

## Evaluating the Probability of CRISPR-based Gene Drive Contaminating Another Species

Virginie Courtier-Orgogozo<sup>1</sup>, Antoine Danchin<sup>2</sup>, Pierre-Henri Gouyon<sup>3</sup> and Christophe Boëte<sup>4</sup>

<sup>1</sup>Institut Jacques Monod, CNRS, UMR 7592, Université de Paris, Paris, France

<sup>2</sup>Institut Cochin INSERM U1016 – CNRS UMR8104 – Université Paris Descartes, Paris, France

<sup>3</sup>Institut de Systématique, Évolution, Biodiversité, Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, UA, Paris, France

<sup>4</sup>ISEM, Université de Montpellier, CNRS, IRD, EPHE, Montpellier, France

### Abstract

The probability  $D$  that a given CRISPR-based gene drive element contaminates another, non-target species can be estimated by the following Drive Risk Assessment Quantitative Estimate (DRAQUE) Equation:

$D = (hyb+transf).express.cut.flank.immune.nonextinct$  with

*hyb* = probability of hybridization between the target species and a non-target species

*transf* = probability of horizontal transfer of a piece of DNA containing the gene drive cassette from the target species to a non-target species (with no hybridization)

*express* = probability that the *Cas9* and guide RNA genes are expressed

*cut* = probability that the CRISPR-guide RNA recognizes and cuts at a DNA site in the new host

*flank* = probability that the gene drive cassette inserts at the cut site

*immune* = probability that the immune system does not reject *Cas9*-expressing cells

*nonextinct* = probability of invasion of the drive within the population

We discuss and estimate each of the seven parameters of the equation, with particular emphasis on possible transfers within insects, and between rodents and humans. We conclude from current data that the probability of a gene drive cassette to contaminate another species is not insignificant. We propose strategies to reduce this risk and call for more work on estimating all the parameters of the formula.

### Abbreviations

CRISPR, Clustered Regularly Interspaced Short Palindromic Repeats

DRAQUE, Drive Risk Assessment Quantitative Estimate

HGT, horizontal gene transfer

HTT, horizontal transfer of transposable element

TE, transposable element

### Introduction

Selfish genetic elements are biological entities that favor their own transmission across generations. Examples include transposons that insert copies of themselves at other places in the genome, homing endonuclease genes that copy themselves at targeted

genomic sites, segregation distorters that destroy competing chromosomes during meiosis and maternally heritable microorganisms such as *Wolbachia* that favor progeny of infected females (Agren and Clark 2018). In recent years, researchers have started to develop CRISPR-based gene drives, named here gene drive for short, with the intention to spread synthetic genetic elements into wild populations. Potential applications of gene drives are numerous and include the elimination of mosquitoes to fight malaria, Zika, and other mosquito-borne diseases, or alternatively the modification of mosquitoes from vector to non-vector so that they no longer transmit human pathogens (Esvelt *et al.* 2014). Applications are not restricted to public health issues and also include agriculture, with for instance the elimination of invasive and pest species such as *Drosophila suzukii* or the suppression of herbicide resistance in weeds (Scott *et al.* 2018). Potential uses of gene drive are also reaching the field of conservation biology, with the targeting of rats (*Rattus rattus* and *Rattus norvegicus*) in New Zealand (Leitschuh *et al.* 2018; Rode *et al.* 2019). So far, CRISPR-based gene drives have only been tested in laboratories or in large indoor cages. They have been shown to efficiently boost their own transmission in yeasts (DiCarlo *et al.* 2015), *Drosophila* flies (Gantz and Bier 2015; Champer *et al.* 2017; KaramiNejadRanjbar *et al.* 2018), mosquitoes (Gantz *et al.* 2015; Hammond *et al.* 2016; Kyrou *et al.* 2018), the pathogenic fungus *Candida albicans* (Shapiro *et al.* 2018) and mice (Grunwald *et al.* 2019).

A CRISPR gene drive cassette is a piece of DNA that comprises several elements: (i) a gene encoding a guide RNA (gRNA) that can recognize a specific target DNA sequence, (ii) a *Cas9* gene encoding a *Cas9* endonuclease that can cut DNA at the site specified by the gRNA, (iii) sequences at the extremities that are homologous to sequences flanking the target site, so that the gene drive cassette can copy itself at the cleavage site via homology-directed repair and (iv) optional sequences, for example conferring a trait of interest such as malaria resistance (Esvelt *et al.* 2014). By converting heterozygotes for the gene drive allele into homozygotes, the gene drive cassette alters Mendelian transmission and can thus spread into wild populations. The release in the wild of a few individuals carrying gene drive constructs is thus expected to be sufficient to transform an entire population after a dozen generations (Deredec *et al.* 2008). Gene drives can be designed to introduce a phenotype of interest in a targeted population either through the introduction of a new gene, or by the inactivation of an endogenous gene via the insertion of the gene drive cassette into it (Esvelt *et al.* 2014). With introduced genetic changes that decrease viability or fertility, a gene drive can be used to eradicate a targeted population or to reduce its size, while with other types of genetic changes it is possible to alter the characteristics of a population.

The possibilities offered to humanity in terms of benefits by the new molecular techniques of genome edition CRISPR-Cas are innumerable but also associated with risks which should be carefully monitored (Zhang 2019). One obvious risk associated with gene drive is that the sequence may escape from the target species and spread into other species. Such spillover could have devastating effects, such as the extinction of a species, or the modification of a large number of individuals, with potentially important ecological consequences. Compared to natural bacterial CRISPR systems, gene drive cassettes are more compact and contain eukaryotic cis-regulatory elements, so that they are one step closer to potential contamination of non-target eukaryote species. The risk of gene drives contaminating another species has been mentioned by several authors (Benedict *et al.* 2008; Esvelt *et al.* 2014; Webber *et al.* 2015; National Academies of Sciences and Medicine 2016; Rode *et al.* 2019) but to our knowledge it has not been examined in detail. Risk assessment studies classically present the value of a risk as a product of two terms: Risk = Probability of occurrence x Damage in case of occurrence. The aim of this paper is to derive a formula to evaluate the probability of occurrence of a drive sequence escaping from the target species and endangering another species. The damage resulting from such an

unwanted event will obviously depend on the concerned species. If it is simply another mosquito, the damage will be limited (or this could even be seen as a serendipitous positive externality). If it were a keystone species or humans for instance, damages could be very important and hard, not to say impossible to mitigate. This paper does not address these questions but concentrates on the probability that a drive escapes from the target species and contaminates another species. Such an event results from a succession of events and the probability is thus the product of the corresponding conditional probabilities. Certain of these probabilities are still poorly known so that the estimates provided here are very rough. However, producing this formula can have two important effects. First, it can help developers of gene drive technology to contemplate the risk associated to their action and consider its potential magnitude. Second, it can trigger further research for better assessment of these probabilities. A famous example of such an approach is the Drake Equation which aimed at estimating the number of active, communicative extraterrestrial civilizations in the Milky Way galaxy (Burchell 2006). The equation was written in 1961 by Frank Drake, mainly as a way to stimulate scientific dialogue. A weakness of the Drake equation is that some factors are poorly known. However, it certainly promoted numerous research studies and thoughts among scientists.

Here we examine CRISPR-based gene drives and we review the different parameters to be considered to evaluate the risk of transfer to another species. We focus on gene drives that can spread autonomously and we do not consider here binary systems, whose genetic elements are located on different chromosomes. Such binary systems have been proposed as a solution to prevent contamination of non-target populations. However, they are expected to spread less efficiently than autonomous gene drives (Esvelt *et al.* 2014; Akbari *et al.* 2015).

## Results

For a given gene drive construct to contaminate another species, six consecutive steps are required (Fig. 1). The probability of contamination can be estimated by multiplying the probability of occurrence of each event, as defined with the following Drive Risk Assessment Quantitative Estimate (DRAQUE) formula:  $D = (hyb + transf) \times express \times cut \times flank \times immune \times nonextinct$ .

The bulk probability  $D$  of the gene drive to contaminate a non-target species includes (i) the probability that the given piece of DNA passes from the target species to the non target species. This can occur by hybridization (*hyb*) or by other means (*transf*). Then (ii) the *Cas9* gene and the guide RNA gene have to be expressed in the new host, with probability *express*. A target sequence must then be recognized and cut by CRISPR-Cas9 (iii) with probability *cut*. The gene drive cassette should insert at the cut site (iv) with probability *flank*; the immune system (v) must not eliminate it, with probability *immune* and finally, it must not be eliminated by stochastic processes (vi) with probability *nonextinct*.

This formula could be estimated for one given non-target species. However, for ease of estimation and practical purposes, we think that it is more relevant to directly define  $D$  as the probability that at least one non-target species is contaminated. We examine below each term of the DRAQUE formula.

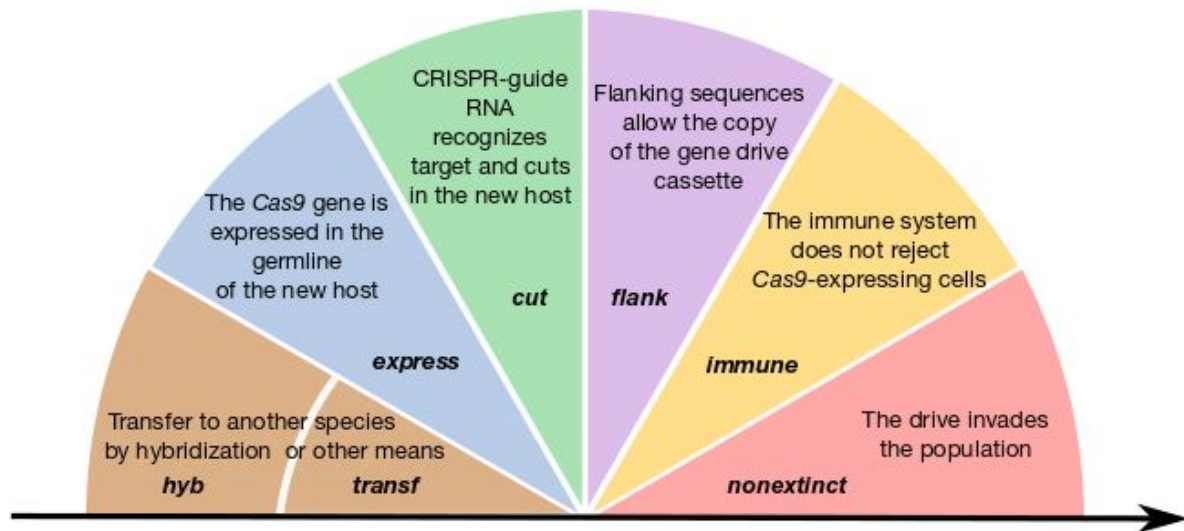


Fig.1. Summary of the different events whose probability must be estimated to assess the risk of gene drive contaminating another species.

### Probability of hybridization between the target species and a non-target species (*hyb*)

The probability that individuals of the target species hybridize with a non-target species and produce fertile progeny has to be evaluated for each target species, as it may vary among species. We treat here two taxa for which gene drive technology is most advanced, *Drosophila* flies and *Anopheles* mosquitoes.

*Drosophila suzukii* is an invasive pest species originating from southeast Asia that invaded both America and Europe since 2008 and that attacks ripe fruits (Scott *et al.* 2018). This species is one of the most advanced systems for potential gene drive applications (Scott *et al.* 2018). Two closely related species have been described, *Drosophila pulchrella* and *Drosophila subpulchrella*; they are found in Japan, China, India and Nepal [(Takamori *et al.* 2006) and references therein]. Recent genome data suggest that hybridization probably occurred recently between *D. suzukii* and *D. subpulchrella*, which diverged about 1-9 million years ago (Conner *et al.* 2017). Furthermore, fertile hybrids between *D. suzukii* and *D. subpulchrella* have been obtained in the laboratory [(Fuyama 1983), note that *D. subpulchrella* was erroneously named *D. pulchrella* in the 1983 paper (Muto *et al.* 2018)]. Because hybridization between closely related species usually produce hybrids with reduced fitness, it can lead to reinforcement, i.e. the increase of reproductive isolation as closely related species diverge (Turelli *et al.* 2014). Considering the prevalence of reinforcement in *Drosophila*, one might reasonably speculate that hybridization is common in the wild. Current data thus suggests that gene drives targeting *D. suzukii* may end up in *D. subpulchrella*/*D. pulchrella* if a gene drive ever reaches areas of contact on the Asian continent.

To control malaria with gene drive, two major strategies are currently being developed. One relies on the reduction or suppression of the population of vectors and the other on genetically modifying populations of wild vectors so that they no longer transmit pathogens. The most technically advanced approach is the one conducted by the Consortium Target Malaria aiming at reducing the population of several mosquito species of the *Anopheles gambiae* complex. This complex consists of at least 8 species of the *Anopheles* genus, morphologically indistinguishable and present in sub-Saharan Africa [(Coetzee *et al.* 2013) and references therein]. Some have a large afro-tropical distribution (*An. gambiae* s.s.) while others are restricted to savannah area (*An. arabiensis*) or coastal regions (*An. merus* and *An. melas*). It also includes the species *An. quadriannulatus* that is

not considered as a malaria vector. The work by Target Malaria is conducted on 3 species of this complex : *Anopheles gambiae* s.s., *Anopheles coluzzii* and *Anopheles arabiensis* (<https://targetmalaria.org/our-work/>). *An. coluzzii*, formerly known as *An. gambiae* M molecular form, is defined as a separate species since 2013 (Coetzee *et al.* 2013). Among the *Anopheles gambiae* complex, the question of hybridization has been a subject of interest for geneticists and public health practitioners for decades (Fontaine *et al.* 2015). Hybridization is of high concern and interest in these mosquito populations due to the potential spread of insecticide resistance between species, and now, with the development of gene drives, of human-made transgenes.

A classical example of transfer is the geographic expansion and adaptation to arid environment by *An. gambiae* that is associated with an introgression from *An. arabiensis* into *An. gambiae*, resulting from past hybridization between the two species (Besansky *et al.* 2003; Sharakhov *et al.* 2006). The presence of the *kdr* resistance (a mutation conferring resistance against pyrethroids, insecticides largely used in impregnated bednets) in the S form of *An. gambiae* (now *An. gambiae* s.s.) and then later in the M form (now called *An. coluzzi*) has also been explained by an introgression rather than by an independant, novel mutation (Weill *et al.* 2000). This highlights the existence of gene flow between these two species. Recent studies have also highlighted high frequency of hybridization between *An. coluzzi* to *An. gambiae* in West Africa (Oliveira *et al.* 2008; Caputo *et al.* 2011; Marsden *et al.* 2011) and an asymmetric introgression from *An. coluzzi* to *An. gambiae* (Mancini *et al.* 2015). Genetic exchanges have also been detected between *An. gambiae* s.s. and *An. arabiensis* and led to the idea of particular genomic regions being more prone to cross species boundaries than others (Crawford *et al.* 2015). In the lab, introgression of a synthetic sex ratio distortion system has even been possible from *An. gambiae* to its sibling species *An. arabiensis* (Bernardini *et al.* 2019). In summary, multiple species inside the *An. gambiae* complex appear to cross-hybridize, and developers of gene drive technology may aim to develop a system able to target several of them. The higher the number of species, and thus individuals, harboring gene drive constructs, the higher the probability that the gene drive contaminates other species.

### **Probability of horizontal transfer of a piece of DNA containing the gene drive cassette from the target species to a non-target species with no hybridization (*transf*)**

DNA can be naturally transferred from one eukaryote species to another via so-called horizontal transfer (HT), through unknown means that may involve vectors such as viruses, microsporidia, mites or parasitoids (Houck *et al.* 1991; Gilbert *et al.* 2010, 2014; Parisot *et al.* 2014). In particular, viruses can carry over nucleic acid loads that do not directly belong to the viral specific genetic setup, but from the virus host (Gasmi *et al.* 2015; Gilbert and Cordaux 2017). Bacteria probably also constitute an important vector to transfer DNA material from one Eukarya species to another. HGT from Bacteria to Eukarya is frequent (Lacroix and Citovsky 2016) and from Bacteria to Bacteria it is the norm. This implies that environmental DNA can be rapidly contaminated by any novel construct. As cases in point, apparition of antibiotic resistance spreads rapidly from locations where it has first appeared [see e.g. (Schultz *et al.* 2017)]. HGT from Eukarya to Bacteria has been illustrated in some cases, with viruses as intermediates (Bordenstein and Bordenstein 2016). Eukaryotic genes in bacteria (EUGENS) are frequent in intracellular parasitic or symbiotic bacteria [e.g. (Hilbi *et al.* 2011)]; they may play an efficient role in gene exchange. The DNA pieces that have been observed to undergo HT are usually between 1 kb and several dozens of kb [see HTT-DB database (Dotto *et al.* 2015)], which is comparable to the size of gene drive constructs, and can go up to 150 kb in plants and animals (Inoue *et al.* 2017; Dunning *et al.* 2019).

Two types of horizontal transfer can be distinguished, horizontal transfer of transposable elements (HTT) and horizontal gene transfer (HGT). Although transposable elements (TE) also carry genes, this distinction reflects the fact that we know many cases of HTT and relatively few cases of HGT. Most HGT events that occurred in the past between distantly related species are not expected to be detected through comparison of present-day genomes because most newly inserted DNA sequences are likely to be lost by genetic drift (i.e. the probability *nonextinct* is zero). The breadth of HGT that can be approached based on comparative studies of genome sequences is thus largely underestimated. As a matter of fact, identified HGT events involve pieces of DNA that appear to increase host fitness [HGT, e.g. carotenoid synthesis genes from fungi to pea aphids (Moran and Jarvik 2010)]. Compared to other DNA sequences, TEs have particular characteristics that allow them to integrate into DNA more frequently, and they can also self-replicate in the new host after HT, so that they are more likely to be noticed. Whether the higher rate observed for HTT than for HGT is only due to the integration and replicative properties of TE is unknown. If viruses are important vectors for HT, the propensity of TE to jump from eukaryote genomes to viruses, and reversely, more frequently than random pieces of DNA (Gilbert *et al.* 2016) may also explain their higher rate of HT. Compared to TE, a gene drive cassette can also insert itself into a host genome, but its integration into DNA may be less likely and its number of copies in a genome should be lower, so that gene drives may transfer horizontally between genomes less frequently than TEs. However, if TEs are present in the vicinity of the gene drive cassette in the target species, they could end up facilitating the transfer and integration of the DNA in another host via a hitchhiking process. To limit this risk, gene drives should be designed to target genomic regions that are devoid of TEs, if such regions exist in the target species.

Better than TEs, human-made gene drive constructs resemble homing endonuclease genes, which are naturally occurring mobile elements that bias their inheritance by cutting and inserting themselves at targeted sites within genomes (Burt and Koufopanou 2004; Agren and Clark 2018). A homing endonuclease gene that specifically targets the *cox1* mitochondrial gene has been transferred independently 70 times between 162 plant species within 45 different families (Sanchez-Puerta *et al.* 2008). This element is also present in several species of fungi, green algae and liverworts, highlighting extensive HT (Cho *et al.* 1998). Unfortunately, no estimate of HT rates for homing endonuclease genes is available.

Cases of HTT have been detected between extremely distantly related species (Gilbert and Feschotte 2018). For example, the *BovB* element moved at least 11 times between snakes, lizards, ruminants and marsupials (Ivancevic *et al.* 2017) and the *Mariner* element moved between nematodes, arthropods, fungi, molluscs, vertebrates and plants [(Palazzo *et al.* 2019) and references therein]. Numerous stable introductions of virus DNA into the germline of various eukaryote species have also been reported (Katzourakis and Gifford 2010; Holmes 2011; Feschotte and Gilbert 2012; Chen *et al.* 2016). In vertebrates, endogenous retroviruses (ERVs) can insert into their host genome and generate copies of themselves through germline reinfections or retrotransposition events (Gilbert and Feschotte 2018; Greenwood *et al.* 2018). A comparison of endogenous retrovirus sequences in 65 genomes identified no less than 1000 HT events between distantly related species of vertebrates (Hayward *et al.* 2015). Pan-phylogenomic analyses of endogenous retrovirus sequences revealed that rodents are a major source of retroviruses which they can transmit to other mammals such as livestock (Cui *et al.* 2015).

HT events are also ongoing now. The best known case is the worldwide invasion of *D. melanogaster* populations by a transposable element (*P*-element) originally present in the distantly-related species *D. willistoni* (Daniels *et al.* 1990; Clark and Kidwell 1997). This invasion has been carefully recorded and occurred within a few decades during the second

half of the last century. Nowadays, the *P*-element is invading two other *Drosophila* fly species worldwide, *D. simulans*, originating from *D. melanogaster* (Hill *et al.* 2016) and *D. yakuba* (Serrato-Capuchina *et al.* 2018). *D. simulans* flies sampled before 2010 do not carry the *P*-element, indicating that this invasion is very recent. The *P*-element arose in *D. simulans* most likely through hybridization with *D. melanogaster*, but could also have occurred via the unintended escape of a few laboratory-raised *D. simulans* flies genetically engineered to carry the *D. melanogaster* *P*-element. In Mammals, the endogenous retrovirus sequence KoRV-A is currently observed to invade natural populations of koala (Xu and Eiden 2015) and herpes virus DNA has started to integrate into human chromosomes (Morissette and Flamand 2010).

Rough estimates of HTT rates have been obtained recently. Systematic surveys of transposable elements in complete genomes have inferred HTT events that essentially occurred during the last 10 My. They counted more than 2000 HTT events in 195 insect species (Peccoud *et al.* 2017) and more than 330 HTT in 460 arthropod species (Reiss *et al.* 2019). El Baidouri *et al.* (2014) estimated that more than 2 million HTT occurred between plant species. A study of three *Drosophila* species estimated an average rate of 0.035 HTT per transposable element family per million years between these three species, with LTR RTs and DNA transposons displaying a higher rate of 0.046 (Bartolomé *et al.* 2009). Note that the power to identify HTT events decreases as the HTT events approach the time of species divergence, so that existing quantifications are conservative.

Based on the inventory of species inhabiting Great Smoky Mountains National Park (which harbors more than 500 vertebrate species and >3900 insect species - <https://dliia.org/smokies-species-tally>), we can assume that a given species can be in contact with approximately 1000 distinct Eukaryote species, either directly or indirectly via viruses, bacteria or other microorganisms. As a result, the rate of HTT can be approximated to a minimum of 0.035 transfer events to at least one species per thousand of years. This estimate is conservative as it does not take into account horizontally transferred DNA that have disappeared from the recipient genome since their transfer nor HTT events that occurred right after reproductive isolation of the two lineages of interest. In addition, due to the broad geographic distribution of viruses and bacteria that can act as vectors, the probability of transfer to any species among the total estimated 10 millions in the world is probably much higher than this figure. Furthermore, there is no reason to believe that the rate of HTT is constant across time. It is possible that the number of HTT events increases under certain conditions, such as ecological stress or pervasive pathogen infections (Horváth *et al.* 2017). This stress sensitivity has been witnessed as a cause of evolution from commensalism to pathogenicity driven by an intra-organism explosion of TEs (Magaziner *et al.* 2019).

The risk of transfer to non-target species also depends on the persistence time of gene drives. For replacement/rescue drives, the drive is expected to reach all individuals of the targeted population. Then, in the long term, since there is no selective pressure to maintain a functional endonuclease, the CRISPR-cas9 cassette can eventually accumulate mutations and thus lose its self-replicating activity, and should eventually disappear from the target population. However, the persistence time of the gene drive cassette should still be relatively long, of the order of several thousands of generations, leaving time at the human scale for possible HGT. The long term presence of rescue/replacement drive constructs in the population increases the odds of transfer to non-target species. As far as we know, no strategy has been proposed to completely remove the gene drive cassette once a population has been fully targeted by a replacement drive.

For eradication/suppression drives, the target population is expected to go extinct, so that no gene drive construct is expected to remain in living organisms. Nevertheless, DNA is a very stable molecule and DNA from dead organisms can make up reservoirs of gene drive

constructs. Just taking into account viruses, aquatic ecosystems typically contain virus-like particles and sediments  $10^8$ - $10^9$  particles (Filippini and Middelboe 2007). DNA is found in all kinds of environments [(see (Hunter *et al.* 2019) for techniques allowing recovery of 1 to 500 ng of DNA per microliter of water]. Some organisms, such as *Acinetobacter baylyi* are able to take it spontaneously into their genome (Mantilla-Calderon *et al.* 2019), further amplifying and propagating DNA sequences. Subsequently, bacteria can transfer DNA into their hosts chromosomes following successful metabolic pathways, in particular via symbiotic organisms, as we have discussed previously. Gram negative bacteria are cases in point as it has been shown that they rapidly transfer antibiotic resistance to a large number of recipients (Oliveira and Reygaert 2019).

In summary, recent data indicate that DNA can transfer extensively between distantly-related species in all taxon groups. Current quantifications of the probability of transfer of a particular DNA piece to another species are rare and underestimated, and provide a probability of a minimum of 0.035 transfer events to at least one species among 1000 per thousand years. Whereas the probability of hybridization discussed above is relatively high and concerns a small number of species, the probability of horizontal transfer discussed here is relatively low but involves a larger number of potential non-target species. These non-target species can be very distant both phylogenetically and geographically, due to vectors which can be ubiquitous.

### **Probability that the guide RNA and Cas9 genes are expressed in the new host (express)**

For the CRISPR-Cas9 gene drive system to be active in the non-targeted species, the guide RNA gene and the Cas9 gene must be expressed in the newly-formed zygote or in the new host germline (Fig. 2). In other words, potent enhancer regions should be present in the vicinity of the two genes, and promoters should be active to drive expression in the new host (Wittkopp and Kalay 2012). Experiments swapping enhancers between species suggest that rodent sequences are active across Mammals, while fly/mosquito sequences function across Diptera and sometimes throughout insects. Human enhancers generally drive similar expression in mice [88 My divergence - all divergence times are from TimeTree (Kumar *et al.* 2017)] (Cheng *et al.* 2014). Enhancers have also been observed to drive expression in corresponding homologous organs in more distantly related species such as mice and bats [94 My divergence (Cretkos *et al.* 2008; Booker *et al.* 2016)] or even fish species and humans [465 My divergence (Yuan *et al.* 2018)]. A particular enhancer was shown to drive comparable expression across 9 species of vertebrates including fishes, lizards, snakes and mice (Kvon *et al.* 2016). In insects, several enhancers have been tested and observed to drive similar expression patterns in *Drosophila*, mosquitoes (248 My divergence) and *Tribolium* [309 My divergence (Cande *et al.* 2009; Lai *et al.* 2018)]. Several native promoters of *Drosophila* have been tested and they have been found to function in Lepidoptera [286 My divergence (Tamura *et al.* 2000; Imamura *et al.* 2003; Ramos *et al.* 2006)] but not in *Tribolium* (Schinko *et al.* 2010). Native mosquito enhancers are thus unlikely to function in humans (790 My divergence) or in the malaria parasite (1552 My divergence).

To activate expression, published gene drive constructs harbor endogenous cis-regulatory sequences of germline-specific genes for the Cas9 gene and of constitutively expressed genes for the guide RNA genes (Table 1). Therefore, present-day constructs designed for flies or mosquitoes can be expected to drive expression in the germline across Diptera/insects, and for mice across Mammals.

Synthetic regulatory elements containing multiple adjacent binding sites for a small number of transcriptional activators have not been used so far in published gene drive constructs but they might in the future. In general, such synthetic pieces tend to be more

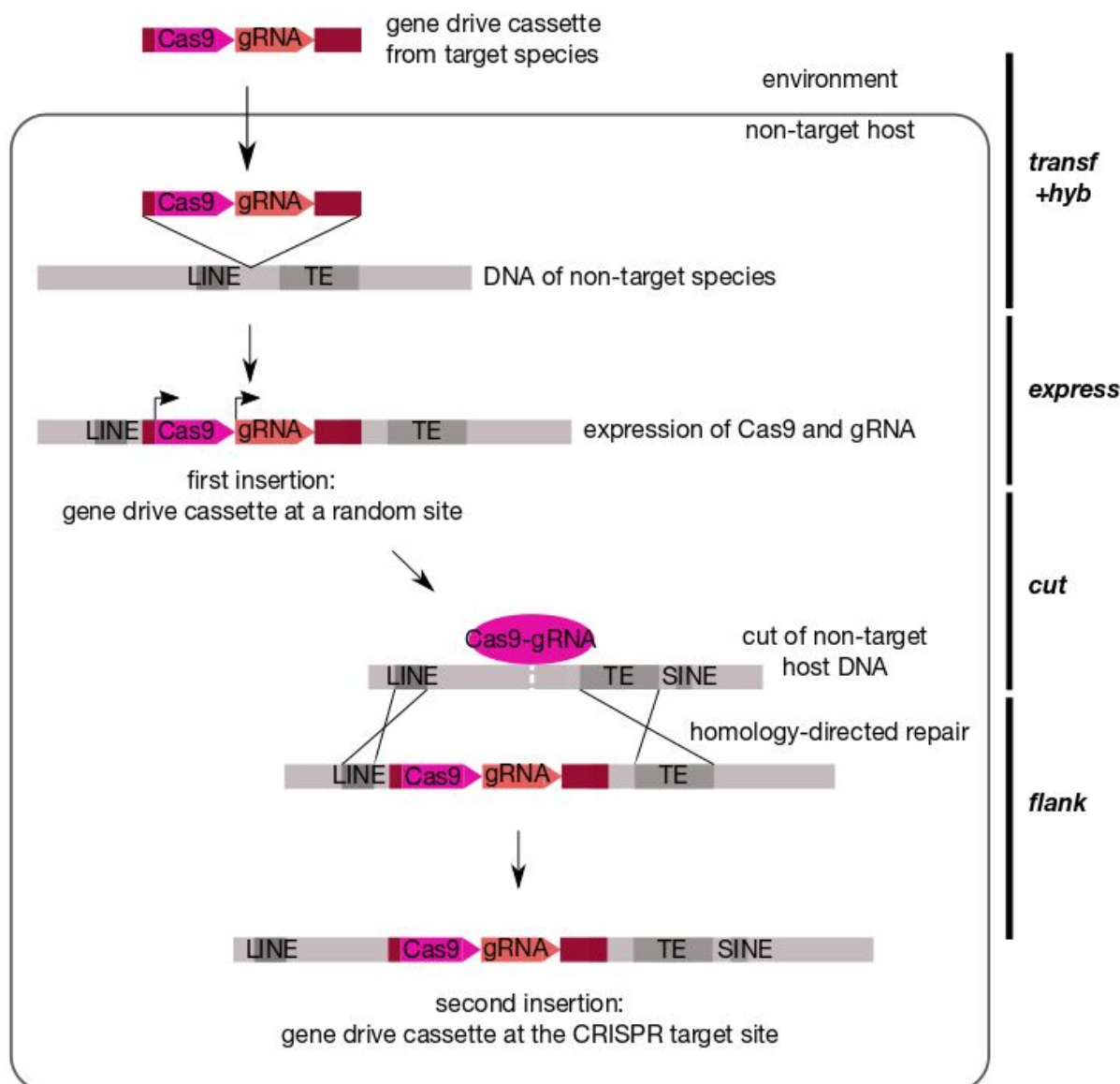


universal than endogenous regulatory elements (Schetelig and Wimmer 2011). For example, the *Drosophila*-derived *3xP3* artificial element drives expression in the eyes not only in insects but also in arthropods (Pavlopoulos and Averof 2005) and planarians (Gonzalez-Estevez *et al.* 2003).

The guide RNA gene and the *Cas9* gene can also be activated by enhancers located further away, outside of the original gene drive cassette. Reciprocally, vectors used for gene therapy have been shown to misregulate endogenous genes neighboring the inserted DNA, thereby causing harmful side effects such as leukemia (Browning and Trobridge 2016). The addition of chromatin insulators into retroviral vectors, to block the activation of nearby genes, seems to be a good solution to this problem, as shown by a recent successful gene therapy (Mamcarz *et al.* 2019). To our knowledge, none of the published gene drive constructs contain insulators. To limit the activity of gene drive constructs in non-targeted species, we suggest that gene drive vectors should include insulators.

Target Species	Cis-regulatory sequences for <i>Cas9</i> expression	Cis-regulatory sequences for guide RNA expression	Reference
<i>Drosophila melanogaster</i>	<i>D. melanogaster vasa</i>	<i>D. melanogaster U6:3</i>	(Gantz and Bier 2015)
<i>Anopheles stephensi</i>	<i>A. stephensi vasa</i>	<i>A. stephensi U6A</i>	(Gantz <i>et al.</i> 2015)
<i>Anopheles gambiae</i>	<i>A. gambiae vasa2</i>	<i>A. gambiae U6</i>	(Hammond <i>et al.</i> 2016)
<i>Drosophila melanogaster</i>	<i>D. melanogaster vasa</i> <i>D. melanogaster nanos</i>	<i>D. melanogaster U6:3</i>	(Champer <i>et al.</i> 2017)
<i>Drosophila melanogaster</i>	<i>D. melanogaster Sry-alpha</i> <i>D. melanogaster DNAPol-α180</i> <i>D. melanogaster Rcd-1r</i>	<i>D. melanogaster U6:3</i>	(KaramiNejadRanjbar <i>et al.</i> 2018)
<i>Anopheles gambiae</i>	<i>A. gambiae zpg</i>	<i>A. gambiae U6</i>	(Kyrou <i>et al.</i> 2018)
<i>Mus musculus</i>	<i>M. musculus vasa</i> <i>M. musculus Stra8</i>	<i>H. sapiens U6</i>	(Grunwald <i>et al.</i> 2019)

**Table 1. List of cis-regulatory sequences used in published gene drive constructs.** Information is presented in the chronological order of the publications. All are endogenous cis-regulatory sequences. The genes whose cis-regulatory sequences were used to drive *Cas9* expression are expressed specifically in the germline. *U6* is a ubiquitously expressed gene encoding a small RNA involved in mRNA splicing. For *Cas9* expression in *M. musculus*, the germline-specific cis-regulatory sequences were used in combination with the Cre-loxP system and constitutive cis-regulatory sequences.



**Fig. 2. Succession of putative events leading to the insertion of a functional gene drive cassette in another distantly related non-target species (in case of horizontal transfer with no hybridization).** DNA of the non-target species is indicated in gray and DNA from the target species in warm colors: pink: Cas9 gene, orange: gRNA gene and brown: neighboring sequences. On the right are different parameters of the DRAQUE equation. In the represented scenario, one LINE (Long INterspersed Element) and one Transposable Element (TE) serve as flanking sequences for homology-directed repair. Note that other sequences present in the non-target host may also be used.

### Probability that the CRISPR-guide RNA recognizes and cuts at a DNA site in the new host (*cut*)

So far, the CRISPR-Cas9 system has been shown to cut at target sites specified by the guide RNA in all species that have been tested, including animals, plants, bacteria and parasites (Zhang 2019). We therefore consider here that the probability *cut* that the CRISPR-guide RNA recognizes and cuts at a DNA site in the new host is simply the probability that the host genome contains a site targeted by the CRISPR-guide RNA of the gene drive construct. The target site recognized by CRISPR-Cas9 is 20 nucleotides long

followed by a protospacer adjacent motif (PAM), typically NGG (Zhang 2019). Importantly, studies in yeast and human cells indicate that CRISPR-Cas9 cleavage activity can still occur with three to five base pair mismatches in the 5' end (Fu *et al.* 2013; Hsu *et al.* 2013; Roggenkamp *et al.* 2018). As a consequence, we need here to estimate the probability of finding a particular sequence of 17-19 nucleotides (plus one nucleotide that can be either base, located two bases before the 3' end of the segment) in genomes of interest. The existence of other off-target cuts with lower sequence similarity to the on-target site is still under exploration and not entirely understood (Zhang *et al.* 2015; Gao *et al.* 2019). The number of off-target cuts tends to accumulate with longer and stronger expression of the Cas9 gene (Kim *et al.* 2014), suggesting that the range of off-targets may differ between gene drive cassettes. We consider here only the sequences that are closely related to the on-target site and therefore provide a lower estimate of the probability *cut*.

A very rough estimate of the probability for a given sequence of 19 nucleotides to be present in a genome of a billion nucleotides can be estimated as follows. Assuming that the 4 nucleotides A, C, G and T are equiprobable, the probability of finding each sequence is  $2/4^{19}$  (the factor 2 stands for the fact that DNA has two strands) which is about  $7 \cdot 10^{-12}$ . In a genome of  $3 \cdot 10^9$  nucleotides such as the human haploid genome, such a sequence should be present with a probability  $7 \cdot 10^{-12} \times 3 \cdot 10^9$ , which is approximately 2% (about 30% for a sequence of 17 nucleotides). Note that this is a very crude estimate. It could be higher if one took into account the fact that the DNA contents of AT and GC are different. Also, it does not take into account that an extant DNA sequence is never random. The non random character of genomes is illustrated by the omnipresence of repeated sequences (for example in the human genome the repeats Alu, SINE, LINE etc.) so that if a given sequence is present in a genome, it is likely to be present more than once. Overall it appeared that genomes behave as n-plications of a core set of sequences followed by reduction. This has been observed in yeasts (Escalera-Fanjul *et al.* 2018), plants (Clark and Donoghue 2018a) and animals (Hermansen *et al.* 2016).

The fact that the target sequence is not an arbitrary sequence can increase or decrease the probability according to whether the bias for the sequence of interest and the bias for the sequences in the genome both go in the same direction or not. The probability *cut* would be lower if the existence of repeated sequences was taken into account and if the targeted sequence did not match some of the repeated sequences. By contrast, it would be higher if the sequence matched the "style" of the DNA of a particular organism, i.e. its non random content in particular motifs (Fertil *et al.* 2005). Different mechanisms can lead to the occurrence of repeated sequences in the genome (gene conversion, unequal gene exchanges, transposition, etc.); they have been grouped under the name of "molecular drive" (Dover 1982). "Molecular drive" is likely to be widespread, as a sizeable proportion of individuals carry local duplications of any sequence of the genome. Furthermore, some conserved repeats may maintain the coexistence of stable rearrangements in some species (Smalec *et al.* 2019), increasing the possibility of unexpected cuts in certain species. All these features may considerably impact the probability of cuts in particular genomes. If the targeted site is present in multiple copies, there is a risk for the gene drive construct to spread across the entire genome. We urge the researchers developing gene drive constructs to make sure that they choose a target site that is very distinct from sequences such as retroelements, LINE, SINE elements that are present in large quantities in genomes [see (Breitwieser *et al.* 2019) for an updated view of repeats in the human genome].

A recently proposed gene drive that aims at sterilizing mosquito females has the following target site, GTTTAACACAGGTCAAGCGGNGG, which is a highly conserved sequence within the *Doublesex* gene, displaying a 3' terminal end that contains either a repeated CGG triplet or variant of it. We should note that CGG repeats are quite frequent in

genomes (Pan *et al.* 2018; Rabeh *et al.* 2018). The 23-bp sequence is present in 7 species of *Anopheles* mosquitoes, and with one mismatch in at least 6 additional *Anopheles* species (Kyrou *et al.* 2018). Sequencing of 765 wild-caught *Anopheles gambiae* mosquitoes identified only one single nucleotide variant within this sequence, and this variant was still permissive to gene drive. Our BLAST searches for this target sequence and for the one used in the published mouse gene drive found fragments of up to 20 bases identical to the 3' part of the sequence in several genomes, belonging to all three domains of life (Table S1). While a target sequence which accommodates little nucleotide variation such as the *Doublesex* locus can be useful to prevent gene drive resistance, it is then also associated with an increased risk of spread to non-target species. This trade-off exists not only for the *Anopheles gambiae* mosquito complex but for any species targeted by a gene drive system.

### **Probability that the gene drive cassette inserts near the targeted site (*flank*)**

The gene drive cassette is designed to bias its transmission by copying itself on the paired chromosome, so that heterozygous individuals carrying initially one copy of the gene drive cassette end up with two copies in their germline cells (Esvelt *et al.* 2014). This homing process occurs through homology-directed repair, using homology arms flanking both the gene drive cassette and the guide RNA target site. Therefore, for the gene drive to be active in non-target species, the DNA containing the *Cas9* gene and the *gRNA* gene should be inserted at the guide RNA target site. *Flank* is the probability that the gene drive cassette lands up at the guide RNA target site.

In case of hybridization, the non-target species being closely related to the species carrying the initial gene drive, the genomic regions harboring the gene drive cassette are expected to be comparable, so that the probability *flank* should be close to 1.

In case of horizontal transfer to a distantly related species with no hybridization, the gene drive cassette is likely to insert at a random site in the genome. In this case, *flank* is the probability that it moves from this initial position to a site targeted by its guide RNA (Fig. 2). There are several ways such a transposition can happen. Double-strand DNA breaks, such as the one created at the guide RNA target site by the CRISPR system, are known to induce recombination and various DNA repair mechanisms (Jasin 1996; Hartlerode and Scully 2009).

First, the gene drive cassette may move to the guide RNA target site via homology-directed repair (Fig. 2). Such phenomenon has been observed with gene drive constructs that were not inserted initially at the target site but at another position in the genome (Gantz and Bier 2015; Guichard *et al.* 2019). The probability of such an event will depend on the length of the flanking homology arms, their percentage of identity, the length of the DNA sequence in between and the position of the donor sequence relative to the cut site (Wang *et al.* 2017). Unfortunately, as far as we know, no extensive survey of these parameters has been done with respect to homology-dependent repair using another chromosomal locus as template. A gene drive cassette of 21 kb was found to spread effectively in mosquitoes (Gantz *et al.* 2015), suggesting that relatively large pieces of DNA can be inserted via homology-directed repair. In any case, larger inserts tend to show lower efficiency of recombination (Li *et al.* 2014). Gene drive constructs published to date have used flanking sequences of about 1kb (Gantz and Bier 2015; Gantz *et al.* 2015; Grunwald *et al.* 2019). When linear double stranded DNA is used as a template for repair, efficient targeted genome integration can be obtained using flanking sequences that are only 50-bp long in mammalian cells (Li *et al.* 2014; Wierson *et al.* 2018) and even 20-40-bp long in zebrafish (Auer and Del Bene 2014; Hisano *et al.* 2015; Wierson *et al.* 2018), *Xenopus laevis* (Nakade *et al.* 2014), *Bombyx mori* (Nakade *et al.* 2014) or the nematode *C. elegans* (Paix *et al.* 2014). Whether such short sequences could also favor transposition of the gene

drive cassette to the guide RNA target site remains unknown. If repeat sequences exist both near the initial insertion site of the gene drive and the guide RNA site, they may facilitate such transposition (Fig. 2). Evaluating the parameter *flank* thus requires an assessment of the distribution of repeats in potential non-target species.

As emphasized by Salzberg and coworkers, the human genome sequence, while far more complete than most animal genomes, is still made of 473 scaffolds and comprises 875 gaps (Breitwieser *et al.* 2019). As expected, the gaps encompass regions with a variety of repeats, some of them still poorly characterized in particular in centromeric and pericentromeric regions. This precludes accurate analysis of their distribution, and the situation is even worse for other genomes. For instance, the transposon-derived Alu repeats (~300 bp long), that are present in primate genomes at more than one million copies, are widely variable, within and among chromosomes (Grover *et al.* 2004). Interestingly this distribution is biased towards proximity of protein-coding gene regulatory regions (Lavi and Carmel 2018). In addition to our somewhat limited knowledge of the distribution of repeats in mammalian genomes we have to consider that when a species is represented by a large number of individuals there are many copy number variants, in particular in repeated regions (Monlong *et al.* 2018). This is an unavoidable consequence of involvement of satellite DNA in recombination at heterologous sites, a process which is still a matter of speculation (Satović *et al.* 2016).

Second, the gene drive cassette may move to the guide RNA target site via a transposable element. It is well established that transposable elements play a considerable role in displacement of genes or regulatory regions across genomes (Chen and Yang 2017). As a matter of fact TEs are recognized as a frequent cause of genetic diseases [see for example (Larsen *et al.* 2018; Song *et al.* 2018) and references therein].

The events discussed above may seem extremely rare, but they do not have to occur right after the gene drive inserted in a genome. The gene drive cassette may stay dormant for a few generations, and there can be many trials and errors in various individuals before an insertion occurs at the guide RNA target site. Of note, gene drive constructs containing several guide RNAs (Esvelt *et al.* 2014) increase the chance that the drive moves to a targeted site.

### **Probability that the immune system does not eliminate Cas9-expressing cells (*immune*)**

The Cas9 protein is derived from the bacteria *Streptococcus pyogenes* (Zhang 2019) and can trigger an immune response in mice (Chew *et al.* 2016; Chew 2018). During the gene drive multiplication process, germline cells produce Cas9 proteins to cut the DNA and insert the gene drive cassette (Fig. 2). These germline cells may thus present Cas9 fragments on their surface.

In insects, the immune system is not known to recognize novel foreign molecules (Lemaitre and Hoffmann 2007), hence the presence of Cas9 is unlikely to trigger an immune response. However, double stranded RNAs larger than 30bp can be recognized by Dicer2 and activate the RNA interference pathway, leading to their degradation (Elbashir *et al.* 2001; Gammon and Mello 2015). Guide RNAs are about 100 nucleotide long including the target site and they contain hairpins shorter than 15bp (Bassett and Liu 2014), so they should not be recognized by Dicer2. In summary, present knowledge thus suggest that gene drives would not be hampered by the immune system in insects.

In vertebrates, if Cas9 proteins accumulate in somatic tissues at a late stage during development due to leakage of the cis-regulatory regions controlling Cas9 gene expression, Cas9 fragments may be recognized as foreign molecules and trigger an adaptive immune T cell response, leading to the potential elimination of the Cas9-expressing cells and a

probable decrease in fitness of the individual carrying the gene drive (Chew 2018). The expression of *Cas9* at an early stage of development may also activate an immune response if the gene drive carrier has inherited anti-*Cas9* antibodies from its mother, as maternal immunoglobulins G have been shown to cross the placental barrier and the intestine, and to be maintained for a long time in the fetus after birth (Madani and Heiner 1989; Roopenian and Akilesh 2007). However, testes - and maybe ovaries - appear to readily accept foreign antigens without the induction of an immune response in several mammals (Simpson 2006; Mellor and Munn 2008; Li *et al.* 2012), so that germline cells expressing *Cas9* may not be eliminated by the immune system. Furthermore, guide RNAs produced from gene drive constructs are not expected to elicit an immune response in vertebrates, as they do not carry 5'-triphosphate ends (Kim *et al.* 2018; Wienert *et al.* 2018). Clearly, our knowledge in immunology is presently too sparse to anticipate how gene drive systems will interact with the immune system in vertebrates.

*S. pyogenes* is a facultative pathogenic bacteria mostly restricted to humans, with about 10-20% of the population being asymptomatic carriers (Shaikh *et al.* 2010; Roberts *et al.* 2012). It is thus no surprise that recent investigations have found anti-*Cas9* antibodies and *Cas9*-reactive T cells in several human populations (Charlesworth *et al.* 2019; Wagner *et al.* 2019). It is thus possible that certain humans are immune to gene drives, and future studies will undoubtedly shed light on this question. The presence of *S. pyogenes* has also been documented in a few animals such as macaques, mice, dogs, hedgehogs, rabbits and sheep [(Vela *et al.* 2017; Chen *et al.* 2019) and references therein]. Whether these vertebrates may also be immune to gene drives remains to be investigated.

So far, all published gene drive constructs have used *Cas9* from *S. pyogenes* (Sp*Cas9*). Other *Cas* proteins with activity similar to Sp*Cas9* are available (Zhang 2019) and could potentially be used for gene drive technology. As they are derived from other types of bacteria, their immunogenicity and their associated probability *immune* would have to be specifically assessed.

### **Probability of non-extinction of the drive (*nonextinct*)**

Once the drive successfully introduced into the genome, its fate will depend on its ability to distort its segregation, on chance and on the probability that the non-target population evolve resistance to the gene drive. For simplicity, we treat here the probability that a drive, initially present in a single individual or so, is not eliminated immediately by mere chance due to sampling effect and reach significant numbers of individuals in the non-target population. Even if an allele manages to be present in more than one half of the gametes from heterozygous individuals, it can still disappear rapidly due to random processes. If not, then it can invade the population. This process has been modelled as a branching process by Bienaymé in 1845 (Bienaymé 1845), actually published by Cournot (1847), and Watson and Galton (1875), see (Kendall 1975; Bru 1991). The probability of extinction, starting with one replicator, is the lowest root of equation  $G(x) = x$ , where  $G(x)$  is the generating function of the law describing the number of copies left per generation by one replicator. Here, it is relevant to assume the distribution to follow a Poisson law. If the population as a whole is stable, each “normal” gene leaves on average one copy of itself in the next generation, *i.e.* on average 1/2 through male gametes and 1/2 through female ones. The drive will then leave a number of copies of itself equal to the sum of the proportions which it represents in male ( $\lambda_m$ ) and in female ( $\lambda_f$ ) gametes:  $\lambda = \lambda_m + \lambda_f$ . Typically,  $\lambda_m$  and  $\lambda_f$  are around 0.7 – 0.9 for gene drives tested in mice and insects (Gantz and Bier 2015; Gantz *et al.* 2015; Hammond *et al.* 2016; KaramiNejadRanjbar *et al.* 2018; Kyrou *et al.* 2018; Grunwald *et al.* 2019) so that a reasonable estimate of  $\lambda$  is 1.4 – 1.8. According to the “Bienaymé, Galton, Watson” model, the extinction probability for a Poisson law of parameter

$\lambda$  is such that  $G(x) = e^{\lambda(x-1)} = x$ . With  $\lambda = 1.4$  (resp. 1.8), it gives  $x \approx 0.5$  (resp. 0.27) and the probability of invasion  $(1-x)$  is thus around 0.5-0.7. For the dynamics of gene drive alleles once they are introduced in a substantial amount in a targeted population, several stochastic and deterministic models have been composed, which take into account possible costs of harboring a gene drive as well as appearance of resistance to gene drive (Marshall 2009; Unckless *et al.* 2015).

Parameter	Rough Estimate
<i>transf</i>	0.035 transfer events or more to at least one species per thousand years for a given transposable element; may be lower for a gene drive construct
<i>hyb</i>	unclear, such events are possible in species closely related to the target species
<i>express</i>	close to 1 in vertebrate species for gene drives targeting mice, close to 1 in Diptera species for gene drives targeting mosquitoes or flies
<i>cut</i>	0.02 in a random genome of human size
<i>flank</i>	unclear
<i>immune</i>	close to 1 for non-target insect species, may be lower for non-target vertebrate species
<i>nonextinct</i>	0.5-0.7

**Table 2. Overview of the various parameters of the DRAQUE Equation.**

## Discussion

Based on our examination of the seven parameters of the DRAQUE Equation (Table 2), the probability that a gene drive transfers to another species can have values ranging from 0 to 0.5 per year in the worst-case scenario (one transfer event occurring per year, the guide RNA site is present in the non-target genome, the non-target species has high levels of homologous recombination). Our current estimate of the overall risk (Table 2) remains nevertheless very crude, and asks for further studies to refine this estimate.

We did not treat here the risk of contamination of another population, within the targeted species, but the same formula could be used in principle for this case. Furthermore, our equation does not take into account the phenotypic effect of the drive on the contaminated species. A drive may display no phenotypic effects in the non-target species that it invaded. If the gene drive was designed to eliminate a target population, then it is more likely to eliminate the non-target species. The drive may also have unexpected effects, for example mosaicism due to perdurance of Cas9 expression (Guichard *et al.* 2019).

To prevent contamination, several containment strategies for laboratory experiments have been proposed (Benedict *et al.* 2008; Akbari *et al.* 2015; National Academies of Sciences and Medicine 2016) as well as gene drives split in two different constructs (Benedict *et al.* 2008; DiCarlo *et al.* 2015) and synthetic target sites (Champer *et al.* 2019).

Based on our survey of the various parameters, we suggest further design strategies to minimize the risk of transfer to a non-target species: the addition of insulators (see above), and the choice of a guide RNA target site that is not close to the centromere, to avoid rearrangements and increased probability of creating an active gene drive in a non-target species.

Here we treated probabilities for a standard gene drive construct, but the risk should be estimated for each particular gene drive construct. New types of self-limiting gene drives have been proposed in recent years to try to limit the spread of gene drives spatially or temporally: toxin-antidote systems including Medea (Buchman *et al.* 2018), CleaveR (Oberhofer *et al.* 2019), Killer-Rescue (K-R) (Gould *et al.* 2008; Webster *et al.* 2019), one or two-locus underdominance (Dhole *et al.* 2018) and Daisy-Chain drives (Noble *et al.* 2019). The DRAQUE parameters would have to be evaluated specifically for these particular cases.

We have restricted our evaluation of the probability of accidental spread of gene drive constructs to genomes considered as fairly stable entities. However, genomes are dynamic structures. The process of gene duplication (or even n-plication) is commonplace (Clark and Donoghue 2018b; Harari *et al.* 2018; Moriyama and Koshiba-Takeuchi 2018). Local amplification of sequences is also very frequent (Liu *et al.* 2019; Traynor *et al.* 2019) and this may increase the probability of accidents to a further unknown level.

The DRAQUE Formula does not cover all the risks associated with gene drive. In addition to the risk of transfer to another species, gene drive designed to eliminate a target population may have additional ecological consequences that are not treated here.

Living with highly evolved technologies entails high risks for individuals and societies. Here we have attempted to evaluate circumstances where risks could be identified, but this assumes that we are aware of all the natural processes coupled with the technologies of interest. Besides the technology itself—which can be properly monitored and steered—there is an additional risk that is seldom taken into account, the risk derived from the way the organizations that implement the technologies manage them (Perrow 2011). We have not tackled this question here, but it is an obvious place of much concern. The way novel biological technologies have been used recently—see modification of human babies with deletion of a surface cell receptor (Sand *et al.* 2019)—should remind us that rogue or inconscient scientists may use gene drive approaches without proper risk assessment.

Overall, our study reveals that there is a need for more detailed investigations of the different factors influencing the probability of a gene drive contaminating another species before any release in the wild population is ever considered. We hope that our paper will trigger discussions and progress in the ethics of gene drive implementation.

## Acknowledgements

We deeply thank Florence Débarre, Clément Gilbert, Arnaud Martin and Nicolas Rode for their comments on a previous draft, and Christophe Antoniewski, Sylvain Charlat, Jean David, Scott Gilbert, Frank Jiggins, Bruno Lemaitre, Olivier Panaud, Thomas Pradeu, Benjamin Prud'homme, Michael Turelli and Mylène Weill for interesting discussions and useful information.

## References

- Agren, J. A., and A. G. Clark, 2018 Selfish genetic elements. *PLoS Genet.* 14: e1007700.
- Akbari, O. S., H. J. Bellen, E. Bier, S. L. Bullock, A. Burt *et al.*, 2015 Safeguarding gene drive experiments in the laboratory. *Science* 349: 927–929.
- Auer, T. O., and F. Del Bene, 2014 CRISPR/Cas9 and TALEN-mediated knock-in approaches in zebrafish. *Methods* 69: 142–150.



- Bartolomé, C., X. Bello, and X. Maside, 2009 Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes. *Genome Biol.* 10: R22.
- Bassett, A. R., and J.-L. Liu, 2014 CRISPR/Cas9 and genome editing in *Drosophila*. *J. Genet. Genomics* 41: 7–19.
- Benedict, M., P. D'Abbs, S. Dobson, M. Gottlieb, L. Harrington *et al.*, 2008 Guidance for contained field trials of vector mosquitoes engineered to contain a gene drive system: recommendations of a scientific working group. *Vector-Borne Zoonotic Dis.* 8: 127–166.
- Bienaymé, I.-J., 1845 De la loi de multiplication et de la durée des familles. *Soc Philomat Paris Extr. Sér* 5: 37–39.
- Booker, B. M., T. Friedrich, M. K. Mason, J. E. VanderMeer, J. Zhao *et al.*, 2016 Bat accelerated regions identify a bat forelimb specific enhancer in the *HoxD* locus. *PLoS Genet.* 12: e1005738.
- Breitwieser, F. P., M. Perteua, A. V. Zimin, and S. L. Salzberg, 2019 Human contamination in bacterial genomes has created thousands of spurious proteins. *Genome Res.* 29: 954–960.
- Browning, D., and G. Trobridge, 2016 Insulators to improve the safety of retroviral vectors for HIV gene therapy. *Biomedicine* 4: 4.
- Bru, B., 1991 A la recherche de la démonstration perdue de Bienaymé. *Mathématiques Sci. Hum.* 114: 5–17.
- Buchman, A., J. M. Marshall, D. Ostrovski, T. Yang, and O. S. Akbari, 2018 Synthetically engineered *Medea* gene drive system in the worldwide crop pest *Drosophila suzukii*. *Proc. Natl. Acad. Sci.* 115: 4725–4730.
- Burchell, M. J., 2006 W (h)ither the Drake equation? *Int. J. Astrobiol.* 5: 243–250.
- Cande, J., Y. Goltsev, and M. S. Levine, 2009 Conservation of enhancer location in divergent insects. *Proc. Natl. Acad. Sci.* 106: 14414–14419.
- Champer, J., J. Chung, Y. L. Lee, C. Liu, E. Yang *et al.*, 2019 Molecular safeguarding of CRISPR gene drive experiments. *Elife* 8: e41439.
- Champer, J., R. Reeves, S. Y. Oh, C. Liu, J. Liu *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 Genet.* 13: e1006796.
- Charlesworth, C. T., P. S. Deshpande, D. P. Dever, J. Camarena, V. T. Lemgart *et al.*, 2019 Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nat. Med.* 25: 249.
- Chen, D.-S., Y.-Q. Wu, W. Zhang, S.-J. Jiang, and S.-Z. Chen, 2016 Horizontal gene transfer events reshape the global landscape of arm race between viruses and homo sapiens. *Sci. Rep.* 6: 26934.
- Chen, Y., P. Xiao, R. Bai, T. Zhang, Y. Zhou *et al.*, 2019 Extensive adaptive immune response of AAVs and Cas proteins in non-human primates. *BioRxiv* 588913.
- Chen, L.-L., and L. Yang, 2017 ALU alternative regulation for gene expression. *Trends Cell Biol.* 27: 480–490.
- Cheng, Y., Z. Ma, B.-H. Kim, W. Wu, P. Cayting *et al.*, 2014 Principles of regulatory information conservation between mouse and human. *Nature* 515: 371.
- Chew, W. L., 2018 Immunity to CRISPR Cas9 and Cas12a therapeutics. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 10: e1408.
- Chew, W. L., M. Tabebordbar, J. K. Cheng, P. Mali, E. Y. Wu *et al.*, 2016 A multifunctional AAV–CRISPR–Cas9 and its host response. *Nat. Methods* 13: 868.
- Clark, J. W., and P. C. Donoghue, 2018a Whole-genome duplication and plant macroevolution. *Trends Plant Sci.*
- Clark, J. W., and P. C. Donoghue, 2018b Whole-genome duplication and plant macroevolution. *Trends Plant Sci.*

- Clark, J. B., and M. G. Kidwell, 1997 A phylogenetic perspective on P transposable element evolution in *Drosophila*. *Proc. Natl. Acad. Sci.* 94: 11428–11433.
- Coetzee, M., R. H. Hunt, R. Wilkerson, A. Della Torre, M. B. Coulibaly *et al.*, 2013 *Anopheles coluzzii* and *Anopheles amharicus*, new members of the *Anopheles gambiae* complex. *Zootaxa* 3619: 246–274.
- Cournot, A. A., 1847 *De l'origine et des limites de la correspondance entre l'algèbre et la géométrie*. L. Hachette.
- Cretekos, C. J., Y. Wang, E. D. Green, J. F. Martin, J. J. Rasweiler *et al.*, 2008 Regulatory divergence modifies limb length between mammals. *Genes Dev.* 22: 141–151.
- Cui, J., G. Tachedjian, and L.-F. Wang, 2015 Bats and rodents shape mammalian retroviral phylogeny. *Sci. Rep.* 5: 16561.
- Daniels, S. B., K. R. Peterson, L. D. Strausbaugh, M. G. Kidwell, and A. Chovnick, 1990 Evidence for horizontal transmission of the P transposable element between *Drosophila* species. *Genetics* 124: 339–355.
- Deredec, A., A. Burt, and H. C. J. Godfray, 2008 The population genetics of using homing endonuclease genes in vector and pest management. *Genetics* 179: 2013–2026.
- Dhole, S., M. R. Vella, A. L. Lloyd, and F. Gould, 2018 Invasion and migration of spatially self-limiting gene drives: A comparative analysis. *Evol. Appl.* 11: 794–808.
- DiCarlo, J. E., A. Chavez, S. L. Dietz, K. M. Esvelt, and G. M. Church, 2015 Safeguarding CRISPR-Cas9 gene drives in yeast. *Nat. Biotechnol.* 33: 1250.
- Dover, G., 1982 Molecular drive: a cohesive mode of species evolution. *Nature* 299: 111.
- Dunning, L. T., J. K. Olofsson, C. Parisod, R. R. Choudhury, J. J. Moreno-Villena *et al.*, 2019 Lateral transfers of large DNA fragments spread functional genes among grasses. *Proc. Natl. Acad. Sci.* 116: 4416–4425.
- El Baidouri, M., M.-C. Carpentier, R. Cooke, D. Gao, E. Lasserre *et al.*, 2014 Widespread and frequent horizontal transfers of transposable elements in plants. *Genome Res.* 24: 831–838.
- Elbashir, S. M., W. Lendeckel, and T. Tuschl, 2001 RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* 15: 188–200.
- Escalera-Fanjul, X., H. Quezada, L. Riego-Ruiz, and A. González, 2018 Whole-Genome duplication and yeast's fruitful way of life. *Trends Genet.*
- Esvelt, K. M., A. L. Smidler, F. Catteruccia, and G. M. Church, 2014 Emerging technology: concerning RNA-guided gene drives for the alteration of wild populations. *Elife* 3: e03401.
- Fertil, B., M. Massin, S. Lespinats, C. Devic, P. Dumeé *et al.*, 2005 GENSTYLE: exploration and analysis of DNA sequences with genomic signature. *Nucleic Acids Res.* 33: W512–W515.
- Feschotte, C., and C. Gilbert, 2012 Endogenous viruses: insights into viral evolution and impact on host biology. *Nat. Rev. Genet.* 13: 283.
- Filippini, M., and M. Middelboe, 2007 Viral abundance and genome size distribution in the sediment and water column of marine and freshwater ecosystems. *FEMS Microbiol. Ecol.* 60: 397–410.
- Fontaine, M. C., J. B. Pease, A. Steele, R. M. Waterhouse, D. E. Neafsey *et al.*, 2015 Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* 347: 1258524.
- Fu, Y., J. A. Foden, C. Khayter, M. L. Maeder, D. Reyon *et al.*, 2013 High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 31: 822.
- Gammon, D. B., and C. C. Mello, 2015 RNA interference-mediated antiviral defense in insects. *Curr. Opin. Insect Sci.* 8: 111–120.
- Gantz, V. M., and E. Bier, 2015 The mutagenic chain reaction: a method for converting

- heterozygous to homozygous mutations. *Science* 348: 442–444.
- Gantz, V. M., N. Jasinskiene, O. Tatarenkova, A. Fazekas, V. M. Macias *et al.*, 2015 Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*. *Proc. Natl. Acad. Sci.* 112: E6736–E6743.
- Gao, Y., G. Chuai, W. Yu, S. Qu, and Q. Liu, 2019 Data imbalance in CRISPR off-target prediction. *Brief. Bioinform.*
- Gasmi, L., H. Boulain, J. Gauthier, A. Hua-Van, K. Musset *et al.*, 2015 Recurrent domestication by Lepidoptera of genes from their parasites mediated by bracoviruses. *PLoS Genet.* 11: e1005470.
- Gilbert, C., A. Chateigner, L. Ernenwein, V. Barbe, A. Bézier *et al.*, 2014 Population genomics supports baculoviruses as vectors of horizontal transfer of insect transposons. *Nat. Commun.* 5: 3348.
- Gilbert, C., and R. Cordaux, 2017 Viruses as vectors of horizontal transfer of genetic material in eukaryotes. *Curr. Opin. Virol.* 25: 16–22.
- Gilbert, C., and C. Feschotte, 2018 Horizontal acquisition of transposable elements and viral sequences: patterns and consequences. *Curr. Opin. Genet. Dev.* 49: 15–24.
- Gilbert, C., J. Peccoud, A. Chateigner, B. Moumen, R. Cordaux *et al.*, 2016 Continuous influx of genetic material from host to virus populations. *PLoS Genet.* 12: e1005838.
- Gilbert, C., S. Schaack, J. K. Pace II, P. J. Brindley, and C. Feschotte, 2010 A role for host–parasite interactions in the horizontal transfer of transposons across phyla. *Nature* 464: 1347.
- Gonzalez-Estevez, C., T. Momose, W. J. Gehring, and E. Salo, 2003 Transgenic planarian lines obtained by electroporation using transposon-derived vectors and an eye-specific GFP marker. *Proc. Natl. Acad. Sci.* 100: 14046–14051.
- Gould, F., Y. Huang, M. Legros, and A. L. Lloyd, 2008 A Killer–Rescue system for self-limiting gene drive of anti-pathogen constructs. *Proc. R. Soc. B Biol. Sci.* 275: 2823–2829.
- Greenwood, A. D., Y. Ishida, S. P. O’Brien, A. L. Roca, and M. V. Eiden, 2018 Transmission, evolution, and endogenization: lessons learned from recent retroviral invasions. *Microbiol Mol Biol Rev* 82: e00044–17.
- Grover, D., M. Mukerji, P. Bhatnagar, K. Kannan, and S. K. Brahmachari, 2004 Alu repeat analysis in the complete human genome: trends and variations with respect to genomic composition. *Bioinformatics* 20: 813–817.
- Grunwald, H. A., V. M. Gantz, G. Poplawski, X.-R. S. Xu, E. Bier *et al.*, 2019 Super-Mendelian inheritance mediated by CRISPR–Cas9 in the female mouse germline. *Nature* 566: 105.
- Guichard, A., T. Haque, M. Bobik, X.-R. S. Xu, C. Klanseck *et al.*, 2019 Efficient allelic-drive in *Drosophila*. *Nat. Commun.* 10: 1640.
- Hammond, A., R. Galizi, K. Kyrou, A. Simoni, C. Siniscalchi *et al.*, 2016 A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. *Nat. Biotechnol.* 34: 78.
- Harari, Y., Y. Ram, and M. Kupiec, 2018 Frequent ploidy changes in growing yeast cultures. *Curr. Genet.* 64: 1001–1004.
- Hartlerode, A. J., and R. Scully, 2009 Mechanisms of double-strand break repair in somatic mammalian cells. *Biochem. J.* 423: 157–168.
- Hayward, A., C. K. Cornwallis, and P. Jern, 2015 Pan-vertebrate comparative genomics unmask retrovirus macroevolution. *Proc. Natl. Acad. Sci.* 112: 464–469.
- Hermansen, R. A., T. R. Hvidsten, S. R. Sandve, and D. A. Liberles, 2016 Extracting functional trends from whole genome duplication events using comparative genomics. *Biol. Proced. Online* 18: 11.
- Hill, T., C. Schlötterer, and A. J. Betancourt, 2016 Hybrid dysgenesis in *Drosophila simulans*

- associated with a rapid invasion of the P-element. *PLoS Genet.* 12: e1005920.
- Hisano, Y., T. Sakuma, S. Nakade, R. Ohga, S. Ota *et al.*, 2015 Precise in-frame integration of exogenous DNA mediated by CRISPR/Cas9 system in zebrafish. *Sci. Rep.* 5: 8841.
- Holmes, E. C., 2011 What does virus evolution tell us about virus origins? *J. Virol.* 85: 5247–5251.
- Horváth, V., M. Merenciano, and J. González, 2017 Revisiting the relationship between transposable elements and the eukaryotic stress response. *Trends Genet.* 33: 832–841.
- Houck, M. A., J. B. Clark, K. R. Peterson, and M. G. Kidwell, 1991 Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis*. *Science* 253: 1125–1128.
- Hsu, P. D., D. A. Scott, J. A. Weinstein, F. A. Ran, S. Konermann *et al.*, 2013 DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31: 827.
- Hunter, M. E., J. A. Ferrante, G. Meigs-Friend, and A. Ulmer, 2019 Improving eDNA yield and inhibitor reduction through increased water volumes and multi-filter isolation techniques. *Sci. Rep.* 9: 5259.
- Imamura, M., J. Nakai, S. Inoue, G. X. Quan, T. Kanda *et al.*, 2003 Targeted gene expression using the GAL4/UAS system in the silkworm *Bombyx mori*. *Genetics* 165: 1329–1340.
- Inoue, Y., T. Saga, T. Aikawa, M. Kumagai, A. Shimada *et al.*, 2017 Complete fusion of a transposon and herpesvirus created the Teratorn mobile element in medaka fish. *Nat. Commun.* 8: 551.
- Ivancevic, A., D. Kortschak, T. Bertozzi, and D. Adelson, 2017 Re-evaluating inheritance in genome evolution: widespread transfer of LINEs between species. *bioRxiv* 106914.
- Jasin, M., 1996 Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet.* 12: 224–228.
- KaramiNejadRanjbar, M., K. N. Eckermann, H. M. Ahmed, S. Dippel, J. M. Marshall *et al.*, 2018 Consequences of resistance evolution in a Cas9-based sex conversion-suppression gene drive for insect pest management. *Proc. Natl. Acad. Sci.* 115: 6189–6194.
- Katzourakis, A., and R. J. Gifford, 2010 Endogenous viral elements in animal genomes. *PLoS Genet.* 6: e1001191.
- Kendall, D. G., 1975 The genealogy of genealogy branching processes before (and after) 1873. *Bull. Lond. Math. Soc.* 7: 225–253.
- Kim, S., D. Kim, S. W. Cho, J. Kim, and J.-S. Kim, 2014 Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* 24: 1012–1019.
- Kim, S., T. Koo, H.-G. Jee, H.-Y. Cho, G. Lee *et al.*, 2018 CRISPR RNAs trigger innate immune responses in human cells. *Genome Res.* 28: 367–373.
- Kumar, S., G. Stecher, M. Suleski, and S. B. Hedges, 2017 TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* 34: 1812–1819.
- Kvon, E. Z., O. K. Kamneva, U. S. Melo, I. Barozzi, M. Osterwalder *et al.*, 2016 Progressive loss of function in a limb enhancer during snake evolution. *Cell* 167: 633–642.
- Kyrou, K., A. M. Hammond, R. Galizi, N. Kranjc, A. Burt *et al.*, 2018 A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged *Anopheles gambiae* mosquitoes. *Nat. Biotechnol.* 36: 1062.
- Lai, Y.-T., K. D. Deem, F. Borràs-Castells, N. Sambrani, H. Rudolf *et al.*, 2018 Enhancer identification and activity evaluation in the red flour beetle, *Tribolium castaneum*. *Development* 145: dev160663.
- Larsen, P. A., K. E. Hunnicutt, R. J. Larsen, A. D. Yoder, and A. M. Saunders, 2018 Warning

- SINEs: Alu elements, evolution of the human brain, and the spectrum of neurological disease. *Chromosome Res.* 26: 93–111.
- Lavi, E., and L. Carmel, 2018 Alu exaptation enriches the human transcriptome by introducing new gene ends. *RNA Biol.* 15: 715–725.
- Leitschuh, C. M., D. Kanavy, G. A. Backus, R. X. Valdez, M. Serr *et al.*, 2018 Developing gene drive technologies to eradicate invasive rodents from islands. *J. Responsible Innov.* 5: S121–S138.
- Lemaitre, B., and J. Hoffmann, 2007 The host defense of *Drosophila melanogaster*. *Annu Rev Immunol* 25: 697–743.
- Li, K., G. Wang, T. Andersen, P. Zhou, and W. T. Pu, 2014 Optimization of genome engineering approaches with the CRISPR/Cas9 system. *PloS One* 9: e105779.
- Li, N., T. Wang, and D. Han, 2012 Structural, cellular and molecular aspects of immune privilege in the testis. *Front. Immunol.* 3: 152.
- Liu, Q., X. Li, X. Zhou, M. Li, F. Zhang *et al.*, 2019 The repetitive DNA landscape in *Avena* (Poaceae): chromosome and genome evolution defined by major repeat classes in whole-genome sequence reads. *BMC Plant Biol.* 19: 226.
- Magaziner, S. J., Z. Zeng, B. Chen, and G. P. Salmond, 2019 The prophages of *Citrobacter rodentium* represent a conserved family of horizontally acquired mobile genetic elements associated with enteric evolution towards pathogenicity. *J. Bacteriol.* 201: e00638–18.
- Mamcarz, E., S. Zhou, T. Lockey, H. Abdelsamed, S. J. Cross *et al.*, 2019 Lentiviral gene therapy combined with low-dose busulfan in infants with SCID-X1. *N. Engl. J. Med.* 380: 1525–1534.
- Mantilla-Calderon, D., M. J. Plewa, G. Michoud, S. Fodelianakis, D. Daffonchio *et al.*, 2019 Water disinfection byproducts increase natural transformation rates of environmental DNA in *Acinetobacter baylyi* ADP1. *Environ. Sci. Technol.*
- Marshall, J. M., 2009 The effect of gene drive on containment of transgenic mosquitoes. *J. Theor. Biol.* 258: 250–265.
- Mellor, A. L., and D. H. Munn, 2008 Creating immune privilege: active local suppression that benefits friends, but protects foes. *Nat. Rev. Immunol.* 8: 74.
- Monlong, J., P. Cossette, C. Meloche, G. Rouleau, S. L. Girard *et al.*, 2018 Human copy number variants are enriched in regions of low mappability. *Nucleic Acids Res.* 46: 7236–7249.
- Morissette, G., and L. Flamand, 2010 Herpesviruses and chromosomal integration. *J. Virol.* 84: 12100–12109.
- Moriyama, Y., and K. Koshiba-Takeuchi, 2018 Significance of whole-genome duplications on the emergence of evolutionary novelties. *Brief. Funct. Genomics* 17: 329–338.
- Nakade, S., T. Tsubota, Y. Sakane, S. Kume, N. Sakamoto *et al.*, 2014 Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nat. Commun.* 5: 5560.
- National Academies of Sciences, and Medicine, 2016 *Gene drives on the horizon: advancing science, navigating uncertainty, and aligning research with public values*. National Academies Press.
- Noble, C., J. Min, J. Olejarz, J. Buchthal, A. Chavez *et al.*, 2019 Daisy-chain gene drives for the alteration of local populations. *Proc. Natl. Acad. Sci.* 116: 8275–8282.
- Oberhofer, G., T. Ivy, and B. A. Hay, 2019 Cleave and Rescue, a novel selfish genetic element and general strategy for gene drive. *Proc. Natl. Acad. Sci.* 116: 6250–6259.
- Oliveira, J., and W. C. Reygaert, 2019 Gram Negative Bacteria, in *StatPearls [Internet]*, StatPearls Publishing.
- Paix, A., Y. Wang, H. E. Smith, C.-Y. S. Lee, D. Calidas *et al.*, 2014 Scalable and versatile genome editing using linear DNAs with microhomology to Cas9 Sites in

- Caenorhabditis elegans. *Genetics* 198: 1347–1356.
- Palazzo, A., P. Lorusso, C. Miskey, O. Walisko, A. Gerbino *et al.*, 2019 Transcriptionally promiscuous “blurry” promoters in Tc1/mariner transposons allow transcription in distantly related genomes. *Mob. DNA* 10: 13.
- Pan, F., V. H. Man, C. Roland, and C. Sagui, 2018 Structure and Dynamics of DNA and RNA Double Helices Obtained from the CCG and GGC Trinucleotide Repeats. *J. Phys. Chem. B* 122: 4491–4512.
- Parisot, N., A. Pelin, C. Gasc, V. Polonais, A. Belkorchia *et al.*, 2014 Microsporidian genomes harbor a diverse array of transposable elements that demonstrate an ancestry of horizontal exchange with metazoans. *Genome Biol. Evol.* 6: 2289–2300.
- Pavlopoulos, A., and M. Averof, 2005 Establishing genetic transformation for comparative developmental studies in the crustacean *Parhyale hawaiiensis*. *Proc. Natl. Acad. Sci.* 102: 7888–7893.
- Peccoud, J., V. Loiseau, R. Cordaux, and C. Gilbert, 2017 Massive horizontal transfer of transposable elements in insects. *Proc. Natl. Acad. Sci.* 114: 4721–4726.
- Perrow, C., 2011 *Normal accidents: Living with high risk technologies-Updated edition*. Princeton university press.
- Rabeh, K., F. Gaboun, B. Belkadi, and A. Filali-Maltouf, 2018 In Silico development of new SSRs primer for aquaporin linked to drought tolerance in plants. *Plant Signal. Behav.* 13: e1536630.
- Ramos, D. M., F. Kamal, E. A. Wimmer, A. N. Cartwright, and A. Monteiro, 2006 Temporal and spatial control of transgene expression using laser induction of the hsp70 promoter. *BMC Dev. Biol.* 6: 55.
- Reiss, D., G. Mialdea, V. Miele, D. M. de Vienne, J. Peccoud *et al.*, 2019 Global survey of mobile DNA horizontal transfer in arthropods reveals Lepidoptera as a prime hotspot. *PLoS Genet.* 15: e1007965.
- Roberts, A. L., K. L. Connolly, D. J. Kirse, A. K. Evans, K. A. Poehling *et al.*, 2012 Detection of group A Streptococcus in tonsils from pediatric patients reveals high rate of asymptomatic streptococcal carriage. *BMC Pediatr.* 12: 3.
- Rode, N. O., A. Estoup, D. Bourguet, V. Courtier-Orgogozo, and F. Débarre, 2019 Population management using gene drive: molecular design, models of spread dynamics and assessment of ecological risks. *Conserv. Genet.* 1–20.
- Roggenkamp, E., R. M. Giersch, M. N. Schrock, E. Turnquist, M. Halloran *et al.*, 2018 Tuning CRISPR-Cas9 gene drives in *Saccharomyces cerevisiae*. *G3 Genes Genomes Genet.* 8: 999–1018.
- Sand, M., A. L. Bredenoord, and K. R. Jongsma, 2019 After the fact—the case of CRISPR babies. *Eur. J. Hum. Genet.* 1.
- Satović, E., T. V. Zeljko, A. Luchetti, B. Mantovani, and M. Plohl, 2016 Adjacent sequences disclose potential for intra-genomic dispersal of satellite DNA repeats and suggest a complex network with transposable elements. *BMC Genomics* 17: 997.
- Schetelig, M. F., and E. A. Wimmer, 2011 Insect transgenesis and the sterile insect technique, pp. 169–194 in *Insect biotechnology*, Springer.
- Schinko, J. B., M. Weber, I. Viktorinova, A. Kiupakis, M. Averof *et al.*, 2010 Functionality of the GAL4/UAS system in *Tribolium* requires the use of endogenous core promoters. *BMC Dev. Biol.* 10: 53.
- Scott, M. J., F. Gould, M. Lorenzen, N. Grubbs, O. Edwards *et al.*, 2018 Agricultural production: assessment of the potential use of Cas9-mediated gene drive systems for agricultural pest control. *J. Responsible Innov.* 5: S98–S120.
- Serrato-Capuchina, A. G., S. Zhang, W. Martin, D. Peede, E. Earley *et al.*, 2018 Recent invasion of P transposable element into *Drosophila yakuba*. bioRxiv 453829.
- Shaikh, N., E. Leonard, and J. M. Martin, 2010 Prevalence of streptococcal pharyngitis and

- streptococcal carriage in children: a meta-analysis. *Pediatrics* 126: e557–e564.
- Shapiro, R. S., A. Chavez, C. B. Porter, M. Hamblin, C. S. Kaas *et al.*, 2018 A CRISPR–Cas9-based gene drive platform for genetic interaction analysis in *Candida albicans*. *Nat. Microbiol.* 3: 73.
- Simpson, E., 2006 A historical perspective on immunological privilege. *Immunol. Rev.* 213: 12–22.
- Smalec, B. M., T. N. Heider, B. L. Flynn, and R. J. O'Neill, 2019 A centromere satellite concomitant with extensive karyotypic diversity across the *Peromyscus* genus defies predictions of molecular drive. *Chromosome Res.* 1–16.
- Song, X., C. R. Beck, R. Du, I. M. Campbell, Z. Coban-Akdemir *et al.*, 2018 Predicting human genes susceptible to genomic instability associated with Alu/Alu-mediated rearrangements. *Genome Res.* 28: 1228–1242.
- Tamura, T., C. Thibert, C. Royer, T. Kanda, A. Eappen *et al.*, 2000 Germline transformation of the silkworm *Bombyx mori* L. using a piggyBac transposon-derived vector. *Nat. Biotechnol.* 18: 81.
- Traynor, S., N. E. Møllegaard, M. G. Jørgensen, N. H. Brückmann, C. B. Pedersen *et al.*, 2019 Remodeling and destabilization of chromosome 1 pericentromeric heterochromatin by SSX proteins. *Nucleic Acids Res.*
- Turelli, M., J. R. Lipkowitz, and Y. Brandvain, 2014 On the Coyne and Orr-igin of species: Effects of intrinsic postzygotic isolation, ecological differentiation, X chromosome size, and sympatry on *Drosophila* speciation. *Evolution* 68: 1176–1187.
- Unckless, R. L., P. W. Messer, T. Connallon, and A. G. Clark, 2015 Modeling the manipulation of natural populations by the mutagenic chain reaction. *Genetics* 201: 425–431.
- Vela, A. I., P. Villalón, J. A. Sáez-Nieto, G. Chacón, L. Domínguez *et al.*, 2017 Characterization of *Streptococcus pyogenes* from Animal Clinical Specimens, Spain. *Emerg. Infect. Dis.* 23: 2011.
- Wagner, D. L., L. Amini, D. J. Wendering, L.-M. Burkhardt, L. Akyüz *et al.*, 2019 High prevalence of *Streptococcus pyogenes* Cas9-reactive T cells within the adult human population. *Nat. Med.* 25: 242.
- Wang, R. W., C.-S. Lee, and J. E. Haber, 2017 Position effects influencing intrachromosomal repair of a double-strand break in budding yeast. *PLOS ONE* 12: e0180994.
- Watson, H. W., and F. Galton, 1875 On the probability of the extinction of families. *J. Anthropol. Inst. G. B. Irel.* 4: 138–144.
- Webber, B. L., S. Raghu, and O. R. Edwards, 2015 Opinion: Is CRISPR-based gene drive a biocontrol silver bullet or global conservation threat? *Proc. Natl. Acad. Sci.* 112: 10565–10567.
- Webster, S. H., M. R. Vella, and M. J. Scott, 2019 Development and testing of a novel Killer-Rescue self-limiting gene drive system in *Drosophila melanogaster*. *bioRxiv* 680629.
- Wienert, B., J. Shin, E. Zelin, K. Pestal, and J. E. Corn, 2018 In vitro–transcribed guide RNAs trigger an innate immune response via the RIG-I pathway. *PLoS Biol.* 16: e2005840.
- Wierson, W. A., J. M. Welker, M. P. Almeida, C. M. Mann, D. A. Webster *et al.*, 2018 GeneWeld: a method for efficient targeted integration directed by short homology. *bioRxiv* 431627.
- Wittkopp, P. J., and G. Kalay, 2012 Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* 13: 59.
- Xu, W., and M. V. Eiden, 2015 Koala retroviruses: evolution and disease dynamics. *Annu. Rev. Virol.* 2: 119–134.

- Yuan, X., M. Song, P. Devine, B. G. Bruneau, I. C. Scott *et al.*, 2018 Heart enhancers with deeply conserved regulatory activity are established early in zebrafish development. *Nat. Commun.* 9: 4977.
- Zhang, F., 2019 Development of CRISPR-Cas systems for genome editing and beyond. *Q. Rev. Biophys.* 52:.
- Zhang, X.-H., L. Y. Tee, X.-G. Wang, Q.-S. Huang, and S.-H. Yang, 2015 Off-target effects in CRISPR/Cas9-mediated genome engineering. *Mol. Ther.-Nucleic Acids* 4: e264.



Table S1. Results of BLAST searches for target sequences of two gene drive constructs published previously (Kyrou *et al.* 2018; Grunwald *et al.* 2019). The BLAST search was performed using 20-nucleotide-long queries on the nr database with default parameters on [https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE\\_TYPE=BlastSearch&LINK\\_LOC=blasthome](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome). Note that the BLAST algorithm does not retrieve results when the query sequence contains an N at the antepenultimate position. In consequence, the “N” in “NGG” was replaced by each of the four nucleotide letters and four BLAST searches were performed for each target sequence. We examined only the first 100 hit sequences and selected here the ones with at least 16/20 nucleotide sites identical to the query and harbouring the protospacer adjacent motif (PAM) NGG. Hit sequences are shown by taxonomic group for each query sequence. Note that sequences are not always deposited in the nr database (for example the *Anopheles* sequences were not retrieved by our BLAST searches), so that this table does not reflect the full extent of all published sequence data. Note also that the target sequence for the tyrosinase locus is incorrectly labelled in Extended Data Figure 2 of Grunwald *et al.* 2019 (H. Grunwald, personal communication).

Target: <i>doublesex</i> in 6 <i>Anopheles</i> species			
TAACACAGGTCAAGCGGT <b>TGG</b>	TAACACAGGTCAAGCGG <b>AGG</b>	TAACACAGGTCAAGCGG <b>CGG</b>	TAACACAGGTCAAGCGG <b>GGG</b>
<b>Insects</b>	<b>Insects</b>	<b>Insects</b>	<b>Insects</b>
<i>Anopheles stephensi</i>	<i>Leptidea sinapis</i>		<i>Spodoptera litura</i>
			<i>Habropoda laboriosa</i>
<b>Fungi</b>	<b>Fungi</b>	<b>Fungi</b>	<b>Fungi</b>
<i>Aspergillus sojae</i>		<i>Uncinocarpus reesii</i>	
<i>Aspergillus nomius</i>			
<i>Aspergillus flavus</i>			
<i>Aspergillus oryzae</i>			
<i>Podospora comata</i>			
<i>Podospora anserina</i>			
<i>Fonsecaea monophora</i>			
<i>Fonsecaea nubica</i>			
<b>Apicomplexans</b>	<b>Apicomplexans</b>	<b>Apicomplexans</b>	<b>Apicomplexans</b>
	<i>Plasmodium cynomolgi</i>		<i>Plasmodium coatneyi</i>
	<i>Plasmodium vivax</i>		
<b>Bacteria</b>	<b>Bacteria</b>	<b>Bacteria</b>	<b>Bacteria</b>
<i>Aerococcus urinaehominis</i>	<i>Desulfitobacterium hafniense</i>	<i>Pseudarthrobacter sp.</i>	<i>Enterococcus gallinarum</i>
<i>Plantomycetes sp. varia</i>	<i>Vampirococcus sp.</i>	<i>Altererythrobacter sp.</i>	<i>Enterococcus sp.</i>
<i>Marinobacter sp.</i>	<i>Bacillus licheniformis</i>	<i>Streptomyces sp.</i>	<i>Serratia sp.</i>
<i>Thermaerobacter marianensis</i>		<i>Nitrospira sp.</i>	<i>Enterococcus casseliflavus</i>
		<i>Alistipes sp.</i>	<i>Streptomyces venezuelae</i>
		<i>Sporosarcina sp.</i>	<i>Desulfovibrio vulgaris</i>
		<i>Sphingomonas paucimobilis</i>	<i>Micromonospora auratinigra</i>
		<i>Sodalis glossinidius</i> (insect symbiont)	<i>Streptomyces venezuelae</i>
<b>Birds</b>	<b>Birds</b>	<b>Birds</b>	<b>Birds</b>
<i>Anser cygnoides domesticus</i>			

<i>Streptopelia turtur</i>			
<b>Mammals</b>	<b>Mammals</b>	<b>Mammals</b>	<b>Mammals</b>
<i>Homo sapiens</i>		<i>Peromyscus leucopus</i>	<i>Homo sapiens</i>
		<i>Rhinopithecus bieti</i>	<i>Pan troglodytes</i>
		<i>Rhinopithecus roxellana</i>	<i>Pongo abelii</i>
			<i>Canis lupus (dingo and familiaris)</i>
<b>Bony fishes</b>	<b>Bony fishes</b>	<b>Bony fishes</b>	<b>Bony fishes</b>
<i>Chanos chanos</i>	<i>Chanos chanos</i>	<i>Myripristis murdjan</i>	<i>Takifugu rubripes</i>
<i>Salarias fasciatus</i>	<i>Salmo trutta</i>	<i>Denticeps clupeoides</i>	<i>Amphiprion ocellaris</i>
<i>Parambassis ranga</i>	<i>Anabas testudineus</i>		<i>Acanthochromis polyacanthus</i>
<i>Betta splendens</i>	<i>Astatotilapia calliptera</i>		<i>Haplochromis burtoni</i>
	<i>Oreochromis niloticus</i>		<i>Oreochromis niloticus</i>
	<i>Scophthalmus maximus</i>		<i>Stegastes partitus</i>
	<i>Maylandia zebra</i>		<i>Maylandia zebra</i>
			<i>Neolamprologus brichardi</i>
			<i>Pundamilia nyererei</i>
			<i>Gasterosteus aculeatus</i>
			<i>Scleropages formosus</i>
			<i>Lateolabrax maculatus</i>
<b>Lancelets</b>	<b>Collembolas</b>	<b>Molluscs</b>	<b>Lancelets</b>
<i>Branchiostoma floridae</i>	<i>Folsomia candida</i>	<i>Crassostrea gigas</i>	<i>Branchiostoma floridae</i>
<b>Plants</b>	<b>Plants</b>	<b>Plants</b>	<b>Plants</b>
<i>Prunus dulcis</i>	<i>Ipomoea trifida</i>	<i>Oryza sativa</i>	
<i>Citrus sinensis</i>		<i>Solanum lycopersicum</i>	

Target: tyrosinase in <i>Mus musculus</i>			
ATGGCCGATAGGTGCAT <b>AGG</b>	ATGGCCGATAGGTGCAT <b>CGG</b>	ATGGCCGATAGGTGCAT <b>GGG</b>	ATGGCCGATAGGTGCAT <b>TGG</b>
<b>Mammals</b>	<b>Mammals</b>	<b>Mammals</b>	<b>Mammals</b>
		<i>Pan troglodytes</i>	<i>Cricetulus griseus</i>
			<i>Grammomys surdaster</i>
			<i>Homo sapiens</i>
	<b>Amphibians</b>	<b>Birds</b>	<i>Mus caroli</i>
	<i>Rhinatrema bivittatum</i>	<i>Streptopelia turtur</i>	<i>Mus musculus</i>
			<i>Mus pahari</i>
			<i>Peromyscus leucopus</i>
<b>Bony fishes</b>	<b>Bony fishes</b>	<b>Bony fishes</b>	
<i>Gadus morhua</i>	<i>Sparus aurata</i>	<i>Anabas testudineus</i>	
<i>Betta splendens</i>	<i>Sphaeramia orbicularis</i>	<i>Cottoperca gobio</i>	
		<i>Gadus morhua</i>	
		<i>Parambassis ranga</i>	
<b>Insects</b>	<b>Insects</b>	<b>Insects</b>	<b>Insects</b>
		<i>Drosophila elegans</i>	<i>Diabrotica virgifera</i>

			<b>Collembola</b>
			<i>Folsomia candida</i>
<b>Fungi</b>	<b>Fungi</b>	<b>Fungi</b>	<b>Fungi</b>
			<i>Cercospora sojina</i>
			<i>Vonarxia vagans</i>
<b>Apicomplexans</b>	<b>Euglenozoans</b>	<b>Apicomplexans</b>	<b>Apicomplexans</b>
	<i>Trypanosoma rangeli</i>		
	<b>Plants</b>		
	<i>Physcomitrella patens</i>		
<b>Bacteria</b>	<b>Bacteria</b>	<b>Bacteria</b>	<b>Bacteria</b>
<i>Pseudomonas monteilii</i>	<i>Arthrobacter sp.</i>	<i>Mycobacterium intracellulare</i>	<i>Candidatus Planktophila dulcis</i>
<i>Pseudomonas putida</i>	<i>Moorella thermoacetica</i>	<i>Pseudomonas arsenicoxydans</i>	<i>Rhodothermus marinus</i>
<i>Streptomyces autolyticus</i>	<i>Pseudomonas lurida</i>		Uncultured bacterium
<i>Streptomyces malaysiensis</i>	<i>Pseudomonas sp.</i>		
<i>Streptomyces sp. M56</i>			
<i>Vibrio coralliilyticus</i>			
<i>Vibrio tubiashii</i>			
	Marine virus		