



**HAL**  
open science

## Evaluating the probability of CRISPR-based gene drive contaminating another species

Virginie Courtier-orgogozo, Antoine Danchin, Pierre-henri Gouyon,  
Christophe Boëte

► **To cite this version:**

Virginie Courtier-orgogozo, Antoine Danchin, Pierre-henri Gouyon, Christophe Boëte. Evaluating the probability of CRISPR-based gene drive contaminating another species. *Evolutionary Applications*, 2020, 10.1111/eva.12939 . mnhn-02559844

**HAL Id: mnhn-02559844**

**<https://mnhn.hal.science/mnhn-02559844>**

Submitted on 4 Jun 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# 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>  |  
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, EPHE, IRD, Montpellier, France

## Correspondence

Pierre-Henri Gouyon, MNHN, 12 rue Buffon, CP39, F75005 Paris, France.  
Email: pierre-henri.gouyon@mnhn.fr

## Abstract

The probability  $D$  that a given clustered regularly interspaced short palindromic repeats (CRISPR)-based gene drive element contaminates another, nontarget species can be estimated by the following Drive Risk Assessment Quantitative Estimate (DRAQUE) Equation:  $D = (hyb + transf) \times express \times cut \times flank \times immune \times nonextinct$  with  $hyb$  = probability of hybridization between the target species and a nontarget species;  $transf$  = probability of horizontal transfer of a piece of DNA containing the gene drive cassette from the target species to a nontarget 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.

## KEYWORDS

biotechnology, genetically modified organisms, hybridization, molecular evolution, population genetics, risk assessment, wildlife management

## 1 | 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 & Clark, 2018). In recent years, researchers have started to

**Abbreviations:** CRISPR, clustered regularly interspaced short palindromic repeats; DRAQUE, drive risk assessment quantitative estimate; GM, genetically modified; HGT, horizontal gene transfer; HTT, horizontal transfer of transposable element; TE, transposable element.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd

develop clustered regularly interspaced short palindromic repeats (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 nonvector so that they no longer transmit human pathogens (Esvelt, Smidler, Catteruccia, & Church, 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, Estoup, Bourguet, Courtier-Orgogozo, & Débarre, 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, Chavez, Dietz, Esvelt, & Church, 2015), *Drosophila* flies (Champer et al., 2017; Gantz & Bier, 2015; 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: (a) a gene encoding a guide RNA (gRNA) that can recognize a specific target DNA sequence, (b) a *Cas9* gene encoding a *Cas9* endonuclease that can cut DNA at the site specified by the gRNA, (c) 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 (d) 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, Burt, & Godfray, 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.

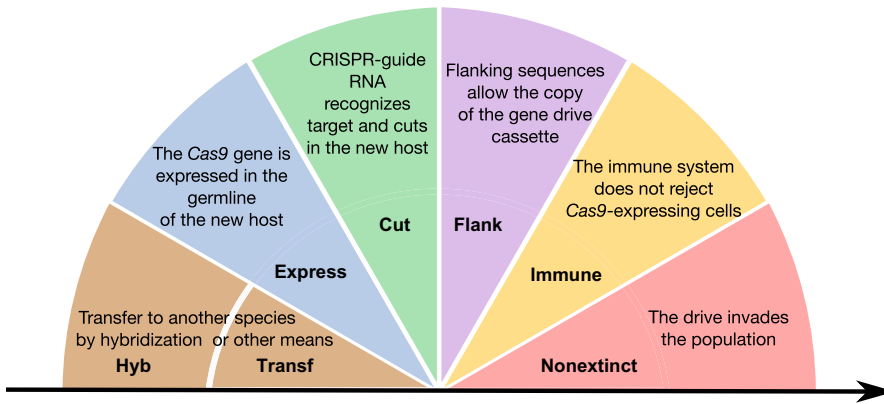
The possibilities offered to humanity in terms of benefits by the new molecular techniques of genome edition CRISPR-Cas are innumerable but also associated with risks which should be carefully monitored (Zhang, 2019). One obvious risk associated with gene drive is that the sequence may escape from the 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 nontarget eukaryote species. The risk of gene drives contaminating another species has been mentioned by several authors (Benedict et al., 2008; Esvelt et al., 2014; National Academies of Sciences & Medicine, 2016; Rode et al., 2019; Webber, Raghu, & Edwards, 2015), 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 × 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 contaminating 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 might be limited or 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 outcome 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 with 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 nontarget populations. However, they are expected to spread less efficiently than autonomous gene drives (Akbari et al., 2015; Esvelt et al., 2014).

## 2 | RESULTS

For a given gene drive construct to contaminate another species, six consecutive steps are required (Figure 1). The probability of



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

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 nontarget species includes (a) the probability that the given piece of DNA passes from the target species to the nontarget species. This can occur by hybridization (*hyb*) or by other means (*transf*). Then, (b) 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 (c) with probability *cut*. The gene drive cassette should insert at the cut site (d) with probability *flank*; the immune system (e) must not eliminate it, with probability *immune*, and finally, it must not be eliminated by stochastic or selective processes (f) with probability *nonextinct*.

This formula could be estimated for one given nontarget 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 nontarget species is contaminated. We examine below each term of the DRAQUE formula.

## 2.1 | Probability of hybridization between the target species and a nontarget species (*hyb*)

The probability that individuals of the target species hybridize with a nontarget 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 pulchella* and *Drosophila subpulchella*; they are found in Japan, China, India, and Nepal ([Takamori, Watabe, Fuyama, Zhang, & Aotsuka, 2006] and references therein). Recent

genome data suggest that hybridization probably occurred recently between *D. suzukii* and *D. subpulchella*, which diverged about 1–9 million years ago (Conner et al., 2017). Furthermore, fertile hybrids between *D. suzukii* and *D. subpulchella* have been obtained in the laboratory ([Fuyama, 1983], note that *D. subpulchella* was erroneously named *D. pulchella* in the 1983 paper [Muto, Kamimura, Tanaka, & Takahashi, 2018]). Because hybridization between closely related species usually produces hybrids with reduced fitness, it can lead to reinforcement, that is, the increase of reproductive isolation as closely related species diverge (Turelli, Lipkowitz, & Brandvain, 2014). Given that reinforcement is prevalent in *Drosophila*, one might reasonably speculate that hybridization is common in the wild. Current data thus suggest that gene drives targeting *D. suzukii* may end up in *D. subpulchella*/*D. pulchella* 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 laboratory-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 independent, 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* and *An. gambiae* in West Africa (Caputo et al., 2011; Marsden et al., 2011; Oliveira et al., 2008) 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 laboratory, introgression of a synthetic sex ratio distortion system has even been possible from *An. gambiae* to its sibling species *An. arabiensis* (Bernardini, Kriezis, Galizi, Nolan, & Crisanti, 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.

Hybridization is common in plants (Goulet, Roda, & Hopkins, 2017; Whitney, Ahern, Campbell, Albert, & King, 2010). Therefore, since the use of genetically modified (GM) plants, the risk of transgene transfer to other species via hybridization has been a general concern (Rizwan et al., 2019; Ryffel, 2014). Unfortunately, despite

extensive discussions about this risk, natural populations have rarely been monitored by researchers to search for such transgene escapes from GM fields. Unintended transgene transfers to other species have been demonstrated in rice, canola, sugar beet, soybean, cotton and bentgrass (Table 1). With the diminishing costs of DNA sequencing it is now possible to quantify more broadly the extent of transgene escape events, but whether the necessary funds will be engaged for such investigations is unclear. Conditions have been proposed to decrease the risk of gene transfer, including change in flower color or flowering date (Rizwan et al., 2019; Ryffel, 2014). However, no system totally preventing gene exchange has yet been proposed.

## 2.2 | Probability of horizontal transfer of a piece of DNA containing the gene drive cassette from the target species to a nontarget 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 (Gilbert et al., 2014; Gilbert, Schaack, Pace, Brindley, & Feschotte, 2010; Houck, Clark, Peterson, & Kidwell, 1991; 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 (Gasmı et al., 2015; Gilbert & Cordaux, 2017). Bacteria probably also constitute an important vector to transfer DNA material from one Eukarya species to another. Horizontal gene transfer (HGT) from Bacteria to Eukarya is frequent (Lacroix & Citovsky, 2016), and from Bacteria to Bacteria, it is the norm. Bacteria release their DNA in the environment because they keep

**TABLE 1** Reported cases of transgene escape from different crops

Crop	From	To	Reference
Bentgrass	<i>Agrostis stolonifera</i>	<i>Polygomon monspeliensis</i>	Zapiola and Mallory-Smith (2012)
Canola	<i>Brassica napus</i>	<i>Brassica rapa</i>	Yoshimura, Beckie, and Matsuo (2006), Warwick, Legere, Simard, and James (2008), Londo, Bautista, Sagers, Lee, and Watrud (2010), Aono et al. (2011)
		<i>Brassica juncea</i>	Zhang et al. (2018)
		<i>Brassica carinata</i>	Séguin-Swartz et al. (2013)
Cotton	<i>Gossypium hirsutum</i>	<i>Gossypium barbadense</i>	Van Deynze, Hutmacher, and Bradford (2011)
Rice	<i>Oryza sativa</i>	weedy rice	Chun et al. (2011)
		<i>Oryza rufipogon</i>	Wang et al. (2006)
		<i>Oryza sativa</i> f. <i>spontanea</i>	Serrat et al. (2013)
Soybean	<i>Glycine max</i>	<i>Glycine soja</i>	Mizuguti et al. (2010)
Sugar beet	<i>Beta vulgaris</i> ssp. <i>vulgaris</i>	<i>Beta vulgaris</i> ssp. <i>maritima</i>	Saeglitz, Pohl, and Bartsch (2000)

lysing either under the action of bacteriophages or during the sporulation process (which entails lysis of the mother cell). This implies that fragments of DNA in the environment or in bacteria 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 & Bordenstein, 2016). Eukaryotic genes in bacteria (EUGENS) are frequent in intracellular parasitic or symbiotic bacteria (e.g., [Hilbi, Hoffmann, & Harrison, 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 (Dunning et al., 2019; Inoue et al., 2017).

Two types of HT can be distinguished, horizontal transfer of transposable elements (HTT) and HGT. Although transposable elements (TEs) 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 or by selection against the insertion if it is deleterious (i.e., the probability *nonextinct* is zero). The breadth of HGT that can be approached based on comparative studies of actual 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 & 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, laboratory-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 (Agren & Clark, 2018; Burt & Koufopanou, 2004). So far, homing endonucleases have only been found in unicellular organisms and in eukaryotes organelles. Compared

to bacterial restriction enzymes, they recognize a long sequence motif, whose size (14–44 bp) is comparable to the one of CRISPR/Cas9 target sites (Hafez & Hausner, 2012). 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, Cho, Mower, Alverson, & Palmer, 2008). This element is also present in several species of fungi, green algae and liverworts, highlighting extensive HT (Cho, Qiu, Kuhlman, & Palmer, 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 & Feschotte, 2018). For example, the *BovB* element moved at least 11 times between snakes, lizards, ruminants and marsupials (Ivancevic, Kortschak, Bertozzi, & Adelson, 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 (Chen, Wu, Zhang, Jiang, & Chen, 2016; Feschotte & Gilbert, 2012; Holmes, 2011; Katzourakis & Gifford, 2010). In vertebrates, endogenous retroviruses (ERVs) can insert into their host genome and generate copies of themselves through germline reinfections or retrotransposition events (Gilbert & Feschotte, 2018; Greenwood, Ishida, O'Brien, Roca, & Eiden, 2018). A comparison of ERV sequences in 65 genomes identified no less than 1,000 HT events between distantly related species of vertebrates (Hayward, Cornwallis, & Jern, 2015). Pan-phylogenomic analyses of ERV sequences revealed that rodents are a major source of retroviruses which they can transmit to other mammals such as livestock (Cui, Tachedjian, & Wang, 2015).

Horizontal transfer events are also ongoing now. The best known case is the worldwide invasion of *D. melanogaster* populations by a TE (*P*-element) originally present in the distantly related species *D. willistoni* (Clark & Kidwell, 1997; Daniels, Peterson, Strausbaugh, Kidwell, & Chovnick, 1990). 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, Schlotterer, & Betancourt, 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 ERV sequence KoRV-A is currently observed to invade natural populations of koala (*Phascolarctos cinereus*) (Xu & Eiden, 2015) and herpes virus DNA has started to integrate into human chromosomes (Morissette & Flamand, 2010).

Rough estimates of HTT rates have been obtained recently. Systematic surveys of TEs 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, Loiseau, Cordaux, & Gilbert, 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 TE family per million years between these three species, with LTR RTs and DNA transposons displaying a higher rate of 0.046 (Bartolomé, Bello, & Maside, 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.

In order to obtain an estimate of the number of contacts a species can have with other eukaryotic species, we used data from the all-taxa biodiversity inventory (ATBI) conducted in the Great Smoky Mountains National Park (Nichols & Langdon, 2007). As it hosts more than 450 vertebrate species and >8,000 insect species (Discover Life in America website, accessed January 2020 <https://dliia.org/smokies-species-tally>), we can assume that a given species can be in contact with approximately 1,000 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 (1,000 species  $\times$  0.035 HTT events per million years—see paragraph above). 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 (Demanèche et al., 2001; Horváth, Merenciano, & González, 2017).

The risk of transfer to nontarget 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 nontarget 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.

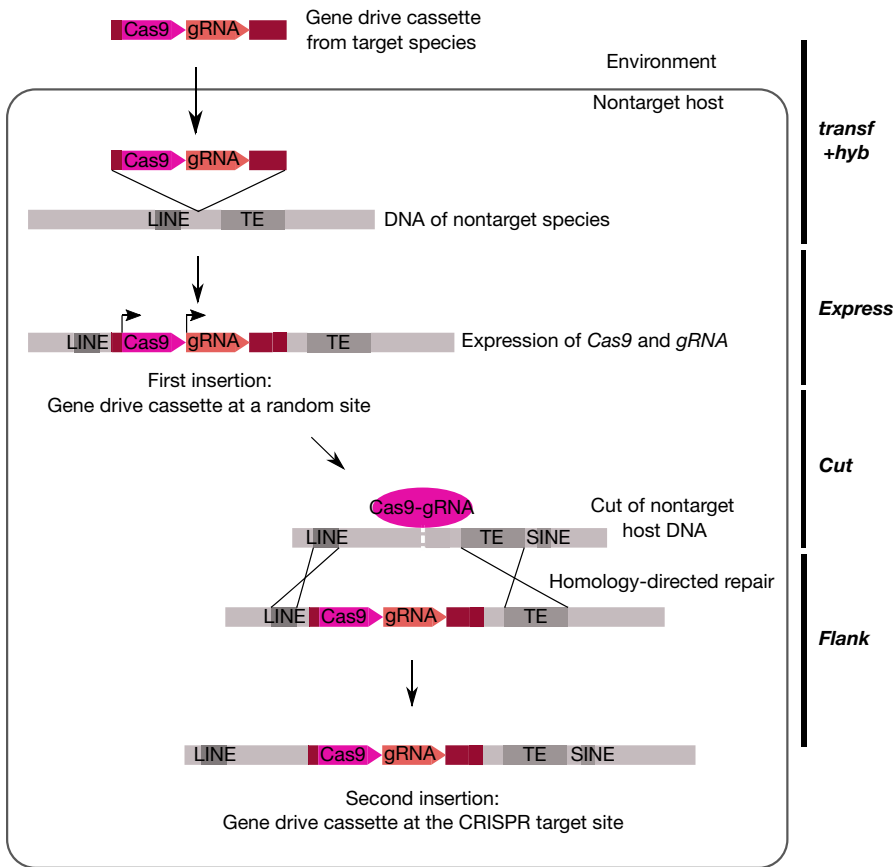
For eradication/suppression drives, the target population is expected to go extinct, so that no gene drive construct is expected to remain in living organisms. Nevertheless, DNA is a very stable molecule and DNA from dead organisms can make up reservoirs

of gene drive constructs. Just taking into account viruses, aquatic ecosystems typically contain  $10^6$ – $10^8$  virus-like particles per mL (Cunningham, Brum, Schwenck, Sullivan, & John, 2015) and sediments  $10^8$ – $10^9$  particles (Filippini & Middelboe, 2007). DNA is found in all kinds of environments (see [Hunter, Ferrante, Meigs-Friend, & Ulmer, 2019] for techniques allowing recovery of 1–500 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. 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 & 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 1,000 per thousand years. Whereas the probability of hybridization discussed above is relatively high and concerns a small number of species, the probability of HT discussed here is relatively low but involves a larger number of potential nontarget species. These nontarget species can be very distant both phylogenetically and geographically, due to vectors which can be ubiquitous.

### 2.3 | 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 nontargeted species, the guide RNA gene and the *Cas9* gene must be expressed in the newly formed zygote or in the new host germline (Figure 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 & 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, Stecher, Suleski, & Hedges, 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 [Booker et al., 2016; Cretokos et al., 2008]) 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, Goltsev, & Levine, 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 [Imamura et al., 2003; Ramos, Kamal, Wimmer, Cartwright, & Monteiro, 2006; Tamura



**FIGURE 2** Succession of putative events leading to the insertion of a functional gene drive cassette in another distantly related nontarget species (in case of horizontal transfer with no hybridization). DNA of the nontarget 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 nontarget host may also be used

et al., 2000]) 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 2). 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 & Wimmer, 2011). For example, the *Drosophila*-derived 3xP3 artificial

**TABLE 2** List of cis-regulatory sequences used in published gene drive constructs

Target Species	Cis-regulatory sequences for Cas9 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 et al. (2015)
<i>Anopheles gambiae</i>	<i>A. gambiae vasa2</i>	<i>A. gambiae U6</i>	Hammond et al. (2016)
<i>Drosophila melanogaster</i>	<i>D. melanogaster vasa</i> <i>D. melanogaster nanos</i>	<i>D. melanogaster U6:3</i>	Champer et al. (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 et al. (2018)
<i>Anopheles gambiae</i>	<i>A. gambiae zpg</i>	<i>A. gambiae U6</i>	Kyrou et al. (2018)
<i>Mus musculus</i>	<i>M. musculus vasa</i> <i>M. musculus Stra8</i>	<i>H. sapiens U6</i>	Grunwald et al. (2019)

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



element drives expression in the eyes not only in insects but also in arthropods (Pavlopoulos & Averof, 2005) and planarians (Gonzalez-Estevez, Momose, Gehring, & Salo, 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 & 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 nontargeted species, we suggest that gene drive vectors should include insulators.

## 2.4 | 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. Constructs with multiple guide RNAs have a greater chance for cutting the DNA in a foreign host compared to single-gRNA constructs. 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 (Gao, Chuai, Yu, Qu, & Liu, 2019; Zhang, Tee, Wang, Huang, & Yang, 2015). The number of off-target cuts tends to accumulate with longer and stronger expression of the *Cas9* gene (Kim, Kim, Cho, Kim, & Kim, 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  $7.10^{-12}$ . In a genome of  $3.10^9$  nucleotides such as the human

haploid genome, such a sequence should be present with a probability  $7.10^{-12} \times 3.10^9$ , 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 nonrandom 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 appears that genomes behave as n-plications of a core set of sequences followed by reduction. This has been observed in yeasts (Escalera-Fanjul, Quezada, Riego-Ruiz, & González, 2018), plants (Clark & Donoghue, 2018) and animals (Hermansen, Hvidsten, Sandve, & Liberles, 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, that is, its nonrandom 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, Heider, Flynn, & O'Neill, 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, Perte, Zimin, & Salzberg, 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, GTTTAACACAGGTCAGCGGNGG, 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, Man, Roland, & Sagui, 2018; Rabeh, Gaboun, Belkadi, & Filali-Maltouf, 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 nontarget species. This trade-off exists not only for the *Anopheles gambiae* mosquito complex but for any species targeted by a gene drive system.

## 2.5 | 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 nontarget 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 nontarget 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 HT 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 (Figure 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 (Hartlerode & Scully, 2009; Jasin, 1996).

First, the gene drive cassette may move to the guide RNA target site via homology-directed repair (Figure 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 & 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 (Kanca et al., 2019; Wang, Lee, & Haber, 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, Wang, Andersen, Zhou, & Pu, 2014). Gene drive constructs published to date have used flanking sequences of about 1 kb (Gantz & 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), 100-bp long in *D. melanogaster* (Kanca et al., 2019) and even 20–40-bp long in zebrafish (Auer & 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 (Figure 2). Evaluating the parameter *flank* thus requires an assessment of the distribution of repeats in potential nontarget 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, Mukerji, Bhatnagar, Kannan, & Brahmachari, 2004). Interestingly this distribution is biased toward proximity of protein-coding gene regulatory regions (Lavi & 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).

Second, the gene drive cassette may move to the guide RNA target site via a TE. It is well established that TEs play a considerable role in displacement of genes or regulatory regions across genomes (Chen & Yang, 2017). As a matter of fact, TEs are recognized as a frequent cause of genetic diseases (see, e.g., [Larsen, Hunnicutt, Larsen, Yoder, & Saunders, 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.

## 2.6 | 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, 2018; Chew et al., 2016). During the gene drive multiplication

process, germline cells produce Cas9 proteins to cut the DNA and insert the gene drive cassette (Figure 2). These germline cells may thus present Cas9 fragments on their surface.

In insects there is no adaptive immune system (Lemaitre & Hoffmann, 2007), so the presence of Cas9 is unlikely to trigger an immune response. However, double-stranded RNAs larger than 30 bp can be recognized by Dicer2 and activate the RNA interference pathway, leading to their degradation (Elbashir, Lendeckel, & Tuschl, 2001; Gammon & Mello, 2015). Guide RNAs are about 100 nucleotide long including the target site and they contain hairpins shorter than 15 bp (Bassett & Liu, 2014), so they should not be recognized by Dicer2. In summary, present knowledge suggests 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 & Heiner, 1989; Roopenian & Akilesh, 2007). However, testes—and maybe ovaries—appear to readily accept foreign antigens without the induction of an immune response in several mammals (Li, Wang, & Han, 2012; Mellor & Munn, 2008; Simpson, 2006), 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, Shin, Zelin, Pestal, & Corn, 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 (Roberts et al., 2012; Shaikh, Leonard, & Martin, 2010). 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 ([Chen et al., 2019; Vela et al., 2017] 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* (SpCas9). Other Cas proteins with activity similar to SpCas9 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.

## 2.7 | Probability of nonextinction 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, on its associated selective value and on the probability that the nontarget population evolve resistance to the gene drive. Several stochastic and deterministic models have been composed to assess the dynamics of gene drive alleles once they are introduced in substantial amount in a targeted population, taking into account the possible costs of harboring a gene drive as well as the appearance of resistance to gene drive (Deredec et al., 2008; Marshall, 2009; Unckless, Messer, Connallon, & Clark, 2015).

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 nontarget 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 modeled as a branching process by Bienaymé (1845), actually published by Cournot (1847), and Watson and Galton (1875), see (Bru, 1991; Kendall, 1975). 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, that is, 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 & Bier, 2015; Gantz et al., 2015; Grunwald et al., 2019; Hammond et al., 2016; KaramiNejadRanjbar et al., 2018; Kyrou et al., 2018) 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  $G(x) = e^{\lambda(x-1)}$ . 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 GM plants methods allowing to reduce the risk of contamination of non-GM plants have been proposed (see above). However, no experimental situation has yet allowed to test the efficiency of such measures. Moreover, despite a strong concern about these questions from citizens and biosafety agencies, none of the methods which could reduce contamination has been developed in commercialized varieties.

A mathematically elegant solution aimed at preventing a drive from invading a nontarget species has been proposed (Barton & Turelli, 2011; Tanaka, Stone, & Nelson, 2017). It is inspired from the population genetics of hybrid zones where a gene can invade

only if it reaches a sufficiently high frequency locally. Such threshold-dependent gene drives appear as a good solution to prevent contamination of nontarget populations. Calculations show that the selective disadvantage of such drives should be more than 0.5 to invade the target population and less than 0.697 so that it cannot invade nontarget populations when starting from a low initial frequency. Note that these values hold only when the segregation distortion of the drive is 100% and that the range is narrower if segregation distortion is lower. One can hope that any drive released in nature will display these characteristics (100% efficiency conversion and  $0.5 < s < 0.697$ ). However, given the narrow range of values for  $s$ , it will not be easy to achieve such drives, especially when gene drive individuals are released in diverse places in nature where various ecological factors can affect fitness values.

### 3 | DISCUSSION

Based on our examination of the seven parameters of the DRAQUE equation (Table 3), 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 hybridization event occurring per year, the guide RNA site is present in the nontarget genome, the nontarget species has high levels of homologous recombination). Our current estimate of the overall risk (Table 3) 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 nontarget species that it invaded. If the gene drive was designed to eliminate a target population, then it is more likely to eliminate the nontarget species. The drive may also have unexpected effects, for example the creation of nontarget mutations

in somatic cells due to perdurance of Cas9 expression (Guichard et al., 2019).

To prevent contamination, several containment strategies for laboratory experiments have been proposed (Akbari et al., 2015; Benedict et al., 2008; National Academies of Sciences & 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 nontarget species: the addition of insulators (see above), and the choice of a guide RNA target site that is not close to repeated sequences or to the centromere, to avoid rearrangements and increased probability of creating an active gene drive in a nontarget 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, Marshall, Ostrovski, Yang, & Akbari, 2018), CleaveR (Oberhofer, Ivy, & Hay, 2019), Killer-Rescue (K-R) (Gould, Huang, Legros, & Lloyd, 2008; Webster, Vella, & Scott, 2019), one or two-locus underdominance (Dhole, Vella, Lloyd, & Gould, 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 & Donoghue, 2018; Harari, Ram, & Kupiec, 2018; Moriyama & 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.

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 nontarget insect species, may be lower for nontarget vertebrate species
<i>nonextinct</i>	0.5–0.7

**TABLE 3** Overview of the various parameters of the DRAQUE equation

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, Bredenoord, & Jongma, 2019)—should remind us that rogue or reckless 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 technology.


## ACKNOWLEDGEMENTS

We deeply thank the reviewers, Andy Clark, 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.

## DATA AVAILABILITY STATEMENT

This article does not contain data.

## ORCID

Virginie Courtier-Orgogozo  <http://orcid.org/0000-0002-9297-9230>

Antoine Danchin  <https://orcid.org/0000-0002-6350-5001>

Pierre-Henri Gouyon  <https://orcid.org/0000-0003-1954-4553>

Christophe Boëte  <https://orcid.org/0000-0001-9043-8652>

## REFERENCES

- Agren, J. A., & Clark, A. G. (2018). Selfish genetic elements. *PLoS Genetics*, *14*, e1007700.
- Akbari, O. S., Bellen, H. J., Bier, E., Bullock, S. L., Burt, A., Church, G. M., ... Wildonger, J. (2015). Safeguarding gene drive experiments in the laboratory. *Science*, *349*, 927–929. <https://doi.org/10.1126/science.aac7932>
- Aono, M., Wakiyama, S., Nagatsu, M., Kaneko, Y., Nishizawa, T., Nakajima, N., ... Saji, H. (2011). Seeds of a possible natural hybrid between herbicide-resistant *Brassica napus* and *Brassica rapa* detected on a riverbank in Japan. *GM Crops*, *2*, 201–210. <https://doi.org/10.4161/gmcr.2.3.18931>
- Auer, T. O., & Del Bene, F. (2014). CRISPR/Cas9 and TALEN-mediated knock-in approaches in zebrafish. *Methods*, *69*, 142–150. <https://doi.org/10.1016/j.jymeth.2014.03.027>
- Bartolomé, C., Bello, X., & Maside, X. (2009). Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes. *Genome Biology*, *10*, R22. <https://doi.org/10.1186/gb-2009-10-2-r22>
- Barton, N. H., & Turelli, M. (2011). Spatial waves of advance with bistable dynamics: Cytoplasmic and genetic analogues of Allee effects. *American Naturalist*, *178*, E48–E75. <https://doi.org/10.1086/661246>
- Bassett, A. R., & Liu, J.-L. (2014). CRISPR/Cas9 and genome editing in *Drosophila*. *Journal of Genetics and Genomics*, *41*, 7–19. <https://doi.org/10.1016/j.jgg.2013.12.004>
- Benedict, M., D'Abbs, P., Dobson, S., Gottlieb, M., Harrington, L., Higgs, S., ... Toure, Y. (2008). Guidance for contained field trials of vector mosquitoes engineered to contain a gene drive system: Recommendations of a scientific working group. *Vector-Borne and Zoonotic Diseases*, *8*(2), 127–166. <https://doi.org/10.1089/vbz.2007.0273>
- Bernardini, F., Kriezis, A., Galizi, R., Nolan, T., & Crisanti, A. (2019). Introgression of a synthetic sex ratio distortion system from *Anopheles gambiae* into *Anopheles arabiensis*. *Scientific Reports*, *9*, 5158.
- Besansky, N. J., Krzywinski, J., Lehmann, T., Simard, F., Kern, M., Mukabayire, O., ... Sagnon, N. (2003). Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: Evidence from multilocus DNA sequence variation. *Proceedings of the National Academy of Sciences*, *100*, 10818–10823. <https://doi.org/10.1073/pnas.1434337100>
- Bienaymé, I.-J. (1845). De la loi de multiplication et de la durée des familles. *Société Philomatique de Paris Extraits*, *5*, 37–39.
- Booker, B. M., Friedrich, T., Mason, M. K., VanderMeer, J. E., Zhao, J., Eckalbar, W. L., ... Ahituv, N. (2016). Bat accelerated regions identify a bat forelimb specific enhancer in the HoxD locus. *PLoS Genetics*, *12*, e1005738. <https://doi.org/10.1371/journal.pgen.1005738>
- Bordenstein, S. R., & Bordenstein, S. R. (2016). Eukaryotic association module in phage WO genomes from *Wolbachia*. *Nature Communications*, *7*, 13155. <https://doi.org/10.1038/ncomms13155>
- Breitwieser, F. P., Perlea, M., Zimin, A. V., & Salzberg, S. L. (2019). Human contamination in bacterial genomes has created thousands of spurious proteins. *Genome Research*, *29*, 954–960. <https://doi.org/10.1101/gr.245373.118>
- Browning, D., & Trobridge, G. (2016). Insulators to improve the safety of retroviral vectors for HIV gene therapy. *Biomedicine*, *4*, 4. <https://doi.org/10.3390/biomedicine4010004>
- Bru, B. (1991). A la recherche de la démonstration perdue de Bienaymé. *Mathématiques Et Sciences Humaines*, *114*, 5–17.
- Buchman, A., Marshall, J. M., Ostrovski, D., Yang, T., & Akbari, O. S. (2018). Synthetically engineered Medea gene drive system in the worldwide crop pest *Drosophila sukuzii*. *Proceedings of the National Academy of Sciences*, *115*, 4725–4730. <https://doi.org/10.1073/pnas.1713139115>
- Burchell, M. J. (2006). W(h)ither the Drake equation? *International Journal of Astrobiology*, *5*, 243–250. <https://doi.org/10.1017/S1473550406003107>
- Burt, A., & Koufopanou, V. (2004). Homing endonuclease genes: The rise and fall and rise again of a selfish element. *Current Opinion in Genetics & Development*, *14*, 609–615. <https://doi.org/10.1016/j.gde.2004.09.010>
- Cande, J., Goltsev, Y., & Levine, M. S. (2009). Conservation of enhancer location in divergent insects. *Proceedings of the National Academy of Sciences*, *106*, 14414–14419. <https://doi.org/10.1073/pnas.0905754106>

- Caputo, B., Santolamazza, F., Vicente, J. L., Nwakanma, D. C., Jawara, M., Palsson, K., ... Torre, A. D. (2011). The "far-west" of *Anopheles gambiae* molecular forms. *PLoS ONE*, 6, e16415. <https://doi.org/10.1371/journal.pone.0016415>
- Champer, J., Chung, J., Lee, Y. L., Liu, C., Yang, E., Wen, Z., ... Messer, P. W. (2019). Molecular safeguarding of CRISPR gene drive experiments. *Elife*, 8, e41439. <https://doi.org/10.7554/eLife.41439>
- Champer, J., Reeves, R., Oh, S. Y., Liu, C., Liu, J., Clark, A. G., & Messer, P. W. (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. <https://doi.org/10.1371/journal.pgen.1006796>
- Charlesworth, C. T., Deshpande, P. S., Dever, D. P., Camarena, J., Lemgart, V. T., Cromer, M. K., ... Porteus, M. H. (2019). Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nature Medicine*, 25, 249. <https://doi.org/10.1038/s41591-018-0326-x>
- Chen, D.-S., Wu, Y.-Q., Zhang, W., Jiang, S.-J., & Chen, S.-Z. (2016). Horizontal gene transfer events reshape the global landscape of arm race between viruses and homo sapiens. *Scientific Reports*, 6, 26934. <https://doi.org/10.1038/srep26934>
- Chen, L.-L., & Yang, L. (2017). ALU alternative regulation for gene expression. *Trends in Cell Biology*, 27, 480–490. <https://doi.org/10.1016/j.tcb.2017.01.002>
- Chen, Y., Xiao, P., Bai, R., Zhang, T., Zhou, Y., Zhou, Z., ... Wu, R. (2019). Extensive adaptive immune response of AAVs and Cas proteins in non-human primates. *BioRxiv*, 588913. <https://doi.org/10.1101/588913>
- Cheng, Y., Ma, Z., Kim, B.-H., Wu, W., Cayting, P., Boyle, A. P., ... Snyder, M. P. (2014). Principles of regulatory information conservation between mouse and human. *Nature*, 515, 371. <https://doi.org/10.1038/nature13985>
- Chew, W. L. (2018). Immunity to CRISPR Cas9 and Cas12a therapeutics. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 10(1), e1408. <https://doi.org/10.1002/wsbm.1408>
- Chew, W. L., Tabebordbar, M., Cheng, J. K. W., Mali, P., Wu, E. Y., Ng, A. H. M., ... Church, G. M. (2016). A multifunctional AAV-CRISPR-Cas9 and its host response. *Nature Methods*, 13, 868. <https://doi.org/10.1038/nmeth.3993>
- Cho, Y., Qiu, Y. L., Kuhlman, P., & Palmer, J. D. (1998). Explosive invasion of plant mitochondria by a group I intron. *Proceedings of the National Academy of Sciences of the United States of America*, 95(24), 14244–14249. <https://doi.org/10.1073/pnas.95.24.14244>
- Chun, Y. J., Kim, D. I., Park, K. W., Kim, H.-J., Jeong, S.-C., An, J. H., ... Kim, C.-G. (2011). Gene flow from herbicide-tolerant GM rice and the heterosis of GM rice-weed F2 progeny. *Planta*, 233, 807–815. <https://doi.org/10.1007/s00425-010-1339-y>
- Clark, J. W., & Donoghue, P. C. (2018). Whole-genome duplication and plant macroevolution. *Trends in Plant Science*, 23(10), 933–945. <https://doi.org/10.1016/j.tplants.2018.07.006>
- Clark, J. B., & Kidwell, M. G. (1997). A phylogenetic perspective on P transposable element evolution in *Drosophila*. *Proceedings of the National Academy of Sciences*, 94, 11428–11433. <https://doi.org/10.1073/pnas.94.21.11428>
- Coetzee, M., Hunt, R. H., Wilkerson, R., Della Torre, A., Coulibaly, M. B., & Besansky, N. J. (2013). *Anopheles coluzzii* and *Anopheles amharicus*, new members of the *Anopheles gambiae* complex. *Zootaxa*, 3619, 246–274. <https://doi.org/10.11646/zootaxa.3619.3.2>
- Conner, W. R., Blaxter, M. L., Anfora, G., Ometto, L., Rota-Stabelli, O., & Turelli, M. (2017). Genome comparisons indicate recent transfer of wR i-like *Wolbachia* between sister species *Drosophila sukuzii* and *D. subpulchrella*. *Ecology and Evolution*, 7, 9391–9404. <https://doi.org/10.1002/ece3.3449>
- Cournot, A. A. (1847). *De l'origine et des limites de la correspondance entre l'algèbre et la géométrie* (399 pp.). Paris: L. Hachette. <https://gallica.bnf.fr/ark:/12148/bpt6k6563896n.texteImage>
- Crawford, J. E., Riehle, M. M., Guelbeogo, W. M., Gnome, A., Sagnon, N., Vernick, K. D., ... Lazzaro, B. P. (2015). Reticulate speciation and barriers to introgression in the *Anopheles gambiae* species complex. *Genome Biology and Evolution*, 7, 3116–3131.
- Cretekos, C. J., Wang, Y., Green, E. D., Martin, J. F., Rasweiler, J. J., & Behringer, R. R. (2008). Regulatory divergence modifies limb length between mammals. *Genes & Development*, 22, 141–151. <https://doi.org/10.1101/gad.1620408>
- Cui, J., Tachedjian, G., & Wang, L.-F. (2015). Bats and rodents shape mammalian retroviral phylogeny. *Scientific Reports*, 5, 16561. <https://doi.org/10.1038/srep16561>
- Cunningham, B. R., Brum, J. R., Schwenck, S. M., Sullivan, M. B., & John, S. G. (2015). An inexpensive, accurate, and precise wet-mount method for enumerating aquatic viruses. *Applied and Environmental Microbiology*, 81(9), 2995–3000. <https://doi.org/10.1128/AEM.03642-14>
- Daniels, S. B., Peterson, K. R., Strausbaugh, L. D., Kidwell, M. G., & Chovnick, A. (1990). Evidence for horizontal transmission of the P transposable element between *Drosophila* species. *Genetics*, 124, 339–355.
- Demanèche, S., Bertolla, F., Buret, F., Nalin, R., Sailland, A., Auriol, P., ... Simonet, P. (2001). Laboratory-scale evidence for light-mediated gene transfer in soil. *Applied and Environmental Microbiology*, 67(8), 3440–3444. <https://doi.org/10.1128/AEM.67.8.3440-3444.2001>
- Deredec, A., Burt, A., & Godfray, H. C. J. (2008). The population genetics of using homing endonuclease genes in vector and pest management. *Genetics*, 179, 2013–2026. <https://doi.org/10.1534/genetics.108.089037>
- Dhole, S., Vella, M. R., Lloyd, A. L., & Gould, F. (2018). Invasion and migration of spatially self-limiting gene drives: A comparative analysis. *Evolutionary Applications*, 11, 794–808. <https://doi.org/10.1111/eva.12583>
- DiCarlo, J. E., Chavez, A., Dietz, S. L., Esvelt, K. M., & Church, G. M. (2015). Safeguarding CRISPR-Cas9 gene drives in yeast. *Nature Biotechnology*, 33, 1250. <https://doi.org/10.1038/nbt.3412>
- Dotto, B. R., Carvalho, E. L., Silva, A. F., Duarte Silva, L. F., Pinto, P. M., Ortiz, M. F., & Wallau, G. L. (2015). HTT-DB: Horizontally transferred transposable elements database. *Bioinformatics*, 31, 2915–2917. <https://doi.org/10.1093/bioinformatics/btv281>
- Dover, G. (1982). Molecular drive: A cohesive mode of species evolution. *Nature*, 299, 111. <https://doi.org/10.1038/299111a0>
- Dunning, L. T., Olofsson, J. K., Parisod, C., Choudhury, R. R., Moreno-Villena, J. J., Yang, Y., ... Christin, P.-A. (2019). Lateral transfers of large DNA fragments spread functional genes among grasses. *Proceedings of the National Academy of Sciences*, 116, 4416–4425. <https://doi.org/10.1073/pnas.1810031116>
- El Baidouri, M., Carpentier, M.-C., Cooke, R., Gao, D., Lasserre, E., Llauro, C., ... Panaud, O. (2014). Widespread and frequent horizontal transfers of transposable elements in plants. *Genome Research*, 24, 831–838. <https://doi.org/10.1101/gr.164400.113>
- Elbashir, S. M., Lendeckel, W., & Tuschl, T. (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes & Development*, 15, 188–200. <https://doi.org/10.1101/gad.862301>
- Escalera-Fanjul, X., Quezada, H., Riego-Ruiz, L., & González, A. (2018). Whole-genome duplication and yeast's fruitful way of life. *Trends in Genetics*, 35(1), 42–54. <https://doi.org/10.1016/j.tig.2018.09.008>
- Esvelt, K. M., Smidler, A. L., Catteruccia, F., & Church, G. M. (2014). Emerging technology: Concerning RNA-guided gene drives for the alteration of wild populations. *Elife*, 3, e03401. <https://doi.org/10.7554/eLife.03401>
- Fertil, B., Massin, M., Lespinats, S., Devic, C., Dumée, P., & Giron, A. (2005). GENSTYLE: Exploration and analysis of DNA sequences with genomic signature. *Nucleic Acids Research*, 33, W512–W515. <https://doi.org/10.1093/nar/gki489>

- Feschotte, C., & Gilbert, C. (2012). Endogenous viruses: Insights into viral evolution and impact on host biology. *Nature Reviews Genetics*, 13, 283. <https://doi.org/10.1038/nrg3199>
- Filippini, M., & Middelboe, M. (2007). Viral abundance and genome size distribution in the sediment and water column of marine and freshwater ecosystems. *FEMS Microbiology Ecology*, 60, 397–410. <https://doi.org/10.1111/j.1574-6941.2007.00298.x>
- Fontaine, M. C., Pease, J. B., Steele, A., Waterhouse, R. M., Neafsey, D. E., Sharakhov, I. V., ... Besansky, N. J. (2015). Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*, 347, 1258524. <https://doi.org/10.1126/science.1258524>
- Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., & Sander, J. D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology*, 31, 822. <https://doi.org/10.1038/nbt.2623>
- Fuyama, Y. (1983). Species-specificity of paragonial substances as an isolating mechanism in *Drosophila*. *Experientia*, 39, 190–192. <https://doi.org/10.1007/BF01958894>
- Gammon, D. B., & Mello, C. C. (2015). RNA interference-mediated antiviral defense in insects. *Current Opinion in Insect Science*, 8, 111–120. <https://doi.org/10.1016/j.cois.2015.01.006>
- Gantz, V. M., & Bier, E. (2015). The mutagenic chain reaction: A method for converting heterozygous to homozygous mutations. *Science*, 348, 442–444. <https://doi.org/10.1126/science.aaa5945>
- Gantz, V. M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V. M., Bier, E., & James, A. A. (2015). Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*. *Proceedings of the National Academy of Sciences*, 112, E6736–E6743. <https://doi.org/10.1073/pnas.1521077112>
- Gao, Y., Chuai, G., Yu, W., Qu, S., & Liu, Q. (2019). Data imbalance in CRISPR off-target prediction. *Briefings in Bioinformatics*, *bbz069*, 1477–4054. <https://doi.org/10.1093/bib/bbz069>
- Gasmi, L., Boulain, H., Gauthier, J., Hua-Van, A., Musset, K., Jakubowska, A. K., ... Drezén, J.-M. (2015). Recurrent domestication by Lepidoptera of genes from their parasites mediated by bracoviruses. *PLoS Genetics*, 11, e1005470. <https://doi.org/10.1371/journal.pgen.1005470>
- Gilbert, C., Chateigner, A., Ernenwein, L., Barbe, V., Bézier, A., Herniou, E. A., & Cordaux, R. (2014). Population genomics supports baculoviruses as vectors of horizontal transfer of insect transposons. *Nature Communications*, 5, 3348. <https://doi.org/10.1038/ncomms4348>
- Gilbert, C., & Cordaux, R. (2017). Viruses as vectors of horizontal transfer of genetic material in eukaryotes. *Current Opinion in Virology*, 25, 16–22. <https://doi.org/10.1016/j.coviro.2017.06.005>
- Gilbert, C., & Feschotte, C. (2018). Horizontal acquisition of transposable elements and viral sequences: Patterns and consequences. *Current Opinion in Genetics & Development*, 49, 15–24. <https://doi.org/10.1016/j.gde.2018.02.007>
- Gilbert, C., Peccoud, J., Chateigner, A., Moumen, B., Cordaux, R., & Herniou, E. A. (2016). Continuous influx of genetic material from host to virus populations. *PLoS Genetics*, 12, e1005838. <https://doi.org/10.1371/journal.pgen.1005838>
- Gilbert, C., Schaack, S., Pace, J. K. II, Brindley, P. J., & Feschotte, C. (2010). A role for host-parasite interactions in the horizontal transfer of transposons across phyla. *Nature*, 464, 1347. <https://doi.org/10.1038/nature08939>
- Gonzalez-Estevéz, C., Momose, T., Gehring, W. J., & Salo, E. (2003). Transgenic planarian lines obtained by electroporation using transposon-derived vectors and an eye-specific GFP marker. *Proceedings of the National Academy of Sciences*, 100, 14046–14051. <https://doi.org/10.1073/pnas.2335980100>
- Gould, F., Huang, Y., Legros, M., & Lloyd, A. L. (2008). A Killer-Rescue system for self-limiting gene drive of anti-pathogen constructs. *Proceedings of the Royal Society B-Biological Sciences*, 275, 2823–2829. <https://doi.org/10.1098/rspb.2008.0846>
- Goulet, B. E., Roda, F., & Hopkins, R. (2017). Hybridization in plants: Old ideas, new techniques. *Plant Physiology*, 173, 65–78. <https://doi.org/10.1104/pp.16.01340>
- Greenwood, A. D., Ishida, Y., O'Brien, S. P., Roca, A. L., & Eiden, M. V. (2018). Transmission, evolution, and endogenization: Lessons learned from recent retroviral invasions. *Microbiology and Molecular Biology Reviews*, 82, e00044–e00117.
- Grover, D., Mukerji, M., Bhatnagar, P., Kannan, K., & Brahmachari, S. K. (2004). Alu repeat analysis in the complete human genome: Trends and variations with respect to genomic composition. *Bioinformatics*, 20, 813–817. <https://doi.org/10.1093/bioinformatics/bth005>
- Grunwald, H. A., Gantz, V. M., Poplawski, G., Xu, X.-R., Bier, E., & Cooper, K. L. (2019). Super-Mendelian inheritance mediated by CRISPR-Cas9 in the female mouse germline. *Nature*, 566, 105. <https://doi.org/10.1038/s41586-019-0875-2>
- Guichard, A., Haque, T., Bobik, M., Xu, X.-R., Klanssek, C., Kushwah, R. B. S., ... Bier, E. (2019). Efficient allelic-drive in *Drosophila*. *Nature Communications*, 10, 1640. <https://doi.org/10.1038/s41467-019-09694-w>
- Hafez, M., & Hausner, G. (2012). Homing endonucleases: DNA scissors on a mission. *Genome*, 55, 553–569. <https://doi.org/10.1139/g2012-049>
- Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., ... Nolan, T. (2016). A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. *Nature Biotechnology*, 34, 78. <https://doi.org/10.1038/nbt.3439>
- Harari, Y., Ram, Y., & Kupiec, M. (2018). Frequent ploidy changes in growing yeast cultures. *Current Genetics*, 64, 1001–1004. <https://doi.org/10.1007/s00294-018-0823-y>
- Hartlerode, A. J., & Scully, R. (2009). Mechanisms of double-strand break repair in somatic mammalian cells. *The Biochemical Journal*, 423, 157–168. <https://doi.org/10.1042/BJ20090942>
- Hayward, A., Cornwallis, C. K., & Jern, P. (2015). Pan-vertebrate comparative genomics unmasks retrovirus macroevolution. *Proceedings of the National Academy of Sciences*, 112, 464–469. <https://doi.org/10.1073/pnas.1414980112>
- Hermansen, R. A., Hvidsten, T. R., Sandve, S. R., & Liberles, D. A. (2016). Extracting functional trends from whole genome duplication events using comparative genomics. *Biological Procedures Online*, 18, 11. <https://doi.org/10.1186/s12575-016-0041-2>
- Hilbi, H., Hoffmann, C., & Harrison, C. F. (2011). *Legionella* spp. outdoors: Colonization, communication and persistence. *Environmental Microbiology Reports*, 3, 286–296. <https://doi.org/10.1111/j.1758-2229.2011.00247.x>
- Hill, T., Schlötterer, C., & Betancourt, A. J. (2016). Hybrid dysgenesis in *Drosophila simulans* associated with a rapid invasion of the P-element. *PLoS Genetics*, 12, e1005920. <https://doi.org/10.1371/journal.pgen.1005920>
- Hisano, Y. U., Sakuma, T., Nakade, S., Ohga, R., Ota, S., Okamoto, H., ... Kawahara, A. (2015). Precise in-frame integration of exogenous DNA mediated by CRISPR/Cas9 system in zebrafish. *Scientific Reports*, 5, 8841. <https://doi.org/10.1038/srep08841>
- Holmes, E. C. (2011). What does virus evolution tell us about virus origins? *Journal of Virology*, 85, 5247–5251. <https://doi.org/10.1128/JVI.02203-10>
- Horváth, V., Merenciano, M., & González, J. (2017). Revisiting the relationship between transposable elements and the eukaryotic stress response. *Trends in Genetics*, 33, 832–841. <https://doi.org/10.1016/j.tig.2017.08.007>
- Houck, M. A., Clark, J. B., Peterson, K. R., & Kidwell, M. G. (1991). Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis*. *Science*, 253, 1125–1128. <https://doi.org/10.1126/science.1653453>
- Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., ... Zhang, F. (2013). DNA targeting specificity of

- RNA-guided Cas9 nucleases. *Nature Biotechnology*, 31, 827. <https://doi.org/10.1038/nbt.2647>
- Hunter, M. E., Ferrante, J. A., Meigs-Friend, G., & Ulmer, A. (2019). Improving eDNA yield and inhibitor reduction through increased water volumes and multi-filter isolation techniques. *Scientific Reports*, 9, 5259. <https://doi.org/10.1038/s41598-019-40977-w>
- Imamura, M., Nakai, J., Inoue, S., Quan, G. X., Kanda, T., & Tamura, T. (2003). Targeted gene expression using the GAL4/UAS system in the silkworm *Bombyx mori*. *Genetics*, 165, 1329–1340.
- Inoue, Y., Saga, T., Aikawa, T., Kumagai, M., Shimada, A., Kawaguchi, Y., ... Takeda, H. (2017). Complete fusion of a transposon and herpesvirus created the Teratorn mobile element in medaka fish. *Nature Communications*, 8, 551. <https://doi.org/10.1038/s41467-017-00527-2>
- Ivančević, A. M., Kortschak, R. D., Bertozzi, T., & Adelson, D. L. (2018). Horizontal transfer of BovB and L1 retrotransposons in eukaryotes. *Genome Biology*, 19(1), 85.
- Jasin, M. (1996). Genetic manipulation of genomes with rare-cutting endonucleases. *Trends in Genetics*, 12, 224–228. [https://doi.org/10.1016/0168-9525\(96\)10019-6](https://doi.org/10.1016/0168-9525(96)10019-6)
- Kanca, O., Zirin, J., Garcia-Marques, J., Knight, S. M., Yang-Zhou, D., Amador, G., ... Bellen, H. J. (2019). An efficient CRISPR-based strategy to insert small and large fragments of DNA using short homology arms. *Elife*, 8, e51539. <https://doi.org/10.7554/eLife.51539>
- KaramiNejadRanjbar, M., Eckermann, K. N., Ahmed, H. M. M., Sánchez C., H. M., Dippel, S., Marshall, J. M., & Wimmer, E. A. (2018). Consequences of resistance evolution in a Cas9-based sex conversion-suppression gene drive for insect pest management. *Proceedings of the National Academy of Sciences*, 115, 6189–6194. <https://doi.org/10.1073/pnas.1713825115>
- Katzourakis, A., & Gifford, R. J. (2010). Endogenous viral elements in animal genomes. *PLoS Genetics*, 6, e1001191. <https://doi.org/10.1371/journal.pgen.1001191>
- Kendall, D. G. (1975). The genealogy of genealogy branching processes before (and after) 1873. *Bulletin of the London Mathematical Society*, 7(3), 225–253. <https://doi.org/10.1112/blms/7.3.225>
- Kim, S., Kim, D., Cho, S. W., Kim, J., & Kim, J.-S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Research*, 24, 1012–1019. <https://doi.org/10.1101/gr.171322.113>
- Kim, S., Koo, T., Jee, H.-G., Cho, H.-Y., Lee, G., Lim, D.-G., ... Kim, J.-S. (2018). CRISPR RNAs trigger innate immune responses in human cells. *Genome Research*, 28, 367–373. <https://doi.org/10.1101/gr.231936.117>
- Kumar, S., Stecher, G., Suleski, M., & Hedges, S. B. (2017). TimeTree: A resource for timelines, timetrees, and divergence times. *Molecular Biology and Evolution*, 34, 1812–1819. <https://doi.org/10.1093/molbev/msx116>
- Kvon, E. Z., Kamneva, O. K., Melo, U. S., Barozzi, I., Osterwalder, M., Mannion, B. J., ... Visel, A. (2016). Progressive loss of function in a limb enhancer during snake evolution. *Cell*, 167, 633–642. <https://doi.org/10.1016/j.cell.2016.09.028>
- Kyrou, K., Hammond, A. M., Galizi, R., Kranjc, N., Burt, A., Beaghton, A. K., ... Crisanti, A. (2018). A CRISPR-Cas9 gene drive targeting doublesex causes complete population suppression in caged *Anopheles gambiae* mosquitoes. *Nature Biotechnology*, 36, 1062. <https://doi.org/10.1038/nbt.4245>
- Lacroix, B., & Citovsky, V. (2016). Transfer of DNA from bacteria to eukaryotes. *MBio*, 7, e00863–e00916. <https://doi.org/10.1128/mBio.00863-16>
- Lai, Y.-T., Deem, K. D., Borràs-Castells, F., Sambrani, N., Rudolf, H., Suryamohan, K., ... Tomoyasu, Y. (2018). Enhancer identification and activity evaluation in the red flour beetle, *Tribolium castaneum*. *Development*, 145, dev160663.
- Larsen, P. A., Hunnicutt, K. E., Larsen, R. J., Yoder, A. D., & Saunders, A. M. (2018). Warning SINEs: Alu elements, evolution of the human brain, and the spectrum of neurological disease. *Chromosome Research*, 26, 93–111. <https://doi.org/10.1007/s10577-018-9573-4>
- Lavi, E., & Carmel, L. (2018). Alu exaptation enriches the human transcriptome by introducing new gene ends. *RNA Biology*, 15, 715–725.
- Leitschuh, C. M., Kanavy, D., Backus, G. A., Valdez, R. X., Serr, M., Pitts, E. A., ... Godwin, J. (2018). Developing gene drive technologies to eradicate invasive rodents from islands. *Journal of Responsible Innovation*, 5, S121–S138. <https://doi.org/10.1080/23299460.2017.1365232>
- Lemaître, B., & Hoffmann, J. (2007). The host defense of *Drosophila melanogaster*. *Annual Review of Immunology*, 25, 697–743.
- Li, K., Wang, G., Andersen, T., Zhou, P., & Pu, W. T. (2014). Optimization of genome engineering approaches with the CRISPR/Cas9 system. *PLoS ONE*, 9, e105779. <https://doi.org/10.1371/journal.pone.0105779>
- Li, N., Wang, T., & Han, D. (2012). Structural, cellular and molecular aspects of immune privilege in the testis. *Frontiers in Immunology*, 3, 152. <https://doi.org/10.3389/fimmu.2012.00152>
- Liu, Q., Li, X., Zhou, X., Li, M., Zhang, F., Schwarzacher, T., & Heslop-Harrison, J. S. (2019). The repetitive DNA landscape in *Avena* (Poaceae): Chromosome and genome evolution defined by major repeat classes in whole-genome sequence reads. *BMC Plant Biology*, 19, 226. <https://doi.org/10.1186/s12870-019-1769-z>
- Londo, J. P., Bautista, N. S., Sagers, C. L., Lee, E. H., & Watrud, L. S. (2010). Glyphosate drift promotes changes in fitness and transgene gene flow in canola (*Brassica napus*) and hybrids. *Annals of Botany*, 106, 957–965. <https://doi.org/10.1093/aob/mcq190>
- Madani, G., & Heiner, D. C. (1989). Antibody transmission from mother to fetus. *Current Opinion in Immunology*, 1, 1157–1164. [https://doi.org/10.1016/0952-7915\(89\)90009-5](https://doi.org/10.1016/0952-7915(89)90009-5)
- Mamcarz, E., Zhou, S., Lockey, T., Abdelsamed, H., Cross, S. J., Kang, G., ... Li, C. (2019). Lentiviral gene therapy combined with low-dose busulfan in infants with SCID-X1. *New England Journal of Medicine*, 380, 1525–1534.
- Mancini, E., Spinaci, M. I., Gordicho, V., Caputo, B., Pombi, M., & Vicente, J. L., ... Pinto, J. (2015). Adaptive potential of hybridization among malaria vectors. *PLoS ONE*, 10(6), e0127804. <https://doi.org/10.1371/journal.pone.0127804>
- Mantilla-Calderon, D., Plewa, M. J., Michoud, G., Fodelianakis, S., Daffonchio, D., & Hong, P.-Y. (2019). Water disinfection byproducts increase natural transformation rates of environmental DNA in *Acinetobacter baylyi* ADP1. *Environmental Science & Technology*, 53, 6520–6528. <https://doi.org/10.1021/acs.est.9b00692>
- Marsden, C. D., Lee, Y., Nieman, C. C., Sanford, M. R., Dinis, J., Martins, C., ... Lanzaro, G. C. (2011). Asymmetric introgression between the M and S forms of the malaria vector, *Anopheles gambiae*, maintains divergence despite extensive hybridization. *Molecular Ecology*, 20, 4983–4994. <https://doi.org/10.1111/j.1365-294X.2011.05339.x>
- Marshall, J. M. (2009). The effect of gene drive on containment of transgenic mosquitoes. *Journal of Theoretical Biology*, 258, 250–265. <https://doi.org/10.1016/j.jtbi.2009.01.031>
- Mellor, A. L., & Munn, D. H. (2008). Creating immune privilege: Active local suppression that benefits friends, but protects foes. *Nature Reviews Immunology*, 8, 74. <https://doi.org/10.1038/nri2233>
- Mizuguti, A., Ohigashi, K., Yoshimura, Y., Kaga, A., Kuroda, Y., & Matsuo, K. (2010). Hybridization between GM soybean (*Glycine max* (L.) Merr.) and wild soybean (*Glycine soja* Sieb. et Zucc.) under field conditions in Japan. *Environmental Biosafety Research*, 9, 13–23. <https://doi.org/10.1051/embr/2010004>
- Monlong, J., Cossette, P., Meloche, C., Rouleau, G., Girard, S. L., & Bourque, G. (2018). Human copy number variants are enriched in regions of low mappability. *Nucleic Acids Research*, 46, 7236–7249. <https://doi.org/10.1093/nar/gky538>
- Moran, N. A., & Jarvik, T. (2010). Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science*, 328, 624–627. <https://doi.org/10.1126/science.1187113>



- Morissette, G., & Flamand, L. (2010). Herpesviruses and chromosomal integration. *Journal of Virology*, *84*, 12100–12109. <https://doi.org/10.1128/JVI.01169-10>
- Moriyama, Y., & Koshihara-Takeuchi, K. (2018). Significance of whole-genome duplications on the emergence of evolutionary novelties. *Briefings in Functional Genomics*, *17*(5), 329–338. <https://doi.org/10.1093/bfgp/ely007>
- Muto, L., Kamimura, Y., Tanaka, K. M., & Takahashi, A. (2018). An innovative ovipositor for niche exploitation impacts genital coevolution between sexes in a fruit-damaging *Drosophila*. *Proceedings of the Royal Society B-Biological Sciences*, *285*, 20181635. <https://doi.org/10.1098/rspb.2018.1635>
- Nakade, S., Tsubota, T., Sakane, Y., Kume, S., Sakamoto, N., Obara, M., ... Suzuki, K.-I. (2014). Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nature Communications*, *5*, 5560. <https://doi.org/10.1038/ncomms6560>
- National Academies of Sciences, and Medicine. (2016). *Gene drives on the horizon: Advancing science, navigating uncertainty, and aligning research with public values* (230 pp.). Washington, DC: National Academies Press. <https://doi.org/10.17226/23405>
- Nichols, B. J., & Langdon, K. R. (2007). The Smokies all taxa biodiversity inventory: History and progress. *Southeastern Naturalist*, *6*, 27–34. [https://doi.org/10.1656/1528-7092\(2007\)6\[27:TSATBI\]2.0.CO;2](https://doi.org/10.1656/1528-7092(2007)6[27:TSATBI]2.0.CO;2)
- Noble, C., Min, J., Olejarz, J., Buchthal, J., Chavez, A., Smidler, A. L., ... Esvelt, K. M. (2019). Daisy-chain gene drives for the alteration of local populations. *Proceedings of the National Academy of Sciences*, *116*, 8275–8282. <https://doi.org/10.1073/pnas.1716358116>
- Oberhofer, G., Ivy, T., & 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*, *116*, 6250–6259. <https://doi.org/10.1073/pnas.1816928116>
- Oliveira, E., Salgueiro, P., Palsson, K., Vicente, J. L., Arez, A. P., Jaenson, T. G., ... Pinto, J. (2008). High levels of hybridization between molecular forms of *Anopheles gambiae* from Guinea Bissau. *Journal of Medical Entomology*, *45*, 1057–1063. <https://doi.org/10.1093/jmedent/45.6.1057>
- Oliveira, J., & Reygaert, W. C. (2020). Gram negative bacteria. In *StatPearls [Internet]*. Treasure Island (FL): StatPearls Publishing. <https://www.ncbi.nlm.nih.gov/books/NBK538213/>
- Paix, A., Wang, Y., Smith, H. E., Lee, C.-Y.-S., Calidas, D., Lu, T., ... Seydoux, G. (2014). Scalable and versatile genome editing using linear DNAs with microhomology to Cas9 Sites in *Caenorhabditis elegans*. *Genetics*, *198*, 1347–1356. <https://doi.org/10.1534/genetics.114.170423>
- Palazzo, A., Lorusso, P., Miskey, C., Walisko, O., Gerbino, A., Marobbio, C. M. T., ... Marsano, R. M. (2019). Transcriptionally promiscuous “blurry” promoters in Tc1/mariner transposons allow transcription in distantly related genomes. *Mobile DNA*, *10*, 13. <https://doi.org/10.1186/s13100-019-0155-6>
- Pan, F., Man, V. H., Roland, C., & Sagui, C. (2018). Structure and dynamics of DNA and RNA double helices obtained from the CCG and GGC trinucleotide repeats. *The Journal of Physical Chemistry B*, *122*, 4491–4512. <https://doi.org/10.1021/acs.jpcc.8b01658>
- Parisot, N., Pelin, A., Gasc, C., Polonais, V., Belkorchia, A., Panek, J., ... Lerat, E. (2014). Microsporidian genomes harbor a diverse array of transposable elements that demonstrate an ancestry of horizontal exchange with metazoans. *Genome Biology and Evolution*, *6*(9), 2289–2300. <https://doi.org/10.1093/gbe/evu178>
- Pavlopoulos, A., & Averof, M. (2005). Establishing genetic transformation for comparative developmental studies in the crustacean *Parhyale hawaiiensis*. *Proceedings of the National Academy of Sciences*, *102*, 7888–7893. <https://doi.org/10.1073/pnas.0501101102>
- Peccoud, J., Loiseau, V., Cordaux, R., & Gilbert, C. (2017). Massive horizontal transfer of transposable elements in insects. *Proceedings of the National Academy of Sciences*, *114*, 4721–4726. <https://doi.org/10.1073/pnas.1621178114>
- Perrow, C. (2011). *Normal accidents: Living with high risk technologies-Updated edition* (464 pp.). Princeton, NJ: Princeton University Press. ISBN 9780691004129.
- Rabeh, K., Gaboun, F., Belkadi, B., & Filali-Maltouf, A. (2018). In Silico development of new SSRs primer for aquaporin linked to drought tolerance in plants. *Plant Signaling & Behavior*, *13*, e1536630. <https://doi.org/10.1080/15592324.2018.1536630>
- Ramos, D. M., Kamal, F., Wimmer, E. A., Cartwright, A. N., & Monteiro, A. (2006). Temporal and spatial control of transgene expression using laser induction of the hsp70 promoter. *BMC Developmental Biology*, *6*, 55.
- Reiss, D., Mialdea, G., Miele, V., de Vienne, D. M., Peccoud, J., Gilbert, C., ... Charlat, S. (2019). Global survey of mobile DNA horizontal transfer in arthropods reveals Lepidoptera as a prime hotspot. *PLoS Genetics*, *15*, e1007965. <https://doi.org/10.1371/journal.pgen.1007965>
- Rizwan, M., Hussain, M., Shimelis, H., Hameed, M. U., Atif, R. M., Azhar, M. T., ... Asif, M. (2019). Gene flow from major genetically modified crops and strategies for containment and mitigation of transgene escape: A review. *Applied Ecology and Environmental Research*, *17*, 11191–11208. [https://doi.org/10.15666/aeer/1705\\_1119111208](https://doi.org/10.15666/aeer/1705_1119111208)
- Roberts, A. L., Connolly, K. L., Kirse, D. J., Evans, A. K., Poehling, K. A., Peters, T. R., & Reid, S. D. (2012). Detection of group A *Streptococcus* in tonsils from pediatric patients reveals high rate of asymptomatic streptococcal carriage. *BMC Pediatrics*, *12*, 3. <https://doi.org/10.1186/1471-2431-12-3>
- Rode, N. O., Estoup, A., Bourguet, D., Courtier-Orgogozo, V., & Débarre, F. (2019). Population management using gene drive: Molecular design, models of spread dynamics and assessment of ecological risks. *Conservation Genetics*, *20*, 671–690. <https://doi.org/10.1007/s10592-019-01165-5>
- Roggenkamp, E., Giersch, R. M., Schrock, M. N., Turnquist, E., Halloran, M., Finnigan, G. C. (2018). Tuning CRISPR-Cas9 gene drives in *Saccharomyces cerevisiae*. *G3: Genes, Genomes, Genetics*, *8*, 999–1018.
- Roopenian, D. C., & Akilesh, S. (2007). FcRn: The neonatal Fc receptor comes of age. *Nature Reviews Immunology*, *7*, 715. <https://doi.org/10.1038/nri2155>
- Ryffel, G. U. (2014). Transgene flow: Facts, speculations and possible countermeasures. *GM Crops & Food*, *5*(4), 249–258. <https://doi.org/10.4161/21645698.2014.945883>
- Saeglitz, C., Pohl, M., & Bartsch, D. (2000). Monitoring gene flow from transgenic sugar beet using cytoplasmic male-sterile bait plants. *Molecular Ecology*, *9*, 2035–2040. <https://doi.org/10.1046/j.1365-294X.2000.01109.x>
- Sanchez-Puerta, M. V., Cho, Y., Mower, J. P., Alverson, A. J., & Palmer, J. D. (2008). Frequent, phylogenetically local horizontal transfer of the *cox1* group I Intron in flowering plant mitochondria. *Molecular Biology and Evolution*, *25*, 1762–1777. <https://doi.org/10.1093/molbev/msn129>
- Sand, M., Bredenoord, A. L., & Jongsma, K. R. (2019). After the fact—The case of CRISPR babies. *European Journal of Human Genetics*, *27*(11), 1621–1624. <https://doi.org/10.1038/s41431-019-0459-5>
- Schetelig, M. F., & Wimmer, E. A. (2011). Insect transgenesis and the sterile insect technique. In A. Vilcinskis (Ed.), *Insect biotechnology* (pp. 169–194). Springer. <https://doi.org/10.1007/978-90-481-9641-8>
- Schinko, J. B., Weber, M., Viktorinova, I., Kiupakis, A., Averof, M., Klingler, M., ... Bucher, G. (2010). Functionality of the GAL4/UAS system in *Tribolium* requires the use of endogenous core promoters. *BMC Developmental Biology*, *10*, 53. <https://doi.org/10.1186/1471-213X-10-53>
- Schultz, E., Barraud, O., Madec, J.-Y., Haenni, M., Cloeckeaert, A., Ploy, M. C., & Doublet, B. (2017). Multidrug resistance *Salmonella* genomic island 1 in a *Morganella morganii* subsp. *morganii* human clinical isolate

- from France. *Msphere*, 2, e00118–e00217. <https://doi.org/10.1128/msphere.00118-17>
- Scott, M. J., Gould, F., Lorenzen, M., Grubbs, N., Edwards, O., & O'Brochta, D. (2018). Agricultural production: Assessment of the potential use of Cas9-mediated gene drive systems for agricultural pest control. *Journal of Responsible Innovation*, 5, S98–S120. <https://doi.org/10.1080/23299460.2017.1410343>
- Séguin-Swartz, G., Beckie, H. J., Warwick, S. I., Roslinsky, V., Nettleton, J. A., Johnson, E. N., & Falk, K. C. (2013). Pollen-mediated gene flow between glyphosate-resistant *Brassica napus* canola and *B. juncea* and *B. carinata* mustard crops under large-scale field conditions in Saskatchewan. *Canadian Journal of Plant Science*, 93, 1083–1087. <https://doi.org/10.4141/cjps2013-129>
- Serrat, X., Esteban, R., Penas, G., Catala, M. M., Mele, E., & Messeguer, J. (2013). Direct and reverse pollen-mediated gene flow between GM rice and red rice weed. *AoB Plants*, 5, plt050. <https://doi.org/10.1093/aobpla/plt050>
- Serrato-Capuchina, A. G., Zhang, S., Martin, W., Peede, D., Earley, E., & Matute, D. R. (2018). Recent invasion of P transposable element into *Drosophila yakuba*. *bioRxiv*, 453829. <https://doi.org/10.1101/453829>
- Shaikh, N., Leonard, E., & Martin, J. M. (2010). Prevalence of streptococcal pharyngitis and streptococcal carriage in children: A meta-analysis. *Pediatrics*, 126, e557–e564. <https://doi.org/10.1542/peds.2009-2648>
- Shapiro, R. S., Chavez, A., Porter, C. B. M., Hamblin, M., Kaas, C. S., DiCarlo, J. E., ... Collins, J. J. (2018). A CRISPR–Cas9-based gene drive platform for genetic interaction analysis in *Candida albicans*. *Nature Microbiology*, 3, 73. <https://doi.org/10.1038/s41564-017-0043-0>
- Sharakhov, I. V., White, B. J., Sharakhova, M. V., Kayondo, J., Lobo, N. F., Santolamazza, F., ... Besansky, N. J. (2006). Breakpoint structure reveals the unique origin of an interspecific chromosomal inversion (2La) in the *Anopheles gambiae* complex. *Proceedings of the National Academy of Sciences*, 103, 6258–6262. <https://doi.org/10.1073/pnas.0509683103>
- Simpson, E. (2006). A historical perspective on immunological privilege. *Immunological Reviews*, 213, 12–22. <https://doi.org/10.1111/j.1600-065X.2006.00434.x>
- Smalec, B. M., Heider, T. N., Flynn, B. L., & O'Neill, R. J. (2019). A centromere satellite concomitant with extensive karyotypic diversity across the *Peromyscus* genus defies predictions of molecular drive. *Chromosome Research*, 27, 237–252. <https://doi.org/10.1007/s10577-019-09605-1>
- Song, X., Beck, C. R., Du, R., Campbell, I. M., Coban-Akdemir, Z., Gu, S., ... Lupski, J. R. (2018). Predicting human genes susceptible to genomic instability associated with Alu/Alu-mediated rearrangements. *Genome Research*, 28, 1228–1242. <https://doi.org/10.1101/gr.229401.117>
- Takamori, H., Watabe, H.-A., Fuyama, Y., Zhang, Y.-P., & Aotsuka, T. (2006). *Drosophila subpulchrella*, a new species of the *Drosophila suzukii* species subgroup from Japan and China (Diptera: Drosophilidae). *Entomological Science*, 9, 121–128. <https://doi.org/10.1111/j.1479-8298.2006.00159.x>
- Tamura, T., Thibert, C., Royer, C., Kanda, T., Eappen, A., Kamba, M., ... Couple, P. (2000). Germline transformation of the silkworm *Bombyx mori* L. using a piggyBac transposon-derived vector. *Nature Biotechnology*, 18, 81. <https://doi.org/10.1038/71978>
- Tanaka, H., Stone, H. A., & Nelson, D. R. (2017). Spatial gene drives and pushed genetic waves. *Proceedings of the National Academy of Sciences*, 114, 8452–8457. <https://doi.org/10.1073/pnas.1705868114>
- Traynor, S., Møllegaard, N. E., Jørgensen, M. G., Brückmann, N. H., Pedersen, C. B., Terp, M. G., ... Gjerstorff, M. F. (2019). Remodeling and destabilization of chromosome 1 pericentromeric heterochromatin by SSX proteins. *Nucleic Acids Research*, 47(13), 6668–6684. <https://doi.org/10.1093/nar/gkz396>
- Turelli, M., Lipkowitz, J. R., & Brandvain, Y. (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. <https://doi.org/10.1111/evo.12330>
- Unckless, R. L., Messer, P. W., Connallon, T., & Clark, A. G. (2015). Modeling the manipulation of natural populations by the mutagenic chain reaction. *Genetics*, 201, 425–431. <https://doi.org/10.1534/genetics.115.177592>
- Van Deynze, A. E., Hutmacher, R. B., & Bradford, K. J. (2011). Gene flow between *Gossypium hirsutum* L. and *Gossypium barbadense* L. is asymmetric. *Crop Science*, 51, 298–305. <https://doi.org/10.2135/crops.c2010.04.0213>
- Vela, A. I., Villalón, P., Sáez-Nieto, J. A., Chacón, G., Domínguez, L., & Fernández-Garayzábal, J. F. (2017). Characterization of *Streptococcus pyogenes* from animal clinical specimens, Spain. *Emerging Infectious Diseases*, 23, 2011.
- Wagner, D. L., Amini, L., Wendering, D. J., Burkhardt, L.-M., Akyüz, L., Reinke, P., ... Schmueck-Henneresse, M. (2019). High prevalence of *Streptococcus pyogenes* Cas9-reactive T cells within the adult human population. *Nature Medicine*, 25, 242. <https://doi.org/10.1038/s41591-018-0204-6>
- Wang, F., Yuan, Q.-H., Shi, L., Qian, Q., Liu, W.-G., Kuang, B.-G., ... Jia, S.-R. (2006). A large-scale field study of transgene flow from cultivated rice (*Oryza sativa*) to common wild rice (*O. rufipogon*) and barnyard grass (*Echinochloa crusgalli*). *Plant Biotechnology Journal*, 4, 667–676. <https://doi.org/10.1111/j.1467-7652.2006.00210.x>
- Wang, R. W., Lee, C.-S., & Haber, J. E. (2017). Position effects influencing intrachromosomal repair of a double-strand break in budding yeast. *PLoS ONE*, 12, e0180994. <https://doi.org/10.1371/journal.pone.0180994>
- Warwick, S. I., Legere, A., Simard, M.-J., & James, T. (2008). Do escaped transgenes persist in nature? The case of an herbicide resistance transgene in a weedy *Brassica rapa* population. *Molecular Ecology*, 17, 1387–1395. <https://doi.org/10.1111/j.1365-294X.2007.03567.x>
- Watson, H. W., & Galton, F. (1875). On the probability of the extinction of families. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 4, 138–144.
- Webber, B. L., Raghun, S., & Edwards, O. R. (2015). Opinion: Is CRISPR-based gene drive a biocontrol silver bullet or global conservation threat? *Proceedings of the National Academy of Sciences*, 112, 10565–10567. <https://doi.org/10.1073/pnas.1514258112>
- Webster, S. H., Vella, M. R., & Scott, M. J. (2019). Development and testing of a novel Killer-Rescue self-limiting gene drive system in *Drosophila melanogaster*. *bioRxiv*, 680629. <https://doi.org/10.1101/680629>
- Weill, M., Chandre, F., Brengues, C., Manguin, S., Akogbeto, M., Pasteur, N., ... Raymond, M. (2000). The *kdr* mutation occurs in the Mopti form of *Anopheles gambiae* s.s. through introgression. *Insect Molecular Biology*, 9, 451–455. <https://doi.org/10.1046/j.1365-2583.2000.00206.x>
- Whitney, K. D., Ahern, J. R., Campbell, L. G., Albert, L. P., & King, M. S. (2010). Patterns of hybridization in plants. *Perspectives in Plant Ecology, Evolution and Systematics*, 12, 175–182. <https://doi.org/10.1016/j.ppees.2010.02.002>
- Wienert, B., Shin, J., Zelin, E., Pestal, K., & Corn, J. E. (2018). In vitro-transcribed guide RNAs trigger an innate immune response via the RIG-I pathway. *PLoS Biology*, 16, e2005840. <https://doi.org/10.1371/journal.pbio.2005840>
- Wierson, W. A., Welker, J. M., Almeida, M. P., Mann, C. M., Webster, D. A., Weiss, T. J., ... Ming, Z. (2018). GeneWeld: A method for efficient targeted integration directed by short homology. *bioRxiv*, 431627. <https://doi.org/10.1101/431627>
- Wittkopp, P. J., & Kalay, G. (2012). Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics*, 13, 59. <https://doi.org/10.1038/nrg3095>

- Xu, W., & Eiden, M. V. (2015). Koala retroviruses: Evolution and disease dynamics. *Annual Review of Virology*, 2, 119–134. <https://doi.org/10.1146/annurev-virology-100114-055056>
- Yoshimura, Y., Beckie, H. J., & Matsuo, K. (2006). Transgenic oilseed rape along transportation routes and port of Vancouver in western Canada. *Environmental Biosafety Research*, 5, 67–75. <https://doi.org/10.1051/ebr:2006019>
- Yuan, X., Song, M., Devine, P., Bruneau, B. G., Scott, I. C., & Wilson, M. D. (2018). Heart enhancers with deeply conserved regulatory activity are established early in zebrafish development. *Nature Communications*, 9, 4977. <https://doi.org/10.1038/s41467-018-07451-z>
- Zapiola, M. L., & Mallory-Smith, C. A. (2012). Crossing the divide: Gene flow produces intergeneric hybrid in feral transgenic creeping bentgrass population. *Molecular Ecology*, 21, 4672–4680. <https://doi.org/10.1111/j.1365-294X.2012.05627.x>
- Zhang, C.-J., Yook, M.-J., Park, H.-R., Lim, S.-H., Kim, J.-W., Nah, G., ... Kim, D.-S. (2018). Assessment of potential environmental risks of transgene flow in smallholder farming systems in Asia: *Brassica napus* as a case study in Korea. *Science of the Total Environment*, 640, 688–695. <https://doi.org/10.1016/j.scitotenv.2018.05.335>
- Zhang, F. (2019). Development of CRISPR-Cas systems for genome editing and beyond. *Quarterly Reviews of Biophysics*, 52, E6. <https://doi.org/10.1017/S0033583519000052>
- Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S., & Yang, S.-H. (2015). Off-target effects in CRISPR/Cas9-mediated genome engineering. *Molecular Therapy - Nucleic Acids*, 4, e264. <https://doi.org/10.1038/mtna.2015.37>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Courtier-Orgogozo V, Danchin A, Gouyon P-H, Boëte C. Evaluating the probability of CRISPR-based gene drive contaminating another species. *Evol Appl*. 2020;13:1888–1905. <https://doi.org/10.1111/eva.12939>