of-function transgene mutations.

12.5.2 UD reversal

Ideally, a gene drive, or any other intervention, should be simple to reverse in the event of any unintended and undesirable consequences. For threshold-dependent drives such as underdominance (UD), the simplest mechanism for this is the release of wild-type individuals. Releasing a sufficiently large number of wild-type individuals can push the UD system below the threshold frequency, thus it undergoes negative selection and is driven out of the population (consider, for example, Fig. 12.2). To our knowledge, only one alternative reversal mechanism has been proposed for reversing UD, namely releasing individuals carrying free suppressors (i.e., individuals carrying just the antidote from one or both of the original transgenic constructs) (Edgington and Alphey, 2019). This was proposed since modelling of mutation in UD systems showed significantly greater accumulation of mutated constructs that retained function of the antidote gene, which undergo positive selection where a UD system is at high frequency. The same study also showed that if free-suppressor elements conferred a non-zero fitness cost, then the positive selection would be lost as the UD system became rare in the population, thus allowing wild-type alleles to outcompete the free suppressor element, returning to a fully wild-type state.

With there being two viable mechanisms for UD reversal, it is instructive to compare the two. Each has benefits, but which would be more useful in real-world scenarios? If a released UD system were to require reversal, then it would be reasonable to assume that introducing large numbers of wild-type individuals would be undesirable, as it could potentially increase the population

above pre-control levels, albeit transiently. Intensive suppression might reduce numbers in the entire population such that the necessary wild-type releases remain below pre-control levels, but that would have its own costs and issues. In some cases, issues around release of wild-type individuals could be alleviated somewhat by releasing individuals of a single sex (as discussed in Leftwich *et al*., 2018). For example, in many insect species (and certainly many of those likely to be gene drive targets) females provide both the reproductive and epidemiological potential of the population; thus, releasing wild-type males should generally be quite benign.

The release of free-suppressor individuals provides an alternative reversal mechanism and addresses some of the issues associated with wild-type release(s) but also has some drawbacks of its own. For example, free-suppressor individuals are theoretically able to function from an extremely small release. In practice, it would be desirable to perform release(s) large enough to avoid stochastic loss at low frequencies. Despite this, it should still be feasible to release far fewer individuals than required for wild-type reversal and these may be of a single sex. However, there are also some potential drawbacks that need to be weighed against these benefits. Firstly, free-suppressor releases have been shown to function much more slowly than wild-type reversal. For instance, Edgington and Alphey (2019) show that a 2:1 (reversal to wild) release of wild-type individuals can (approximately) eliminate the UD gene drive in about 20 generations whereas an equal release of free suppressor individuals took about 130 generations to reach the same point, with about a further 150 generations required for the free-suppressor individuals to be (approximately) eliminated. Another potential issue is whether the appropriate regulatory body would approve the release of further transgenic insects if the original system required reversal due to unintended negative effects.

12.5.3 Spatial effects

A key feature of UD gene drives is their threshold-dependent nature, since this is often stated to be capable of preventing the system from establishing in non-target neighbouring populations. It may even prevent the system reaching an appreciable frequency, due to negative selection when present at sub-threshold frequencies. The modelling of such spatial factors is therefore important in assessing how well confined UD systems will remain and under what conditions this confinement could potentially fail. These questions are expected to prove important when seeking to attain regulatory approval for UD releases into the field; with highly robust confinement, regulatory approval far beyond the release site(s), for example regional or multi-national approval, may not be required, in contrast to current thought regarding more invasive approaches.

Spatial effects can conceivably be studied in a variety of ways, including *n*-deme population genetics (difference equations), *n*-deme population dynamics (ordinary or delay differential equations), reaction–diffusion (partial differential equations) or individual-based models. Each of these has been used in the study of gene drive approaches, although not all in the context of UD, and possesses its own positive and negative features. Here we discuss a range of these approaches in the context of UD gene drives, focusing primarily on findings rather than technical details. While we cover a broad range of studies here, this is by no means intended as a comprehensive review.

Perhaps the most commonly used technique for assessing spatial properties of gene drives are *n*-deme population genetics models. Briefly, these consider two or more demes (semi-isolated (sub)populations), each of which has its own set of difference equations of the form outlined in section 12.4. For simplicity, the literature primarily considers a scenario with just two demes. These models capture the migration of individuals between demes via a simple exchange of a proportion of individuals in each generation. It is commonly assumed that the two populations are of equal size, such that the number of each migrant type is simply based on the haplotype frequencies in each population – a reasonable assumption where large (modelled as infinite) populations

are considered. This approach has been used to consider spatial aspects for a wide range of gene drive classes, including UD (Marshall and Hay, 2012; Harvey-Samuel *et al*., 2019; Edgington *et al*., 2020b). One such model estimated that a UD system with a homozygote fitness cost of 5% (applied additively) and a bidirectional migration rate of 1% per generation would reach near-fixation in the target population (e.g. results in Fig. 12.2) but reaches a frequency of just 0.032 in a non-target neighbouring population (Marshall and Hay, 2012). The same work then went on to estimate that the same UD system would require a bidirectional migration rate of 4.3% per generation to become established in both populations, thus supporting the notion that UD is robustly confineable.

The approximation of equal population sizes has also been relaxed in a number of studies focusing both on UD (Dhole *et al*., 2018) and on other gene drive classes (Dhole *et al*., 2019, 2020). These studies include an approximate scaling of migration rates to account for the differences in respective population sizes that result from gene drive fitness cost and lethal effects. While this captures an additional level of realism absent in the non-scaled migration models, the results obtained do not suggest that this will have a large impact on the ability of UD systems to remain confined. In fact, we would suggest that this is likely to improve the confinement, since transgene fitness costs and lethal effects will reduce the target population size, meaning fewer migrants into the non-target population. However, this will likely increase the influence of wild-type migrants from the non-target population, potentially creating a necessity for slightly larger UD release(s) to ensure the system remains above the introduction threshold.

Several studies have also considered extensions to these population genetics models by taking account of various ecological factors affecting the life cycle and size of the target and non-target insect populations when subject to the release of a UD gene drive (e.g., Edgington and Alphey, 2018; Khamis *et al*., 2018, 2020). These each consider their own model structures to capture density-dependent effects during the immature stages of the insect life cycle, each with its respective advantages and disadvantages. One density-dependence function used in such models is that of Maynard Smith and Slatkin (1973) and is of the form:

$$f(N) = \left(1 + (aN)^b\right)^{-1},$$

where N denotes the size of the insect population, a is a density parameter ($1/a$ relates to the number of breeding sites) and b defines the strength of density-dependent competition. This function is known to be flexible in that it can capture a range of density-dependence scenarios (Bellows, 1981) and has been used in the study of UD (Edgington and Alphey, 2018) and other gene drive classes (Alphey and Bonsall, 2014). This work outlines a variety of possibilities not extensively discussed in the results of population genetics models. In particular it identifies three possible outcomes of a UD release: (1) no introgression in either population; (2) establishment in both populations; and (3) introgression into the target population with extremely limited spread into the non-target population, with the latter usually considered the most desirable outcome for UD (Edgington and Alphey, 2018). Sánchez *et al*. (2020a) considered similar effects for UD^{MEL} and reciprocal chromosome translocations using a computational framework called MGDriveE (Sánchez *et al*., 2020b; Wu *et al*., 2021).

The above approaches consider spatial effects via an exchange of individuals between two (sub)populations, which are assumed to be well mixed. A possible extension to this work is to consider spatial effects explicitly, using a model defined over a continuous spatial domain (e.g., Champer *et al*., 2020d). This work takes a fully computational approach, using an individual-based model implemented in the open-source software package SLiM (Haller and Messer, 2016, 2019). Here, two circular regions (subpopulations) are linked by a narrow migration corridor, with movement assumed to result from the birth of new offspring (Champer *et al*., 2020d). This showed that UD is robust against re-invasion by wild-type but may

display a greater degree of invasiveness into neighbouring populations than predicted with the spatially implicit model structures discussed above. However, the narrow migration corridor essentially forces migrating individuals to encounter those moving in the opposite direction, creating an approximately linear boundary between the two populations – a scenario shown to facilitate easier gene drive invasion (Champer *et al*., 2020d).

An alternative scenario would allow migration to occur over a wider space, meaning migrants encounter those moving in the opposite direction far less frequently, thus eliminating the linear boundary within the migration corridor. Here migrants would first encounter individuals when arriving at the boundary of the opposite population, meaning they would encounter either a very high or very low local gene drive frequency, likely producing results closer to those from spatially implicit model structures. Such variation highlights the importance of understanding a wide range of species, location and ecological traits when predicting the outcome of a real-world UD release.

While the above approaches assess the likelihood of UD invading non-target populations, spatial effects are also important in determining the ability of UD to spread in a given target population. This has also been addressed using a variety of different modelling approaches.

One possibility is to consider a lattice-based model in which the target region is discretized into a collection of cells, each containing a well-mixed pool of individuals (Huang *et al*., 2011). Individuals move between cells according to a dispersal kernel, defining the probability of an individual moving between any two cells in the lattice on any given day. This model structure was used to compare the relative efficacy of two UD release methods: (i) release into one large area; and (ii) release into many smaller, equally distributed areas. Interestingly, this work showed that either release method could be more effective, depending on the degree of mobility exhibited by individuals and the fitness costs associated with the UD system.

The individual-based model of Champer *et al*. (2020d) has also been used to explore the ability of UD to spread within a single population. This considered two spatial scenarios based on the shape of the UD release area, namely a scenario with either a straight-line (linear) scenario or a circle dividing regions of high and low/zero gene drive frequency. Interestingly, the UD system was able to spread or persist more readily in the linear scenario than the circular one. This was proposed to be a result of the local gene drive frequency being lower for the circular case, since the wild-type partially wraps around the high UD region. However, we would expect this effect to diminish rapidly as the circular region increases in size (reducing the curvature of the boundary).

Finally, another potential approach for considering spatial effects in either single or linked populations is the use of reaction–diffusion equations (i.e., partial differential equation models). These have been explored in the context of highly invasive CRISPR-based gene drives (Beaghton *et al*., 2016; Tanaka *et al*., 2017), but to our knowledge have yet to be widely applied to threshold-dependent systems. Such models should enable a wide range of spatial scenarios to be considered, while allowing various sources of heterogeneity to be considered across the spatial domain. These models can also allow an explicit representation of insect migration to be incorporated into population dynamics model structures, such as those in Edgington and Alphey (2018) and Khamis *et al*. (2018, 2020).

12.6 Linking Theory and Experimentation

The development of gene drive technologies, including UD, has generally followed a design–build–test cycle (Fig. 12.6). At each stage, modelling can play an important role in designing, understanding and analysing experimental work. Thus, in addition to providing insights, modelling can save a significant amount of research time, effort and

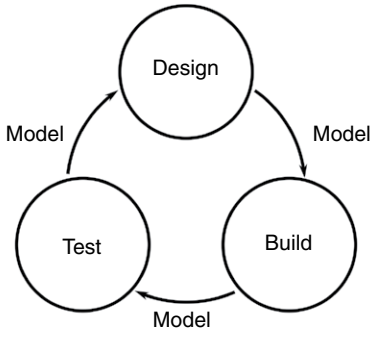

Fig. 12.6. A cartoon showing the design–build–test cycle commonly followed in the development of gene drive technologies and that mathematical modelling is an important tool within each phase.

money. To date there has been relatively little published experimental work on UD and so this section focuses on how one could link theory and experiment when experimental data becomes available. This will partially be informed by studies on alternative gene drive classes and will describe each phase of the design–build–test cycle; however, since the 'design–build' aspect of this cycle has largely been covered in previous sections, it will not be discussed further here.

In the 'build–test' phase, experimentation commonly focuses on discrete generation laboratory cage-based experiments. Models of the form shown in section 12.4 assume discrete generations and so they are ideal for predicting and analysing experiments of that form. In the first instance, specific test crosses between transgenic insect strains can be performed and the number of offspring of each resulting genotype counted/screened (as in Hammond *et al.,* 2016, 2017, 2020; Kyrou *et al*., 2018; Adolfi *et al*., 2020; Champer *et al.,* 2020b,c; Simoni *et al*., 2020), with fluorescent markers commonly used to distinguish between types. This allows a first approximation of various system parameters (for example, relative fitness, toxin penetrance and strength of the antidote effect (Buchman *et al*., 2018b; Webster *et al*., 2020)) to be generated by calculating ratios between the mean number of each genotype produced. These can then be used to parameterize models and predict the expected outcomes of gene drive cage trials. Note that other gene drive classes may approximate fitness costs by fitting models to data from cage trials with only a single transgenic construct present (e.g. Webster *et al*., 2020), but this is not feasible for UD where a single transgenic construct produces a lethal phenotype.

Following the 'test' phase, i.e., after laboratory cage trial experiments, prior predictions of gene drive behaviour can be compared with the actual cage trial data. Using the same mathematical models (and potentially stochastic versions also, especially given the relatively small numbers of individual mosquitoes typical of a laboratory cage experiment), one can then assess whether model predictions are consistent with the observed outcomes (as in Hammond *et al*., 2016, 2020; Buchman *et al*., 2018b; Kyrou *et al*., 2018; Pham *et al*., 2019; Adolfi *et al*., 2020; Champer *et al*., 2020c; Simoni *et al*., 2020; Webster *et al*., 2020). If model and experimental results are consistent, then parameters can be varied to identify potential areas for improvement in the gene drive design. Conversely, if model and experimentation are not consistent, then further modelling may be required to identify sources of this mismatch, potentially informing future models and experimental designs (as in Hammond *et al*., 2017, 2020, for CRISPR-based gene drives).

12.7 Alternative Configurations of UD

Previous sections focused on a specific configuration of UD, namely that based on two mutually repressing bi-sex toxin genes, inserted into two unlinked and independently segregating genomic loci. Similarly, we have primarily focused on the release of both transgenic males and females. However, there is a wide range of alternative methods for engineering/releasing UD systems, a variety of which have been studied previously.

For the UD design considered thus far, a potential variation is in the time at which the toxin takes effect. This has been studied for a UD system with toxins and transgene fitness costs acting either in early (before

density-dependent competition, for example eggs or early instar larvae for mosquitoes) or late (following density-dependent competition, for example pupae or pharate adults in mosquitoes) developmental stages (Edgington and Alphey, 2018; Khamis *et al*., 2018). While early- or late-acting lethal/fitness effects do not have any impact on threshold introduction frequencies, they can have more impact on other traits. For instance, Khamis *et al*. (2018) found that, for a system spreading a cargo gene conferring refractoriness to a pathogen, late-acting lethality produced a slightly larger reduction in disease burden. This is likely due to the greater reduction in both equilibrium and (transiently attained) minimum population sizes observed with late-acting lethality (as seen in Edgington and Alphey, 2018). Despite this potential epidemiological benefit, a greater reduction in population size may not always be good news. For example, *Ae. aegypti* mosquitoes are known to compete with *Aedes albopictus* (Edgerly *et al*., 1993; Juliano *et al*., 2002; Armistead *et al*., 2008). Therefore, an *Ae. aegypti* population may be displaced during the period in which the population is reduced by a late-acting UD – potentially reducing the epidemiological benefit as *Ae. albopictus* are also competent vectors of a similar set of pathogens, including dengue viruses (WHO, 2011). This necessitates some knowledge of ecological factors in the vicinity of gene drive target areas, and could be addressed by modelling approaches similar to those used for sterile insect technique (SIT) and release of insects carrying a dominant lethal (RIDL)-based control (Bonsall *et al*., 2010).

Other possible UD configurations revolve around the use of sex-specific toxins or insect releases, rather than the bi-sex versions considered above. Such considerations have been studied in terms of their effect on release thresholds and degrees of tolerable transgene fitness costs (Edgington and Alphey, 2017). These results showed that considering either male-only release(s) of gene drive-carrying individuals or female-specific toxins results in a lesser ability to tolerate fitness costs and higher introduction thresholds.

In a two-locus UD configuration, it is possible that the suppressor element from one transgenic construct is not sufficient to inactivate two copies of the toxin gene from the other transgenic construct. In the context of UD, this has been referred to as 'weak suppression' (Edgington and Alphey, 2017, 2018); however, the mathematical models and predicted dynamics are equally applicable to systems based on reciprocal chromosome translocations (Buchman *et al*., 2018b). These studies showed that reciprocal chromosome translocations (or weakly suppressed UD) generally have a higher introduction threshold than the UD systems discussed here. As discussed previously, this represents a trade-off between the increased cost/difficulty of gene drive introgression and the increased reliability of gene drive confinement to the target population.

Several gene drive concepts are based on toxin–antidote systems. The UD system considered thus far assumes two mutually suppressing lethals, each of which comprise a 'toxin' gene and an antidote that suppresses its effect, perhaps RNAi targeting the toxin gene. Other toxin–antidote concepts can also provide threshold-dependent gene drives. For example, a synthetic *Medea* drive was constructed in *Drosophila* using a maternally contributed (RNAi) toxin with zygotic expression of the antidote only in those offspring inheriting the *Medea* element (Buchman *et al*., 2018a). *Medea* is a low-threshold drive, zero-threshold in the absence of fitness costs, but a mutually repressing pair of such elements can provide a threshold-dependent drive known as UD^{MEL} (Akbari *et al*., 2013) or double *Medea* (Wimmer, 2013). For a rather different molecular basis, a CRISPR/Cas9 system (toxin) can be used to disrupt an essential endogenous gene, which can then be rescued with a recoded – and therefore toxin-resistant – antidote. A range of threshold-dependent gene drives using this technology have previously been modelled (Champer *et al*., 2021) and are based on the use of cleave and rescue (Oberhofer *et al*., 2019) or CRISPR toxin–antidote (Champer *et al*., 2021; Champer *et al*., 2020a; Champer *et al*., 2020b) elements. These have been discussed extensively

in the original sources and produce broadly similar behaviour to the approach(es) discussed here. Additionally, the mathematical modelling frameworks considered in the studies listed above are similar to those explored throughout this chapter and so we do not discuss results of these studies any further.

12.8 Areas of Future Interest

Despite all the modelling work discussed above, there remain several areas in which further modelling could elucidate various characteristics of UD gene drives. Some have briefly been mentioned in the relevant sections above and so we focus predominantly on areas not yet discussed.

Above, we discussed the use of laboratory cage trial experiments for inferring parameters of the UD system. While these are useful for predicting system performance, these estimates are inherently flawed when moving into the field since they assume that laboratory wild-type strains – and environments – are a good approximation of insects in the wild. In practice, laboratory wild-type strains are recognized as having lower fitness than their wild counterparts (Leftwich *et al*., 2021), likely due to many generations of laboratory adaptation (Leftwich *et al*., 2016; Ross *et al*., 2019). Models can capture the impact of this to a certain degree by considering variation of relative fitness parameters about laboratory-derived estimates. However, the transition toward field-based experiments will potentially necessitate more detailed models capturing a range of ecological, behavioural and fitness effects (some of which have been discussed above). This enhanced modelling can then help to inform the design of UD releases as they progress from small-scale field cage trials right up to the eventual release in full large-scale control programmes.

A feature of most of the modelling discussed here is that it is deterministic and so does not account for the stochasticity inherent in the real world. In the context of UD, this will be important when the release of transgenic insects results in a gene drive frequency close to (or even below) the introduction threshold calculated from deterministic mathematical models (i.e., the unstable equilibrium discussed above). Here stochastic models can provide insight into the expected likelihood of success or failure (i.e., the probability that a UD system increases or decreases in frequency) of a given release strategy. To date, stochastic modelling of UD systems has been limited, to our knowledge, to only Marshall and Hay (2012) for this UD configuration. However, some stochastic modelling frameworks have been used to study other gene drive classes, from which such work could take a lead (for example: Magori *et al*., 2009; Champer *et al*., 2020a; Edgington *et al*., 2020a,b; Sánchez *et al*., 2020a; Wu *et al*., 2021).

A common feature in the modelling of many gene drive classes, and, in particular, toxin–antidote-based approaches, is an assumption that toxins and antidotes are fully penetrant (i.e., that toxins kill 100% of target genotypes and antidotes rescue 100% of carriers to full fitness). However, gene drive components engineered in the laboratory may not give this degree of efficacy. Laboratory experiments, for example life history analysis of different genotypes, can provide initial estimates of such incomplete penetrance. These data could be incorporated into models similar to those in section 12.4. In the absence of working gene drive components to test in the laboratory, one can use the same model to explore the expected behaviour for a range of toxin and antidote penetrance parameters, thus setting performance targets for laboratory-engineered gene drive components. This could be used to assess performance metrics including threshold frequencies, the speed of spread, the system invasiveness (with n-deme versions of the models) and tolerable fitness costs – all of which are likely to be vital when transitioning from laboratory to field-based testing.

The motivation for genetic control of mosquitoes is to reduce or prevent morbidity and mortality from mosquito-borne diseases. Thus, it is important to explore the anticipated epidemiological impact(s) expected from a given gene drive and release

strategy. This can be explored by incorporating a gene drive model into a standard epidemiological modelling framework, for example susceptible–exposed–infectious–recovered (S-E-I-R) or a variety of extensions/modifications as previously considered for *Wolbachia* (Ndii *et al*., 2015, 2016a,b; Zhang and Lui, 2020) or RIDL (Atkinson *et al*., 2007) control approaches in *Ae. aegypti* mosquitoes. This will likely require the use of a population dynamics model similar to those of Khamis *et al*. (2018) and Edgington and Alphey (2018) for two main reasons: (i) models must produce results for gene drive and epidemiological dynamics at all time points; and (ii) the respective sizes of human and insect (vector) populations are important in determining a pathogen's force of infection. Such models can provide important insights into potential epidemiological impacts. However, various factors required to formulate these models (such as infection numbers, exact population sizes, transmissibility of pathogen(s) and biting frequency) can be extremely difficult to measure, meaning that the consideration of model uncertainty will be important when interpreting results.

This chapter has focused on the use of mathematical modelling to predict the efficacy of UD gene drives from molecular design and laboratory testing right through to field testing and final applications. Such studies will likely be important in providing an evidential basis upon which regulatory decisions can be made. As further laboratory-based testing provides more and higher-quality data, we would anticipate that more detailed and species-specific models will be developed, providing greater insight into the anticipated efficacy of UD systems. Likewise, as more field studies into the ecology of potential gene drive target species and field-trial releases (of this or other technologies) become available, more detailed ecological, epidemiological and behavioural factors can be studied and incorporated into models, enabling the best possible predictions of gene drive function following release of transgenic insects. The previous literature and future focus areas discussed here demonstrate the key role that modelling plays in the development of gene drive technologies and emphasizes the necessity for gene drive research and development to follow an interdisciplinary approach. This will ensure that any future gene drive release has the greatest opportunity to function as intended, thus providing the maximum possible beneficial impact.

Acknowledgements

MPE is funded through a Wellcome Trust Investigator Award [110117/Z/15/Z] made to LA. LA is supported by core funding from the UK Biotechnology and Biological Sciences Research Council (BBSRC) to The Pirbright Institute [BBS/E/I/00007033 and BBS/E/I/00007038].

References

Adolfi, A., Gantz, V.M., Jasinskiene, N., Lee, H.-F., Hwang, K. *et al.* (2020) Efficient population modification gene-drive rescue system in the malaria mosquito *Anopheles stephensi*. *Nature Communications* 11, 5553.

Akbari, O.S., Matzen, K.D., Marshall, J.M., Huang, H., Ward, C.M. and Hay, B.A. (2013) A synthetic gene drive system for local, reversible modification and suppression of insect populations. *Current Biology* 23, 671–677.

Alphey, L. (2014) Genetic control of mosquitoes. *Annual Review of Entomology* 59, 205–224.

Alphey, L.S., Crisanti, A., Randazzo, F. and Akbari, O.S. (2020) Opinion: Standardizing the definition of gene drive. *Proceedings of the National Academy of Sciences USA* 117, 30864–30867.

Alphey, N. and Bonsall, M.B. (2014) Interplay of population genetics and dynamics in the genetic control of mosquitoes. *Journal of The Royal Society Interface* 11, 20131071.

Altrock, P.M., Traulsen, A. and Reed, F.A. (2011) Stability properties of underdominance in finite subdivided populations. *PLoS Computational Biology* 7, e1002260.

Altrock, P.M., Traulsen, A., Reeves, R.G. and Reed, F.A. (2010) Using underdominance to bi-stably transform local populations. *Journal of Theoretical Biology* 267, 62–75.

Armistead, J.S., Arias, J.R., Nishimura, N. and Lounibos, L.P. (2008) Interspecific larval competition between *Aedes albopictus* and *Aedes japonicus* (Diptera: Culicidae) in northern Virginia. *Journal of Medical Entomology* 45, 629–637.

Atkinson, M.P., Su, Z., Alphey, N., Alphey, L.S., Coleman, P.G. and Wein, L.M. (2007) Analyzing the control of mosquito-borne diseases by a dominant lethal genetic system. *Proceedings of the National Academy of Sciences USA* 104, 9540–9545.

Beaghton, A., Beaghton, P.J. and Burt, A. (2016) Gene drive through a landscape: reaction–diffusion models of population suppression and elimination by a sex ratio distorter. *Theoretical Population Biology* 108, 51–69.

Bellows, T.S. (1981) The descriptive properties of some models for density dependence. *Journal of Animal Ecology* 50, 139–156.

Bonsall, M.B., Yakob, L., Alphey, N. and Alphey, L. (2010) Transgenic control of vectors: the effects of interspecific interactions. *Israel Journal of Ecology and Evolution* 56, 353–370.

Buchman, A., Marshall, J.M., Ostrovski, D., Yang, T. and Akbari, O.S. (2018a) Synthetically engineered *Medea* gene drive system in the worldwide crop pest *Drosophila suzukii*. *Proceedings of the National Academy of Sciences USA* 115, 4725–4730.

Buchman, A.B., Ivy, T., Marshall, J.M., Akbari, O.S. and Hay, B.A. (2018b) Engineered reciprocal chromosome translocations drive high threshold, reversible population replacement in *Drosophila*. *ACS Synthetic Biology* 7, 1359–1370.

Buchman, A., Gamez, S., Li, M., Antoshechkin, I., Li, H.-H. *et al.* (2019) Engineered resistance to Zika virus in transgenic *Aedes aegypti* expressing a polycistronic cluster of synthetic small RNAs. *Proceedings of the National Academy of Sciences USA* 116, 3656–3661.

Buchman, A., Shriner, I., Yang, T., Liu, J., Antoshechkin, I. *et al.* (2021) Engineered reproductively isolated species drive reversible population replacement. *Nature Communications* 12, 3281.

Champer, J., Buchman, A. and Akbari, O.S. (2016) Cheating evolution: engineering gene drives to manipulate the fate of wild populations. *Nature Reviews Genetics* 17, 146–159.

Champer, J., Reeves, R., Oh, S.Y., Liu, C., Liu, J. *et al.* (2017) Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. *PLoS Genetics* 13, e1006796.

Champer, J., Kim, I.K., Champer, S.E., Clark, A.G. and Messer, P.W. (2020a) Performance analysis of novel toxin-antidote CRISPR gene drive systems. *BMC Biology* 18, 27.

Champer, J., Lee, E., Yang, E., Liu, C., Clark, A.G. and Messer, P.W. (2020b) A toxin–antidote CRISPR gene drive system for regional population modification. *Nature Communications* 11, 1082.

Champer, J., Yang, E., Lee, E., Liu, J., Clark, A.G. and Messer, P.W. (2020c) A CRISPR homing gene drive targeting a haplolethal gene removes resistance alleles and successfully spreads through a cage population. *Proceedings of the National Academy of Sciences USA* 117, 24377–24383.

Champer, J., Zhao, J., Champer, S.E., Liu, J. and Messer, P.W. (2020d) Population dynamics of underdominance gene drive systems in continuous space. *ACS Synthetic Biology* 9, 779–792.

Champer, J., Champer, S.E., Kim, I.K., Clark, A.G. and Messer, P.W. (2021) Design and analysis of CRISPR-based underdominance toxin-antidote gene drives. *Evolutionary Applications* 14, 1052–1069.

Curtis, C. (1968) Possible use of translocations to fix desirable genes in insect pest populations. *Nature* 218, 368–369.

Davis, S., Bax, N. and Grewe, P. (2001) Engineered underdominance allows efficient and economical introgression of traits into pest populations. *Journal of Theoretical Biology* 212, 83–98.

Dhole, S., Vella, M.R., Lloyd, A.L. and Gould, F. (2018) Invasion and migration of spatially self-limiting gene drives: a comparative analysis. *Evolutionary Applications* 11, 794–808.

Dhole, S., Lloyd, A.L. and Gould, F. (2019) Tethered homing gene drives: a new design for spatially restricted population replacement and suppression. *Evolutionary Applications* 12, 1688–1702.

Dhole, S., Lloyd, A.L. and Gould, F. (2020) Gene drive dynamics in natural populations: the importance of density dependence, space, and sex. *Annual Review of Ecology, Evolution, and Systematics* 51, 505–531.

Edgerly, J.S., Willey, M.S. and Livdahl, T.P. (1993) The community ecology of *Aedes* egg hatching: implications for a mosquito invasion. *Ecological Entomology* 18, 123–128.

Edgington, M.P. and Alphey, L.S. (2017) Conditions for success of engineered underdominance gene drive systems. *Journal of Theoretical Biology* 430, 128–140.

Edgington, M.P. and Alphey, L.S. (2018) Population dynamics of engineered underdominance and killer-rescue gene drives in the control of disease vectors. *PLoS Computational Biology* 14, e1006059.

Edgington, M.P. and Alphey, L.S. (2019) Modeling the mutation and reversal of engineered underdominance gene drives. *Journal of Theoretical Biology* 479, 14–21.

Edgington, M.P., Harvey-Samuel, T. and Alphey, L. (2020a) Population-level multiplexing: a promising strategy to manage the evolution of resistance against gene drives targeting a neutral locus. *Evolutionary Aplications* 13, 1939–1948.

Edgington, M.P., Harvey-Samuel, T. and Alphey, L. (2020b) Split drive killer–rescue provides a novel threshold-dependent gene drive. *Scientific Reports* 10, 20520.

Ermentrout, B. (2002) *Simulating, Analyzing, and Animating Dynamical Systems: A Guide to XPPAUT for Researchers and Students*. SIAM, Philadelphia, Pennsylvania.

Franz, A.W., Sanchez-Vargas, I., Adelman, Z.N., Blair, C.D., Beaty, B.J. *et al.* (2006) Engineering RNA interference-based resistance to dengue virus type 2 in genetically modified *Aedes aegypti*. *Proceedings of the National Academy of Sciences USA* 103, 4198–4203.

Franz, A.W., Sanchez-Vargas, I., Raban, R.R., Black IV, W.C., James, A.A. and Olson, K.E. (2014) Fitness impact and stability of a transgene conferring resistance to dengue-2 virus following introgression into a genetically diverse *Aedes aegypti* strain. *PLoS Neglected Tropical Diseases* 8, e2833.

Haag-Liautard, C., Dorris, M., Maside, X., Macaskill, S., Halligan, D.L. *et al.* (2007) Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature* 445, 82–85.

Haller, B.C. and Messer, P.W. (2016) SLiM 2: flexible, interactive forward genetic simulations. *Molecular Biology and Evolution* 34, 230–240.

Haller, B.C. and Messer, P.W. (2019) SLiM 3: forward genetic simulations beyond the Wright–Fisher Model. *Molecular Biology and Evolution* 36, 632–637.

Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C. *et al.* (2016) A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. *Nature Biotechnology* 34, 78–83.

Hammond, A.M., Kyrou, K., Bruttini, M., North, A., Galizi, R. *et al.* (2017) The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *PLoS Genetics* 13, e1007039.

Hammond, A., Karlsson, X., Morianou, I., Kyrou, K., Beaghton, A. *et al.* (2020) Regulation of gene drive expression increases invasive potential and mitigates resistance. *bioRxiv* 360339.

Harvey-Samuel, T., Campbell, K., Edgington, M.P. and Alphey, L. (2019) Trialling gene drives to control invasive species: what, where and how? In Veitch, C.R., Clout, M.N., Martin, A.R., Russell, J.C. and West C.J. (eds) *Island Invasives: Scaling Up to Meet the Challenge*. Occasional Paper SSC no. 662. IUCN, Gland, Switzerland, pp. 618–627.

Huang, Y., Lloyd, A.L., Legros, M. and Gould, F. (2011) Gene-drive into insect populations with age and spatial structure: a theoretical assessment. *Evolutionary Applications* 4, 415–428.

James, A.A. (2005) Gene drive systems in mosquitoes: rules of the road. *Trends in Parasitology* 21, 64–67.

Juliano, S.A., O'Meara, G.F., Morrill, J.R. and Cutwa, M.M. (2002) Desiccation and thermal tolerance of eggs and the coexistence of competing mosquitoes. *Oecologia* 130, 458–469.

Keightley, P.D., Ness, R.W., Halligan, D.L. and Haddrill, P.R. (2014) Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. *Genetics* 196, 313–320.

Khamis, D., El Mouden, C., Kura, K. and Bonsall, M.B. (2018) Ecological effects on underdominance threshold drives for vector control. *Journal of Theoretical Biology* 456, 1–15.

Khamis, D., El Mouden, C., Kura, K. and Bonsall, M. (2020) The effect of dispersal and preferential mating on the genetic control of mosquitoes. *bioRxiv* 114413.

Kyrou, K., Hammond, A.M., Galizi, R., Kranjc, N., Burt, A. *et al.* (2018) A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged *Anopheles gambiae* mosquitoes. *Nature Biotechnology* 36(11), 1062–1066.

Lanzaro, G.C., Campos, M., Crepeau, M., Cornel, A., Estrada, A. *et al.* (2021) Selection of sites for field trials of genetically engineered mosquitoes with gene drive. *Evolutionary Applications* 14, 2147–2161.

Leftwich, P.T., Bolton, M. and Chapman, T. (2016) Evolutionary biology and genetic techniques for insect control. *Evolutionary Applications* 9, 212–230.

Leftwich, P.T., Edgington, M.P., Harvey-Samuel, T., Carabajal Paladino, L.Z., Norman, V.C. and Alphey, L. (2018) Recent advances in threshold-dependent gene drives for mosquitoes. *Biochemical Society Transactions* 46, 1203–1212.

Leftwich, P.T., Spurgin, L.G., Harvey-Samuel, T., Thomas, C.J.E., Paladino, L.C. *et al.* (2021) Genetic pest management and the background genetics of release strains. *Philosophical Transactions of the Royal Society B: Biological Sciences* 376, 20190805.

Long, K.C., Alphey, L., Annas, G.J., Bloss, C.S., Campbell, K.J. *et al.* (2020) Core commitments for field trials of gene drive organisms. *Science* 370(6523), 1417–1419.

Magori, K. and Gould, F. (2006) Genetically engineered underdominance for manipulation of pest populations: a deterministic model. *Genetics*, 172(4), 2613–2620.

Magori, K., Legros, M., Puente, M.E., Focks, D.A., Scott, T.W., Lloyd, A.L. and Gould, F. (2009) Skeeter Buster: a stochastic, spatially explicit modeling tool for studying *Aedes aegypti* population replacement and population suppression strategies. *PLoS Neglected Tropical Diseases* 3, e508.

Marshall, J.M. and Hay, B.A. (2012) Confinement of gene drive systems to local populations: a comparative analysis. *Journal of Theoretical Biology* 294, 153–171.

Maselko, M., Feltman, N., Upadhyay, A., Hayward, A., Das, S. *et al.* (2020) Engineering multiple species-like genetic incompatibilities in insects. *Nature Communications* 11(1), 4468.

Maynard Smith, J. and Slatkin, M. (1973) The stability of predator–prey systems. *Ecology* 54, 384–391.

NASEM (National Academies of Sciences, Engineering, and Medicine) (2016) *Gene Drives on the Horizon: Advancing Science, Navigating Uncertainty, and Aligning Research with Public Values*. The National Academies Press, Washington, DC.

Ndii, M.Z., Hickson, R.I., Allingham, D. and Mercer, G.N. (2015) Modelling the transmission dynamics of dengue in the presence of *Wolbachia*. *Mathematical Biosciences* 262, 157–166.

Ndii, M.Z., Allingham, D., Hickson, R.I. and Glass, K. (2016a) The effect of *Wolbachia* on dengue dynamics in the presence of two serotypes of dengue: symmetric and asymmetric epidemiological characteristics. *Epidemiology and Infection* 144, 2874–2882.

Ndii, M. Z., Allingham, D., Hickson, R.I. and Glass, K. (2016b) The effect of *Wolbachia* on dengue outbreaks when dengue is repeatedly introduced. *Theoretical Population Biology* 111, 9–15.

Oberhofer, G., Ivy, T. and Hay, B.A. (2019) Cleave and Rescue, a novel selfish genetic element and general strategy for gene drive. *Proceedings of the National Academy of Sciences USA* 116(13), 6250–6259.

Pham, T.B., Phong, C.H., Bennett, J.B., Hwang, K., Jasinskiene, N. *et al.* (2019) Experimental population modification of the malaria vector mosquito, *Anopheles stephensi*. *PLoS Genetics* 15, e1008440.

Reeves, R.G., Bryk, J., Altrock, P.M., Denton, J.A. and Reed, F.A. (2014) First steps towards underdominant genetic transformation of insect populations. *PLoS ONE* 9, e97557.

Ross, P.A., Endersby-Harshman, N.M. and Hoffmann, A.A. (2019) A comprehensive assessment of inbreeding and laboratory adaptation in *Aedes aegypti* mosquitoes. *Evolutionary Applications* 12(3), 572–586.

Sánchez Castellanos, H.M., Bennett, J.B., Wu, S.L., Rašić, G., Akbari, O.S. and Marshall, J.M. (2020a) Modeling confinement and reversibility of threshold-dependent gene drive systems in spatially-explicit *Aedes aegypti* populations. *BMC Biology* 18(1), 50.

Sánchez Castellanos, H.M., Wu, S.L., Bennett, J.B. and Marshall, J.M. (2020b) MGDrivE: A modular simulation framework for the spread of gene drives through spatially explicit mosquito populations. *Methods in Ecology and Evolution* 11, 229–239.

Simoni, A., Hammond, A.M., Beaghton, A.K., Galizi, R., Taxiarchi, C. *et al.* (2020) A male-biased sex-distorter gene drive for the human malaria vector *Anopheles gambiae*. *Nature biotechnology* 38(9), 1054–1060.

Sinkins, S.P. and Gould, F. (2006) Gene drive systems for insect disease vectors. *Nature Reviews Genetics* 7(6), 427–435.

Tanaka, H., Stone, H.A. and Nelson, D.R. (2017) Spatial gene drives and pushed genetic waves. *Proceedings of the National Academy of Sciences USA* 114, 8452–8457.

Webster, S.H., Vella, M.R. and Scott, M.J. (2020) Development and testing of a novel killer rescue self-limiting gene drive system in *Drosophila melanogaster*. *Proceedings of the Royal Society B: Biological Sciences* 287, 20192994.

Wilson, D.S. and Turelli, M. (1986) Stable underdominance and the evolutionary invasion of empty niches. *The American Naturalist* 127, 835–850.

Wimmer, E.A. (2013) Insect biotechnology: controllable replacement of disease vectors. *Current Biology* 23(10), R453-R456.

Windbichler, N., Menichelli, M., Papathanos, P.A., Thyme, S.B., Li, H. *et al.* (2011) A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature* 473(7346), 212–215.

WHO (2011) *Comprehensive Guideline for Prevention and Control of Dengue and Dengue Haemorrhagic Fever*. Revised and expanded edition. World Health Organization Regional Office for South-East Asia, New Delhi, India.

Wu, S.L., Bennett, J.B., Sánchez Castellanos, H.M., Dolgert, A.J., León, T.M. and Marshall, J.M. (2021) MGDrivE 2: A simulation framework for gene drive systems incorporating seasonality and epidemiological dynamics. *PLoS Computational Biology* 17, e1009030.

Zhang, H. and Lui, R. (2020) Releasing *Wolbachia*-infected *Aedes aegypti* to prevent the spread of dengue virus: a mathematical study. *Infectious Disease Modelling* 5, 142–160.