



HAL
open science

Mathematical models of gene drive for population management

Léna Klay

► **To cite this version:**

Léna Klay. Mathematical models of gene drive for population management. Analysis of PDEs [math.AP]. Sorbonne Université, 2023. English. NNT : 2023SORUS338 . tel-04334913

HAL Id: tel-04334913

<https://theses.hal.science/tel-04334913>

Submitted on 11 Dec 2023

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.

THÈSE

en vue de l'obtention du grade de

Docteur en Sciences de l'évolution
délivré par Sorbonne Université

Discipline : Biologie des populations et écologie

Institut d'écologie et des sciences de l'environnement de Paris

École doctorale 227 : Sciences de la nature et de l'Homme

Présentée et soutenue publiquement le 15 novembre 2023
par **Léna KLÄY**

Modèles mathématiques de forçage génétique pour la gestion de populations

Devant le jury composé de :

Vincent CALVEZ	<i>CNRS, Université de Bretagne Occidentale</i>	Directeur de thèse
Florence DÉBARRE	<i>CNRS, Sorbonne Université</i>	Directrice de thèse
Léo GIRARDIN	<i>CNRS, Université Lyon 1</i>	Membre invité
Eric MAROIS	<i>Inserm, IBMC</i>	Examinateur
Emmanuelle PORCHER	<i>MNHN</i>	Présidente du jury
Lionel ROQUES	<i>INRAE Avignon</i>	Rapporteur
Amandine VÉBER	<i>CNRS, Université de Paris</i>	Examinatrice
Elodie VERCKEN	<i>INRAE, Institut Sophia Agrobiotech</i>	Rapportrice

Institut d'écologie et des sciences de
l'environnement de Paris
Campus Pierre et Marie Curie,
Barre 44 – 45, 4e étage – CC 237,
4 place Jussieu, 75005 PARIS

École doctorale 227 : Sciences de la nature
et de l'Homme
Muséum national d'Histoire naturelle,
57 rue Cuvier - CP 27, 75005 Paris

*Je dédie cette thèse à mes parents, mélangeant maths et
anglais à l'image de ce que vous m'avez transmis.
Je vous aime fort.*

*Je te dédie également cette thèse Antoine. Je sais que
tu n'en as que faire, il manque certainement quelques
dates et une barrière pour attiser ton intérêt... mais j'ai
réussi à placer une frise chronologique et à écrire un
chapitre entier sur le fait de stopper une invasion, tu
devrais jeter un oeil, nos deux matières ne sont pas
aussi éloignées que tu ne le crois !*

Remerciements

Tout d'abord je tiens à remercier mes deux directeurs de thèse, Florence Débarre et Vincent Calvez, non pas parce que cela est d'usage, mais bien parce qu'ils ont été des encadrants exceptionnels tant d'un point de vue humain que scientifique. Merci Florence d'avoir été toujours si présente et disponible (malgré une activité covid bien chargée !), attentive à mon travail comme à mon moral, et bienveillante sur tout ce qui tenait de l'extra thèse ! Merci Vincent pour ton incroyable enthousiasme dans notre travail, c'est si agréable d'avancer dans de telles conditions, merci pour tous ces échanges mathématiques, ces tableaux noirs remplis d'équations, mais aussi ces conversations anodines remplies de lectures et d'anecdotes ! Merci également à Léo Girardin, d'avoir été si présent et d'une grande aide sur ces fameux tableaux, c'était un plaisir de travailler ensemble !

Je remercie également les membres de mon jury de thèse, Éric Marois, Emmanuelle Porcher, Lionel Roques, Amandine Véber et Elodie Vercken, qui ont pris du temps sur leur travail de chercheur et chercheuse pour lire ce long manuscrit et rien que pour cela, je leur en suis extrêmement reconnaissante ! Merci aux membres de mon comité de suivi, Nicolas Rode, Natacha Kremer et Isabelle Dajoz pour toutes vos remarques pertinentes sur mes travaux et la direction que prenait cette thèse. I would also like to thank Sarah P. Otto and Frederik J.H. de Haas for the rich and very interesting work we carried together and your great enthusiasm in every step of it.

Merci à toutes les personnes de l'IEES-Paris et plus particulièrement l'équipe EERI. Merci à Carole Bousquet, Johanna Azzi, Véronique Marciat, Paola Paradisi et Julie Legoupi d'avoir été si présentes et d'une efficacité redoutable pour tout ce qui concernait les démarches administratives qui d'habitude font très peur... Merci Julien pour ta gentillesse à toute épreuve et ton soutien dans les méandres d'ADUM. Merci également à Nathalie Machon d'avoir su me prêter une oreille attentive lors d'une formation de l'école doctorale, elle me fut d'une grande aide. Merci à Martine Maïbèche et Emma Rochelle-Newall, ancienne et nouvelle directrices de l'IEES-Paris, qui soutiennent toujours les projets des jeunes chercheurs. Je remercie enfin les chercheurs et chercheuses de l'Institut Camille Jordan pour leur accueil toujours chaleureux lorsque je séjournais à Lyon.

Merci aux jeunes chercheurs qui m'ont accompagnée dans cette grande aventure qu'est la thèse, pour cette ambiance décontractée et joyeuse qui me faisait oublier le temps d'une pause que mon code qui ne fonctionnait plus ! Merci pour toutes ces discussions, nos repas au CROUS, les goûters à la cafet' et les pots du vendredi sur les quais de Seine. Merci Arthur pour m'avoir sauvé la mise plus d'une fois et pour ta bonne humeur constante, au labo comme à la coloc, Alice pour avoir été une super partenaire de sensations fortes, Basile et Vincent pour vos conseils variés et toujours très au fait ! Merci Antoine pour nos chouettes discussions et ton soutien aéronautique, Fatima pour ta pêche et ta gentillesse inconditionnelle, Samuel pour tes énigmes qui nous ont fait bouger jusqu'au jardin des plantes, Chloé, Thomas, Souleyman et Amandine pour avoir partagé de jolies soirées ! Un grand merci également à Paul, Hajar, Nicolas, Lauren, Youssef, Marie B., Fatima E., Marie G., Aurélie, Théo B., Théo V., Elsa et bien d'autres que j'oublie sûrement !

Merci aux enseignants qui ont su attiser ma curiosité tout au long de ma scolarité, je remercie en particulier M. Millour pour sa joie de vivre et ses blagues, le projet algues vertes fut mon premier pas dans le monde de la recherche. Merci à Mme Bellec pour ce dynamisme et ces cours de maths qui resteront mémorables ! J'en oublie tellement, merci à tous ceux qui de près ou de loin ont rendu ces années d'étude agréables et instructives.

Je souhaite également remercier toutes les personnes qui ont égaillé mon quotidien au cours de ces trois belles années de doctorat, à Paris ou lors de mes différentes vadrouilles. Merci à Alan, Alix, Anna, Annaëlle, Arthur, Edern, Elena, Hermine, Hugo, Juliette, Léo, Louise, Madeleine, Marieke, Orlanne, Phong, Samuel, Sylvia et Tanguy pour nos retrouvailles toujours joyeuses et nos nouvel ans aux ambiances roscovites. Merci Victor, Olivier et Luna pour ces belles soirées à Cité U ou ailleurs, Claire pour ton franc parlé que rien n'égale, Erwan pour ces nouvelles aventures en perspective alors qu'elles auraient du se cantonner à une formation doctorale. Merci Yuna, Théo et Juliette pour nos baignades dans les eaux de Norvège et nos fous rires, Évaël pour nos lectures et aventures océaniques, Ariane, Anne-Laure, Jean-Baptiste, Louise, Emmanuel, Mathieu, Raph, Damien, Carla et tant d'autres pour un camping au bord d'une rivière ensoleillée. Merci Barnabé pour la découverte des multiples recoins et détours de Paris en musique, Jason pour nos chemins emmêlés qui se sont croisés et recroisés de milles façons ! Merci les filles, Yasmin, Louise et Amalia pour ces étés en bivouac dans les alpages, Manu pour tes blagues et nos expéditions dans les calanques, dans la neige, à l'UICN... Merci à la famille Breysse pour votre superbe accueil à Belle-Ile, Antoine pour ce bonheur et nos rires quotidiens ainsi que ton soutien infailible en cette fin de thèse, Morgane pour avoir été une amie et alliée. Thanks Diane and Jo for this amazing adventure abroad, for the tropical fruits, the paddle session and the colorful markets. Merci Alexis pour la superbe vue depuis la terrasse du Collège de France, et Rémi pour quelques belles prises de Jujitsu ! Merci à Gaston et Janine, Murielle, Jean Marie, Henriette, Caro, Bowie et tant d'autres pour votre accueil toujours si chaleureux dans les montagnes !

Enfin je remercie de tout coeur mes parents. Ma mère, lumineuse et rieuse, si pleine d'idées parfois farfelues mais toujours si drôles, avec qui jamais l'on ne s'ennuie. Et mon papa, la force tranquille, la discussion facile, que rien ni personne n'ébranle. Si ce n'est peut-être la soupe du lundi et du mardi. Merci à tous les deux de m'avoir fait grandir libre et confiante, heureuse. Et puisque j'ai décidé d'élargir la famille aux amies de longue date, je souhaite également remercier Maëllou, pour un paquet de choses mais rappelons la plus importante : pour avoir viré en maternelle un pauvre gamin (qui n'avait rien demandé) de son vélo à roulettes afin que je puisse en avoir un aussi... merci de m'avoir fait tellement rire durant toutes ces belles années ! Merci Kezya d'avoir partagé une bonne partie de mon enfance, et pour ton soutien inconditionnel lorsqu'il fallait organiser une mutinerie en chaussons dans le salon ! Et enfin merci Laurane et Ines pour toutes nos batailles d'eau, sorties rollers, kayak, soirées pyjama... c'est toujours un plaisir de vous retrouver après toutes ces années !

Contents

1	Introduction	1
1.1	Foreword	1
1.2	Biological background	2
1.2.1	Motivation and definition of gene drive	2
1.2.2	Two engineering approaches to bias transmission	3
1.2.3	A short chronology	6
1.2.4	Concrete and speculative applications	7
1.2.5	Risks with a focus on CRISPR-based homing drives	10
1.2.6	Technological fix criticisms	10
1.2.7	Moral considerations	11
1.2.8	How to achieve confinement?	11
1.3	Mathematical and numerical analysis	17
1.3.1	Introduction to gene drive modelling	17
1.3.2	Main results of this thesis	23
1.3.3	Perspectives	42
2	Pulled, pushed or failed: the demographic impact of a gene drive can change the nature of its spatial spread	45
2.1	Introduction	46
2.2	Methodology	47
2.2.1	Models	47
2.2.2	Setting of the problem	50
2.2.3	Glossary	51
2.3	Results	52
2.3.1	Model with perfect conversion in the zygote	52
2.3.2	Models with partial conversion	56

2.4	Discussion	67
2.5	Acknowledgements	68
	Appendices	69
2.A	Model with partial conversion: growth term details	69
2.B	System rewritten with variables (n, p_D)	71
2.C	Proofs for model (2.12) with perfect conversion in the zygote	74
2.D	Critical traveling wave for an SI similar model.	77
2.E	Study of the reaction term when $r = +\infty$ in section 2.3.2	83
2.F	Heatmap supplementary materials	91
3	The speed of advance of a gene drive is affected by density dependence	95
3.1	Introduction	96
3.2	Models and methods	98
3.2.1	Models	98
3.2.2	Traveling waves	100
3.3	Results	101
3.3.1	Demography and dominance can affect the final allelic proportions	101
3.3.2	The Allee effect makes the eradication easier and reduces the final density in case of drive persistence	103
3.3.3	A density-dependence constraint on the deaths instead of the births results in a faster invasion.	105
3.3.4	The Allee effect might cause the failure of threshold dependent drives	107
3.4	Discussion	108
	Appendices	111
3.A	Allee effect	111
3.B	Models	112
3.C	Final density after a drive invasion	116
4	Stochastic dynamics at the back of gene drive propagation: drive eradication, wild-type recolonisation or chasing	119
4.1	Introduction	120
4.2	Models and methods	122
4.2.1	Continuous deterministic model	122
4.2.2	Discrete stochastic model	124

4.2.3	Parameters and initial conditions for numerical simulations	127
4.2.4	Setting of the problem	128
4.3	Results	131
4.3.1	Determining the almost deterministic distance between the drive and wild-type wave at a level line of 100 individuals	131
4.3.2	Characterising the stochastic fluctuations of the last wild-type individual	132
4.3.3	Numerical conclusions	136
4.3.4	Relation between drive intrinsic fitness, diffusion and risk of wild-type recolonisation.	137
4.4	Discussion	138
Appendices		141
4.A	Speed and exponential approximations of the wave	141
4.B	Implementation of the stochastic model	146
5	Reducing risk of spillover using daisy quorum drive	147
5.1	Abstract	147
5.2	Introduction	148
5.3	Methods and results	151
5.3.1	Drive design	151
5.3.2	Dynamics in an isolated population	152
5.3.3	Invasion analysis	155
5.3.4	Spatial spread of daisy quorum drive in a discrete environment	156
5.3.5	Spatial spread of daisy quorum drive in a continuous environment	158
5.4	Discussion	160
5.5	Acknowledgements	161
Appendices		163
5.A	Daisy quorum drive	163
5.B	Individual-based simulations with discrete patches	168
5.C	Numerical simulations in continuous space	171

Chapter 1

Introduction

1.1 Foreword

This PhD thesis deals with a new genetic engineering technology from which arises a multitude of fascinating questions on possible applications, conditions for a successful use, risks, ethics, regulation and governance... The topic elevates numerous and complex debates, and it is important to develop a clear understanding of the potential outcomes of drive release in nature. Across these pages, I study the spatial spread of gene drive through the analysis of mathematical models, answering some of the scientific questions on the subject. However, it is important to remain aware of the limitations of this approach: a model is only a projection and a simplification of the reality as we understand it, not reality itself.

People often tell me: "Modelling is so complex!"

To which I usually answer: "Models are a simplification of reality: what's complex is the real world!"

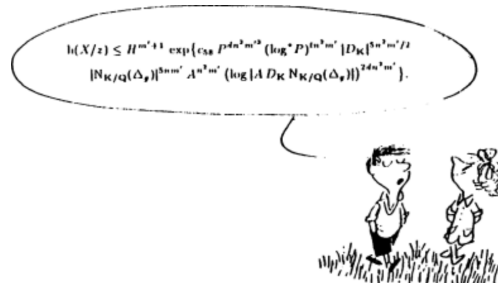


Figure 1.1: From "Little Nicholas, PhD Student" [English] [French].
Drawing: J.J. Sempé, Formulas: Y. Bugeaud - M. Mignotte - F. Normandin,
Text: G. Taviot, Mise en page: G. Taviot, A. Maes

1.2 Biological background

1.2.1 Motivation and definition of gene drive

Some humanity's most pressing public health, ecological and agricultural problems result from interactions with other species of the ecosystem, such as disease vectors, crop pests, invasive animals and plants... For centuries, humans have sought to reduce these "pest populations"* through a variety of methods: chemical control (pesticides and poisons), physical control (nets, barriers, traps...) and biological control (introduction of a natural predator or parasite in the environment). Genome edition has also been considered to render these populations harmless or cause extinction. However, this idea remained technically impossible until 2012, when the field underwent a technical revolution. The successful use of the CRISPR-Cas9 system in-vivo provided a cheap and efficient tool to introduce a trait of interest in the genome of one individual.

However this tool alone is not sufficient to alter wild sexually reproducing populations. Mendelian inheritance (ensuring genetic mixing) combined with natural selection guarantees that an allele fails to spread to fixation in large populations unless it offers a fitness benefit. To bypass this mechanism, recent studies have focused on gene drives: genetic elements biasing their inheritance towards a super-Mendelian rate, therefore driving themselves to spread faster through a wild population [34]. Super-Mendelian inheritance can either be symbiont based (such as *Wolbachia* [23, 75, 117]) or genome engineering based. In this thesis, I focus on the genome engineering based approach.

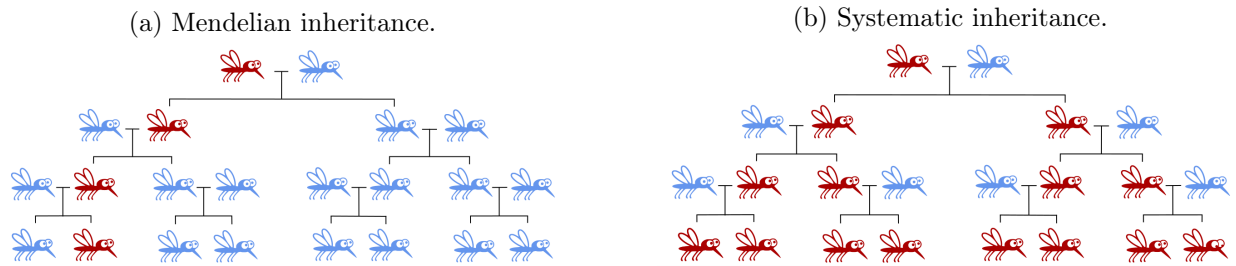


Figure 1.2: Comparison between a Mendelian and a super-Mendelian (here systematic) inheritance. Blue mosquitoes represent wild-type individuals while red mosquitoes represent gene drive individuals (at least one drive allele in the genome). In the case of Mendelian inheritance (a), a heterozygous individual has a one-half chance of transmitting the altered copy to its offspring, whereas with super-Mendelian inheritance (b), the mutation is much more often transmitted to the descendants, here for the illustration, systematically.

A gene drive allele, thanks to its super-Mendelian inheritance, has the potential to spread through a wild population and fully replace the original trait with limited impact on the population size (*replacement drive*) [34]. If the gene drive allele induces an evolutionary disadvantage for the individual carrying it, for example if it alters an essential fertility or viability gene, its super-Mendelian spread can lead to the complete extinction of the population (*eradication drive*) [34, 200, 133, 104]. If the drive causes a significant decrease in population size without extinction, it is a *suppression drive*.

*Note that the concept of pest species is scientifically outdated, as it is based on a utilitarian vision of nature that invisibilises the complex interactions between animals [144].

1.2.2 Two engineering approaches to bias transmission

Gene drive constructs are usually composed of a cargo (a gene of interest for population replacement, or an inhibitory cargo with a fitness cost for population suppression/eradication), and a mechanism that increases the transmission rate of the whole construct to offspring. There are two strategies to increase the transmission rate: either to convert non-modified alleles into drive ones (*homing drives*), or to disable non-drive alleles resulting in a relatively higher transmission of the former ones (*toxin-antidote drives*).

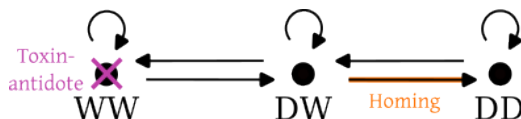


Figure 1.3: Strategies to increase the transmission of alleles D (drive) to the detriment of alleles W (wild-type). The black arrows represent the possible genotypes a parent can give birth to. Homing drives convert wild-type alleles into drive alleles while toxin-antidote drives disable wild-type gametes or offspring.

Homing drives: increasing the number of drive alleles

The first strategy consists in converting heterozygous cells into drive homozygous cells (in the germline or in the new zygote), to increase the number of drive alleles transmitted to offspring. These drives are called *homing gene drives*. In a heterozygous cell, the drive placed in the middle of its own recognition sequence induces a cut in the homologous chromosome. If the damage is repaired by homology direct-repair (HDR), it results in the copy of the drive sequence into the second chromosome. The previously heterozygous cell is now drive homozygous. This duplication repeats through generations and largely benefits the drive propagation.

Another possible repair pathway in the cell is non-homologous end joining (NHEJ), which consists in directly ligating the break ends. Consequently, the success of the homing gene drive technique relies on the fact that double-strand breaks will preferably be repaired by homology direct-repair rather than non-homologous end joining. This affirmation is complex and depends on the cell type, the developmental stage, the species and the phase of the cell cycle, which may require in-depth preliminary studies [81].

Theoretically, gene conversion is possible in the germ cells of the parents or in the zygote cells of the offspring, but, in practice, only the first solution has been successfully implemented. Up to now, scientists have failed to convert heterozygous zygotes into drive homozygous zygotes [43].

Homing drive dynamics were first modelled by Burt, who considered the spread of homing endonuclease genes (HEGs) [34]. HEGs naturally bias their transmission rate and have been observed to spread in fungi, plants, and bacteria [96, 51]. There have been some attempts to engineer similar mechanisms in insects and vertebrates with zinc finger or TALE nucleases [199], but CRISPR-based nucleases, discovered in 2012, proved to be the easiest and most efficient way to construct homing gene drives [93, 91, 103].

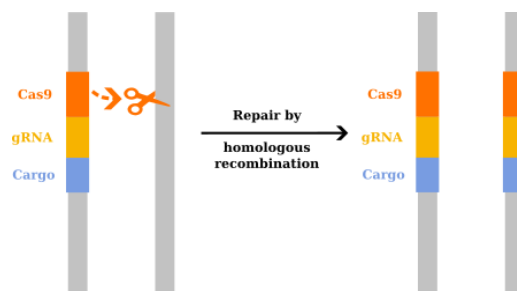


Figure 1.4: CRISPR-Cas9 homing gene drive, studied in Chapters 2, 3 and 4.

A current obstacle to CRISPR homing drives is their propensity to generate resistance alleles. Natural variation exists in every population: such resistance alleles can arise from de novo mutations in wild-type alleles, but they can also be produced by the drive itself when cleavage is repaired by non-homologous end joining (or rarely, incomplete homology direct-repair) instead of successful homology direct-repair [183, 212, 106].

Toxin-antidote drives: reducing the number of wild-type alleles

In contrast to homing drives, which spread by directly increasing the number of drive alleles, toxin-antidote drives spread by reducing the number of wild-type alleles over time (and therefore increasing the relative frequency of the drive). These drives are composed of two elements: i) a toxin targeting an essential gene by either blocking its expression or damaging its DNA sequence and ii) a corresponding antidote providing a recoded version of the essential gene (no more matching the toxin). A lack of expression from the essential gene in the gametes or in the zygote might be lethal, unless the drive allele provides the rescuing antidote.

In a heterozygous sexually reproducing individual, meiosis separates the drive from the wild-type allele therefore the antidote will only be passed through half of the gametes, while the toxin can be active before or after the gamete formation. If selection occurs among the gametes, the toxin is active prior to the gamete formation. It disrupts the wild-type allele in the heterozygous germline cells and meiosis results in disrupted wild-type gametes which are non-viable as they lack the antidote (contained in the drive allele). Selection can also occur among the zygotes and then the toxin can be active prior to the gamete formation, after the gamete formation (in the egg), or both. If the toxin is active after the gamete formation, it is usually deposited in the egg by one of the parents such as in *Medea systems* (maternal microRNA-mediated silencing) [50], or in *Toxin-Antidote Recessive Embryo systems* (maternally deposited Cas9) [44]. Regarding the genotype of the mating partner and the toxin timing, zygotes containing the toxin without its corresponding antidote are non-viable. There exists subtleties if the essential gene targeted by the toxin is haplosufficient (one functional copy of the gene is required for viability), or haplolethal (two functional copies of the gene are required for viability) [41, 44].

Toxin-antidote systems are widely distributed in prokaryotes and classified into eight different classes [122]. Several of them exhibit activity in yeast and mammalian cells, which makes their application possible in eukaryotic systems [214, 215]. Several designs have been suggested, based on *Medea* toxin-antidote systems / UD^{MEL} [50, 2, 221, 31], *Wolbachia* toxin-antidote elements [197, 112, 113, 161], or RPM-Drive [185, 7, 6, 184]. However, their need for highly specific targets, promoters and RNAi elements, make them difficult to engineer and adapt to various species. To overcome this

problem, toxin-antidote systems can also be designed using the CRISPR-Cas9 complex (*Cleave and rescue systems*). The toxin, a Cas9 with the adapted gRNA, is programmed to cut an essential gene while its corresponding antidote is a copy of the gene, recoded so it does not match the toxin. The targeted essential gene can be on the same pair of chromosomes than the toxin-antidote complex, or on a different one (see Figure 1.5).

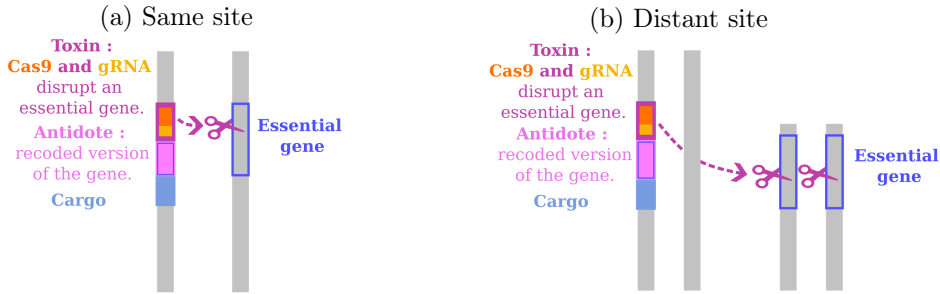


Figure 1.5: Cleave and rescue systems.

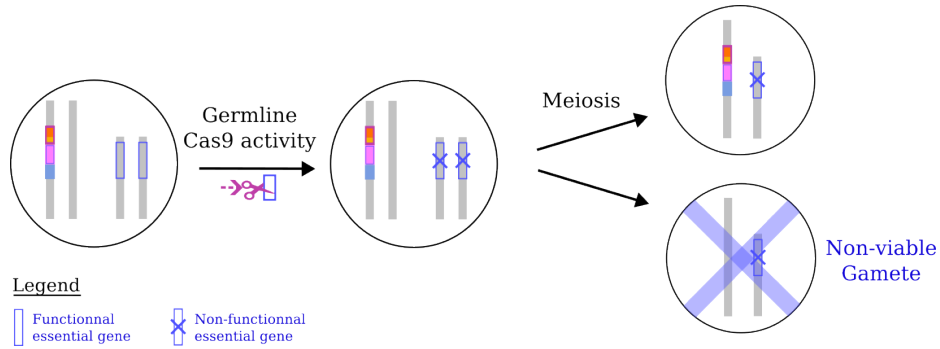


Figure 1.6: Illustration of a Cleave and rescue system (distant site) with a Cas9 toxin being active before the gamete formation and a selection occurring in the gametes. A gamete with a non-functional essential gene and no antidote is non-viable, consequently, only the drive alleles of a heterozygous cell are transmitted to the offspring.

The *X-shredder* drive is a particular case of cleave and rescue system with selection occurring in the gametes. A toxin is introduced in the Y-chromosome and this toxin shreds the X-chromosome in heterozygous (XY) male cells prior to the gametes formation, resulting in the non-viable X-gametes. The Y-chromosome is the drive component: the goal of such construct is to enhance the transmission of the Y-chromosome to offspring in order to bias the population sex ratio (more male XY than female XX). Among the gametes produced by heterozygous (XY) male cells, only the Y-gametes are viable and transmitted to offspring, resulting in male progeny. We observe naturally occurring *X-shredder* in *Aedes aegypti* and *Culex quinquefasciatus* mosquitoes [63, 222]. Efforts have been put to introduce *X-shredder* drives in *Anopheles gambiae*, an important vector of malaria, to reduce the mosquito population by male-biasing the sex ratio (males do not bite) [90, 89]. However, expressing the Cas9 toxin in the Y chromosome is difficult probably because of meiotic sex chromosome inactivation (MSCI) [24, 211]. The toxin has successfully been inserted and expressed on autosomal chromosomes, but a driving Y chromosome has not yet been developed, although some progress has been made in that direction [21, 102].

Combinations of homing and toxin-antidote drives

Possible combinations between the two strategies are possible, for instance X-shredder homing gene drives [198] or haplolethal homing drive mechanisms [40]. Should one system fail, the other will function (homing is especially sensitive to resistance [183, 212, 45]), and the combined construct is expected to spread more efficiently than either of them alone.

1.2.3 A short chronology

Since the first agrarian societies, humans have intentionally genetically modified their environment through selection process, plant breeding, and more recently mutagenesis. In 1953, the discovery of the DNA molecule came as a breakthrough and, with it, the idea to modify living organisms by editing their genome [217]. Interestingly selfish elements (or *gene drive*, elements that bias their own inheritance) were observed before the discovery of DNA molecule, in 1928 [94].

The first genetically modified organism (GMO) was developed in 1973 [54] with restriction endonucleases and DNA ligase. Since then, different families of engineered nucleases have been used to create GMOs: engineered meganuclease (MegaN) discovered in 1985 [120], Zinc finger nucleases (ZFNs) also discovered in 1985 [156], transcription activator-like effector nucleases (TALENs) discovered in 2010 [52].

In parallel in the second half of the 20th century, natural homing endonucleases drew the attention of scientists [163], but they could not alter its specificity to other DNA sequences than the one initially targeted. In 2003, Austin Burt proposed the first gene drive theoretical model based on natural homing endonuclease [34].

In 2012 however, the clustered regularly interspaced short palindromic repeat/CRISPR-associated protein 9 (CRISPR-Cas9) nuclease system was shown as being applicable for genome editing [121]. The efforts to insert, delete or modify genetic material into specific spots on chromosomes got a major boost with the use CRISPR-Cas9 *in vivo*: it opened up an incredible number of new perspectives to genetically modify organisms, and in particular, create gene drive organisms. In 2015, the first gene drive fly was produced in the laboratory [93]. In that same year, scientists were also able to propagate a gene for resistance to the malaria parasite through a mosquito population [91] and another gene drive was developed to reduce female mosquito fertility with the motivation to eradicate major malaria vector species [103].

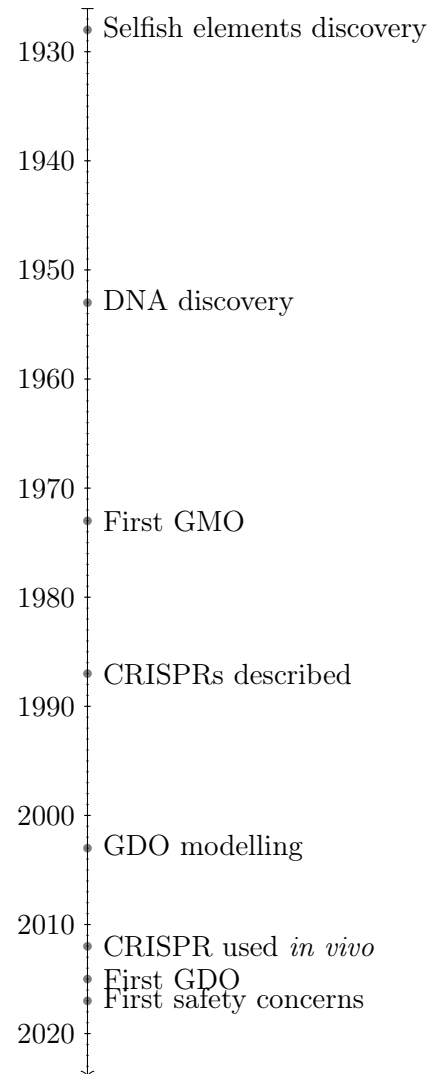


Figure 1.7: Gene drive timeline, where DNA stands for Deoxyribonucleic acid, GMO for Genetically Modified Organisms and GDO for Gene Drive Organisms.

Despite the numerous promising applications of gene drive, the trivialisation of this powerful technique also raised concerns [61]. In 2016, 160 civil society organisations called for a global moratorium on the development of the gene drive technology to have time for assessing environmental and health risks, establishing missing regulations and debating publicly [83]. This moratorium was rejected that same year by the UN Convention on Biological Diversity [36].

1.2.4 Concrete and speculative applications

Gene drives offer the possibility to genetically modify any wild population at the condition that it is diploid, sexually reproducing, and preferentially with short generation times. The possible applications of this technology are numerous and mainly fall into three categories: public health, conservation and agriculture.

Public health

During the last two decades, the world has witnessed the resurgence and the expansion of vector-borne diseases. Rapid urbanisation and long-distance travel and trade have brought humans into more frequent contact with vectors, while climate change is responsible for disease spread in new regions, especially in temperate areas [174].

Physical control such as bed nets is efficient but offers only a limited protection (at night), while insecticide-based control strategies come with important drawbacks. The incidence of resistance to insecticides has rapidly increased in recent years leading to the long term failure of the method [158], some repellents have been shown to be toxic to human health [196] and pesticides can also affect non-target insects such as pollinators [216].

Gene drive might represent an alternative solution to fight vector-borne diseases, although the safest and most effective strategy might certainly be to coordinate gene drive with existing control techniques [218]. Two main approaches are considered: the spread of an resistance allele to the pathogen in the vector species (*replacement drive*) or the reduction of the population size of the vector species (*suppression or eradication drives*) [200, 24]. This approach could theoretically target the major vector-borne diseases of humans (malaria, dengue, lymphatic filariasis, Chagas disease, onchocerciasis, leishmaniasis, chikungunya, Zika virus disease, yellow fever, Japanese encephalitis and schistosomiasis) and other local important vector-borne diseases such as tick-borne diseases [174]. Today ongoing research mainly focuses on two diseases, malaria and dengue, I present three non exhaustive projects below.

A first project, Target Malaria, works on suppression drives in *Anopheles gambiae* mosquito, the first African malaria vector. Two approaches were considered in this project: X-shredder and homing gene drive. Despite preliminary promising results [90, 89], the X-shredder linked to the Y chromosome approach encountered difficulties, most likely due to meiotic sex chromosome inactivation (MSCI) [24, 211]. The homing gene drive approach however, creating sterile females by altering a highly conserved sequence in the *doublesex* gene, was successfully experimented in laboratory: 150 modified individuals introduced in a cage of 600 wild-type mosquitoes led the whole population to extinction in 8 generations in the first experiment, and in 12 generations in the second [133]. This homing gene drive has also been successfully tested in large indoor cages which allow complex feeding and reproductive behaviours [105]. This drive must now fulfil risk assessment studies [57, 58] and regulation conditions before a potential field release.

Another research team at the University of California in San Diego also works on Malaria, but on replacement drive. This project is trying to introduce a resistance gene in *Anopheles* mosquitoes that would kill the malaria pathogen [91]. However, primary laboratory experiments have proven limited viability of those mosquitoes in cages [179].

A third project also at the University of California in San Diego focuses on dengue. Scientists have engineered mosquitoes that express an antibody against all four major strains of dengue, and are thinking of combining this ability with a gene drive construct [30]. One ambitious and speculative idea for the future would be to build an all-purpose gene that would release a toxin when any virus infects the mosquito (dengue, yellow fever, chikungunya, zika...) [195].

Conservation

Biodiversity is declining at unprecedented rates and across ecological scales [73]. Among the biggest threat comes the reshaping of natural habitats by humans (farming, urbanisation, extracting resources...), alien invasive species (species introduced outside their natural ranges and out-competing native species for food, water and space [19, 182]) and climate change. These causes are often intertwined: for instance, globalisation and climate change have led to a rise of alien invasive species (facilitating their spread and establishment) while these alien species can reduce the resilience of natural habitats to climate change [119]. Some species are also impacted by resurgence and expansion of vector-borne diseases.

Gene drive could help conserve biodiversity in several ways: remove invasive alien species that contribute to biodiversity loss, spread resilience in keystone species threatened by climate change or other anthropogenic pressures, or fight diseases targeting endangered species... If such conservative applications spark interest [180, 81], they also are a source of significant concern [203, 82].

A few projects aiming to target alien species with gene drive emerged in 2016, just after the first gene drive organism was engineered in laboratory. The Genetic Biocontrol of Invasive Rodents (GBIRD) program investigates how to eradicate invasive rodents such as mice through gene drives, especially on islands [98, 37] and the program "Predator Free 2050" in New Zealand aims to eliminate rats, possums and stoats in New Zealand by 2050 with the potential use of gene drive [172, 68]. Even if islands represent only 6.7% of Earth's landmass, their conservation is of main importance as they harbor approximately 20% of the Earth's biodiversity [86]. Rodents have been introduced by humans on numerous islands and are now responsible for many endemic species extinctions [207]. Toxin-antidote drives biasing the sex ratio to eradicate rodent populations have been proposed, such as the t-Sry system producing "daughterless mice" [135, 100, 99, 38].

Other gene drives applications in conservation remain speculative or at an early stage of technical development (non exhaustive list). In Hawai'i, the avian malaria threatens endangered native birds and could be controlled by similar gene drive approaches to the ones developed in the health section to eradicate diseases [187, 162]. In New Zealand, scientists are considering a gene drive based on spermatogenesis knockdown to eradicate the common wasp *Vespula vulgaris* [137]. In the United Kingdom, proposition of a localised suppression drive have been illustrated with the case of invasive grey squirrels, contributing to the decline of native red squirrels and damaging the forest [85].

Scientists have also found a mutation inducing heat tolerance in corals and there is hope that gene drive could help spreading this mutation in coral populations to cope with climate change [53, 108].

Agriculture

Last but not least, gene drives could be used in agriculture, a sector also impacted by climate change, diseases and crop pests (old and new). The foundational patent application on RNA-guided gene drives by Kevin Esvelt lists more than 180 agricultural weed species as well as 160 animal pest species relevant to agriculture. A similarly foundational gene drive patent application by Ethan Bier and Valentino Gantz lists more than 600 agricultural pests [84] (detailed species in [111]). Gene drive could be used to eliminate these pest species [194, 164] or, more likely in the current business model, to eliminate resistance to pesticides or other repellents produced by the agrochemical firms [111, 154, 164]. Finally, gene drive could help increase stress tolerance in plants, in response to climate change [187].

The field of synthetic gene drives in plants is very young: the first gene drive construct with the explicit goal of modifying wild plant populations was successfully engineered in 2023 [141]. Non-homologous end joining (NHEJ) is the preferential repair pathway used in plants [150] which poses a significant challenge to the development of homing-based gene drives. To bypass this difficulty, scientists have engineered a Cleave and Rescue system: a Cas9-toxin is designed to cut an essential gene for pollen germination (NPG1), while a recoded CRISPR-resistant NPG1 is the antidote [141].

Replacement, suppression and eradication drives

Table 1.1 and 1.2 summarise the main possible applications of gene drive. They are categorised into replacement drives (spreading a trait of interest in a population without significantly affecting its size) or suppression/eradication drives (spreading a fecundity or survival cost leading to the reduction of the size of the population or to its eradication).

Replacement drives

Introduce a specific function inside a population. This modified population will persist over time.

Public health	Conservation	Agriculture
★ Modify disease vector species so they become resistant to a pathogen [91, 30].	★ Help endangered species to adapt to changing environments [81, 187]	★ Eliminate resistances to pesticides [111, 154, 164]

Table 1.1: Main applications of replacement drives in health, conservation and agriculture sectors.

Suppression and Eradication drives

Reduce or eradicate an undesirable population.

Public health	Conservation	Agriculture
★ Reduce or eradicate disease vector species [133, 105]	★ Reduce or eradicate disease vector species [187, 108] ★ Reduce or eradicate invasive alien species [85, 137, 135]	★ Reduce or eradicate disease vectors species [84, 111] ★ Reduce or eradicate crops pests [194, 164, 84, 111]

Table 1.2: Main applications of suppression and eradication drives in health, conservation and agriculture sectors.

1.2.5 Risks with a focus on CRISPR-based homing drives

Risk assessment is a central question before any field releases. Most reports on gene drive emphasise on the lack of knowledge regarding complex potential consequences that could impact every scale of biodiversity from the ecosystem down to the molecular level [110]

Interaction networks within ecosystems are multiple and we only know the tip of the iceberg. The consequences of releasing a gene drive in a complex ecosystem are incredibly difficult to predict: the Norwegian Biotechnology Advisory Board warns about the "so-called known unknowns (expected or foreseeable) and unknown unknowns (unexpected or unforeseeable)" [27]. An eradication drive causing the extinction of a species will necessarily introduce a disequilibrium in the networks, potentially affecting other species than the one initially targeted. A report by the National Academies (US) warned about the "cascades of population dynamics and evolutionary processes" that could be initiated this way. [162].

In contrast to eradication drives, a suppression or replacement drive does not eliminate a species: it introduces a long-lasting modification in the population. In other words, the gene drive cassette remains indefinitely in the genome of the species, and the risk of it being disrupted in its function by emerging resistances is increased [183, 212, 45]. We also know that CRISPR-Cas9 does not work perfectly and might lead to large deletions, complex rearrangements [132], or even unintended changes to non-target sequences [128].

Although the drive is initially designed to target only one species, there is also an unlikely but very consequence-costly scenario of out-crossing across species boundaries not targeted initially [62, 56, 162].

1.2.6 Technological fix criticisms

In the agriculture field especially, productivism comes with the unquestioned belief that science and technology are progressive [206]. This way of thinking leads to the research of technological fixes where there may exist alternative and more appropriate ways of conceptualising the problem [153, 193]. Gene drive applications are often linked to deep social, cultural, legal and economic causes, and technologists might be blind to all others potential solutions, such as existing social practices or institutions.

Moreover, technological solutions are designed and built to solve narrowly defined problems [153], but "taking a wider and longer view, they tend to delay, transform and relocate problems, as well as creating new ones" [193]. In the bigger picture, technological fixes prove to not always be the most suitable solution, and might lead to important side effects. It is worrying because there is a risk that this aspect might not be taken into account as "the economic development strategies dominant today in agribusiness focus on short-term return on investments and [has a] disdain for long-term issues" [61]. It raises fears about the care and the rigour put into the risk assessment studies prior to the release of gene drive individuals.

Finally, technological fixes are often presented as progressive, but it is actually quite incorrect. Such solutions tend to "preserve, or fix, systems that should be abandoned in favour of better alternatives. They are in this sense conservative" [193]. Today, scientists are exploring ways of making threatened species more resilient to climate change through gene drive, such as preventing coral bleaching [108]. In agriculture, they work on increasing the abiotic stress tolerance in crops threatened by climate change [127]. All these gene drive researches illustrate the fact that technological fixing offers a trade-off, not a solution to climate change, with the risk of drawing attention away from the real problem.

Considering the future commercialisation of gene drive, patents mainly focus on eliminating resistance to herbicides or insecticides [84, 111]. These applications are rooted in the current agribusiness system and reinforce farmers' dependence on toxic agrochemicals instead of offering more sustainable solutions, when it might be time to reconsider our productivist model of agriculture.

1.2.7 Moral considerations

Gene drive use is justified by its purposes: improving human health, protecting the economy or/and promoting biodiversity. These purposes take humans' needs and interests as central and all means are legitimate to control or eradicate other species for humanity's greater good. But we should question the right of humans to dominate nature: do we dispose too easily of other living organisms to solve our problems or to satisfy our private interest? Are we not blinded in a technocratic and instrumentalist paradigm of nature [219]? Beyond this philosophical question also stands the economic vision dominant today that sees nature only as a source of profits. There is a need to take a step back and consider the many facets of the problem, including what cannot be quantified financially.

1.2.8 How to achieve confinement?

Gene drives spread limitless if their fitness cost is not too high. Therefore they are highly invasive: a small release of modified individuals will likely result in the colonisation of all the connected populations without distinction [5, 34, 69, 165]. Those geographic habitats might cover countries or even continents, spreading the gene drive across international borders. This propagation can be accelerated with long distance migrations: mosquitoes for example, can benefit from fast air currents (transporting them for hundreds of kilometres in a few hours) [49, 115, 116] or human-based modes of travel such as cars [79] or planes [80].

The property of limitless invasion might be appropriate and interesting in some specific cases, such as for large-scale disease eradication, at the condition that there exists high and wide social approval. However, in most cases, the possibility of drive confinement is highly desirable. It has been deemed safer both for laboratory research [3, 1] and first field testing releases [139]. Gene drive might also be used to solve local issues, for instance the control of invasive species in specific areas such as islands [202, 175]. Furthermore, regulatory issues might arise if gene drives spread across international borders without consent (violating the Cartagena Protocol [59]).

Ways to confine gene drive broadly fall into three categories: temporary drives, threshold-dependent drives and population specific drives.

Temporary drives

A simple way to limit the super-Mendelian inheritance property in time is to segregate the drive construct in two or more independent loci in the genome. Thanks to genetic mixing, the drive construct is doomed to be separated and the transmission advantage conferred to the cargo disabled over time. Consequently, we expect these constructs to locally spread at high frequencies for only a limited period of time.

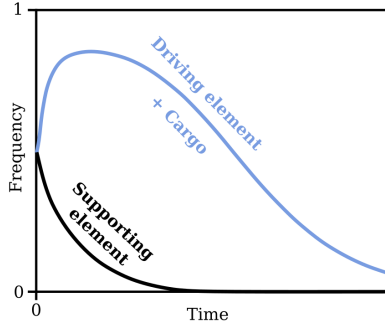


Figure 1.8: Driving element frequency and supporting element (linked to the cargo) frequency over time. This figure has been reproduced from [77].

In the simplest constructs, the drive is segregated in two parts: a driving element and a supporting element. The supporting element cannot increase in frequency: it will progressively be removed from the population due to its fitness cost. The driving element increases in frequency only if it happens to be with the supporting element: together in the genome, they confer a super-Mendelian advantage to the driving element. The driving element first benefits from the large frequency of supporting elements in the population to increase in turn in frequency, until the supporting element becomes too rare and the advantage is less profitable. At this point, the driving element, due to its fitness cost, starts decreasing in frequency too (Figure 1.8). The cargo and the driving element are linked together on a single locus so that they have the same dynamics (Figure 1.9).

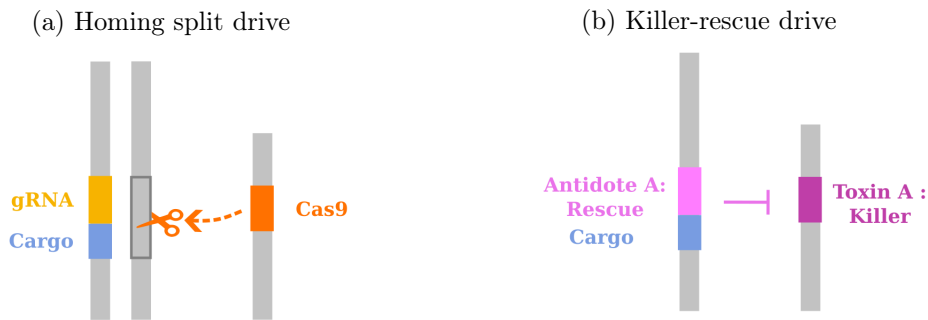


Figure 1.9: Schematic illustrations of a homing split drive (a) and a killer-rescue drive (b). In homing split drives, the supporting component is a Cas9 protein gene and the driving element is its corresponding gRNA-coded gene. In the killer-rescue drives, the supporting component is a toxin and the driving element is its corresponding antidote.

As in Section 1.2.2, the super-Mendelian advantage can be of two types: gene conversion (*homing split drives*) or toxin-antidote (*killer-rescue drives*). In homing split drives, the supporting element is a Cas9 protein gene while the driving element is its corresponding gRNA-coded gene. Together they allow the copy of the cargo complex in the homologous chromosome, following the mechanisms detailed in Section 1.2.2. In killer-rescue drives, the supporting element is a toxin while the driving element is its corresponding antidote. Together they reduce the fitness of the wild-type gametes or offspring, following the mechanisms detailed in Section 1.2.2. Both constructs are illustrated in Figure 1.9 and their effect on genotypes over time are detailed in Figure 1.10. Note that a combination of homing

split drive and killer-rescue drive (*Split drive killer-rescue*) have also been studied [78].

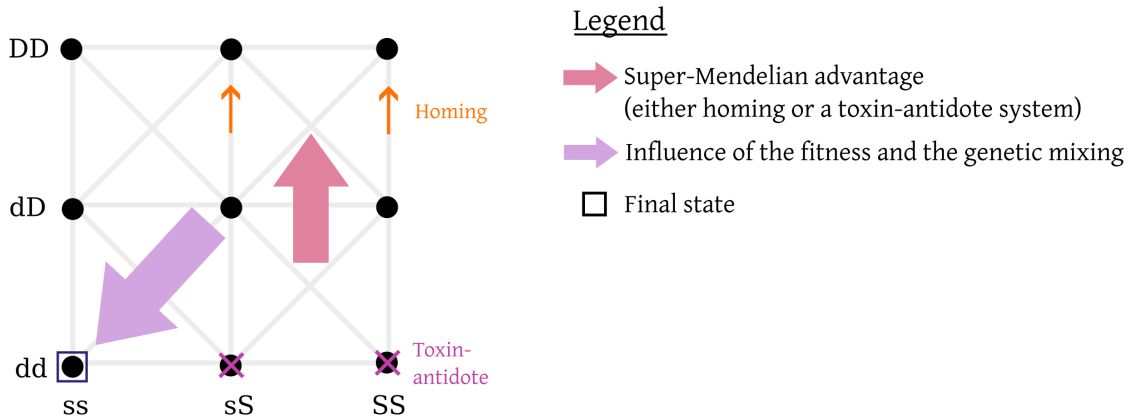


Figure 1.10: Diagram of the possible genotypes and their variations in frequency function of time. The supporting allele is denoted by S while its wild-type version is s. The driving allele is denoted by D while its wild-type version is d. The super-Mendelian advantage (homing or toxin-antidote system) increases the production of D alleles to the detriment of d alleles (vertical arrow from the bottom to the top). But this advantage is doomed to decrease over time as the S allele gradually disappears from the environment, due to its fitness cost. The genetic mixing tends to separate the transgenic S and D alleles and this tendency is enhanced by the difference of fitness between wild-type (d and s) and transgenic alleles (D and S), resulting in a full wild-type final state.

Homing split drives have been tested in laboratory on yeast [74], flies [42, 205] and mosquitoes [139]. Their time-limited inheritance advantage makes them safer in case of accidental releases and they could also be useful for short-term field trials before considering long-term releases.

To extend the moment when the super-Mendelian advantage disappears, it is possible to separate the drive in more than two loci. Daisy drive systems are similar to homing split drives, except that they contain several supporting and driving couples instead of one. These couples are arranged in a chain such that each one drives the next. Because the first element cannot drive, each element will in turn eventually decline after losing its supporting element [166]. The last link might either be a homing gene drive [166] or a toxin-antidote system [85]. In Figure 1.11, we illustrate the second possibility, a combination of a daisy chain and a cleave and rescue system (*HD-ClvR drive*) [85].

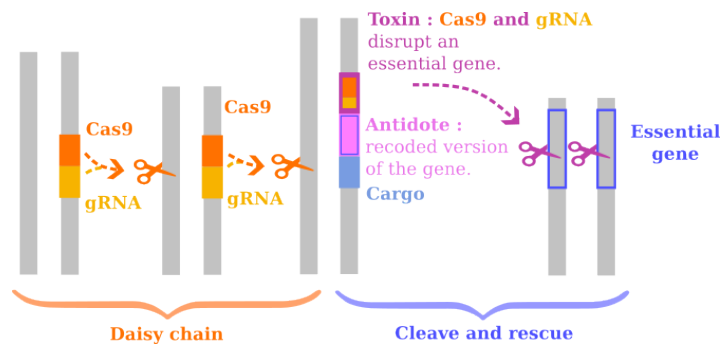


Figure 1.11: HD-ClvR: combination of a daisy chain and a cleave and rescue system.

Threshold-dependent drives

Threshold-dependent gene drives spread within a population only when introduced above some required threshold frequency. Therefore, if a few transgenic individuals happen to spread into a non-targeted area or population, either by dispersal or by accidental releases, we expect that this quantity would be too small to initiate a drive invasion. However, it appears that spatial confinement of threshold-dependent drives only works in discrete environments with large enough spatial steps, as shown in Chapter 5.

Most homing drive constructs are threshold-dependent but only under specific conditions over the drive fitness cost, the conversion rate and/or the dominance [69, 213, 187, 204, 95, 129] (the threshold-dependence is referred to as "bistability" or "unstable" in those articles). However, there exists a mechanism that can be combined to any kind of drive to ensure the threshold-dependence: underdominance. If the fitness of heterozygous individuals is lower than the fitness of homozygous ones, the invasion success depends on the initial drive frequency [107].

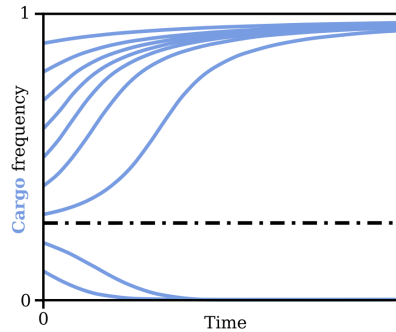


Figure 1.12: Cargo frequency over time in a threshold-dependent drive. The drive invasion is successful only if the initial drive frequency is high enough. This figure has been reproduced from [77].

A first option to reduce the fitness of heterozygotes is the use of translocation: a chromosomal rearrangement characterised by the reciprocal exchange of chromosomal material between non-homologous chromosomes. Translocation heterozygotes are usually partially sterile (low fitness) while translocation homozygotes if viable are usually fully fertile (high fitness) [64]. This interesting property could be used to fix desirable genes in pest populations. Experiments were conducted in that direction on *Aedes aegypti* [145], *Anopheles* [126, 125, 124] and *Drosophila* [32], but the chosen method faced engineering difficulties with especially high fitness costs for modified individuals.

Another way to achieve underdominance is to use toxin-antidote systems. Strictly speaking, we here extend the definition of underdominance as we consider two types of modified alleles and sometimes even two loci: the notion of heterozygote is generalised with the notion of *intermediate genotype*. Two chosen genotypes of interest, one fully wild-type and the other fully modified, have the highest fitnesses among all the genotypes (the threshold-dependent invasion dynamic occurs between them). The others, the intermediate genotypes have equal or lower fitnesses. A set of two engineered constructs is needed to obtain this dynamic, each of them containing the cargo, a toxin and the antidote corresponding to the toxin of the other engineered construct. The intermediate genotypes carrying a toxin without the corresponding antidote suffer a genetic load, while the genotypes of interest have either none or both of the engineered constructs. This toxin-antidote system can be designed over one locus or two loci (see Figure 1.13). The notion of intermediate genotypes with lower fitness is detailed for both cases in Figure 1.14.

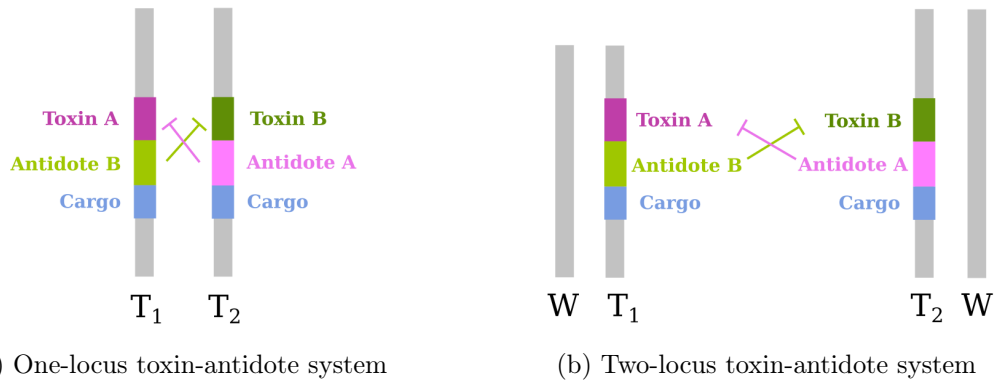


Figure 1.13: Schematic illustration of (a) an one-locus and (b) a two-locus toxin-antidote underdominance system introduced by [66].

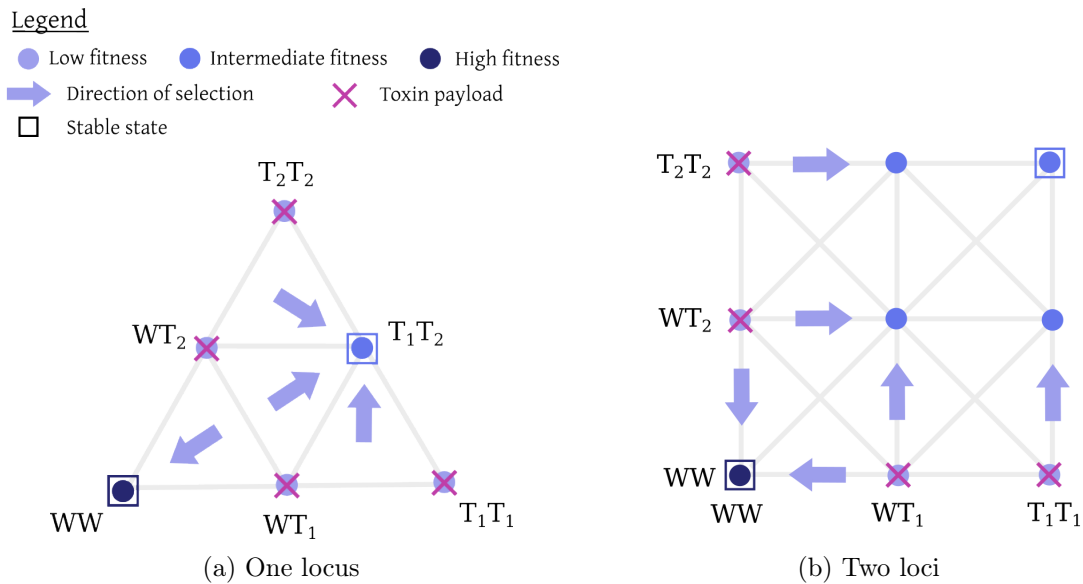


Figure 1.14: Diagram of the possible genotypes in a (a) one-locus and (b) two-locus toxin-antidote underdominance system. T_1 and T_2 are the two engineered constructs illustrated in Figure 1.13 while W represents the wild-type alleles. In sub-figure (a), the genotypes are given by the letter next to each point, while in sub-figure (b), the genotypes are the combination of a line and a column. The intermediate genotypes with lower fitness are in (a) WT_1 , WT_2 , T_1T_1 , T_2T_2 , and in (b) $WW-WT_1$, $WW-WT_2$, $WW-T_1T_1$, $WW-T_2T_2$. Depending on the initial condition and the fitness of each genotype, the final state can either be a fully wild-type or a fully modified genotype.

The invasion threshold can be fine-tuned by adjusting the fitness cost of the toxin, however the threshold must not be too low for the confinement to be efficient. Consequently, this method usually requires a large release of modified animals simultaneously. This might be a technical challenge, as well as a significant pressure for the local ecosystems: if the population is close to the carrying capacity, it may lead to starvation or mass migration. To circumvent these problems, combinations of both temporary and threshold-dependent drives have been proposed.

Combination of both temporary and threshold-dependent drives

A gene drive construct combining a super-Mendelian transmission limited in time with a threshold-dependent dynamic could prove to be very promising. Requiring a small initial release thanks to the "super-Mendelian boost", such drive should locally spread before fixing in the areas where its frequency is above the threshold. Outside these areas, if the environment is discrete with large enough spatial steps (see Chapter 2), it is supposed to disappear. The drive modification would then theoretically maintain indefinitely in a spatially limited environment, if we can ensure that unexpected large events of migration will never happen.

One construct of this type as been proposed recently, the *tethered homing gene drive* [71]. It is composed of a homing split drive and a two-locus underdominance drive (linked with the Cas9 endonuclease of the split drive, see Figure 1.15). The homing split drive limits the super-Mendelian transmission in time while the underdominance system provides the threshold-dependent dynamic. This construct has been tested on flies and proven to be easily engineered and efficient in laboratory [155].

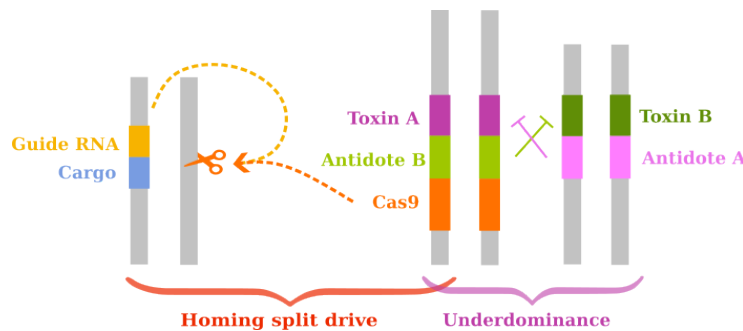


Figure 1.15: Tethered homing gene drive: combination of a homing split drive and a 2-locus underdominance system.

The *daisy quorum drive* is another combination of both temporary and threshold-dependent drives, composed of a daisy chain and a 2-locus underdominance system (Figure 1.16). Conceptually proposed by Min et al. [157], we have modelled its propagation in Chapter 5.

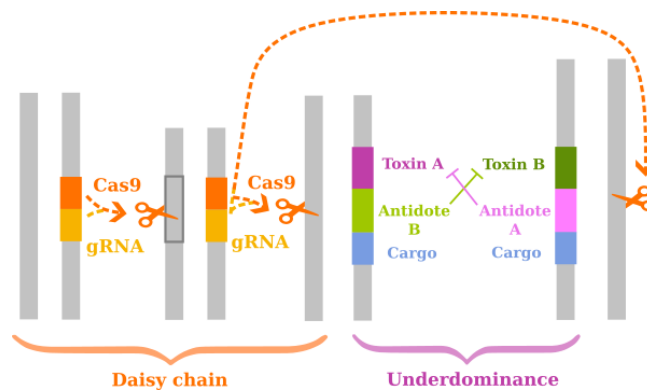


Figure 1.16: Daisy quorum drive: combination of a daisy chain and a 2-locus underdominance system. This construct is studied in Chapter 5.

Population-specific confinement

Finally, a third way to confine gene drive spread at a population scale is to ensure that it will only target a defined set of individuals. To serve this purpose, homing can be conditioned to the presence of unique and highly conserved DNA sequences, observed in the target population but absent in all the others [202, 220]. It would be interesting to integrate this additional safety barrier in drive constructs if such sequences exist.

1.3 Mathematical and numerical analysis

This PhD thesis mainly focuses on homing gene. Key to my analysis are the drive fitness cost, the gene conversion rate and the gene conversion timing (when does the conversion of alleles W into alleles D happen in the life cycle). I present step-by-step early gene drive models and then more complex models that I studied during this thesis.

1.3.1 Introduction to gene drive modelling

Early panmictic models

Let consider the propagation of two alleles (A and B) with Mendelian inheritance and equal fitnesses for all genotypes. In a well-mixed population mating at random with discrete non-overlapping generations, the frequency of alleles A at time t is denoted by p_A^t and $p_B^t = 1 - p_A^t$. Time is discrete; the allele frequency is considered just before fertilisation (among the produced gametes), so that the frequency of newly born homozygotes AA is $(p_A^t)^2$, newly born homozygotes BB is $(1 - p_A^t)^2$, and newly born heterozygotes AB is $2 p_A^t (1 - p_A^t)$. At each new generation, we have:

$$p_A^{t+1} = (p_A^t)^2 + (2 p_A^t (1 - p_A^t)) \frac{1}{2}. \quad (1.1)$$

The coefficient $\frac{1}{2}$ at the end of the formula indicates that only half of the gametes produced by heterozygotes AB are A gametes.

If we now consider the propagation of drive alleles with super Mendelian inheritance, more than half of the gametes produced by heterozygous individuals DW will be drive (D : drive, W : wild-type). We denote by c the conversion rate, i.e. the rate of successful conversion of an allele W into an allele D in a heterozygous cell. This rate is given by the frequency of Cas9 efficient cuts times the frequency of double-strand break being repaired by homology direct-repair. With equal fitness for all genotypes, the frequency p_D^{t+1} of drive gametes at time $t + 1$ is given by:

$$\begin{aligned} p_D^{t+1} &= (p_D^t)^2 + c (2 p_D^t (1 - p_D^t)) + (1 - c) (2 p_D^t (1 - p_D^t)) \frac{1}{2} \\ &= (p_D^t)^2 + \frac{c+1}{2} (2 p_D^t (1 - p_D^t)). \end{aligned} \quad (1.2)$$

A proportion c of heterozygotes becomes homozygous drive, while a proportion $1 - c$ is not converted. In laboratories, experiments have shown drive transmission rates in heterozygotes ($\frac{c+1}{2}$) of 99% in yeast *Saccharomyces cerevisiae* [74], more than 90% in mosquito *Anopheles gambiae* [88], and more than 85%

in fruit flies *Drosophila melanogaster* [224]. These rates are relatively high and sometimes simplified to a systematic conversion ($c = 1$).

To model suppression or eradication drives, I consider a fitness coefficient between 0 and 1 reducing the number of offspring. The more drive alleles an individual carries, the more its fecundity is negatively affected. The fitness of wild-type homozygotes is $f_{WW} = 1$ (not impacted). The fitness of drive homozygotes is $f_{DD} = 1 - s$ where $s \in (0, 1)$ is the fitness cost of the drive. The fitness of drive heterozygotes is $f_{DW} = 1 - sh$, where $h \in (0, 1)$ is the dominance parameter.

Gene conversion can happen in the germline (successfully experimented in laboratory) or in the zygote (still speculative for now). This conversion timing significantly impacts the drive propagation as the fitness is based on the adult genotype. If we consider a heterozygous egg, the adult is homozygous drive with fitness $1 - s$ if conversion occurs in the zygote, while it remains heterozygote with fitness $1 - sh$ if conversion occurs in the germline. A schematic illustration is given in Figure 1.17 and the drive dynamics are detailed in equation (1.8) for zygote conversion and equation (1.9) for germline conversion.

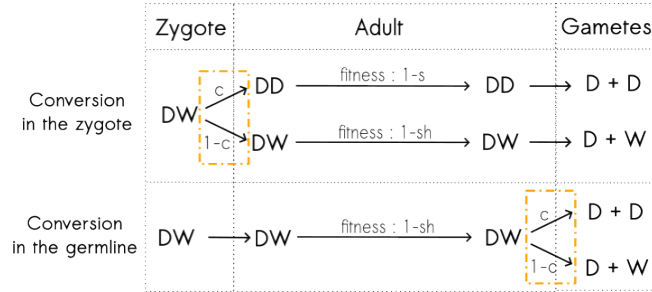


Figure 1.17: Schematic illustration of the impact of conversion timing (rectangle in orange) on fitness.

We define the mean fitness (weighted by the proportion of each pair of alleles i.e. each genotype) when conversion occurs in the zygote \mathcal{F}^z and when conversion occurs in the germline \mathcal{F}^g :

$$\mathcal{F}^z(p_D) = (1 - s) (p_D)^2 + [(1 - s)c + (1 - sh)(1 - c)] (2 p_D (1 - p_D)) + (1 - p_D)^2, \quad (1.3)$$

$$\mathcal{F}^g(p_D) = (1 - s) (p_D)^2 + (1 - sh) (2 p_D (1 - p_D)) + (1 - p_D)^2. \quad (1.4)$$

We also define the mean drive fitness \mathcal{F}_D^z and \mathcal{F}_D^g and the mean wild-type fitness \mathcal{F}_W^z and \mathcal{F}_W^g as follow:

$$\mathcal{F}_D^z(p_D) = (1 - s) (p_D)^2 + (1 - s) c (2 p_D (1 - p_D)) + (1 - sh) (1 - c) (2 p_D (1 - p_D)) \frac{1}{2}, \quad (1.5)$$

$$\mathcal{F}_D^g(p_D) = (1 - s) (p_D)^2 + (1 - sh) c (2 p_D (1 - p_D)) + (1 - sh) (1 - c) (2 p_D (1 - p_D)) \frac{1}{2}, \quad (1.6)$$

$$\mathcal{F}_W^z(p_D) = \mathcal{F}_W^g(p_D) = (1 - sh) (1 - c) (2 p_D (1 - p_D)) \frac{1}{2} + (1 - p_D)^2, \quad (1.7)$$

which verify $\mathcal{F}_D^z + \mathcal{F}_W^z = \mathcal{F}^z$ and $\mathcal{F}_D^g + \mathcal{F}_W^g = \mathcal{F}^g$.

The frequency of drive and wild-type alleles is given by

$$p_D^{t+1} = \frac{\mathcal{F}_D^z(p_D^t)}{\mathcal{F}^z(p_D^t)} \quad \text{and} \quad p_W^{t+1} = \frac{\mathcal{F}_W^z(p_D^t)}{\mathcal{F}^z(p_D^t)}, \quad (1.8)$$

when conversion occurs in the zygote, and by

$$p_D^{t+1} = \frac{\mathcal{F}_D^g(p_D^t)}{\mathcal{F}^g(p_D^t)} \quad \text{and} \quad p_W^{t+1} = \frac{\mathcal{F}_W^g(p_D^t)}{\mathcal{F}^g(p_D^t)}, \quad (1.9)$$

when conversion occurs in the germline. Note that the renormalisation by \mathcal{F}^z and \mathcal{F}^g is necessary to fulfil the condition $p_D^{t+1} + p_W^{t+1} = 1$. Previous studies have analysed equation (1.8) [69, 213] and equation (1.9) [69, 187]. They have determined for each value of c , s and h , the type of invasion (drive, wild-type or coexistence final state) and characterised if the drive was threshold-dependent or not, among other results.

Spatial models

Understanding the spatial spread of gene drives and how demographic features might affect it is essential before any field release. If the drive invades the population in space, this occurs through a traveling wave: a wave of change in genotype densities through space. Traveling waves propagate with a constant speed while maintaining their shape in space (for a rigorous definition of traveling waves, see Chapter 2).

Spatially explicit models can be discrete in time and space with punctual events of migration, or continuous in time and space assuming that the movement of individuals is described by a diffusion term. In the discrete stepping stone version, a proportion m of every population migrates in the neighbouring populations at each time step, while in the continuous reaction-diffusion version the individuals move in a Brownian motion modelled by a second derivative of density in space multiplied by a diffusion rate \mathcal{D} (for details, see [173]).

In spatially explicit models, it is important to consider the allelic density instead of the allelic frequency, because all spatial sites do not necessarily contain the same number of individuals. We denote n_D the drive allelic density, n_W the wild-type allelic density, and $n = n_D + n_W$ the total allelic density. As before, \mathcal{F} is the mean fitness, \mathcal{F}_D the mean drive fitness and \mathcal{F}_W the mean wild-type fitness. If we consider the discrete panmictic model:

$$\begin{cases} n_D^{t+1} = \mathcal{F}_D(p_D^t) n^t, \\ n_W^{t+1} = \mathcal{F}_W(p_D^t) n^t, \end{cases} \iff \begin{cases} n^{t+1} = \mathcal{F}(p_D^t) n^t, \\ p_D^{t+1} = \frac{n_D^{t+1}}{n^{t+1}} = \frac{\mathcal{F}_D(p_D^t)}{\mathcal{F}(p_D^t)}, \end{cases} \quad (1.10)$$

then the spatial allelic density in the discrete stepping stones version of model (1.10) (with x the spatial variable) is given by:

$$n_i^{t+1,x} = (1 - m) \mathcal{F}_i(p_D^{t,x}) n^{t,x} + \frac{m}{2} \left(\mathcal{F}_i(p_D^{t,x+1}) n^{t,x+1} + \mathcal{F}_i(p_D^{t,x-1}) n^{t,x-1} \right) \quad \forall i \in \{D, W, \emptyset\}, \quad (1.11)$$

and the corresponding spatial drive frequency is:

$$p_D^{t+1,x} = \frac{n_D^{t+1,x}}{n^{t+1,x}} = \frac{(1 - m) \mathcal{F}_D(p_D^{t,x}) n^{t,x} + \frac{m}{2} \left(\mathcal{F}_D(p_D^{t,x+1}) n^{t,x+1} + \mathcal{F}_D(p_D^{t,x-1}) n^{t,x-1} \right)}{(1 - m) \mathcal{F}(p_D^{t,x}) n^{t,x} + \frac{m}{2} \left(\mathcal{F}(p_D^{t,x+1}) n^{t,x+1} + \mathcal{F}(p_D^{t,x-1}) n^{t,x-1} \right)}. \quad (1.12)$$

It could be tempting to not take into account demography in the model and directly apply migration on the drive frequency dynamic in system (1.10) :

$$p_D^{t+1,x} = (1 - m) \frac{\mathcal{F}_D(p_D^{t,x})}{\mathcal{F}(p_D^{t,x})} + \frac{m}{2} \frac{\mathcal{F}_D(p_D^{t,x+1})}{\mathcal{F}(p_D^{t,x+1})} + \frac{m}{2} \frac{\mathcal{F}_D(p_D^{t,x-1})}{\mathcal{F}(p_D^{t,x-1})}, \quad (1.13)$$

however the drive frequency dynamics in equations (1.12) and (1.13) are equivalent if and only if $n^{t,x-1} = n^{t,x} = n^{t,x+1} \quad \forall t, x$ i.e. if and only if the population remains homogeneously distributed across space. This assumption is wrong in case of suppression or eradication drive invasion, which significantly reduce the population size while spreading.

The same reasoning holds if we consider the continuous panmictic model:

$$\begin{cases} \partial_t n_D(t) = \mathcal{F}_D(p_D(t)) n(t) - n_D(t), \\ \partial_t n_W(t) = \mathcal{F}_W(p_D(t)) n(t) - n_W(t), \end{cases} \iff \begin{cases} \partial_t n(t) = \mathcal{F}(p_D(t)) n(t) - n(t), \\ \partial_t p(t) = \partial_t \left(\frac{n_D(t)}{n(t)} \right) = \mathcal{F}_D(p_D(t, x)) - p_D(t, x) \mathcal{F}(p_D(t, x)), \end{cases} \quad (1.14)$$

then the spatial allelic density in the continuous reaction-diffusion version of model (1.14) is given by:

$$\partial_t n_i(t, x) - \mathcal{D} \partial_{xx}^2 n_i(t, x) = \mathcal{F}_i(p_D(t)) n(t, x) - n_i(t, x) \quad \forall i \in \{D, W, \emptyset\}, \quad (1.15)$$

and the corresponding spatial drive frequency is (for calculation details see Appendix of Chapter 2):

$$\begin{aligned} \partial_t p_D(t, x) - \mathcal{D} \partial_{xx}^2 p_D(t, x) &= \partial_t \left(\frac{n_D(t, x)}{n(t, x)} \right) - \mathcal{D} \partial_{xx}^2 \left(\frac{n_D(t, x)}{n(t, x)} \right) \\ &= \mathcal{F}_D(p_D(t, x)) - p_D(t, x) \mathcal{F}(p_D(t, x)) + 2 \partial_x(\log n(t, x)) \partial_x p_D(t, x). \end{aligned} \quad (1.16)$$

Again, it could be tempting to not take into account demography in the model and directly apply diffusion on the drive frequency dynamic in system (1.14):

$$\partial_t p_D(t, x) - \mathcal{D} \partial_{xx}^2 p_D(t, x) = \mathcal{F}_D(p_D(t, x)) - p_D(t, x) \mathcal{F}(p_D(t, x)). \quad (1.17)$$

however the drive frequency dynamics in equations (1.16) and (1.17) are equivalent if and only if $\partial_x(\log n(t, x)) = 0 \quad \forall t, x$, i.e. if and only if the population remains homogeneously distributed across space. The transport term $2 \partial_x(\log n) \partial_x p_D$ illustrates the demographic flux from denser to less dense areas naturally generated by variations in population density, especially important in case of suppression and eradication drives. This flux is directed in opposition to the spread of the drive allele and may counteract the drive progression [95].

For the sake of clarity in the following, I omit variables in the notations ($n_i = n_i(t, x)$ and $p_i = p_i(t, x)$).

Methods and models

This PhD thesis studies the spatial and temporal spread of gene drives, more specifically of homing gene drives. These gene drives can significantly reduce the population size (suppression and eradication drive) and create important gradients of density in space while spreading. Nevertheless demography is usually ignored in spatial models (see e.g. [204, 167, 166]). To investigate how it might impact the drive dynamics, I will focus on the following outcomes: i) the density of the final population and its composition in terms of genotype densities, ii) the speed of invasion if there is one, and iii) the possibility to spatially confine the drive spread. In Chapter 2 and 3, I study the influence of demography and population dynamics over the first two features, with a deterministic approach. In Chapter 4, I explore stochastic dynamics beyond this deterministic approach when the population size gets to be small after the propagation of an eradication drive. In particular, I study wild-type recolonising events that might prevent the eradication of the population within a targeted area. Finally, in Chapter 5, I

study necessary conditions for a drive underdominant construct to fix inside a targeted area, but to fail outside.

In Chapter 2, I assume that the wild-type population without drive follows a logistic growth with an intrinsic growth rate r and a maximum carrying capacity of 1. I denote this density \tilde{n}_{WW} (only wild-type homozygotes) which dynamics is given by the standard Fisher–KPP equation [87, 130]:

$$\partial_t \tilde{n}_{\text{WW}} - \mathcal{D} \partial_{xx}^2 \tilde{n}_{\text{WW}} = r(1 - \tilde{n}_{\text{WW}}) \tilde{n}_{\text{WW}} = \overbrace{r(1 - \tilde{n}_{\text{WW}}) + 1}^{\text{births}} \tilde{n}_{\text{WW}} - \overbrace{\tilde{n}_{\text{WW}}}^{\text{deaths}}. \quad (1.18)$$

Then, I modify the birth term to take into account the super-Mendelian advantage, fitness differences and mating probabilities. I denote the density of drive homozygotes n_{DD} , heterozygotes n_{DW} , wild-type homozygotes n_{WW} , and the total density $n = n_{\text{DD}} + n_{\text{DW}} + n_{\text{WW}}$. I assume that all genotypes disperse at the same rate, arbitrary set to $\mathcal{D} = 1$. I obtain two systems (one for each conversion timing) of the form:

$$\begin{cases} \partial_t n_{\text{DD}} - \partial_{xx}^2 n_{\text{DD}} &= B_{\text{DD}}(n_{\text{DD}}, n_{\text{DW}}, n_{\text{WW}}) n_{\text{DD}} - n_{\text{DD}} \\ \partial_t n_{\text{DW}} - \partial_{xx}^2 n_{\text{DW}} &= B_{\text{DW}}(n_{\text{DD}}, n_{\text{DW}}, n_{\text{WW}}) n_{\text{DW}} - n_{\text{DW}} \\ \partial_t n_{\text{WW}} - \partial_{xx}^2 n_{\text{WW}} &= B_{\text{WW}}(n_{\text{DD}}, n_{\text{DW}}, n_{\text{WW}}) n_{\text{WW}} - n_{\text{WW}} \end{cases} \quad (1.19)$$

The birth terms B_{DD} , B_{DW} and B_{WW} are detailed in Chapter 2 for each conversion timing, either in the zygote or in the germline. From these systems of three equations, I deduced two systems on the allelic (half-)densities $(n_{\text{D}}, n_{\text{W}})$ containing only two equations. The relationship between genotype and allele densities depends on the timing of conversion: $n_{\text{D}} = n_{\text{DD}} + \alpha n_{\text{DW}}$ and $n_{\text{W}} = n_{\text{WW}} + (1 - \alpha) n_{\text{DW}}$, with $\alpha = \frac{1}{2}$ when conversion occurs in the zygote, and $\alpha = \frac{1+c}{2}$ when conversion occurs in the germline. Both systems are described in systems (1.20) and (1.21).

Allele density (conversion in the zygote):

$$\begin{cases} \partial_t n_{\text{D}} - \partial_{xx}^2 n_{\text{D}} &= (r(1 - n) + 1) \left[(1 - s) \frac{n_{\text{D}}^2}{n} + (1 - s)c \frac{2n_{\text{D}}n_{\text{W}}}{n} + (1 - sh) \frac{1 - c}{2} \frac{2n_{\text{D}}n_{\text{W}}}{n} \right] - n_{\text{D}}, \\ \partial_t n_{\text{W}} - \partial_{xx}^2 n_{\text{W}} &= (r(1 - n) + 1) \left[\frac{n_{\text{W}}^2}{n} + (1 - sh) \frac{1 - c}{2} \frac{2n_{\text{W}}n_{\text{D}}}{n} \right] - n_{\text{W}}. \end{cases} \quad (1.20)$$

Allele density (conversion in the germline):

$$\begin{cases} \partial_t n_{\text{D}} - \partial_{xx}^2 n_{\text{D}} &= (r(1 - n) + 1) \left[(1 - s) \frac{n_{\text{D}}^2}{n} + (1 - sh) \frac{1 + c}{2} \frac{2n_{\text{D}}n_{\text{W}}}{n} \right] - n_{\text{D}}, \\ \partial_t n_{\text{W}} - \partial_{xx}^2 n_{\text{W}} &= (r(1 - n) + 1) \left[\frac{n_{\text{W}}^2}{n} + (1 - sh) \frac{1 - c}{2} \frac{2n_{\text{W}}n_{\text{D}}}{n} \right] - n_{\text{W}}. \end{cases} \quad (1.21)$$

Equivalently, I rewrite models (1.20) and (1.21) so that they follow the drive frequency $p_{\text{D}} = \frac{n_{\text{D}}}{n_{\text{D}} + n_{\text{W}}}$ and the total population density $n = n_{\text{W}} + n_{\text{D}}$.

Allele frequency (conversion in the zygote):

$$\left\{ \begin{array}{l} \partial_t n - \partial_{xx}^2 n = (r(1-n) + 1) \left((1-s)p_D^2 + 2p_D(1-p_D)[c(1-s) + (1-c)(1-sh)] + (1-p_D)^2 \right) n - n, \\ \partial_t p_D - \partial_{xx}^2 p_D = 2 \partial_x \log(n) \partial_x p_D \\ \quad + (r(1-n) + 1) \left([1 - 2(1-c)(1-h)] s p_D - s[1 - (1-c)(1-h)] + c(1-s) \right) (1-p_D)p_D. \end{array} \right. \quad (1.22)$$

Allele frequency (conversion in the germline):

$$\left\{ \begin{array}{l} \partial_t n - \partial_{xx}^2 n = (r(1-n) + 1) \left((1-s)p_D^2 + 2(1-sh)p_D(1-p_D) + (1-p_D)^2 \right) n - n, \\ \partial_t p_D - \partial_{xx}^2 p_D = 2 \partial_x \log(n) \partial_x p_D + (r(1-n) + 1) \left((2h-1) s p_D + (1-sh)(1+c) - 1 \right) p_D (1-p_D). \end{array} \right. \quad (1.23)$$

In Chapter 2, I consider a general and theoretical approach to study gene drive propagation for both conversion timings (in the zygote or in the germline). In Chapter 3, I generalise the results of Chapter 2 taking into account several other biological assumptions on population dynamics, for practical use by biologists. The wild-type population dynamics before any drive introduction is modified, therefore I study variations of equation (1.18). As the conversion in the zygote remains speculative today, I focus on germline conversion in this third chapter. I introduce:

An Allee effect, with $a \in [-1, 1]$ being the Allee threshold:

$$\partial_t \tilde{n}_{\text{WW}} - \partial_{xx}^2 \tilde{n}_{\text{WW}} = \overbrace{(r(1-\tilde{n}_{\text{WW}}) (\tilde{n}_{\text{WW}} - a) + 1) \tilde{n}_{\text{WW}}}^{\text{births}} - \overbrace{\tilde{n}_{\text{WW}}}^{\text{deaths}}, \quad (1.24)$$

A logistic density dependence affecting the deaths instead of the births:

$$\partial_t \tilde{n}_{\text{WW}} - \partial_{xx}^2 \tilde{n}_{\text{WW}} = \overbrace{(r+1) \tilde{n}_{\text{WW}}}^{\text{births}} - \overbrace{(1+r\tilde{n}_{\text{WW}}) \tilde{n}_{\text{WW}}}^{\text{deaths}}, \quad (1.25)$$

Or both of them:

$$\partial_t \tilde{n}_{\text{WW}} - \partial_{xx}^2 \tilde{n}_{\text{WW}} = \overbrace{(r+1) \tilde{n}_{\text{WW}}}^{\text{births}} - \overbrace{(r(\tilde{n}_{\text{WW}} - 1) (\tilde{n}_{\text{WW}} - a) + r + 1) \tilde{n}_{\text{WW}}}^{\text{deaths}}. \quad (1.26)$$

As before, these equations form the basis of my models, I then consider the influence of the drive introduction on each genotype density first, on each allele density second, and on each allele frequency finally.

In Chapter 4, I present a stochastic model to highlight possible stochastic events such as the reemergence of small wild-type populations close to extinction (Chapter 4). This model is discrete in time and space; we denote dt the size of a temporal step and dx the size of a spatial step. I again consider that conversion happens in the germline and I set an identical migration proportion for all genotypes ($m = 0.2$). The model is illustrated in Figure 1.18 and the mean allele density dynamics are given by system (4.43) and (4.44).

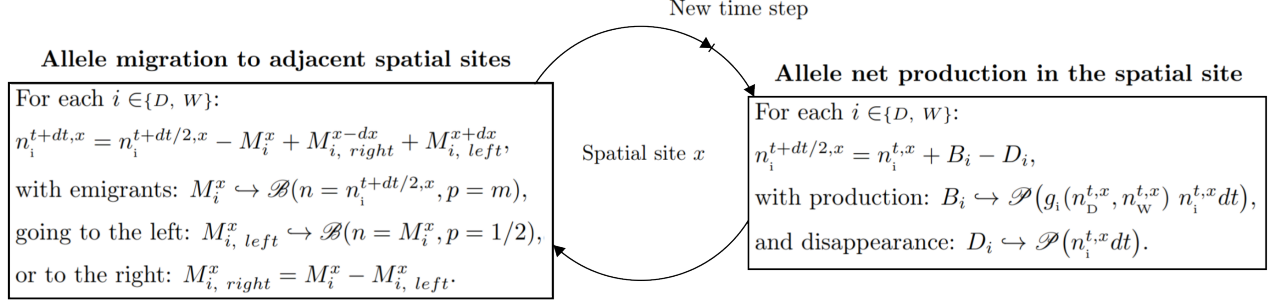


Figure 1.18: Illustration of the different steps in the stochastic discrete model, in one spatial dimension.

Mean drive allele density:

$$\begin{cases} n_D^{t+\frac{dt}{2},x} &= \left((g_D(n_D^{t,x}, n_W^{t,x}) - 1) n_D^{t,x} dt + n_D^{t,x}, \right. \\ n_D^{t+dt,x} &= (1 - m) n_D^{t+\frac{dt}{2},x} + \frac{m}{2} (n_D^{t+\frac{dt}{2},x+dx} + n_D^{t+\frac{dt}{2},x-dx}), \end{cases} \quad (1.27)$$

Mean wild-type allele density:

$$\begin{cases} n_W^{t+\frac{dt}{2},x} &= (g_W(n_D^{t,x}, n_W^{t,x}) - 1) n_W^{t,x} dt + n_W^{t,x}, \\ n_W^{t+dt,x} &= (1 - m) n_W^{t+\frac{dt}{2},x} + \frac{m}{2} (n_W^{t+\frac{dt}{2},x+dx} + n_W^{t+\frac{dt}{2},x-dx}), \end{cases} \quad (1.28)$$

with:

$$g_D(n_D, n_W) = (r(1-n) + 1) \left[(1-s) \frac{n_D}{n} + (1-sh)(1+c) \frac{n_W}{n} \right], \quad (1.29)$$

$$g_W(n_D, n_W) = (r(1-n) + 1) \left[\frac{n_W}{n} + (1-sh)(1-c) \frac{n_D}{n} \right]. \quad (1.30)$$

Note that there exists a relationship between the diffusion rate \mathcal{D} and the migration rate m (for details, see Appendix of Chapter 5):

$$\mathcal{D} = \frac{m(\Delta x)^2}{2\Delta t}. \quad (1.31)$$

In Chapter 5, I study the spread of a daisy quorum gene drive construct. I do not detail the model in this synthesis as it is very complex and quite different from Models (1.21), (1.20) and from the model illustrated in Figure 1.18: it is composed of 256 equations thanks to the four possible allele combinations on four different loci. The details of the model are available in the main text of the Chapter 5.

1.3.2 Main results of this thesis

Different population dynamics can emerge from the models presented in the previous section. I detail them all at once, before presenting the parameters conditions leading to such dynamics.

Overview of the possible dynamics

I consider an initial condition in which the left half of the domain is full of drive ($n_{DD} = 1$), and the right half is full of wild-type ($n_{WW} = 1$), illustrated in Figure 1.19. In this PhD thesis, I am not exploring the effect of inoculum size and distribution, which is a question in itself, and arises in the case of threshold-dependent drives. I therefore chose an initial condition maximising the possibility of drive spread.

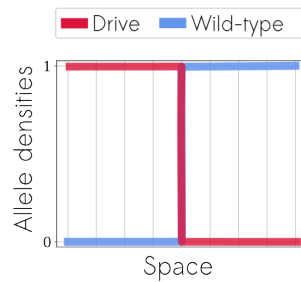


Figure 1.19: Initial allelic densities conditions in a one-dimensional space. The drive allele density is in red while the wild-type allele density is in blue.

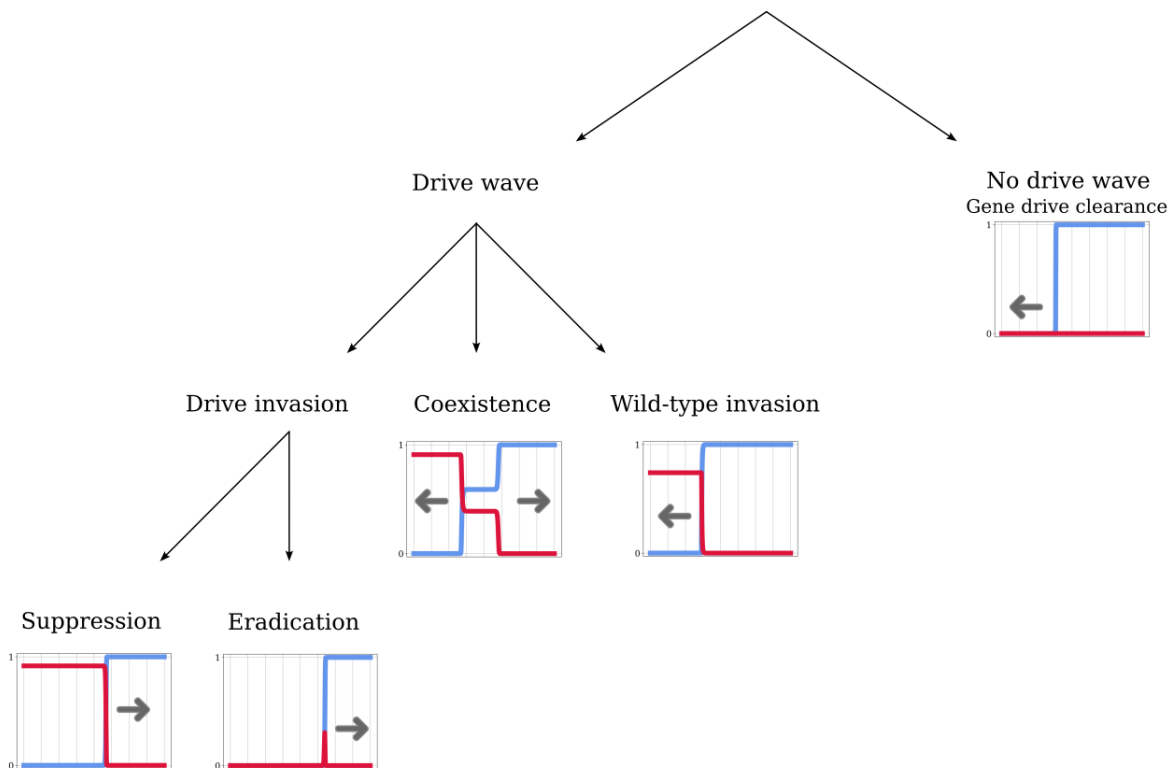


Figure 1.20: Overview of the different allelic density dynamics, in a one-dimensional space. The drive allele density is in red while the wild-type allele density is in blue. The arrows indicate the direction in which the wave moves as a function of time.

The arrows indicate the direction in which the wave moves as a function of time.

At high drive fitness costs (s) and low intrinsic growth rates (r) we observe the decay of the drive allele uniformly in space: a case we call *gene drive clearance*. No drive traveling wave is formed and the problem boils down to a standard Fisher-KPP traveling wave problem as the wild-type population colonises an empty environment.

Except for this particular situation, the drive traveling wave exists. We distinguish between two cases depending on the sign of the speed. When $v > 0$, the wave moves to the right leading to a *drive invasion*. When $v < 0$, the wave moves to the left leading to a *wild-type invasion*.

In some specific cases, drive and wild-type invasions can happen simultaneously: the waves decompose into two sub-traveling wave solutions over half of the domain. They move in opposite directions and lead to the coexistence of both alleles in-between.

In the case of drive invasion, we distinguish several cases depending on the state of the population in the wake of the front(s): i) *eradication* drives are those for which the population vanishes in the wake of the front(s); ii) *suppression* drives are those for which population persists in the wake of the front(s).

Other dynamics might appear when considering stochastic models. More specifically, eradication gene drives might experience events of wild-type reemergence in areas previously cleared by the drive [46]. These events often prevent the population from complete elimination and might lead to infinite colonisation-extinction events: the wild-type genotype invades empty areas, the drive genotype invades wild-type colonised areas, and the drive invasion leads to local eradication. We denote these reemergence dynamics as *chasing dynamics*.

In addition to the type of invasion (or the absence if there exists no drive wave) and its speed, we can describe the dynamics through several other criteria.

An equation / a system is monostable if it has exactly one stationary stable state, and bistable if it has exactly two stationary stable states. This distinction is important as it indicates whether or not the fate of the invasion can be changed by introducing more or less gene drive individuals. It is the case in the bistable configuration, also referred to as threshold-dependent dynamics and detailed in Section 1.2.8 (density-dependent confinement part).

A wave is said to be pulled if the wave speed coincides with the minimal speed of the linearised problem at low density (resp. low frequency). This occurs when the population at low density (resp. low frequency) has sufficient reproductive success to determine the dynamics of the full invasion. A wave is said to be pushed if the wave speed is strictly larger than the minimal speed of the linearised problem. In contrast to pulled waves, the whole population contributes to the dynamics of invasion [188].

The demographic impact of a gene drive can change the nature of its spatial spread

In Chapter 2, I study the reaction-diffusion models (1.20) and (1.21) describing the interplay between demographic and allelic dynamics, in a one-dimensional spatial context. I focus on the traveling wave solutions, and more specifically, on the speed of gene drive invasion (if successful).

In this synthesis, I detail results for the model assuming systematic conversion in the zygote ($c = 1$ in system (1.21)) but generalisations exist for system with partial conversion ($c \in [0, 1]$ in systems (1.20) and (1.21)). I briefly mention them along the text, for details see Chapter 2. These generalisations were

made possible by rewriting the system on genotypes density (three equations) into a system on allelic densities (two equations). Interestingly, this loss of information does not prevent us from describing the drive propagation and its wave properties.

The analysis of reaction-diffusion systems even with only two equations remains complex. But studying these systems for extreme values of r (the intrinsic growth rate), I was able to draw parallels with already existing mathematical models or results.

On the one hand, when $r = 0$ system (1.20) with $c = 1$ becomes:

$$\begin{cases} \partial_t n_W - \partial_{xx}^2 n_W &= \frac{-n_W n_D}{n_W + n_D}, \\ \partial_t n_D - \partial_{xx}^2 n_D &= (1-s) \frac{n_W n_D}{n_W + n_D} - s n_D. \end{cases} \quad (1.32)$$

Noticeably, system (1.32) shares some features with density-dependent epidemiological SI models. By changing notations $n_{WW} \leftrightarrow S$ (susceptible individuals), and $n_{DD} \leftrightarrow I$ (infected individuals), it can be recast as follows:

$$\begin{cases} \partial_t S - \partial_{xx}^2 S &= -\beta_1 \frac{S I}{S + I}, \\ \partial_t I - \partial_{xx}^2 I &= \beta_2 \frac{S I}{S + I} - \gamma I. \end{cases} \quad (1.33)$$

with $\beta_1 = 1$, $\beta_2 = (1-s)$ (transmission parameters), and $\gamma = s$ (disease clearance). Usually, in SI models, individuals of type S are all transformed into individuals of type I at infection, hence $0 < \beta_1 = \beta_2$. In our case, these two rates are distinct because of the fitness cost of the drive. The existence and characterisation of traveling waves for model (1.32) with $\beta_1 = \beta_2$ has been recently studied in the literature [225]. I adapted these proofs and generalised the results for $\beta_1 \neq \beta_2$ (with $0 < \beta_1$ and $0 < \beta_2$), leading to the characterisation in Table 1.3.

On the other hand, when $r \rightarrow +\infty$ system (1.23) with $c = 1$ reduces to one equation on p :

$$\partial_t p_D - \partial_{xx}^2 p_D = \frac{(1-s) p_D^2 + (1-s) 2p_D(1-p_D)}{(1-s) p_D^2 + (1-s) 2p_D(1-p_D) + (1-p_D)^2} - p_D = \frac{s p_D (1-p_D) \left(p_D - \frac{2s-1}{s} \right)}{1-s+s(1-p_D)^2}, \quad (1.34)$$

while n is constant equal to 1 everywhere in the environment. Interestingly when $r \rightarrow +\infty$ the dynamics are described by only one equation (1.34): this was previously shown in [95] using the Strugarek-Vauchelet rescaling [201]. Note that this equation is independent of the population density n and does not contain the term $2 \partial_x(\log n) \partial_x p_D$. This is due to the fact that the population size $n(t, x)$ remains spatially homogeneous when $r \rightarrow +\infty$. Indeed so many offspring are produced at each generation that the carrying capacity is instantaneously restored at all time. This represents a rare case where the variations in population density across space (n) are negligible, and consequently, so is the demographic flux.

The characterisation of solutions of equation (1.34) is given in Table 1.3. The numerical value of the threshold for the transition from positive to negative speed ($s \approx 0.70$) was already known [204, 95]. The numerical value of the threshold for the transition from pulled to pushed ($s \approx 0.35$ up to two digits) is numerically computed by a continuation method following [13, 114].

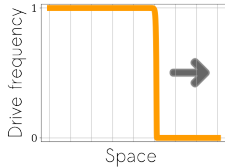
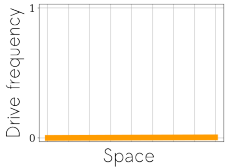
s value	$0 < s < 1/2$	$1/2 < s < 1$
Stability	Monostable	Degenerate case
Speed	$v = 2\sqrt{1 - 2s}$	No wave
Wave	Pulled wave	
Invasion	<p>Drive invasion</p> 	<p>Gene drive clearance</p> 

Table 1.3: When $r = 0$: traveling waves under system (1.32).

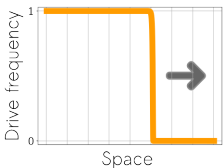
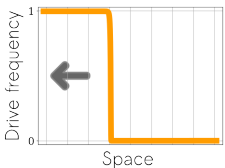
s value	$0 < s \lesssim 0.35$	$0.35 \lesssim s < 1/2$	$1/2 < s \lesssim 0.70$	$0.70 \lesssim s < 1$
Stability	Monostable		Bistable	
Speed	$v = 2\sqrt{1 - 2s}$	$v > 2\sqrt{1 - 2s}$	$v > 0$	$v < 0$
Wave	Pulled wave	Pushed wave		
Invasion	<p>Drive invasion</p> 			<p>Wild-type invasion</p> 

Table 1.4: When $r = +\infty$: traveling waves under equation (1.34).

Tables 1.3 and 1.4 show that the demographic parameter r can have a major influence on the drive spread dynamics, and especially on intermediate values of the fitness cost s . It is however rarely introduced in the gene drive modelling studies.

For $1/2 < s \lesssim 0.70$, the value of the intrinsic growth rate r makes the difference between a threshold-dependent drive invasion (large values of r) and the decay of the drive allele uniformly in space leaving only wild-type individuals in the environment (small values of r). These two drastically different outcomes illustrate the importance of taking demography into account. On the one hand, threshold-dependent drive invasions are often recommended, considered more socially responsible than threshold independent drive invasions, as they could potentially be localised and reversible [204, 134]. On the other hand, the systematic decay of gene drive alleles means the failure of drive invasion. This important result is generalised with different s threshold values in Chapter 2 for a partial conversion occurring in the zygote or in the germline. Note that models assuming a partial conversion, do not necessary conclude on a bistable case for the considered values of s : if $(1 - c)(1 - h) > \frac{1}{2}$ (conversion in the zygote) or if $h < \frac{1}{2}$ (conversion in the germline), we instead observe coexistence.

For s small ($s \lesssim 0.35$), we do not observe a significant impact of r in tables 1.3 and 1.4. The

drive wave is pulled, traveling at a speed $2\sqrt{1-2s}$ and there exists only one stable final state: the drive always invades. This result is generalised with different s -threshold values for partial conversion occurring in the zygote, with speed:

$$2\sqrt{2c(1-s) + (1-sh)(1-c) - 1}, \quad (1.35)$$

and for partial conversion occurring in the germline with speed:

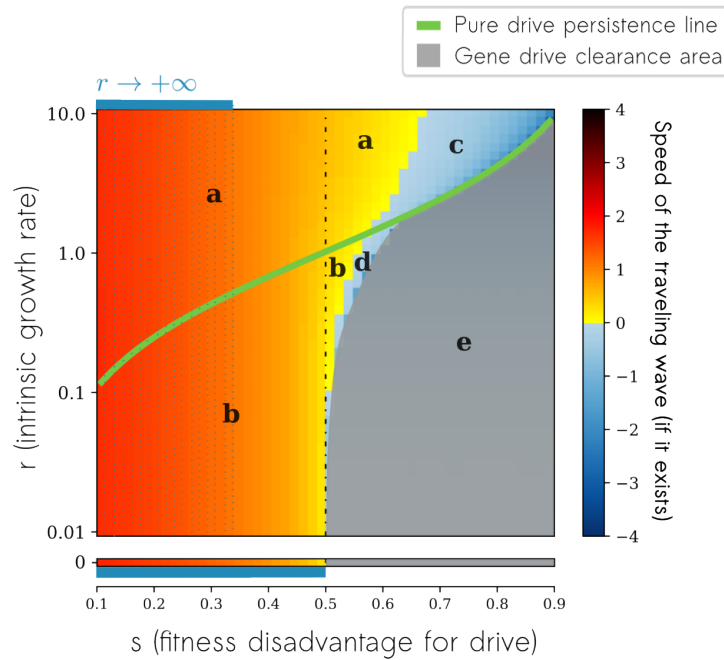
$$2\sqrt{(1-sh)(1+c) - 1}. \quad (1.36)$$

Note that speed (1.36) only exists for $(1-sh)(1+c) > 1$, which is the necessary condition to have a strictly positive drive alleles production at the front of the wave in case of drive invasion. To understand why, first note that the density of drive alleles is very low at the front of the wave. Therefore, we can assume that at least one parent in each couple formed at the front of the wave has a genotype WW . Consequently, the offspring carrying a drive allele are necessarily heterozygotes: in the front of the wave, the production of drive alleles only relies on the heterozygotes. These heterozygotes have a fitness of $(1-sh)$ and produce drive alleles at rate $(1+c)$: therefore for a drive invasion to be possible, the production rate $(1-sh)(1+c)$ of drive alleles should be above the rate 1 at which they disappear. The higher the production rate is the faster the wave moves. A similar reasoning holds for speed (1.35) when conversion occurs in the zygote, except that a heterozygous egg will become a homozygous drive adult with probability c and a heterozygous adult with probability $1-c$.

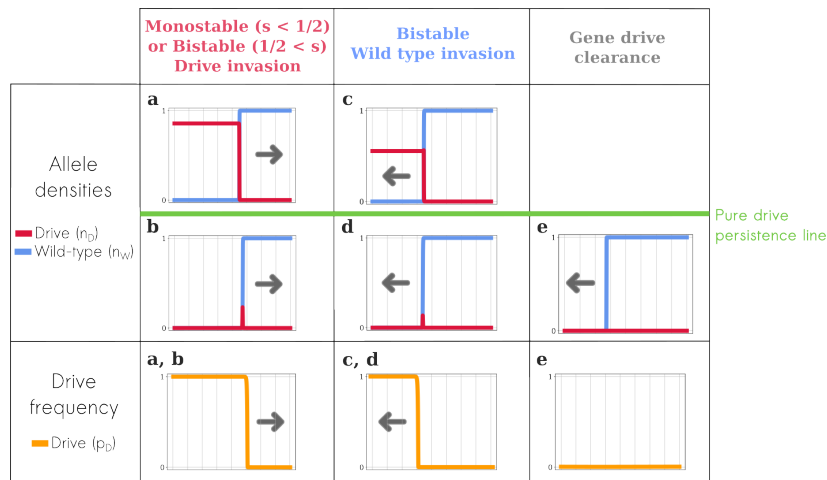
Noticeably, these speed values are always independent of r ; the movement induced by the few drive individuals at the front of the wave is not faster if the intrinsic growth rate is higher. A small fitness cost seems to provide sufficient reproductive success for them to determine the dynamics of the invasion.

To investigate if this conclusion holds for intermediate values of r , we compute numerically the speed of the traveling wave (if it exists) for a large range of s and r values (see Figure 1.21). When the drive invades, the speed is positive (in yellow-orange) and when the wild type invades, the speed is negative (in blue). The vertical level lines for small values of s seem to verify our intuition of a wave traveling at speed $2\sqrt{1-2s}$ whatever the value of r is.

In this figure, we also indicate in a table the dynamics observed. We computed the pure persistence line $r = \frac{1-s}{s}$: a line below which in case of a drive invasion there are no individuals left after the wave has passed (eradication drive). Above the line, the population persists (suppression or replacement drive). A similar line can be computed in case of coexistence for partial conversion occurring in the zygote or in the germline, see Chapter 2.



(a) Heatmap representing the speed of the wave for $c = 1$ when conversion occurs in the zygote.



(b) Illustrations of the shape of the wave (allelic densities and drive frequency) for each corresponding case in the heatmap.

Figure 1.21: (A) Heatmap representing the speed of the waves for $c = 1$ when conversion occurs in the zygote. When the drive invades the population, the speed is positive (in yellow-orange). On the contrary, when the wild-type invades the population, the speed is negative (in blue). Numerically (and analytically for $r \rightarrow +\infty$), we observe that when $s < \frac{1}{2}$ the system is monostable and the drive always invades. Otherwise when $\frac{1}{2} < s$, we observe a bistable system and the invasion type (drive or wild-type) depends on the initial condition for $r > \frac{1-s}{s}$ while we observe gene drive clearance for $r < \frac{1-s}{s}$. The turquoise horizontal lines at the bottom and at the top of the heatmap indicate the theoretical values of s for which we know that the wave travels at speed $2\sqrt{1-2s}$ when $r = +\infty$ or $r = 0$. When $0 < s \lesssim 0.35$, the level lines are apparently vertical: this is in agreement with the intuition that the wave travels at speed $2\sqrt{1-2s}$ for any $r > 0$. Below the pure drive persistence line (light green), a drive invasion leads to the population eradication (B) Shape of the wave for each case indicated by a letter in the heatmap above. The position of the graphs in the table reflects the position in the heatmap with respect to the pure drive persistence line.

Questioning modelling choices on population dynamics

In Chapter 3, I question several modelling choices made in Chapter 2. More specifically, I study variations of Model (1.21) with conversion in the germline. I consider various population dynamics assumptions based on the biology of the species. Note that the fitness cost reducing the fecundity of a drive individual, requires separating the birth and the death terms in the equations.

For large population densities, if the maximum capacity of the environment is reached ($K = 1$ in our models), the rarefaction of the resources is such that the population can not increase in density anymore, as we considered a logistic growth. This logistic density dependence can be modeled in two ways: either the number of deaths increases sufficiently to compensate the number of births, or the fecundity decreases leading to fewer births, just enough to compensate the number of deaths. In Model (1.21), the second option (density dependence constrain on births) was chosen, however the first (density dependence constrain on deaths) might make more sense in certain biological contexts.

For small population densities, growth can also be altered in a way that the per capita growth decreases as population density declines. It is a well-known biological effect called the *Allee effect* [146], and this might result from inbreeding depression or difficulties to find a mate when population density is low [60]. It is frequently observed in the wild; for instance some mosquito species suffer inbreeding depression at low population density [10, 14, 189], however it is not often taken into account in the models. The Allee effect is another density dependence constraint and has to be combined to the previous one in the equation. Consequently, the Allee effect will impact the births or the deaths depending on the modelling choice above.

We denote the different models according to the biological assumptions taken into account. Model (1.21) with a logistic density dependence on the birth term and no Allee effect is called Model \mathcal{BN} . The equivalent version of this model with Allee effect is called Model \mathcal{BA} . Models with a logistic density dependence on the death term are called Model \mathcal{DN} (without Allee effect) and Model \mathcal{DA} (with Allee effect). In equations (1.38) and (1.40), $a \in [-1, 1]$ is the Allee threshold.

Model \mathcal{BN}

$$\begin{cases} \partial_t n_D - \partial_{xx}^2 n_D = n_D \left[\frac{r(1-n) + 1}{n} \left[(1-s)n_D + (1-sh)(1+c)n_W \right] - 1 \right], \\ \partial_t n_W - \partial_{xx}^2 n_W = n_W \left[\frac{r(1-n) + 1}{n} \left[n_W + (1-sh)(1-c)n_D \right] - 1 \right]. \end{cases} \quad (1.37)$$

Model \mathcal{BA}

$$\begin{cases} \partial_t n_D - \partial_{xx}^2 n_D = n_D \left[\frac{\max(r(1-n)(n-a) + 1, 0)}{n} \left[(1-s)n_D + (1-sh)(1+c)n_W \right] - 1 \right], \\ \partial_t n_W - \partial_{xx}^2 n_W = n_W \left[\frac{\max(r(1-n)(n-a) + 1, 0)}{n} \left[n_W + (1-sh)(1-c)n_D \right] - 1 \right]. \end{cases} \quad (1.38)$$

Model \mathcal{DN}

$$\begin{cases} \partial_t n_D - \partial_{xx}^2 n_D = n_D \left[\frac{r+1}{n} \left[(1-s)n_D + (1-sh)(1+c)n_W \right] - (rn+1) \right], \\ \partial_t n_W - \partial_{xx}^2 n_W = n_W \left[\frac{r+1}{n} \left[n_W + (1-sh)(1-c)n_D \right] - (rn+1) \right]. \end{cases} \quad (1.39)$$

Model \mathcal{DA}

$$\begin{cases} \partial_t n_D - \partial_{xx}^2 n_D = n_D \left[\frac{r+1}{n} \left[(1-s)n_D + (1-sh)(1+c)n_W \right] - (r(n-1)(n-a) + r + 1) \right], \\ \partial_t n_W - \partial_{xx}^2 n_W = n_W \left[\frac{r+1}{n} \left[n_W + (1-sh)(1-c)n_D \right] - (r(n-1)(n-a) + r + 1) \right]. \end{cases} \quad (1.40)$$

In Figure 1.22, we compute numerically the speed of the traveling wave for various values of s and r . We arbitrarily choose $c = 0.85$ and $h = 0.9$. When the drive invades the speed is positive (in yellow-orange) and when the wild-type invades the speed is negative (in blue). In our analysis, we focus on the yellow-orange section, in which the drive invades the environment. In Figure 1.23, we plot the final population sizes for the same values of s and r .

As in the previous section, we observe numerically and show analytically for $r \rightarrow +\infty$ that the system is: i) monostable for $s \in (0, s_{2,g})$, consequently, the drive always invades and ii) bistable for $s \in (s_{2,g}, s_1)$, consequently, the drive invades at the condition that enough drive individuals have been introduced in the environment. The initial conditions used in the simulations are always the following: the left half of the domain is full of drive homozygotes and the right half of the domain is full of wild-type homozygotes.

In the pure drive eradication area, the drive invasion leads to the complete extinction of the population (according to this deterministic model i.e without the possibility of stochastic reemerging events, see Chapter 4). In the pure drive persistence area, the drive invasion leads to a reduction of the population size but this population (fully drive) remains indefinitely in the environment. Finally, in the pure drive bistable area, the persistence of the final population depends on the nature of the drive introduction: for a dense enough introduction, the final population will persist indefinitely in the environment, otherwise it will not. The borders of these areas were all determined analytically in Chapter 3. Note that in my simulations, the population always persists in the pure drive bistable area as the drive was initially introduced at the maximum carrying capacity (largest possible density) in one half of the domain.

In Figure 1.22, the Allee effect seems to widen the range of s (fitness disadvantage for drive) and r (intrinsic growth rate) leading to population eradication after a drive invasion. This first impression is confirmed analytically in Chapter 3 as well as the fact that it helps reduce the final population density in case of persistence (Figure 1.23). For $s > s_{2,g}$ and r large enough however, the Allee effect might prevent the invasion of a threshold-dependent drive, leading instead to a wild-type invasion (yellow-orange areas becoming blue when adding Allee effect in the model, in Figure 1.22). All these influences are accentuated when the effect gets stronger (for larger values of a). In conclusion, the Allee effect is a non negligible force which might help eradicate or suppress the population. However, it is important to keep in mind that this effect might also reduce the range of s (fitness disadvantage for drive) and r (intrinsic growth rate) leading to a threshold-dependent drive invasion.

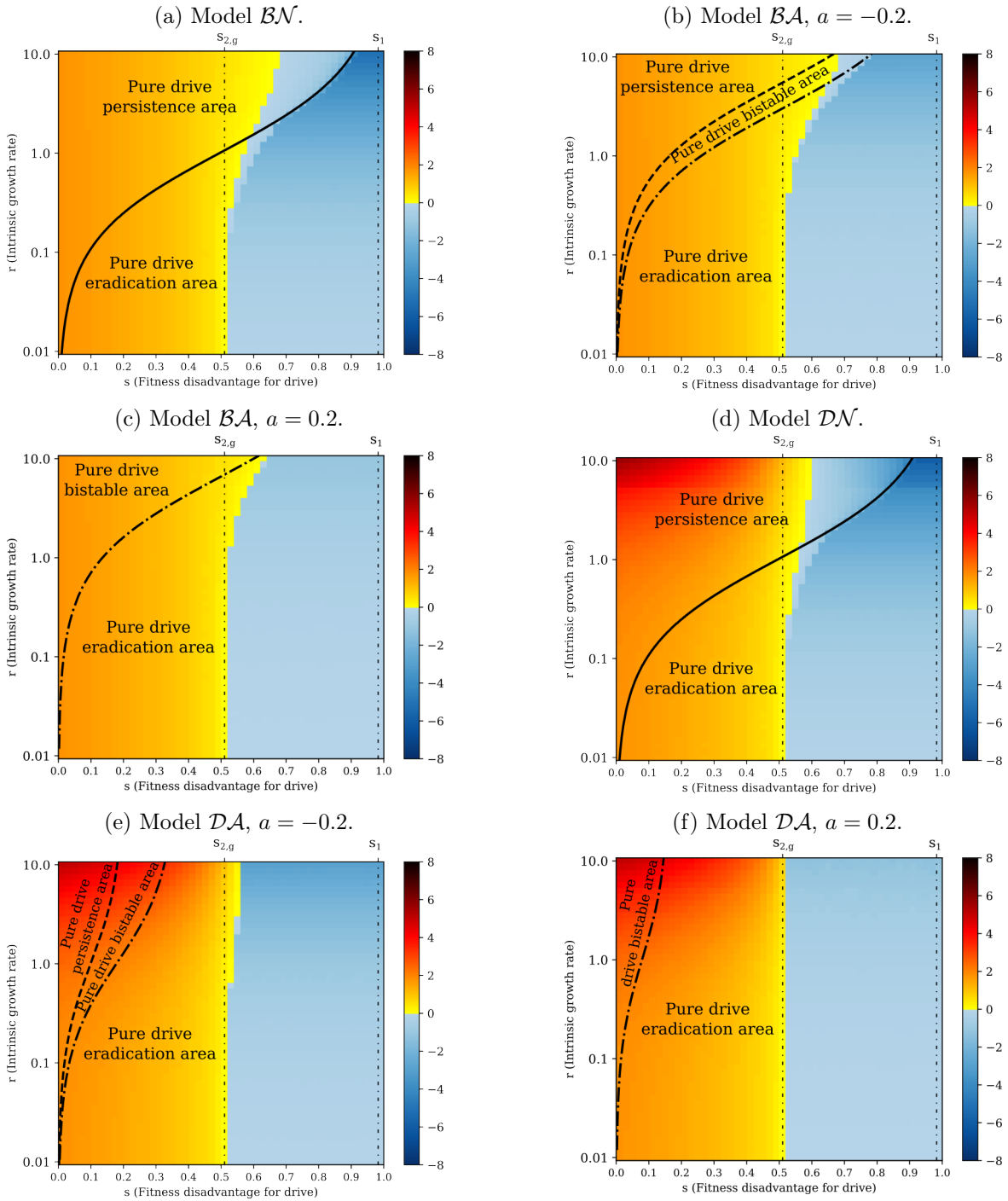


Figure 1.22: Speed of the wave when $c = 0.85, h = 0.9$. When the drive invades the population, the speed is positive (in yellow-orange). On the contrary, when the wild-type invades the population, the speed is negative (in blue).

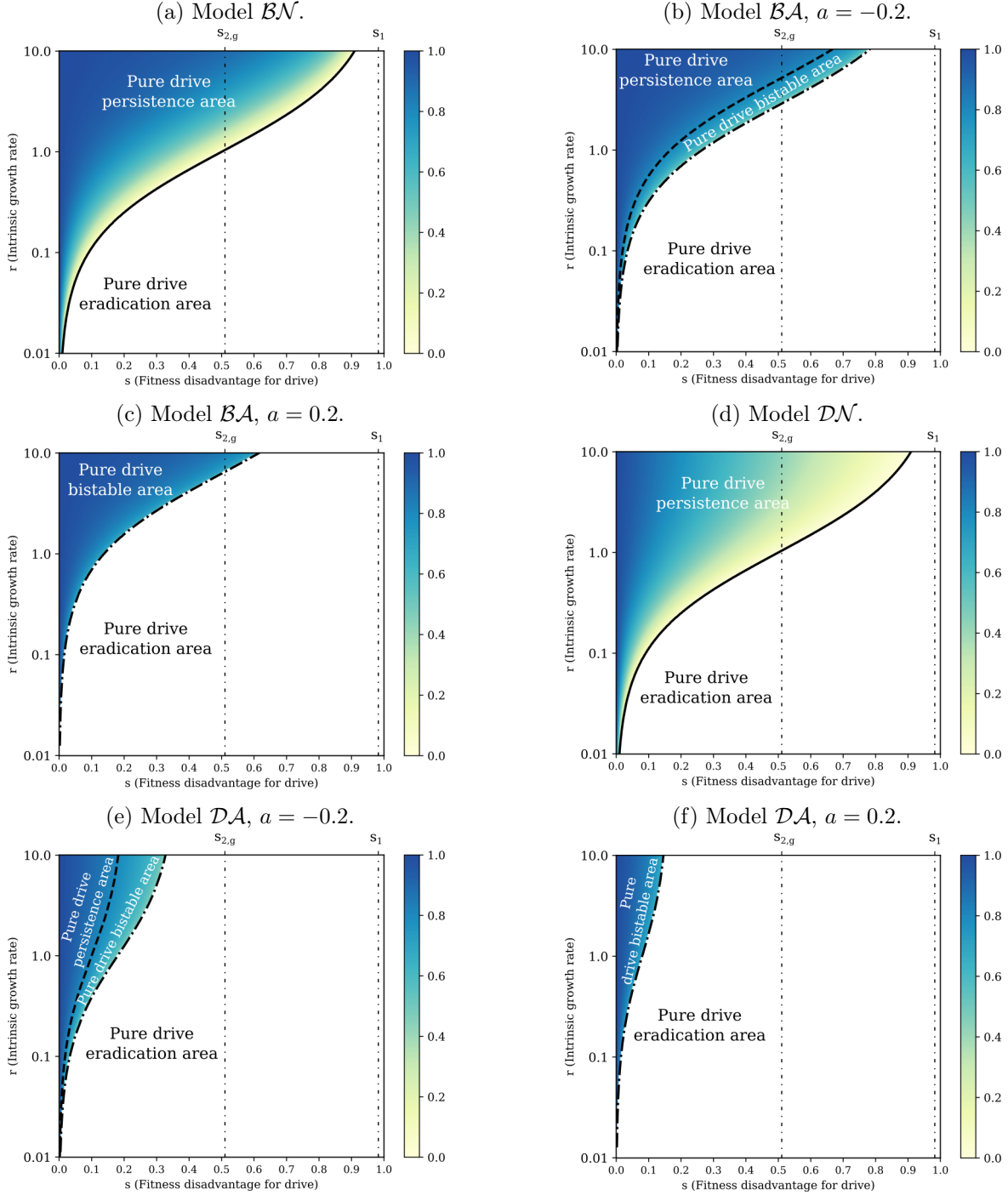


Figure 1.23: Density of the final population, with parameters $c = 0.85, h = 0.9$.

In Chapter 2 we showed that for a small enough s value in Model \mathcal{BN} , the speed is independent of the intrinsic growth rate r . When the conversion occurs in the germline, this speed value is given by:

$$2\sqrt{(1 - sh)(1 + c) - 1}. \quad (1.41)$$

The same conclusion holds for Model \mathcal{BA} however it is not true for Models \mathcal{DN} and \mathcal{DA} : in Figure

1.22, we observe that the speed increases with r for small values of s . In Chapter 3, I show that in Models \mathcal{DN} and \mathcal{DA} , the speed for small enough s values is given by:

$$2\sqrt{(1+r) [(1-sh)(1+c) - 1]}. \quad (1.42)$$

Consequently, the drive invasion for small s values is $\sqrt{r+1}$ times faster for a logistic density dependence targeting the deaths instead of the births. At the front of the wave, the density of the population, which is composed nearly only of wild-type individuals, reaches the maximum carrying capacity. Consequently, the logistic density-dependent constraint prohibits any increase in the population density either by limiting the births so that they do not exceed the deaths (Models \mathcal{BN} and \mathcal{BA}), or by increasing the death rate so that it compensates the births (Models \mathcal{DN} and \mathcal{DA}). As a result, the turnover rate is greater in models \mathcal{DN} and \mathcal{DA} , which induces a faster invasion as the wave movement is mainly driven by reproduction. This conclusion holds for a drive fitness cost targeting fecundity i.e. reducing the birth rate. However, if the drive was affecting a different fitness component, for instance decreasing survival, this could have led to a different outcome [186].

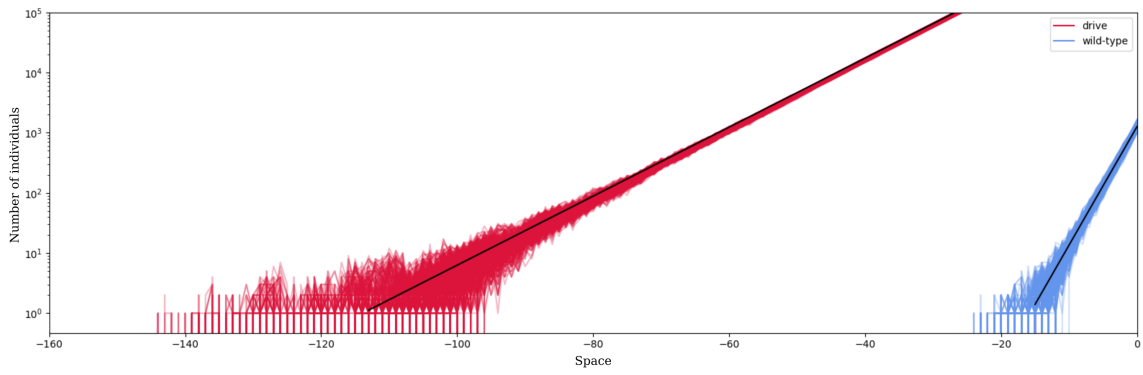
To conclude, in this chapter I have generalised some results of Chapter 2 by taking into account other realistic biological assumptions on population dynamics. According to the modelling study, the Allee effect helps eradicate or reduce in density the targeted population, however it might also lead to the failure of threshold-dependent drive invasions. In case of a successful drive invasion, the speed of the wave is $\sqrt{r+1}$ times faster for small s values if the logistic density-dependent constraint targets the deaths instead of the births (with r being the intrinsic growth rate). Interestingly this conclusion puts into perspective the previous result of Chapter 2 stating that the speed value was independent of r for small enough values of s : this result seems only true for a logistic density-dependent constraint targeting the births and not the deaths.

A larger carrying capacity and a fitter drive decrease the chances of chasing

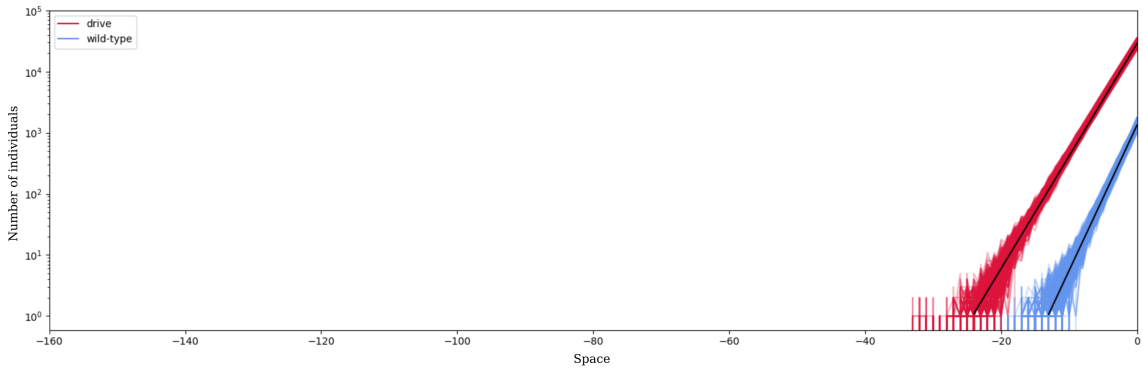
In Chapter 4, I investigate a major challenge faced by eradication drives: chasing events i.e. wild-type recolonising events of areas previously cleared by the drive. Such events of recolonisation might delay or even prevent the eradication of the population [46]. In this chapter, I study these small population dynamics with the stochastic model illustrated in Figure 1.18. This model is population-based: it follows the number of individuals in each spatial site instead of the position of each living individual, allowing to simulate very large populations.

I focus on how chasing dynamics might be influenced by the local carrying capacity K and the drive fitness cost s . In Figure 1.24, I plot the back of the stochastic wave for $s \in \{0, 3, 0.7\}$ and $K \in \{10^5, 10^8\}$: we observe a chasing event for $s = 0.7$ and $K = 10^5$. I illustrate the shape of the full eradication wave in Figure 1.25 in arithmetic scale, without stochastic fluctuations.

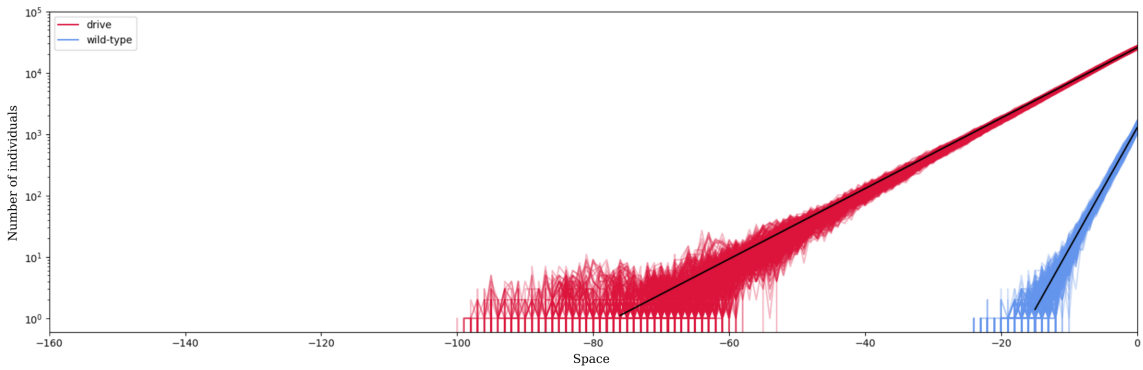
In Figure 1.24, we observe that the local carrying capacity K does not influence the slope (confirmed analytically in Chapter 4). However it seems to bring the wild-type and drive curves closer together when it decreases. Thanks to stochastic fluctuations, this might result in the last wild-type individual being positioned on the left of the last drive individual at some point, a necessary (but not sufficient) condition for chasing to occur. We also observe that the drive slope at the back of the wave increase with the drive fitness cost s . In other words, the fitter the drive is, the longer it will stay in an empty environment after the wave has passed which might reduce chasing probabilities.



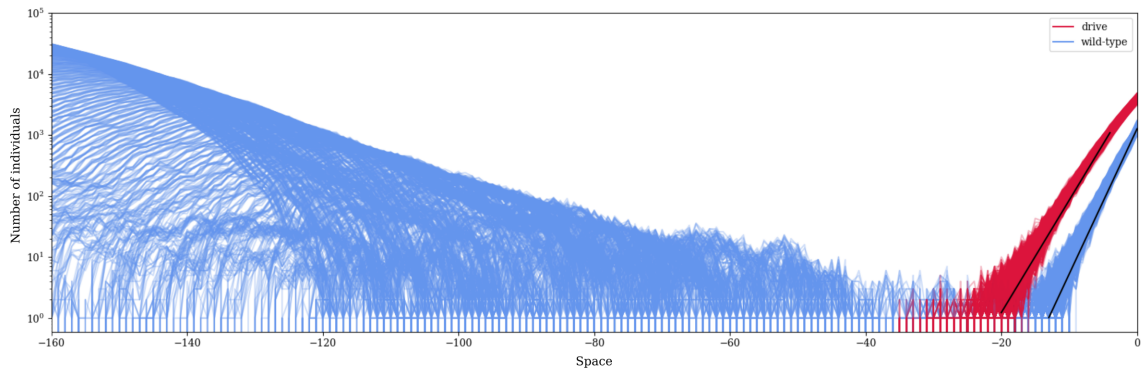
(a) $K = 10^8, s = 0.3$



(b) $K = 10^8, s = 0.7$

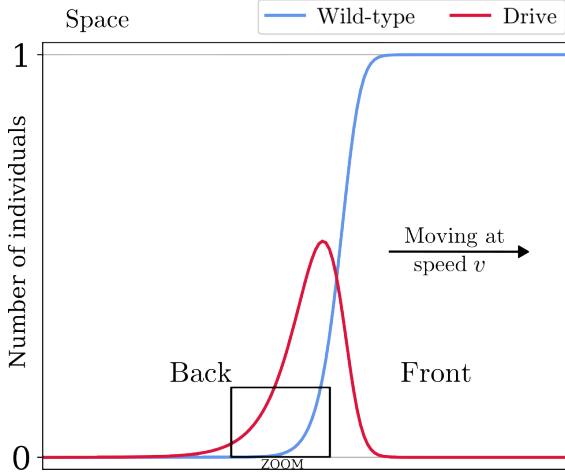


(c) $K = 10^5, s = 0.3$

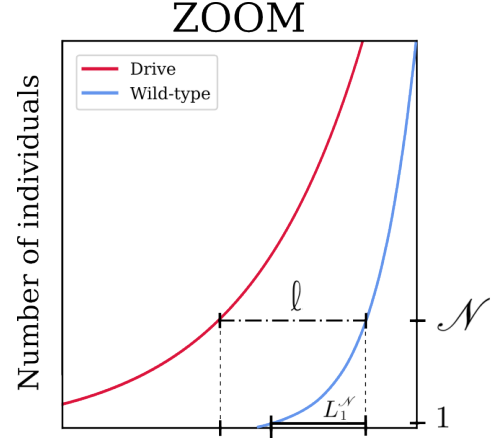


(d) $K = 10^5, s = 0.7$ - Chasing event

Figure 1.24: Back of the drive wave (in red) and the wild-type wave (in blue) in log scale. I superimpose multiple curves corresponding to different times to observe the stochastic variations. The back of each wave can be approximated by an exponential function (in black), see Chapter 4 for calculations. We observe a chasing event for $s = 0.7$ and $K = 10^5$.



(a) Illustration of the wave



(b) Zoom at the back of the wave

Figure 1.25: Schematic illustration of an eradication drive wave in arithmetic scale. If \mathcal{N} is large enough and ℓ is always strictly larger than $L_1^{\mathcal{N}}$, then chasing is very unlikely because the last wild-type individual is always surrounded by a large number of drive individuals.

Quantifying the probability that at least one chasing event occurs in a determined period of time is a very complex problem. However, I posit that if the last wild-type individual is always surrounded by a large enough number of drive individuals denoted \mathcal{N} , then chasing is very unlikely. With ℓ and $L_1^{\mathcal{N}}$ the distances defined in Figure 1.25, this is verified only if we always have $\ell > L_1^{\mathcal{N}}$.

If \mathcal{N} is large enough, the distance ℓ is nearly deterministic: we can easily approximate it with a deterministic simulation. In this work, I chose $\mathcal{N} = 100$ individuals based on Figure 1.24 (low stochastic variations). However the distance L_1^{100} is more complex to determine.

A first approach is based on the exponential increase at the back of the wave. Let \mathcal{N} be reached at an arbitrary $z = 0$, as in Figure 1.26. In an exponential profile, the distance $L_\eta^{\mathcal{N}}$ between the two spatial sites with η and \mathcal{N} individuals is given by :

$$\mathcal{N}e^{\lambda(-L_\eta^{\mathcal{N}})} = \eta \iff L_\eta^{\mathcal{N}} = \frac{\log(\mathcal{N}) - \log(\eta)}{\lambda} \quad (1.43)$$

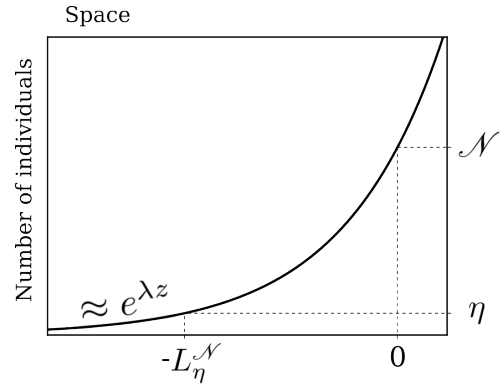


Figure 1.26: Schematic illustration at the back of the wild-type wave.

The position of the last spatial site with more than $\mathcal{N} = 100$ wild-type individuals is almost deterministic. However the position of the last wild-type individual ($\eta = 1$) is on the contrary, highly stochastic. The exponential profile at the back of the wave gives us a rough approximation of L_1^{100} , but we are interested in characterising the "worst case", the longest L_1^{100} in a determined period of time, the further away the wild-type individual can be found at the back of the wave.

To characterise extreme values of L_1^{100} , I study in Chapter 2 a spatial Galton-Watson process in a bounded domain with appropriate initialisation (the exponential approximation stated before). As this process models an isolated population, I consider the wild-type dynamics when the drive proportion is one, an approximation of the condition at the back of the wave. I accumulate numerical pieces of evidence that a specific time of extinction in this spatial Galton-Watson process, multiplied by the speed of the traveling wave, is fitting very well the numerical distribution of L_1 (see Chapter 2 for details). However I was not able to characterise this distribution analytically; this is left for future work.

In Figure 1.27, I superimpose this distribution at the back of the wave for two values of s (0.3 and 0.7). It seems clear that there is no chance of chasing events to occur when $s = 0.3$ in a realisable time window: the further a wild-type individual might be is a spatial site containing more than 10^4 drive individuals. This statement is however less clear when $s = 0.7$ where the end of the light green histogram (distribution of last wild-type individual position) mixes with the drive last individuals: we cannot guarantee the absence of chasing at all in this configuration.

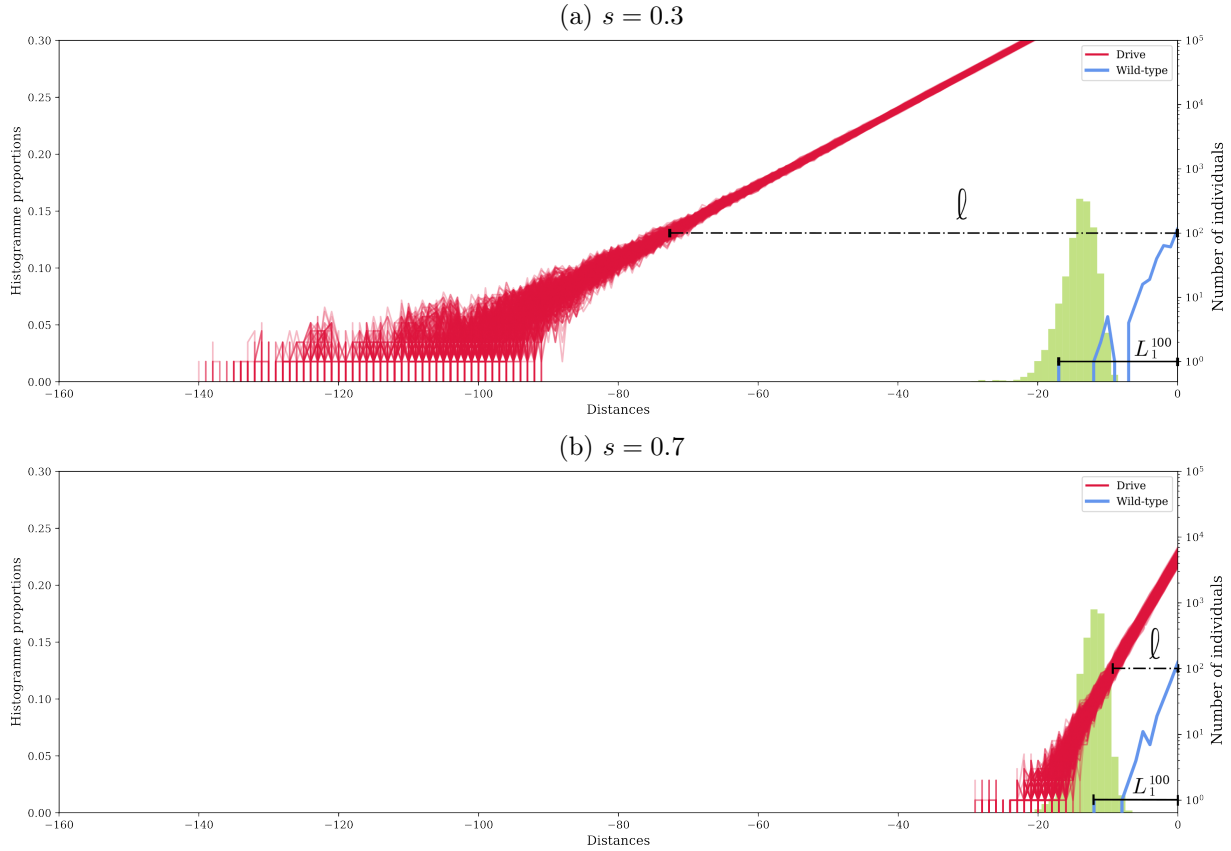


Figure 1.27: Superimposition of the distribution of the distance L_1^{100} (in light green), on the wild-type (in blue) and drive (in red) wave, when $K = 10^8$ in log scale. Several drive waves are plotted, whereas only one wild-type wave appears in the graph for clarity. The last individual of this wild-type wave determine the distance L_1^{100} : it is one realisation of the light green histogram. When $s = 0.3$ chasing seems very unlikely as the furthest a wild-type individual might be is a spatial site containing more than 10^4 drive individuals. When $s = 0.7$ however, chasing dynamics might appear as the histogram mixes with the end of the drive wave: a wild-type individual could possibly be beyond the drive last individual at one point and recolonise the empty area. Parameters are $c = 0.9$, $h = 0.4$, $r = 0.1$.

I also numerically compute the chasing probability for different values of s the drive fitness cost and K the local carrying capacity, over 100 replicates, each of 1000 units of time in Figure 1.28. This probability increases with s the drive fitness cost and decreases with K the local carrying capacity.

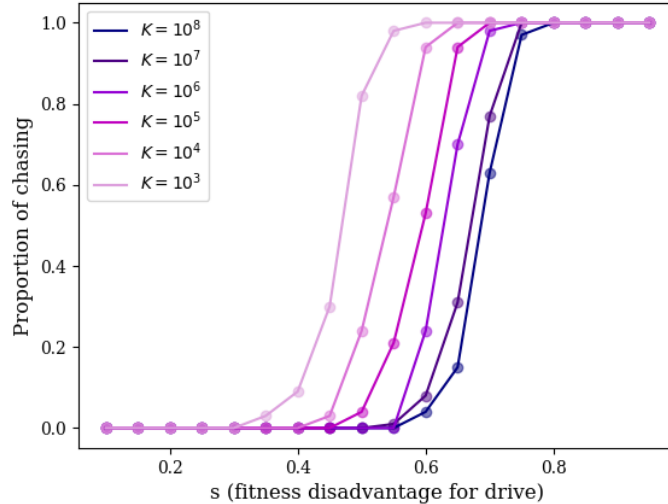


Figure 1.28: Chasing proportions as a function of the fitness cost for drive s , and the local carrying capacity K . For each point, we ran the simulation 100 times and computed the proportion of simulations where at least one chasing event occurred.

To conclude, this study paves the way for further analysis of chasing events at the back of eradication traveling waves. It establish a connection between i) the absence of chasing events in a realisable time window, ii) the position of the last wild-type in the wave, and iii) a specific extinction time relative to a spatial Galton-Watson process. Quantifying analytically this last element would enable us to determine conditions under which chasing is very unlikely: this problem is left open.

In the biological point of view, my conclusions are the following: the number of chasing events decreases as the local carrying capacity gets larger. A higher fitness for drive individuals also reduces the likelihood of chasing. Noticeably for a given local carrying capacity K , the transition between very low ($< 10\%$) and very high ($> 90\%$) chances of chasing within 1000 units when the fitness cost s varies, is relatively restricted: these two extreme conditions can be reached at s values within a range of 0.2.

Space discretisation in modelling: how it impacts drive confinement

In Chapter 5, I contributed to an ongoing project modelling the spread of a daisy quorum gene drive construct. Designed to spread locally and then stay localised in a spatial area without overflowing in non target regions, the daisy quorum drive is a combination of temporal and density-dependent confinement mechanisms (details in section 1.2.8).

The construct we studied is illustrated in Figure 1.29. It is segregated on four independent loci in the genome, all carrying fitness costs (1.29a). The fitness costs are multiplicative: if the individual carries at least an A modified allele the fitness is multiplied by $1 - s_d$, if it carries at least a B modified

allele, the fitness is again multiplied by $1 - s_d$ and if it carries at least an C, or D, or both C and D modified alleles, the fitness is multiplied by $1 - s_p$. The daisy chain takes place on the first two chromosomes: a transgenic allele allows the gene conversion of the following allele in the chain, A targets B, and B targets by C and D (1.29b). An underdominance mechanism, more precisely a two-locus toxin-antidote system, is introduced in the last two chromosomes: if the genome contains a transgenic allele on one chromosome but none on the other one (C only, or D only), the individual will suffer an additional fitness cost s_t (1.29a).

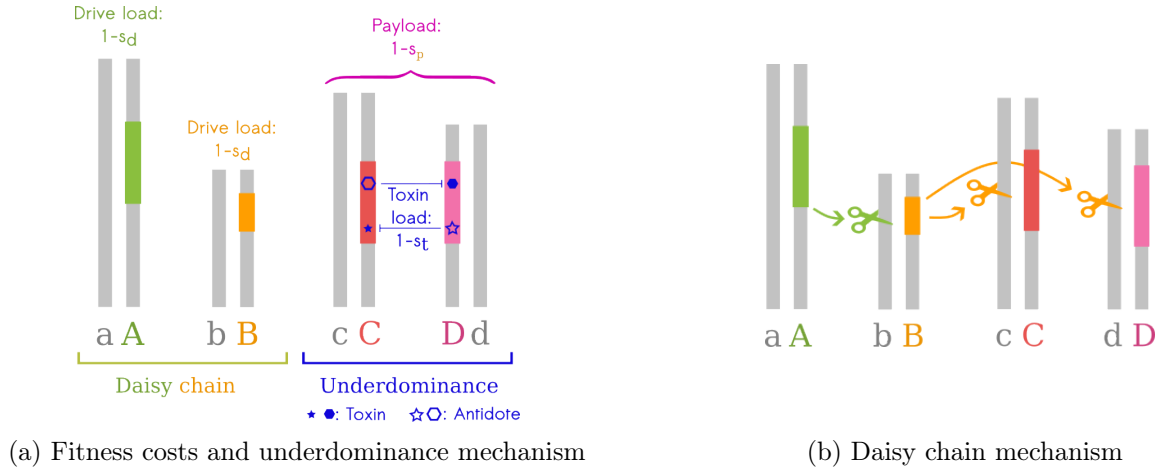


Figure 1.29: Details of the daisy quorum drive studied in Chapter 3.

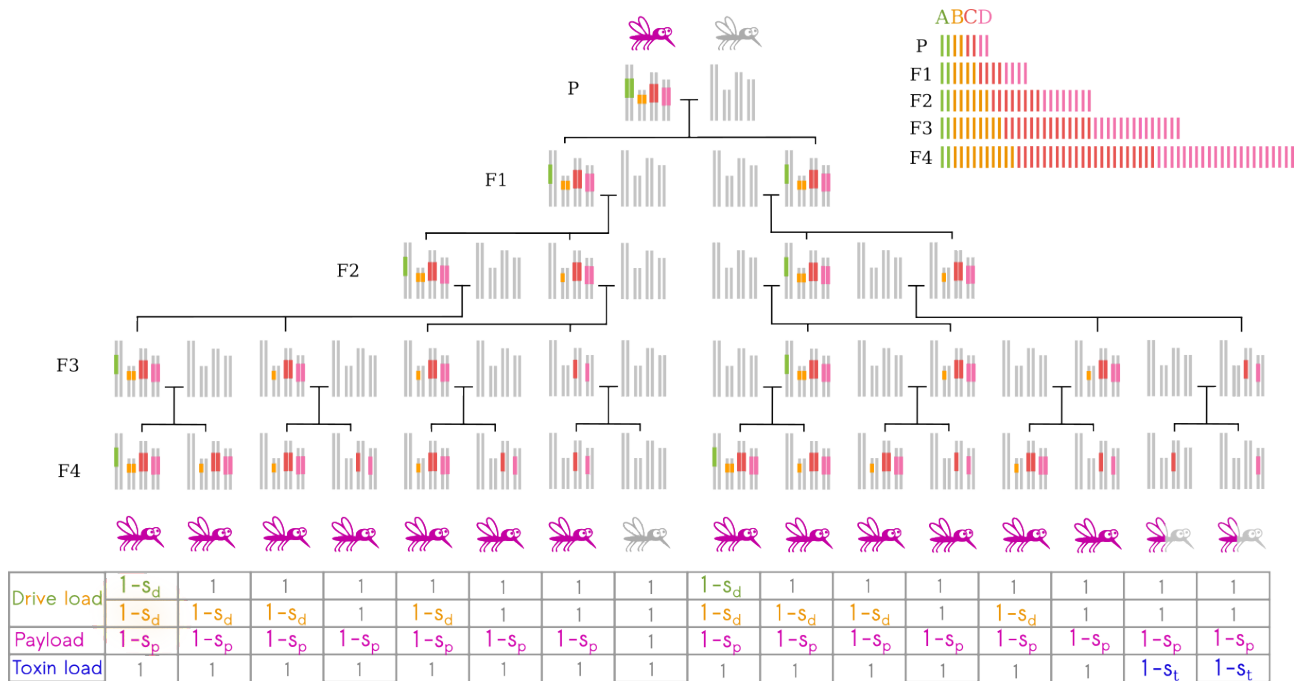


Figure 1.30: Illustration of the daisy quorum propagation and fitness computation. Wild-type mosquitoes are in grey while gene drive mosquitoes carrying both C and D alleles are in pink. Mosquitoes carrying only one type of modified alleles C or D, are in grey and pink.

In Figure 1.29, I illustrate the spread of the daisy quorum drive and I indicate the fitness costs associated with each genotype at generation F4. The number of modified allele A remains constant at each generation as there exists no previous chromosome in the chain to allow this allele duplication. In fact it will even be removed due to its fitness cost s_d , not taken into account in Figure 1.29. The number of modified alleles B however first increases through the generations, and the number of modified alleles C and D increases even faster. As detailed in the temporal confinement part (section 1.2.8), this dynamic is however doomed to vanish over time because of the genetic mixing and the fitness cost of each transgenic allele.

The modelling results obtained by Frederik J.H. de Haas and Sarah P. Otto in a discrete spatial domain were promising: C and D alleles spread for a while and then remain spatially restricted despite migration events, as long as the drive is built with a light enough toxin load (s_t) and a heavy enough payload fitness cost (s_p). If the fitness cost causes local population eradication, wildtype individuals from neighbouring patches will eventually recolonise the area. By contrast, constructs that involve population suppression or population replacement persist for longer because the underdominance mechanism acts as a safeguard. Thus this engineered construct is most suited for population replacement or suppression, rather than population eradication (see Chapter 5 for details on these results).

One question remained, however: do these conclusions hold in continuous environments? Depending on the landscape, the environment might be more or less discrete, for instance islands or forest patches are great examples of separated spatial areas with possible migration events between them. But if we consider mosquitoes flying all over a country and even across international borders, continuous environment may be a more appropriate modelling assumption.

I showed that this type of confinement is impossible in a continuous environment. Indeed the slightest advantage from one genotype on another induces a movement when the spatial sites are sufficiently close to each other (in a continuous environment, the spatial step tends to zero). Unless there exists a perfect balance between the forces involved, the wave travels at a strictly positive (or negative if it goes in the other direction) speed. Figure 1.31 shows the influence of the payload fitness cost s_p over the speed of the daisy quorum traveling wave. We observe that for each value of s_p , the absolute speed value reduces as the spatial step increases. Above a critical step size, the wave is stopped; if there exists too much distance between two spatial sites, the migrants are not numerous enough to trigger an invasion. Note that this figure does not take into account the initial propagation "boost" given by the daisy chain, so that in reality, the wave can move for a while before being stopped. In heterogeneous environments with various spatial step sizes, we also observe that the wave travels until it faces a spatial step larger than the critical size.

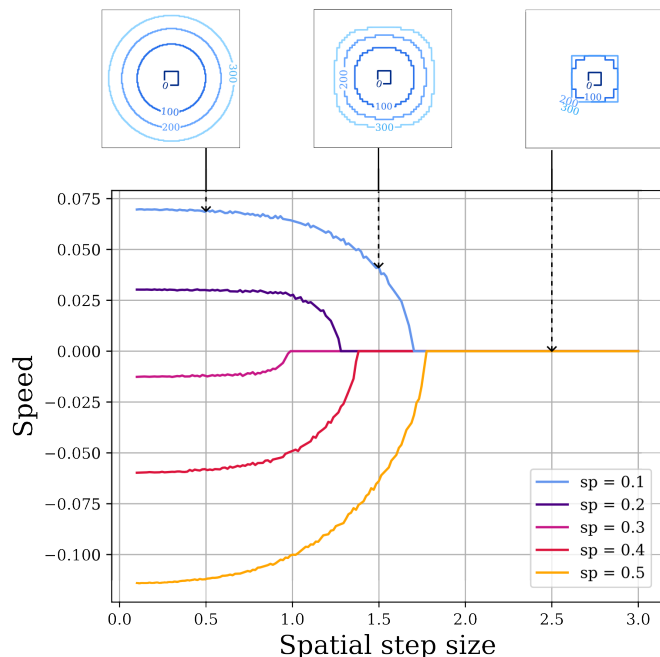


Figure 1.31: Asymptotic wave speed for a daisy quorum drive as a function of the spatial step size for different values of the payload fitness cost s_p across a two-dimensional area. The three panels above the graph show the convex hull containing 80% of the population for either the C or D allele at $t = 0$ (dark blue contour), $t = 100$, $t = 200$ and $t = 300$ (light blue), with $s_p = 0.1$. The diffusion rate is 1.

Consequently, and according to these modelling studies, a daisy quorum construct is not a convincing candidate for spatial confinement in continuous environment, or in discrete environments with small enough spatial steps (all spatial step values below the critical size). This result can in fact be generalised to any underdominant construct.

We investigate the problem based on a classical underdominance equation studied by Barton in [15]:

$$\partial_t p = \mathcal{D} \partial_{xx}^2 p + sp(1-p)(2p-1) + \alpha sp(1-p). \quad (1.44)$$

This equation considers two alleles, P and Q, and the following fitnesses for each genotype: $f_{PP} = 1 + 2\alpha s$, $f_{PQ} = 1 - (1 - \alpha)s$ and $f_{QQ} = 1$, with $\alpha \in [0, 1]$ and $s \in [0, 1]$. No gene drive is involved. Note that the underdominance criteria is verified since $f_{PQ} \leq f_{QQ} \leq f_{PP}$.

This equation holds for s small: initially, the formula contains a numerator $1 - 2ps(1 - p - \alpha)$ which is approximated by one when $s \ll 1$. Also note that, as it is usually the case in classical population genetics frameworks, this model does not take into account the spatial variations in the population size and the resulting demographic flux (Section 1.3.1, spatial deterministic models).

Barton obtains a similar conclusion to ours analytically, using several approximations in the calculations: he conceptualised the fact that a traveling wave, however the strength of the disequilibrium between the two homozygote fitnesses, could be stopped if the spatial step was large enough. Taking the opposite conclusion, if there exists a disequilibrium between the forces involved, the wave will always travel for a small enough spatial step.

I computed the speed of the traveling wave resulting from equation (1.44) and plotted the results the same way as in Figure 1.31. We observe the exact same qualitative dynamics, which seems to generalise our conclusion to any underdominant construct. The dashed lines represent the predictions made by Barton about the critical spatial step above which the wave speed is zero. Unless we misunderstood an aspect of his approach, the approximations he used in his calculations might have led to an underestimation of the value, concluding that a confinement was possible where in reality it was not.

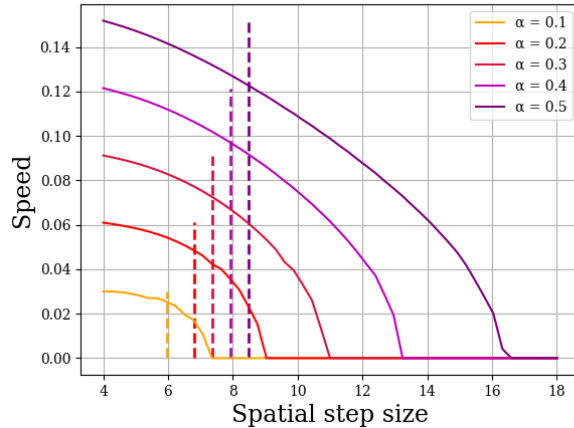


Figure 1.32: Wave speed function of the spatial step size for different values of α , with $s = 0.1$ and $\mathcal{D} = 1$. The dashed lines represent the predictions of the critical spatial step above which the wave speed is zero in [15], for each value of α .

1.3.3 Perspectives

Given that my models aim to provide relevant insights on the potential outcomes of drive release in nature, it is important to assess how their results depend on modelling choices. In this PhD thesis, I study the impact of demography and population dynamics over the drive spread, I then focus on small populations considering the effect of large carrying capacity and various fitness costs on chasing dynamics, and finally I investigate how the environment might impact the possibility to spatially confine the spread of an underdominant construct.

It would be highly relevant in future works to further improve modelling by assessing more modelling choices and study their possible combined effects. The features that can be considered mainly fall into two categories: i) investigating the characteristics of the targeted species and the effect of the genetic modification on individuals, and ii) considering a bigger picture including the ecosystem or/and other already existing strategies for population control. Realistic predictions are based on models that succeed in prioritising the features needed to be taken into account and conserve only the relevant ones depending on the case study. Indeed, if very complex models are more likely to capture the truth of a specific system, they are also more difficult to analyse or interpret. In addition, each new feature makes the construction and parameterisation of the model more difficult, leading to more uncertainty.

Among the biological characteristics of the targeted species some have already been proven to be key to the analyses, such as complex life cycles, feeding or reproductive behaviours. In mouse populations, the spread of gene drive can be limited by the polyandrous mating system [149, 148] or mate search

capabilities [25]. In mosquito populations, the plural life stages (egg, larva, pupa and adults) might influence the modelling conclusions and need to be taken into account by including corresponding age structure in models [152, 191, 48]. In bee populations, the haploid phases of the life cycle result in less powerful drives: the conditions for fixation are narrower and the spread is slower [143, 138]. It would be wise to combine our results on demography and population dynamics with these natural behaviours in order to obtain more realistic conclusions for specific systems.

Characteristics of the gene drive construct might also require modelling details. In this thesis, we assumed a constant drive fitness cost on the fecundity of adult individuals, but in reality this cost might vary in time and with environmental variations [55], target embryo survival or/and adult death rate, potentially changing the demographic dynamics [186] and affect differently males and females like for instance intentionally in some transgenic mosquitoes [17, 171, 133, 103]. In addition, unintentional side effects might have to be taken into account, such as a decline in competitiveness for transgenic individuals regarding mating or resource finding [133, 187]. If their dispersal capacity is also altered, it would be highly relevant to consider the consequences on chasing dynamics as the wild-type individuals might have more opportunities to access drive-free areas. We also did not take into account the possible emergence of resistance alleles in our studies, which can significantly alter the drive spread [18, 106, 181, 187]. More generally, research might be needed to better characterise the behaviour of transgenic individuals, although it is very difficult to reproduce realistic conditions in laboratories or outside cages. An alternative solution might be to investigate already existing natural drive systems in the wild [140].

For the sake of realism, gene drive releases have to be considered also as part of a larger environment. Species do not live in isolation, and interactions of the targeted species within its biotic and abiotic environment are of major importance. Competing species or predators can facilitate drive-based suppression [143]. In terms of risk assessment, it is essential to model and quantify the possibility of unexpected spread in non-targeted populations through hybridisation [56], or potential cascade of population dynamics initiated by rapid population declines [162]. The abiotic environment might also strongly influence drive spread. For instance, seasonality (dry or wet season) highly impacts the eradication of mosquito populations [76, 170, 171]. A spatially realistic environment with variable habitat quality or even maps of the targeted area, might be important to obtain reliable predictions, and it would be interesting to discuss our results on spatially confined gene drives in this context.

As the CRISPR-based gene drive technique is relatively recent, it is also usually envisaged in areas where other strategies for population control are already implemented. Plus there is a risk that eradication drives alone might not lead to complete extinction after having efficiently reduce the population size [136, 46]. The safest and most effective strategy would certainly be to coordinate gene drive with already existing control techniques [218] and modelling efforts in that direction are highly relevant.

Finally the cooperation of gene drive researchers with various backgrounds seems essential before any potential drive release in nature. Laboratories experiments provide information in confined and relatively small environments, mathematical models can help gain further insights at small and larger scales, but these studies have to be run in parallel with others, including risk assessment, social and ethical issues and epistemological considerations. Interactions and exchanges between these fields are essential and it would be highly relevant for modelers to develop accessible tools that enable the exploration and the clear understanding of their modelling results to all.

Chapter 2

Pulled, pushed or failed: the demographic impact of a gene drive can change the nature of its spatial spread

Léna KLÄY^{*,1}, Léo GIRARDIN², Vincent CALVEZ², Florence DÉBARRE¹

¹ Institute of Ecology and Environmental Sciences Paris (IEES Paris), Sorbonne Université, CNRS, IRD, INRAE, Université Paris Est Creteil, Université de Paris, Paris Cedex 5, France.

² Institut Camille Jordan, UMR 5208 CNRS and Université Claude Bernard Lyon 1, France

Abstract

Understanding the temporal and spatial spread of gene drive alleles – alleles that bias their own transmission – through modeling is essential before any field experiments. In this paper, we present a deterministic reaction-diffusion model describing the interplay between demographic and allelic dynamics, in a one-dimensional spatial context. We focused on the traveling wave solutions, and more specifically, on the speed of gene drive invasion (if successful). We considered various timings of gene conversion (in the zygote or in the germline) and different probabilities of gene conversion (instead of assuming 100% conversion as done in a previous work). We compared the types of propagation when the intrinsic growth rate of the population takes extreme values, either very large or very low. When it is infinitely large, the wave can be either successful or not, and, if successful, it can be either pulled or pushed, in agreement with previous studies (extended here to the case of partial conversion). In contrast, it cannot be pushed when the intrinsic growth rate is vanishing. In this case, analytical results are obtained through an insightful connection with an epidemiological SI model. We conducted extensive numerical simulations to bridge the gap between the two regimes of large and low growth rate. We conjecture that, if it is pulled in the two extreme regimes, then the wave is always pulled, and the wave speed is independent of the growth rate. This occurs for instance when the fitness cost is small enough, or when there is stable coexistence of the drive and the wild-type in the population after successful drive invasion. Our model helps delineate the conditions under which demographic dynamics can affect the spread of a gene drive.

*Corresponding author: lena.klay@sorbonne-universite.fr

Contents

2.1	Introduction	46
2.2	Methodology	47
2.2.1	Models	47
2.2.2	Setting of the problem	50
2.2.3	Glossary	51
2.3	Results	52
2.3.1	Model with perfect conversion in the zygote	52
2.3.2	Models with partial conversion	56
2.4	Discussion	67
2.5	Acknowledgements	68

2.1 Introduction

A highly accurate, cost-effective and easy-to-use technology, the CRISPR-Cas genome editing system has been favoring the development of promising innovations [121]. Among them, CRISPR-Cas9 gene drive [4], which aims to spread a trait of interest in a wild type population in a relatively short number of generations [133]. Application fields are numerous, and include i) the eradication of insect-borne diseases [30, 91, 133]; ii) the elimination of herbicide and pesticide resistance in pest populations [164]; iii) the control of destructive invasive species [93, 100]; iv) the conservation of biodiversity by spreading beneficial traits in endangered species [81, 187].

Targeting sexually reproducing species, CRISPR-Cas9 gene drive biases the transmission of an allele from a parent to its offspring. This biased inheritance occurs through gene conversion (also called “homing” [69]): in a heterozygous cell, the gene drive cassette present on one chromosome induces a double-strand break at a specific target site on the homologous chromosome, and the repair process duplicates the cassette. Overall, this process increases the chances of transmitting the gene drive cassette compared to its wild-part counterpart, and the mechanism repeats through the generations. Gene conversion can potentially take place at different timings of the life cycle: from very early on, in the zygote, meaning that potentially every single cell of the individual could become homozygous for the gene drive, to, in the germline, where only the gametes are converted.

Gene drives can be classified into two main categories depending on the purpose of their use [70, 95]. A “replacement drive” is aimed at spreading a genetic modification in order to introduce an important and durable feature in the natural population. Population size is then not significantly affected and the drive construct may in principle persist indefinitely in the environment. A “suppression drive” on the other hand is meant to reduce population size by spreading a detrimental trait, such as a sex ratio distorter [147] or by altering fertility [133], for example. The term “eradication drive” can be used for the extreme case where population extinction is the aim.

As with any new tool, it is essential to balance risks (safety) and benefits (efficacy) of the technique before running any field trials. Experiments currently conducted in laboratories provide small- to medium-scale information; mathematical models can help to extend these empirical results and identify the features that are the most important in determining the dynamics at larger scales [162].

Early gene drive models [34, 69, 213] used classical population genetics frameworks, and considered discrete non-overlapping generations in a well-mixed population. These simplifications helped to draw general conclusions, but it is important to challenge them. First of all, most of the species targeted in the context of gene drive do not have synchronous generations (for instance mosquitoes [91, 103, 30, 133], flies [93], mice [100]). Secondly, the assumption of a single well-mixed collection of individuals living across a uniform space is usually not realistic. In fact, most of the natural landscapes are heterogeneous. Individuals are also more likely to interact with others that are in closer proximity, which might result in local genetic variations. Finally, releases of transgenic individuals are limited in range, which is another factor of spatial heterogeneity.

Taking into account spatio-temporal dynamics of the population size is another key step towards more realistic models. For the sake of simplicity, most early models focused on allele frequencies and considered a constant population density. However in the context of gene drive, the introduction of maladapted transgenic individuals can lead to the reduction (or even extinction) of the population [70]. When considering a spatially structured population, variations in population density naturally generate a demographic flux from denser to less dense areas. This demographic flux is directed in opposition to the spread of the drive allele. It was previously shown [95] that the advantage conferred by gene conversion may nevertheless counteract the demographic effect linked to the fitness cost.

The main goal of this paper is to clarify the impact of variations in population density over the course of drive propagation over space.

We study partial differential equations which follow the propagation of the drive in space and time. We explore numerically and analytically two models: a first model based on perfect conversion in the zygote, already introduced in [95] in a spatially structured population, corresponding to an idealized case where gene conversion always succeeds; second, a more realistic model with partial conversion and presence of heterozygous individuals, already studied in [187] in a well-mixed, non spatial population. In order to investigate the possible spreading of gene drives through space after local introduction, we focus on the description of traveling waves solutions, that is, particular solutions which are stationary in a frame moving at constant speed. Our analysis goes beyond [95] by several means: we extend it to the case of partial conversion, and we systematically analyze the case where the demographic effects are the strongest, in the regime of vanishing growth rate. The latter is possible through an insightful connection with an epidemiological SI model.

2.2 Methodology

2.2.1 Models

We present our model step-by-step. For a genetically and spatially homogeneous population, we consider the following (non-dimensionalized) equation:

$$\partial_t n(t) = \left(1 + r(1 - n(t))\right) f n(t) - n(t) \quad (\forall t > 0). \quad (2.1)$$

where the unit of time is generations. Fecundity is density-dependent, and parametrized by the fitness f and the rate r at which the (f -dependent) carrying capacity is restored. When $f = 1$, the carrying capacity is 1, and we recover the logistic equation $\partial_t n(t) = r(1 - n(t))n(t)$. Other modeling options are discussed in [95].

Then, we add genetic diversity in the population. We still denote by n the total density, and by n_i the density of individuals with genotype i . The population we consider is diploid, sexually reproducing, and the fitness f_i depends on the genotype. The dynamics are given by the following equations:

$$\partial_t n_i(t) = \left(1 + r(1 - n(t))\right) f_i n(t) \underbrace{\sum_{l,k} \pi_{l,k}^i \frac{n_l(t)}{n(t)} \frac{n_k(t)}{n(t)}}_{\text{Mating term}} - n_i(t) \quad (\forall t > 0) \quad (\forall i). \quad (2.2)$$

The mating term takes into account the probability for each couple of parents l,k to have offspring of type i ($\pi_{l,k}^i$), multiplied by the probability of a mating event l,k ($\frac{n_l(t)n_k(t)}{n(t)^2}$), assuming random mating.

Last but not least, we consider a spatially structured population. We assume that the movement of individuals is described by a diffusion term with equal diffusion coefficients, normalized to 1. Since we focus on traveling wave solutions, we restrict our analysis to a one-dimensional space. We obtain the following equations:

$$\partial_t n_i(t, x) - \partial_{xx}^2 n_i(t, x) = \left(1 + r(1 - n(t, x))\right) f_i n(t) \sum_{l,k} \pi_{l,k}^i \frac{n_l(t)}{n(t)} \frac{n_k(t)}{n(t)} - n_i(t, x) \quad (\forall t > 0) \quad (\forall x \in \mathbb{R}) \quad (\forall i). \quad (2.3)$$

There are two possible alleles at the locus that we consider: the wild-type allele (W) and the drive allele (D). We have three genotypes: wild-type homozygotes ($i = WW$), drive homozygotes ($i = DD$) and heterozygotes ($i = DW$). Wild-type homozygotes have fitness $f_{WW} = 1$, drive homozygotes have fitness $f_{DD} = 1 - s$, where s is the fitness cost of the drive, and drive heterozygotes have fitness $f_{DW} = 1 - sh$, where h is the dominance parameter (see Table 2.1).

	Density	Adult genotype	Fitness
Drive Homozygote	n_{DD}	D D	$1 - s$
Heterozygote	n_{DW}	W D	$1 - sh$
Wild-type Homozygote	n_{WW}	W W	1

Table 2.1: Population characteristics (D: Drive allele, W: Wild-type allele).

All along the paper, we assume $s \in (0, 1)$, corresponding to a fitness cost carried by the drive alleles. Furthermore, we assume that the fitness of heterozygotes cannot be greater than the fitness of either homozygote ($h \in [0, 1]$).

Gene conversion turns a heterozygous cell into a drive homozygous cell. To determine the probability $\pi_{l,k}^i$ (probability for a couple l,k to have offspring of type i), we need to take into account both the probability $c \in [0, 1]$ with which gene conversion occurs in heterozygotes, and the stage of the life cycle at which it occurs: either in the zygote, or in the germline. This last feature modifies significantly the probabilities: for example, a couple W, D of gametes has a probability $1 - c$ to lead to heterozygous offspring if conversion occurs in the zygote, whereas this probability becomes one if conversion occurs in the germline. We detail all $\pi_{l,k}^i$ values in Appendix 2.A. For the sake of clarity, we now omit variables in the notation ($n_i = n_i(t, x)$).

The parameters are summarized in Table 2.2.

Parameters	Range values	Description
r	$(0, +\infty)$	Intrinsic growth rate
c	$[0, 1]$	Conversion rate
s	$(0, 1)$	Fitness cost of drive homozygotes
h	$[0, 1]$	Drive dominance

Table 2.2: Model parameters.

In this article, we will analyse the three following versions/variations of model (2.3):

Partial conversion occurring in the zygote:

$$\left\{ \begin{array}{l} \partial_t n_{DD} - \partial_{xx}^2 n_{DD} = (1-s)(r(1-n)+1) \frac{c n_{WW} n_{DW} + 2c n_{WW} n_{DD} + (\frac{1}{2}c + \frac{1}{4}) n_{DW}^2 + (c+1) n_{DW} n_{DD} + n_{DD}^2}{n} - n_{DD}, \\ \partial_t n_{DW} - \partial_{xx}^2 n_{DW} = (1-sh)(r(1-n)+1) (1-c) \frac{n_{WW} n_{DW} + 2n_{WW} n_{DD} + \frac{1}{2} n_{DW}^2 + n_{DW} n_{DD}}{n} - n_{DW}, \\ \partial_t n_{WW} - \partial_{xx}^2 n_{WW} = (r(1-n)+1) \frac{n_{WW}^2 + n_{WW} n_{DW} + \frac{1}{4} n_{DW}^2}{n} - n_{WW}. \end{array} \right. \quad (2.4)$$

Partial conversion occurring in the germline:

$$\left\{ \begin{array}{l} \partial_t n_{DD} - \partial_{xx}^2 n_{DD} = (1-s)(r(1-n)+1) \frac{\frac{1}{4}(1+c)^2 n_{DW}^2 + (1+c) n_{DW} n_{DD} + n_{DD}^2}{n} - n_{DD}, \\ \partial_t n_{DW} - \partial_{xx}^2 n_{DW} = (1-sh)(r(1-n)+1) \frac{(1+c) n_{WW} n_{DW} + 2n_{WW} n_{DD} + \frac{1}{2}(1-c^2) n_{DW}^2 + (1-c) n_{DW} n_{DD}}{n} - n_{DW}, \\ \partial_t n_{WW} - \partial_{xx}^2 n_{WW} = (r(1-n)+1) \frac{n_{WW}^2 + (1-c) n_{WW} n_{DW} + \frac{1}{4}(1-c)^2 n_{DW}^2}{n} - n_{WW}. \end{array} \right. \quad (2.5)$$

Perfect conversion occurring in the zygote (no heterozygotes):

For a perfect conversion occurring in the zygote ($c = 1$), model (2.4) reduces to the following set of two equations, which was introduced in [95]:

$$\left\{ \begin{array}{l} \partial_t n_{DD} - \partial_{xx}^2 n_{DD} = (1-s) \left(r(1-n_{DD}-n_{WW}) + 1 \right) \frac{n_{DD}^2 + 2n_{WW} n_{DD}}{n_{WW} + n_{DD}} - n_{DD} = F_D(n_{DD}, n_{WW}) \\ \partial_t n_{WW} - \partial_{xx}^2 n_{WW} = \left(r(1-n_{DD}-n_{WW}) + 1 \right) \frac{n_{WW}^2}{n_{WW} + n_{DD}} - n_{WW} = F_W(n_{DD}, n_{WW}). \end{array} \right. \quad (2.6)$$

This last model only follows the two homozygous genotypes, drive and wild-type. Due to perfect gene conversion ($c = 1$), no heterozygous individuals are ever produced: heterozygous eggs are all transformed into homozygotes. Further assuming that there are no heterozygotes initially, we only need to follow the densities of homozygotes.

Note that system (2.6) can also be obtained from model (4.1) by assuming perfect conversion in the germline ($c = 1$) and drive dominance ($h = 1$). In this case, heterozygotes and drive homozygotes have the same fitnesses, and both only produce gametes with the drive allele. We can then group them together and follow their density $n_{\text{DW}} + n_{\text{DD}}$, whose dynamics are given by the first line of (2.6).

2.2.2 Setting of the problem

Traveling waves

We seek stationary solutions in a reference frame moving at speed v , where v is some unknown:

$$\begin{cases} n_{\text{DD}}(t, x) = n_{\text{DD}}(x - vt) & (\forall t > 0) (\forall x \in \mathbb{R}), \\ n_{\text{DW}}(t, x) = n_{\text{DW}}(x - vt) & (\forall t > 0) (\forall x \in \mathbb{R}), \\ n_{\text{WW}}(t, x) = n_{\text{WW}}(x - vt) & (\forall t > 0) (\forall x \in \mathbb{R}). \end{cases} \quad (2.7)$$

Traveling wave solutions contain important information for the biological interpretation of the results, such as the speed of invasion v , the genetic composition of the expanding population, or the final equilibrium. In this paper, we focus our study on this mathematical object and detail below the vocabulary we use. Key to our analysis are the notions of monostable or bistable systems, and whether the traveling wave is pulled or pushed. There may be confusion around these concepts in the literature, so we clarify their definitions below.

Numerical simulations

We complement our mathematical analysis with numerical simulations of the Cauchy problem, using a Crank-Nicolson finite difference method with initial conditions for each genotype specified as in Figure 2.1. The outcomes of the simulations are heatmaps of the expansion speed over a wide range of parameters.

Initial conditions for numerical simulations are as follows: the left half of the domain is full of drive ($n_{\text{DD}} = 1$), and the right half is full of wild-type ($n_{\text{WW}} = 1$) (see Figure 2.1).

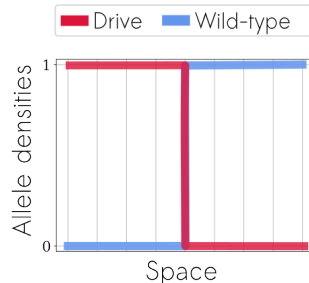


Figure 2.1: Initial conditions for numerical simulations. The left half of the domain is full of drive ($n_{\text{DD}} = n_D = 1$), and the right half is full of wild-type ($n_{\text{WW}} = n_W = 1$).

The code is available on GitHub (https://github.com/LenaKlay/gd_project_1, in the folder:

deterministic). We ran our simulations in Python 3.6, with the Spyder environment. Heavy heatmaps 2.4, 2.5, 2.6a, 2.6b; 2.C.1 have been computed thanks to the INRAE Migale bioinformatics facility (doi: 10.15454/1.5572390655343293E12). We are grateful to them for providing these computing resources.

2.2.3 Glossary

Allelic densities and frequencies

For our analysis, it is convenient to introduce the allelic (half-) densities (n_D, n_W) . The precise definition depends on the model, and more specifically on the timing of conversion. In fact, we have $n_D = n_{DD} + \alpha n_{DW}$ and $n_W = n_{WW} + (1 - \alpha) n_{DW}$, with $\alpha = \frac{1}{2}$ when conversion occurs in the zygote, and $\alpha = \frac{1+c}{2}$ when conversion occurs in the germline (see section 2.3.2). Depending on the regime of parameters, it may be more appropriate to study the allelic frequencies $p_D = \frac{n_D}{n_D + n_W}$, $p_W = \frac{n_W}{n_D + n_W}$.

Classification of the dynamics

It can happen that the dynamics lead to the decay of the drive allele uniformly in space. In this case, there cannot exist a traveling wave for the drive population: we use the term *gene drive clearance* to describe this situation. Then, the problem boils down to the standard Fisher-KPP traveling wave problem for the expansion of the wild-type in the absence of a drive (see [95]).

When traveling waves do exist, we distinguish between two cases depending on the sign of the speed. When $v > 0$, the wave moves to the right: it is a *drive invasion*. When $v < 0$, the wave moves to the left: it is a *wild-type invasion*. In some specific cases, drive and wild-type invasions can happen simultaneously: the waves decompose into two sub-traveling wave solutions over half of the domain. They move in opposite directions and lead to the coexistence of both alleles in-between.

In case of drive invasion, we distinguish several cases depending on the state of the population in the wake of the front(s): i) *eradication* drives are those for which the population vanishes in the wake of the front(s); ii) *suppression* drives are those for which population persists in the wake of the front(s). In the latter case, two scenarios are possible: persistence of drive homozygotes only; persistence of all genotypes.

Monostable / Bistable systems

To illustrate useful concepts in the theory of reaction-diffusion equations, we consider the following standard equation of population genetics [176] describing the dynamics of the frequency p of an allele of interest:

$$\partial_t p - \partial_{xx}^2 p = p(1-p)\sigma(p) \quad \text{with } p \in [0, 1], \quad (2.8)$$

where $\sigma(p)$ is the selection term, which we consider frequency-dependent (i.e., function of p).

If σ is of constant sign, say $\sigma > 0$, this equation is referred to as a monostable case. Then, the solution converges locally to the unique stable equilibrium $p = 1$ (or $p = 0$ if $\sigma < 0$). If σ is changing signs once in $(0, 1)$, being negative below some threshold, and positive above, it is referred to as a bistable case. In the latter case, the solution converges locally to one of the two stable equilibria $p = 0$ or $p = 1$, depending on the initial condition. Moreover, each equilibria has a basin of attraction and

there is a threshold effect – hence the name “threshold-dependent drives” in the gene drive literature to describe this kind of case (for example in reference [204]).

In both cases, there exist traveling waves connecting the two equilibria $p = 0$ and $p = 1$. A straightforward integration by parts shows that, whatever the stability, the sign of the wave speed satisfies

$$\text{sign}(v) = \text{sign} \left(\int_0^1 p(1-p)\sigma(p)dp \right). \quad (2.9)$$

In monostable cases with $\sigma > 0$, this sign is positive; in bistable cases, however, it depends on the details of the frequency-dependence σ . Moreover, under some circumstances (bistable case, or degenerate monostable case), the invasion outcome for the Cauchy problem can be changed by modifying the inoculum size. Even if traveling waves exist such that $p = 1$ is invading $p = 0$, small initial conditions may not succeed in propagating in space, see the discussion in [204, 209, 160].

By analogy with the scalar case, we consider that a system is monostable if it has exactly one stationary stable state, and bistable if it has exactly two stationary stable states.

Pulled and pushed waves

Usually, a wave is said to be pulled if the wave speed coincides with the minimal speed of the linearized problem at low density (resp. low frequency). This occurs when the population at low density (resp. low frequency) has sufficient reproductive success to determine the dynamics of the full invasion.

Conversely, a wave is said to be pushed if the wave speed is strictly larger than the minimal speed of the linearized problem. In contrast with pulled waves, the whole population contributes to the dynamics of invasion.

A bistable wave is clearly pushed [101]. However, a monostable wave can be either pulled or pushed, see [9, 114, 26] and discussion therein. Nonetheless, a monostable wave is necessarily pulled if the per-capita growth rate is maximal at low density (resp. low frequency). In the particular case of the scalar problem (2.8), this criterion simply writes:

$$\sigma(0) \geq (1-p)\sigma(p) \quad (\forall p \in [0, 1]). \quad (2.10)$$

2.3 Results

In part 2.3.1, we study the model with perfect conversion in the zygote (2.6) and compare the qualitative behavior of the solution when $r = 0$ and $r = +\infty$. In part 2.3.2, we proceed the same way on models with partial conversion (2.4) and (4.1), obtaining more general results.

2.3.1 Model with perfect conversion in the zygote

2.3.1.1 Preliminary statements on the model

We introduce a few general results on model (2.6) when $r > 0$, which will be useful in the study.

When $s \leq \frac{1}{2}$, system (2.12) is monostable: the only stable state is $(n_{\text{DD}} = n_{\text{DD}}^*, n_{\text{WW}} = 0)$ with $n_{\text{DD}}^* = \min(0, 1 - \frac{s}{r(1-s)})$ [95], leading to a drive invasion if any. We introduce the minimal speed

of problem (2.6) linearized at low drive density, i.e. the speed of any pulled wave in case of a drive invasion:

$$2\sqrt{\partial_{n_{\text{DD}}} F_D(0, 1)} = 2\sqrt{1 - 2s}. \quad (2.11)$$

When $s > \frac{1}{2}$, system (2.12) is bistable. Consequently traveling waves are either semi-trivial ($n_{\text{DD}} = 0$ identically, standard Fisher-KPP problem for n_{WW}) or pushed.

For our analysis, it will be convenient to rewrite model (2.6) so that it follows the frequency of the drive $p_{\text{D}} = \frac{n_{\text{D}}}{n_{\text{D}} + n_{\text{W}}} = \frac{n_{\text{DD}}}{n_{\text{WW}} + n_{\text{DD}}}$ (because $n_{\text{DW}} = 0$) and total population density $n = n_{\text{WW}} + n_{\text{DD}}$ (details in 2.B.1):

$$\begin{cases} \partial_t p_{\text{D}} - \partial_{xx}^2 p_{\text{D}} &= 2 \partial_x(\log n) \partial_x p_{\text{D}} + (r(1-n) + 1) s p_{\text{D}} (1 - p_{\text{D}}) \left(p_{\text{D}} - \frac{2s-1}{s} \right), \\ \partial_t n - \partial_{xx}^2 n &= (r(1-n) + 1) (1 - s + s(1 - p_{\text{D}})^2) n - n. \end{cases} \quad (2.12)$$

System (2.12) differs from standard equations often used in populations genetics as it contains an advection term $2 \partial_x(\log n) \partial_x p_{\text{D}}$. This term appears when calculating $\partial_{xx}^2 p_{\text{D}} = \partial_{xx}^2 \frac{n_{\text{DD}}}{n}$ (details in 2.B.1) and represents a demographic flux from denser to less dense areas, due to variations in population density. It is opposed to the spread of the (costly) drive allele (see Figure 2 [95]). We observe a singularity for $n = 0$ in both formulations of the system: in (2.6) due to $\frac{1}{n_{\text{DD}} + n_{\text{WW}}}$ and in (2.12) due to $\log(n)$. This should be handled with care.

2.3.1.2 $r = +\infty$

The limit of system (2.12) when $r \rightarrow +\infty$ has already been determined in [95]. Using the Strugarek-Vauchelet rescaling [201], the following limit equation is obtained, which was also previously introduced in [204]:

$$\partial_t p_{\text{D}} - \partial_{xx}^2 p_{\text{D}} = \frac{s p_{\text{D}} (1 - p_{\text{D}}) \left(p_{\text{D}} - \frac{2s-1}{s} \right)}{1 - s + s(1 - p_{\text{D}})^2}. \quad (2.13)$$

Interestingly, equation (2.13) is independent of the population density n and it does not contain the advection term $2 \partial_x(\log n) \partial_x p_{\text{D}}$. This is due to the fact that the population size $n(t, x)$ remains spatially homogeneous after the introduction of drive individuals, when $r \rightarrow +\infty$. Intuitively, so many offspring are produced at each generation that the carrying capacity is instantaneously restored, and losing a fraction s of these offspring by selection has no consequence. Therefore the variations in population density (n), and consequently the demographic flux, are negligible.

Equation (2.13) has a single parameter, the fitness cost of the drive s . The numerical value of the threshold for the transition from positive to negative speed (≈ 0.70) was already known [204, 95], and can be computed to arbitrary precision by the formula (2.9). The numerical value of the threshold for the transition from pulled to pushed (≈ 0.35 up to two digits) was numerically computed by a continuation method following [13, 114].

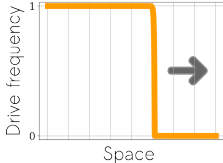
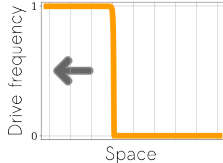
s value	$0 < s \lesssim 0.35$	$0.35 \lesssim s < 1/2$	$1/2 < s \lesssim 0.70$	$0.70 \lesssim s < 1$
Stability	Monostable		Bistable	
Speed	$v = 2\sqrt{1-2s}$	$v > 2\sqrt{1-2s}$	$v > 0$	$v < 0$
Wave	Pulled wave	Pushed wave		Pushed wave
Invasion	<p style="text-align: center;">Drive invasion</p> 			<p style="text-align: center;">Wild-type invasion</p> 

Table 2.3: Traveling waves study for Model (2.13), limit of system (2.6) when $r = +\infty$. All statements in the table are proved in Appendix 2.C.2.

Note that equation (2.13) shows qualitative similarities with a common equation in the population genetics literature [15] (and which is actually an approximation of (2.13) under weak selection, i.e. $s \rightarrow 0$):

$$\partial_t p_D - \partial_{xx}^2 p_D = s p_D (1 - p_D) \left(p_D - \frac{2s-1}{s} \right). \quad (2.14)$$

Quantitatively, the thresholds are $\frac{2}{3}$ instead of 0.35, and $\frac{2}{3}$ instead of 0.70 (analytical values) [101].

2.3.1.3 $r = 0$

When the intrinsic growth rate r is finite, it is expected that the final population density after the invasion of the drive (if any) is strictly below 1, because of the fitness cost. The smaller r , the lower the final size. The spatial effect of demography on gene drive expansion is expected to be maximal as r vanishes, when the population can hardly restore its carrying capacity, leading to a high amplitude of the population size gradient $2 \partial_x (\log n)$. In this section we focus on the limit $r = 0$, which maximizes the demographic impact of the fitness cost on drive propagation.

In a purely wild-type population, the case $r = 0$ corresponds to a number of births balancing exactly the number of deaths. As soon as the drive allele is introduced, this balance is locally broken, yielding a net decrease in the population size. Then, the drive can either propagate by leaving empty space behind, or disappear. The same conclusion holds as long as $r < \frac{s}{1-s}$, see Section 2.3.1.1.

We checked numerically that the wave speed is continuous in the limit $r \rightarrow 0$. Therefore, each conclusion on the case $r = 0$ sheds some light on the case of small r (see heatmap in Appendix 2.C.1).

As discussed above, we cannot just consider a single equation on the drive frequency p_D when r is finite because of the demographic contribution $2\partial_x \log n$. Interestingly, in the case $r = 0$, the demographic system (2.6) reduces to the following pair of equations:

$$\begin{cases} \partial_t n_{\text{WW}} - \partial_{xx}^2 n_{\text{WW}} &= \frac{-n_{\text{WW}} n_{\text{DD}}}{n_{\text{WW}} + n_{\text{DD}}}, \\ \partial_t n_{\text{DD}} - \partial_{xx}^2 n_{\text{DD}} &= (1-s) \frac{n_{\text{WW}} n_{\text{DD}}}{n_{\text{WW}} + n_{\text{DD}}} - s n_{\text{DD}}. \end{cases} \quad (2.15)$$

Noticeably, the previous system shares some features with density-dependent epidemiological SI models. In particular, the dynamics of n_{WW} is always decreasing. The dynamics of n_{DD} is the balance of creation and linear decay. By changing notations $n_{\text{WW}} \leftrightarrow S$ (susceptible individuals), and $n_{\text{DD}} \leftrightarrow I$ (infected individuals), (2.15) can be recast as follows:

$$\begin{cases} \partial_t S - \partial_{xx}^2 S &= -\beta_1 \frac{S I}{S + I}, \\ \partial_t I - \partial_{xx}^2 I &= \beta_2 \frac{S I}{S + I} - \gamma I. \end{cases} \quad (2.16)$$

with $\beta_1 = 1$, $\beta_2 = (1-s)$ (transmission parameters), and $\gamma = s$ (disease clearance). Usually, in SI models, individuals of type S are all transformed into individuals of type I at infection, hence $\beta_1 = \beta_2$. In our case, these two rates are distinct because of the fitness cost of the drive. The existence of traveling waves for model (2.15) with $\beta_1 = \beta_2$ has been studied recently in the literature [225]. Here, we extend the results of reference [225] to a more general case $0 < \beta_1$ and $0 < \beta_2$. This leads to the characterization in Table 2.4 and Appendix 2.D.

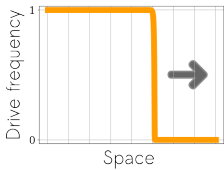
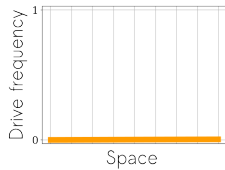
s value	$0 < s < 1/2$	$1/2 < s < 1$
Stability	Monostable	Degenerate case
Speed	$v = 2\sqrt{1-2s}$	No wave
Wave	Pulled wave	
Invasion	Drive invasion 	Gene drive clearance 

Table 2.4: Traveling waves study for Model (2.15), limit of system (2.6) when $r = 0$. All statements in the table are proved in Appendix 2.C.2.

In contrast to the results obtained when $r = +\infty$, when $r = 0$ there is only one threshold value of s determining the outcome of the model (Table 2.4). When $0 < s < 1/2$, the system is monostable, the drive necessarily invades. Moreover, the wave is pulled and travels at speed $v = 2\sqrt{1-2s}$ (2.11). When $1/2 < s < 1$, the problem is degenerate: there exists a family of steady states, corresponding to homogeneous $n_{\text{WW}} \in [0, 1]$ and $n_{\text{DD}} = 0$. It is a case of gene drive clearance, as n_{DD} converges to zero uniformly in space (at rate at least $1-2s$). However, the final density of wild-type is not clearly

determined, as it boils down to diffusion only in the large time asymptotics (details in Appendix 2.C.2.1). Note that this conclusion holds in a well-mixed population (without spatial consideration): the drive decays uniformly and the final density of wild type depends on the initial data.

2.3.1.4 Comparison between the outcomes when $r = +\infty$ and $r = 0$

The differences between the two regimes are strongest for intermediate values of s . When $1/2 < s \lesssim 0.70$, the drive can spread when the demographic consequences are negligible ($r = +\infty$). However, such a costly drive cannot invade when the intrinsic growth rate r is very low ($r = 0$). When $0.35 \lesssim s < 1/2$, the drive wave advances for both $r = +\infty$ and $r = 0$. However, it is of different nature: the wave is pulled when $r = 0$, while it is pushed when $r = \infty$.

By providing analytical results for $r = 0$, our study is complementary to [95], where the invasion outcome was described numerically in [95, Figure 3.A] together with a series of analytical estimates of the sign of the speed.

2.3.2 Models with partial conversion

In this section, we extend the study to models (2.4) and (4.1) with partial conversion. The models are reformulated in terms of allelic densities (n_D and n_W) rather than genotype densities (n_{DD} , n_{DW} and n_{WW}). This reformulation enables reducing the number of equations from three to two equations. Even if we are not able to determine the genotypic composition of the population (which individual genotype a gamete comes from, either homozygote or heterozygote), the spreading properties are equivalent. Interestingly, the same roadmap as in the full conversion model can be followed. Again, we focus on the two extreme regimes $r = +\infty$ and $r = 0$.

2.3.2.1 Conversion occurring in the zygote

When conversion occurs in the zygote, we can deduce the following system from model (2.4), with $n_D = n_{DD} + \frac{1}{2} n_{DW}$ and $n_W = n_{WW} + \frac{1}{2} n_{DW}$:

$$\begin{cases} \partial_t n_D - \partial_{xx}^2 n_D = n_D \left[\frac{r(1-n)+1}{n} [(1-s)(n_D + 2c n_W) + (1-sh)(1-c)n_W] - 1 \right] = F_D^z(n_D, n_W), \\ \partial_t n_W - \partial_{xx}^2 n_W = n_W \left[\frac{r(1-n)+1}{n} [n_W + (1-sh)(1-c)n_D] - 1 \right] = F_W^z(n_D, n_W). \end{cases} \quad (2.17)$$

The density n_W (resp. n_D) corresponds to one half of the wild-type (resp. drive) allele density at the time of zygote formation. When conversion happens in the zygote, heterozygous individuals are the result of conversion failures and produce one half of each type of gamete, drive or wild-type.

2.3.2.1.1 Preliminary statements on the model

This model brings more variety in terms of traveling waves than the previous one (2.6). Cases of monostable wild-type invasion can occur, as well as cases of monostable coexistence. We introduce

first all the possible minimal speeds of the problem linearized at low densities and detail later under which parameters they arise.

The minimal speed of the problem linearized at low drive density, i.e. the speed of any pulled monostable wave with a positive speed is given by:

$$2\sqrt{\partial_{n_D} F_D^z(0,1)} = 2\sqrt{2c(1-s) + (1-sh)(1-c) - 1} = 2\sqrt{c(1-2s) - sh(1-c)}, \quad (2.18)$$

provided that the quantity is non-negative. The minimal speed of the problem linearized at low wild-type density depends on the stable steady state of a population only bearing drive alleles (i.e. a population of drive homozygotes): $n_D^* = n_{DD}^* = \min(0, 1 - \frac{s}{r(1-s)})$ [95]. If $n_D^* = 1 - \frac{s}{r(1-s)}$, this minimal speed is given by:

$$-2\sqrt{\partial_{n_W} F_W^z(n_D^*, 0)} = -2\sqrt{\frac{(1-sh)(1-c)}{1-s} - 1}. \quad (2.19)$$

If $n_D^* = 0$, the minimal speed of the problem linearized at low wild-type density is given by the classical Fisher-KPP formula:

$$-2\sqrt{\partial_{n_W} F_W^z(0,0)} = -2\sqrt{r}. \quad (2.20)$$

Note that (2.20) is the only minimal speed depending on parameter r : it corresponds to the case of gene drive clearance, the only configuration where the drive allele disappears uniformly in space.

For our analysis, it will be convenient to rewrite model (2.17) so that it follows the frequency of the drive $p_D = \frac{n_D}{n_W + n_D}$ and the total population density $n = n_{WW} + n_{DW} + n_{DD} = n_W + n_D$ (details in 2.B.2.1):

$$\begin{cases} \partial_t n - \partial_{xx}^2 n = & (r(1-n) + 1) \left((1-s)p_D^2 + 2p_D(1-p_D)[c(1-s) + (1-c)(1-sh)] + (1-p_D)^2 \right) n - n, \\ \partial_t p_D - \partial_{xx}^2 p_D = & 2\partial_x \log(n) \partial_x p_D \\ & + (r(1-n) + 1) \left([1 - 2(1-c)(1-h)] s p_D - s[1 - (1-c)(1-h)] + c(1-s) \right) (1-p_D) p_D. \end{cases} \quad (2.21)$$

2.3.2.1.2 $r = +\infty$

Similarly as in Section 2.3.1.2, we compute formally the limiting equation on (2.21) when $r \rightarrow +\infty$:

$$\partial_t p_D - \partial_{xx}^2 p_D = \frac{\left([1 - 2(1-c)(1-h)] s p_D - s[1 - (1-c)(1-h)] + c(1-s) \right) (1-p_D) p_D}{(1-s)p_D^2 + 2p_D(1-p_D)[c(1-s) + (1-c)(1-sh)] + (1-p_D)^2}. \quad (2.22)$$

Note that as in Section 2.3.1, this equation does not depend on n . We introduce:

$$\mathcal{A}_z := s [2(1-c)(1-h) - 1], \quad s_1 := \frac{c}{1-h(1-c)}, \quad s_{2,z} := \frac{c}{2c+h(1-c)}. \quad (2.23)$$

where z stands for zygote. Note that $\mathcal{A}_z > 0 \iff s_1 < s_{2,z}$.

We distinguish between two cases, depending on the sign of \mathcal{A}_z . If $\mathcal{A}_z > 0$, which can only happen if we have both $c < \frac{1}{2}$ and $h < \frac{1}{2}$, the system is always monostable. We observe a drive invasion for $s < s_1$, a coexistence state for $s_1 < s < s_{2,z}$ and a wild-type invasion for $s_{2,z} < s$. Criterion (2.10) is always verified (see Appendix 2.E.1), consequently every traveling wave (or sub-traveling wave in case of coexistence) is pulled, moving at speed (2.18) or (2.19). These statements are summarized in Table 2.5.

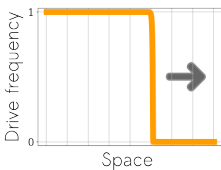
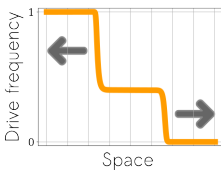
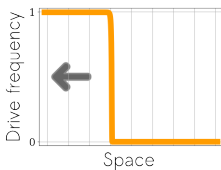
s value	$s < s_1$	$s_1 < s < s_{2,z}$	$s_{2,z} < s$
Stability	Monostable		
Speed	$v = v_{lin+}$	$v = v_{lin+}$ and $v = v_{lin-}$	$v = v_{lin-}$
Wave	Pulled Wave		
Invasion	<p>Drive invasion</p> 	<p>Coexistence</p> 	<p>Wild-type invasion</p> 

Table 2.5: Traveling waves study for Model (2.22) (limit of system (2.21) when $r = +\infty$) when $\mathcal{A}_z > 0$, with $v_{lin+} = 2\sqrt{c(1-2s) - sh(1-c)}$ (2.18) and $v_{lin-} = -2\sqrt{\frac{(1-sh)(1-c)}{1-s}} - 1$ (2.19). For all values of s , there exists only one stable state (monostability). In particular in the case of coexistence, the stable state (in the center) invades two unstable states (on the right and left)

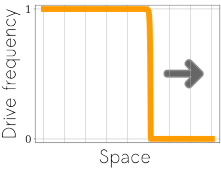
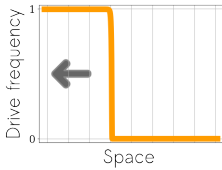
s value	$s \in \mathcal{S}_z$	$s \in (0, s_{2,z}) \setminus \mathcal{S}_z$	$s_{2,z} < s < s_1$	$s_1 < s < 1$
Stability	Monostable		Bistable	Monostable
Speed	$v = v_{lin+}$	$v \geq v_{lin+}$		$v < 0$
Wave	Pulled wave		Pushed wave	
Invasion	<p>Drive invasion</p> 	<p>Drive or Wild-type invasion</p>	<p>Wild-type invasion</p> 	

Table 2.6: Traveling waves study for Model (2.22) (limit of system (2.21) when $r = +\infty$) when $\mathcal{A}_z < 0$, with $v_{lin+} = 2\sqrt{c(1-2s) - sh(1-c)}$ (2.18).

If we consider the case $\mathcal{A}_z < 0$. The system is monostable for $s < s_{2,z}$ (drive invasion), bistable for $s_{2,z} < s < s_1$ and monostable for $s_1 < s$ (wild-type invasion). In case of monostable drive invasion, we

define a set \mathcal{S}_z of s values:

$$\mathcal{S}_z := \left\{ s \in (0, 1) \mid \left(1 - 2s[1 - (1-h)(1-c)]\right) \left((-2c - h + ch)s + c\right) + s [2(1-c)(1-h) - 1] > 0 \right\}. \quad (2.24)$$

For all s in \mathcal{S}_z and $\mathcal{A}_z < 0$, criterion (2.10) is verified and consequently, there exists a pulled monostable traveling wave with positive speed (see Appendix 2.E.1). Note that such s values are necessarily strictly below $s_{2,z}$, condition for a monostable drive invasion. In case of wild-type invasion, criterion (2.10) is never verified (see Appendix 2.E.1). These statements are summarized in Table 2.6.

2.3.2.1.3 $r = 0$

Using the relation $n = n_{\text{WW}} + n_{\text{DW}} + n_{\text{DD}} = n_{\text{W}} + n_{\text{D}}$, system (2.17) can be rewritten as follows when $r = 0$:

$$\begin{cases} \partial_t n_{\text{D}} - \partial_{xx}^2 n_{\text{D}} = \left(c(1-s) + s(1-c)(1-h) \right) \frac{n_{\text{D}} n_{\text{W}}}{n_{\text{D}} + n_{\text{W}}} - s n_{\text{D}}, \\ \partial_t n_{\text{W}} - \partial_{xx}^2 n_{\text{W}} = - \left(1 - (1-sh)(1-c) \right) \frac{n_{\text{D}} n_{\text{W}}}{n_{\text{D}} + n_{\text{W}}}. \end{cases} \quad (2.25)$$

We apply the results of Appendix 2.D with $\beta_1 = 1 - (1-sh)(1-c)$ and $\beta_2 = c(1-s) + s(1-c)(1-h)$. There exists a monostable and pulled drive invasion wave if:

$$\beta_2 > \gamma \iff s < s_{2,z} = \frac{c}{2c + h(1-c)}. \quad (2.26)$$

On the other hand when $\beta_2 < \gamma$, the reaction term of n_{D} in (2.25) is strictly negative. As before, the density n_{D} converges to zero uniformly in space at rate at least $\beta_2 - \gamma$ (gene drive clearance) and the final density of wild-type is not clearly determined: the problem boils down to diffusion only in the large time asymptotics (details in Appendix 2.C.2.1). These statements are summarized in Table 2.7.

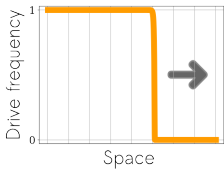
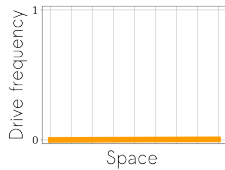
s value	$0 < s < s_{2,z}$	$s_{2,z} < s < 1$
Stability	Monostable	Degenerate case
Speed	$v = v_{\text{lin}+}$	No wave
Wave	Pulled wave	
Invasion	Drive invasion 	Gene drive clearance 

Table 2.7: Traveling waves study for Model (2.25) (limit of system (2.21) when $r = 0$), with $v_{\text{lin}+} = 2\sqrt{c(1-2s) - sh(1-c)}$ (2.18).

Note that, when $\mathcal{A}_z > 0$, a condition for having a pulled wave with positive speed for both $r = 0$ and $r = +\infty$ is $s < s_{2,z}$. When $\mathcal{A}_z < 0$, a condition for having a pulled wave with positive speed for

both $r = 0$ and $r = +\infty$ is $s \in \mathcal{S}_z \subseteq [0, s_{2,z}]$. This suggests that, under those conditions, whatever the value of the demographic parameter r is, the drive invasion wave is always pulled and consequently, travels at a speed which does not depend on r either (speed given by (2.18)). We verify this intuition numerically (vertical level lines) in the following section.

2.3.2.1.4 Numerical illustrations

$\mathcal{A}_z > 0$

In a first example we choose $c = 0.25$ and $h = 0.1$ such that $\mathcal{A}_z > 0$. The s threshold values are $s_1 \approx 0.27$ and $s_{2,z} \approx 0.43$. As discussed in the previous sections, when $s < s_{2,z}$, all waves are pulled (sub-)traveling waves for $r = +\infty$ and $r = 0$. Note that $s > s_1$ is the condition for the existence of pulled (sub-)traveling waves with negative speed only when $r = +\infty$.

We show the value of the speed (2.18) and (2.19) of the pulled waves as a function of s when $r = +\infty$, with $c = 0.25$ and $h = 0.1$ (in Figure 2.2).

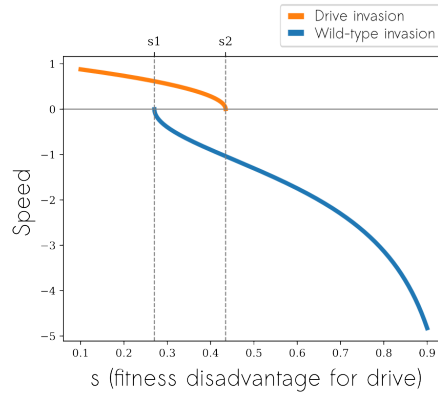


Figure 2.2: Speed of the drive invasion (2.18) in orange and speed of the wild-type invasion (2.19) in blue, as a function of s when $r = +\infty$, with $c = 0.25$ and $h = 0.1$.

Note that for $0.27 \approx s_1 < s < s_{2,z} \approx 0.43$, we have both a drive and a wild-type invasion, leading to a stable coexistence state. This case is illustrated in Figure 2.3 with $s = 0.35$.

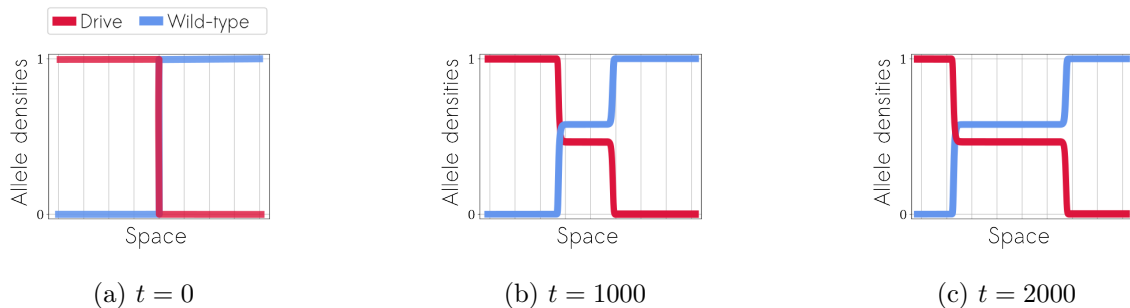
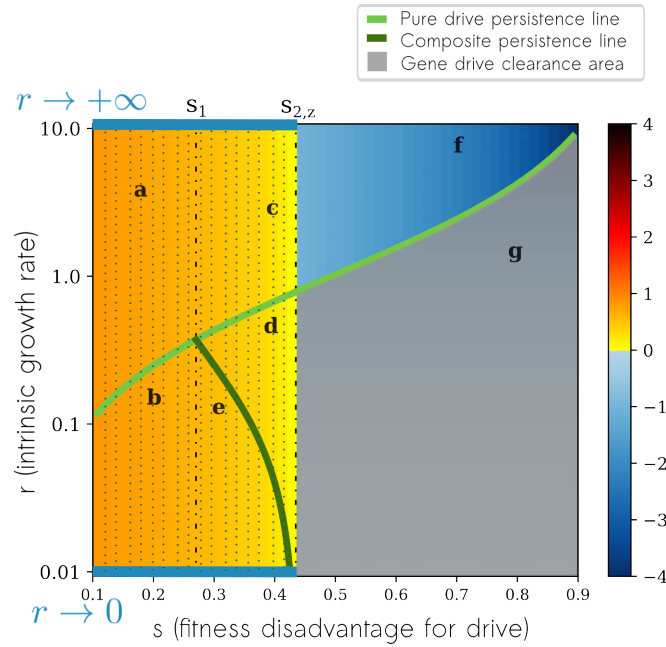
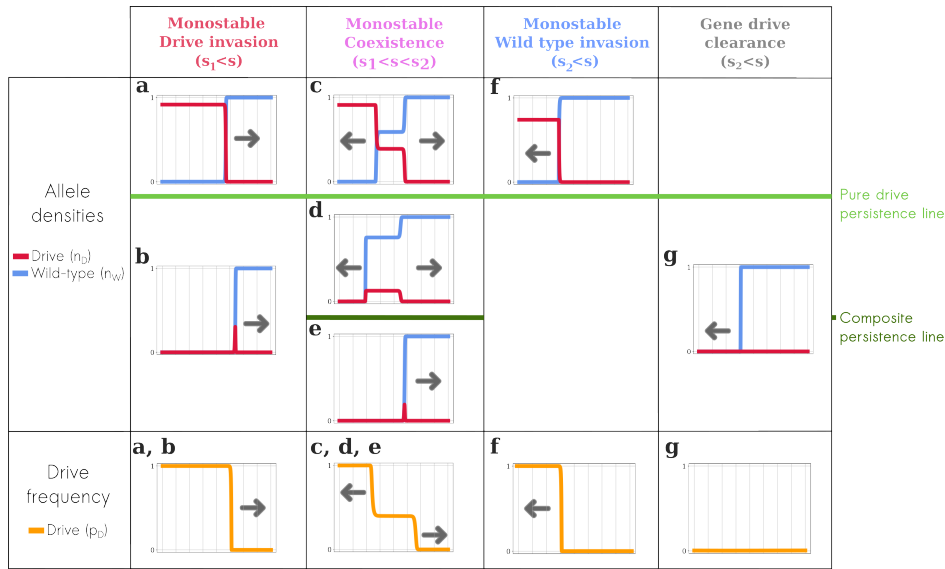


Figure 2.3: Allele densities as a function of space, at different times, for $s = 0.35$, $c = 0.25$, $h = 0.1$ and $r = 3$.

We now compute numerically the values of the wave speed for intermediate values of r (in Figure 2.4). In case of coexistence, for $s_1 < s < s_{2,z}$, we choose to show only the positive speed value.



(a) Heatmap representing the speed of the wave for $c = 0.25$, $h = 0.1$ when conversion occurs in the zygote.



(b) Illustrations of the shape of the wave (allele densities and drive frequency) for each corresponding case in the heatmap.

Figure 2.4: (a) Heatmap representing the speed of the waves for $c = 0.25$, $h = 0.1$ when conversion occurs in the zygote. When the drive invades the population, the speed is positive (in yellow-orange). On the contrary, when the wild-type invades the population, the speed is negative (in blue). When both drive and wild types invade (coexistence), only the speed of the drive is shown in the heatmap, resulting in an apparent discontinuity at $s = s_{2,z}$. As $\mathcal{A}_z > 0$, the system is always monostable for $r = +\infty$: when $s < s_1$ the drive always invades; when $s_1 < s < s_{2,z}$ the final state is a coexistence state; when $s > s_{2,z}$ the wild-type invades or there is gene drive clearance. The turquoise horizontal lines at the bottom and at the top of the heatmap indicate the theoretical values of s such that there exists a pulled wave with positive speed, respectively for $r = +\infty$ and $r = 0$. Below the pure drive persistence line (light green), a well-mixed population containing only drive homozygous individuals will necessarily go extinct. Below the composite persistence line (dark green), it is the whole population that goes extinct (calculations for both lines available in Appendix 2.F.2). The gray zone corresponds to the gene drive clearance area. Outside the gray zone, the level lines are apparently vertical, meaning that the wave speed would be independent of r . This is in agreement with the fact that the values of the speed coincide when $r = +\infty$ and $r = 0$ for $s < s_{2,z}$. If correct, the value of the speed can be found in Figure 2.2. (b) Shape of the wave for each case indicated by a letter in the heatmap above. The position of the graphs in the table reflects the position in the heatmap with respect to the persistence lines.

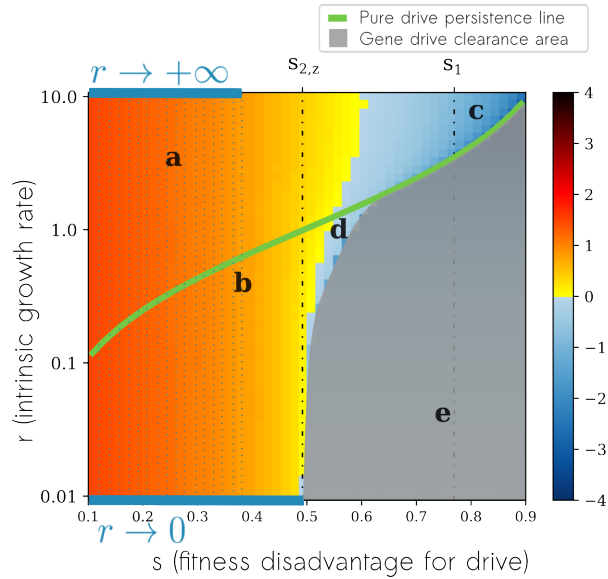
For a better understanding of Figure 2.4, we detail the effect of fitness disadvantage s and dominance coefficient h on drive dynamics for $r = +\infty$ and $c = 0.25$, without spatial structure, in Appendix 2.F.1 (in Figure 2.F.1a).

In Figure 2.4, the speed value for $s < s_{2,z}$ seems not to depend on the demographic parameter r : whatever the final equilibrium is, going from population extinction to full replacement of the wild-type genotypes by drive genotypes, the invasion occurs at the same speed. This is in agreement with the fact that the values of the speed coincide when $r = +\infty$ and $r = 0$ for $s < s_{2,z}$. If correct, the value of the speed can be found in Figure 2.2.

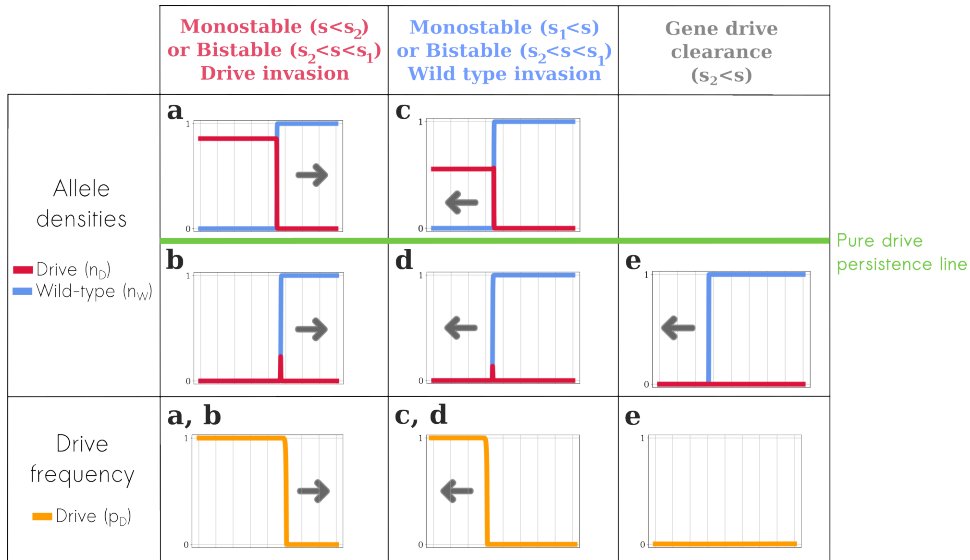
$\mathcal{A}_z < 0$

In a second example, we choose $c = 0.75$ and $h = 0.1$ such that $\mathcal{A}_z < 0$. The s threshold values are $s_1 \approx 0.77$ and $s_{2,z} \approx 0.49$. As discussed in the previous section, when $s \in \mathcal{S}_z$ (in our case $s \lesssim 0.38$), all waves are pulled traveling waves. However, the latter criterion is not a sufficient condition. It is expected that waves are indeed pulled beyond this approximate value of 0.38. However, this would require to use numerical continuation methods as in Section 2.3.1.2. We computed numerically the speed values for intermediate values of r , as shown in Figure 2.5. We believe that the wave speed is independent of the demographic parameter r when the wave is pulled (visual observation for $s \lesssim 0.38$).

For a better understanding of Figure 2.5, we detail the effect of fitness disadvantage s and dominance coefficient h on drive dynamics for $r = +\infty$ and $c = 0.75$, without spatial structure, in Appendix 2.F.1 (in Figure 2.F.1b).



(a) Heatmap representing the speed of the wave for $c = 0.75$, $h = 0.1$ when conversion occurs in the zygote.



(b) Illustrations of the shape of the wave (allele densities and drive frequency) for each corresponding case in the heatmap.

Figure 2.5: (a) Heatmap representing the speed of the wave for $c = 0.75$, $h = 0.1$ when conversion occurs in the zygote. When the drive invades the population, the speed is positive (in yellow-orange-red). On the contrary, when the wild-type invades the population, the speed is negative (in blue). We have $\mathcal{S}_z < 0$, therefore when $r = +\infty$: when $s < s_{2,z}$ the system is monostable and the drive always invades; when $s_{2,z} < s < s_1$ the system is bistable and the final state depends on the initial condition; when $s > s_1$ the system is monostable and the wild-type invades or there is gene drive clearance. The turquoise horizontal lines at the bottom and at the top of the heatmap indicate the theoretical values of s such that there exists a pulled wave with positive speed, respectively for $r = +\infty$ and $r = 0$. Below the pure drive persistence line (light green), a well-mixed population containing only drive homozygous individuals will necessarily go extinct (calculations for this line available in Appendix 2.F.2). For $s \in \mathcal{S}_z$, i.e. $s \lesssim 0.38$, the level lines are apparently vertical: this is in agreement with the fact that the values of the speed coincide when $r = +\infty$ and $r = 0$ in this area. (b) Shape of the wave for each case indicated by a letter in the heatmap above. The position of the graphs in the table reflects the position in the heatmap with respect to the pure drive persistence line.

2.3.2.2 Conversion occurring in the germline

When conversion occurs in the germline, we can deduce the following system from model (4.1), with $n_D = n_{DD} + (1+c) \frac{1}{2} n_{DW}$ and $n_W = n_{WW} + (1-c) \frac{1}{2} n_{DW}$:

$$\begin{cases} \partial_t n_D - \partial_{xx}^2 n_D = n_D \left[\frac{r(1-n) + 1}{n} \left[(1-s)n_D + (1-sh)(1+c)n_W \right] - 1 \right] = F_D^g(n_D, n_W), \\ \partial_t n_W - \partial_{xx}^2 n_W = n_W \left[\frac{r(1-n) + 1}{n} \left[n_W + (1-sh)(1-c)n_D \right] - 1 \right] = F_W^g(n_D, n_W). \end{cases} \quad (2.27)$$

The density n_W (resp. n_D) corresponds to one half of the wild-type (resp. drive) allele density at the time of zygote formation. When conversion happens in the germline, heterozygous individuals undergo a conversion of their wild-type alleles with probability c , and produce a fraction $(1+c)/2$ of drive-carrying gametes.

2.3.2.2.1 Preliminary statements on the model

As before, we detail the minimal speed of the problem linearized at low densities, for both drive and wild-type alleles.

In case of drive invasion, the minimal speed of the problem linearized at low drive density, i.e. the speed of any pulled monostable wave with positive speed is given by:

$$2\sqrt{\partial_{n_D} F_D^g(0, 1)} = 2\sqrt{(1-sh)(1+c) - 1}. \quad (2.28)$$

Note that $F_W^g(n_D, n_W) = F_W^z(n_D, n_W)$: in case of a wild-type invasion, the minimal speeds are already given by (2.19) and (2.20) (Section 2.3.2.1).

For our analysis, it will be convenient to rewrite model (4.4) so that it follows the frequency of the drive $p_D = \frac{n_D}{n_W + n_D}$ and the total population density $n = n_{WW} + n_{DW} + n_{DD} = n_W + n_D$ (details in Appendix 2.B.2.2):

$$\begin{cases} \partial_t n - \partial_{xx}^2 n = (r(1-n) + 1) \left((1-s)p_D^2 + 2(1-sh)p_D(1-p_D) + (1-p_D)^2 \right) n - n, \\ \partial_t p_D - \partial_{xx}^2 p_D = 2 \partial_x \log(n) \partial_x p_D + (r(1-n) + 1) \left((2h-1) s p_D + (1-sh)(1+c) - 1 \right) p_D (1-p_D). \end{cases} \quad (2.29)$$

2.3.2.2.2 $r = +\infty$

Similarly as in Section 2.3.1.2, we can compute formally the limiting equation on p_D when $r = +\infty$:

$$\partial_t p_D - \partial_{xx}^2 p_D = \frac{\left(-(1-2h) s p_D + [(1-sh)(1+c) - 1] \right) p_D (1-p_D)}{(1-s) p_D^2 + 2(1-sh) p_D (1-p_D) + (1-p_D)^2}. \quad (2.30)$$

Note that as in section 2.3.1, this equation does not depend on n . We introduce:

$$\mathcal{A}_g := s(1 - 2h), \quad s_1 := \frac{c}{1 - h(1 - c)}, \quad s_{2,g} := \frac{c}{2ch + h(1 - c)} = \frac{c}{h(1 + c)}. \quad (2.31)$$

where g stands for germline. Note that $\mathcal{A}_g > 0 \iff s_1 < s_{2,g}$. We define a set \mathcal{S}_g of s values:

$$\mathcal{S}_g := \left\{ s \in (0, 1) \mid (1 - 2sh)(c - sh(c + 1)) + s(1 - 2h) > 0 \right\}. \quad (2.32)$$

Results are exactly the same as in Section 2.3.2.1.2, substituting \mathcal{A}_z by \mathcal{A}_g , $s_{2,z}$ by $s_{2,g}$, \mathcal{S}_z by \mathcal{S}_g , and the minimal speed of the problem linearized at low drive density (2.18) by (2.28) (see Appendix 2.E.2).

2.3.2.2.3 $r = 0$

Using the relation $n = n_W + n_D$, system (4.4) can be rewritten as follows when $r = 0$:

$$\begin{cases} \partial_t n_D - \partial_{xx}^2 n_D = \left(c(1 - sh) + s(1 - h) \right) \frac{n_D n_W}{n_D + n_W} - s n_D, \\ \partial_t n_W - \partial_{xx}^2 n_W = -\left(1 - (1 - sh)(1 - c) \right) \frac{n_D n_W}{n_D + n_W}. \end{cases} \quad (2.33)$$

We apply the results of Appendix 2.D with $\beta_1 = 1 - (1 - sh)(1 - c)$ and $\beta_2 = c(1 - sh) + s(1 - h)$. There exists a monostable and pulled drive invasion wave if:

$$\beta_2 > \gamma \iff s < s_{2,g} = \frac{c}{h(1 + c)} = \frac{c}{2ch + h(1 - c)}. \quad (2.34)$$

On the other hand when $\beta_2 < \gamma$, the reaction term of n_D in (2.33) is strictly negative. As before, the density n_D converges to zero uniformly in space at rate at least $\beta_2 - \gamma$ (gene drive clearance) and the final density of wild-type is not clearly determined: the problem boils down to diffusion only in the large time asymptotics (details in Appendix 2.C.2.1).

Note that the same intuitions as in section 2.3.2.1.3 hold: when $\mathcal{A}_g > 0$, a condition for having a pulled wave with positive speed for both $r = 0$ and $r = +\infty$ is $s < s_{2,g}$; when $\mathcal{A}_g < 0$, a condition for having a pulled wave with positive speed for both $r = 0$ and $r = +\infty$ is $s \in \mathcal{S}_g \subseteq [0, s_{2,g}]$. This suggests that, under those conditions, whatever the value of the demographic parameter r is, the drive invasion wave is always pulled and consequently, travels at a speed which does not depend on r either (speed given by (2.28)). As before, we verify this intuition numerically (vertical level lines) in the following section.

2.3.2.2.4 Numerical illustrations

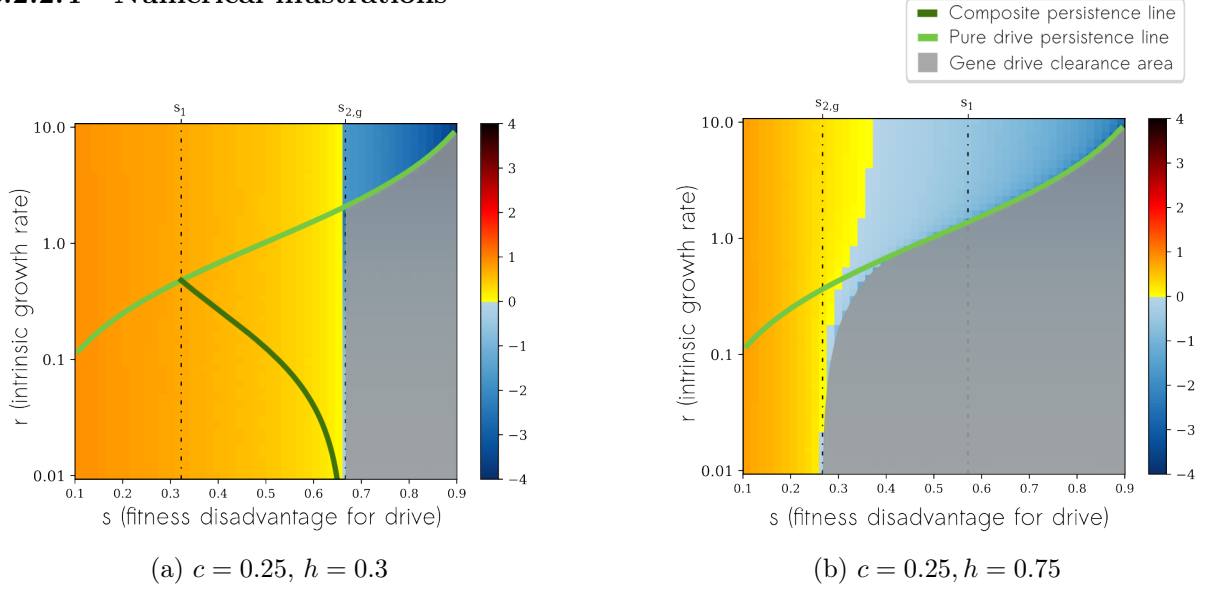


Figure 2.6: Heatmap representing the speed of the wave when conversion occurs in the germline. When the drive invades the population, the speed is positive (in yellow-orange-red). On the contrary, when the wild-type invades the population, the speed is negative (in blue). Below the pure drive persistence line (light green), a well-mixed population containing only drive homozygous individuals will necessarily go extinct. Below the composite persistence line (dark green), it is the whole population that goes extinct (calculations for both lines available in Appendix 2.F.2). The gray zone corresponds to the gene drive clearance area. In (a) we have $\mathcal{A}_g > 0$, therefore the system is always monostable for $r = +\infty$: when $s < s_1$ the drive always invades; when $s_1 < s < s_{2,g}$ the final state is a coexistence state; when $s > s_{2,g}$ the wild-type invades or there is gene drive clearance. When both drive and wild types invade (coexistence), only the speed of the drive is shown in the heatmap, resulting in an apparent discontinuity at $s = s_{2,z}$. In (b) we have $\mathcal{A}_z < 0$, therefore when $r = +\infty$: when $s < s_{2,g}$ the system is monostable and the drive always invades; when $s_{2,g} < s < s_1$ the system is bistable and the final state depends on the initial condition; when $s > s_1$ the system is monostable and the wild-type invades or there is gene drive clearance.

For a better understanding of Figures 2.6a and 2.6b, we detail the effect of fitness disadvantage s and dominance coefficient h on drive dynamics for $r = +\infty$ and $c = 0.25$, without spatial structure, in Appendix 2.F.1 (in Figure 2.F.2a).

2.3.2.3 Conclusion

When conversion occurs in the zygote (resp. in the germline) for $\mathcal{A}_z < 0$ (resp. $\mathcal{A}_g < 0$), demographics influence the speed of the drive propagation at least for $s \in (s_{2,z}, s_1)$ (resp. $s \in (s_{2,g}, s_1)$). More precisely, the sign of the speed can switch, changing the type of the invasion (drive or wild-type). When $\mathcal{A}_z > 0$ (resp. $\mathcal{A}_g > 0$) however, a model following only frequencies will always predict the correct speed of expansion. However, a model following only frequencies will not provide information on population size, and in particular whether the population is suppressed or eradicated, while this point is of great biological relevance.

For both zygote and germline conversion timings, the critical values of \mathcal{A}_z and \mathcal{A}_g can be interpreted

as the values at which the fitness of adults who were born heterozygous (f'_H) is the arithmetic mean of the fitness of adults born homozygote ($(f_D + f_W)/2$). The fitness of adults who were born heterozygous depends on the timing of gene conversion. For germline conversion, $f'_H = f_H$, and $\mathcal{A}_g = 0$ when $h = 1/2$, i.e. when there is co-dominance between the drive and wild-type alleles, i.e. when $f'_H = f_H = (f_D + f_W)/2$. For zygote conversion, the fitness of adults born heterozygous depends on whether gene conversion has taken place or not ($f'_H = (1 - c)(1 - hs) + c(1 - s)$). The condition $\mathcal{A}_z = 0$ is equivalent to $(1 - c)(1 - h) = 1/2$, which happens for $f'_H = (f_D + f_W)/2$.

2.4 Discussion

Following [95], we quantified the impact of demography in the case of the propagation of a super-Mendelian drive. We extended the analysis of reference [95] to the case of partial conversion ($0 < c < 1$), implying the presence of heterozygotes.

On the final state of the population

The final size of the population naturally varies. In case where no wild type can survive, the final size is the same regardless of the details of gene conversion (timing nor probability): $n_D^* = \min(0, 1 - \frac{s}{r(1-s)})$. In case of coexistence between wild-type and drive alleles, the final size depends on all parameters (see Appendix 2.F.2.2). Interestingly, in the case of coexistence, the drive allele can persist in the population even if a pure drive population would not ($n_D^* = 0$), see Figure 2.4 (note the area between the *composite persistence line* and the *pure drive persistence line*). In contrast with standard Mendelian genetics (corresponding to $c = 0$), coexistence can occur even if the dominance parameter is such that $h \in (0, 1)$ [69, 187]. More precisely, when conversion is partial and, either $h < 1 - \frac{1}{2(1-c)}$ (zygote conversion), or $h < \frac{1}{2}$ (germline conversion), there exists a stable coexistence state if s takes intermediate values $s \in (s_1, s_{2,z})$ (zygote conversion), or $s \in (s_1, s_{2,g})$ (germline conversion), where $s_1, s_{2,z}, s_{2,g}$ depend on (c, h) but do not depend on the demographic parameter r (see details in Section 2.3.2.1 and 2.3.2.2). While the final size of the population naturally depends on r .

On the transient regime (propagation of waves)

In order to evaluate the impact of demography on the dynamics of drive expansion, we compared the extreme cases $r \rightarrow \infty$ and $r \rightarrow 0$ (resp. low demographic variations versus large demographic variations).

For $r = 0$, we found that, when the drive propagates, it does so through a monostable and pulled wave. This happens when the drive is not too costly. In contrary, the drive gets uniformly extinct if it is too costly. The threshold on the fitness cost $s_{2,z}$ (zygote conversion), or $s_{2,g}$ (germline conversion), depends on (c, h) . The situation is analogous to the spatial spreading of an epidemic following a SI type model.

The case $r = 0$ gives the possibility to measure the importance of the demographic advection term $2\partial_x \log(n)\partial_x p$ when the problem is formulated in frequency, see equations (2.12), (2.21), (2.29). In fact, we show that ignoring this term can lead to an overestimation of the wave speed. This happens, for instance, in case of perfect conversion in the zygote, when $s \in (\frac{2}{5}, \frac{1}{2})$, then the equation (2.12) without $2\partial_x \log(n)\partial_x p$ would lead to a pushed front with velocity $\frac{2-3s}{\sqrt{2s}}$ [101]. However, we proved that

the front is actually pulled with velocity $2\sqrt{1-2s} < \frac{2-3s}{\sqrt{2s}}$. Intuitively, advection due to demographic variations slows down the expansion of the bulk. Noticeably, the effect is so strong that it prevents the front from being pushed.

In contrast, for $r = \infty$, the analysis boils down to a single equation on the drive allele frequency [201, 95]. According to [187], where the case of germline conversion was investigated, there is a large panel of frequency-dependence relationships, including monostable fixation of one allele, bi-stability, and stable coexistence between the two alleles, even if $h \in (0, 1)$. This leads to a variety of propagation phenomena, either pulled or pushed, as described in Section 2.3.2.2. The same panel of relationships arises in the case of zygote conversion, with qualitative similarities but quantitative differences in the thresholds and in the wave speeds, compare Section 2.3.2.2 with Section 2.3.2.1.

To connect $r = 0$ and $r = \infty$, we conjecture that, if the wave of the drive is pulled at $r = \infty$, then it is pulled for any value of $r > 0$, and the wave speed is independent of r . In particular, this occurs when the frequency-dependence term induces monostable dynamics and s is small enough, or when there is stable co-existence. This conjecture is supported by numerical investigations (Figures 2.4, 2.5, 2.6a and 2.6b). Still, the final size of population naturally depends on r .

Perspectives

We have focused on the classical dichotomy between pulled and pushed waves, even if the transition between the two is subject to current research both in theoretical studies [8, 13, 26], and in experimental works [65].

Pulled and pushed waves are associated with different outcomes on the maintenance of neutral diversity (which was not considered in our study). The genetic diversity of a population expanding by a pulled wave is very limited (with possible accumulation of deleterious mutations [178]), while more diversity is maintained under a pushed wave [188]. It could be interesting to investigate how gene conversion influences the maintenance of diversity along an expanding wave. More generally, the bottleneck following spread of a suppression drive will affect diversity, which may have long-lasting consequences even if wild-type individuals later recolonize the area.

It would be highly relevant to explore stochastic dynamics beyond our deterministic approach. When population sizes get to be small, as in the drive eradication case, large fluctuations and even chasing events are expected, as described in [46]. It would also be extremely interesting to extend the scope of the model, including by distinguishing between males and females which may have different fitnesses (especially in transgenic mosquitoes [17, 171, 133, 103]). Plural life stages or haploid phases might also influence modeling conclusions [143, 138].

2.5 Acknowledgements

This work is funded by ANR-19-CE45-0009-01 "TheoGeneDrive".

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 865711).

We are grateful to the INRAE MIGALE bioinformatics facility (MIGALE, INRAE, 2020. Migale bioinformatics Facility, doi: 10.15454/1.5572390655343293E12) for providing computing resources.

Appendices

2.A Model with partial conversion: growth term details

To obtain the global growth term for each genotypes in models (2.4) and (4.1), we calculate type proportions among the offspring for each possible couple, and then we sum the corresponding terms. The calculations follow standard lines of population genetics, differing only by the timing of gene conversion. When conversion occurs in the zygote, the parameter c appears in between the gametes and the offspring production, whereas when conversion occurs in the germline, it appears before gametes production. These equations in densities are consistent with the one obtained in frequency in the literature (when conversion occurs in the zygote [69, 213], or in the germline [69, 187]).

2.A.1 Conversion occurring in the zygote

Parents	Egg	Adult	Fitness	Growth term
WW + WW	$- 1 \rightarrow$ WW	$— 1 \rightarrow$ WW	1	$\frac{n_{WW}n_{WW}}{n}$
WW + WD	$\begin{array}{l} \swarrow \frac{1}{2} \rightarrow \text{WD} \\ \searrow \frac{1}{2} \rightarrow \text{WW} \end{array}$	$\begin{array}{l} \text{WD} \begin{array}{l} \swarrow c \rightarrow \text{DD} \\ \searrow 1-c \rightarrow \text{WD} \end{array} \\ \text{WW} \text{ — } 1 \rightarrow \text{WW} \end{array}$	$1-s$ $1-sh$ 1	$\frac{1}{2} c (1-s) \frac{2n_{WW}n_{DW}}{n}$ $\frac{1}{2} (1-c) (1-sh) \frac{2n_{WW}n_{DW}}{n}$ $\frac{1}{2} \frac{2n_{WW}n_{DW}}{n}$
WW + DD	$- 1 \rightarrow$ WD	$\begin{array}{l} \swarrow c \rightarrow \text{DD} \\ \searrow 1-c \rightarrow \text{WD} \end{array}$	$1-s$ $1-sh$	$c (1-s) \frac{2n_{WW}n_{DD}}{n}$ $(1-c) (1-sh) \frac{2n_{WW}n_{DD}}{n}$
WD + WD	$\begin{array}{l} \swarrow \frac{1}{4} \rightarrow \text{WW} \\ \swarrow \frac{1}{2} \rightarrow \text{WD} \\ \searrow \frac{1}{4} \rightarrow \text{DD} \end{array}$	$\begin{array}{l} \text{WW} \text{ — } 1 \rightarrow \text{WW} \\ \text{WD} \begin{array}{l} \swarrow c \rightarrow \text{DD} \\ \searrow 1-c \rightarrow \text{WD} \end{array} \\ \text{DD} \text{ — } 1 \rightarrow \text{DD} \end{array}$	1 $1-s$ $1-sh$ $1-s$	$\frac{1}{4} \frac{n_{DW}n_{DW}}{n}$ $\frac{1}{2} c (1-s) \frac{n_{DW}n_{DW}}{n}$ $\frac{1}{2} (1-c) (1-sh) \frac{n_{DW}n_{DW}}{n}$ $\frac{1}{4} (1-s) \frac{n_{DW}n_{DW}}{n}$
WD + DD	$\begin{array}{l} \swarrow \frac{1}{2} \rightarrow \text{WD} \\ \searrow \frac{1}{2} \rightarrow \text{DD} \end{array}$	$\begin{array}{l} \text{WD} \begin{array}{l} \swarrow c \rightarrow \text{DD} \\ \searrow 1-c \rightarrow \text{WD} \end{array} \\ \text{DD} \text{ — } 1 \rightarrow \text{DD} \end{array}$	$1-s$ $1-sh$ $1-s$	$\frac{1}{2} c (1-s) \frac{2n_{DW}n_{DD}}{n}$ $\frac{1}{2} (1-c) (1-sh) \frac{2n_{DW}n_{DD}}{n}$ $\frac{1}{2} (1-s) \frac{2n_{DW}n_{DD}}{n}$
DD + DD	$- 1 \rightarrow$ DD	$— 1 \rightarrow$ DD	$1-s$	$(1-s) \frac{n_{DD}n_{DD}}{n}$

Table 2.A.1: Growth term details when conversion occurs in the zygote.

2.A.2 Conversion occurring in the germline

Parents	Gametes	Adult	Fitness	Growth term
WW + WW	$\text{--- } 1 \rightarrow \text{W,W} + \text{W,W} \text{--- } 1 \rightarrow \text{WW}$		1	$\frac{n_{\text{WW}}n_{\text{WW}}}{n}$
WW + WD	$\begin{array}{l} \swarrow c \rightarrow \text{W,W} + \text{D,D} \text{--- } 1 \rightarrow \text{WD} \\ \searrow 1-c \rightarrow \text{W,W} + \text{W,D} \begin{array}{l} \text{--- } \frac{1}{2} \rightarrow \text{WW} \\ \text{--- } \frac{1}{2} \rightarrow \text{WD} \end{array} \end{array}$		$1 - sh$ 1 $1 - sh$	$c (1 - sh) \frac{2n_{\text{WW}}n_{\text{DW}}}{n}$ $(1 - c) \frac{1}{2} \frac{2n_{\text{WW}}n_{\text{DW}}}{n}$ $(1 - c) \frac{1}{2} (1 - sh) \frac{2n_{\text{WW}}n_{\text{DW}}}{n}$
WW + DD	$\text{--- } 1 \rightarrow \text{W,W} + \text{D,D} \text{--- } 1 \rightarrow \text{WD}$		$1 - sh$	$(1 - sh) \frac{2n_{\text{WW}}n_{\text{DD}}}{n}$
WD + WD	$\begin{array}{l} \swarrow c^2 \rightarrow \text{D,D} + \text{D,D} \text{--- } 1 \rightarrow \text{DD} \\ \begin{array}{l} \leq 2c(1-c) \rightarrow \text{D,D} + \text{W,D} \begin{array}{l} \text{--- } \frac{1}{2} \rightarrow \text{DD} \\ \text{--- } \frac{1}{2} \rightarrow \text{WD} \end{array} \\ \searrow (1-c)^2 \rightarrow \text{W,D} + \text{W,D} \begin{array}{l} \text{--- } \frac{1}{4} \rightarrow \text{DD} \\ \text{--- } \frac{1}{2} \rightarrow \text{WD} \\ \text{--- } \frac{1}{4} \rightarrow \text{WW} \end{array} \end{array}$		$1 - s$ $1 - s$ $1 - sh$ $1 - s$ $1 - sh$ 1	$c^2 (1 - s) \frac{n_{\text{DW}}n_{\text{DW}}}{n}$ $c (1 - c) (1 - s) \frac{n_{\text{DW}}n_{\text{DW}}}{n}$ $c (1 - c) (1 - sh) \frac{n_{\text{DW}}n_{\text{DW}}}{n}$ $(1 - c)^2 \frac{1}{4} (1 - s) \frac{n_{\text{DW}}n_{\text{DW}}}{n}$ $(1 - c)^2 \frac{1}{2} (1 - sh) \frac{n_{\text{DW}}n_{\text{DW}}}{n}$ $(1 - c)^2 \frac{1}{4} \frac{n_{\text{DW}}n_{\text{DW}}}{n}$
WD + DD	$\begin{array}{l} \swarrow c \rightarrow \text{D,D} + \text{D,D} \text{--- } 1 \rightarrow \text{DD} \\ \searrow 1-c \rightarrow \text{W,D} + \text{D,D} \begin{array}{l} \text{--- } \frac{1}{2} \rightarrow \text{WD} \\ \text{--- } \frac{1}{2} \rightarrow \text{DD} \end{array} \end{array}$		$1 - s$ $1 - sh$ $1 - s$	$c (1 - s) \frac{2n_{\text{DW}}n_{\text{DD}}}{n}$ $(1 - c) \frac{1}{2} (1 - sh) \frac{2n_{\text{DW}}n_{\text{DD}}}{n}$ $(1 - c) \frac{1}{2} (1 - s) \frac{2n_{\text{DW}}n_{\text{DD}}}{n}$
DD + DD	$\text{--- } 1 \rightarrow \text{D,D} + \text{D,D} \text{--- } 1 \rightarrow \text{DD}$		$1 - s$	$(1 - s) \frac{n_{\text{DD}}n_{\text{DD}}}{n}$

Table 2.A.2: Growth term details when conversion occurs in the germline.

2.B System rewritten with variables (n, p_D)

Below, we present the details of the reformulation from models (2.6), (2.17) and (4.4) in terms of total population density n and drive allele frequency p_D .

2.B.1 Model with perfect conversion

We rewrite model (2.6) with variables:

$$n = n_{\text{WW}} + n_{\text{DD}}, \quad p_D = \frac{n_D}{n_{\text{W}} + n_D} = \frac{n_{\text{DD}}}{n_{\text{WW}} + n_{\text{DD}}}. \quad (2.35)$$

where n is the total population density and p_D is the drive allele frequency, or equivalently in this model, the frequency of drive homogeneous individuals.

Equation on n :

$$\begin{aligned} \partial_t n - \partial_{xx}^2 n &= \left(r(1-n) + 1 \right) \left((1-s) \frac{n_{\text{DD}}^2 + 2 n_{\text{WW}} n_{\text{DD}}}{n_{\text{WW}} + n_{\text{DD}}} + \frac{n_{\text{WW}}^2}{n_{\text{WW}} + n_{\text{DD}}} \right) - n, \\ &= \left(r(1-n) + 1 \right) \left((1-s) p_D^2 + (1-s) 2 p_D (1-p_D) + (1-s)(1-p_D)^2 + s(1-p_D)^2 \right) n - n, \\ &= \left(r(1-n) + 1 \right) \left((1-s)(p_D + 1 - p_D)^2 + s(1-p_D)^2 \right) n - n, \\ &= \left(r(1-n) + 1 \right) \left((1-s) + s(1-p_D)^2 \right) n - n, \\ &= \left(r(1-n) + 1 \right) \left((1-s) p_D (2-p_D) + (1-p_D)^2 \right) n - n. \end{aligned} \quad (2.36)$$

Equation on n_{DD} :

$$\begin{aligned} \partial_t n_{\text{DD}} &= \partial_{xx}^2 n_{\text{DD}} + (1-s) \left(r(1-n_{\text{DD}} - n_{\text{WW}}) + 1 \right) \frac{n_{\text{DD}}^2 + 2 n_{\text{WW}} n_{\text{DD}}}{n_{\text{WW}} + n_{\text{DD}}} - n_{\text{DD}}, \\ &\stackrel{*}{=} p_D \partial_{xx}^2 n + 2 \partial_x n \partial_x p_D + n \partial_{xx}^2 p_D + (1-s) \left(r(1-n) + 1 \right) (2-p_D) n p_D - n p_D. \end{aligned} \quad (2.37)$$

Equation on $p_D = \frac{n_{\text{DD}}}{n}$:

$$\begin{aligned} \partial_t p_D &= \frac{n (\partial_t n_{\text{DD}}) - n_{\text{DD}} (\partial_t n)}{n^2} = \frac{n (\partial_t n_{\text{DD}}) - n p_D (\partial_t n)}{n^2} = \frac{1}{n} (\partial_t n_{\text{DD}} - p_D (\partial_t n)), \\ &= \frac{1}{n} \left[p_D \partial_{xx}^2 n + 2 \partial_x n \partial_x p_D + n \partial_{xx}^2 p_D + \left(r(1-n) + 1 \right) (1-s) (2-p_D) n p_D - n p_D \right] \\ &\quad - \frac{1}{n} \left[p_D \partial_{xx}^2 n + \left(r(1-n) + 1 \right) \left((1-s) p_D (2-p_D) + (1-p_D)^2 \right) n p_D - n p_D \right], \\ &= \partial_{xx}^2 p_D + 2 \partial_x \log(n) \partial_x p_D + \left(r(1-n) + 1 \right) p_D \left((1-s)(2-p_D) - (1-s)(2-p_D)p_D - (1-p_D)^2 \right), \\ &= \partial_{xx}^2 p_D + 2 \partial_x \log(n) \partial_x p_D + \left(r(1-n) + 1 \right) s p_D (1-p_D) \left(p_D - \frac{2s-1}{s} \right). \end{aligned} \quad (2.38)$$

Combining equations on n and p_D , we obtain model (2.12).

$$*\partial_{xx}^2 n_{\text{DD}} = \partial_{xx}^2 n p_D = \partial_x (p_D \partial_x n + n \partial_x p_D) = p_D \partial_{xx}^2 n + 2 \partial_x p_D \partial_x n + n \partial_{xx}^2 p_D$$

2.B.2 Model with partial conversion

2.B.2.1 Conversion in the zygote

We rewrite model (2.17) with variables:

$$n = n_W + n_D, \quad p_D = \frac{n_D}{n_W + n_D}. \quad (2.39)$$

where n is the total population density and p_D is the drive allele frequency.

Equation on n :

$$\begin{aligned} \partial_t n - \partial_{xx}^2 n &= \frac{r(1-n)+1}{n} \left((1-s)n_D^2 + [2c(1-s) + 2(1-sh)(1-c)]n_D n_W + n_W^2 \right) - n, \\ &= \frac{r(1-n)+1}{n} \left((1-s)(np_D)^2 + [2c(1-s) + 2(1-sh)(1-c)]p_D(1-p_D)n^2 + (1-p_D)^2 n^2 \right) - n, \\ &= (r(1-n)+1) \left((1-s)p_D^2 + [2c(1-s) + 2(1-sh)(1-c)]p_D(1-p_D) + (1-p_D)^2 \right) n - n, \\ &= (r(1-n)+1) \left((1-s)p_D^2 + 2p_D(1-p_D)[c(1-s) + (1-c)(1-sh)] + (1-p_D)^2 \right) n - n. \end{aligned} \quad (2.40)$$

Equation on n_D :

$$\begin{aligned} \partial_t n_D &= \partial_{xx}^2 n_D + \frac{r(1-n)+1}{n} \left[(1-s)(n_D + 2cn_W) + (1-sh)(1-c)n_W \right] n_D - n_D, \\ &= \partial_{xx}^2 n_D + \frac{r(1-n)+1}{n} \left[(1-s)(np_D + 2cn(1-p_D)) + (1-sh)(1-c)n(1-p_D) \right] n p_D - n p_D, \\ &\stackrel{\dagger}{=} p_D \partial_{xx}^2 n + 2 \partial_x n \partial_x p_D + n \partial_{xx}^2 p_D \\ &\quad + (r(1-n)+1) \left[(1-s)(p_D + 2c(1-p_D)) + (1-sh)(1-c)(1-p_D) \right] n p_D - n p_D. \end{aligned} \quad (2.41)$$

Equation on $p_D = \frac{n_D}{n}$:

$$\begin{aligned} \partial_t p_D &= \frac{n(\partial_t n_D) - n_D(\partial_t n)}{n^2} = \frac{n(\partial_t n_D) - n p_D(\partial_t n)}{n^2} = \frac{1}{n} (\partial_t n_D - p_D(\partial_t n)), \\ &= \frac{1}{n} \left[2 \partial_x n \partial_x p_D + n \partial_{xx}^2 p_D + (r(1-n)+1) \left((1-s)(p_D + 2c(1-p_D)) + (1-sh)(1-c)(1-p_D) \right. \right. \\ &\quad \left. \left. - (1-s)p_D^2 - 2p_D(1-p_D)[c(1-s) + (1-c)(1-sh)] - (1-p_D)^2 \right) p_D n \right], \\ &= 2 \partial_x \log(n) \partial_x p_D + \partial_{xx}^2 p_D + (r(1-n)+1) \left(p_D(1-s)(1-p_D) + 2(1-s)c(1-p_D) \right. \\ &\quad \left. + (1-sh)(1-c)(1-p_D) - 2p_D c(1-s)(1-p_D) - 2p_D(1-c)(1-sh)(1-p_D) - (1-p_D)^2 \right) p_D \Big], \\ &= 2 \partial_x \log(n) \partial_x p_D + \partial_{xx}^2 p_D + (r(1-n)+1) \left(p_D [(1-s)(1-2c) - 2(1-sh)(1-c) + 1] + 2(1-s)c \right. \\ &\quad \left. + (1-sh)(1-c) - 1 \right) (1-p_D) p_D, \\ &= 2 \partial_x \log(n) \partial_x p_D + \partial_{xx}^2 p_D \\ &\quad + (r(1-n)+1) \left([1-2(1-c)(1-h)]s p_D - s[1-(1-c)(1-h)] + c(1-s) \right) (1-p_D) p_D. \end{aligned} \quad (2.42)$$

Combining equations on n and p_D , we obtain model (2.21).

[†] $\partial_{xx}^2 n_D = \partial_{xx}^2 n p_D = \partial_x(p_D \partial_x n + n \partial_x p_D) = p_D \partial_{xx}^2 n + 2 \partial_x p_D \partial_x n + n \partial_{xx}^2 p_D$

2.B.2.2 Conversion in the germline

We rewrite model (4.4) with variables:

$$n = n_W + n_D, \quad p_D = \frac{n_D}{n_W + n_D}. \quad (2.43)$$

where n is the total population density and p_D is the drive allele frequency.

Equation on n :

$$\begin{aligned} \partial_t n - \partial_{xx}^2 n &= \frac{r(1-n)+1}{n} \left((1-s)n_D^2 + 2(1-sh)n_D n_W + n_W^2 \right) - n, \\ &= \frac{r(1-n)+1}{n} \left((1-s)(np_D)^2 + 2(1-sh)p_D(1-p_D)n^2 + ((1-p_D)n)^2 \right) - n, \\ &= (r(1-n)+1) \left((1-s)p_D^2 + 2(1-sh)p_D(1-p_D) + (1-p_D)^2 \right) n - n. \end{aligned} \quad (2.44)$$

Equation on n_D :

$$\begin{aligned} \partial_t n_D &= \partial_{xx}^2 n_D + \frac{r(1-n)+1}{n} \left[(1-s)n_D + (1-sh)(1+c)n_W \right] n_D - n_D, \\ &= p_D \partial_{xx}^2 n + 2 \partial_x n \partial_x p_D \\ &\quad + n \partial_{xx}^2 p_D + \frac{r(1-n)+1}{n} \left[(1-s)np_D + (1-sh)(1+c)n(1-p_D) \right] n p_D - n p_D, \\ &\stackrel{\ddagger}{=} p_D \partial_{xx}^2 n + 2 \partial_x n \partial_x p_D + n \partial_{xx}^2 p_D \\ &\quad + \left(r(1-n)+1 \right) \left[(1-s)p_D + (1-sh)(1+c)(1-p_D) \right] n p_D - n p_D. \end{aligned} \quad (2.45)$$

Equation on $p_D = \frac{n_D}{n}$:

$$\begin{aligned} \partial_t p_D &= \frac{n(\partial_t n_D) - n_D(\partial_t n)}{n^2} = \frac{n(\partial_t n_D) - n p_D(\partial_t n)}{n^2} = \frac{1}{n} \left(\partial_t n_D - p_D(\partial_t n) \right), \\ &= \frac{1}{n} \left[2 \partial_x n \partial_x p_D + n \partial_{xx}^2 p_D + (r(1-n)+1) \left((1-s)p_D + (1-sh)(1+c)(1-p_D) \right. \right. \\ &\quad \left. \left. - (1-s)p_D^2 - 2(1-sh)p_D(1-p_D) - (1-p_D)^2 \right) p_D n \right], \\ &= 2 \partial_x \log(n) \partial_x p_D \\ &\quad + \partial_{xx}^2 p_D + (r(1-n)+1) \left((1-s)p_D + (1-sh)(1+c) - 2(1-sh)p_D - (1-p_D) \right) p_D (1-p_D), \\ &= 2 \partial_x \log(n) \partial_x p_D + \partial_{xx}^2 p_D + (r(1-n)+1) \left((2h-1)sp_D + (1-sh)(1+c) - 1 \right) p_D (1-p_D). \end{aligned} \quad (2.46)$$

Combining equations on n and p_D , we obtain model (2.29).

$\ddagger \partial_{xx}^2 n_D = \partial_{xx}^2 n p_D = \partial_x(p_D \partial_x n + n \partial_x p_D) = p_D \partial_{xx}^2 n + 2 \partial_x p_D \partial_x n + n \partial_{xx}^2 p_D$

2.C Proofs for model (2.12) with perfect conversion in the zygote

2.C.1 Numerical evidence for the continuity when $r \rightarrow 0$

In Figure 2.C.1, we plot the speed of the traveling wave solutions of model (2.12) for a range of r and s values, and for $r = 0$. A positive speed correspond to drive invasion.

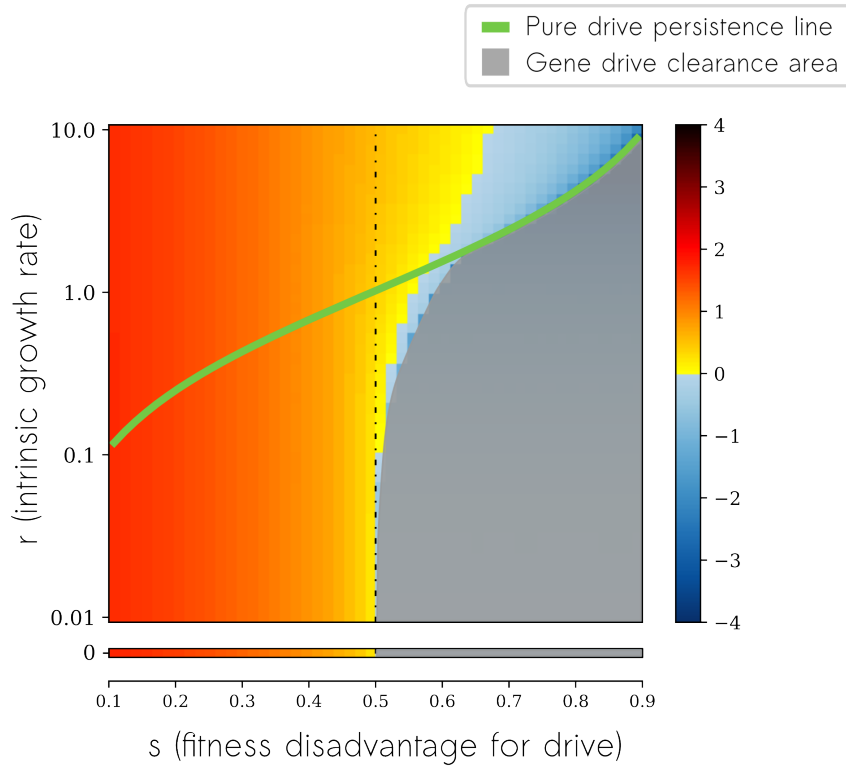


Figure 2.C.1: Wave speed values in model with perfect conversion in the zygote (2.12), regarding parameters r the intrinsic growth rate (log scale in between 0.01 and 10, plus the exact value $r = 0$ in the bottom color line) and s the fitness disadvantage for drive (normal scale). Below the pure drive persistence line (light green), a well-mixed population containing only drive homozygous individuals will necessarily go extinct.

We observe continuity in the speed value when $r \rightarrow 0$ away from $s = \frac{1}{2}$, meaning that the case $r = 0$ is relevant to approximate very small intrinsic growth rates.

2.C.2 Proof of the statements in Tables 2.3 and 2.4 when perfect conversion occurs in the zygote

In this section we prove the statements of Tables 2.3 and 2.4 on the two models of interest:

$\mathbf{r} = \infty$

$$\partial_t p - \partial_{xx}^2 p = \frac{s p (1-p) \left(p - \frac{2s-1}{s}\right)}{1-s+s(1-p)^2} = f^\infty(p). \quad (2.47)$$

$\mathbf{r} = \mathbf{0}$

$$\begin{cases} \partial_t n_{\text{DD}} - \partial_{xx}^2 n_{\text{DD}} &= (1-s) \frac{n_{\text{WW}} n_{\text{DD}}}{n_{\text{WW}} + n_{\text{DD}}} - s n_{\text{DD}} = f^0(n_{\text{DD}}, n_{\text{WW}}) \\ \partial_t n_{\text{WW}} - \partial_{xx}^2 n_{\text{WW}} &= \frac{-n_{\text{WW}} n_{\text{DD}}}{n_{\text{WW}} + n_{\text{DD}}}. \end{cases} \quad (2.48)$$

Monostable / Bistable

$\mathbf{r} = \infty$

$0 < s < 0.5$ The equation admits two admissible steady states 0 and 1. As $(f^\infty)'(0) > 0$ and $(f^\infty)'(1) < 0$, the only stable state is $p = 1$.

$0.5 < s < 1$ The equation admits three admissible steady states 0, $\frac{2s-1}{s}$ and 1. As $(f^\infty)'(0) < 0$, $(f^\infty)'(\frac{2s-1}{s}) > 0$ and $(f^\infty)'(1) < 0$, both $p = 0$ and $p = 1$ are stable states.

$\mathbf{r} = \mathbf{0}$

$0 < s < 0.5$ The system admits $(n_{\text{DD}} = 0, n_{\text{WW}} \in [0, 1])$ as admissible steady states. The Jacobian matrix, when switching to n and p_{D} variables, indicates that the only stable state is $(n = 0, p_{\text{D}} = 1)$, i.e. $(n_{\text{DD}} = 0, n_{\text{WW}} = 0)$.

$0.5 < s < 1$ The system admits $(n_{\text{DD}} = 0, n_{\text{WW}} \in [0, 1])$ as admissible steady states. The Jacobian matrix, when switching to n and p_{D} variables, indicates that the stable states are $(n = 0, p_{\text{D}} = 1)$ and $(n \in [0, 1], p_{\text{D}} = 0)$, i.e. $(n_{\text{DD}} = 0, n_{\text{WW}} \in [0, 1])$.

Existence of critical traveling waves

$\mathbf{r} = \infty$

The existence of traveling waves for the scalar equation (2.13) in both monostable and bistable cases is a classical result in the theory of reaction-diffusion equations, see for instance the seminal works in [11, 12].

$\mathbf{r} = \mathbf{0}$

$0 < s < 0.5$ We apply the results of Appendix 2.D with $\beta_1 = 1$ and $\beta_2 = 1 - s$. Therefore system (2.48) admits a traveling wave when $0 < s < 0.5$.

$0.5 < s < 1$ There is no drive propagation due to the gene drive clearance: the drive allele density decreases uniformly in space (details in section 2.C.2.1). Regarding the wild-type alleles, their dynamic is given by the heat equation implying only diffusion and no growth. It cannot admit traveling wave solutions.

Pulled/pushed waves and speed values

For both models, the speed of the linearized problem around zero density of drive allele is given by $2\sqrt{1-2s} = 2\sqrt{(f^\infty)'(0)} = 2\sqrt{\partial_p f^0(0,1)}$.

$r = \infty$

$0 < s \lesssim 0.35$ Numerically, we observe that the speed of the wave is equal to the minimal speed of the linearized problem: the wave is pulled (detail in section 2.C.2.2)

$0.35 \lesssim s < 0.5$ Numerically, we observe that the speed of the wave is strictly above the minimal speed of the linearized problem: the wave is pushed (detail in section 2.C.2.2).

$0.5 < s < 1$ As the system is bistable, the wave is necessarily pushed. The numerical approximation $s \approx 0.70$ indicating whether the drive of the wild-type population will invade the environment was already determined in the work of Tanaka et al [204].

$r = 0$

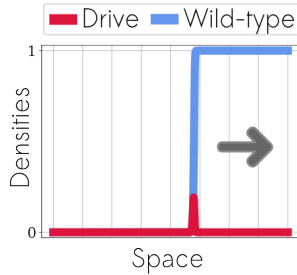
$0 < s < 0.5$ We apply the results of Appendix 2.D with $\beta_1 = 1$ and $\beta_2 = 1 - s$. Therefore system (2.48) admits a traveling wave with speed $v = 2\sqrt{1-2s}$ when $0 < s < 0.5$. This value corresponds to the KPP speed, by definition the wave is pulled.

$0.5 < s < 1$ No wave (see above, in Existence of critical traveling waves).

2.C.2.1 Gene drive clearance for $s \in (0.5, 1)$ when $r = 0$

Consider the model with perfect conversion in the zygote (2.6). The densities n_{WW} and n_{DD} dynamics are qualitatively given in Figure 2.C.2 for $r = 0$.

(a) Spreading eradication drive when $0 < s < 0.5$.



(b) Gene drive clearance when $0.5 < s < 1$

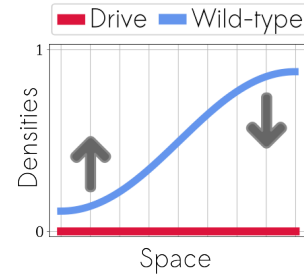


Figure 2.C.2: Qualitative dynamics of the drive homozygotes density n_{DD} (red line) and the wild-type homozygotes density n_{WW} (blue line) in space.

When $s > 0.5$ and $r = 0$, we observe gene drive clearance (in Figure 2.C.2b). More precisely, we have the following estimate, deduced from (2.15):

$$\partial_t n_{\text{DD}} - \partial_{xx}^2 n_{\text{DD}} \leq (1 - 2s) n_{\text{DD}}, \quad (2.49)$$

Therefore, n_{DD} is exponentially decaying in time, uniformly in space. The dynamics of the wild type then boil down to the standard heat equation, there cannot exist a traveling wave.

2.C.2.2 Numerical approximation of s threshold value for the pulled/pushed wave when $r = +\infty$

In order to determine an approximation of the threshold value at which the wave switches from a pulled wave to a pushed wave, we used the recent continuation procedure published in [13]. Figure 2.C.3 presents the value of the wave speed obtained via the latter continuation numerical scheme [114], for a wide range of s values. Notice the transition between pulled fronts (plain red) and pushed fronts (plain green). For the sake of clarity, the value of the minimal speed of the linearized problem $v = 2\sqrt{1-2s}$ is shown in red for $s \in (0, \frac{1}{2})$. Notice that the speed of the pushed front changes sign approximately at $s \approx 0.70$, in agreement with the theoretical criterion (2.9).

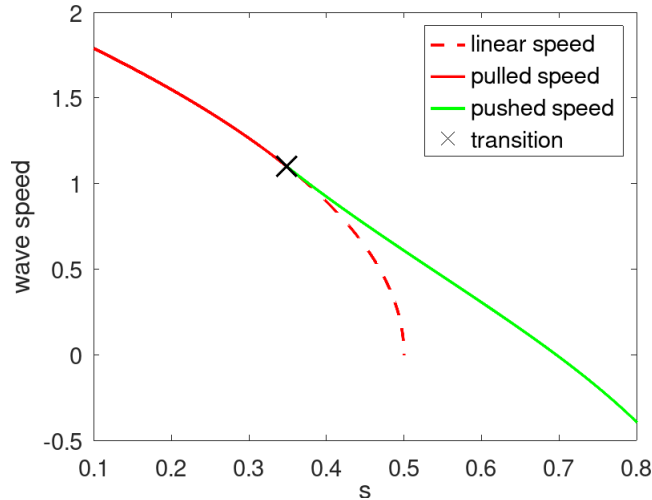


Figure 2.C.3: Value of the wave speed when $r = +\infty$ obtained via the numerical scheme in [114], for a wide range of s values. The transition between pulled fronts (plain red) and pushed fronts (plain green) is approximately 0.35. For the sake of clarity, the value of the minimal speed of the linearized problem $v = 2\sqrt{1-2s}$ is shown in red for $s \in (0, \frac{1}{2})$.

2.D Critical traveling wave for an SI similar model.

Consider the following epidemiological model,

$$\begin{cases} \partial_t S - \partial_{xx}^2 S &= -\beta \frac{S I}{S+I}, \\ \partial_t I - \partial_{xx}^2 I &= \beta \frac{S I}{S+I} - \gamma I. \end{cases} \quad (2.50)$$

where S is the density of susceptible individuals, I is the density of infected individuals, γ is the mortality of infected individuals and β is the transmission coefficient. This model has already been studied in the literature, see [225] and references therein. In particular, the existence of a minimal traveling wave has been established in the latter reference.

Models (2.15), (2.25) and (2.33) are very similar to the above SI model (2.50), except that the

coefficient β is different in the first and the second equation of the system. We write this new system with two coefficients β_1, β_2 :

$$\begin{cases} \partial_t S - \partial_{xx}^2 S &= -\beta_1 \frac{S I}{S + I}, \\ \partial_t I - \partial_{xx}^2 I &= \beta_2 \frac{S I}{S + I} - \gamma I. \end{cases} \quad (2.51)$$

2.D.1 Existence of critical traveling wave solutions

We are able to establish the following Theorem by adapting the proof in [225].

Theorem. *Suppose that $\beta_1 > 0$, and $\beta_2 > \gamma$, then system (2.51) admits a positive and bounded traveling wave solution with profile (S^*, I^*) , and speed $v = 2\sqrt{\beta_2 - \gamma}$. Furthermore, both S^* and I^* are positive, and bounded by 1 and $\frac{\beta_2 - \gamma}{\gamma}$ respectively.*

By adapting further the elements of [225], it would be possible to prove that the profile S^* is increasing, whereas the profil I^* is unimodal, as shown in Figure 2.D.1.

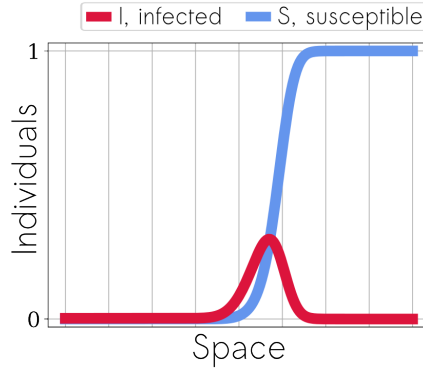


Figure 2.D.1: Qualitative shape of the solution (I^* in red, S^* in blue).

2.D.2 Proof of the theorem

We proceed as follows:

1. Although the system does not satisfy the comparison principle due to a lack of monotonicity, the construction of traveling waves is performed through a construction of sub-solutions ($\underline{S}, \underline{I}$) and super-solutions (\bar{S}, \bar{I}) for the system.
2. Using Schauder's fixed point theorem, we prove the existence of a critical traveling wave solution (S^*, I^*) with speed v such that $\underline{S}(z) \leq S^*(z) \leq \bar{S}(z)$ and $\underline{I}(z) \leq I^*(z) \leq \bar{I}(z)$ for all z in \mathbb{R} .
3. Finally, we conclude with the positivity of the critical traveling wave solution thanks to the strong maximum principle.

2.D.2.1 Construction of sub- and super-solutions

We are seeking sub- and super-solutions, respectively $(\underline{S}, \underline{I})$, (\bar{S}, \bar{I}) . Because of the non-monotonic coupling in the system, the following set of cross-relationships must be satisfied:

1. $-v \bar{S}' - \bar{S}'' \geq -\beta_1 \frac{\bar{S} I}{\bar{S} + I} \quad \forall \underline{I} \leq I \leq \bar{I}$;
2. $-v \underline{S}' - \underline{S}'' \leq -\beta_1 \frac{\underline{S} I}{\underline{S} + I} \quad \forall \underline{I} \leq I \leq \bar{I}$;
3. $-v \bar{I}' - \bar{I}'' \geq \beta_2 \frac{S \bar{I}}{S + \bar{I}} - \gamma \bar{I} \quad \forall \underline{S} \leq S \leq \bar{S}$;
4. $-v \underline{I}' - \underline{I}'' \leq \beta_2 \frac{S \underline{I}}{S + \underline{I}} - \gamma \underline{I} \quad \forall \underline{S} \leq S \leq \bar{S}$.

Inequalities 1,2,3 and 4 are valid in a weak sense. As \underline{S} is a piece-wise differentiable function, the quantity \underline{S} consists of functions on each sub-interval with a Dirac mass at the point of C^1 discontinuity. However, the Dirac mass has the good sign in this case (the transition has a convex shape), hence the second derivative \underline{S}'' is non-negative in the sense of a measure. All signs are correct: \underline{I} has a convex transition also, and \bar{I} has a concave transition. This key point is equivalent to the standard principle in the theory of parabolic equations (widely used for reaction-diffusion equations involving comparison techniques): "the maximum of sub-solutions is a sub-solution" (here, $\underline{S}, \underline{I}$) and "the minimum of super-solutions is a super-solution" (here, \bar{S}, \bar{I}).

To define our sub and super-solutions, it is useful to introduce the the following family of functions $\mathcal{I}(z) = e^{-\lambda^* z}$, where λ^* is solution of the following dispersion equation:

$$(\lambda^*)^2 - v\lambda^* + (\beta_2 - \gamma) = 0. \quad (2.52)$$

They are solutions of the linearized problem

$$v\mathcal{I}' + \mathcal{I}'' + (\beta_2 - \gamma) \mathcal{I} = 0. \quad (2.53)$$

For the critical speed $v = 2\sqrt{(\beta_2 - \gamma)}$, the corresponding double root is $\lambda^* = \frac{v}{2} = \sqrt{(\beta_2 - \gamma)}$.

Lemma. *There exist two large enough constants $L_1 > 0$ and $L_2 > 0$, such that the functions $\bar{S}, \underline{S}, \bar{I}, \underline{I}$ defined below satisfy the conditions 1. 2. 3. and 4.:*

$$\bar{S} = 1. \quad (2.54) \quad \bar{I} = \begin{cases} M & \forall z \leq \frac{1}{\lambda^*}, \\ eM\lambda^* z e^{-\lambda^* z} & \forall z > \frac{1}{\lambda^*}. \end{cases} \quad (2.56)$$

$$\underline{S} = \begin{cases} 0 & \forall z \leq L_1 \log(L_1), \\ 1 - L_1 e^{-\frac{z}{L_1}} & \forall z > L_1 \log(L_1). \end{cases} \quad (2.55) \quad \underline{I} = \begin{cases} 0 & \forall z \leq \left(\frac{L_2}{eM\lambda^*}\right)^2, \\ (eM\lambda^* z - L_2\sqrt{z})e^{-\lambda^* z} & \forall z > \left(\frac{L_2}{eM\lambda^*}\right)^2. \end{cases} \quad (2.57)$$

$$\text{with } M = \frac{\beta_2 - \gamma}{\gamma} = \frac{(\lambda^*)^2}{\gamma}.$$

Proof. Before we proceed with the proof, we introduce the following set of conditions, which are more restrictive than 1,2,3,4, but may appear more useful at some point in the calculations. When satisfied, they clearly imply 1,2,3,4.

$$(i) \quad -v \bar{S}' - \bar{S}'' \geq 0;$$

$$(ii) \quad -v \underline{S}' - \underline{S}'' \leq -\beta_1 \bar{I};$$

$$(iii) \quad -v \bar{I}' - \bar{I}'' \geq (\beta_2 - \gamma) \bar{I};$$

$$(iv) \quad -v \underline{I}' - \underline{I}'' \leq \beta_2 \frac{\underline{S} \underline{I}}{\underline{S} + \underline{I}} - \gamma \underline{I}.$$

We verify each of the four conditions on \bar{S} , \underline{S} , \bar{I} , \underline{I} :

$$\underline{\text{Condition 1:}} \quad -v \bar{S}' - \bar{S}'' \geq -\beta_1 \frac{\bar{S} \bar{I}}{\bar{S} + \bar{I}}.$$

The constant function $\bar{S} = 1$ satisfies the more restrictive condition (i) $-v \bar{S}' - \bar{S}'' \geq 0$.

$$\underline{\text{Condition 2:}} \quad \underline{S}' - \underline{S}'' \leq -\beta_1 \frac{\underline{S} \underline{I}}{\underline{S} + \underline{I}}.$$

Let us take L_1 sufficiently large such that $\frac{1}{\lambda^*} < L_1 \log(L_1)$.

- For $z > L_1 \log(L_1)$:

Since $\bar{I} = e M \lambda^* z e^{\lambda^* z}$ and $\underline{S} = 1 - L_1 e^{-\frac{z}{L_1}}$, the condition (ii) $-v \underline{S}' - \underline{S}'' \leq -\beta_1 \bar{I}$ holds for a sufficiently large $L_1 > 0$:

$$-v \underline{S}' - \underline{S}'' = \left(\frac{1}{L_1} - v\right) e^{-\frac{z}{L_1}} \leq -\beta_1 e M \lambda^* z e^{-\lambda^* z} = -\beta_1 \bar{I}, \quad (2.58)$$

$$\iff \beta_1 e M \lambda^* z e^{\left(\frac{1}{L_1} - \lambda^*\right)z} \leq \left(v - \frac{1}{L_1}\right). \quad (2.59)$$

- For $z \leq L_1 \log(L_1)$:

The condition 2. is verified since $\underline{S} = 0$.

$$\underline{\text{Condition 3:}} \quad -v \bar{I}' - \bar{I}'' \geq \beta_2 \frac{\underline{S} \bar{I}}{\underline{S} + \bar{I}} - \gamma \bar{I}.$$

- For $z \leq \frac{1}{\lambda^*}$:

With $\bar{I} = M = \frac{\beta_2 - \gamma}{\gamma}$:

$$-v \bar{I}' - \bar{I}'' = 0 = \beta_2 \frac{M}{1+M} - \gamma M = \beta_2 \frac{\bar{S} \bar{I}}{\bar{S} + \bar{I}} - \gamma \bar{I} \geq \beta_2 \frac{S \bar{I}}{S + \bar{I}} - \gamma \bar{I} \quad \forall S \leq \bar{S}. \quad (2.60)$$

- For $z > \frac{1}{\lambda^*}$:

Since \bar{I} is proportional to $ze^{-\lambda^*z}$, and λ^* is precisely the double root of the characteristic equation (2.52), we deduce that condition (iii) $-v \bar{I}' - \bar{I}'' - (\beta_2 - \gamma) \bar{I} \geq 0$ is verified.

Condition 4: $-v \underline{I}' - \underline{I}'' \leq \beta_2 \frac{S \underline{I}}{S + \underline{I}} - \gamma \underline{I}$. Let us take L_2 sufficiently large such that $L_2 > eM\lambda^* \sqrt{L_1 \log(L_1)}$.

- For $z \leq \left(\frac{L_2}{e M \lambda^*}\right)^2$:

$\underline{I} = 0$ so condition 4. is satisfied.

- For $z > \left(\frac{L_2}{e M \lambda^*}\right)^2$:

The choice of L_2 implies $z > L_1 \log(L_1)$, which means $\underline{S} = 1 - L_1 e^{-\frac{z}{L_1}}$.

We can reformulate condition (iv) as follows:

$$-v \underline{I}' - \underline{I}'' \leq \beta_2 \frac{\underline{S} \underline{I}}{\underline{S} + \underline{I}} - \gamma \underline{I} \iff -v \underline{I}' - \underline{I}'' - (\beta_2 - \gamma) \underline{I} \leq -\beta_2 \frac{\underline{I}^2}{\underline{S} + \underline{I}}. \quad (2.61)$$

With $L_3 = eM\lambda^*$, and:

$$\underline{I} = [L_3 z - L_2 \sqrt{z}] e^{-\lambda^*z}, \quad (2.62)$$

$$\underline{I}' = [L_3 - L_2 \frac{1}{2\sqrt{z}}] e^{-\lambda^*z} - [L_3 z - L_2 \sqrt{z}] \lambda^* e^{-\lambda^*z}, \quad (2.63)$$

$$\underline{I}'' = [L_2 \frac{1}{4z\sqrt{z}}] e^{-\lambda^*z} - 2 [L_3 - L_2 \frac{1}{2\sqrt{z}}] \lambda^* e^{-\lambda^*z} + [L_3 z - L_2 \sqrt{z}] (\lambda^*)^2 e^{-\lambda^*z} \quad (2.64)$$

On the one hand, we obtain the following identities:

$$-v \underline{I}' - \underline{I}'' - (\beta_2 - \gamma) \underline{I} = e^{-\lambda^* z} \left[L_3 \left(-v + v \lambda^* z + 2 \lambda^* - (\lambda^*)^2 z - (\beta_2 - \gamma) z \right) \right. \quad (2.65)$$

$$\left. + L_2 \left(v \frac{1}{2 \sqrt{z}} - v \sqrt{z} \lambda^* - \frac{1}{4 z \sqrt{z}} - \lambda^* \frac{1}{\sqrt{z}} + (\lambda^*)^2 \sqrt{z} + (\beta_2 - \gamma) \sqrt{z} \right) \right], \quad (2.66)$$

$$= e^{-\lambda^* z} \left[L_3 \left((2\lambda^* - v) - z \left(-v\lambda^* + (\lambda^*)^2 + (\beta_2 - \gamma) \right) \right) \right. \quad (2.67)$$

$$\left. + L_2 \left(\sqrt{z} \left(-v \lambda^* + (\lambda^*)^2 + (\beta_2 - \gamma) \right) + \frac{1}{2 \sqrt{z}} (v - 2\lambda^*) - \frac{1}{4 z \sqrt{z}} \right) \right], \quad (2.68)$$

$$= -L_2 e^{-\lambda^* z} \frac{1}{4 z \sqrt{z}} \quad (2.69)$$

On the other hand, we have:

$$-\beta_2 \frac{\underline{I}^2}{\underline{S} + \underline{I}} = -\beta_2 \frac{[L_3 z - L_2 \sqrt{z}]^2 e^{-2\lambda^* z}}{1 - L_1 e^{-\frac{z}{L_1}} + [L_3 z - L_2 \sqrt{z}] e^{-\lambda^* z}}. \quad (2.70)$$

We resume with the reformulation (2.61), which is now equivalent to the following:

$$-L_2 e^{-\lambda^* z} \left(1 - L_1 e^{-\frac{z}{L_1}} + [L_3 z - L_2 \sqrt{z}] e^{-\lambda^* z} \right) \leq -4\beta_2 [L_3 z - L_2 \sqrt{z}]^2 e^{-2\lambda^* z} z \sqrt{z} \quad (2.71)$$

$$\iff 4\beta_2 [L_3 z - L_2 \sqrt{z}]^2 e^{-\lambda^* z} z \sqrt{z} - L_2 [L_3 z - L_2 \sqrt{z}] e^{-\lambda^* z} \leq L_2 \left(1 - L_1 e^{-\frac{z}{L_1}} \right), \quad (2.72)$$

$$\iff 4\beta_2 e^{-\lambda^* z} z^3 \sqrt{z} (L_3)^2 + e^{-\lambda^* z} \left((1 - 8\beta_2 z^2) L_3 L_2 z + (L_2)^2 \sqrt{z} (1 - 4\beta_2 z^2) \right) \leq L_2 \left(1 - L_1 e^{-\frac{z}{L_1}} \right). \quad (2.73)$$

We may increase L_2 such that $1 - 4\beta_2 \left(\frac{L_2}{eM\lambda^*} \right)^4 \leq 0$. Then, since $z > \left(\frac{L_2}{eM\lambda^*} \right)^2$:

$$1 - 8\beta_2 z^2 \leq 1 - 4\beta_2 z^2 < 1 - 4\beta_2 \left(\frac{L_2}{eM\lambda^*} \right)^4 \leq 0. \quad (2.74)$$

Since $(1 - 8\beta_2 z^2)$ and $(1 - 4\beta_2 z^2)$ are negative terms, we need to show $L_2 \left(1 - L_1 e^{-\frac{z}{L_1}} \right) \geq 4\beta_2 e^{-\lambda^* z} z^3 \sqrt{z} (L_3)^2$.

Let $g(z) = \beta_2 e^{-\lambda^* z} z^3 \sqrt{z} (L_3)^2$ be a $\mathcal{C}^1([0; +\infty[)$ function. Since $\lim_{z \rightarrow 0} (g(z)) = 0$ and $\lim_{z \rightarrow +\infty} (g(z)) = 0$ there exists a constant C (which is independent from L_2) such that $g(z) < C \quad \forall z \geq 0$. We finally increase L_2 so that condition (iv) is verified. \square

2.D.2.2 Existence and positivity of a critical traveling wave solution

Now, exactly as in [225], we are in a position to define a set of functions

$$\Gamma = \{(S, I) \in B_\mu(\mathbb{R}, \mathbb{R}^2) \mid \underline{S} \leq S \leq \bar{S}, \underline{I} \leq I \leq \bar{I}\},$$

where $B_\mu(\mathbb{R}, \mathbb{R}^2)$ is the set of two-component continuous functions with each component growing at infinity slower than $e^{\mu|z|}$, as well as an operator $F : \Gamma \rightarrow C(\mathbb{R}, \mathbb{R}^2)$ that will satisfy the assumptions of the Schauder fixed point theorem and whose fixed point in Γ will precisely be the solution (S, I) we seek. Note that the inequalities 1., 2., 3., 4. (beginning of Section 2.D.2.1) are precisely what we use to prove that $F(\Gamma) \subset \Gamma$. Details can be found in [225].

The positivity of both S and I comes from the use of the strong maximum principle, again exactly as in [225].

2.E Study of the reaction term when $r = +\infty$ in section 2.3.2

We are searching for conditions implying a pulled monostable wave, using criterion (2.10).

2.E.1 Conversion occurring in the zygote

We rewrite limit equation (2.22):

$$\partial_t p_D - \partial_{xx}^2 p_D = \frac{\left(- [2(1-c)(1-h) - 1] s p_D - s[1 - (1-c)(1-h)] + c(1-s) \right) (1-p_D) p_D}{-s[2(1-c)(1-h) - 1] p_D^2 - 2s[1 - (1-c)(1-h)] p_D + 1}. \quad (2.75)$$

With $\mathcal{A}_z := s [2(1-c)(1-h) - 1] \in [-s, s]$:

$$\partial_t p_D - \partial_{xx}^2 p_D = \frac{\left(-\mathcal{A}_z p_D + \frac{1}{2}(\mathcal{A}_z - s) + c(1-s) \right) (1-p_D) p_D}{-\mathcal{A}_z p_D^2 + (\mathcal{A}_z - s) p_D + 1}. \quad (2.76)$$

Note that the mean fitness $\mathcal{F}^z(p_D) = -\mathcal{A}_z p_D^2 + (\mathcal{A}_z - s) p_D + 1 \in [1-s, 1]^\S$. When $\mathcal{A}_z \neq 0$, equation (2.76) can be rewritten:

$$\partial_t p_D - \partial_{xx}^2 p_D = \frac{-\mathcal{A}_z (p_D - p_{Dz}^*) (1-p_D) p_D}{-\mathcal{A}_z p_D^2 + (\mathcal{A}_z - s) p_D + 1} \quad \text{with} \quad p_{Dz}^* := \frac{1}{2} + \frac{2c(1-s) - s}{2\mathcal{A}_z}. \quad (2.77)$$

Let us introduce $s_1 := \frac{c}{1-h(1-c)}$ and $s_{2,z} := \frac{c}{2c+h(1-c)}$. Note that $\mathcal{A}_z > 0 \iff s_1 < s_{2,z}$. We draw the reaction term regarding the sign of \mathcal{A}_z and the s values in Figure 2.E.1.

When $\mathcal{A}_z < 0$ and $s \in (s_{2,z}, s_1)$, equation (2.22) admits two stable steady states (bistability). The final proportion will then strongly depend on the initial condition. On the other hand, when $\mathcal{A}_z > 0$ and $s \in (s_1, s_{2,z})$, the only possible equilibrium state is a coexistence state: the final proportion p_D will be strictly in between 0 and 1.

Independently of the sign of \mathcal{A}_z , if $s < \min(s_1, s_{2,z})$ the only stable steady state is $p_D = 1$ meaning that for an initial condition outside of the steady states, we expect that the drive always invades the population. If $s > \max(s_1, s_{2,z})$, the only stable steady state is $p_D = 0$ meaning that for an initial condition outside of the steady states, we expect that the wild-type always invades the population.

^{\S} $\mathcal{F}^{z'}(p_D) = \mathcal{A}_z(1 - 2p_D) - s \leq 0$ therefore $\mathcal{F}^z(1) \leq \mathcal{F}^z(p_D) \leq \mathcal{F}^z(0)$.

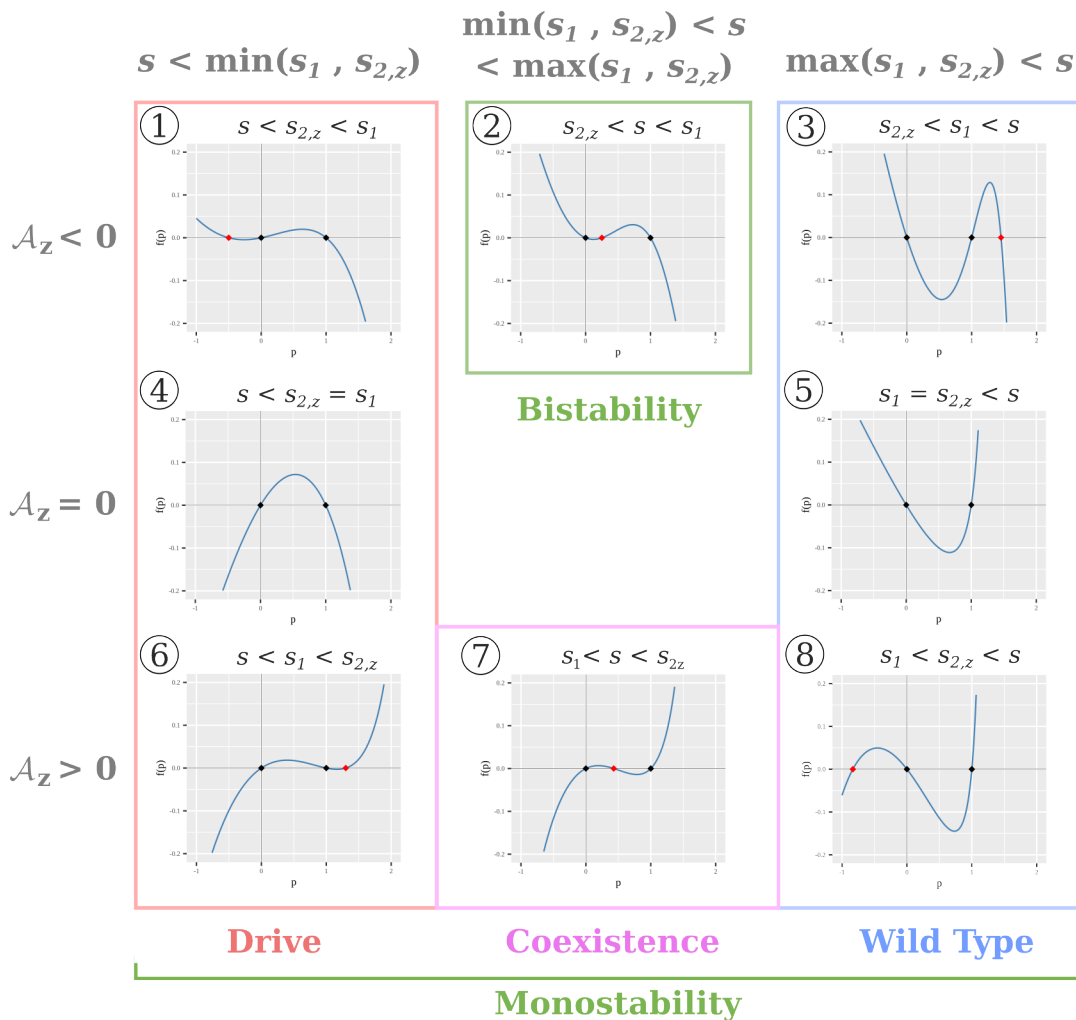


Figure 2.E.1: Reaction term of equation (2.76) regarding the sign of $\mathcal{A}_z = s[2(1-c)(1-h) - 1]$ and the s values. The s threshold values are $s_1 = \frac{c}{1-h(1-c)}$ and $s_{2,z} = \frac{c}{2c+h(1-c)}$. The dots on the x axis correspond to the steady states: $p_D = 0$ and $p_D = 1$ in black, $p_{Dz}^* = \frac{1}{2} + \frac{2c(1-s)-s}{2\mathcal{A}_z}$ in red when it exists.

In case of bistability, the wave is always pushed [101]; we can dismiss condition ② in the research of pulled monostable waves. We use criterion (2.10) on monostable cases, i.e. drive invasion ① ④ ⑥, wild type invasion ③ ⑤ ⑧, and coexistence state ⑦ (the numbers refer to the subgraphs in Figure 2.E.1).

2.E.1.1 Monostable drive invasion

④ **When** $\mathcal{A}_z = 0$ **and** $s < s_1 = s_{2,z} = \frac{2c}{2c+1} \iff s < 2c(1-s)$

From equation (2.76), we have for all $p_D \in [0, 1]$:

$$\sigma(0) - (1 - p_D) \sigma(p_D) = \left(c(1-s) - \frac{s}{2} \right) - \frac{\left(c(1-s) - \frac{s}{2} \right) (1 - p_D)}{1 - sp_D} = \left(c(1-s) - \frac{s}{2} \right) \frac{(1-s)p_D}{1 - sp_D} \geq 0, \quad (2.78)$$

where σ is the selection term defined by equation (2.8). Criterion (2.10) is verified.

⑥ **When** $\mathcal{A}_z > 0$ **and** $p_{Dz}^* > 1$

From equation (2.77), we have for all $p_D \in [0, 1]$:

$$\sigma(0) - (1 - p_D) \sigma(p_D) = \mathcal{A}_z p_{Dz}^* - \frac{-\mathcal{A}_z (p_D - p_{Dz}^*) (1 - p_D)}{-\mathcal{A}_z p_D^2 + (\mathcal{A}_z - s) p_D + 1} = \mathcal{A}_z p_D \frac{-(\mathcal{A}_z p_{Dz}^* + 1) p_D + (\mathcal{A}_z + 1 - s) p_{Dz}^* + 1}{-\mathcal{A}_z p_D^2 + (\mathcal{A}_z - s) p_D + 1}. \quad (2.79)$$

Note that $-\mathcal{A}_z p_D^2 + (\mathcal{A}_z - s) p_D + 1 > (1 - s) > 0$ and $\mathcal{A}_z p_D > 0$. The affine term $-(\mathcal{A}_z p_{Dz}^* + 1) p_D + (\mathcal{A}_z + 1 - s) p_{Dz}^* + 1$ decreases with p_D . In order to show that it is positive for all $p_D \in [0, 1]$, we just need to verify that this it is true for $p_D = 1$:

$$-(\mathcal{A}_z p_{Dz}^* + 1) + (\mathcal{A}_z + 1 - s) p_{Dz}^* + 1 = (1 - s) p_{Dz}^* \geq 0 \quad \Rightarrow \quad \sigma(0) - (1 - p_D) \sigma(p_D) \geq 0 \quad \forall p_D \in [0, 1]. \quad (2.80)$$

Criterion (2.10) is verified.

① **When** $\mathcal{A}_z < 0$ **and** $p_{Dz}^* < 0$ **and** $s < s_{2,z} \iff 0 < c - 2sc - sh + sch$

We consider equation (2.79) with $-\mathcal{A}_z p_D^2 + (\mathcal{A}_z - s) p_D + 1 > 1 - s > 0$ and $\mathcal{A}_z p_D < 0$. The affine term $-(\mathcal{A}_z p_{Dz}^* + 1) p_D + (\mathcal{A}_z + 1 - s) p_{Dz}^* + 1$ decreases with p_D . In order to show that it is negative for all $p_D \in [0, 1]$, we introduce a condition implying the negativity for $p_D = 0$:

$$(\mathcal{A}_z + 1 - s) p_{Dz}^* + 1 = \frac{(\mathcal{A}_z + 1 - s) (\mathcal{A}_z + 2c(1-s) - s + 2\mathcal{A}_z)}{2\mathcal{A}_z} < 0 \quad (2.81)$$

$$\iff \left(1 - 2s[1 - (1-h)(1-c)] \right) \left(c - 2sc - sh + sch \right) + s [2(1-c)(1-h) - 1] > 0 \quad (2.82)$$

Criterion (2.10) is verified when condition (2.82) is true.

2.E.1.2 Monostable wild-type invasion

In case of a monostable wild-type invasion, we need to consider the wild-type proportion $p_W = 1 - p_D \in [0, 1]$ and rewrite the equation (2.76):

$$\begin{aligned} -\partial_t p_W + \partial_{xx}^2 p_W &= \frac{(-\mathcal{A}_z (1 - p_W) + \frac{1}{2} (\mathcal{A}_z - s) + c(1 - s)) (1 - p_W) p_W}{-\mathcal{A}_z (1 - p_W)^2 + (\mathcal{A}_z - s)(1 - p_W) + 1} \\ \iff \partial_t p_W - \partial_{xx}^2 p_W &= \frac{(-\mathcal{A}_z p_W + \frac{1}{2} (\mathcal{A}_z + s) - c(1 - s)) (1 - p_W) p_W}{-\mathcal{A}_z p_W^2 + (\mathcal{A}_z + s) p_W + (1 - s)} \end{aligned} \quad (2.83)$$

When $\mathcal{A}_z \neq 0$, equation (2.83) can be rewritten:

$$\partial_t p_W - \partial_{xx}^2 p_W = \frac{-\mathcal{A}_z (p_W - p_{Wz}^*) (1 - p_W) p_W}{-\mathcal{A}_z p_W^2 + (\mathcal{A}_z + s) p_W + (1 - s)} \quad \text{with} \quad p_{Wz}^* = \frac{1}{2} - \frac{2c(1 - s) - s}{2\mathcal{A}_z} = 1 - p_{Dz}^* \quad (2.84)$$

⑤ **When $\mathcal{A}_z = 0$ and $s < s_1 = s_{2,z} = \frac{2c}{2c+1} \iff 2c(1 - s) < s$**

From equation (2.83) we have for all $p_W \in [0, 1]$:

$$\sigma(0) - (1 - p_W) \sigma(p_W) = \left(\frac{s}{2} - c(1 - s)\right) \left(\frac{1}{1 - s} - \frac{1 - p_W}{sp_W + 1 - s}\right) \geq \frac{p_W \left(\frac{s}{2} - c(1 - s)\right)}{1 - s} \geq 0 \quad (2.85)$$

Criterion (2.10) is verified.

⑧ **When $\mathcal{A}_z > 0$ and $p_{Wz}^* = 1 - p_{Dz}^* > 1$**

From equation (2.84), we have for all $p_W \in [0, 1]$:

$$\begin{aligned} \sigma(0) - (1 - p_W) \sigma(p_W) &= \frac{\mathcal{A}_z p_{Wz}^*}{1 - s} - \frac{-\mathcal{A}_z (p_W - p_{Wz}^*) (1 - p_W)}{-\mathcal{A}_z p_W^2 + (\mathcal{A}_z + s) p_W + (1 - s)} \\ &= \mathcal{A}_z p_W \frac{(-p_{Wz}^* + 1 - s) p_W + p_{Wz}^* (\mathcal{A}_z + 1) + (1 - s)}{(1 - s)(-\mathcal{A}_z p_W^2 + (\mathcal{A}_z + s) p_W + (1 - s))}. \end{aligned} \quad (2.86)$$

Note that $(1 - s)(-\mathcal{A}_z p_W^2 + (\mathcal{A}_z + s) p_W + (1 - s)) > (1 - s)^2 > 0$ and $\mathcal{A}_z p_W > 0$. As $p_{Wz}^* > 1$, the affine term $(-p_{Wz}^* + 1 - s) p_W + p_{Wz}^* (\mathcal{A}_z + 1) + (1 - s)$ decreases with p_W . In order to show that it is positive for all $p_W \in [0, 1]$, we just need to verify that this is true for $p_W = 1$:

$$(-p_{Wz}^* + 1 - s) + p_{Wz}^* (\mathcal{A}_z + 1) + (1 - s) = 2(1 - s) + \mathcal{A}_z p_{Wz}^* \geq 0 \quad \Rightarrow \quad \sigma(0) - (1 - p_W) \sigma(p_W) \geq 0 \quad \forall p_W \in [0, 1]. \quad (2.87)$$

Criterion (2.10) is verified.

③ **When** $\mathcal{A}_z < 0$ **and** $p_{Wz}^* = 1 - p_{Dz}^* < 0$

We consider equation (2.86) with $(1-s)(-\mathcal{A}_z p_W^2 + (\mathcal{A}_z + s) p_W + (1-s)) > (1-s)^2 > 0$ and $\mathcal{A}_z p_W < 0$. The affine term $(-p_{Wz}^* + 1 - s) p_W + p_{Wz}^* (\mathcal{A}_z + 1) + (1-s)$ is strictly positive for $p_W = 1$, therefore criterion (2.10) is not verified.

2.E.1.3 Monostable coexistence state

⑦ **When** $\mathcal{A}_z > 0$ **and** $0 < p_{Dz}^* = 1 - p_{Wz}^* < 1$

In the coexistence case, we have to verify that both waves, the drive invasion wave going to the right and the wild-type invasion wave going to the left, are pulled waves (see Figure 2.3).

For the drive invasion wave we consider equation (2.79) with $0 < p_{Dz}^* < 1$ and $p_D \in [0, p_{Dz}^*]$ (the term drive wave implies that the proportion of wild type increases after the wave passes; therefore the global stable steady state p_{Dz}^* is also the maximum proportion). Once again, we need to prove that the affine term $-(\mathcal{A}_z p_{Dz}^* + 1) p_D + (\mathcal{A}_z + 1 - s) p_{Dz}^* + 1$ is positive. As it decreases with $p_D \in [0, p_{Dz}^*]$, we determine its sign for $p_D = p_{Dz}^*$:

$$\begin{aligned} -(\mathcal{A}_z p_{Dz}^* + 1) p_{Dz}^* + (\mathcal{A}_z + 1 - s) p_{Dz}^* + 1 &= -\mathcal{A}_z (p_{Dz}^*)^2 + (\mathcal{A}_z - s) p_{Dz}^* + 1 \geq 1 - s \geq 0 \\ \Rightarrow \sigma(0) - (1 - p_D)\sigma(p_D) &\geq 0 \quad \forall p_D \in [0, p_{Dz}^*]. \end{aligned} \quad (2.88)$$

Criterion (2.10) is verified for the drive wave.

For the wild-type invasion wave, we consider equation (2.86) with $0 < p_{Wz}^* = 1 - p_{Dz}^* < 1$ and $p_W \in [0, p_{Wz}^*]$ (the term wild-type wave implies that the proportion of wild type increases after the wave passes; therefore the global stable steady state p_{Wz}^* is also the maximum proportion). Once again, we need to prove that the affine term $(-p_{Wz}^* + 1 - s) p_W + p_{Wz}^* (\mathcal{A}_z + 1) + (1-s)$ is positive. As it decreases with $p_W \in [0, p_{Wz}^*]$, we determine its sign for $p_W = p_{Wz}^*$:

$$\begin{aligned} (-p_{Wz}^* + 1 - s) p_{Wz}^* + (\mathcal{A}_z + 1) p_{Wz}^* + (1 - s) &= -(p_{Wz}^*)^2 + (\mathcal{A}_z + 2 - s) p_{Wz}^* + 1 - s \\ \geq \min(1, \mathcal{A}_z + 2(1 - s)) &\geq 0 \quad \Rightarrow \quad \sigma(0) - (1 - p_W)\sigma(p_W) \geq 0 \quad \forall p_W \in [0, p_{Wz}^*]. \end{aligned} \quad (2.89)$$

Criterion (2.10) is verified for the wild-type wave.

2.E.2 Conversion occurring in the germline

We rewrite limit equation (2.30):

$$\partial_t p_D - \partial_{xx}^2 p_D = \frac{\left(-(1-2h) s p_D + [(1-sh)(1+c) - 1] \right) p_D (1-p_D)}{-s(1-2h)p_D^2 - 2shp_D + 1} \quad (2.90)$$

With $\mathcal{A}_g := s(1-2h) \in [-s, s]$:

$$\partial_t p_D - \partial_{xx}^2 p_D = \frac{\left(-\mathcal{A}_g p_D + \frac{1}{2}(\mathcal{A}_g - s) + c(1-sh) \right) p_D (1-p_D)}{-\mathcal{A}_g p_D^2 + (\mathcal{A}_g - s) p_D + 1}. \quad (2.91)$$

Note that the mean fitness $\mathcal{F}^g(p_D) = -\mathcal{A}_g p_D^2 + (\mathcal{A}_g - s) p_D + 1 \in [1 - s, 1]$ [¶]. When $\mathcal{A}_g \neq 0$, equation (2.91) can be rewritten:

$$\partial_t p_D - \partial_{xx}^2 p_D = \frac{-\mathcal{A}_g (p_D - p_{D,g}^*) (1 - p_D) p_D}{-\mathcal{A}_g p_D^2 + (\mathcal{A}_g - s) p_D + 1} = f_g(p_D) \quad \text{with} \quad p_{D,g}^* := \frac{1}{2} + \frac{2c(1-sh) - s}{2\mathcal{A}_g}. \quad (2.92)$$

Let us introduce $s_1 := \frac{c}{1-h(1-c)}$ and $s_{2,g} := \frac{c}{2ch+h(1-c)} = \frac{c}{h(1+c)}$. We draw the reaction term regarding the sign of \mathcal{A}_g and the s values (in Figure 2.E.2).

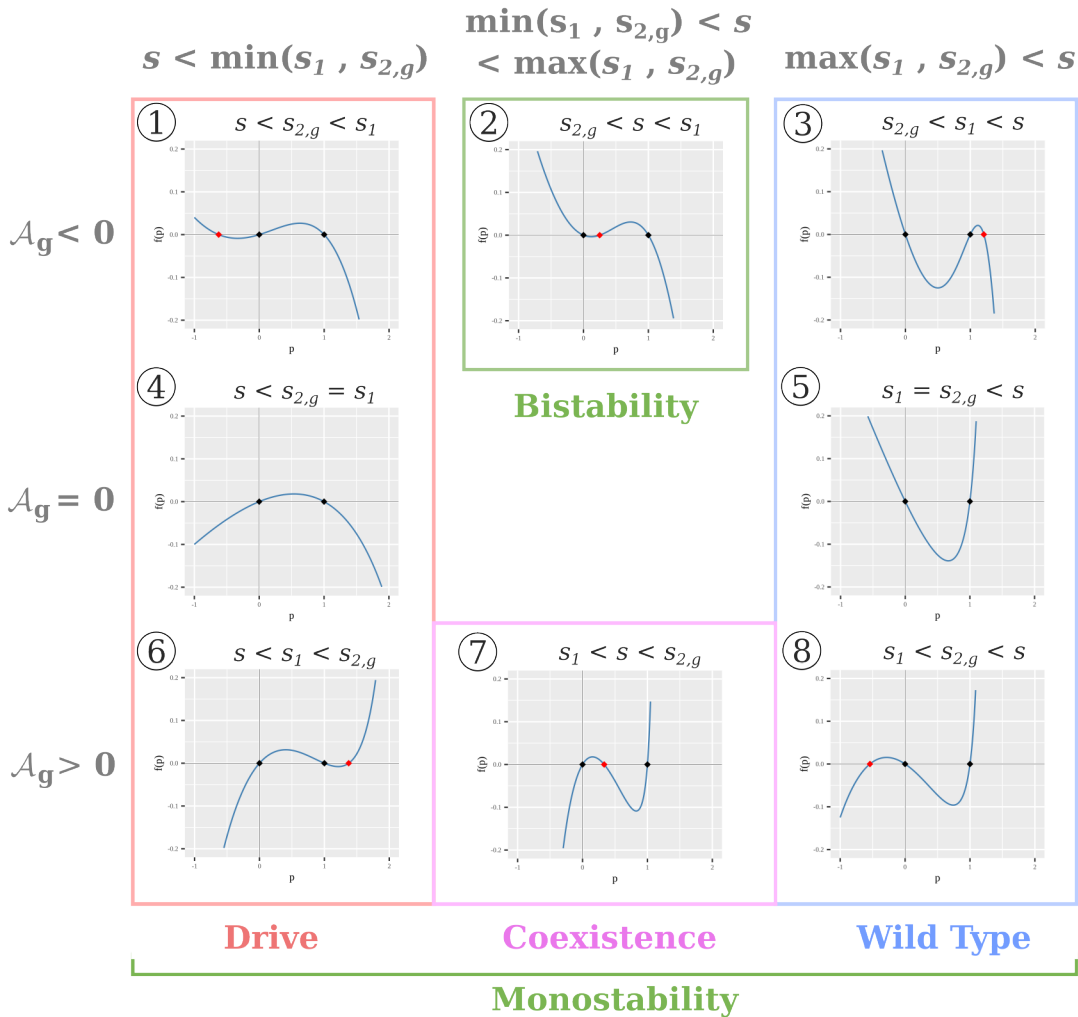


Figure 2.E.2: Reaction term of equation (2.91) regarding the sign of $\mathcal{A}_g = s(1 - 2h)$ and the s values. The s threshold values are $s_1 = \frac{c}{1-h(1-c)}$ and $s_{2,g} = \frac{c}{2ch+h(1-c)} = \frac{c}{h(1+c)}$. The dots on the x axis correspond to the steady states: $p_D = 0$ and $p_D = 1$ in black, $p_{D,g}^* = \frac{1}{2} + \frac{2c(1-sh) - s}{2\mathcal{A}_g}$ in red when it exists.

[¶] $\mathcal{F}^{g'}(p_D) = \mathcal{A}_g(1 - 2p_D) - s \leq 0$ therefore $\mathcal{F}^g(1) \leq \mathcal{F}^g(p_D) \leq \mathcal{F}^g(0)$.

In case of bistability, the wave is always pushed [101]; we can dismiss condition ② in the research of pulled monostable waves. We use criterion (2.10) on monostable cases, i.e. drive invasion ① ④ ⑥, wild type invasion ③ ⑤ ⑧, and coexistence state ⑦ (the numbers refer to the subgraphs in Figure 2.E.2).

2.E.2.1 Monostable drive invasion

$$\textcircled{4} \text{ When } \mathcal{A}_g = 0 \iff h = \frac{1}{2} \text{ and } s < s_1 = s_{2,g} = \frac{2c}{c+1} \iff \frac{s}{2} (c+1) < c$$

From equation (2.91), we have for all $p_D \in [0, 1]$:

$$\sigma(0) - (1-p_D)\sigma(p) = \left(c - \frac{s}{2}(c+1)\right) - \frac{\left(c - \frac{s}{2}(c+1)\right)(1-p_D)}{1-sp_D} = \left(c - \frac{s}{2}(c+1)\right) \frac{(1-s)p_D}{1-sp_D} \geq 0, \quad (2.93)$$

where σ is the selection term defined by equation (2.8). Criterion (2.10) is verified.

$$\textcircled{6} \text{ When } \mathcal{A}_g > 0 \text{ and } p_{Dg}^* > 1$$

From equation (2.92), we have for all $p_D \in [0, 1]$:

$$\sigma(0) - (1-p_D)\sigma(p) = \mathcal{A}_g p_{Dg}^* - \frac{-\mathcal{A}_g (p_D - p_{Dg}^*) (1-p_D)}{-\mathcal{A}_g p_D^2 + (\mathcal{A}_g - s) p_D + 1} = \mathcal{A}_g p_D \frac{-(\mathcal{A}_g p_{Dg}^* + 1) p_D + (\mathcal{A}_g + 1 - s) p_{Dg}^* + 1}{-\mathcal{A}_g p_D^2 + (\mathcal{A}_g - s) p_D + 1}. \quad (2.94)$$

Note that $-\mathcal{A}_g p_D^2 + (\mathcal{A}_g - s) p_D + 1 > (1-s) > 0$ and $\mathcal{A}_g p_D > 0$. The affine term $-(\mathcal{A}_g p_{Dg}^* + 1) p_D + (\mathcal{A}_g + 1 - s) p_{Dg}^* + 1$ decreases with p_D . In order to show that it is positive for all $p_D \in [0, 1]$, we just need to verify that this it is true for $p_D = 1$:

$$-(\mathcal{A}_g p_{Dg}^* + 1) + (\mathcal{A}_g + 1 - s) p_{Dg}^* + 1 = (1-s) p_{Dg}^* \geq 0 \implies \sigma(0) - (1-p_D)\sigma(p) \geq 0 \quad \forall p_D \in [0, 1]. \quad (2.95)$$

Criterion (2.10) is verified.

$$\textcircled{1} \text{ When } \mathcal{A}_g < 0 \text{ and } p_{Dg}^* < 0 \text{ and } s < s_{2,g} \iff 0 < c - sh(1+c)$$

We consider equation (2.94) with $-\mathcal{A}_g p_D^2 + (\mathcal{A}_g - s) p_D + 1 > 1 - s > 0$ and $\mathcal{A}_g p_D < 0$. The affine term $-(\mathcal{A}_g p_{Dg}^* + 1) p_D + (\mathcal{A}_g + 1 - s) p_{Dg}^* + 1$ decreases with p_D . In order to show that it is negative for all $p_D \in [0, 1]$, we introduce a condition implying the negativity for $p = 0$:

$$(\mathcal{A}_g + 1 - s) p_{Dg}^* + 1 = \frac{(1-2sh)(\mathcal{A}_g + 2c(1-sh) - s) + 2\mathcal{A}_g}{2\mathcal{A}_g} < 0 \quad (2.96)$$

$$\iff (1-2sh)(c - sh(c+1)) + s(1-2h) > 0 \quad (2.97)$$

Criterion (2.10) is verified when condition (2.97) is true.

2.E.2.2 Monostable wild-type invasion

In case of a monostable wild-type invasion, we need to consider the wild-type proportion $p_W = 1 - p_D \in [0, 1]$ and rewrite the equation (2.91):

$$\begin{aligned} -\partial_t p_W + \partial_{xx}^2 p_W &= \frac{(-\mathcal{A}_g (1 - p_W) + \frac{1}{2} (\mathcal{A}_g - s) + c(1 - sh)) (1 - p_W) p_W}{-\mathcal{A}_g (1 - p_W)^2 + (\mathcal{A}_g - s)(1 - p_W) + 1} \\ \iff \partial_t p_W - \partial_{xx}^2 p_W &= \frac{(-\mathcal{A}_g p_W + \frac{1}{2} (\mathcal{A}_g + s) - c(1 - sh)) (1 - p_W) p_W}{-\mathcal{A}_g p_W^2 + (\mathcal{A}_g + s) p_W + (1 - s)} \end{aligned} \quad (2.98)$$

When $\mathcal{A}_g \neq 0$, equation (2.98) can be rewritten:

$$\partial_t p_W - \partial_{xx}^2 p_W = \frac{-\mathcal{A}_g (p_W - p_{Wg}^*) (1 - p_W) p_W}{-\mathcal{A}_g p_W^2 + (\mathcal{A}_g + s) p_W + (1 - s)} \quad \text{with} \quad p_{Wg}^* = \frac{1}{2} - \frac{2c(1 - sh) - s}{2\mathcal{A}_g} = 1 - p_{Dg}^* \quad (2.99)$$

⑤ **When $\mathcal{A}_g = 0 \iff h = \frac{1}{2}$ and $s_1 = s_{2,g} = \frac{2c}{c+1} < s \iff c < \frac{s}{2} (c+1)$**

From equation (2.98) we have for all $p_W \in [0, 1]$:

$$\sigma(0) - (1 - p_W)\sigma(p_W) = \left(\frac{s}{2} (c+1) - c\right) \left(\frac{1}{1-s} - \frac{1-p_W}{sp_W + 1-s}\right) \geq \frac{p_W \left(\frac{s}{2} (c+1) - c\right)}{1-s} \geq 0 \quad (2.100)$$

Criterion (2.10) is verified.

⑧ **When $\mathcal{A}_g > 0$ and $p_{Wg}^* = 1 - p_{Dg}^* > 1$**

From equation (2.99), we have for all $p_W \in [0, 1]$:

$$\begin{aligned} \sigma(0) - (1 - p_W)\sigma(p_W) &= \frac{\mathcal{A}_g p_{Wg}^*}{1-s} - \frac{-\mathcal{A}_g (p_W - p_{Wg}^*) (1 - p_W)}{-\mathcal{A}_g p_W^2 + (\mathcal{A}_g + s) p_W + (1 - s)} \\ &= \mathcal{A}_g p_W \frac{(-p_{Wg}^* + 1 - s) p_W + p_{Wg}^* (\mathcal{A}_g + 1) + (1 - s)}{(1-s)(-\mathcal{A}_g p_W^2 + (\mathcal{A}_g + s) p_W + (1 - s))}. \end{aligned} \quad (2.101)$$

Note that $(1-s)(-\mathcal{A}_g p_W^2 + (\mathcal{A}_g + s) p_W + (1-s)) > (1-s)^2 > 0$ and $\mathcal{A}_g p_W > 0$. As $p_{Wg}^* > 1$, the affine term $(-p_{Wg}^* + 1 - s) p_W + p_{Wg}^* (\mathcal{A}_g + 1) + (1 - s)$ decreases with p_W . In order to show that it is positive for all $p_W \in [0, 1]$, we just need to verify that this it is true for $p_W = 1$:

$$(-p_{Wg}^* + 1 - s) + p_{Wg}^* (\mathcal{A}_g + 1) + (1 - s) = 2(1-s) + \mathcal{A}_g p_{Wg}^* \geq 0 \quad \Rightarrow \quad \sigma(0) - (1 - p_W)\sigma(p_W) \geq 0 \quad \forall p_W \in [0, 1]. \quad (2.102)$$

Criterion (2.10) is verified.

③ **When** $\mathcal{A}_g < 0$ **and** $p_{wg}^* = 1 - p_{Dz}^* < 0$

We consider equation (2.101) with $(1-s)(-\mathcal{A}_g p_w^2 + (\mathcal{A}_g + s) p_w + (1-s)) > (1-s)^2 > 0$ and $\mathcal{A}_g p_w < 0$. The affine term $(-p_{wg}^* + 1 - s) p_w + p_{wg}^*(\mathcal{A}_g + 1) + (1 - s)$ is strictly positive for $p_w = 1$, therefore criterion (2.10) is not verified.

2.E.2.3 Monostable coexistence state

⑦ **When** $\mathcal{A}_g > 0$ **and** $0 < p_{Dg}^* = 1 - p_{wg}^* < 1$

In the coexistence case, we have to verify that both sub-traveling waves, the drive invasion wave going to the right and the wild-type invasion wave going to the left, are pulled waves (see Figure 2.3).

For the drive invasion wave we consider equation (2.94) with $0 < p_{Dg}^* < 1$ and $p_D \in [0, p_{Dg}^*]$ (the term drive wave implies that the proportion of wild type increases after the wave passes; therefore the global stable steady state p_{Dg}^* is also the maximum proportion). Once again, we need to prove that the affine term $(-\mathcal{A}_g p_{Dg}^* + 1) p_D + (\mathcal{A}_g + 1 - s) p_{Dg}^* + 1$ is positive. As it decreases with $p_D \in [0, p_{Dg}^*]$, we determine its sign for $p_D = p_{Dg}^*$:

$$\begin{aligned} -(\mathcal{A}_g p_{Dg}^* + 1) p_{Dg}^* + (\mathcal{A}_g + 1 - s) p_{Dg}^* + 1 &= -\mathcal{A}_g (p_{Dg}^*)^2 + (\mathcal{A}_g - s) p_{Dg}^* + 1 \geq 1 - s \geq 0 \\ &\Rightarrow \sigma(0) - (1 - p_D)\sigma(p_D) \geq 0 \quad \forall p_D \in [0, p_{Dg}^*]. \end{aligned} \quad (2.103)$$

Criterion (2.10) is verified for the drive wave.

For the wild-type invasion wave, we consider equation (2.101) with $0 < p_{wg}^* = 1 - p_{Dg}^* < 1$ and $p_w \in [0, p_{wg}^*]$ (the term wild-type wave implies that the proportion of wild type increases after the wave passes; therefore the global stable steady state p_{wg}^* is also the maximum proportion). Once again, we need to prove that the affine term $(-p_{wg}^* + 1 - s) p_w + p_{wg}^*(\mathcal{A}_g + 1) + (1 - s)$ is positive. As it decreases with $p_w \in [0, p_{wg}^*]$, we determine its sign for $p_w = p_{wg}^*$:

$$\begin{aligned} (-p_{wg}^* + 1 - s) p_{wg}^* + (\mathcal{A}_g + 1) p_{wg}^* + (1 - s) &= -(p_{wg}^*)^2 + (\mathcal{A}_g + 2 - s) p_{wg}^* + 1 - s \\ &\geq \min(1, \mathcal{A}_g + 2(1 - s)) \geq 0 \quad \Rightarrow \sigma(0) - (1 - p_w)\sigma(p_w) \geq 0 \quad \forall p_w \in [0, p_{wg}^*]. \end{aligned} \quad (2.104)$$

Criterion (2.10) is verified for the wild-type wave.

2.F Heatmap supplementary materials

2.F.1 Effect of fitness disadvantage (s) and dominance coefficient (h) on drive dynamics, for $r = +\infty$.

In Figure 2.F.1 and 2.F.2, we compute heatmaps indicating the stability regime of systems (2.4) and (4.1) when $r = +\infty$, depending on the values of (h, s) and for a fixed value of c .

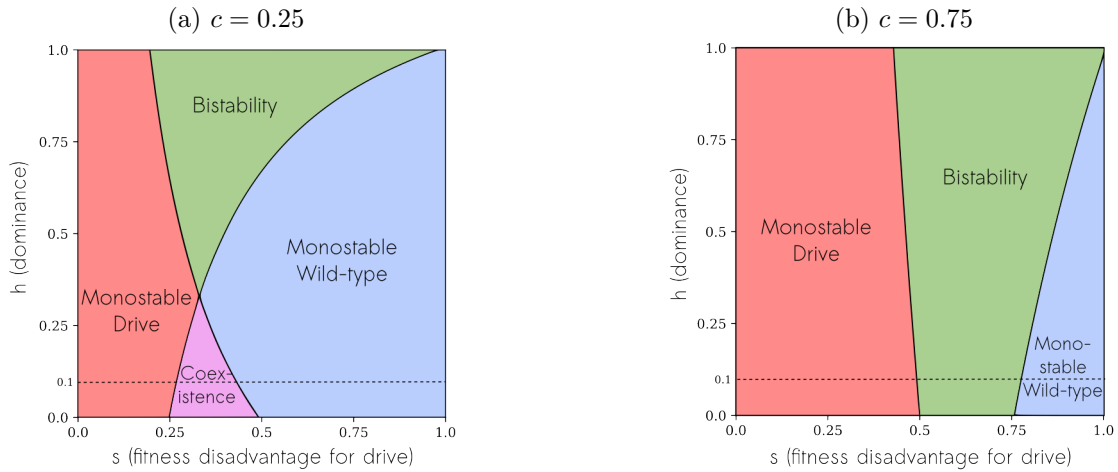


Figure 2.F.1: Effect of fitness disadvantage (s) and dominance coefficient (h) on drive dynamics for system (2.4) (when conversion occurs in the zygote) when $r = +\infty$. Parameters for Figure 2.4 ($c = 0.25$ and $h = 0.1$) and Figure 2.5 ($c = 0.75$ and $h = 0.1$) are materialized by dotted lines.

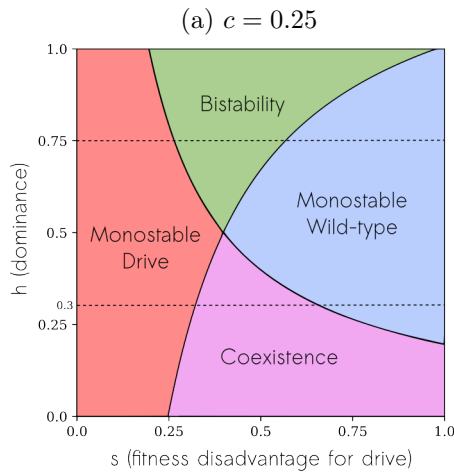


Figure 2.F.2: Effect of fitness disadvantage (s) and dominance coefficient (h) on drive dynamics for system (4.1) (when conversion occurs in the germline) when $r = +\infty$. Parameters for Figure 2.6a ($c = 0.25$ and $h = 0.3$) and Figure 2.6b ($c = 0.25$ and $h = 0.75$) are materialized by dotted lines.

A similar figure has already been computed in [187] for $c = 0.85$, and conversion occurring in the germline.

2.F.2 Heatmap lines

2.F.2.1 Pure drive line

Consider model (2.6) with $n_{\text{WW}} = 0$. A well-mixed population containing only drive homozygous individuals will persist in the environment if its equilibrium state n_{DD}^* is strictly positive, i.e. if:

$$n_{\text{DD}}^* = \min\left(0, 1 - \frac{s}{r(1-s)}\right) > 0 \iff r > \frac{s}{1-s} \quad (2.105)$$

In case of partial conversion, calculations give the same threshold (consider models (2.4) and (4.1) with $n_{\text{DW}} = 0$ and $n_{\text{WW}} = 0$).

2.F.2.2 Composite persistence line

Similarly, in case of coexistence, a well-mixed population will persist in the environment only if its equilibrium state n^* is strictly positive. Using Mathematica, we compute this population density equilibrium when conversion occurs in the zygote (n_z^*) or in the germline (n_g^*) based on systems (2.21) and (2.29). We obtain the following:

$$n_z^* = \min\left(0, 1 - \frac{1 - \mathcal{F}^z(p_{\text{Dz}}^*)}{r\mathcal{F}^z(p_{\text{Dz}}^*)}\right) \quad \text{and} \quad n_g^* = \min\left(0, 1 - \frac{1 - \mathcal{F}^g(p_{\text{Dg}}^*)}{r\mathcal{F}^g(p_{\text{Dg}}^*)}\right), \quad (2.106)$$

where the mean fitness \mathcal{F}^z and \mathcal{F}^g (already defined in Appendix 2.E) are given by:

$$\mathcal{F}^z(p_{\text{D}}) = -\mathcal{A}_z p_{\text{D}}^2 + (\mathcal{A}_z - s) p_{\text{D}} + 1 \quad \text{and} \quad \mathcal{F}^g(p_{\text{D}}) = -\mathcal{A}_g p_{\text{D}}^2 + (\mathcal{A}_g - s) p_{\text{D}} + 1, \quad (2.107)$$

and the proportions p_{Dz}^* and p_{Dg}^* (already defined in Appendix 2.E) are given by:

$$p_{\text{Dz}}^* = \frac{1}{2} + \frac{2c(1-s) - s}{2\mathcal{A}_z} \quad \text{and} \quad p_{\text{Dg}}^* := \frac{1}{2} + \frac{2c(1-sh) - s}{2\mathcal{A}_g} \quad (2.108)$$

with,

$$\mathcal{A}_z = s [2(1-c)(1-h) - 1] \quad \text{and} \quad \mathcal{A}_g = s(1-2h). \quad (2.109)$$

Finally, the threshold values for r are given by:

$$n_z^* > 0 \iff r > \frac{1 - \mathcal{F}^z(p_{\text{Dz}}^*)}{\mathcal{F}^z(p_{\text{Dz}}^*)} \quad (2.110)$$

when conversion occurs in the zygote and,

$$n_g^* > 0 \iff r > \frac{1 - \mathcal{F}^g(p_{\text{Dg}}^*)}{\mathcal{F}^g(p_{\text{Dg}}^*)} \quad (2.111)$$

when conversion occurs in the germline.

Chapter 3

The speed of advance of a gene drive is affected by density dependence

Léna KLÄY^{*,1}, Léo GIRARDIN², Vincent CALVEZ², Florence DÉBARRE¹

¹ Institute of Ecology and Environmental Sciences Paris (IEES Paris), Sorbonne Université, CNRS, IRD, INRAE, Université Paris Est Creteil, Université de Paris, Paris Cedex 5, France.

² Institut Camille Jordan, UMR 5208 CNRS and Université Claude Bernard Lyon 1, France

Abstract

Homing gene drive alleles bias their own transmission by converting wild-type alleles into drive alleles. If introduced in a natural population, they might fix within a relatively small number of generations, even if they are deleterious. No engineered homing gene drive organisms have been released in the wild so far, and modelling is essential to develop a clear understanding of the potential outcomes of such releases. We use deterministic models to investigate how different demographic features affect the spatial spread of a gene drive. Building on previous work, we first consider the effect of the intrinsic population growth rate on drive spread, and we confirm that including demography can change outcomes compared to a model ignoring changes in population sizes, opposing the spatial spread of a drive. Second, we study the consequences of including an Allee effect, and find that the inclusion of an Allee effect makes a population more prone to extinction following drive spread. Finally, we investigate the effects of the fitness component on which density dependence operates (birth or death), and find that it affects the speed of drive invasion in space, and can accentuate the consequences of an Allee effect.

*Corresponding author: lena.klay@sorbonne-universite.fr

Contents

3.1	Introduction	96
3.2	Models and methods	98
3.2.1	Models	98
3.2.2	Traveling waves	100
3.3	Results	101
3.3.1	Demography and dominance can affect the final allelic proportions	101
3.3.2	The Allee effect makes the eradication easier and reduces the final density in case of drive persistence	103
3.3.3	A density-dependence constraint on the deaths instead of the births results in a faster invasion.	105
3.3.4	The Allee effect might cause the failure of threshold dependent drives	107
3.4	Discussion	108

3.1 Introduction

A promising but controversial new strategy for the control of natural populations, CRISPR-based gene drive biases the transmission of particular alleles to the offspring, over expectations of regular Mendelian transmission [4, 34, 35]. Such alleles can be detrimental to the individuals carrying them, and yet spread in a population thanks to their transmission advantage. Lab experiments have achieved transmission rates of 99% in yeast *Saccharomyces cerevisiae* [74], more than 90% in mosquito *Anopheles gambiae* [88], and more than 85% in fruit flies *Drosophila melanogaster* [224].

In “homing drives”, biased inheritance relies on gene conversion: in a heterozygous cell, the gene drive cassette present on one chromosome induces a double-strand break on the homologous chromosome, and repair by homologous recombination duplicates the cassette. The repetition of this process through generations favour the propagation of the drive allele in the population. Conversion can theoretically happen at different steps of the life-cycle, like in the germline of the parents, or in the zygote. However, practical implementations in the lab have focused on conversion in the germline [43].

Biased transmission via gene conversion can lead to the spread of new, potentially deleterious traits in a population within a relatively small number of generations. Two main types of drive can be distinguished: *replacement drives*, aiming to change features of the target population without directly affecting its size, and *suppression drives*, aiming to reduce population size (an extreme being *eradication drives*). Experimental proofs of principle of the latter type have been obtained with cage populations [133, 104] and the feasibility in large populations had been confirmed by theoretical studies [34, 97, 95].

Artificial gene drive like CRISPR-based homing drive holds promise for addressing a number of real-world problems [24, 109, 168], among which vector-borne diseases, i.e. infectious diseases spread by insects or other arthropods, such as Malaria. Artificial gene drive could be used to spread a new trait rendering mosquitoes progeny unable to transmit disease [91], or simply leading to the reduction of vector mosquitoes population size over time [103, 133]. Applications of artificial gene drive are however not limited to human health. Gene drive could help conserve or even partially restore native ecosystems by disadvantaging invasive species or favouring endemic ones [81, 187]. It could also be used

in agriculture to reverse insecticide resistance in pest animal species [123] or make weeds susceptible again to herbicides [164].

As of today, no artificial gene drive organisms have been released in the wild. Lab experiments, as well as mathematical and computational models, are crucial to evaluate the risks and benefits of gene drive, and assess the safety of potential releases. Models are however simplifications of the living world, and it is crucial to understand the impact and importance of various modelling choices, and to test the robustness of results to changes in modelling assumptions.

The simplest theoretical models of gene drive often represent well-mixed populations [70], and focus on allele frequencies changes over time [69, 213, 67, 187, 204]. Here, we investigate the spatial spread of a gene drive allele, and how demographic features affect it. Previous work has shown that propagation of a drive in a well-mixed population did not necessarily imply that the drive would spread. This is the case when the drive is threshold-dependent, i.e. when in a well-mixed population it needs to be introduced in a high enough amount to increase in proportion [134, 39]. While changes in population density may be ignored when a drive barely affects reproduction or survival, it becomes important to consider them in the case of a suppression drive, because its increase in proportion directly affects population size. Previous work on a specific model [95, 129] however found that demographic features barely affects the speed of advance of a drive wave over a continuous space. Here, we will assess the robustness of this result to different modelling assumptions.

A population's growth rate is determined by birth and death rates [190]. Density regulation may affect the two differently, which has consequences for overall demographic dynamics [208]. Likewise, which fitness component is affected by the drive (i.e., whether the drive reduces fecundity or decreases survival) can also influence outcome [186]. Finally, growth at low population density may be different from growth at high population densities, i.e. Allee effects may operate [146]. This can be caused by inbreeding depression, or difficulties to find a mate when the population density is low, for example [60]. Allee effects are frequently observed in the wild, including for animals considered as potential targets of control by artificial gene drive, like mosquito species affected by inbreeding depression [10, 14, 189]. The existence of Allee effects may also influence the outcome of the release of a drive affecting population size.

In this article, we consider a one-dimensional continuous environment, and study the spatial spread (or not) of a drive allele invading an established wild-type population. We follow the densities of the different genotypes (drive homozygous, wild-type homozygous and heterozygous) over space and time using partial differential equations. We compare four demographic models, depending on the presence or absence of an Allee effect, and the fitness component (birth or death) on which density dependence operates. We find that the Allee effect might help eradicate or reduce in density the targeted population, however it might also lead the failure of threshold-dependent drive invasions. We also find that the effect of demography on drive spread is limited in the case of density regulation on the birth rate, but are not when density regulation affects the death rate, where wave speed increases with intrinsic growth rate. This difference emerges because drive invasion over space primarily relies on the birth of new individuals, and highlights the importance of ecological details on the outcome of the release of a drive.

3.2 Models and methods

3.2.1 Models

In this section, we will build step-by-step the different models that we will compare. These models differ in their demographic components, which we first introduce.

3.2.1.1 Demographic terms

To assess how sensitive results might be to different demographic modelling choices, we will consider four models differing in their birth and death terms. We first illustrate these four demographic models in the case of a genetically and spatially homogeneous population, composed only of wild-type individuals. We will compare density dependence acting on the birth term (Models \mathcal{BN} and \mathcal{BA}) or death term (Models \mathcal{DN} and \mathcal{DA}), and the absence (Models \mathcal{BN} and \mathcal{DN}) and the presence of an Allee effect (Models \mathcal{BA} and \mathcal{DA}).

We denote by r the population's intrinsic growth rate, and by a the parameter controlling the Allee effect threshold (when there is an Allee effect; $-1 \leq a \leq 1$). In these models, the population initial growth rate (i.e. when $n \rightarrow 0$) is r in the absence of Allee effect, and $-ar$ in the presence of Allee effect. When $-1 < a < 0$, the Allee effect is said to be weak (the initial growth rate remains positive), while when $0 < a < 1$, the Allee effect is said to be strong (the initial growth rate is negative; the population only grows if already at high enough density; see appendix 3.A for details).

Population density is scaled so that the carrying capacity in all models is 1, and time is scaled so the death rate in the absence of density regulation is 1. Denoting by $n(t)$ population density at time t , the four models read:

Model \mathcal{BN}

$$\partial_t n(t) = \overbrace{(r(1-n(t)) + 1)n(t)}^{\text{births}} - \overbrace{n(t)}^{\text{deaths}} \quad (\forall t > 0), \quad (3.1a)$$

Model \mathcal{BA}

$$\partial_t n(t) = \overbrace{(r(1-n(t))(n(t)-a) + 1)n(t)}^{\text{births}} - \overbrace{n(t)}^{\text{deaths}} \quad (\forall t > 0), \quad (3.1b)$$

Model \mathcal{DN}

$$\partial_t n(t) = \overbrace{(r+1)n(t)}^{\text{births}} - \overbrace{(1+rn(t))n(t)}^{\text{deaths}} \quad (\forall t > 0), \quad (3.1c)$$

Model \mathcal{DA}

$$\partial_t n(t) = \overbrace{(r+1)n(t)}^{\text{births}} - \overbrace{(r(n(t)-1)(n(t)-a) + r + 1)n(t)}^{\text{deaths}} \quad (\forall t > 0). \quad (3.1d)$$

In the following, we denote the birth rate by $B(n(t))$ and the death rate by $D(n(t))$ such that the four equations can all be written as:

$$\partial_t n(t) = \overbrace{B(n(t))n(t)}^{\text{births}} - \overbrace{D(n(t))n(t)}^{\text{deaths}} \quad (\forall t > 0). \quad (3.2)$$

3.2.1.2 Drive and wild-type

The demographic models being defined, we now add genetic diversity to the models, following the same approach as in [95, 129]. The variable $n(t)$ becomes the total density, and we denote by n_i the density of individuals with genotype i . There are two possible alleles at the locus that we consider: the wild-type allele (W) and the drive allele (D), so that there are three different genotypes: wild-type homozygotes ($i = WW$), drive homozygotes ($i = DD$) and heterozygotes ($i = DW$). The fitness effect of a genotype is represented by a coefficient f_i acting on the birth term. It represents the selective disadvantage conferred by the drive to the individual carrying it. Wild-type homozygotes have fitness $f_{WW} = 1$, drive homozygotes have fitness $f_{DD} = 1 - s$, where s is the fitness cost of the drive, and drive heterozygotes have fitness $f_{DW} = 1 - sh$, where h is the dominance parameter. We assume that mating occurs at random: the probability that a genotype l mates with a genotype k is equal to $\frac{n_l n_k}{n^2}$. Finally, we denote by $\pi_{l,k}^i$ the probability for a couple of parents with genotypes l and k to have offspring of type i . This probability depends on the moment at which gene conversion takes place and on the probability that gene conversion takes place and is successful c ($0 \leq c \leq 1$). Here we assume that gene conversion takes place in the germline, because this is the timing currently successfully implemented in the lab [43, 47], unlike gene conversion in the zygote. With these assumptions, the dynamics are now given by the following equations:

$$\partial_t n_i(t) = B(n(t)) f_i n(t) \underbrace{\sum_{l,k} \pi_{l,k}^i \frac{n_l(t)}{n(t)} \frac{n_k(t)}{n(t)}}_{\text{Mating term}} - D(n(t)) n_i(t) \quad (\forall t > 0) (\forall i). \quad (3.3)$$

The formulas for $\pi_{l,k}^i$ are included in the full equations in Appendix 3.B.1.

3.2.1.3 Space

Our equations so far did not include space, we now add this component. We assume that the movement of individuals is described by a diffusion term with equal diffusion coefficients. Space is scale such that these coefficients are normalised to 1. We obtain the following equations:

$$\partial_t n_i(t, x) = B(n(t)) f_i n(t) \sum_{l,k} \pi_{l,k}^i \frac{n_l(t)}{n(t)} \frac{n_k(t)}{n(t)} - D(n(t)) n_i(t, x) + \partial_{xx}^2 n_i(t, x) \quad (\forall t > 0) (\forall x \in \mathbb{R}) (\forall i). \quad (3.4)$$

Finally, all parameters of the models are summarised in Table 3.1.

Parameters	Range values	Description
r	$(0, +\infty)$	Intrinsic growth rate
c	$[0, 1]$	Conversion rate
s	$(0, 1)$	Fitness cost of drive homozygotes
h	$[0, 1]$	Drive dominance
a	$[-1, 1]$	Allee effect threshold

Table 3.1: Model parameters.

We have presented equations with genotype densities n_i (condensed model in equation (3.4); full equations for each model are given in appendix 3.B). The model can be rewritten to follow allele densities instead (see appendix 3.B.2), or total population size and allele frequencies (see appendix 3.B.3); different steps of the analysis may require different formulations of the model.

3.2.2 Traveling waves

The introduction of drive individuals in a wild-type population will give rise to a wave of change in genotype densities through space: a traveling wave (except in the *gene drive clearance* case, see below). Traveling waves have the particularity to propagate with a constant speed, while maintaining their shape in space. We consider an initial condition in which the left half of the domain is full of drive ($n_{DD} = 1$), and the right half is full of wild-type ($n_{WW} = 1$), illustrated in Figure 3.1. In this article, we are not exploring the effect of inoculum size and distribution, which is a question in itself, and arises in the case of threshold-dependent drives. We therefore chose an initial condition maximising the possibility of drive spread. The model is then solved numerically. We classify the outcomes into five categories, present in the four models, depending on: the existence or not of a traveling wave; whether the population persists or is eradicated; and in the former case, the genotype(s) present at the end. These outcomes are illustrated in Figure 3.2.

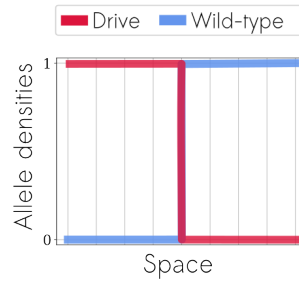


Figure 3.1: Initial conditions used in the simulations. The left half of the domain is full of drive ($n_{DD} = n_D = 1$), and the right half is full of wild-type ($n_{WW} = n_W = 1$).

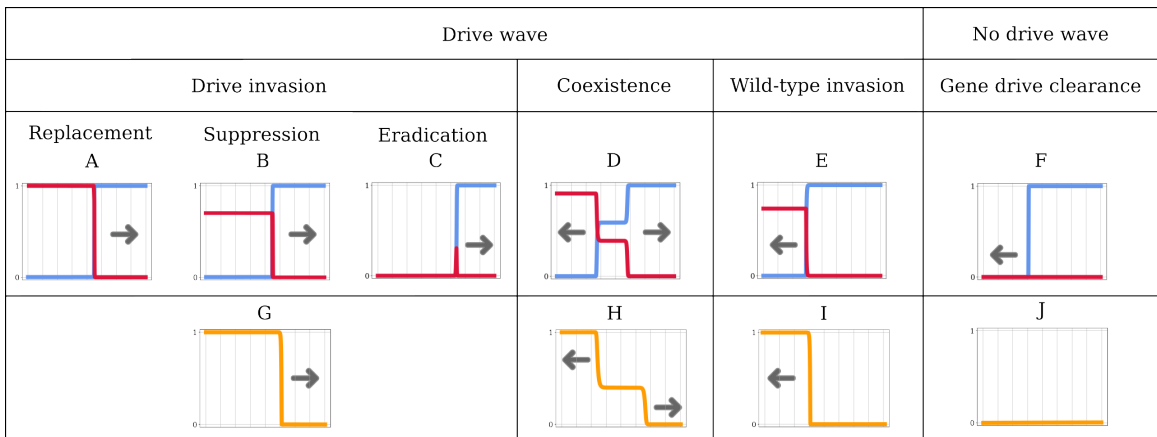


Figure 3.2: Types of spatial dynamics. Panels A–F correspond to allele densities, with the drive allele in red and the wild-type allele in blue. Arrows represent the direction of advance of the wave. Panels G–J show the equivalent with the drive allele frequency. The x-axis represents space.

It can happen that the model lead to the decay of the drive allele uniformly in space. This case arises in particular when a well-mixed population composed only of drive individuals is not sustainable. The introduced drive subpopulation just dies out, freeing space; in this case, there is no drive traveling wave. We describe this as *gene drive clearance*. The wild-type population then recolonises the emptied space, at a speed described in the standard Fisher-KPP traveling wave problem (see [87, 130, 12]).

When the drive traveling wave does exist, we distinguish between two cases, depending on the sign of the speed v . When $v > 0$, the wave moves to the right: it is a *drive invasion*. When $v < 0$, the wave moves to the left: it is a *wild-type invasion*. In some specific cases, drive and wild-type invasions can happen simultaneously: the waves decompose into two sub-traveling wave solutions over half of the domain. They move in opposite directions and lead to the coexistence of both alleles in-between.

In case of drive invasion, we distinguish three cases depending on the state of the population in the wake of the front(s): i) in the case of *replacement* drives, the population persists in the wake of the front(s) at the same density as the original wild-type population; ii) in the case of *suppression* drives, the population persists in the wake of the front(s), albeit at a lower density than the original wild-type population; iii) in the case of *eradication* drives, the population is eradicated in the wake of the drive invasion front(s), just leaving empty space.

The code for these simulations is available on GitHub (https://github.com/LenaKlay/gd_project_1, in the folder: `deterministic`). We ran our simulations in Python 3.6, with the Spyder environment. Heatmaps in Figures 3.4 have been computed thanks to the INRAE Migale bioinformatics facility, doi: 10.15454/1.5572390655343293E12.

3.3 Results

3.3.1 Demography and dominance can affect the final allelic proportions

Here, we focus on the importance of demography in the model, i.e. on the role played by the intrinsic growth rate r over the final allelic proportions. Analytical results can be obtained for $r = 0$ and for $r \rightarrow \infty$; intermediate cases are investigated numerically.

When $r = 0$, deaths and births compensate each other in a fully wild-type population. In this limit case, Models \mathcal{BN} , \mathcal{BA} , \mathcal{DN} and \mathcal{DA} are the same (given in equation (3.19)). Both the final densities of all genotypes and the speed of the drive wave are therefore the same, which we will characterise below, recalling results from our previous work [129].

Leaving aside the density-dependence constraint, the bigger r is, the faster the wild-type population grows. When $r \rightarrow \infty$, final allelic proportions are the same in models \mathcal{BN} , \mathcal{BA} , \mathcal{DN} and \mathcal{DA} (see Section 3.B.4). This is however not necessarily the case for the total population density and for the wave speed.

Following previous work [129], let us introduce:

$$s_1 := \frac{c}{1 - h(1 - c)}, \quad s_{2,g} := \frac{c}{2ch + h(1 - c)} = \frac{c}{h(1 + c)}. \quad (3.5)$$

These are threshold values of the fitness cost s determining qualitatively different outcomes. When the drive allele is recessive ($h < 1/2$), $s_1 < s_{2,g}$, and when the drive allele is dominant ($h > 1/2$), $s_1 > s_{2,g}$.

When the fitness cost s is low enough ($s < \min(s_1, s_{2,g})$), there is a wave of advance of the drive for both $r = 0$ and $r \rightarrow \infty$ (drive invasion, as in Fig. 3.2G).

When the fitness cost s is high enough ($s > \max(s_1, s_{2,g})$), and the intrinsic growth rate is high ($r \rightarrow \infty$), the drive wave retreats (wild-type invasion, as in Fig. 3.2I). When the intrinsic growth rate is low ($r = 0$), $s > \min(s_1, s_{2,g})$ results in drive clearance (as in Fig. 3.2J): the drive is just too costly even for a full-drive population.

What happens for intermediate fitness cost ($\min(s_1, s_{2,g}) < s < \max(s_1, s_{2,g})$) and high growth rate depends on the dominance parameter h . If $h < 1/2$, drive and wild-type alleles coexist eventually (coexistence, as in Fig. 3.2H). If $h > 1/2$, there is a bistability, the drive is threshold-dependent: the final outcome is either drive invasion or wild-type invasion, and depends on the initial conditions.

These results are summarised in Table 3.2 and illustrate the importance of taking demography into account. Threshold-dependent drives (i.e. drives leading to bistabilities), are considered more socially responsible than threshold independent drives, as they are potentially localised and reversible [204, 134]. The intrinsic growth rate r is a key component to reach this bistable condition (threshold dependence), as r has to be sufficiently large for the bistability to happen. Indeed, a small r would result in the systematic decay of gene drive alleles (Table 3.2) and no possibility of drive invasion at all.

As in models without demography nor spatial structure [69, 213], the dominance parameter h conditions whether threshold-dependence can be attained or not: a bistable outcome only exists when $h > \frac{1}{2}$, i.e. when the fitness of heterozygous individuals is closest to the fitness of drive homozygous individuals than wild-type homozygous individuals. This result was already given in a simpler panmictic model [69]: indeed, the birth and death terms in our models \mathcal{BN} and \mathcal{BA} tend to this panmictic model for large values of r .

(a) When $h < 1/2$			
	$0 < s < s_1$	$s_1 < s < s_{2,g}$	$s_{2,g} < s < 1$
$r \rightarrow \infty$	Drive invasion	Coexistence	Wild-type invasion
$r = 0$	Drive invasion**		Gene drive clearance

(b) When $h > 1/2$			
	$0 < s < s_{2,g}$	$s_{2,g} < s < s_1$	$s_1 < s < 1$
$r \rightarrow \infty$	Drive invasion	Bistability	Wild-type invasion
$r = 0$	Drive invasion**	Gene drive clearance	

Table 3.2: Types of model outcomes for Models \mathcal{BN} , \mathcal{BA} , \mathcal{DN} and \mathcal{DA} , depending on the fitness cost s , intrinsic growth rate r and dominance parameter h . The outcomes are in terms of allele proportions, as in Fig. 3.2G–J

**When $r = 0$ and $s < s_{2,g}$, we do not know for sure the final drive frequency as the invasion leads to the eradication of the population. We call it a drive eradication although it might also be coexistence with no individual left at the equilibrium state.

3.3.2 The Allee effect makes the eradication easier and reduces the final density in case of drive persistence

In the previous section, we have only described outcomes in terms of allele frequencies. In this section, we compare the final population density n^* in the four models, and in particular conditions for which the population goes extinct ($n^* = 0$). We detail the final densities in all three types of invasions: drive invasion, wild-type invasion and coexistence. In case of gene drive clearance (decay of the drive allele uniformly in space), the final density is equivalent to the one obtained after a wild-type invasion: population size goes back to carrying capacity 1).

In all three types of invasion, there are up to three possible regimes: population eradication ($n^* = 0$); population persistence ($n^* = n^+ > 0$); and bistability (the final total population size is either 0 or n^+ depending on the initial condition relative to a specific density n^τ). Note that "bistability" here is different from bistability on allele frequencies as seen in the previous section; this bistability is about population densities.

We can write the final population densities in a generic manner for the three types of invasion. We define the mean fitness \mathcal{F} :

$$\mathcal{F}(p_D) = (1 - s) (p_D)^2 + 2 (1 - sh) p_D (1 - p_D) + (1 - p_D)^2, \quad (3.6)$$

and p_D^* the final proportion of the drive allele in the population. This last one verifies:

$$((2h - 1) s p_D^* + (1 - sh)(1 + c) - 1) p_D^* (1 - p_D^*) = 0, \quad (3.7)$$

according to the allelic frequency systems detailed in Section 3.B.3. For wild-type invasion, $p_D^* = 0$ and $\mathcal{F}(p_D^*) = 1$; for drive invasion, $p_D^* = 1$ and $\mathcal{F}(p_D^*) = 1 - s$, and for coexistence,

$$p_D^* = \frac{1 - (1 - sh)(1 + c)}{s(2h - 1)} \in (0, 1). \quad (3.8)$$

The final densities n^* are then computed by solving the allelic frequency systems detailed in Section 3.B.3 with the relevant value of p_D^* . These results, holding for all values of the intrinsic growth rate r , are summarised in Table 3.3 and illustrated in Figure 3.3 with $c = 0.85$ and $h = 0.9$.

In the models without Allee effects, there is no bistability for the final population size. The condition for eradication is the same in models \mathcal{BN} and \mathcal{DN} , i.e. does not depend on whether density dependence acts on births or deaths. In the case of population persistence, final population size is lower in Model \mathcal{DN} . This is because the fitness cost in our models acts on births. Density dependence on births (as in Model \mathcal{BN}) affects a lower number of individuals than density dependence on deaths (as in Model \mathcal{DN}).

In the models with Allee effects, the outcome depends on the value of a . When $a = -1$, the final densities in models \mathcal{BA} and \mathcal{DA} are the same as in models \mathcal{BN} and \mathcal{DN} , respectively. In case of a weak Allee effect ($-1 < a < 0$), the three regimes are possible: eradication, persistence, bistability (i.e. eradication or persistence, depending on the initial conditions). Finally, in the case of a strong Allee effect ($0 < a < 1$), the "persistence" regime disappears.

Model	Regime	n^+ and n^τ (if it exists)
\mathcal{BN}	Eradication if $r < \frac{1-\mathcal{F}(p_D^*)}{\mathcal{F}(p_D^*)}$ Persistence if $r > \frac{1-\mathcal{F}(p_D^*)}{\mathcal{F}(p_D^*)}$	$n^{+(\mathcal{BN})} = 1 - \frac{1-\mathcal{F}(p_D^*)}{r\mathcal{F}(p_D^*)}$
\mathcal{BA}	Eradication if $r < \frac{1-\mathcal{F}(p_D^*)}{\left(\frac{1-a}{2}\right)^2 \mathcal{F}(p_D^*)}$ Bistability if $r > \frac{1-\mathcal{F}(p_D^*)}{\left(\frac{1-a}{2}\right)^2 \mathcal{F}(p_D^*)}$ and $ra > \frac{\mathcal{F}(p_D^*)-1}{\mathcal{F}(p_D^*)}$ Persistence if $r > \frac{1-\mathcal{F}(p_D^*)}{\left(\frac{1-a}{2}\right)^2 \mathcal{F}(p_D^*)}$ and $ra < \frac{\mathcal{F}(p_D^*)-1}{\mathcal{F}(p_D^*)}$	$n^{\tau(\mathcal{BA})} = \frac{1+a-\sqrt{(1+a)^2-4\left(a+\frac{1-\mathcal{F}(p_D^*)}{r\mathcal{F}(p_D^*)}\right)}}{2}$ $n^{+(\mathcal{BA})} = \frac{1+a+\sqrt{(1+a)^2-4\left(a+\frac{1-\mathcal{F}(p_D^*)}{r\mathcal{F}(p_D^*)}\right)}}{2}$
\mathcal{DN}	Eradication if $r < \frac{1-\mathcal{F}(p_D^*)}{\mathcal{F}(p_D^*)}$ Persistence if $r > \frac{1-\mathcal{F}(p_D^*)}{\mathcal{F}(p_D^*)}$	$n^{+(\mathcal{DN})} = 1 - \frac{(1-\mathcal{F}(p_D^*))(r+1)}{r}$
\mathcal{DA}	Eradication if $r\left[\left(\frac{1-a}{2}\right)^2 - (1-\mathcal{F}(p_D^*))\right] < (1-\mathcal{F}(p_D^*))$ Bistability if $r\left[\left(\frac{1-a}{2}\right)^2 - (1-\mathcal{F}(p_D^*))\right] > (1-\mathcal{F}(p_D^*))$ and $r[a+1-\mathcal{F}(p_D^*)] > (\mathcal{F}(p_D^*)-1)$ Persistence if $r\left[\left(\frac{1-a}{2}\right)^2 - (1-\mathcal{F}(p_D^*))\right] > (1-\mathcal{F}(p_D^*))$ and $r[a+1-\mathcal{F}(p_D^*)] < (\mathcal{F}(p_D^*)-1)$	$n^{\tau(\mathcal{DA})} = \frac{1+a-\sqrt{(1+a)^2-4\left(a+\frac{(r+1)(1-\mathcal{F}(p_D^*))}{r}\right)}}{2}$ $n^{+(\mathcal{DA})} = \frac{1+a+\sqrt{(1+a)^2-4\left(a+\frac{(r+1)(1-\mathcal{F}(p_D^*))}{r}\right)}}{2}$

Table 3.3: Final densities for Models \mathcal{BN} , \mathcal{BA} , \mathcal{DN} and \mathcal{DA} , with p_D^* the final proportion of the drive allele in the population, $\mathcal{F}(p_D^*)$ the mean fitness, n^+ the final non-zero density and n^τ the threshold density in case of bistability.

Adding an Allee effect in the model always results in a bigger chance of extinction, but also, in our models, in a smaller final density in case of persistence when the drive allele is still present (Appendix 3.C.2 and 3.C.3). When the Allee effect gets stronger, these effects are accentuated (Appendix 3.C.4). The density dependence constraint placed on the death term also accentuate these effects when coupled with an Allee effect (model \mathcal{DA}) compared to the birth term (model \mathcal{BA}) (see Appendix 3.C.4). Interestingly, when $r \rightarrow \infty$, eradication is still possible in model \mathcal{DA} (for $\mathcal{F}(p_D^*) < 1 - \left(\frac{1-a}{2}\right)^2$) while it is not in model \mathcal{BA} .

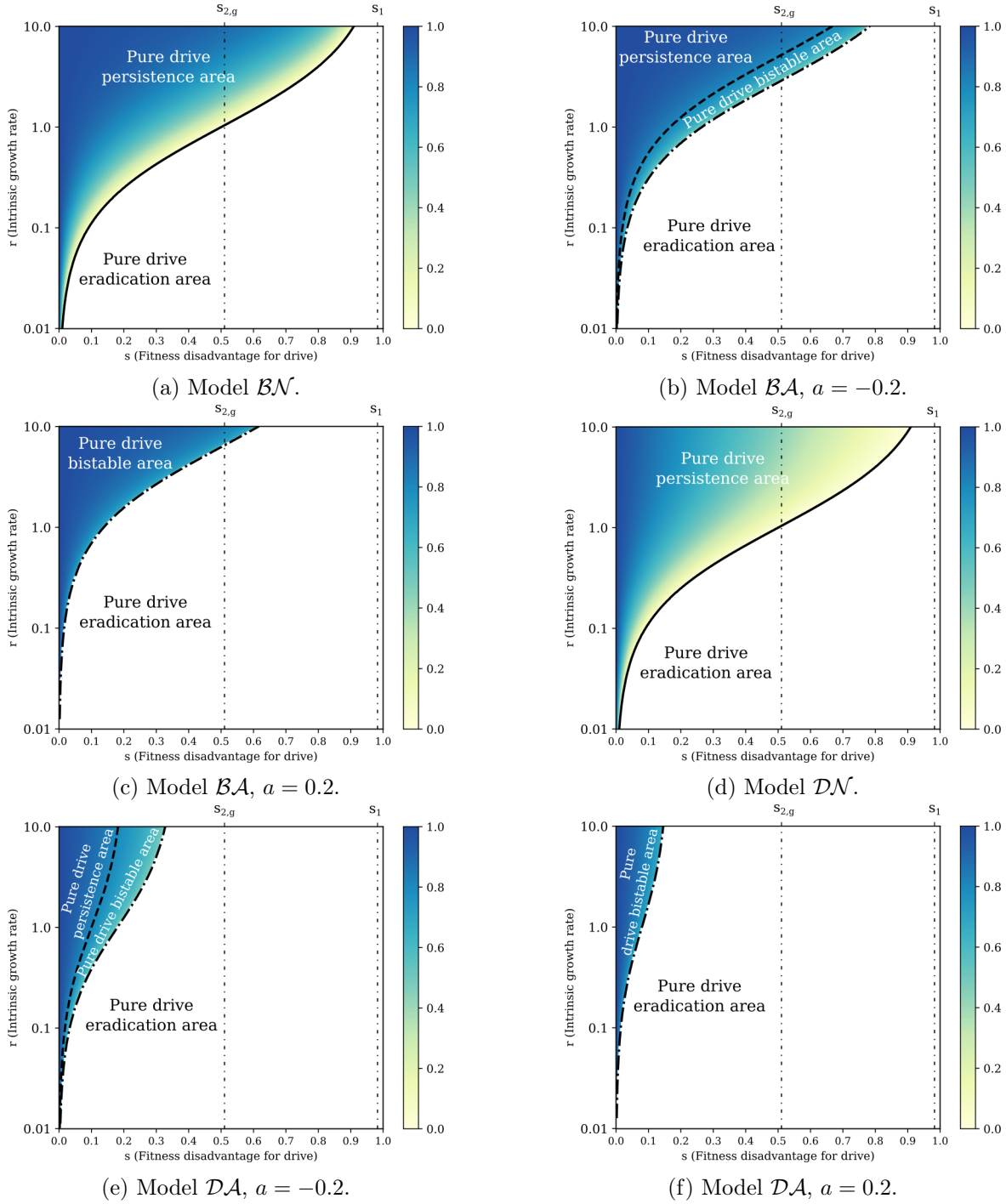


Figure 3.3: Density of the final population, with parameters $c = 0.85, h = 0.9$.

3.3.3 A density-dependence constraint on the deaths instead of the births results in a faster invasion.

We now focus on the speed of a drive invasion, i.e. the speed of the traveling wave emerging from a drive invasion (see Section 3.2.2).

A speed v for the drive wave can be calculated when the models are simplified (linearised) assuming low drive density. It corresponds to the speed of drive invasion when the movement of individuals is caused by the few drive individuals at the expansion edge, where the drive density is low (pulled wave). This happens if such small populations have high growth rates, as the movement is then mainly driven by the reproduction. The calculated speed corresponds to a minimal speed when movement is brought about by individuals in the bulk of the wave (pushed wave), i.e. the real speed is higher, but cannot be calculated. In a previous article [129], we showed that the calculated speed v corresponds to the speed of a drive invasion when the dominance parameter $h < \frac{1}{2}$, and for a drive fitness cost s small enough when $h > \frac{1}{2}$ (for a precise condition, see [129]). This result was rigorously proven for large and small values of the intrinsic growth rate r , and numerically observed for all r . We calculate and compare this speed value in our four models (details in Appendix 3.B.5).

In models \mathcal{BN} and \mathcal{BA} with density dependence acting on the birth term, this speed is given by:

$$v_{\mathcal{BN},\mathcal{BA}} = 2\sqrt{(1 - sh)(1 + c) - 1}. \quad (3.9)$$

In models \mathcal{DN} and \mathcal{DA} with density dependence acting on the death term, it becomes :

$$v_{\mathcal{DN},\mathcal{DA}} = 2\sqrt{(1 + r)[(1 - sh)(1 + c) - 1]}. \quad (3.10)$$

The speeds $v_{\mathcal{BN},\mathcal{BA}}$ and $v_{\mathcal{DN},\mathcal{DA}}$ are very similar but differ by one coefficient: $v_{\mathcal{DN},\mathcal{DA}}$ is $\sqrt{r + 1}$ times bigger than $v_{\mathcal{BN},\mathcal{BA}}$. This difference relies on the density-dependence constraint, affecting either the births or the deaths. At the front of the wave, the population density, composed nearly only of wild-type individuals, reaches the maximum carrying capacity. Consequently, the density-dependence constraint prohibits any increase in the population density and this happens in two different ways: in models \mathcal{BN} and \mathcal{BA} , it limits the births so that they do not exceed the deaths, whereas in models \mathcal{DN} and \mathcal{DA} , it increases the death rate to compensate the births. As a result, the turnover rate is greater in models \mathcal{DN} and \mathcal{DA} , which induces a faster invasion as the wave movement is mainly driven by the reproduction. Details of the speed calculations are given in Appendix 3.B.5. To illustrate this result, we plot the speed of the wave for the four models in Figure 3.4 and observe that the speed of the drive invasion always increases with r in models \mathcal{DN} and \mathcal{DA} , in contrast with models \mathcal{BN} and \mathcal{BA} .

Note that speeds $v_{\mathcal{BN},\mathcal{BA}}$ and $v_{\mathcal{DN},\mathcal{DA}}$ only exist for $(1 - sh)(1 + c) > 1$ (or equivalently $s < s_{2,g}$, with $s_{2,g}$ given in equation (3.5)), which is the necessary condition to have a strictly positive drive alleles production at the front of the wave. To understand why, first note that the density of drive alleles is very low at the front of the wave. Therefore, we can make the approximation that at least one parent in each couple formed at the front of the wave has a genotype WW . Consequently, the offspring carrying a drive allele are necessarily heterozygotes: in the front of the wave, the production of drive alleles only relies on the heterozygotes. These heterozygotes have a fitness of $(1 - sh)$ and produce drive alleles at rate $(1 + c)$: therefore, for a drive invasion to be possible, the production rate $(1 - sh)(1 + c)$ of drive alleles should be above the rate 1 at which they disappear. The higher the production rate is, the faster the wave moves.

Also note that while the speed $v_{\mathcal{BN},\mathcal{BA}}$ is independent of r for a drive fitness cost s small enough, it is not the case for the final density n_D^* (Table 3.3 and Figure 3.3, models \mathcal{BN} and \mathcal{BA}). As a result, for a small enough s in models \mathcal{BN} and \mathcal{BA} , the wave travels at a constant speed no matter the density of population left behind.

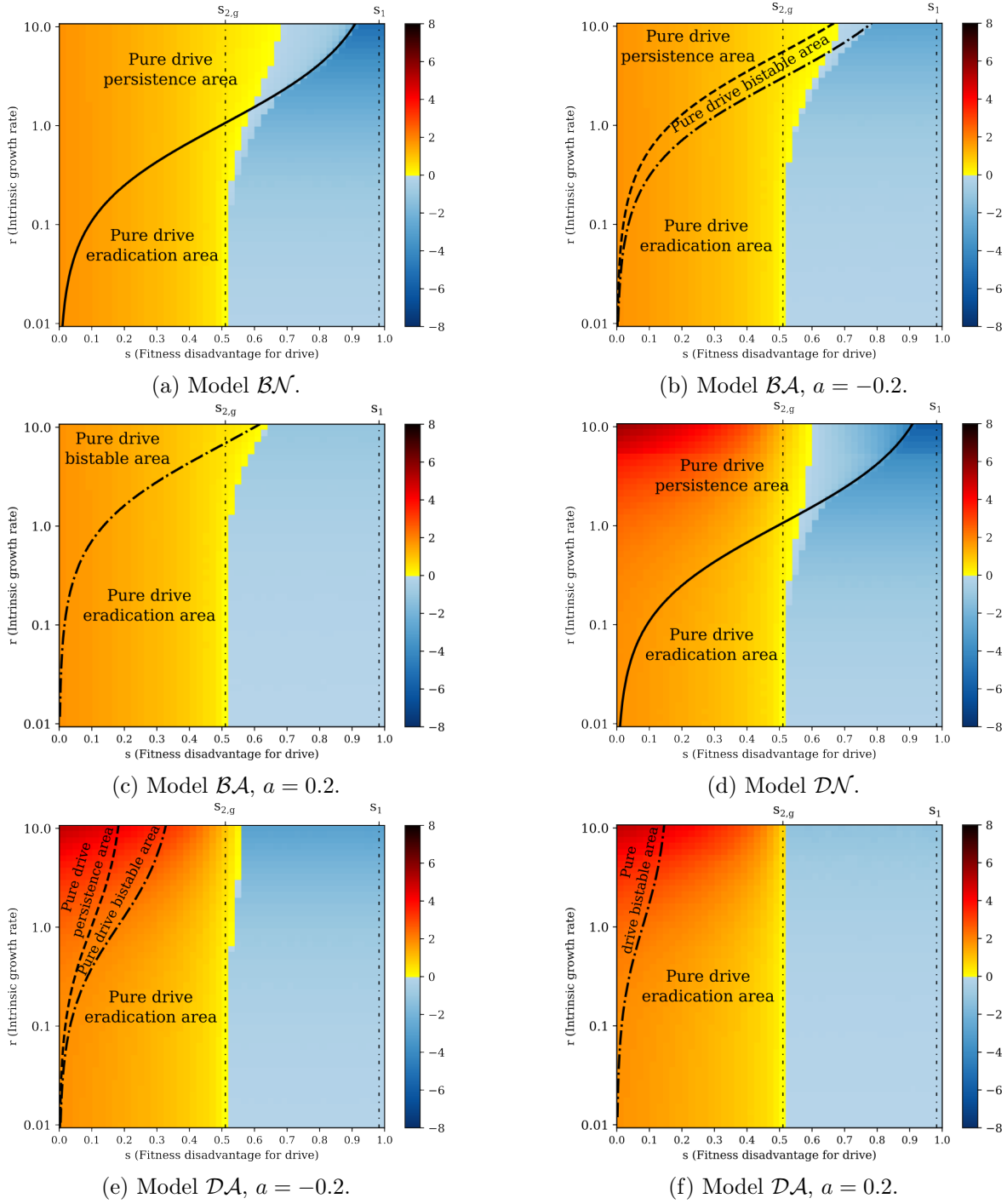


Figure 3.4: $c = 0.85, h = 0.9$

3.3.4 The Allee effect might cause the failure of threshold dependent drives

Finally in Figure 3.4, we observe that for $s > s_{2,g}$, the Allee effect might prevent drive invasion. Yellow-orange areas in heatmaps (a) (resp. (d)) are becoming blue in heatmaps (b) and (c) (resp. (e) and (f)) due to the Allee effect in the model. Noticeably, the drive invasions for $s > s_{2,g}$ are threshold-

dependent drive invasions, often considered as more socially responsible than threshold independent drives invasions [204, 134]. Once again this influence is accentuated when the Allee effect gets stronger (for larger values of a).

3.4 Discussion

Understanding the conditions for the spatial spread of an artificial gene drive and its consequences on a targeted population is essential before considering any field release. Laboratory experiments provide information on gene drive dynamics in a small confined and controlled environment, and mathematical models can help gain further insights at small and larger scales.

Theoretical models are meant to provide insights on real-world dynamics, so it is important to assess how a model’s result depends on modelling choices. In this article, we investigate the influence of considering i) demography, and more precisely different values of the target population’s intrinsic growth rate, ii) the presence/absence of an Allee effect and iii) which fitness component (birth or death) is affected by density dependence. We considered the effects of these features on the type of outcome, on final population density, and on the speed of the drive wave.

We first described the different qualitative outcomes, extending results from our previous studies [95, 129] on the importance of taking into account demography in the models. We confirm that the intrinsic growth rate r qualitatively affects results at intermediate values of the fitness cost s . A high intrinsic growth rate leads to a threshold-dependant drive invasion, while a low intrinsic growth rate results in the decay of drive alleles uniformly in space. Models not considering population densities but focusing on frequencies [69, 213, 67, 187, 204, e.g.] have dynamics similar to when $r \rightarrow \infty$ in our models.

We also showed that an Allee effect makes the population more susceptible to eradication, widening the range of s (fitness disadvantage for drive) and r (intrinsic growth rate) leading to population extinction after a drive invasion. It also reduces the final population density in case of persistence, meaning that an Allee effect might represent a non-negligible helping force to eradicate or suppress natural populations. However, we also showed that it might reduce the range of s and r values leading to a threshold-dependent drive invasion, often considered as more socially responsible than threshold independent drives invasions [204, 134]. Both are accentuated when the Allee effect gets stronger (for larger values of a).

Finally, we consider the impact of density-dependence constraint whether it targets the birth or the death: close to the maximal carrying capacity, in case of rarefaction of the resources, the net growth of the population is limited by either a low number of offspring per generation or a high death rate. This choice is usually made to simplify the model construction. In this study, we show that when targeting the death, this constraint supports the Allee effect by enlarging the eradication conditions and reducing the final density. It also strongly impacts the speed of propagation: a drive invasion would be $\sqrt{r+1}$ times faster for a density-dependence constraint over the death instead of the birth rate. This prediction holds for a fitness cost reducing the birth rate (individuals carrying drive alleles have fewer offspring than wilt-type ones). However, the conclusions might change for a fitness cost increasing the death rate instead, as shown in a different model of CRISPR-based homing drives [186].

Our models are deterministic. They can describe population dynamics at large scales, but cannot take into account stochastic effects such as large fluctuations or “chasing” events, which can arise at low population densities [169, 76, 46, 170, 171]. Stochastic fluctuations are likely to be important in

particular in the case of suppression and eradication drives, and are left for future investigation.

Among the deterministic models in the literature, the models we develop are generalist: they could be applied to different species, and any gene drive construct reducing the fitness of the individual carrying it. These models do not aim to bring precise and quantitative predictions, for which more specific models need to be developed, but rather get some insights into the possible outcomes, and dissect the roles played by different model elements. However, this generalist approach naturally come with simplifications.

In our models, we assume that gene conversion either successfully takes places, or does not take place. We did not include resistance alleles which can emerge when conversion fails and repairs by non-homologous end-joining occur, or resistance due to standing genetic variation at the target locus. The emergence of resistance alleles can alter the drive's propagation [18, 106, 181, 187].

Some other simplifications are directly related to the biological characteristics of the species. In mouse populations, the gene drive spread can be limited by their polyandrous mating system [149, 148] or mate search capabilities [25]. In mosquito populations, the plural life stages (egg, larva, pupa and adults) might influence the modelling conclusions and need to be taken into account by including corresponding age structure in models [152, 192, 48]. In bee populations, the haploid phases of the life cycle result in less powerful drives: the conditions for fixation are narrower and the spread is slower [143, 138]. Finally, it is not rare that males and females have different fitnesses in transgenic mosquitoes [17, 171, 133, 103]: more specific models than ours would need to include sex differences.

Finally and more broadly, species do not live in isolation, and interactions of the targeted species within its ecosystem would need to be considered. Competing species or predators can facilitate drive-based suppression [143], and environmental conditions such as seasonality (dry or wet season) can highly impact the eradication of mosquito populations, for example [76, 170, 171]. It is of public utility to also consider the impact of gene drive on the whole ecosystem and anticipate the potential risks: the probability of transmit the gene drive cassette to another species [56], or the cascade of population dynamics and evolutionary processes potentially initiated by the eradication of a species [162].

Overall, we have shown the importance of considering precise population dynamics, on the outcome of the release of a drive. This approach though theoretical models give first interesting insights that now need to be enhanced with ecological knowledge on specific systems.

Appendices

3.A Allee effect

We consider the equation describing the dynamics of the population density n :

$$\partial_t n - \partial_{xx}^2 n = n(1-n)\sigma(n) = f(n) \quad \text{with } n \in [0, 1]. \quad (3.11)$$

The Allee effect characterises a correlation between population density and the per capita population growth rate. Without Allee effect, the per capita population growth rate is always positive and is maximum as the population density tends to zero. This happens for example when $\sigma(n) = 1$ in Equation (3.11) (Figure 3.A.1a). Mathematically, we write:

$$\max_{n \in]0,1]} \left(\frac{f(n)}{n} \right) \leq f'(0). \quad (3.12)$$

With a weak Allee effect, the per capita population growth rate is still positive, but the maximum is reached at a strictly positive population density. This happens for example when $\sigma(n) = (n - a)$ with $-1 < a < 0$ in Equation (3.11) (Figure 3.A.1b). Mathematically:

$$\max_{n \in]0,1]} \left(\frac{f(n)}{n} \right) > f'(0) > 0. \quad (3.13)$$

Finally, with a strong Allee effect, the per capita population growth rate is negative for small population density, and positive after. This happens for example when $\sigma(n) = (n - a)$ with $0 < a < 1$ in Equation (3.11) (Figure 3.A.1c). Mathematically:

$$\exists a > 0 \text{ such that } \forall n \in]0, a[\quad \frac{f(n)}{n} < 0. \quad (3.14)$$

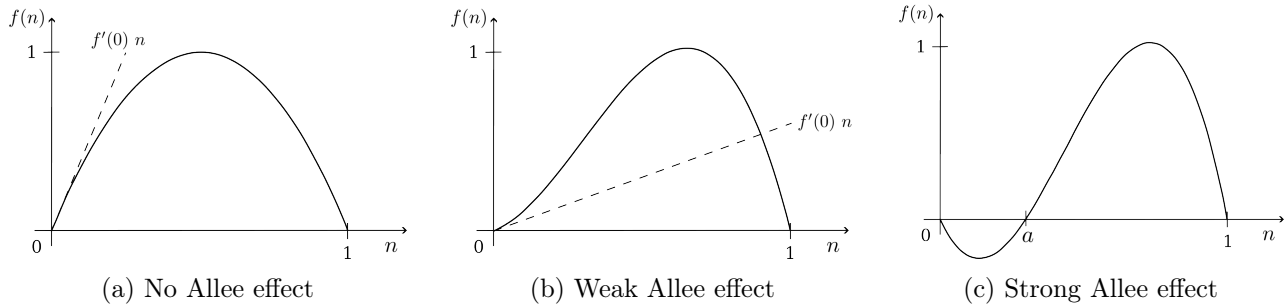


Figure 3.A.1: Illustration of the three cases concerning the Allee effect.

3.B Models

3.B.1 Genotype densities

For the sake of clarity, we omit variables in the notation ($n_i = n_i(t, x)$) in the following. Each model contains three equations for the three genotype densities: homozygote drive n_{DD} , heterozygote n_{DW} and homozygote wild-type n_{WW} .

Model \mathcal{BN}

$$\left\{ \begin{array}{l} \partial_t n_{DD} - \partial_{xx}^2 n_{DD} = (1-s)(r(1-n)+1) \frac{\frac{1}{4}(1+c)^2 n_{DW}^2 + (1+c) n_{DW} n_{DD} + n_{DD}^2}{n} - n_{DD}, \\ \partial_t n_{DW} - \partial_{xx}^2 n_{DW} = (1-sh)(r(1-n)+1) \frac{(1+c) n_{WW} n_{DW} + 2 n_{WW} n_{DD} + \frac{1}{2}(1-c^2) n_{DW}^2 + (1-c) n_{DW} n_{DD}}{n} - n_{DW}, \\ \partial_t n_{WW} - \partial_{xx}^2 n_{WW} = (r(1-n)+1) \frac{n_{WW}^2 + (1-c) n_{WW} n_{DW} + \frac{1}{4}(1-c)^2 n_{DW}^2}{n} - n_{WW}. \end{array} \right. \quad (3.15)$$

Model \mathcal{BA}

$$\left\{ \begin{array}{l} \partial_t n_{DD} - \partial_{xx}^2 n_{DD} = (1-s)(\max(r(1-n)(n-a)+1, 0)) \frac{\frac{1}{4}(1+c)^2 n_{DW}^2 + (1+c) n_{DW} n_{DD} + n_{DD}^2}{n} - n_{DD}, \\ \partial_t n_{DW} - \partial_{xx}^2 n_{DW} = (1-sh)(\max(r(1-n)(n-a)+1, 0)) \frac{(1+c) n_{WW} n_{DW} + 2 n_{WW} n_{DD} + \frac{1}{2}(1-c^2) n_{DW}^2 + (1-c) n_{DW} n_{DD}}{n} - n_{DW}, \\ \partial_t n_{WW} - \partial_{xx}^2 n_{WW} = (\max(r(1-n)(n-a)+1, 0)) \frac{n_{WW}^2 + (1-c) n_{WW} n_{DW} + \frac{1}{4}(1-c)^2 n_{DW}^2}{n} - n_{WW}. \end{array} \right. \quad (3.16)$$

Model \mathcal{DN}

$$\left\{ \begin{array}{l} \partial_t n_{DD} - \partial_{xx}^2 n_{DD} = (1-s)(r+1) \frac{\frac{1}{4}(1+c)^2 n_{DW}^2 + (1+c) n_{DW} n_{DD} + n_{DD}^2}{n} - (rn+1)n_{DD}, \\ \partial_t n_{DW} - \partial_{xx}^2 n_{DW} = (1-sh)(r+1) \frac{(1+c) n_{WW} n_{DW} + 2 n_{WW} n_{DD} + \frac{1}{2}(1-c^2) n_{DW}^2 + (1-c) n_{DW} n_{DD}}{n} - (rn+1)n_{DW}, \\ \partial_t n_{WW} - \partial_{xx}^2 n_{WW} = (r+1) \frac{n_{WW}^2 + (1-c) n_{WW} n_{DW} + \frac{1}{4}(1-c)^2 n_{DW}^2}{n} - (rn+1)n_{WW}. \end{array} \right. \quad (3.17)$$

Model \mathcal{DA}

$$\left\{ \begin{array}{l} \partial_t n_{DD} - \partial_{xx}^2 n_{DD} = (1-s)(r+1) \frac{\frac{1}{4}(1+c)^2 n_{DW}^2 + (1+c) n_{DW} n_{DD} + n_{DD}^2}{n} - (r(n-1)(n-a)+r+1) n_{DD}, \\ \partial_t n_{DW} - \partial_{xx}^2 n_{DW} = (1-sh)(r+1) \frac{(1+c) n_{WW} n_{DW} + 2 n_{WW} n_{DD} + \frac{1}{2}(1-c^2) n_{DW}^2 + (1-c) n_{DW} n_{DD}}{n} - (r(n-1)(n-a)+r+1) n_{DW}, \\ \partial_t n_{WW} - \partial_{xx}^2 n_{WW} = (r+1) \frac{n_{WW}^2 + (1-c) n_{WW} n_{DW} + \frac{1}{4}(1-c)^2 n_{DW}^2}{n} - (r(n-1)(n-a)+r+1) n_{WW}. \end{array} \right. \quad (3.18)$$

Note that all four models reduce to a single model for $r = 0$. This model is :

$$\left\{ \begin{array}{l} \partial_t n_{\text{DD}} - \partial_{xx}^2 n_{\text{DD}} = (1-s) \frac{\frac{1}{4}(1+c)^2 n_{\text{DW}}^2 + (1+c) n_{\text{DW}} n_{\text{DD}} + n_{\text{DD}}^2}{n} - n_{\text{DD}}, \\ \partial_t n_{\text{DW}} - \partial_{xx}^2 n_{\text{DW}} = (1-sh) \frac{(1+c) n_{\text{WW}} n_{\text{DW}} + 2 n_{\text{WW}} n_{\text{DD}} + \frac{1}{2}(1-c^2) n_{\text{DW}}^2 + (1-c) n_{\text{DW}} n_{\text{DD}}}{n} - n_{\text{DW}}, \\ \partial_t n_{\text{WW}} - \partial_{xx}^2 n_{\text{WW}} = \frac{n_{\text{WW}}^2 + (1-c) n_{\text{WW}} n_{\text{DW}} + \frac{1}{4}(1-c)^2 n_{\text{DW}}^2}{n} - n_{\text{WW}}. \end{array} \right. \quad (3.19)$$

3.B.2 Allelic densities

For our analysis, it is convenient to introduce the allelic (half-) densities $(n_{\text{D}}, n_{\text{W}})$. For a conversion occurring in the germline, we have $n_{\text{D}} = n_{\text{DD}} + \frac{1+c}{2} n_{\text{DW}}$ and $n_{\text{W}} = n_{\text{WW}} + (1 - \frac{1+c}{2}) n_{\text{DW}}$ (see section 3.2 in [129] for more details). We deduce the following systems:

Model \mathcal{BN}

$$\left\{ \begin{array}{l} \partial_t n_{\text{D}} - \partial_{xx}^2 n_{\text{D}} = n_{\text{D}} \left[\frac{r(1-n)+1}{n} \left[(1-s)n_{\text{D}} + (1-sh)(1+c)n_{\text{W}} \right] - 1 \right], \\ \partial_t n_{\text{W}} - \partial_{xx}^2 n_{\text{W}} = n_{\text{W}} \left[\frac{r(1-n)+1}{n} \left[n_{\text{W}} + (1-sh)(1-c)n_{\text{D}} \right] - 1 \right]. \end{array} \right. \quad (3.20)$$

Model \mathcal{BA}

$$\left\{ \begin{array}{l} \partial_t n_{\text{D}} - \partial_{xx}^2 n_{\text{D}} = n_{\text{D}} \left[\frac{\max(r(1-n)(n-a)+1, 0)}{n} \left[(1-s)n_{\text{D}} + (1-sh)(1+c)n_{\text{W}} \right] - 1 \right], \\ \partial_t n_{\text{W}} - \partial_{xx}^2 n_{\text{W}} = n_{\text{W}} \left[\frac{\max(r(1-n)(n-a)+1, 0)}{n} \left[n_{\text{W}} + (1-sh)(1-c)n_{\text{D}} \right] - 1 \right]. \end{array} \right. \quad (3.21)$$

Model \mathcal{DN}

$$\left\{ \begin{array}{l} \partial_t n_{\text{D}} - \partial_{xx}^2 n_{\text{D}} = n_{\text{D}} \left[\frac{r+1}{n} \left[(1-s)n_{\text{D}} + (1-sh)(1+c)n_{\text{W}} \right] - (rn+1) \right], \\ \partial_t n_{\text{W}} - \partial_{xx}^2 n_{\text{W}} = n_{\text{W}} \left[\frac{r+1}{n} \left[n_{\text{W}} + (1-sh)(1-c)n_{\text{D}} \right] - (rn+1) \right]. \end{array} \right. \quad (3.22)$$

Model \mathcal{DA}

$$\left\{ \begin{array}{l} \partial_t n_{\text{D}} - \partial_{xx}^2 n_{\text{D}} = n_{\text{D}} \left[\frac{r+1}{n} \left[(1-s)n_{\text{D}} + (1-sh)(1+c)n_{\text{W}} \right] - (r(n-1)(n-a) + r + 1) \right], \\ \partial_t n_{\text{W}} - \partial_{xx}^2 n_{\text{W}} = n_{\text{W}} \left[\frac{r+1}{n} \left[n_{\text{W}} + (1-sh)(1-c)n_{\text{D}} \right] - (r(n-1)(n-a) + r + 1) \right]. \end{array} \right. \quad (3.23)$$

3.B.3 Allelic frequencies

It may sometimes be more appropriate to study the allelic frequencies $p_{\text{D}} = \frac{n_{\text{D}}}{n_{\text{D}}+n_{\text{W}}}$, $p_{\text{W}} = \frac{n_{\text{W}}}{n_{\text{D}}+n_{\text{W}}}$. The models become:

Model \mathcal{BN}

$$\begin{cases} \partial_t n - \partial_{xx}^2 n = (r(1-n) + 1) \left((1-s) p_D^2 + 2(1-sh) p_D (1-p_D) + (1-p_D)^2 \right) n - n, \\ \partial_t p_D - \partial_{xx}^2 p_D = 2 \partial_x \log(n) \partial_x p_D + (r(1-n) + 1) \left((2h-1) s p_D + (1-sh)(1+c) - 1 \right) p_D (1-p_D). \end{cases} \quad (3.24)$$

Model \mathcal{BA}

$$\begin{cases} \partial_t n - \partial_{xx}^2 n = (\max(r(1-n)(n-a) + 1, 0)) \left((1-s) p_D^2 + 2(1-sh) p_D (1-p_D) + (1-p_D)^2 \right) n - n, \\ \partial_t p_D - \partial_{xx}^2 p_D = 2 \partial_x \log(n) \partial_x p_D + (\max(r(1-n)(n-a) + 1, 0)) \left((2h-1) s p_D + (1-sh)(1+c) - 1 \right) p_D (1-p_D). \end{cases} \quad (3.25)$$

Model \mathcal{DN}

$$\begin{cases} \partial_t n - \partial_{xx}^2 n = (r+1) \left((1-s) p_D^2 + 2(1-sh) p_D (1-p_D) + (1-p_D)^2 \right) n - (rn+1) n, \\ \partial_t p_D - \partial_{xx}^2 p_D = 2 \partial_x \log(n) \partial_x p_D + (r+1) \left((2h-1) s p_D + (1-sh)(1+c) - 1 \right) p_D (1-p_D). \end{cases} \quad (3.26)$$

Model \mathcal{DA}

$$\begin{cases} \partial_t n - \partial_{xx}^2 n = (r+1) \left((1-s) p_D^2 + 2(1-sh) p_D (1-p_D) + (1-p_D)^2 \right) n - (r(n-1)(n-a) + r+1) n, \\ \partial_t p_D - \partial_{xx}^2 p_D = 2 \partial_x \log(n) \partial_x p_D + (r+1) \left((2h-1) s p_D + (1-sh)(1+c) - 1 \right) p_D (1-p_D). \end{cases} \quad (3.27)$$

Equations on p_D differ from the standard equation often used in populations genetics, as it contains an advection term $2 \partial_x(\log n) \partial_x p_D$. This term appears when calculating $\partial_{xx}^2 p_D = \partial_{xx}^2 \frac{n_{DD}}{n}$ and represents a demographic flux from denser to less dense areas, due to variations in population density. It is opposed to the spread of the costly drive allele (see Figure 2 [95]).

3.B.4 Final allelic proportions for r small and large

In models \mathcal{BN} and \mathcal{BA} for large values of r , using the Strugarek-Vauchelet rescaling [201] in (3.24) and (3.25), the systems reduce to one limit equation on p_D :

$$\partial_t p_D - \partial_{xx}^2 p_D = \frac{\left((2h-1) s p_D + (1-sh)(1+c) - 1 \right) p_D (1-p_D)}{(1-s) p_D^2 + 2(1-sh) p_D (1-p_D) + (1-p_D)^2 - 1}. \quad (3.28)$$

Equation (3.28) has already been studied in [187]: a heatmap illustrates the final proportions in the case $c = 0.85$ (Figure 4). In models \mathcal{DN} and \mathcal{DA} , the equation on p_D in (3.24) and (3.25) is:

$$\partial_t p_D - \partial_{xx}^2 p_D = 2 \partial_x \log(n) \partial_x p_D + \underbrace{(r+1) \left((2h-1) s p_D + (1-sh)(1+c) - 1 \right) p_D (1-p_D)}_{\text{reaction term}}. \quad (3.29)$$

The reaction term in equation (3.29) becomes larger as r increases: this indicates us that the traveling wave has an infinite speed when r tends to infinity, meaning that the equilibrium is reached

instantaneously. Therefore, the term $2 \partial_x \log(n) \partial_x p_D$ is instantaneously zero and the final proportions are the same as for models \mathcal{BN} and \mathcal{BA} .

As a consequence, all models \mathcal{BN} , \mathcal{BA} , \mathcal{DN} and \mathcal{DA} share the same final proportions for large values of r . This conclusion also holds for $r = 0$, as the models are equal (see Appendix 3.B.1). These proportions have already been determined in a previous article [129], in \mathcal{BN} . We recall these results and generalise them to our four models.

3.B.5 Speed of the problem simplified at low drive density

In Section 3.3.3, we focus on drive invasion and therefore consider low drive density and high wild-type density at the front of the wave. The speed v for the drive wave can be calculated when the models are simplified (linearised) at low drive density: it is deduced from the reproduction of the few drive individuals at the front wave,

$$v = 2 \sqrt{\lim_{n_D \rightarrow 0} \left(\frac{F(n_D)}{n_D} \right)} \quad (3.30)$$

where F represents the net production of drive alleles.

Model \mathcal{BN}

$$F^{(\mathcal{BN})}(n_D) = n_D \left[(r(1-n) + 1) \left[(1-s) \frac{n_D}{n} + (1-sh)(1+c) \frac{n_W}{n} \right] - 1 \right] \quad (3.31)$$

Model \mathcal{BA}

$$F^{(\mathcal{BA})}(n_D) = n_D \left[\max(r(1-n)(n-a) + 1, 0) \left[(1-s) \frac{n_D}{n} + (1-sh)(1+c) \frac{n_W}{n} \right] - 1 \right]. \quad (3.32)$$

Model \mathcal{DN}

$$F^{(\mathcal{DN})}(n_D) = n_D \left[(r+1) \left[(1-s) \frac{n_D}{n} + (1-sh)(1+c) \frac{n_W}{n} \right] - (rn+1) \right]. \quad (3.33)$$

Model \mathcal{DA}

$$F^{(\mathcal{DA})}(n_D) = n_D \left[(r+1) \left[(1-s) \frac{n_D}{n} + (1-sh)(1+c) \frac{n_W}{n} \right] - (r(n-1)(n-a) + r + 1) \right]. \quad (3.34)$$

Considering high wild-type density at the front of the wave ($n_W \approx n \approx 1$), the speed in models \mathcal{BN} and \mathcal{BA} is given by:

$$v_{\mathcal{BN},\mathcal{BA}} = 2 \sqrt{(1-sh)(1+c) - 1} \quad (3.35)$$

and becomes in models \mathcal{DN} and \mathcal{DA} :

$$v_{\mathcal{DN},\mathcal{DA}} = 2 \sqrt{(1+r) [(1-sh)(1+c) - 1]} \quad (3.36)$$

To understand why $v_{\mathcal{DN},\mathcal{DA}}$ is greater by a coefficient of $\sqrt{1+r}$ than $v_{\mathcal{BN},\mathcal{BA}}$, we have to understand the population dynamics at the front of the wave. There, the density is close to the maximum carrying capacity 1, with low drive density ($n_D \approx 0$) and high wild-type density ($n_W \approx n \approx 1$). On one hand,

in models \mathcal{BN} and \mathcal{BA} , the density-dependence constraint is placed on the birth term, reducing the production rate of drive alleles to $(1 - sh)(1 + c)$ while they disappear at rate 1 (3.31, 3.32). On the other hand, in models \mathcal{DN} and \mathcal{DA} , the density-dependence constraint is placed on the death term increasing to $(r + 1)$ the rate at which the drive alleles disappears, while they are produced at rate $(r + 1)(1 - sh)(1 + c)$ (3.33, 3.34). Consequently, the net production remains constant $(1 - sh)(1 + c)$, but the turnover rate is $r + 1$ times greater. As the wave movement largely relies on the reproduction, this reflects in the speed formula: the propagation is $\sqrt{r + 1}$ times faster.

3.C Final density after a drive invasion

3.C.1 Comparison of the final density in models \mathcal{BN} and \mathcal{DN}

We compare the final density for model \mathcal{BN} and \mathcal{DN} in case of persistence.

$$n_{\text{D}}^{*(\mathcal{BN})} = 1 - \frac{-s}{r(1-s)} = \frac{r(1-s) - s}{r(1-s)} \quad \text{and} \quad n_{\text{D}}^{*(\mathcal{DN})} = 1 - \frac{s(r+1)}{r} = \frac{r(1-s) - s}{r} \quad (3.37)$$

Therefore, the final density is $1 - s$ times lower in model \mathcal{DN} than in model \mathcal{BN} , in case of persistence.

3.C.2 Comparison of the final density in models \mathcal{BN} and \mathcal{BA}

We compare the final density for model \mathcal{BN} and \mathcal{BA} in case of persistence:

$$n_{\text{D}}^{*(\mathcal{BN})} - n_{\text{D}}^{*(\mathcal{BA})} = \frac{2r(1-s) - 2s - r(1-s)(1+a) - r(1-s)\sqrt{(1-a)^2 - 4\frac{s}{r(1-s)}}}{2r(1-s)} \quad (3.38a)$$

$$> \frac{-2s + r(1-s)(1-a) - r(1-s)(1-a) + r(1-s)\sqrt{4\frac{s}{r(1-s)}}}{2r(1-s)} \quad (3.38b)$$

$$= \frac{-2s + 2\sqrt{r(1-s)}s}{2r(1-s)} \quad (3.38c)$$

$$> \frac{-2s + 2\sqrt{s^2}}{2r(1-s)} = 0. \quad (3.38d)$$

because $\sqrt{\alpha - \beta} > \sqrt{\alpha} - \sqrt{\beta}$ for α and β positive real numbers (3.38a to 3.38b), and $r(1-s) > s$ outside the eradication area since $n_{\text{D}}^{*(\mathcal{BN})} > 0$ (3.38c to 3.38d).

Since the strictly positive final density $n_{\text{D}}^{*(\mathcal{BN})}$ is always higher than $n_{\text{D}}^{*(\mathcal{BA})}$, and because the eradication in model \mathcal{BN} necessarily implies the eradication in model \mathcal{BA} , we conclude that the final density after a drive invasion is always smaller in model \mathcal{BA} compare to model \mathcal{BN} .

3.C.3 Comparison of the final density in models \mathcal{DN} and \mathcal{DA}

In Figure 3.C.1, we compare the final density for model \mathcal{DN} and \mathcal{DA} in case of persistence:

$$n_{\text{D}}^{*(\mathcal{DN})} - n_{\text{D}}^{*(\mathcal{DA})} = \frac{2r - 2s(r+1) - r(1+a) - r\sqrt{(1-a)^2 - 4\frac{s(r+1)}{r}}}{2r} \quad (3.39a)$$

$$> \frac{-2s(r+1) + r(1-a) - r(1-a) + 2r\sqrt{\frac{s(r+1)}{r}}}{2r(1-s)} \quad (3.39b)$$

$$= \frac{-2s(r+1) + 2\sqrt{s(r+1)r}}{2r(1-s)} \quad (3.39c)$$

$$> \frac{-2s(r+1) + 2\sqrt{s^2(r+1)^2}}{2r(1-s)} = 0. \quad (3.39d)$$

because $\sqrt{\alpha - \beta} > \sqrt{\alpha} - \sqrt{\beta}$ for α and β positive real numbers (3.39a to 3.39b), and $r > s(r+1)$ outside the eradication area since $n_{\text{D}}^{*(\mathcal{DN})} > 0$ (3.39c to 3.39d).

Since the strictly positive final density $n_{\text{D}}^{*(\mathcal{DN})}$ is always higher than $n_{\text{D}}^{*(\mathcal{DA})}$, and because the eradication in model \mathcal{DN} necessarily implies the eradication in model \mathcal{DA} , we conclude that the final density after a drive invasion is always smaller in model \mathcal{DA} compare to model \mathcal{DN} .

3.C.4 Comparison of the final density in models \mathcal{BA} and \mathcal{DA}

We compare the final density for model \mathcal{BA} and \mathcal{DA} in case of persistence.

$$n_{\text{D}}^{*(\mathcal{BA})} = \frac{1 + a + \sqrt{(1+a)^2 - 4(a + \frac{s}{r(1-s)})}}{2} \quad \text{and} \quad n_{\text{D}}^{*(\mathcal{DA})} = \frac{1 + a + \sqrt{(1+a)^2 - 4(a + \frac{s(r+1)}{r})}}{2} \quad (3.40)$$

As $r(1-s) > s$ outside the eradication area, we have:

$$\frac{s}{r(1-s)} < \frac{s(r+1)}{r} \quad \iff \quad n_{\text{D}}^{*(\mathcal{BA})} > n_{\text{D}}^{*(\mathcal{DA})} \quad (3.41)$$

Since the strictly positive final density $n_{\text{D}}^{*(\mathcal{BA})}$ is always higher than $n_{\text{D}}^{*(\mathcal{DA})}$, and because the eradication in model \mathcal{BA} necessarily implies the eradication in model \mathcal{DA} , we conclude that the final density after a drive invasion is always smaller in model \mathcal{DA} compare to model \mathcal{BA} .

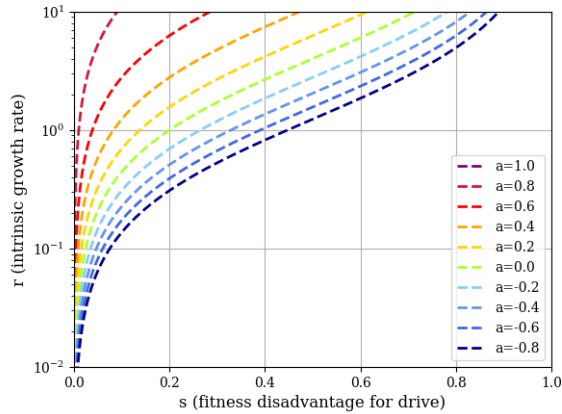
Note that:

$$\partial_a n_{\text{D}}^{*(\mathcal{BA})} = 1 - \frac{2(1-a)}{2\sqrt{(1-a)^2 - 4\frac{s}{r(1-s)}}} \leq 0 \quad \text{and} \quad \partial_a n_{\text{D}}^{*(\mathcal{DA})} = 1 - \frac{2(1-a)}{2\sqrt{(1-a)^2 - 4\frac{s(r+1)}{r}}} \leq 0 \quad (3.42)$$

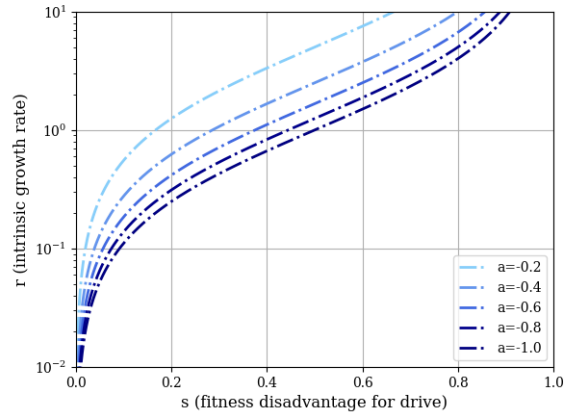
Therefore, in both models \mathcal{BA} and \mathcal{DA} , the stronger the Allee effect, the smaller the final density.

3.C.4.1 Pure drive eradication, bistable and persistence boundary lines for different values of a in models \mathcal{BA} and \mathcal{DA}

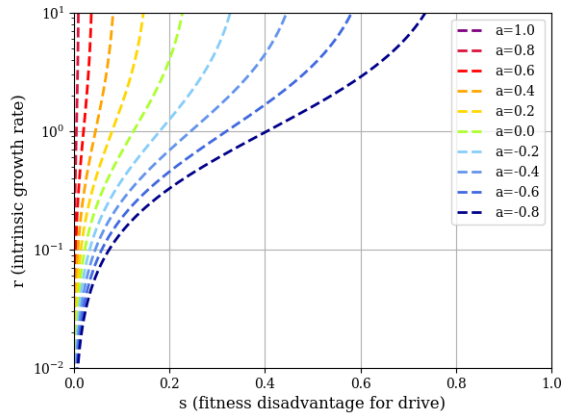
We compare boundary lines in models with Allee effect (\mathcal{BA} and \mathcal{DA}) for different values of a .



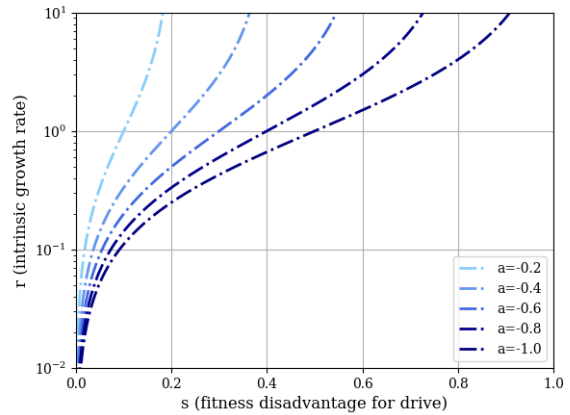
(a) Boundary line between pure drive eradication and the bistable areas in model \mathcal{BA} , for different values of a .



(b) Boundary line between pure drive bistable and persistence areas in model \mathcal{BA} , for different values of a .



(c) Boundary line between pure drive eradication and bistable areas in model \mathcal{DA} , for different values of a .



(d) Boundary line between the pure drive bistable and persistence areas in model \mathcal{DA} , for different values of a .

Figure 3.C.1: Boundary lines between the pure drive eradication, bistable and persistence areas. Without Allee effect, there only exists pure drive eradication and persistence areas (Models \mathcal{BN} and \mathcal{DN}). However, if we consider an Allee effect (Models \mathcal{BA} and \mathcal{DA}), an other area appears: the pure drive bistable area. In this area, the population only persists if the drive was introduced in a large enough density. We observe that the larger a is, the more persistence and bistable areas are restricted to high values of r and small values of s .

Chapter 4

Stochastic dynamics at the back of gene drive propagation: drive eradication, wild-type recolonisation or chasing

Léna KLÄY^{*,1}, Léo GIRARDIN², Florence DÉBARRE¹, Vincent CALVEZ²

¹ Institute of Ecology and Environmental Sciences Paris (IEES Paris), Sorbonne Université, CNRS, IRD, INRAE, Université Paris Est Creteil, Université de Paris, Paris Cedex 5, France.

² Institut Camille Jordan, UMR 5208 CNRS and Université Claude Bernard Lyon 1, France

Abstract

Gene drive alleles positively bias their own inheritance to offspring, possibly resulting in allele fixation inside a wild-type population despite fitness cost. If the fitness cost is high enough, one potential intended outcome is the eradication of the population. However, this outcome might be prevented or delayed by local wild-type recolonisation in areas previously cleared by the drive alleles. In this paper, we dissect the conditions under which these stochastic wild-type recolonising events are likely, or rather unlikely, to occur in one spatial dimension. More precisely, we examine the conditions ensuring that the last wild-type individual is surrounded by a large enough number of drive individuals, resulting in a very low chance of wild-type recolonisation. We accumulate numerical evidence that characterising the absence of recolonising events can be reduced heuristically to the extinction time of a spatial Galton-Watson process in a bounded domain of a suitable size with appropriate initialisation. However, we were not able to characterise this extinction time analytically: this problem is left open. Numerically, we show that the number of wild-type recolonising events increases as the fitness of drive individuals gets smaller and decreases as the local carrying capacity gets larger. Overall, this study paves the way for further analysis of wild-type recolonisation at the back of eradication traveling waves.

*Corresponding author: lena.klay@sorbonne-universite.fr

Contents

4.1	Introduction	120
4.2	Models and methods	122
4.2.1	Continuous deterministic model	122
4.2.2	Discrete stochastic model	124
4.2.3	Parameters and initial conditions for numerical simulations	127
4.2.4	Setting of the problem	128
4.3	Results	131
4.3.1	Determining the almost deterministic distance between the drive and wild-type wave at a level line of 100 individuals	131
4.3.2	Characterising the stochastic fluctuations of the last wild-type individual	132
4.3.3	Numerical conclusions	136
4.3.4	Relation between drive intrinsic fitness, diffusion and risk of wild-type recolonisation.	137
4.4	Discussion	138

4.1 Introduction

Artificial gene drive is a genetic engineering technology that could be used for the control of natural populations. Gene drive alleles bias their inheritance ratio towards a super-Mendelian rate, therefore driving themselves to spread quickly through a population despite a potential fitness cost [4, 34, 35]. Homing gene drives rely on gene conversion to bias their transmission. In a heterozygous cell, the gene drive cassette located on one chromosome induces a double-strand break on the homologous chromosome. This damage is repaired by the cell through homology direct-repair, which duplicates the cassette. This gene conversion repeats through the generations and largely benefits to the drive propagation. The conversion can theoretically take place either in the germline or in the zygote.

Gene drive constructs can be designed to either spread a gene of interest in a population (*population replacement*), or to reduce the population size by lowering the fitness of drive individuals (*population suppression* if the intended goal is to reduce the population density or *population eradication* if the intended goal is to eradicate the population). This fitness cost, usually the alteration of an essential fertility or viability gene, associated with the super-Mendelian propagation can lead to the complete extinction of the population [133, 104, 34, 97, 95].

However, eradication may fail because of the recolonisation of wild-type individuals which prevents the total elimination of the target population. Some wild-type individuals might then stay indefinitely in the environment, or be again invaded by drive individuals, leading to local extinction of the population, but the resulting cleared area can afterwards be again recolonised by some wild-type individuals, and so on. These infinite dynamics have been referred to as “colonisation-extinction” dynamics [171] or “chasing” dynamics [46]. In this paper, we use the term “wild-type recolonisation” if the recolonising wild-type individuals stay indefinitely in the environment, and “chasing” if at least another drive recolonising event follows, leading to potential infinite re-invasions (see illustrations in Figure 4.1).

Wild-type recolonising dynamics have been observed in compartmental models [33] and discrete individual-based models [171, 170, 25, 46, 47, 143, 142, 177, 226]. Previous studies have already shown

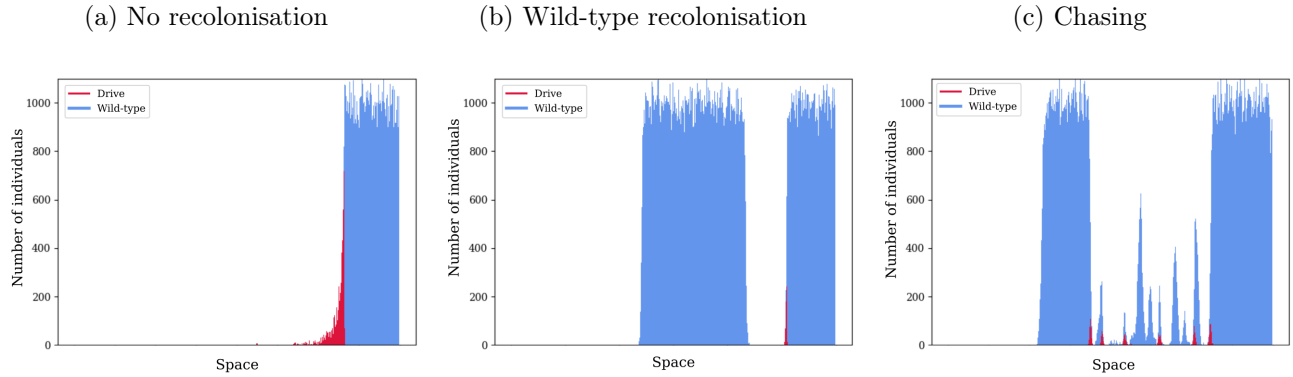


Figure 4.1: The drive eradication wave (moving from the left to the right) might lead to different dynamics in space: (a) no recolonisation event at all, (b) one wild-type recolonisation event only or (c) successive drive and wild-type recolonisation events (*chasing*). The drive individuals are plotted in red while the wild-type individuals are plotted in blue.

the influence of some ecological factors over these recolonising probability. High rates of dispersal reduce the chance of recolonisation [177, 46], high levels of inbreeding increase the likelihood of recolonisation [46], and the presence of a competing species or predator facilitate extinction without recolonisation [143]. The release pattern seems to have little effect on recolonisation outcomes [46]. The drive fitness might also impact the risk of recolonisation: based on simulations, Champer et al. found that a higher fitness reduces the chance of recolonisation [46], while Paril and Phillips obtained contrasted results on X-shredder and W-shredder drives [177]. Our study continues this work by investigate the impact of fitness though an analytical approach supported by simulations.

However, the precise mechanism leading to wild-type recolonising events remains unclear, and we wonder how and when it is initiated. Simulations in one-dimensional space suggest that if the last wild-type individual is surrounded by a large number of drive individuals, then wild-type recolonising events are very unlikely to occur. This statement seems strongly linked with population density, but recolonising dynamics has always been studied on relatively small populations, around 10^5 individuals over the whole domain, in one and two dimensions (one spatial dimension: 3925 individuals [46] or 12500 individuals [177] and two spatial dimensions: 50000 individuals [46, 47, 143, 142, 226]).

In this study, we investigate how the carrying capacity and the drive fitness influence the wild-type recolonising probability. We study a stochastic model which follows the propagation of the drive in space and time. We implement this model in a population-based way, such that we follow the density of each spatial site instead of the position of each living individual. The main advantage of this approach is that it allows to simulate systems with very large number of individuals. In one spatial dimension, we accumulate evidence that characterising the absence of wild-type recolonising events can be reduced heuristically to the extinction time of a spatial Galton-Watson process in a bounded domain of a suitable size with appropriate initialisation. We leave open the quality of the approximation in mathematical terms, as well as the analysis of the distribution of the extinction time for the reduced Galton-Watson process. In the biological point of view, our conclusions are the following: the number of wild-type recolonising events decreases as the local carrying capacity gets larger, and a higher fitness for drive individuals also reduces the likelihood of wild-type recolonisation .

4.2 Models and methods

4.2.1 Continuous deterministic model

We extend a model developed in a previous article [129], assuming a partial conversion of rate $c \in (0, 1)$ occurring in the germline. We denote by n the total density, and by n_i the density of individuals with genotype i . We follow three genotypes: wild-type homozygotes ($i = WW$), drive homozygotes ($i = DD$) and heterozygotes ($i = DW$). Wild-type homozygotes have fitness $f_{WW} = 1$, drive homozygotes have fitness $f_{DD} = 1 - s$, where $s \in (0, 1)$ is the fitness cost of the drive, and drive heterozygotes have fitness $f_{DW} = 1 - sh$, where $h \in (0, 1)$ is the dominance parameter. The intrinsic growth rate is $r \in (0, +\infty)$. We assume a random mixing locally and do not distinguish sexes.

$$\left\{ \begin{array}{l} \partial_t n_{DD} - \partial_{xx}^2 n_{DD} = (1-s)(r(1-n)+1) \frac{\frac{1}{4}(1+c)^2 n_{DW}^2 + (1+c)n_{DW}n_{DD} + n_{DD}^2}{n} - n_{DD}, \\ \partial_t n_{DW} - \partial_{xx}^2 n_{DW} = (1-sh)(r(1-n)+1) \frac{(1+c)n_{WW}n_{DW} + 2n_{WW}n_{DD} + \frac{1}{2}(1-c^2)n_{DW}^2 + (1-c)n_{DW}n_{DD}}{n} \\ \quad - n_{DW}, \\ \partial_t n_{WW} - \partial_{xx}^2 n_{WW} = (r(1-n)+1) \frac{n_{WW}^2 + (1-c)n_{WW}n_{DW} + \frac{1}{4}(1-c)^2 n_{DW}^2}{n} - n_{WW}. \end{array} \right. \quad (4.1)$$

4.2.1.1 From genotypes to alleles

We previously showed that model (4.1) can be reduced to two equations instead of three, focusing on the allele densities (n_D, n_W) instead of the genotype densities (n_{DD}, n_{DW}, n_{WW}) [129]. The transformation is given by $n_D = n_{DD} + \alpha n_{DW}$ and $n_W = n_{WW} + (1-\alpha)n_{DW}$, with $\alpha = \frac{1+c}{2}$ when conversion occurs in the germline [129]. We deduce from system (4.1):

$$g_D(n_D, n_W) = (r(1-n)+1) \left[(1-s) \frac{n_D}{n} + (1-sh)(1+c) \frac{n_W}{n} \right] \quad (4.2)$$

$$g_W(n_D, n_W) = (r(1-n)+1) \left[\frac{n_W}{n} + (1-sh)(1-c) \frac{n_D}{n} \right] \quad (4.3)$$

with the mean positive allele production during one unit of time, g_D for drive and g_W for wild-type given by:

$$\left\{ \begin{array}{l} \partial_t n_D - \partial_{xx}^2 n_D = n_D \left[g_D(n_D, n_W) - 1 \right], \\ \partial_t n_W - \partial_{xx}^2 n_W = n_W \left[g_W(n_D, n_W) - 1 \right]. \end{array} \right. \quad (4.4)$$

4.2.1.2 Traveling waves

We seek stationary solutions in a reference frame moving at speed v_{co} (**co** stands for **continuous model**), where v_{co} is some unknown:

$$\begin{cases} n_D(t, x) = N_D(x - v_{co}t) = N_D(z) & (\forall t > 0) (\forall x \in \mathbb{R}), \\ n_W(t, x) = N_W(x - v_{co}t) = N_W(z) & (\forall t > 0) (\forall x \in \mathbb{R}). \end{cases} \quad (4.5)$$

Traveling wave solutions contain important information for the biological interpretation of the results: in this paper, we focus on the position of the last wild-type individual relatively to the moving frame, to forecast the wild-type recolonising possibility. From equation (4.4), we deduce the equation for the traveling wave density profiles (N_D , N_W):

$$\begin{cases} -v_{co}N'_D - N''_D &= N_D \left[(r(1-N) + 1) \left[(1-s)\frac{N_D}{N} + (1-sh)(1+c)\frac{N_W}{N} \right] - 1 \right], \\ -v_{co}N'_W - N''_W &= N_W \left[(r(1-N) + 1) \left[\frac{N_W}{N} + (1-sh)(1-c)\frac{N_D}{N} \right] - 1 \right]. \end{cases} \quad (4.6)$$

Wild-type recolonisation is possible only when the drive eradicates the population, as it leaves empty areas potentially recolonised by the wild type. Thus, we study the dynamics at the back of the drive eradication wave. Our study requires knowing the speed at which the wave propagates, which is explicitly known only if the wave is pulled: therefore, we make this assumption. As shown in [129], when conversion occurs in the germline, the drive wave is always pulled for $h < 0.5$. In this paper we consider $h = 0.4$ for the numerical illustrations, however all the results hold for any pulled eradication drive wave. If the wave is pulled, the speed is given by the minimal speed v_{co}^{lin} of the problem linearised at low drive density (for details, see Section 4.A.1):

$$v_{co} = v_{co}^{lin} = 2\sqrt{(1-sh)(1+c) - 1}, \quad (4.7)$$

A schematic illustration of the wave is drawn in Figure 4.2. The slopes at the back and at the front of the wave can be approximated by exponential functions:

$$\text{At the front of the wave: } N_D(z) \approx \exp(\lambda_{D, co}^{front} z), \quad (4.8a)$$

$$\text{At the back of the wave: } N_D(z) \approx \exp(\lambda_{D, co}^{back} z), \quad (4.8b)$$

$$\text{At the back of the wave: } N_W(z) \approx \exp(\lambda_{W, co}^{back} z). \quad (4.8c)$$

whose exponents are given by (details in Section 4.A.1):

$$\lambda_{D, co}^{front} = -\sqrt{(1-sh)(1+c) - 1}, \quad (4.9a)$$

$$\lambda_{D, co}^{back} = -\sqrt{(1-sh)(1+c) - 1} + \sqrt{(1-sh)(1+c) - (r+1)(1-s)}, \quad (4.9b)$$

$$\lambda_{W, co}^{back} = -\sqrt{(1-sh)(1+c) - 1} + \sqrt{(1-sh)(1+c) - (r+1)(1-sh)(1-c)}. \quad (4.9c)$$

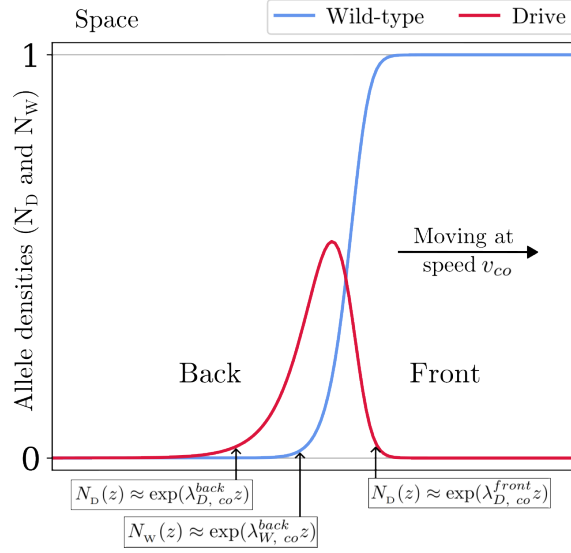


Figure 4.2: Schematic illustration of a pulled eradication drive wave. The slopes at the back and at the front of the wave can be approximated by exponential functions.

4.2.1.3 Approximations of the distances at the back of the wave

We consider η and \mathcal{N} , two numbers of individuals with $\eta < \mathcal{N}$. Let \mathcal{N} be reached in the wave at an arbitrary $z = 0$, as in Figure 4.3. In an exponential profile, the distance $L_\eta^\mathcal{N}$ between the two spatial sites with respectively η and \mathcal{N} individuals is given by :

$$\mathcal{N} e^{\lambda(-L_\eta^\mathcal{N})} = \eta \iff L_\eta^\mathcal{N} = \frac{\log(\mathcal{N}) - \log(\eta)}{\lambda}. \quad (4.10)$$

In one spatial dimension, wild-type recolonisation is possible only if the last wild-type individual is further away than (i.e. to the left of) the last drive individual at some point. Setting $\eta = 1$ gives an approximation of the relative position of the last wild-type individual in the wave. However, the stochastic fluctuations that enable wild-type recolonising dynamics might also degrade the quality of this approximation.

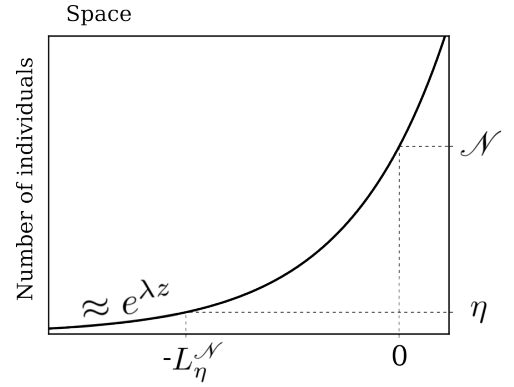


Figure 4.3: Schematic illustration of the back of a wave (drive or wild-type).

4.2.2 Discrete stochastic model

To observe wild-type recolonising dynamics, we need to consider a stochastic discrete model, based on the allele dynamics described by system (4.4). We denote the density of each allele at spatial site x and time t as:

$$n_D(x, t) = n_D^{t,x}, \quad n_W(t, x) = n_W^{t,x} \quad \text{and} \quad n(x, t) = n^{t,x} \quad \forall t \in \{0, dt, 2dt, \dots\}, \quad \forall x \in \{0, dx, 2dx, \dots\}. \quad (4.11)$$

In our stochastic simulations, each individual reproduces, dies and disperses independently of the reproduction, death, and dispersal of others. We alternate between two types of events: 1) allele production/disappearance in each spatial site and 2) allele migration in the neighbouring spatial sites. During one unit of time, a drive allele duplicates in average $g_D(n_D^{t,x}, n_W^{t,x})$ times defined in equation (4.2) and a wild-type allele duplicates in average $g_W(n_D^{t,x}, n_W^{t,x})$ defined in equation (4.3). One allele (drive or wild-type) disappears at rate 1 within one unit of time. To model the number of new alleles and removed alleles, we use Poisson distributions with these respective means. A Poisson distribution corresponds to the number of events observed during a time interval of given length, when the waiting time between two events is given by an exponential distribution. Under an exponential distribution, the expected future waiting time is independent of the past waiting time for one type of event.

Second, we consider migration: at each time step, an allele migrates with probability m outside its original site. It goes to one neighbouring spatial site, either on the right or on the left, with equal probability. To model this event we use two Bernoulli distributions: one with probability m to determine if the individual migrates, the second with probability $\frac{1}{2}$ in case of success, to determine the welcoming site (right or left).

We implement this model in a population-based way: we follow the density of each spatial site instead of following the position of each living individual (which would be an individual-based model). Therefore, the mean of the Poisson law for each event of allele production/disappearance is multiplied by the allele number in the spatial site, and the Bernoulli laws become Binomial laws with the number of alleles being the number of independent experiments replicated. These dynamics are summarised in Figure 4.4 and the detailed code is provided in Appendix 4.B. The main advantage of this approach is that it allows to simulate systems with very large number of individuals.

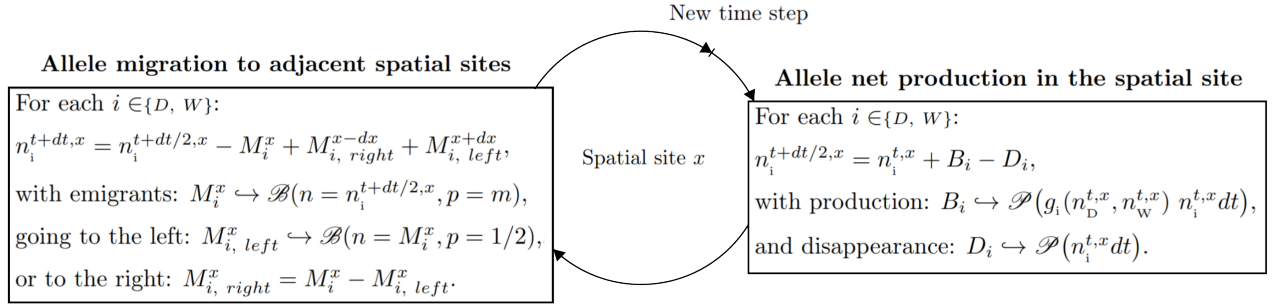


Figure 4.4: Consecutive steps in the stochastic discrete model.

4.2.2.1 Traveling waves in the discrete, deterministic model

The speed formula of the traveling wave slightly changes in the discrete deterministic model (mean of the stochastic dynamics) because of the discretisation. We now denote the speed by v_{di} in the discrete model.

$$\begin{cases} n_D^{t,x} = N_D^{x-v_{di}t} = N_D^z, \\ n_W^{t,x} = N_W^{x-v_{di}t} = N_W^z. \end{cases} \quad (4.12)$$

If the wave is pulled, we know that this speed is given by the minimal speed v_{di}^{lin} of the discrete

problem linearized at low drive density (for details, see Section 4.A.2):

$$v_{di} = v_{di}^{lin} = \min_{\lambda < 0} \left(\mathcal{V}(\lambda) \right) = \min_{\lambda < 0} \left(\frac{\log \left(\left[((1 - sh)(1 + c) - 1) dt + 1 \right] \left[1 - m + m \cosh(\lambda dx) \right] \right)}{-\lambda dt} \right) \quad (4.13)$$

In case of a pulled wave without wild-type recolonising event, we approximate the slopes at the back and at the front of the traveling wave with exponential functions:

$$\text{At the front of the wave: } N_D^z \approx \exp(\lambda_{D, di}^{front} z), \quad (4.14a)$$

$$\text{At the back of the wave: } N_D^z \approx \exp(\lambda_{D, di}^{back} z), \quad (4.14b)$$

$$\text{At the back of the wave: } N_W^z \approx \exp(\lambda_{W, di}^{back} z). \quad (4.14c)$$

with the exponents verifying:

$$e^{-\lambda_{D, di}^{front} v_{di} dt} = \left(((1 - sh)(1 + c) - 1) dt + 1 \right) \left(1 - m + m \cosh(\lambda_{D, di}^{front} dx) \right) \quad (4.15a)$$

$$e^{-\lambda_{D, di}^{back} v_{di}^{lin} dt} = \left(((r + 1)(1 - s) - 1) dt + 1 \right) \left(1 - m + \frac{m}{2} (e^{\lambda_{D, di}^{back} dx} + e^{-\lambda_{D, di}^{back} dx}) \right). \quad (4.15b)$$

$$e^{-\lambda_{W, di}^{back} v_{di}^{lin} dt} = \left(((r + 1)(1 - sh)(1 - c) - 1) dt + 1 \right) \left(1 - m + \frac{m}{2} (e^{\lambda_{W, di}^{back} dx} + e^{-\lambda_{W, di}^{back} dx}) \right). \quad (4.15c)$$

The minimum in (4.13) is reached for $\lambda = \lambda_{D, di}^{front}$, from which we deduce equation (4.15a).

4.2.2.2 Correction of the speed due to stochastic variations at the front of the wave

The wave is pulled, therefore its speed is determined by the few drive individuals at the front of the wave. The model being stochastic, small densities exhibit stochastic fluctuations in their dynamics, so that the speed value slightly differ from (4.13). This difference has been analytically determined in [28, 29] (see also [20, 159] for the mathematical perspective) to be :

$$v_{di}^{(cor)} \approx v_{di} - \frac{\mathcal{V}''(\lambda_{D, di}^{front}) \pi^2 (\lambda_{D, di}^{front})^2}{2 \left(\log \left(\frac{1}{K} \right) \right)^2}, \quad (4.16)$$

where $v_{di}^{(cor)}$ is the corrected discrete speed, and \mathcal{V} is given in equation (4.13). In Table 4.1, we compare for $s = 0.3$ and $s = 0.7$: i) the continuous speed v_{co} given by (4.7), ii) the discrete speed not corrected v_{di} given by (4.13), iii) the discrete speed corrected $v_{di}^{(cor)}$ given by (4.16) and iv) the numerical speed v_{num} given by the simulation. The best approximation of the speed computed numerically is indeed the discrete speed corrected by (4.16).

s value	Speed			
	v_{co}	v_{di}	$v_{di (cor)}$	v_{num}
0.3	1.64	1.63	1.62	1.60
0.7	1.21	1.21	1.20	1.19

Table 4.1: Comparison of the speed values for $s = 0.3$ and $s = 0.7$, for the other parameter values, see Table 4.2. The continuous speed v_{co} is given by (4.7), the discrete speed not corrected v_{di} is given by (4.13), the discrete speed corrected $v_{di (cor)}$ is given by (4.16) and the numerical speed v_{num} is computed numerically.

4.2.3 Parameters and initial conditions for numerical simulations

All the numerical simulations of this article are performed with the same set of parameters, summarised in Table 4.2. In [129], we showed that the drive always invades under these conditions, and we identified threshold values for s delimiting different regimes. For $0 < s \lesssim 0.938$, invasion is monostable, whereas for, $0.938 \lesssim s < 1$ we observe a coexistence stable final state ($s_1 \approx 0.938$ defined by (30) in [129]). For $0.091 \lesssim s < 1$, we have eradication of the population at the end ($\frac{r}{r+1} \approx 0.091$ with $r = 0.1$, result based on (105) in [129]). We focus on this interval in the following.

Parameter	Range value	Value	Description
r	$(0, +\infty)$	0.1	Intrinsic growth rate
c	$[0, 1]$	0.9	Conversion rate
s	$(0, 1)$	0.3 or 0.7	Fitness cost of drive homozygotes
h	$[0, 1]$	0.4	Drive dominance
m	$[0, 1]$	0.2	Migration rate
K	$(0, +\infty)$	from 10^3 to 10^8	Local carrying capacity
dx	$(0, 1]$	1	Spatial step between two sites
dt	$(0, 1]$	0.1	Temporal step
T	$(0, +\infty)$	1000	Simulation time

Table 4.2: Model and numerical parameters.

Initial conditions for our numerical simulations are illustrated in Figure 4.5. The left half of the domain is full of drive ($n_D = 1$), and the right half is full of wild-type ($n_W = 1$). Reference to right or left positions are made in the context of this initial state.

The code is available on GitHub (https://github.com/LenaKlay/gd_project_1, in the folder: `stochastic`). We ran our simulations in Python 3.6, with the Spyder environment.

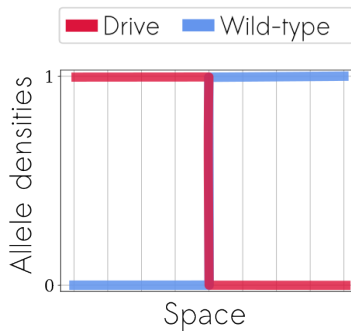


Figure 4.5: Initial conditions for numerical simulations in one spatial dimension.

4.2.4 Setting of the problem

In Figure 4.6, we plot the drive and the wild-type densities at the back of the wave (in red and blue) in log scale for different times, as well as the exponential approximations for each (in black). For a large drive fitness cost ($s = 0.7$) and a small carrying capacity ($K = 10^5$), we observe one wild-type recolonising event within this window of time.

From equations (4.15b) and (4.15c), we know that the slopes at the back of the wave increase with s . The fitter the drive is, the longer it will stay in an empty environment after the wave has passed: we expect that wild-type recolonisation would then be less likely. The carrying capacity K does not influence the value of the slope, however we observe numerically that it brings the waves closer together when it decreases (Figure 4.6): we explain analytically why in Section 4.3.1. Thanks to the stochastic fluctuations, this might result in more wild-type recolonising events, as the last wild-type individual is more likely to be the left of the last drive individual at some point.

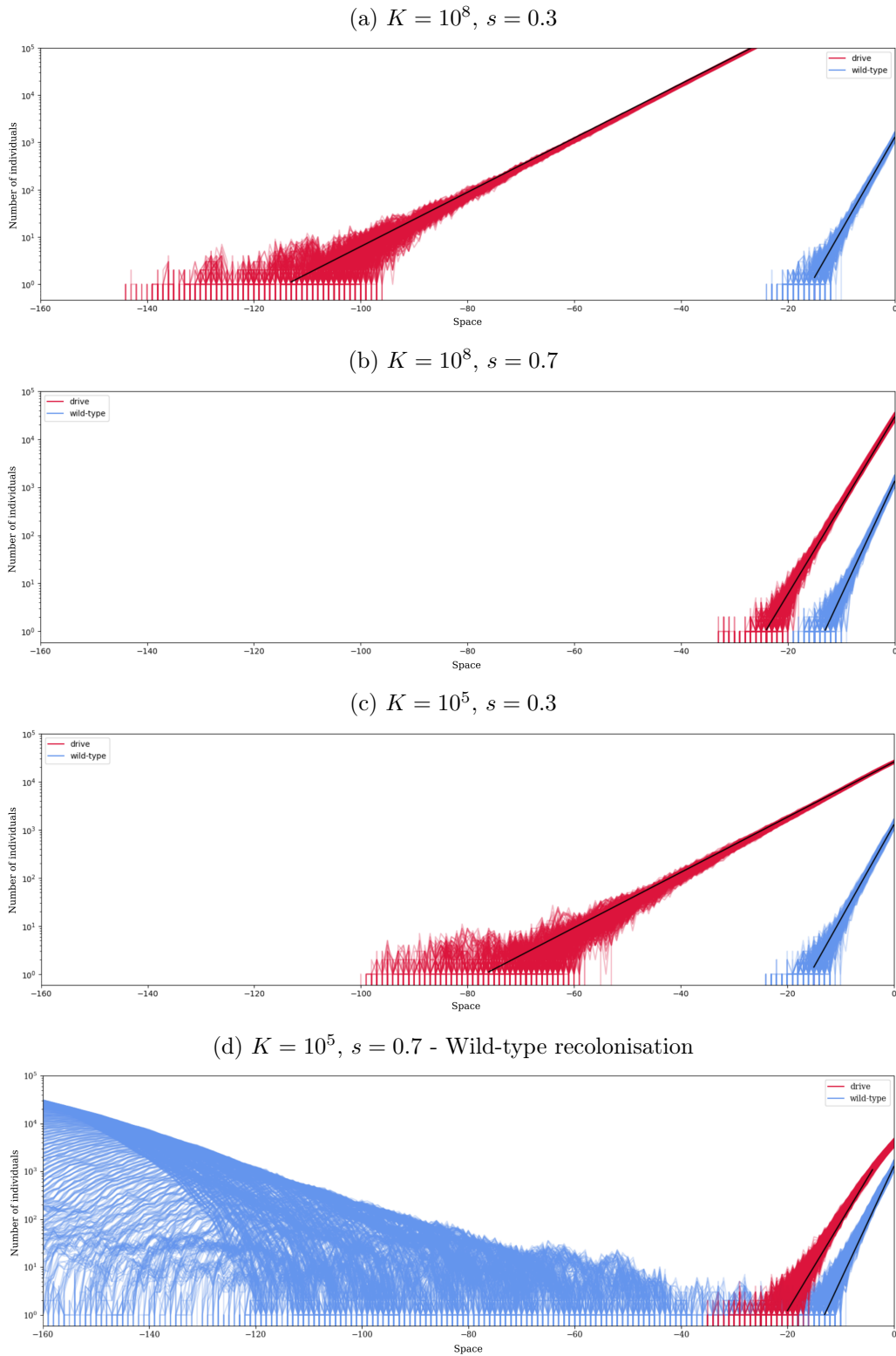


Figure 4.6: Drive densities (in red) and wild-type densities (in blue) at the back of the wave, in log scale. We superimpose multiple realisations corresponding to 2000 different times of the simulation, to observe the stochastic variations. Thus, we add the exponential approximations from Section 4.A.2 for each wave (in black). Both slopes depend on s , here more weakly, for drive. We observe a wild-type recolonising event for $s = 0.7$ and $K = 10^5$.

In this work, we ask the following question: what are the conditions ensuring that the last wild-type individual is surrounded by an important number \mathcal{N} of drive individuals. We believe this is a strong indicator for the prevention of wild-type recolonisation in the stochastic simulations.

We denote ℓ the distance between: i) the last position with more than \mathcal{N} drive individuals and ii) the last position with more than \mathcal{N} wild-type individuals, at the back of the wave. We also denote $L_1^{\mathcal{N}}$ the distances between: i) the position of the last wild-type individual and ii) the last position with more than \mathcal{N} wild-type individuals, at the back of the wave. These distances are illustrated in Figure 4.7. $\ell > L_1^{\mathcal{N}}$ indicates that the last wild-type individual is surrounded by more than \mathcal{N} drive individuals. In the following, we try to characterise the absence of wild-type recolonising events in a realisable time window by determining the distributions of ℓ and $L_1^{\mathcal{N}}$.

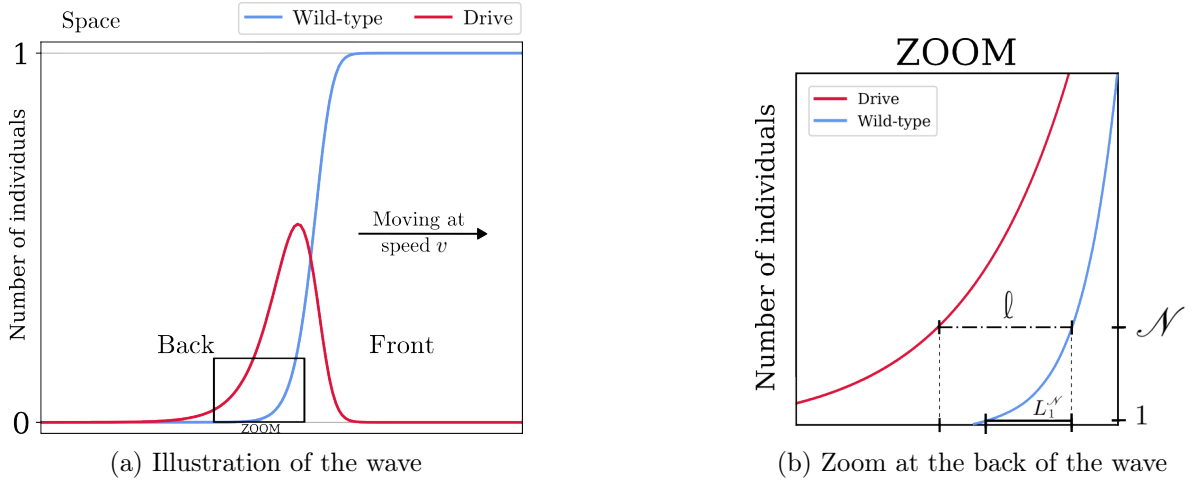


Figure 4.7: Schematic illustration of an eradication drive wave in arithmetic scale. If \mathcal{N} is large enough and ℓ is strictly larger than $L_1^{\mathcal{N}}$, then wild-type recolonisation is very unlikely because the last wild-type individual is surrounded by a large number of drive individuals.

In Figure 4.6, we observe moderate stochastic fluctuations around the level line $\mathcal{N} = 100$, so that ℓ is accurately determined by the exponential decay. This observation is confirmed by Figure 4.8, in which we plot the relative position of the last spatial site containing more than 100 drive (resp. wild-type) individuals at the back of the wave in red (resp. dark blue). We decide to choose the reference value $\mathcal{N} = 100$ in the rest of the paper. This is quite an arbitrary choice at this point, and much of our analysis could be done for other reference values.

On the contrary, the position of the last wild-type individual is highly stochastic, as shown in Figure 4.8. The approach exposed in Section 4.2.1.3 gives us a rough approximation of its mean position (i.e. of the mean value of $L_1^{\mathcal{N}}$), but we are interested in characterising the "worst case", the further away the wild-type individual can be found at the back of the wave.

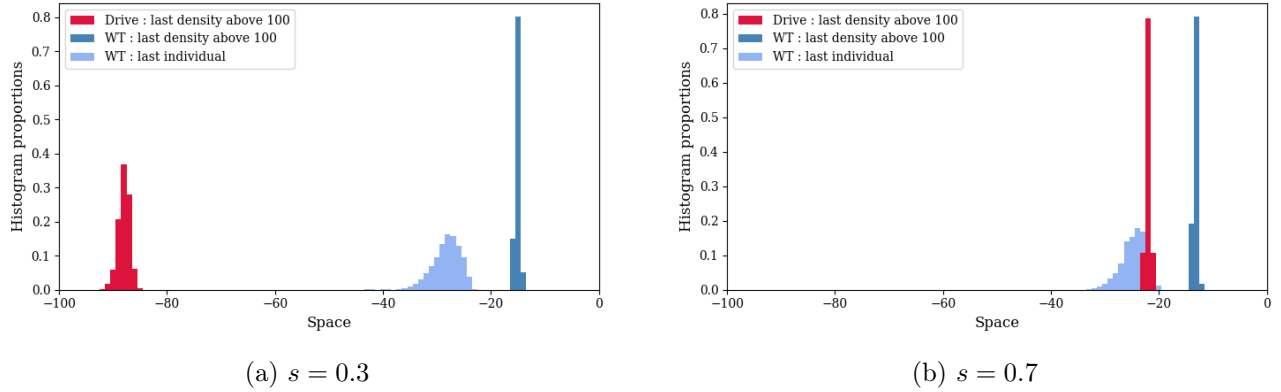


Figure 4.8: Relative positions of the last spatial site with more than 100 drive individuals at the back of the wave (in red), the last spatial site with more than 100 wild-type individuals at the back of the wave (in dark blue) and the last wild-type individual at the back of the wave (in light blue). We arbitrarily set the last spatial site with more than 10^5 wild-type individuals at the position $z = 0$. The statistical distribution of the last wild-type individuals is highly stochastic in contrast with the two other involving more than 100 individuals. We also observe that this distribution is not symmetric, indicating rare events of last wild-type individuals being far away at the back of the wave.

4.3 Results

4.3.1 Determining the almost deterministic distance between the drive and wild-type wave at a level line of 100 individuals

The distance ℓ between the drive and wild-type wave at a level line of 100 individuals (at the back of the wave) is almost deterministic and well approximated by the deterministic model. In Figure 4.9, we compute this value for different values of s and K . As expected, this distance increases for large K and small s . Note that for $s \leq 0.2$, the smallest reference density (with a position almost deterministic) might be higher than 100, as the stochastic fluctuations affect larger densities.

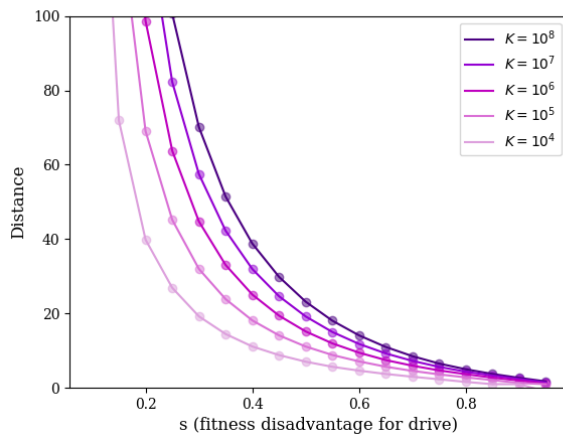


Figure 4.9: Distances between the reference densities in the drive and the wild-type waves, computed through deterministic simulations as in [129]. This distance increases when K increases and s decreases.

For a given s , we define $K_i = 10^i$ with $i \in \{4, 5, 6, 7\}$, and ℓ_i the corresponding ℓ distance. In Figure 4.9, we observe that $\ell_{i+1} - \ell_i$ is constant for all $i \in \{4, 5, 6, 7\}$: we show this result analytically below.

If we consider the relative density instead of the number of individuals, the local carrying capacity becomes one, consequently the shape of the wave is the same for all values of i and the level line is given by $\mathcal{N}_i = 100/K_i$. We have the following equality:

$$\ell_{i+1} = \ell_i + (z_{D,i} - z_{D,i+1}) - (z_{W,i} - z_{W,i+1}) \quad (4.17)$$

From equation (4.10), we have:

$$\begin{aligned} \ell_{i+1} &= \ell_i + \left(\log(\mathcal{N}_i) - \log(\mathcal{N}_{i+1}) \right) \left(\frac{1}{\lambda_{D, di}^{back}} - \frac{1}{\lambda_{W, di}^{back}} \right) \\ &= \ell_i + \left(\log\left(\frac{K^{i+1}}{K^i}\right) \right) \left(\frac{1}{\lambda_{D, di}^{back}} - \frac{1}{\lambda_{W, di}^{back}} \right) \\ &= \ell_i + \log(10) \left(\frac{1}{\lambda_{D, di}^{back}} - \frac{1}{\lambda_{W, di}^{back}} \right) \end{aligned} \quad (4.18)$$

From equations (4.15b) and (4.15c), we know that the quantity $\log(10) \left(\frac{1}{\lambda_{D, di}^{back}} - \frac{1}{\lambda_{W, di}^{back}} \right)$ is constant regarding K .

4.3.2 Characterising the stochastic fluctuations of the last wild-type individual

The position of the last wild-type individual is highly stochastic. Consequently, the exponential approach exposed in Section 4.2.1.3 is not accurate to determine the further away the last wild-type individual can be found at the back of the wave. In order to simplify the problem, we draw a parallel with Galton-Watson processes, which model a single isolated population in time. This process is based on the assumption that individuals give birth and die independently of each other, and follow the same distribution for each of these events.

There exists a massive literature on spatial branching stochastic processes (of which the Galton-Watson spatial process is part) in which most of the results are obtained using a generating function approach [22]. If a Galton-Watson process gives a good approximation of the fluctuations at the back of the wave, we could use it to approximate the distance between the last wild-type individual and the last wild-type reference density. It appears however that Galton-Watson processes with migration are very sensitive to spatial parameters: as we work on a limited spatial domain, we focus on extinction time instead of distance to the last individual, less affected by this restriction.

We need to verify two assumptions before heuristically reducing our problem to a spatial Galton-Watson process. First, that the question about the last wild-type individual distance can be transformed into an extinction time problem. Second, that the dynamics of the wild-type genotype at the back of the wave can be approximated by the dynamics of an isolated population.

In the following, the positions of the last wild-type individual and the extinction times are all com-

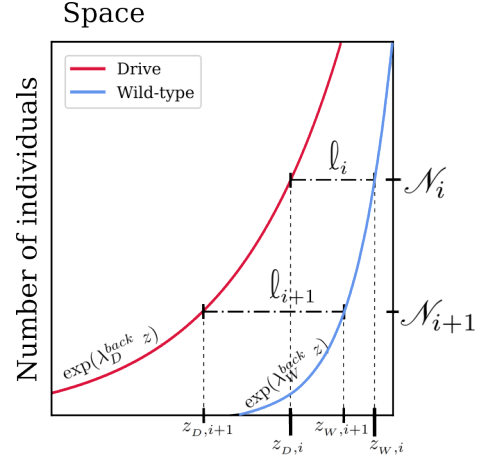


Figure 4.10: Schematic illustration at the back of the waves with relative densities.

puted conditionally to the absence of wild-type recolonisation in the global simulation: our reasoning breaks down if this condition is not verified.

4.3.2.1 Converting distances to extinction times

As the wave is moving at a constant speed v_{num} , we wonder if the distance L_1^{100} follows the same distribution as the time that the last spatial site with more than 100 wild-type individuals at the back of the wave takes to go extinct, multiplied by the speed of the wave (v_{num}). In Figure 4.11, we plot the two distributions for $s = 0.3$ and $s = 0.7$. In both cases, they appear very close from each other, which validates the first assumption.

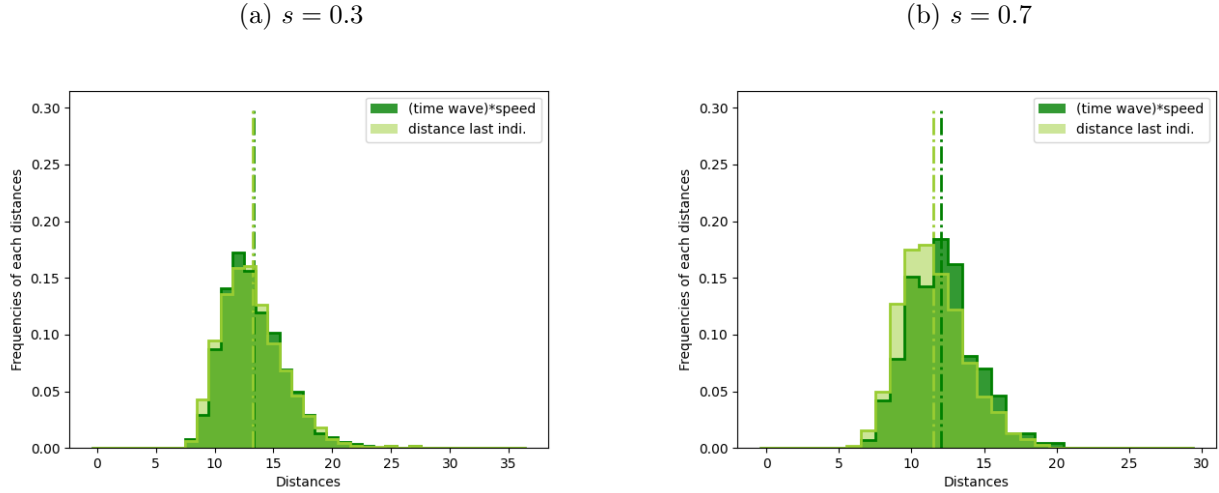


Figure 4.11: Comparison between the distribution of distance L_1^{100} (in light green) and the distribution of the time that the last spatial site with more than 100 wild-type individuals at the back of the wave takes to go extinction, multiplied by the speed of the wave (in dark green). These distributions are calculated over 500 different times. The two distributions are very close for $s = 0.3$, and for $s = 0.7$. We find the same asymmetry as in Figure 4.8 (in the statistical distribution of the relative position of the last wild-type individual) but reversed, asymmetry also preserved by the dark green distribution here.

4.3.2.2 From a global dynamics to an isolated population

To study a spatial Galton-Watson process, we need to approximate the wild-type dynamics at the back of the wave by a single isolated population. At the back of the wave and in the absence of wild-type recolonisation, we assume that:

$$\frac{n_D^{t,x}}{n^{t,x}} \approx 1, \quad \frac{n_W^{t,x}}{n^{t,x}} \approx 0 \quad \text{and} \quad n^{t,x} \approx 0. \quad (4.19)$$

We know from (4.3) that a wild-type allele at the back of the wave lead in average to the production of :

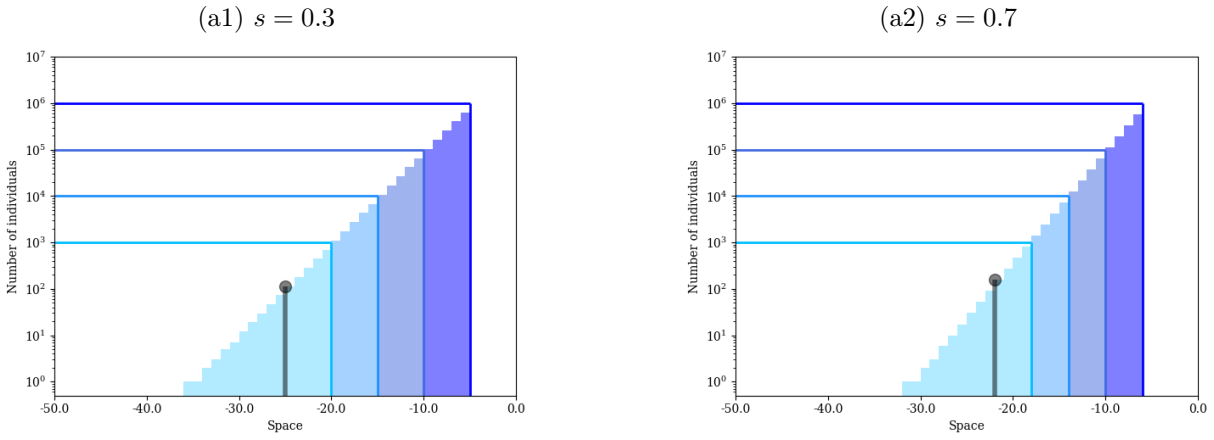
$$g_W(n_D^{t,x}, n_W^{t,x}) = (r(1 - n^{t,x}) + 1) \left[\frac{n_W^{t,x}}{n^{t,x}} + (1 - sh)(1 - c) \frac{n_D^{t,x}}{n^{t,x}} \right] \approx (r + 1)(1 - sh)(1 - c) \quad (4.20)$$

alleles (in offspring) during one unit of time, and disappears on average at rate 1. Thanks to this approximation, the dynamics does not depend on n_D any more, and we can simulate a single isolated wild-type population. As before, at each time step, a wild-type allele migrates to the adjacent site on the right with probability $\frac{m}{2}$ and to the left with probability $\frac{m}{2}$.

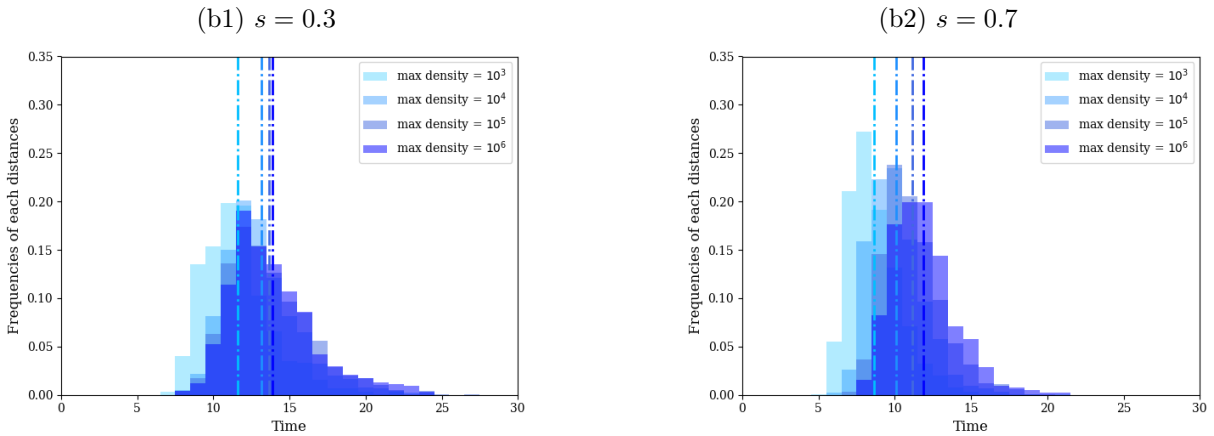
To be consistent in the comparison between the global dynamics and the single population approximation, we initiate the simulation with the exponential function $\exp(z \lambda_{W, di}^{back})$ given in equation (4.14c). We record the extinction time of the last spatial site with more than 100 individuals in the initial condition and run the simulation 500 times to get the statistical distribution of this extinction time.

First, we observe that the migration from dense areas to less dense areas significantly increases the extinction time of the site, initially just above 100 individuals. Indeed, we crop the initial condition on the right so that the maximum number of individuals in one spatial site gradually decreases to 10^3 (Figure 4.12a) and observe that the mean of the extinction distribution consequently decreases from 2 to 3 units of time (Figure 4.12b). Thus, it is very important to consider a large exponential initial condition to fully capture the wild-type dynamics at the end of the wave.

In Figure 4.13, we superimpose the distribution (with the maximum number of individuals per site being 10^6), with the extinction times recorded in the simulation with the drive individuals. The histograms are very close for $s = 0.3$ and $s = 0.7$: our second assumption is verified.



(a) Initial conditions considered for the isolated wild-type population. The black line with a circle at the top represents the last spatial site with more than 100 individuals in the initial condition. We record the extinction time of this site, and run the simulation 500 times to obtain the distributions illustrated in Figure 4.12b.



(b) Extinction times distribution (multiplied by the same coefficient: the speed value of the wave obtained from the global simulation), for initial conditions given in Figure 4.12a. We focus on the last spatial site with more than 100 individuals in the initial condition, and record the time at which it goes extinct (definitively). The dashed lines are the means values of each histogram.

Figure 4.12: Initial conditions and extinction times for the isolated population.

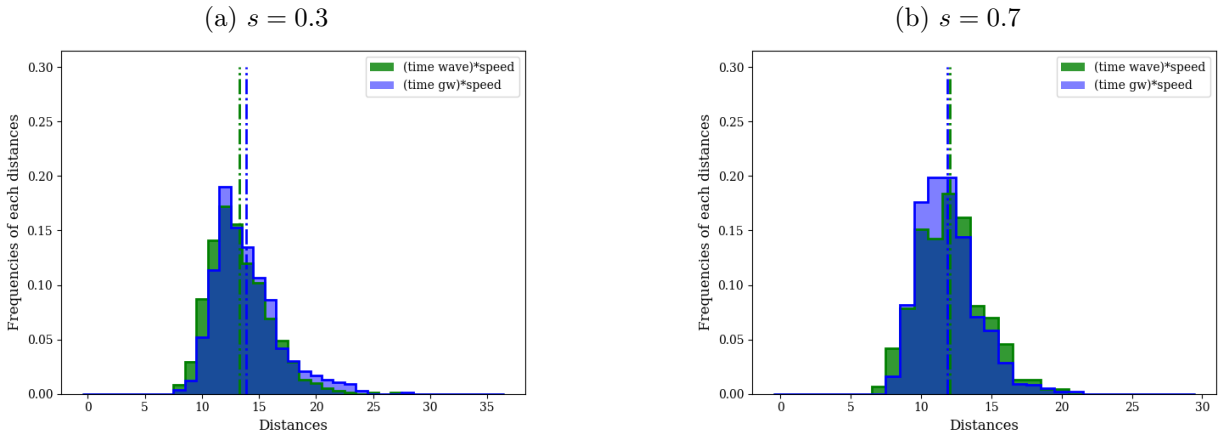


Figure 4.13: Comparison of the extinction time distributions (multiplied by the same coefficient: the speed value of the wave obtained from the global simulation). In dark blue, the time at which the last spatial site with initial more than 100 individuals goes extinct in the isolated population, with a maximum of 10^6 individual per spatial site in the initial condition (isolated population). In dark green, the time at which the last spatial site with more than 100 wild-type individuals goes extinct in the simulation with drive individuals (global population). The distributions fit very well and the asymmetry is again preserved.

4.3.3 Numerical conclusions

In Figure 4.14, we superimpose the distribution of L_1^{100} at the back of the wave for $s = 0.3$ and $s = 0.7$. The probability to observe one wild-type recolonising event in a realisable time window is extremely low when $s = 0.3$: the further away a wild-type individual might be is a spatial site with more than 10^4 drive individuals. However, when $s = 0.7$, this very last position (end of the light green histogram) corresponds to a number of drive individuals between 0 and 5: there is a significant chance of wild-type recolonisation.

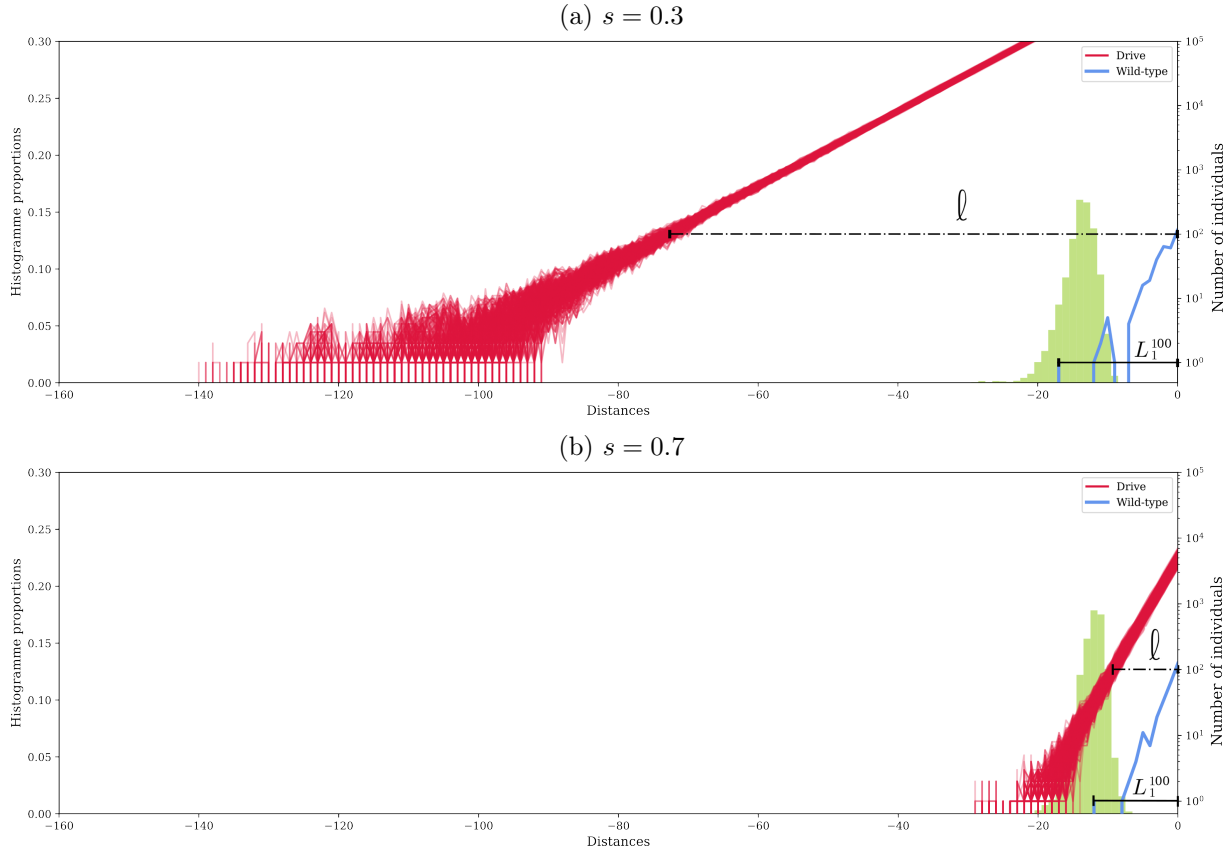


Figure 4.14: Superimposition of the distribution of the distance L_1^{100} (in light green), on the wild-type (in blue) and drive (in red) wave, when $K = 10^8$ in log scale. Several drive waves are plotted, whereas only one wild-type wave appears in the graph for clarity. The last individual of this wild-type wave determine the distance L_1^{100} : it is one realisation of the light green histogram. When $s = 0.3$, wild-type recolonisation is very unlikely as the furthest a wild-type individual might be is a spatial site containing more than 10^4 drive individuals. However, when $s = 0.7$, wild-type recolonisation dynamics might appear as the histogram mixes with the end of the drive wave: a wild-type individual could possibly be beyond the drive last individual at one point and recolonise the empty area. Based on Figure 4.15, the probability to observe wild-type recolonisation over 1000 units of time is close to 0 for $s = 0.3$, and approximately 0.63 when $s = 0.7$.

In Figure 4.15, we numerically approximate the probability to observe wild-type recolonisation within 1000 units of time, for different values of s the drive fitness cost and K the local carrying capacity. Each point of the graph is determined by the proportion of replicates where we observe

wild-type recolonisation, over 100 replicates. This probability increases with s and decreases with K . Noticeably for a given local carrying capacity K , the transition between very low ($< 10\%$) and very high ($> 90\%$) chances of wild-type recolonisation within 1000 units of time when the fitness cost s varies, is relatively restricted: these two extreme conditions can be reached at two different s values within a range of 0.2.

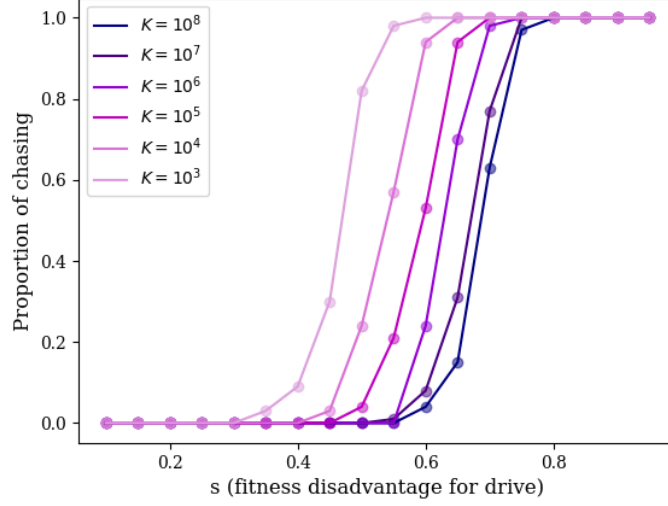


Figure 4.15: Proportion of wild-type recolonisation function of s , the fitness disadvantage for drive, and K , the carrying capacity on one spatial step within 1000 units of time. For each point, we ran the simulation 100 times and computed the proportion of simulations in which there was wild-type recolonisation.

4.3.4 Relation between drive intrinsic fitness, diffusion and risk of wild-type recolonisation.

We investigate analytically how the drive intrinsic fitness f_D and the diffusion σ influence the exponent $\lambda_{D, co}^{back}$ of the exponential drive profile at the back of the wave. $\lambda_{D, co}^{back}$ is a good indicator of the risk of wild-type reemergence: the smaller $\lambda_{D, co}^{back}$ is, the less wild-type recolonisation we observe as drive individuals are present further away at the back.

Drive intrinsic fitness

The drive intrinsic fitness (f_D) is the overall ability of drive alleles to increase in frequency when the drive proportion is close to zero. Its value is given by the first line of system (4.4):

$$f_D = (1 - sh)(1 + c) - 1 \quad \iff \quad s = \frac{c - f_D}{h(1 + c)}. \quad (4.21)$$

From equation (4.9b), we can rewrite $\lambda_{D, co}^{back}$ using the variable f_D the drive intrinsic fitness, instead

of s the drive fitness cost :

$$\begin{aligned}
\lambda_{D, co}^{back} &= \sqrt{\mathfrak{f}_D + 1 - (r+1)\left(1 - \frac{c - \mathfrak{f}_D}{h(1+c)}\right)} - \sqrt{\mathfrak{f}_D}, \\
&= \sqrt{\left(1 - \frac{r+1}{h(1+c)}\right)\mathfrak{f}_D + 1 - (r+1)\left(1 - \frac{c}{h(c+1)}\right)} - \sqrt{\mathfrak{f}_D}, \\
&= \sqrt{\mathcal{A}\mathfrak{f}_D + \mathcal{B}} - \sqrt{\mathfrak{f}_D}.
\end{aligned} \tag{4.22}$$

Since the drive wave is increasing at the back and decreasing at the front:

$$\mathcal{A}\mathfrak{f}_D + \mathcal{B} > 0 \quad , \quad \mathfrak{f}_D > 0 \quad \text{and} \quad \mathcal{A}\mathfrak{f}_D + \mathcal{B} > \mathfrak{f}_D. \tag{4.23}$$

Note that:

$$\mathcal{A} = 1 - \frac{r+1}{h(1+c)} < 1. \tag{4.24}$$

Since we study a drive eradication wave, the following inequality is verified (see [129] for details):

$$r > \frac{s}{1-s} \iff \frac{r}{r+1} < s = \frac{c - \mathfrak{f}_D}{h(1+c)} < \frac{c}{h(1+c)} \iff r < (r+1) \frac{c}{h(1+c)}, \tag{4.25}$$

from which we deduce:

$$\mathcal{B} = 1 - (r+1)\left(1 - \frac{c}{h(c+1)}\right) = (r+1)\frac{c}{h(c+1)} - r > 0. \tag{4.26}$$

We study the following derivative:

$$\partial_{\mathfrak{f}_D} \left(\lambda_{D, co}^{back} \right) = \frac{\mathcal{A}}{2\sqrt{\mathcal{A}\mathfrak{f}_D + \mathcal{B}}} - \frac{1}{2\sqrt{\mathfrak{f}_D}}, \tag{4.27}$$

which is strictly positive for $\mathcal{A} \leq 0$, but also for $\mathcal{A} > 0$ as $\mathfrak{f}_D > 0$, $\mathcal{B} > 0$ and $\mathcal{A} \in (0, 1)$. $\lambda_{D, co}^{back}$ always decreases when \mathfrak{f}_D increases: in other words, the better the drive intrinsic fitness, the wider the back of drive wave lowering the risk of wild-type recolonization.

Diffusion

In the paper, we always considered a constant diffusion rate of 1 for all genotypes. However if we now consider the diffusion rate σ as a parameter of the model, the exponents $\lambda_{D, co}^{front}$, $\lambda_{D, co}^{back}$ and $\lambda_{W, co}^{back}$ are all multiplied by $\sqrt{\sigma}$. As a consequence, the bigger the diffusion σ , the larger the distance between the last wild-type individual and the last drive individual lowering the risk of wild-type recolonisation.

4.4 Discussion

Given both the promise and risk of eradication drives, their dynamics are worthy of careful examination. In this work, we introduce a theoretical framework in order to characterise the conditions leading to

the absence of wild-type recolonisation in a pulled drive eradication wave and in a realisable time window. More precisely, we consider that if the last wild-type individual is surrounded by a large number of drive individuals (set to $\mathcal{N} = 100$ individuals in this work) then wild-type recolonisation is very unlikely. To attest if this is the case, we determine the distance between the last wild-type individual and the last spatial site with more than 100 drive individuals, at the back of the wave. This value is not straight forward to find, so we proceed in two steps.

First, we determine the distance ℓ between the drive and the wild-type wave at the level line corresponding to 100 individuals. This distance is almost deterministic, as we choose $\mathcal{N} = 100$ large enough for stochastic fluctuations to be negligible. Thus, we easily deduce this distance from a deterministic simulation and observe that it increases with the carrying capacity K , and decreases with the drive fitness cost s . Analytically, we find that if the carrying capacity K is multiplied by a factor 10, the distance ℓ increases by $(\frac{1}{\lambda_{D, di}^{back}} - \frac{1}{\lambda_{W, di}^{back}}) \log(10)$, with $\lambda_{D, di}^{back}$, resp. $\lambda_{W, di}^{back}$, the exponent of the exponential profile at the back of the drive, resp. wild-type, wave.

Second, we determine the distance L_1^{100} between the position of the last wild-type individual and the last position with more than 100 wild-type individuals. This distance highly depends on the stochastic fluctuations of the last individual, and its distribution is difficult to characterise. To simplify the problem, we consider an isolated wild-type population, and we reduce heuristically the distance distribution to an extinction time distribution of a spatial Galton-Watson process, with the appropriate initialisation. Multiplying this time by the speed of the wave, we get a good approximation of L_1^{100} . We leave open the quality of this approximation in mathematical terms, as well as the analysis of the distribution of the extinction time.

Combining the two distances ℓ and L_1^{100} allow us to determine precisely the conditions under which wild-type recolonisation is highly unlikely. Numerically, in one spatial dimension, we show that the larger the carrying capacity is, the less recolonisation events we observe. The drive fitness also influences the wild-type recolonisation: the fitter the drive, the fewer risk of wild-type recolonisation in agreement with the results obtained in [46]. Noticeably in our study, the transition between very low ($< 10\%$) and very high ($> 90\%$) chances of wild-type recolonisation when the drive fitness cost varies is relatively restricted: for s the drive fitness cost varying between 0 (no drive fitness cost) and 1 (no possible survival for drive individuals), this happens within a range of 0.2.

If we focus on the drive intrinsic fitness $(1 - sh)(1 + c) - 1$, in other words the overall ability of drive to increase in frequency when the drive proportion is close to zero, we agree on the observations made by [177]: the fitter the drive, the sharper the drive wave at the front (4.9a), and the faster the speed (4.7) as initially showed in [87]. In contrast at the back of the wave, the fitter the drive, the wider the drive wave, reducing the risk of wild-type recolonisation (see Section 4.3.4). We also demonstrate analytically that if we consider the same diffusion rate for all genotypes, high rates of dispersal reduce the chance of wild-type recolonisation. This result is in agreement with the numerical observations made in [177, 46].

This study focus on a one-dimensional space, however it might be interesting to extend these results to a two-dimensional space, where wild-type recolonisation could be caused by stochastic local eliminations of the drive genotype. If we consider a single release at the centre of the domain, we would expect the drive wave to gradually eliminate the wild-type individuals through a growing circle. However if the drive density is low enough, "holes" in the drive wave could result in migration routes for wild-type individuals, allowing the colonisation of the previously cleared centre. This statement is again closely linked with population densities and requires to study the influence of various carrying capacities.

The models we use in this study are generalist: they could be applied to different species and gene drive constructs. They provide general conclusions but also come with necessary simplifications that we are aware of. For instance, we assume a uniform landscape with random movement, which is extremely rare in the wild. Realistic migration patterns over both small and large scales need to be taken into account to obtain better predictions. In mosquito populations, the gene drive propagation can be accelerated by long distance migrations: mosquitoes can benefit from fast air currents (transporting them for hundreds of kilometres in a few hours) [49, 115, 116] or human-based modes of travel such as cars [79] or planes [80]. This could make it easier for wild-type individuals to permeate the drive wave, and initiate the recolonisation of empty areas.

A variety of other ecological parameters must also be taken into account before any field release. A previous study showed that the presence of a competing species or predator in the ecosystem could make eradication substantially easier than anticipated and prevent wild-type recolonisation [143]. The life cycle of the species, the survival characteristics, the mating system, the ecological differences between males and females, the seasonal population fluctuations are all biological characteristics that might also influence our general conclusions.

Finally, the type of gene drive constructs might also bias our prediction depending on how the fitness disadvantage impacts the individual (in our model, it impacts the birth rate) and the possible emerging resistances that might alter the gene conversion ability of the drive (we consider a constant conversion rate c) [18, 106, 181, 187].

Appendices

4.A Speed and exponential approximations of the wave

4.A.1 Continuous model

Because we know the wave is pulled, we can deduce from system (4.6) the speed v_{co} and the exponential approximations at the back and at the front of the wave.

4.A.1.1 At the front of the wave

At the front of the wave, we assume that:

$$\frac{N_D}{N} \approx 0, \quad \frac{N_W}{N} \approx 1 \quad \text{and} \quad N \approx 1. \quad (4.28)$$

Consequently, from system (4.6), we know the solution of following equation (4.29) is an approximation of N_D at the front of the wave:

$$0 = N_D'' + v_{co}N_D' + N_D \left[(1 - sh)(1 + c) - 1 \right]. \quad (4.29)$$

In case of a pulled wave, the speed v_{co} is given by the minimal speed v_{co}^{lin} of the problem linearised at low drive density at the front of the wave. We approximate the decreasing drive section at the front of the wave by an exponential function of the form $N_D(z) \approx e^{\lambda_{D, co}^{front} z}$, and deduce from (4.29):

$$0 = (\lambda_{D, co}^{front})^2 + v_{co} \lambda_{D, co}^{front} + \left[(1 - sh)(1 + c) - 1 \right], \quad (4.30)$$

with the determinant $\Delta = (v_{co})^2 - 4 \left[(1 - sh)(1 + c) - 1 \right]$. The minimal speed v_{co}^{lin} is given by $\Delta = 0$, thus:

$$v_{co}^{lin} = 2\sqrt{(1 - sh)(1 + c) - 1}, \quad (4.31)$$

and the corresponding exponent is given by:

$$\lambda_{D, co}^{front} = -\frac{v_{co}^{lin}}{2} = -\sqrt{(1 - sh)(1 + c) - 1}. \quad (4.32)$$

The quantity $(1 - sh)(1 + c) - 1$ is always strictly positive for a pulled wave in case of drive invasion [129].

4.A.1.2 At the back of the wave

Note that if the back of the wave can be defined as $z \rightarrow -\infty$ in the continuous model, it does not make sense any more in the discrete model. After a certain spatial site, there is no more individual on the left due to eradication: the exponential approximations do not hold any more. In the discrete stochastic model, the back of the wave corresponds to the non-empty spatial sites in the early increasing section of the wave.

At the back of the wave, we assume that:

$$\frac{N_D}{N} \approx 1, \quad \frac{N_W}{N} \approx 0 \quad \text{and} \quad N \approx 0. \quad (4.33)$$

By definition of a traveling wave, the speed at the front and at the back of the wave is equal.

Drive increasing section

Using (4.33) in system (4.6), we know the solution of the following equation (4.34) is an approximation of N_D at the back of the wave:

$$0 = N_D'' + v_{co}^{lin} N_D' + N_D \left[(r+1)(1-s) - 1 \right]. \quad (4.34)$$

We approximate the increasing drive section at the back of the wave by an exponential function of the form $N_D(z) \approx e^{\lambda_{D, co}^{back} z}$ and deduce from (4.34):

$$0 = (\lambda_{D, co}^{back})^2 + v_{co}^{lin} \lambda_{D, co}^{back} + \left[(r+1)(1-s) - 1 \right] r. \quad (4.35)$$

The solutions are given by:

$$\begin{aligned} \{\lambda_{D, co}^{back}\}_{+,-} &= \frac{1}{2} \left(-v_{co}^{lin} \pm \sqrt{(v_{co}^{lin})^2 - 4 \left[(r+1)(1-s) - 1 \right]} \right), \\ &= \frac{1}{2} \left(-2\sqrt{(1-sh)(1+c) - 1} \pm \sqrt{4 \left[(1-sh)(1+c) - 1 \right] - 4 \left[(r+1)(1-s) - 1 \right]} \right), \\ &= -\sqrt{(1-sh)(1+c) - 1} \pm \sqrt{(1-sh)(1+c) - 1 + 1 - (r+1)(1-s)}. \end{aligned} \quad (4.36)$$

Since we study an eradication drive, we have $r < \frac{s}{1-s} \iff 1 - (r+1)(1-s) > 0$ [129]. Therefore, the two solutions of (4.36) are of opposite sign. At the back of the wave, the drive density tends to zero when $z \rightarrow -\infty$ therefore we only conserve the positive solution of (4.36).

$$\lambda_{D, co}^{back} = -\sqrt{(1-sh)(1+c) - 1} + \sqrt{(1-sh)(1+c) - (r+1)(1-s)}. \quad (4.37)$$

Wild-type increasing section

Using (4.33) in system (4.6), we know the solution of the following equation (4.38) is an approximation of N_W at the back of the wave:

$$0 = N''_w + vN'_w + N_w \left[(r+1)(1-sh)(1-c) - 1 \right]. \quad (4.38)$$

We approximate the increasing wild-type section at the back of the wave by an exponential function of the form $N_w(z) \approx e^{\lambda_{W,co}^{back} z}$ and deduce from (4.38):

$$0 = (\lambda_{W,co}^{back})^2 + v_{co}^{lin} \lambda_{W,co}^{back} + \left[(r+1)(1-sh)(1-c) - 1 \right]. \quad (4.39)$$

The solutions are given by:

$$\begin{aligned} & \frac{1}{2} \left(-v_{co}^{lin} \pm \sqrt{(v_{co}^{lin})^2 - 4 \left[(r+1)(1-sh)(1-c) - 1 \right]} \right), \\ & = \frac{1}{2} \left(-2\sqrt{(1-sh)(1+c) - 1} \pm \sqrt{4 \left[(1-sh)(1+c) - 1 \right] - 4 \left[(r+1)(1-sh)(1-c) - 1 \right]} \right), \quad (4.40) \\ & = -\sqrt{(1-sh)(1+c) - 1} \pm \sqrt{(1-sh)(1+c) - 1 + 1 - (r+1)(1-sh)(1-c)}. \end{aligned}$$

Since we study an eradication drive, we have $r < \frac{s}{1-s} \iff r+1 < \frac{1}{1-s}$ [129] and therefore :

$$(r+1)(1-sh)(1-c) < \frac{1-sh}{1-s}(1-c) < 1-c < 1 \quad (4.41)$$

The two solutions of (4.40) are of opposite sign. At the back of the wave, the wild-type density tends to zero when $z \rightarrow -\infty$: we only conserve the positive solution of (4.40).

$$\lambda_{W,co}^{back} = -\sqrt{(1-sh)(1+c) - 1} + \sqrt{(1-sh)(1+c) - (r+1)(1-sh)(1-c)} \quad (4.42)$$

4.A.2 Discrete model

To determine the speed of the wave and the exponential approximations in the discrete stochastic model, we focus on the mean dynamics. For drive alleles, we have:

$$\begin{cases} n_D^{t+\frac{dt}{2},x} & = \left((g_D(n_D^{t,x}, n_W^{t,x}) - 1) dt + 1 \right) n_D^{t,x}, \\ n_D^{t+dt,x} & = (1-m) n_D^{t+\frac{dt}{2},x} + \frac{m}{2} (n_D^{t+\frac{dt}{2},x+dx} + n_D^{t+\frac{dt}{2},x-dx}), \end{cases} \quad (4.43)$$

and for wild-type alleles:

$$\begin{cases} n_W^{t+\frac{dt}{2},x} & = \left((g_W(n_D^{t,x}, n_W^{t,x}) - 1) dt + 1 \right) n_W^{t,x}, \\ n_W^{t+dt,x} & = (1-m) n_W^{t+\frac{dt}{2},x} + \frac{m}{2} (n_W^{t+\frac{dt}{2},x+dx} + n_W^{t+\frac{dt}{2},x-dx}). \end{cases} \quad (4.44)$$

where the first lines correspond to the birth and death dynamics, and the second lines correspond to the migration. Combining the two lines in each system, we obtain:

$$\left\{ \begin{array}{l} n_D^{t+dt,x} = \left((g_D(n_D^{t,x}, n_W^{t,x}) - 1) dt + 1 \right) (1 - m) n_D^{t,x} + \left((g_D(n_D^{t,x+dx}, n_W^{t,x+dx}) - 1) dt + 1 \right) \frac{m}{2} n_D^{t,x+dx} \\ \quad + \left((g_D(n_D^{t,x-dx}, n_W^{t,x-dx}) - 1) dt + 1 \right) \frac{m}{2} n_D^{t,x-dx}, \\ n_W^{t+dt,x} = \left((g_W(n_D^{t,x}, n_W^{t,x}) - 1) dt + 1 \right) (1 - m) n_W^{t,x} + \left((g_W(n_D^{t,x+dx}, n_W^{t,x+dx}) - 1) dt + 1 \right) \frac{m}{2} n_W^{t,x+dx} \\ \quad + \left((g_W(n_D^{t,x-dx}, n_W^{t,x-dx}) - 1) dt + 1 \right) \frac{m}{2} n_W^{t,x-dx}. \end{array} \right. \quad (4.45)$$

with:

$$g_D(n_D, n_W) = (r(1-n) + 1) \left[(1-s) \frac{n_D}{n} + (1-sh)(1+c) \frac{n_W}{n} \right], \quad (4.46)$$

$$g_W(n_D, n_W) = (r(1-n) + 1) \left[\frac{n_W}{n} + (1-sh)(1-c) \frac{n_D}{n} \right]. \quad (4.47)$$

The speed of the wave in the discrete model is denoted v_{di} .

4.A.2.1 At the front of the wave

At the front of the wave, we assume that:

$$\frac{n_D^{t,x}}{n^{t,x}} \approx 0, \quad \frac{n_W^{t,x}}{n^{t,x}} \approx 1 \quad \text{and} \quad n^{t,x} \approx 1. \quad (4.48)$$

These approximations are also true around x , at the spatial sites $x+dx$ and $x-dx$. Using (4.48) in the first line of system (4.45), we know the solution of the following equation (4.49) is an approximation of n_D at the front of the wave:

$$n_D^{t+dt,x} = \left(((1-sh)(1+c) - 1) dt + 1 \right) \left((1-m) n_D^{t,x} + \frac{m}{2} (n_D^{t,x+dx} + n_D^{t,x-dx}) \right). \quad (4.49)$$

In case of a pulled wave, the speed v_{di} is given by the minimal speed v_{di}^{lin} of the problem linearised at low drive density at the front of the wave. We approximate the decreasing drive section at the front of the wave by an exponential function of the form $n_D^{x-v_{di}t} = N_D^z \approx e^{\lambda_{D, di}^{front} z}$, and deduce from (4.49):

$$\begin{aligned} e^{-\lambda_{D, di}^{front} v_{di} dt} &= \left(((1-sh)(1+c) - 1) dt + 1 \right) \left(1 - m + m \frac{e^{(\lambda_{D, di}^{front} dx)} + e^{-(\lambda_{D, di}^{front} dx)}}{2} \right) \\ &= \left(((1-sh)(1+c) - 1) dt + 1 \right) \left(1 - m + m \cosh(\lambda_{D, di}^{front} dx) \right) \end{aligned} \quad (4.50)$$

The minimal speed v_{di}^{lin} of the problem linearised at low drive density is given by:

$$v_{di}^{lin} = \min_{(-\lambda_{D, di}^{front}) > 0} \left(\frac{\log \left(\left[((1-sh)(1+c) - 1) dt + 1 \right] \left[1 - m + m \cosh(\lambda_{D, di}^{front} dx) \right] \right)}{-\lambda_{D, di}^{front} dt} \right). \quad (4.51)$$

4.A.2.2 At the back of the wave

At the back of the wave, we have:

$$\frac{n_D^{t,x}}{n^{t,x}} \approx 1, \quad \frac{n_W^{t,x}}{n^{t,x}} \approx 0 \quad \text{and} \quad n^{t,x} \approx 0. \quad (4.52)$$

These approximations are also true around x , at the spatial sites $x + dx$ and $x - dx$. By definition of a traveling wave, the speed at the front and at the back of the wave is the same, in our case v_{di}^{lin} .

Drive increasing section

Using (4.52) in the first line of system (4.45), we know the solution of the following equation (4.53) is an approximation of n_D at the back of the wave:

$$n_D^{t+dt,x} = \left(((r+1)(1-s) - 1) dt + 1 \right) \left((1-m) n_D^{t,x} + \frac{m}{2} (n_D^{t,x+dx} + n_D^{t,x-dx}) \right). \quad (4.53)$$

We approximate the increasing drive section at the back of the wave by an exponential function of the form $n_D^{x-v_{di}^{lin}t} = N_D^z \approx e^{\lambda_{D, di}^{back} z}$, and deduce from (4.53):

$$e^{-\lambda_{D, di}^{back} v_{di}^{lin} dt} = \left(((r+1)(1-s) - 1) dt + 1 \right) \left(1 - m + \frac{m}{2} (e^{\lambda_{D, di}^{back} dx} + e^{-\lambda_{D, di}^{back} dx}) \right). \quad (4.54)$$

We solve equation (4.54) numerically to obtain the value of $\lambda_{D, di}^{back}$.

Wild-type increasing section

Using (4.52) in the second line of system (4.45), we know the solution of the following equation (4.55) is an approximation of n_W at the back of the wave:

$$n_W^{t+dt,x} = \left(((r+1)(1-sh)(1-c) - 1) dt + 1 \right) \left((1-m) n_W^{t,x} + \frac{m}{2} (n_W^{t,x+dx} + n_W^{t,x-dx}) \right). \quad (4.55)$$

We approximate the increasing drive section at the back of the wave by an exponential function of the form $n_W^{x-v_{di}^{lin}t} = N_W^z \approx e^{\lambda_{W, di}^{back} z}$ and deduce from (4.55):

$$e^{-\lambda_{W, di}^{back} v_{di}^{lin} dt} = \left(((r+1)(1-sh)(1-c) - 1) dt + 1 \right) \left(1 - m + \frac{m}{2} (e^{\lambda_{W, di}^{back} dx} + e^{-\lambda_{W, di}^{back} dx}) \right). \quad (4.56)$$

We solve equation (4.56) numerically to obtain the value of $\lambda_{W, di}^{back}$.

4.B Implementation of the stochastic model

```
1 ### Initialization
2
3 nD = np.zeros(nb_sites).astype(int); nD[:nb_sites//2] = K*dx # Drive density
4 nW = np.zeros(nb_sites).astype(int); nW[nb_sites//2:] = K*dx # Wild-type density
5 fD = np.zeros(nb_sites) # Drive fecundity
6 fW = np.zeros(nb_sites) # Wild-type fecundity
7
8
9 ### Evolution in time
10
11 for t in np.arange(0, T, dt):
12
13     ### Stop the simulation if the wave goes outside the window
14
15     if np.where(nD==max(nD))[0][0] > len(nD)-10 :
16         print("t =",t)
17         break
18
19     ### Birth and Death
20
21     # Index for empty and non empty sites
22     extinct_index = np.where(nD+nW==0)[0]
23     survive_index = np.delete(np.arange(nb_sites), extinct_index)
24     # Fecundity for non empty sites
25     sv_pop = nD[survive_index] + nW[survive_index]
26     sv_nD = nD[survive_index]; sv_nW = nW[survive_index]
27     fD[survive_index] = ( 1+r*(1-sv_pop/(K*dx)) ) * ( (1-s)*sv_nD + (1-s*h)*(1+c)*
sv_nW ) /sv_pop
28     fW[survive_index] = ( 1+r*(1-sv_pop/(K*dx)) ) * ( (1-c)*(1-s*h)*sv_nD + sv_nW ) /
sv_pop
29     # For empty sites, the fecundity is 0.
30     fD[extinct_index] = 0
31     fW[extinct_index] = 0
32     # Add births and subtract deaths (mortality = 1)
33     nD = nD + np.random.poisson(fD*nD*dt) - np.random.poisson(nD*dt)
34     nW = nW + np.random.poisson(fW*nW*dt) - np.random.poisson(nW*dt)
35     # Transform negative number of individuals into 0
36     nD[np.where(nD<0)[0]]=0
37     nW[np.where(nW<0)[0]]=0
38
39
40     ### Migration
41
42     # Number of migrants in each site
43     nD_mig = np.random.binomial(nD,m)
44     nW_mig = np.random.binomial(nW,m)
45     # Half migrate to the right, half to the left
46     nD_mig_left = np.random.binomial(nD_mig,0.5); nD_mig_right = nD_mig - nD_mig_left
47     nW_mig_left = np.random.binomial(nW_mig,0.5); nW_mig_right = nW_mig - nW_mig_left
48     # Subtract the migrants leaving
49     nD -= nD_mig
50     nW -= nW_mig
51     # ... except for those going outside the windows (they stay in the border site)
52     nD[0] += nD_mig_left[0]; nW[0] += nW_mig_left[0]
53     nD[-1] += nD_mig_right[-1]; nW[-1] += nW_mig_right[-1]
54     # Add the migrants in the neighbor sites
55     nD[1:] += nD_mig_right[:-1]; nW[1:] += nW_mig_right[:-1]
56     nD[:-1] += nD_mig_left[1:]; nW[:-1] += nW_mig_left[1:]
```

Chapter 5

Reducing risk of spillover using daisy quorum drive

Frederik J.H. DE HAAS¹, Léna KLÄY², Florence DÉBARRE², Sarah P. OTTO¹

¹ Biodiversity Research Center, Department of Zoology, University of British Columbia, Canada.

² Institute of Ecology and Environmental Sciences Paris (IEES Paris), Sorbonne Université, CNRS, IRD, INRAE, Université Paris Est Creteil, Université de Paris, Paris Cedex 5, France.

5.1 Abstract

Engineered gene-drive techniques for population alteration and/or suppression have the potential for tackling complex challenges, including reducing the spread of diseases and invasive species. Unfortunately, the self-propelled behavior of drives can lead to the spread of transgenic elements beyond the target population. Gene-drive systems with a low threshold frequency for invasion, such as homing-based gene drive, require initially few transgenic individuals to spread and are therefore easy to implement. However, their ease of spread presents a double-edged sword; their low threshold makes these drives much more susceptible to spread outside of the target population (spillover). We model a proposed drive system, called “daisy quorum drive,” that uses daisy-chain technology to transition over time from a low threshold drive system (homing-based gene drive) to a high threshold drive system (two-locus underdominance). This combination tends to restrict the drive strategy spatially, while maintaining an attainable release threshold. Developing and analyzing a proof of concept model, we find that daisy quorum drive is a promising technique for altering local population characteristics, and we explore how different attributes of the system affect the risk of transgenic elements spreading beyond the target area. Daisy quorum drive is less appropriate when population suppression is the goal, because the construct is easily swamped by immigration from other regions once the population size declines.

Contents

5.1	Abstract	147
5.2	Introduction	148
5.3	Methods and results	151
5.3.1	Drive design	151
5.3.2	Dynamics in an isolated population	152
5.3.3	Invasion analysis	155
5.3.4	Spatial spread of daisy quorum drive in a discrete environment	156
5.3.5	Spatial spread of daisy quorum drive in a continuous environment	158
5.4	Discussion	160
5.5	Acknowledgements	161

5.2 Introduction

Engineered gene drives are self-replicating genetic constructs that can bias transmission of desired alleles (usually called payload genes) to progeny, allowing them to rapidly increase in frequency even when the alleles are selected against [34]. Introgression of these transgenic genes into sexually reproducing species could potentially solve many pressing environmental and humanitarian problems, ranging from public health and agriculture to conservation [162]. Examples include modifications of the mosquito genome to reduce its capacity to serve as a disease vector (e.g., reducing the number of female mosquitoes, their longevity or their ability to support development and transmission of the pathogen) [118]. Other examples include driving genes that protect species at risk by eradicating invasive species or reducing pest damage in agriculture by replacing resistant alleles with their ancestral equivalents to restore vulnerability to pesticides or herbicides [162].

There are two general goals to gene drive: population suppression and population alteration. Population suppression makes use of gene-drive systems to purposefully reduce the size of the target population by lowering fertility or survival, while population alteration uses gene-drive systems to alter the genomic composition without the aim to sterilise or kill. A successful example of population suppression in the laboratory was shown in [133] for *Anopheles gambiae*. They used a CRISPR-Cas9 gene drive targeting a doublesex gene (*Agdsex*) that leads to a bias in the effective sex ratio by rendering homozygous females completely sterile. Within 7–11 generations egg production was reduced to the point of total population collapse. Extreme caution with population suppression is required as it could send a ripple through ecosystems, endangering many other plants and animals than just the target species. For example, mosquitoes that live in the Arctic of Canada & Russia fly around in thick swarms, act as pollinators for Arctic plants, and are a prey source, so their loss could destabilize the ecosystem [131].

By contrast, the goal of population alteration is to change the composition, not size, of a target population. Population alteration potentially causes less ecological harm, but it might require multiple drives, e.g., one for each disease carried by a mosquito vector. For example, [118] discovered that *Anopheles stephensi* mosquitoes expressing m1C3, m4B7, or m2A10 single-chain antibodies (scFvs) have significantly lower levels of infection compared to controls when challenged with *Plasmodium falciparum*, a human malaria pathogen. Further research needs to show if expression of a single copy of

a dual scFv transgene can completely inhibit parasite development, potentially without major ecological effects on mosquito populations.

Although gene drives hold much promise, they also raise the strong possibility of unwanted ecological impacts. Major risks include the unintended spread of the payload beyond the area of interest due to unwanted migration of individuals and the spread of the payload beyond the target species due to hybridization. Low threshold gene-drive systems, which can spread even from very low frequency such as homing endonuclease-based gene drive (HEG), are especially susceptible to such spillovers. These systems exploit homology-directed repair (HDR) to replace a targeted, naturally occurring genetic sequence with an engineered construct. Homing-based gene drives work by transcribing an endonuclease (often Cas9) and a guide RNA (gRNA) in the germline. The endonuclease then triggers a double-strand break at the complementary site of the gRNA, which can be repaired by either one of two pathways: non-homologous end joining (NHEJ) or homologous recombination via HDR. In NHEJ the break ends are directly ligated without the use of a homologous template. This often leads to mutations arising in the process, preventing the gRNA from recognizing the sequence for future double-strand breaks. However, when the double-strand break is repaired by HDR, which requires recombination with the homologous chromosome carrying the construct, the result is a copy of the homologous chromosome at the double-strand break: creating effectively homozygous cells from heterozygous cells. The result is equivalent to meiotic drive, δ , leading to a greater fraction of gametes carrying the construct than expected under Mendelian inheritance (specifically, heterozygotes produce a fraction $\frac{1}{2}(1 + \delta)$ of gametes carrying the driven allele).

High conversion rates have been achieved in the lab by the CRISPR-Cas9 system for yeast ($\delta > 0.99$, [74]), fruit flies ($\delta > 0.85$, [224]), and the malaria vector mosquito, *Anopheles stephensi* ($\delta > 0.90$, [88]), although high levels of drive are not always seen and drive can depend strongly on the genetic background [43]. Drive systems require the introduction of few transgenic individuals for spread to occur (low threshold), making them very capable of spilling over to neighboring populations by migration or to related species by hybridization [151]. Numerous countermeasures have been proposed to reverse gene drives such as synthetic CRISPR-Cas9 reversal drives like CATCHA [223] and ERACR [92]. However, reversal drives are techniques to fully eliminate rather than locally confine the drive, which is often the desired outcome.

To reduce the risk of unintended spread, a new self-exhausting CRISPR-Cas9 technique called “daisy-chain drive” was considered [166]. The design of daisy-chain drive spreads multiple interdependent homing-based drive components across the organism’s genome, each bearing a cost (drive load). Each genetic element drives the next in the chain. The final element carries the payload and is driven to higher and higher frequencies due to the preceding drive elements. Because no element can drive itself and the first element is not driven at all, all elements are ultimately expected to be lost from the population due to natural selection against the drive load, returning the population to its original wildtype state. Unfortunately, as the drive disappears, so does the payload gene, rendering this system ineffective in the long term unless regularly reintroduced [166].

Persistent and localized introgression would be possible if the daisy-chain alleles drive a high threshold system (such as underdominance) past its threshold frequency. The advantage of such a construct is that it is expected to be eliminated from adjacent populations once the daisy chain is exhausted due to the low frequency of the construct outside of the target population (below the threshold). This design was first proposed conceptually by Min et al., under the name “daisy quorum drive” [157] but has not yet been modeled, which is our goal here. This new drive design is predicted to spread through a population until all of its daisy elements have been lost, at which point its fitness becomes frequency-dependent. The result is an engineered population surrounded by wildtype populations with

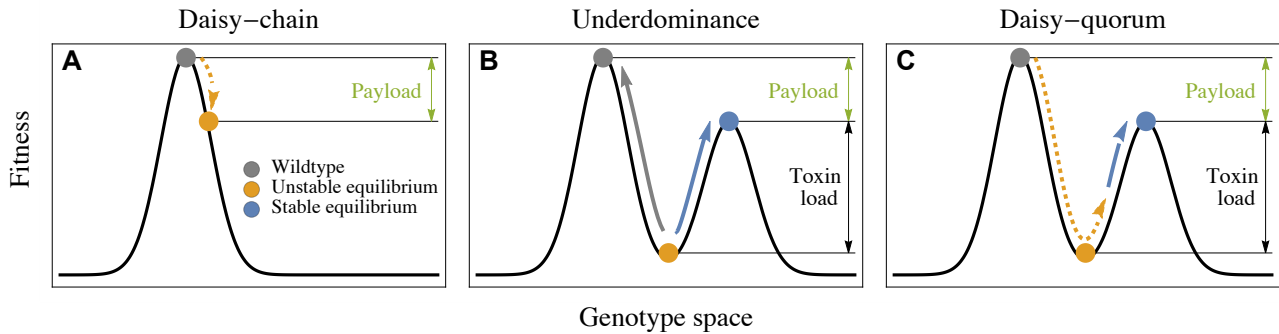


Figure 5.1: Conceptual overview of the different control systems. (A) A daisy-chain drive introgresses a payload fitness cost temporarily, ultimately reverting back to the wildtype state (grey). (B) The dynamics of underdominant gene drive is frequency dependent: left of the valley the population is driven back to a wildtype composition, while right of the valley it evolves toward the lower peak expressing the payload fitness cost. (C) The proposed design (daisy-quorum drive) includes both systems and uses the temporary low-threshold daisy chain to drive an underdominant construct past the valley, after which it evolves toward the other peak with the desired payload fitness cost.

limited mixing at the boundaries due to the reduced fitness of crosses between wildtype individuals and individuals carrying the engineered alleles.

Drives that exhibit a high invasion threshold, such as engineered underdominance and Wolbachia drives, provide a natural means to achieve localization. These drives exhibit frequency-dependent dynamics where the drive can only spread if its frequency exceeds the invasion threshold. Below this level, the frequency of the drive will decrease, leading to its loss from the population. Spread of a threshold drive across a patchy environment is more difficult when the invasion threshold is 50 percent or higher [16]. In mosquitoes, high-threshold Wolbachia drive, introduced by transinfection from another host species, has already been used to control dengue, Zika, and other arboviruses in the wild [210]. Because organisms that migrate to other populations will be in the minority and consequently are selected against, underdominance with high thresholds poses a substantially lower risk of rapid, unintended spread.

An ideal gene-drive system to alter natural populations would exclusively affect organisms within a confined area. Here we present a population genetics model to elucidate the evolutionary dynamics of daisy quorum drive that may help satisfy this criterion by combining daisy-chain drive with a two-locus underdominant system (Figure 5.1). The daisy-chain drive system provides a short window of time to drive the frequency-dependent underdominant system before drive is exhausted, effectively transitioning from a low to a high threshold control system. By exploring spatial models in the face of recurrent migration, we find theoretical support that this new design could reduce the risk of gene-drive spillovers while providing persistent payload introgression to the target population(s), and we determine the conditions that reduced risk of spillover. We caution that the risks of gene drive cannot be entirely eliminated given the possibility of off-target effects and compensatory mutations, but a daisy quorum drive system would serve to lower the risks.

Daisy-chain Component				Underdominant Component			
Locus A		Locus B		Locus C		Locus D	
<i>a</i>	Wildtype	<i>b</i>	Wildtype	<i>c</i>	Wildtype	<i>d</i>	Wildtype
<i>A</i>	CRISPR-Cas9 gRNA targetting <i>b</i> Drive load (s_d)	<i>B</i>	CRISPR-Cas9 gRNA targetting <i>c</i> gRNA targetting <i>d</i> Drive load (s_d)	<i>C</i>	Payload (s_p) Suppressor of <i>D</i> Promoter Toxin load (s_t)	<i>D</i>	Payload (s_p) Suppressor of <i>C</i> Promoter Toxin load (s_t)

Table 5.1: Four-locus daisy-quorum drive. Loci **A** and **B** are part of the daisy-chain component and loci **C** and **D** make up the two-locus underdominance component. The daisy-chain transgenic alleles (**A** and **B**) have a fitness cost of s_d (drive load), and the engineered underdominant alleles (**C** and **D**) have a toxin load s_t and a payload fitness cost s_p , which interact according to Table 5.2. The gRNA of locus **A** is complementary to the wildtype allele *b*, and locus **B** carries two gRNAs to cut the wildtype alleles *c* and *d* at the underdominant loci.

5.3 Methods and results

We formulate a deterministic model that considers the evolution of a large population of diploid organisms. We first introduce the model for a single population and then build up to an array of populations under recurrent migration to explore the spatial spread of the engineered construct. We refer to the loci in our model by bold letters (e.g., **A**) and to the alleles by the italic letters (e.g., *A* or *a*)

5.3.1 Drive design

We model a multilocus bi-allelic population with random mating using discrete-time dynamics. Our proposed drive system consists of two components. The first component encodes a daisy-chain gene drive, which can be of arbitrary length, and the second component is a two-locus engineered underdominant system with an added payload gene on each transgenic allele. We explore here a four-locus genetic architecture as an example case (Table 5.1), although the daisy chain could contain more loci if a longer acting drive is needed [157, 166]. Unless otherwise stated, we assume that the loci involved are freely recombining ($r = \frac{1}{2}$) and that all dynamical systems are initialized at a proportion $1 - f_0$ of *abcd* gametes and f_0 of the introduced *ABCD* gametes.

The first component (daisy-chain) consists of a linear series of n loci arranged such that each element drives the next in the chain. The final element in the chain drives the “cargo”, which in our model are the two underdominant loci **C** and **D**. Each transgenic allele of the daisy chain consists of a CRISPR-Cas9 complex and a drive load, which bears a fitness cost of s_d . The drive load helps ensure that all CRISPR-Cas9 elements are counterselected and eventually disappear from the system.

In the case of the daisy-chain, the gRNA at locus i in the chain guides the Cas9 nuclease to the wildtype allele at locus $i + 1$, inducing a double-strand break, which is repaired by homologous recombination with probability δ . We do not incorporate natural resistance alleles to the daisy-drive elements because their effects are less critical due to the transient nature of the daisy-chain drive (see

discussion).

The two-locus underdominant component is structured following the design proposed by Davis et al. [66]. The two transgenic alleles consist of four elements each: a payload, a suppressor, a promoter, and a toxin. The two engineered alleles are carried on non-homologous chromosomes, each of which produces a toxin unless the suppressor on the other transgenic allele is also present. Specifically, in the absence of the matching suppressor, the promoter causes the toxin gene to be expressed, while individuals who carry one or more copies of both constructs are viable because both suppressors are present and neither toxin gene is expressed (Table 5.2). (As an aside, by using different toxins and matching suppressors, a two-locus underdominant system is more straightforward to engineer than a one-locus underdominant system, because of the lack of good models to engineer heterozygous disadvantage at a single locus.) This scheme creates a frequency-dependent fitness regime, promoting the state of carrying both engineered constructs or none and selecting against individuals carrying one but not the other transgenic allele. As shown in [72], this two-locus underdominant component requires a high initial frequency to establish within a population.

	cd	cD	Cd	CD
cd	1	$(1 - s_t)(1 - s_p)$	$(1 - s_t)(1 - s_p)$	$(1 - s_p)$
cD	$(1 - s_t)(1 - s_p)$	$(1 - s_t)(1 - s_p)$	$(1 - s_p)$	$(1 - s_p)$
Cd	$(1 - s_t)(1 - s_p)$	$(1 - s_p)$	$(1 - s_t)(1 - s_p)$	$(1 - s_p)$
CD	$(1 - s_p)$	$(1 - s_p)$	$(1 - s_p)$	$(1 - s_p)$

Table 5.2: Relative fitness for each diploid genotype in the two-locus underdominant component. The toxin s_t is expressed in individuals carrying either a C or D allele but not both. The payload fitness cost s_p is expressed if the individual carries at least one engineered allele (dominantly expressed). The resulting fitness landscape has two peaks (cd or CD fixed), separated by a fitness valley caused by the toxin load, which we refer to as a two-locus underdominant form of epistasis.

The fitness effects resulting from the toxin $(1 - s_t)$, drive load $(1 - s_d)$, and payload $(1 - s_p)$ are assumed to multiply together across components to determine individual fitness. Table 5.2 illustrates the effect of loci **C** and **D** on individual fitness when the payload cost (s_p) is dominantly expressed. In the Appendix, we also explore different assumptions about the expression of the payload cost, including multiplicative (5.A.4) and recessive (5.A.5) expression. To derive the relative fitness for the full genotype, including the drive load, the fitness of the **CD** genotype (Table 5.2) is multiplied by $(1 - s_d)^\rho$ where ρ equals the number of A and B alleles.

5.3.2 Dynamics in an isolated population

The recursion equations for the 16 (2^4) gamete types are too elaborate so we report them in the supplementary *Mathematica* file. Here, we describe the dynamical equations for the daisy-chain component (**A** and **B**) and the underdominant components (**C** and **D**) in isolation (Table 5.1). We use recursion equations to take the population through diploid selection followed by meiosis with drive and recombination.

Assuming random mating in a single isolated population, the recursion equations for the frequency X_{ij} of the gamete ij at the drive loci **A** and **B** are:

$$\begin{aligned}
X'_{ab} &= \frac{1}{\bar{W}}(X_{ab}^2 + X_{aB}X_{Ab}(1-s_d)^2(1-\delta)r + X_{ab}X_{aB}(1-s_d) + X_{ab}X_{Ab}(1-s_d) \\
&\quad + X_{ab}X_{AB}(1-s_d)^2(1-\delta)(1-r)) \\
X'_{aB} &= \frac{1}{\bar{W}}(X_{aB}^2(1-s_d)^2 + X_{aB}X_{Ab}(1-s_d)^2(1-r(1-\delta)) + X_{aB}X_{AB}(1-s_d)^3 + X_{ab}X_{aB}(1-s_d) \\
&\quad + X_{ab}X_{AB}(1-s_d)^2(1-r(1-\delta))) \\
X'_{Ab} &= \frac{1}{\bar{W}}(X_{Ab}^2(1-s_d)^2 + X_{aB}X_{Ab}(1-s_d)^2(1-\delta)(1-r) + X_{Ab}X_{AB}(1-s_d)^3(1-\delta) \\
&\quad + X_{ab}X_{Ab}(1-s_d) + X_{ab}X_{AB}(1-s_d)^2(1-\delta)r) \\
X'_{AB} &= \frac{1}{\bar{W}}(X_{AB}X_{ab}(1-s_d)^2(1-r(1-\delta)) + X_{AB}X_{Ab}(1-s_d)^3(1-\delta) + X_{AB}^2(1-s_d)^4 \\
&\quad + X_{aB}X_{Ab}(1-s_d)^2(r + \delta(1-r)) + X_{aB}X_{AB}(1-s_d)^3)
\end{aligned} \tag{5.1}$$

where $X_{ab} + X_{aB} + X_{Ab} + X_{AB} = 1$ and \bar{W} is the average fitness (the sum of the numerators). These recursions follow from the mating and gamete production Table S1 and are equivalent to equations analysed in [166] except in discrete time. An automated algorithm for deriving the dynamical equations for longer daisy chains is provided in the supplementary *Mathematica* file.

The recursion equations for the underdominant component, describing the frequency X_{ij} of the gamete ij at loci **C** and **D**, are:

$$\begin{aligned}
X'_{cd} &= \frac{1}{\bar{W}}(X_{cd}^2 + X_{cd}X_{cD}(1-s_t)(1-s_p) + X_{cd}X_{Cd}(1-s_t)(1-s_p) + X_{cD}X_{Cd}(1-s_p)r \\
&\quad + X_{cd}X_{CD}(1-s_p)(1-r)) \\
X'_{cD} &= \frac{1}{\bar{W}}(X_{cd}X_{cD}(1-s_t)(1-s_p) + X_{cD}^2(1-s_t)(1-s_p) + X_{cD}X_{Cd}(1-s_p)(1-r) + X_{cd}X_{CD}(1-s_p)r \\
&\quad + X_{cD}X_{CD}(1-s_p)) \\
X'_{Cd} &= \frac{1}{\bar{W}}(X_{cd}X_{Cd}(1-s_t)(1-s_p) + X_{Cd}^2(1-s_t)(1-s_p) + X_{cD}X_{Cd}(1-s_p)(1-r) + X_{cd}X_{CD}(1-s_p)r \\
&\quad + X_{Cd}X_{CD}(1-s_p)) \\
X'_{CD} &= \frac{1}{\bar{W}}(X_{cd}X_{CD}(1-s_p)(1-r) + X_{cD}X_{CD}(1-s_p) + X_{Cd}X_{CD}(1-s_p) + X_{CD}^2(1-s_p) \\
&\quad + X_{cD}X_{Cd}(1-s_p)r)
\end{aligned} \tag{5.2}$$

where $X_{cd} + X_{cD} + X_{Cd} + X_{CD} = 1$ and \bar{W} is the average fitness (see Table S2).

Assuming that the construct is introduced in individuals bearing the CD haplotype, the C and D alleles are initially equally frequent ($X_{cD} = X_{Cd}$ at $t = 0$) and remain so over time according to Equation 5.2 ($X'_{cD} = X'_{Cd}$). Both $X_{cd} = 1$ and $X_{CD} = 1$ are fixed point solutions to the underdominant component of the dynamical equations 5.2, and a local stability analysis indicates that both are locally stable when the payload fitness cost is dominant. A third fixed point represents the co-existence of alleles C and c , as well as D and d , and is locally unstable (see Appendix for ternary plots of alternative payload fitness regimes).

In Figure 5.2, we illustrate the location of the stable equilibria (blue and grey vertices) and the separatrix (the boundary separating two basins of attraction in a dynamical model) for different payload fitness costs s_p when the toxin creating the fitness valley is relatively weak ($s_t = 0.1$) or very strong ($s_t = 1.0$). The separatrix is calculated numerically. Example dynamics confirm that the system converges to either fixed point depending on whether initialized to the left or right side of the separatrix.

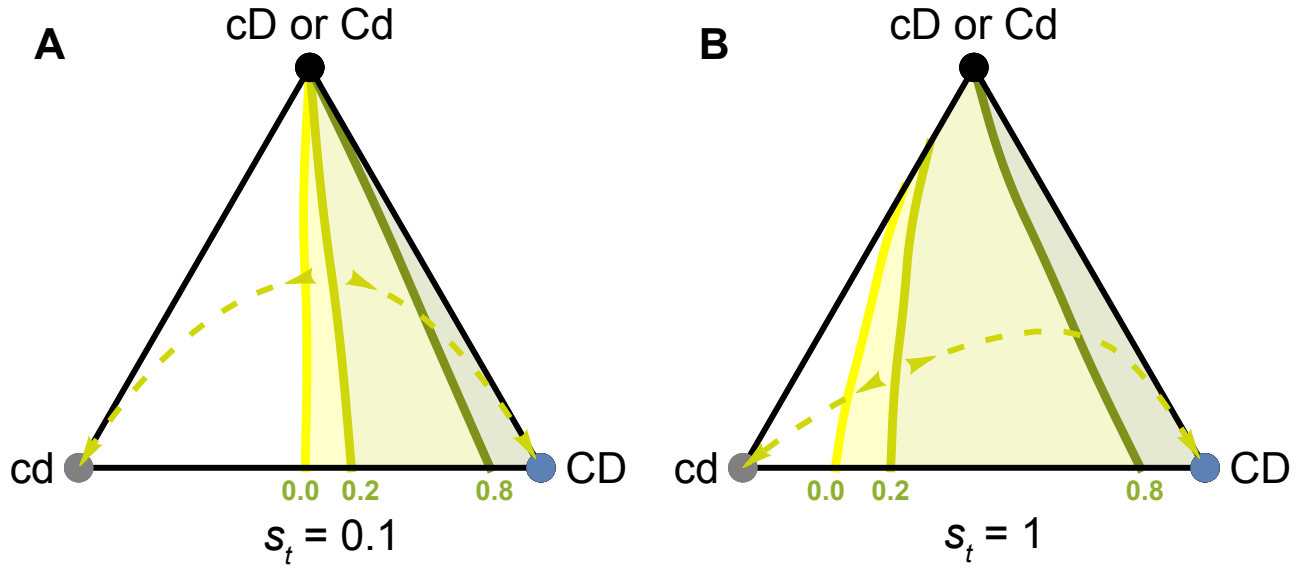


Figure 5.2: Ternary plot illustrating the separatrix (the boundary separating basins of attraction to different fixed points) for the two-locus underdominant construct with a toxin load of (A) $s_t = 0.1$ or (B) $s_t = 1.0$ and a payload fitness cost of $s_p = 0.0, 0.2$ and 0.8 (green curves). Locus **C** and **D** recombine freely ($r = \frac{1}{2}$). The dashed curves are example dynamics for $s_p = 0.2$.

The unstable equilibrium lies on the separatrix and is centred, with wildtype and cargo alleles equally frequent ($p_C = p_c = p_D = p_d = 1/2$), when:

$$s_t = \sqrt{\frac{rs_p^2}{2(1-s_p)(2r(1-s_p) - s_p)}} \quad (5.3)$$

As the payload fitness cost increases (larger s_p), the separatrix moves to the right in Figure 5.2, reducing the basin of attraction of the engineered underdominant construct (haplotype CD). Higher payload fitness costs thus make it harder to drive such constructs into a population, but they also provide stronger fitness effects if the construct is able to fix. By contrast, as the toxin load increases (larger s_t), the separatrix moves left, and the basin of attraction to CD increases. While somewhat counterintuitive, this occurs because the toxin load can be suppressed by even one copy of the alternate allele, making haplotypes Cd and cD strongly selected against when cd is common (the wildtype) but only weakly selected against when CD is common, causing the fitness surface to fall faster near cd than near CD when the toxin load is stronger.

To prevent unintentional spread into neighbouring populations, it is desirable to design a drive where the separatrix is shifted towards the transgenic haplotype (higher threshold), so invasion will only occur once a significant number of transgenic haplotypes have been introduced into a new population. However, shifting the unstable equilibrium too far to the transgenic haplotype allows the population to revert to the wildtype state more easily due to drift and incoming wildtype migrants from neighbouring populations. Adjusting the toxin load s_t relative to the payload fitness cost s_p is required to balance these benefits and risks.

To derive the dynamics for all loci, the daisy-chain and underdominant components are linked by allele A driving allele B , which in turn drives alleles C and D .

5.3.3 Invasion analysis

We next perform an invasion analysis to determine the drive rate δ_c needed for the transgenic alleles C and D to spread. Here δ_c is treated as a constant value, equalling the drive strength δ times the frequency of individuals bearing the driving allele B among heterozygotes at the driven loci (**C** and **D**). Assuming that the daisy quorum drive construct is introduced with all components in the same individuals, initially all heterozygotes at loci **C** and **D** bear the driver allele B , so $\delta_c = \delta$. The genetic association between loci **B** and **C** or **D** decreases over time due to recombination, however, reducing the rate at which the underdominant construct is driven over time, as discussed in Appendix 5.A.3.

For a given initial level of drive δ_c , the invasion fitness of the payload (loci **C** and **D**) is given by the larger of the two eigenvalues in a stability analysis, λ_L :

$$\lambda_L = \max[(1 + \delta_c)(1 - s_p)(1 - s_t), (1 + \delta_c^2 - r(1 - \delta_c)^2)(1 - s_p)] \quad (5.4)$$

As long as there is a payload fitness cost ($s_p > 0$), drive is necessary for the spread of the cargo ($\delta_c > 0$). Because recombination separates the toxin load from its antidote (separating alleles C and D), looser linkage reduces the risk that the construct spreads when rare outside of the target area (reducing the second term).

Focusing on unlinked constructs ($r = \frac{1}{2}$), the cargo will spread when rare if:

$$1 + \delta_c > \min\left[\frac{1}{(1 - s_p)(1 - s_t)}, \sqrt{\frac{2}{1 - s_p}}\right] \quad (5.5)$$

The minimum term describes the conditions under which either allele C or D could spread on their own (first criterion) or the combined CD haplotype could spread (second criterion). With a low toxin load (s_t), the first criterion is easier to satisfy, and we expect the alleles to spread individually. With a high toxin load, however, only the combined haplotype CD will spread, as long as drive is strong enough to counter the haplotype being broken down by recombination.

Given that $\delta_c \leq 1$, a two-locus underdominant construct carrying a dominant payload fitness cost with $s_p > 0.5$ can never invade, no matter the strength of the drive. With multiplicative or recessive payload fitness regimes, a higher payload cost can spread initially, but the construct does not reach fixation because individuals such as Cd/cD bear a lower payload cost while also avoiding the toxin load and so have higher fitness than CD/CD homozygotes (see Appendix 5.A.4 and 5.A.5).

Figure 5.3 illustrates the contrasting effects of a daisy chain driving a constant ‘‘cargo’’ load into a population, compared to a two-locus underdominant ‘‘cargo’’ for different initial release frequencies f_0 . A constant cargo load is modeled as a third locus C' that only contains a payload, as is generally considered with daisy-chain drives. Such a constant cargo does not spread the payload to fixation as it is always selected against once the drive ends, while the underdominant system allows the fixation of the payload provided that it is released at a high enough initial frequency. Panel D illustrates the scenario for a lower initial release frequency where the daisy drive proved to be not powerful enough to push the underdominant construct past the separatrix. Longer daisy-chains can drive the cargo to fixation with even lower initial frequencies (e.g., $f_0 = 0.02$ is sufficient with $n = 3$). With daisy-chain drive, an underdominant construct carrying a payload can thus spread at a far lower initial release frequency than predicted for the underdominant component on its own, which requires a starting frequency of over 0.36 for the parameters considered in Figure 5.3.

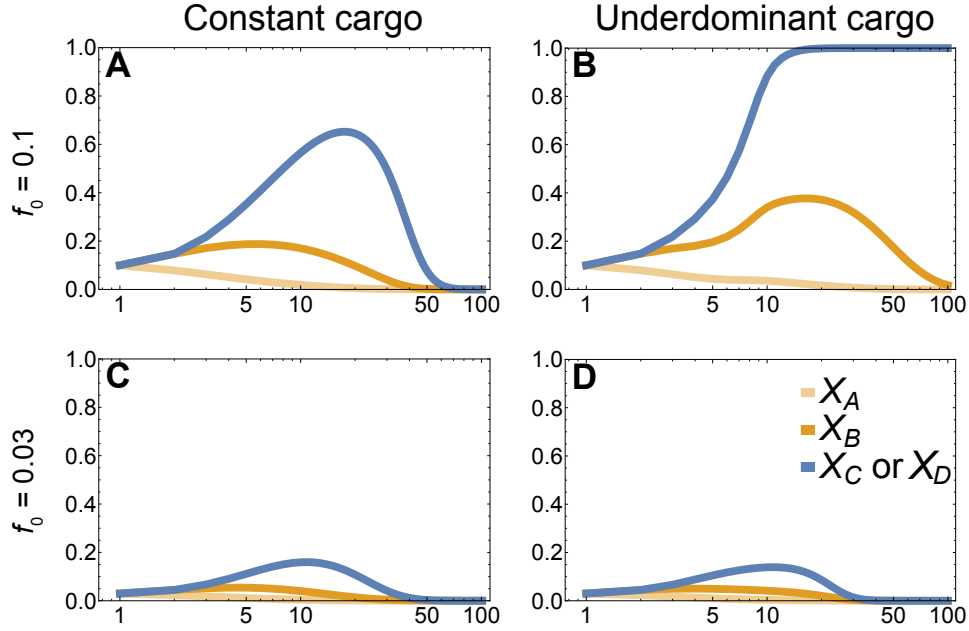


Figure 5.3: Left panels illustrate dynamics of a daisy chain pushing a constant payload fitness cost $s_p = 0.15$, while right panels show the daisy chain pushing an underdominant construct, also with a payload fitness cost of $s_p = 0.15$. The panels are initialized with gametes $ABCD$ at frequency $f_0 = 0.1$ (top) and $f_0 = 0.03$ (bottom). Other parameters are: $s_t = 0.9$, $\delta = 0.95$, $s_d = 0.05$ and $r = \frac{1}{2}$.

5.3.4 Spatial spread of daisy quorum drive in a discrete environment

Given the risk of spread to nearby populations, we now consider the drive system in a spatial context, first with discrete spatial patches and then in continuous space. To begin, we investigate a linear stepping-stone model of $M = 101$ interconnected populations with migration rate m between adjacent patches. Migration is followed by dynamics within each patch, as described in section 2.2. The frequency of gametes of type i after migration for a non-boundary population j is given by: $p_i^{j'} = (1 - m)p_i^j + \frac{1}{2}mp_i^{j-1} + \frac{1}{2}mp_i^{j+1}$. The boundary populations only give and receive migrants from the interior: $p_i^{0'} = (1 - \frac{1}{2}m)p_i^0 + \frac{1}{2}mp_i^1$ and $p_i^{M'} = (1 - \frac{1}{2}m)p_i^M + \frac{1}{2}mp_i^{M-1}$.

Parameter	default value
Initial frequency (f_0)	0.05
Drive rate (δ)	0.9
Recombination rate (r)	0.5
Drive load (s_d)	0.02
payload fitness cost (s_p)	0.1
Toxin load (s_t)	0.9

Table 5.3: Default parameter values for Figures 5.4 and 5.5

The migration rate m is a crucial parameter determining the spread of the payload through interconnected populations (Figure 5.4), as found in [72] in their simulations of a two-locus underdominant

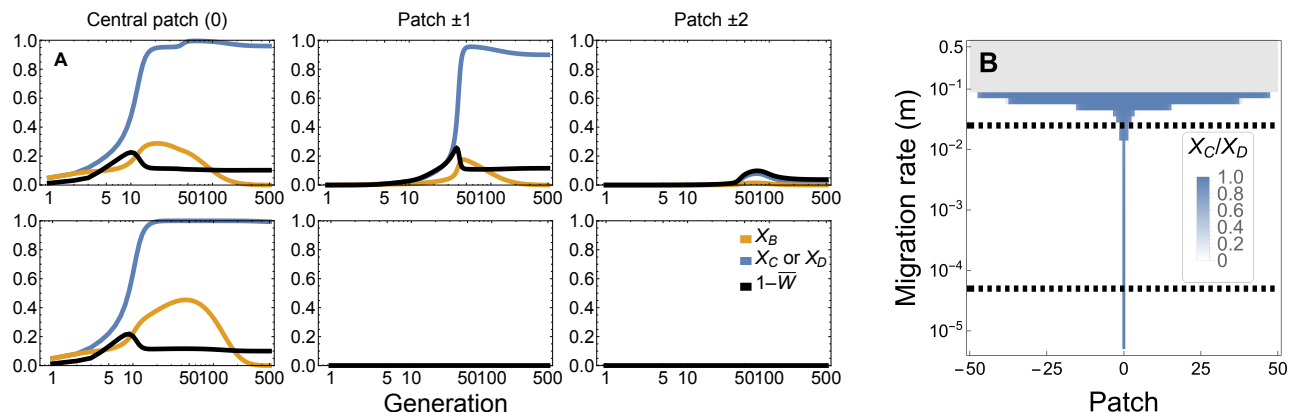


Figure 5.4: Dynamics of spread of a payload for different migration rates when $ABCD$ gametes are introduced into the center patch (patch 0) at an initial frequency of $f_0 = 0.05$. Top and bottom row of panel A show 3 interconnected patches and their dynamics for a migration rate of $m = 0.05$ and $m = 0.00005$, respectively (see Table 5.3 for other parameters). For a relatively high migration rate (top row), spread to neighbouring populations is more likely than for lower migration rates (bottom row). As predicted, the drive disappears from the system due to the drive load $s_d = 0.02$ (orange curves). Panel B shows the frequency of the payload alleles, X_C or X_D , at 1000 generations (blue shading) across all patches (x axis) for a variety of migration rates (y axis). In the gray area of panel B, migration into the center patch (patch 0) is high enough to swamp the engineered construct causing no spread to occur anywhere (dashed lines represent the parameters in panel A).

construct on its own. For moderate to low migration rates (m less than or near 0.05), the construct spreads to high frequency locally but remains spatially restricted, as intended. By contrast, if migration is so high that the population is nearly panmictic ($m \approx 0.5$), the construct fails to spread due to gene swamping, causing the initial frequency at the point of introduction, f_0 , to fall below the threshold needed for invasion (grey area in panel B). Intermediate migration rates pose the greatest risk (m roughly between 0.05 and 0.1), because migration is not too high to swamp the construct but it is high enough to lead to broad spread of the construct (blue horizontal stripe in panel B).

Given a particular migration rate m , the power of the daisy chain combined with the location of the separatrix determine the risk of spillover into neighbouring populations. Lowering the drive load s_d or increasing the drive rate δ results in a daisy chain that persists for longer and risks spreading the underdominant construct to more populations (Figure 5.5, panel A versus E and B versus F), as would lengthening the daisy chain n (S1). By contrast, raising the payload fitness cost s_p reduces the risk of spillover (Figure 5.5, panel D versus H). Interestingly, decreasing the toxin load s_t , which raises the fitness valley, reduces the risk of spread of the payload to neighbouring populations, even under high migration rates (Figure 5.5, panel C versus G). This is because the lower the toxin load, the closer the separatrix becomes to the state where the underdominant construct CD is fixed (Figure 5.2). Thus, somewhat counterintuitively, a lower toxin load provides a higher degree of safety in isolating the underdominant construct from neighbouring patches once the daisy-chain drive has exhausted itself (this assumes that the toxin is neutralized by a single antidote allele, as in Table 5.2).

Stochastic simulations confirm that these results are robust to finite population sizes in cases where the population persists (Figure S1). In cases where the payload drives the local population extinct, recolonization eventually occurs from neighbouring populations (Figure S2).

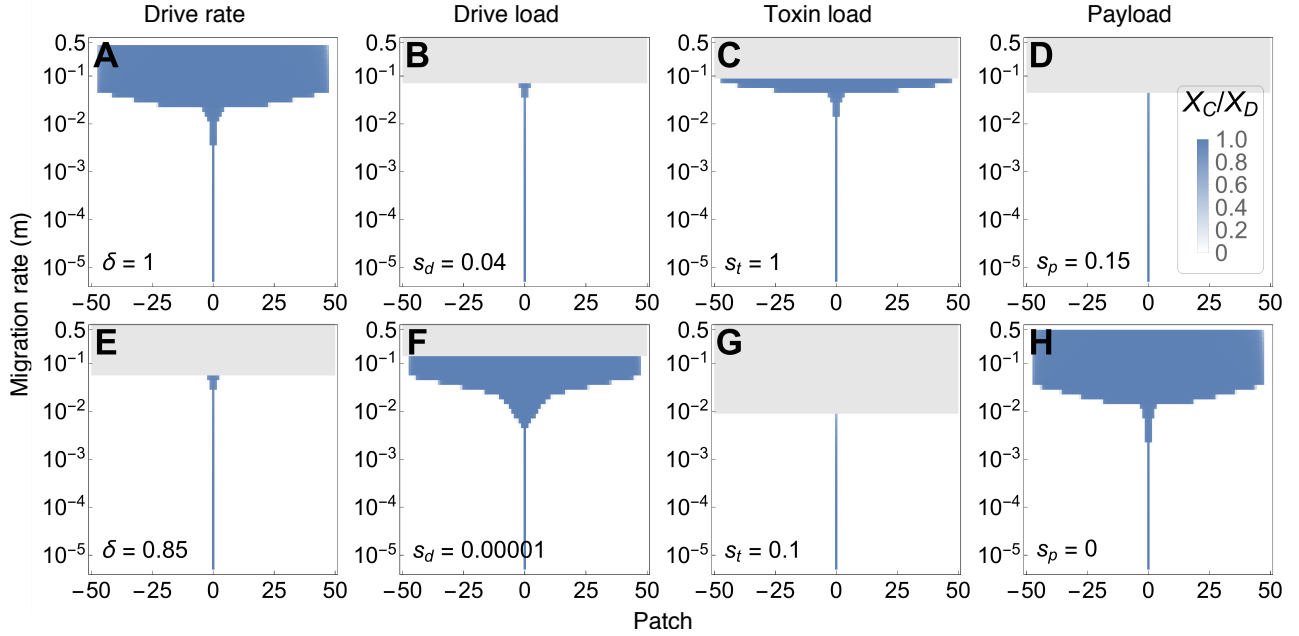


Figure 5.5: Illustration of the effect of changing a single parameter on the results shown in Figure 5.4 (panel B) for the default parameters (Table 5.3). The frequency of the payload is shown in blue after 1000 generations when $ABCD$ gametes are introduced into the center patch (patch 0) at frequency $f_0 = 0.05$. Patches are linearly connected and exchange migrants every generation at a rate m . Subfigures from left to right change the following single parameter up (top figures) or down (bottom figures): drive rate δ , drive load s_d , toxin load s_t and the payload fitness cost s_p

5.3.5 Spatial spread of daisy quorum drive in a continuous environment

We now ask whether the construct can be confined even if space is continuous and under which conditions. We continue to use the dynamics described in section 5.3.2, with diffusion terms added in the same way for all genotypes (at rate $\mathcal{D} = 0.2$ per time unit and spatial unit), unless otherwise mentioned (see Appendix 5.C.1 for details).

In a continuous and homogeneous environment, a daisy quorum drive system does not reach a stable and spatially-restricted equilibrium (Appendix 5.C.3). Either the construct is driven outward in an expanding wave or the wave collapses in on itself. In either case, the wave speed approaches a non-zero constant (Figure S1). Importantly, the asymptotic wave speed is the same whether or not alleles A and B are initially present, because the drive is eventually exhausted. As in the discrete-patch model, the risk of spread of the underdominant construct is then higher when the separatrix is closer to the wildtype (high s_t and low s_p), than when it is closer to fixation of the underdominant construct (low s_t and high s_p). Thus, it is critical that the toxin load be light enough and the payload fitness cost heavy enough to ensure that the daisy quorum drive is not expected to spread unless driven (i.e., has negative asymptotic wave speeds in Figure S1).

To better understand the differences between discrete patches and continuous space, we discretized space by concentrating individuals into fewer and fewer patches, with larger steps between them, but holding the diffusion rate constant (Figure 5.6). As expected from the continuous-space model, with many patches and small steps between them (near zero on the x-axis), the wave reaches a steady speed, either expanding or contracting. As the step size increases further, the wave speed decreases in absolute

value, until a critical size is reached, above which the speed becomes zero. In other words, in a discrete environment with large enough gaps between patches, the drive is confined and maintains itself around the area of introduction, whereas in continuous space it either expands (small payload cost s_p given the toxin load s_t) or contracts (large payload cost s_p). These results echo those of Barton [15], who determined the critical size of spatial steps analytically in the case of a single-locus underdominant model in one spatial dimension.

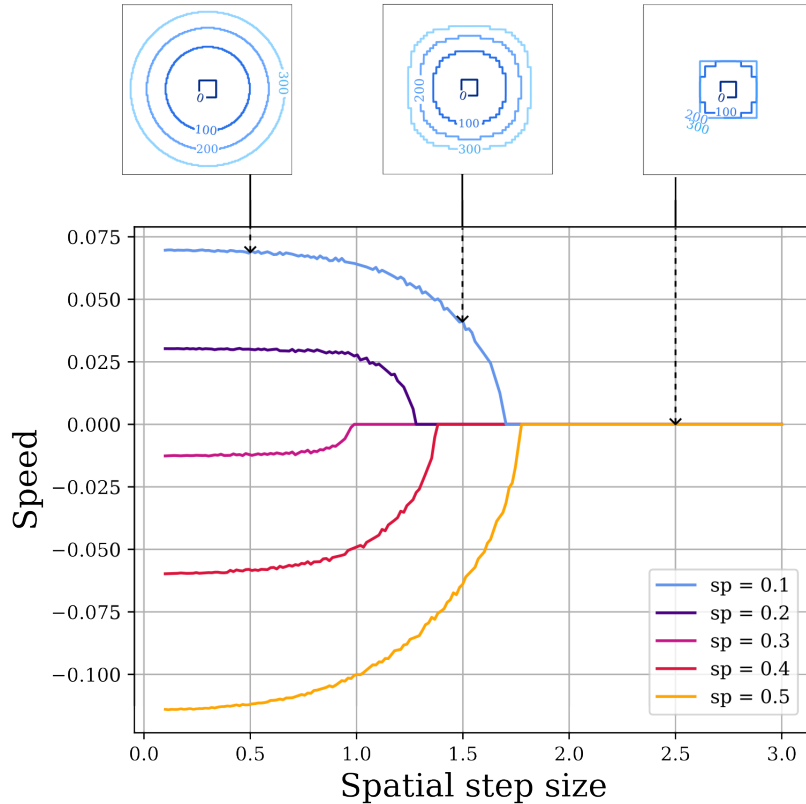


Figure 5.6: The asymptotic wave speed for daisy quorum drive as a function of the spatial step size for different values of the payload cost s_p across a two-dimensional area. The three panels above the graph show the convex hull containing 80% of the population for either the C or D allele at $t = 0$ (dark blue contour), $t = 100$, $t = 200$ and $t = 300$ (light blue), with $s_p = 0.1$. Parameters are as follows: $r = 0.5$ (recombination rate), $s_d = 0.02$ (drive load), $s_t = 0.9$ (toxin load), $\delta = 0.9$ (drive rate), $T = 4000$ (final time), $L = 800$ (length of the spatial domain), $\mathcal{D} = 0.2$ (diffusion rate), with each generation split into ten time steps to mimic continuous time and the spatial domain split into a series of patches (from 8000^2 down to 267^2 , across the two dimensions) at increasing distances apart (from 0.1 to 3 spatial units).

To avoid spread of the underdominant construct across environments that are spatially continuous, payload fitness costs should be high enough and toxin loads low enough that negative asymptotic wave speeds are expected. While such waves would eventually be expected to disappear in a continuous environment, we find that they can be stably maintained when initially driven if there is spatial heterogeneity in the ease of dispersal. Specifically, if the diffusion rate varies across space, a wave that

would otherwise collapse can expand temporarily due to gene drive and then stabilize (Figure 5.7). Regions that are hard to traverse or that support fewer individuals can thus act as barriers, stabilizing the daisy quorum drive system in space. While some wild-type individuals can enter a region where the construct has stabilised, they are never numerous enough to trigger an invasion. Similar behaviour has been observed in the one-locus underdominant system studied in [15].

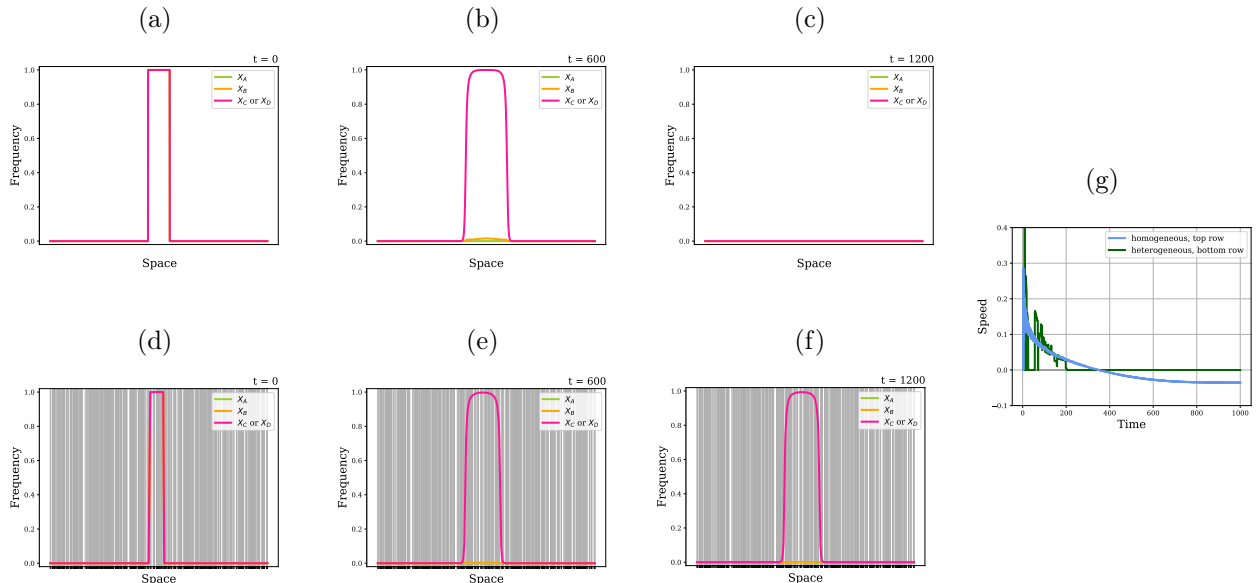


Figure 5.7: Comparing the spread of introduced genotypes across homogeneous or heterogeneous landscapes. The first three panels in each row represent allele frequencies in one-dimensional space at times $t = 0, 600, 1200$. The rows illustrate simulations with homogeneous distances between adjacent sites (top row) or heterogeneous distances (bottom row, with step sizes determined by randomly choosing 400 boundaries across the spatial domain). The last panel describes the speed of the cargo wave measuring the frequency of either C or D (i.e., the speed of the pink curve in previous panels), either in a homogeneous (blue) or heterogeneous (green) environment. The payload fitness cost is $s_p = 0.35$, parameters are otherwise the same as Figure S1. This payload fitness cost value was chosen so that the wave collapses in a homogeneous environment (with a negative asymptotic speed; top row) but can first expand due to drive and then be stably maintained with spatial heterogeneity (bottom row).

5.4 Discussion

Low threshold gene drives such as homing-based gene drives risk spillover to neighbouring populations due to migration and hybridization. The gene-drive design that we study here transitions in time from a low threshold drive (homing-based gene drive), which is especially susceptible to these spillovers, to a high threshold drive (two-locus underdominance), which is more resilient to spillover. This design, called “daisy quorum drive” in [157], couples a daisy-chain technique which limits the temporal extent of gene drive [166], and a two-locus underdominant construct, which stabilizes the construct in space [72]. Here, we have developed and analysed the first model of daisy quorum drive. We find that this design can lead to the spread and maintenance of a payload that alters a target population, while limiting further spread because the driver alleles are eliminated over time. The underdominant

construct remains spatially restricted as long as it is built with a light enough toxin load and a heavy enough payload fitness cost to avoid spread through more homogeneous environments (Figures 5.4, 5.5, and 5.7).

Previous models of homing-based gene drives have focused on the rise of resistance alleles that would prevent the drive from completely reaching fixation in the target population and ultimately revert the system back to the wildtype state. Here we have not taken resistance alleles into account because the daisy-chain drive system, as used in a daisy quorum drive, stops driving and declines in frequency soon after its introduction, before resistance is likely to rise substantially. (This attribute differs from the daisy-chain drive system on its own, where the drive construct has to be repeatedly introduced to maintain the cargo over time [166].) However, we note that the two-locus underdominant construct is susceptible to mutations that decrease or compensate for the payload fitness cost, which would affect the long-term persistence of the underdominant construct. By contrast mutations that reduce the toxin load have little effect because the toxin load is transient and disappears once the underdominant constructs (loci C and D) are fixed. Unlike [72], we recommend that both transgenic alleles C and D carry a dominant payload fitness cost s_p , which better resists the evolution of a compensatory mutation within the payload genotype due to the existence of multiple gene copies.

In the main text, we have focused on strategies intended to alter a population, rather than eliminate it. In the Appendix 5.B, we also consider daisy quorum drives designed for population suppression. If the payload fitness cost causes local populations to decline to extinction (see Figure S2), wildtype individuals from neighbouring patches will eventually recolonize and return the system to its original state. By contrast, strategies that involve population alteration persist for longer because migrants continue to face competition and the toxin load when interbreeding with individuals in the target population carrying the underdominant construct (see Figure S1). Thus the engineered construct investigated here is most suited for population alteration, rather than population suppression, whenever a persistent and localized effect is sought in a non-isolated population.

Implementing gene drives, accidentally or on purpose, can have long lasting effects on biological systems. It is clear that gene drives have the potential to solve many ecological challenges, however, all consequences of triggering a gene drive cannot be foreseen. Models like the one we present here are important first steps, but they can only be used as a rough guide, while more experiments and refined models are required before release of gene drives in the wild [162].

In summary, we modeled a potential solution to the spillover and persistence problems posed by gene-drive systems using a combination of homing-based gene drive and two-locus underdominance. Lab-based research is needed to determine the feasibility of developing an underdominant daisy-quorum drive with the features needed to reduce the risk of spillover (e.g., Figure 5.5). In particular, controlled experiments are needed to determine the likelihood of mutations reversing the payload fitness cost (reducing persistence) and the risk of unintended side effects, such as self-driving elements arising from the daisy-chain genes (increasing spillover risk).

5.5 Acknowledgements

We would like to thank the Otto lab and two previous reviewers for helpful feedback on this project. This work was supported by a Discovery grant from the Natural Sciences and Engineering Research Council of Canada to SPO (NSERC RGPIN-2016-03711), stipend support to FJHdH from an NSERC CREATE grant (BIOS2), and an Agence Nationale de la Recherche grant of France from the Programme Jeunes Chercheuses et Jeunes Chercheurs to FD (ANR-19-CE45-0009-01 “TheoGeneDrive”).

Appendices

5.A Daisy quorum drive

5.A.1 Daisy-chain length

The length of the daisy chain is an important determinant of the maximum attainable frequency of the payload fitness cost. Here we illustrate this effect for daisy chains of different lengths by tracking the maximal frequency attained by the last element in the chain. As the number of loci involved in the daisy chain increases, there is a stronger and longer force pushing the last drive element to a higher frequency (Figure S1). The inset figure shows an example of the dynamics of a daisy-chain system with three drive loci.

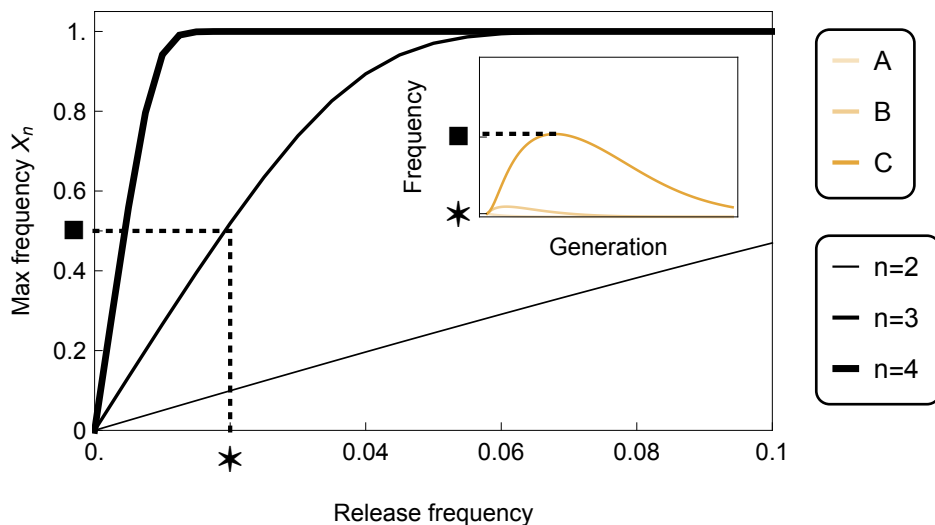


Figure S1: Maximum frequency of the last element in an n length daisy-chain construct like locus **A** and **B** in Table 5.1 in a single population. All elements of the daisy-chain carry a drive load $s_d = 0.05$. The frequency of the n^{th} transgenic allele is denoted by X_n . Other parameters are $\delta = 1$, $r = 0.5$.

5.A.2 Mating and gamete production

Here we enumerate the various possible matings and gametes produced for the 4-locus daisy-quorum system explored in the main text (Table 5.1). Table S1 gives the mating table for the dynamics at loci **A** and **B** (the daisy-chain component). Table S2 provides the mating table for the dynamics at loci **C**

and **D** (the underdominant component).

Gamete 1	Gamete 2	Fitness	Freq	Gametes produced			
				<i>ab</i>	<i>aB</i>	<i>Ab</i>	<i>AB</i>
<i>ab</i>	<i>ab</i>	1	X_{ab}^2	1			
<i>ab</i>	<i>aB</i>	$(1 - s_d)$	$2X_{ab}X_{aB}$	$\frac{1}{2}$	$\frac{1}{2}$		
<i>ab</i>	<i>Ab</i>	$(1 - s_d)$	$2X_{ab}X_{Ab}$	$\frac{1}{2}$		$\frac{1}{2}$	
<i>ab</i>	<i>AB</i>	$(1 - s_d)^2$	$2X_{ab}X_{AB}$	$\frac{1}{2}(1 - \delta)(1 - r)$	$\frac{1}{2}(\delta + (1 - \delta)r)$	$\frac{1}{2}(1 - \delta)r$	$\frac{1}{2}(1 - (1 - \delta)r)$
<i>aB</i>	<i>aB</i>	$(1 - s_d)^2$	X_{aB}^2		1		
<i>aB</i>	<i>Ab</i>	$(1 - s_d)^2$	$2X_{aB}X_{Ab}$	$\frac{1}{2}(1 - \delta)r$	$\frac{1}{2}(1 - (1 - \delta)r)$	$\frac{1}{2}(1 - \delta)(1 - r)$	$\frac{1}{2}(\delta + (1 - \delta)r)$
<i>aB</i>	<i>AB</i>	$(1 - s_d)^3$	$2X_{aB}X_{AB}$		$\frac{1}{2}$	$\frac{1}{2}$	
<i>Ab</i>	<i>Ab</i>	$(1 - s_d)^2$	X_{Ab}^2			1	
<i>Ab</i>	<i>AB</i>	$(1 - s_d)^3$	$2X_{Ab}X_{AB}$			$\frac{1}{2}(1 - \delta)$	$\frac{1}{2}(1 + \delta)$
<i>AB</i>	<i>AB</i>	$(1 - s_d)^4$	X_{AB}^2				1

Table S1: Mating table for loci **A** and **B**, illustrating the gametes that come together to make a diploid individual (first two columns), their fitness (third column), frequency at birth (fourth column), and gametes produced (last four columns).

Gamete 1	Gamete 2	Fitness	Freq	Gametes produced			
				<i>cd</i>	<i>cD</i>	<i>Cd</i>	<i>CD</i>
<i>cd</i>	<i>cd</i>	1	X_{cd}^2	1			
<i>cd</i>	<i>cD</i>	$(1 - s_t)(1 - s_p)$	$2X_{cd}X_{cD}$	$\frac{1}{2}$	$\frac{1}{2}$		
<i>cd</i>	<i>Cd</i>	$(1 - s_t)(1 - s_p)$	$2X_{cd}X_{Cd}$	$\frac{1}{2}$		$\frac{1}{2}$	
<i>cd</i>	<i>CD</i>	$(1 - s_p)$	$2X_{cd}X_{CD}$	$\frac{1}{2}(1 - r)$	$\frac{1}{2}r$	$\frac{1}{2}r$	$\frac{1}{2}(1 - r)$
<i>cD</i>	<i>cD</i>	$(1 - s_t)(1 - s_p)$	X_{cD}^2		1		
<i>cD</i>	<i>Cd</i>	$(1 - s_p)$	$2X_{cD}X_{Cd}$	$\frac{1}{2}r$	$\frac{1}{2}(1 - r)$	$\frac{1}{2}(1 - r)$	$\frac{1}{2}r$
<i>cD</i>	<i>CD</i>	$(1 - s_p)$	$2X_{cD}X_{CD}$		$\frac{1}{2}$	$\frac{1}{2}$	
<i>Cd</i>	<i>Cd</i>	$(1 - s_t)(1 - s_p)$	X_{Cd}^2			1	
<i>Cd</i>	<i>CD</i>	$(1 - s_p)$	$2X_{Cd}X_{CD}$			$\frac{1}{2}$	$\frac{1}{2}$
<i>CD</i>	<i>CD</i>	$(1 - s_p)$	X_{CD}^2				1

Table S2: Mating table for locus **C** and **D**, illustrating the gametes that come together to make a diploid individual (first two columns), their fitness (third column), frequency at birth (fourth column), and gametes produced (last four columns).

5.A.3 Full dynamics

By assuming a constant initial level of drive δ_C , the invasion analysis in the main text (Equation 5.4) ignores changes to the genetic associations between the daisy-chain and underdominant components. Here we compare the predictions from this invasion analysis to full numerical analyses.

Figure S2 illustrates the dynamics for the full four-locus model of Table 5.1, using the default parameters (Table 5.3) with either a low toxin load ($s_t = 0.1$, panels A,B) or a high toxin load ($s_t = 0.9$, panels C,D). We initiate the dynamics by introducing the construct **ABCD** at a frequency that is either slightly below ($f_0 = 0.015$ left panels) or above ($f_0 = 0.02$ right panels) the initial frequency needed to drive the underdominant component (blue curve) to fixation.

Drive only occurs in individuals carrying the driver allele B that are also heterozygous for a driven allele (C or D). The frequency of such individuals is given by the solid orange curve, which starts at one when the alleles are introduced together in the **ABCD** construct. The driver phenotype thus starts out high enough (orange curve above the threshold, equation 5.5, black line) to lead the underdominant construct to rise in frequency (blue curve initially rising), as predicted by the two-locus invasion analysis. Simulations with weaker drive were conducted to confirm that the underdominant construct declines initially if the strength of drive is below the threshold needed to ensure $\lambda > 1$ (here, $\delta \leq 0.49$).

Over time, recombination breaks down the strong initial genetic association between the driver (B) and driven alleles (C or D). If the underdominant construct starts at too low a frequency ($f_0 = 0.015$) and is not driven to a high enough frequency before the drive dissipates, then the underdominant construct is lost (left hand panels). Starting at a slightly higher initial frequency ($f_0 = 0.02$), the underdominant rises above the unstable equilibrium and enters the basin of attraction to the fixation equilibrium where $X_C = X_D = 1$ (blue curve). In the terminology of [16], the underdominant construct initially rides a “Fisherian” wave, rising from low initial frequency due to gene drive, but then completes the transition to fixation by riding a “bistable” wave if it passes the unstable equilibrium before drive is exhausted.

To guide the eye, the green curve gives the position of the unstable equilibrium for the two-locus underdominant construct if the drive phenotype were fixed at its current value ($X_B^2 + 2X_B X_b$, given by the orange solid curve). While this is not precisely the separatrix in the full four-locus system, we can see that the underdominant construct must be introduced at a high enough frequency that it remains near or above the unstable equilibrium in the two-locus system to rise to fixation (blue must remain high enough relative to the green curve).

5.A.4 Multiplicative payload fitness cost

In the main text we focus on fitnesses with a dominant cost (Table 5.2), in which case fixation of the CD haplotype is locally stable. If we assume different cost regimes (e.g., multiplicative or recessive effects on individual fitness), then fixation of the CD haplotype is no longer a stable equilibrium, as described below.

Table S3 gives individual fitness under a multiplicative regime for the payload fitness cost.

The governing equations for the underdominant component with a multiplicative payload fitness cost (Table S3) are:

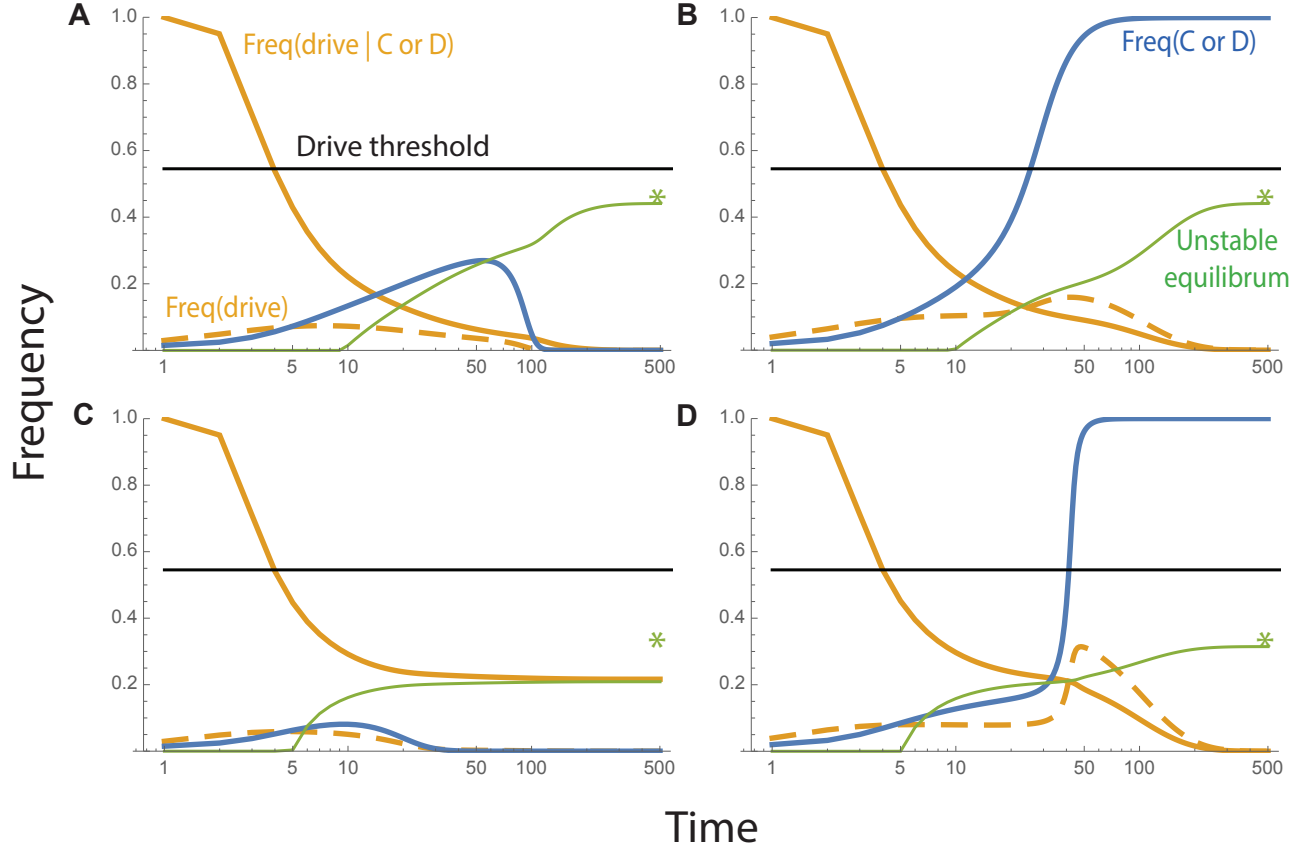


Figure S2: The dynamics of the full four-locus daisy quorum drive with low (top row) or high (bottom row) toxin loads. The frequency of the underdominant allele C or D is shown in blue, which rises to fixation only when started at a high enough initial frequency (right column). The frequency of the drive phenotype (homozygotes or heterozygotes carrying allele B) is shown among carriers of the underdominant alleles C or D (solid orange) and in the full population (dashed orange). The green curve shows the position of the unstable equilibrium of the underdominant construct, given the current frequency of the drive phenotype (solid orange), with the star indicating the position once drive has disappeared. Parameters: $\delta = 0.9$, $r = 0.5$, $s_d = 0.02$, $s_p = 0.1$, $s_t = 0.1$ (panels A,B) or $s_t = 0.9$ (panels C,D), with $f_0 = 0.015$ (panels A,C) or $f_0 = 0.02$ (panels B,D).

	cd	cD	Cd	CD
cd	1	$(1 - s_t)(1 - s_p)^{1/4}$	$(1 - s_t)(1 - s_p)^{1/4}$	$(1 - s_p)^{1/2}$
cD	$(1 - s_t)(1 - s_p)^{1/4}$	$(1 - s_t)(1 - s_p)^{1/2}$	$(1 - s_p)^{1/2}$	$(1 - s_p)^{3/4}$
Cd	$(1 - s_t)(1 - s_p)^{1/4}$	$(1 - s_p)^{1/2}$	$(1 - s_t)(1 - s_p)^{1/2}$	$(1 - s_p)^{3/4}$
CD	$(1 - s_p)^{1/2}$	$(1 - s_p)^{3/4}$	$(1 - s_p)^{3/4}$	$1 - s_p$

Table S3: Expression of payload fitness cost (s_p) is multiplicative within and between loci

$$\begin{aligned}
 X'_{cd} &= \frac{1}{W} (X_{cd}X_{cD} \sqrt[4]{1 - s_p}(1 - s_t) + X_{cd}X_{Cd} \sqrt[4]{1 - s_p}(1 - s_t) + X_{cd}X_{CD}(1 - r) \sqrt[2]{1 - s_p} + X_{cd}X_{cd} + X_{cD}X_{Cd}r \sqrt[2]{1 - s_p}) \\
 X'_{cD} &= \frac{1}{W} (X_{cd}X_{CD}r \sqrt[2]{1 - s_p} + X_{cd}X_{cD}(1 - s_t) \sqrt[4]{1 - s_p} + X_{cD}X_{Cd}(1 - r) \sqrt[2]{1 - s_p} + X_{cD}X_{CD} \sqrt[3]{1 - s_p}) \\
 X'_{Cd} &= \frac{1}{W} (X_{cd}X_{CD}r \sqrt[2]{1 - s_p} + X_{cd}X_{Cd}(1 - s_t) \sqrt[4]{1 - s_p} + X_{cD}X_{Cd}(1 - r) \sqrt[2]{1 - s_p} + X_{Cd}X_{CD} \sqrt[3]{1 - s_p} +) \\
 X'_{CD} &= \frac{1}{W} (X_{CD}X_{cd}(1 - r) \sqrt[2]{1 - s_p} + X_{CD}X_{cD} \sqrt[3]{1 - s_p} - X_{CD}X_{CD}(1 - s_p) + X_{cD}X_{Cd}r \sqrt[2]{1 - s_p} + X_{cD}X_{CD} \sqrt[3]{1 - s_p})
 \end{aligned}
 \tag{5.6}$$

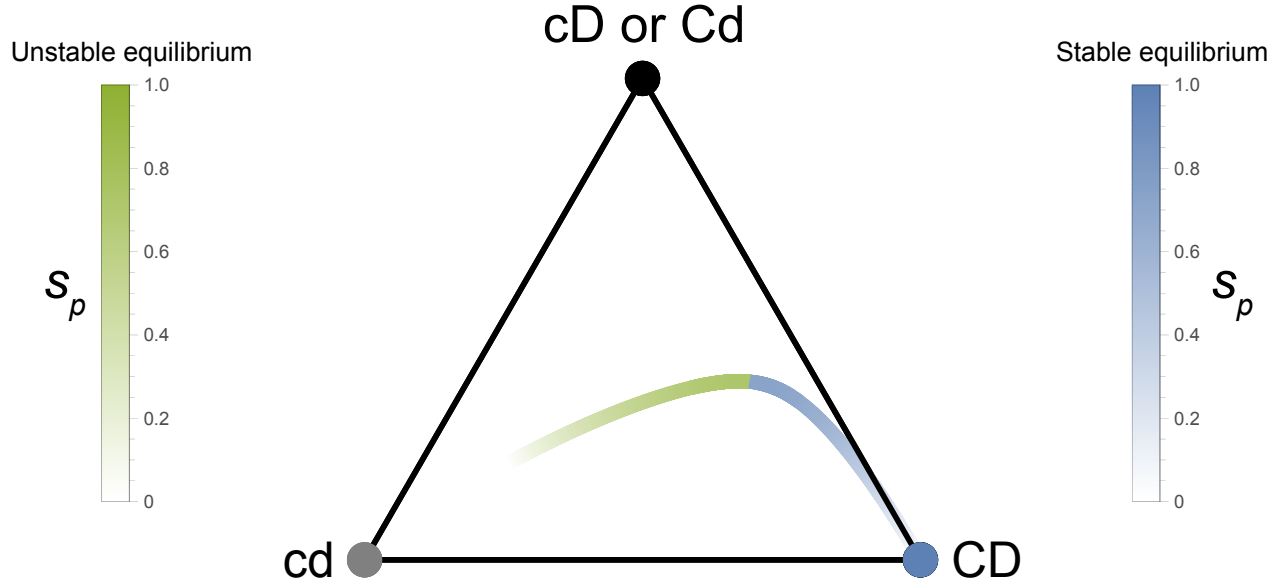


Figure S3: Equilibria with multiplicative expression of the payload fitness cost (Table S3). The coloured curves indicate the internal unstable (green) and stable (blue) equilibria for different payload fitness cost values s_p . For a given payload fitness cost, s_p , starting to the right of the green point will lead the system to approach the blue point with the same shade (s_p value). Other parameters: toxin load $s_t = 1$ and recombination $r = \frac{1}{2}$.

Analysis of the equilibria of this underdominant system reveals that fixation of the wildtype alleles ($X_{cd} = 1$) is stable, but fixation of the underdominant construct is not ($X_{CD} = 1$). In addition, there are two internal equilibria, one of which is unstable (green) and one stable (blue), as illustrated in Figure S3. The stable and unstable equilibrium move towards each other as the payload fitness cost increases. This means that a drive requires a higher initial frequency to invade and will spread to a lower equilibrium frequency as the payload fitness cost increases in strength.

The same invasion analysis as in section 3.3 reveals that the conditions for the drive rate δ_c needed for the transgenic alleles C and D to spread are:

$$\delta_c > \min\left[\frac{\frac{1}{\sqrt[4]{1-s_p}} - (1-s_t)}{1-s_t}, \frac{\frac{\sqrt{\sqrt{1-s_p} - (1-s_p) + r(2-2s_p - \sqrt{1-s_p})}}{\sqrt{1-s_p}} - r}{1-r}\right] \quad (5.7)$$

This shows that the maximum payload fitness cost that can be carried by an invading two-locus underdominant construct with multiplicative cost is $s_p = 15/16$ with maximal drive rate $\delta_c = 1$ and $s_t \approx 0$.

5.A.5 Recessive payload fitness cost

Table S4 presents the relative fitness of individuals when the payload fitness cost is recessive, in which case an individual needs two copies of C and two copies of D alleles before expressing the payload fitness cost.

	<i>cd</i>	<i>cD</i>	<i>Cd</i>	<i>CD</i>
<i>cd</i>	1	$1 - s_t$	$1 - s_t$	1
<i>cD</i>	$1 - s_t$	$1 - s_t$	1	1
<i>Cd</i>	$1 - s_t$	1	$1 - s_t$	1
<i>CD</i>	1	1	1	$1 - s_p$

Table S4: Expression of payload fitness cost (s_p) is recessive

The governing equations for the underdominant component with a recessive payload fitness cost are:

$$\begin{aligned}
X'_{cd} &= \frac{1}{\bar{W}} X_{cd} X_{cD} (1 - s_t) + X_{cd} X_{CD} (1 - r) + X_{cd}^2 + X_{cD} X_{Cd} r \\
X'_{cD} &= \frac{1}{\bar{W}} X_{cD} X_{cd} (1 - s_t) + X_{cD}^2 (1 - s_t) + X_{cD} X_{Cd} (1 - r) + X_{CD} X_{cd} r + X_{CD} X_{cD} \\
X'_{Cd} &= \frac{1}{\bar{W}} X_{cd} X_{CD} r + X_{cd} X_{Cd} (1 - s_t) + X_{Cd} X_{cD} (1 - r) + X_{Cd}^2 (1 - s_t) + X_{Cd} X_{CD} \\
X'_{CD} &= \frac{1}{\bar{W}} X_{CD} X_{cd} (1 - r) + X_{CD} X_{Cd} + X_{CD}^2 (1 - s_p) + X_{cD} X_{Cd} r + X_{cD} X_{CD}
\end{aligned} \tag{5.8}$$

In Figure S4, we illustrate the location of the equilibria of this model for $s_t = 1.0$ and $s_t = 0.1$, where there are again two internal equilibrium (one unstable in green and one stable in blue). We see that similar to the analysis in the main text for a dominant payload fitness cost, an increase in the payload fitness cost or a reduction in the toxin load moves the unstable equilibrium further from the wildtype (Figure S4 (a,b)).

The results of the invasion analysis show that this construct can invade when:

$$\delta_c > \min\left[\frac{s_t}{1 - s_t}, \frac{\sqrt{r}}{\sqrt{r} + 1}\right], \tag{5.9}$$

which is satisfied when drive is strong enough relative to the toxin load.

5.B Individual-based simulations with discrete patches

Individual-based simulations (implemented in C^{++}) were developed to explore whether our results hold for finite population sizes. We use the same fitness regime as in the main text (dominant payload fitness cost expression). We implement the following life cycle in a single isolated population with carrying capacity K and then expand to consider migration among 101 patches in a linear stepping-stone array with the construct introduced into the central patch:

1. Calculate the expected frequency of genotypes predicted under the deterministic model.
2. Calculate mean fitness, \bar{W} , and draw from a Poisson random variate with rate parameter $\lambda = \bar{W} \times F \times N_t$ the number of offspring in the next generation N_{t+1}^o . Here, F represents the average number of offspring per individual. Density dependence with a hard carrying capacity was then imposed, so that the total number of offspring in the next generation was set to $N_{t+1} = \min[N_{t+1}^o, K]$.

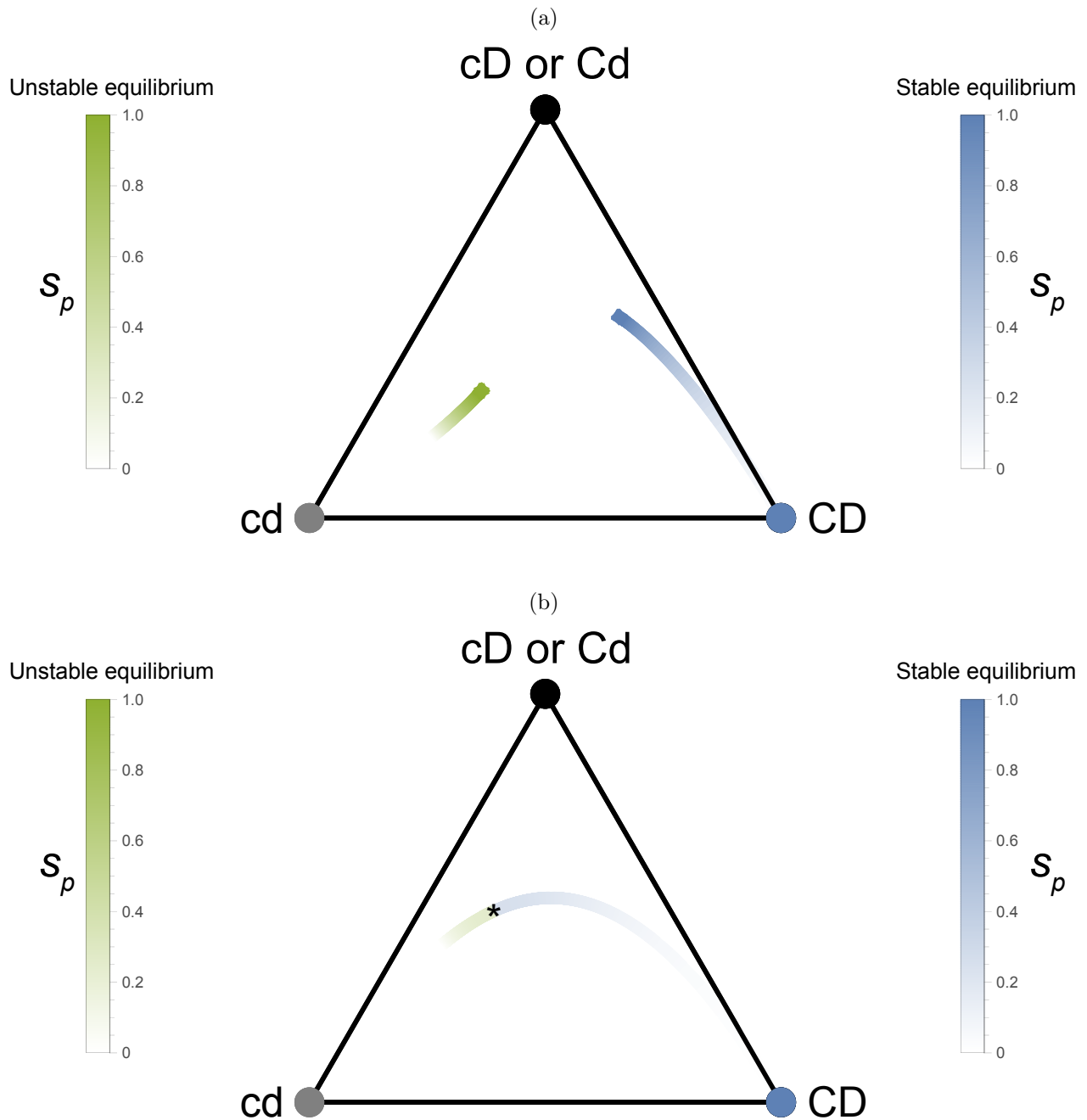


Figure S4: Equilibria with recessive expression of the payload fitness cost (Table S4) for (a) $s_t = 1.0$ and (b) $s_t = 0.1$. The coloured curves indicate the internal unstable (green) and stable (blue) equilibria for different payload fitness cost values s_p with $r = \frac{1}{2}$. For a given payload fitness cost, s_p , starting to the right of the green point will lead the system to approach the blue point with the same shade (s_p value). In panel A, the internal equilibria always exist, but in panel B they do not exist if the payload fitness cost is too strong relative to the toxin load (becoming complex at the * for $s_p \geq 0.25$ with $s_t = 0.1$).

- Genotypes were then drawn for the N_{t+1} offspring from a multinomial distribution with the deterministic predicted genotype frequencies.

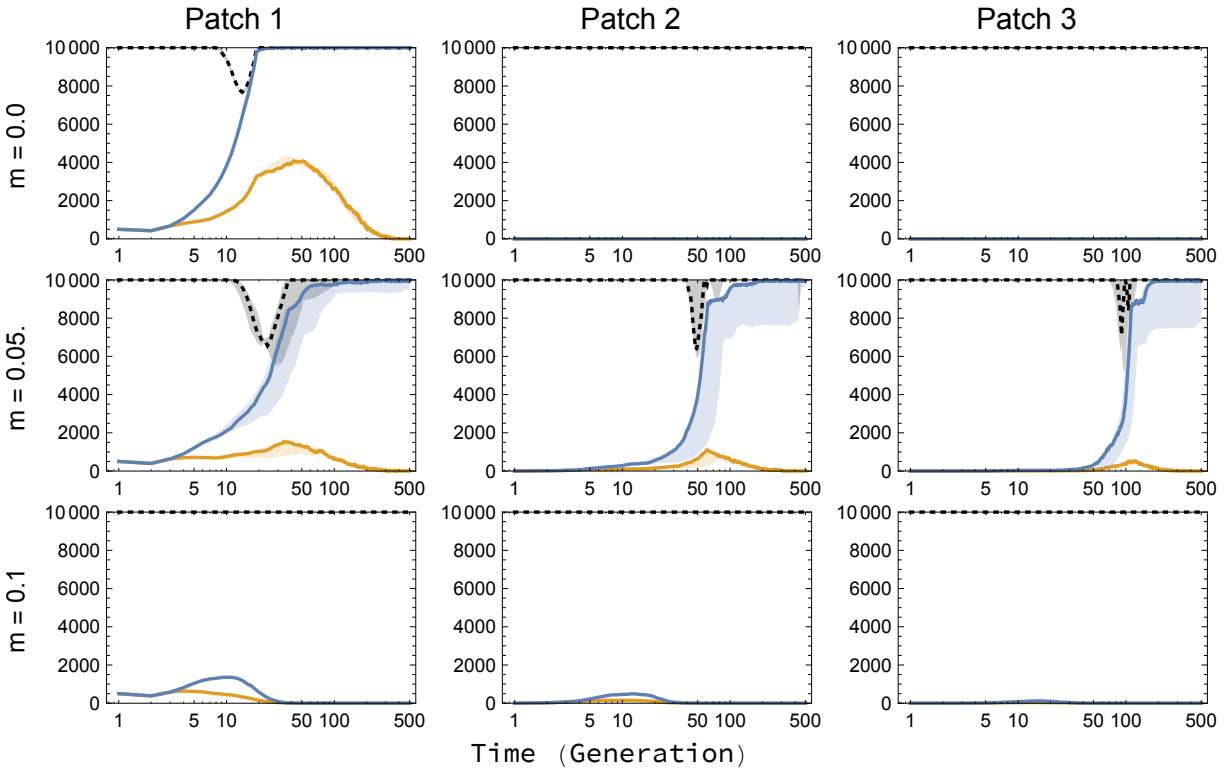


Figure S1: Population alteration with daisy quorum drive in finite populations. Black dotted curve is the total population size at time t , blue solid curve represents the number of individuals carrying the payload allele, and orange solid curve represents the number of individuals carrying the B allele. In this case, fertility is set to $F = 1.2$ and the payload fitness cost to $s_p = 0.1$, so that subpopulations can persist even if the underdominant construct fixes. Each patch has a carrying capacity of $K = 10000$. Rows have a migration probability of 0, 0.05, 0.1 from top to bottom. Drive is released on the left. Results are shown for 50 replicates. The solid curves represent the median of the replicates, and the shaded regions the first and third quartiles.

We repeat these steps for t_{max} generations and show results in Figure S1 and S2 for $F = 1.2$ and $F = 1.05$, respectively. With $F = 1.2$, and a payload fitness cost of $s_p = 0.1$, the average number of offspring per individual carrying the payload remains above unity: $\lambda = 0.9 \times 1.2 = 1.08$. However, the toxin load $s_t = 0.9$ causes a dip in population size (black dashed curve in Figure S1) as the payload increases in frequency. Once the payload has fixed, the toxin load is no longer expressed, and the population recovers to its carrying capacity, regardless of the migration rate. For $F = 1.05$, however, the expected number of offspring per parent is $\lambda = 0.9 \times 1.05 = 0.945$, which is less than one, meaning that the population is unable to sustain itself once the payload becomes common and declines to extinction in the absence of migration (Figure S2). With migration, however, population suppression is only transient. As the population declines in size, proportionally more migrants enter the patch than when the population is large. This proportionally high inflow of wildtype individuals prevents fixation of the payload and eventually helps the population recover to its carrying capacity

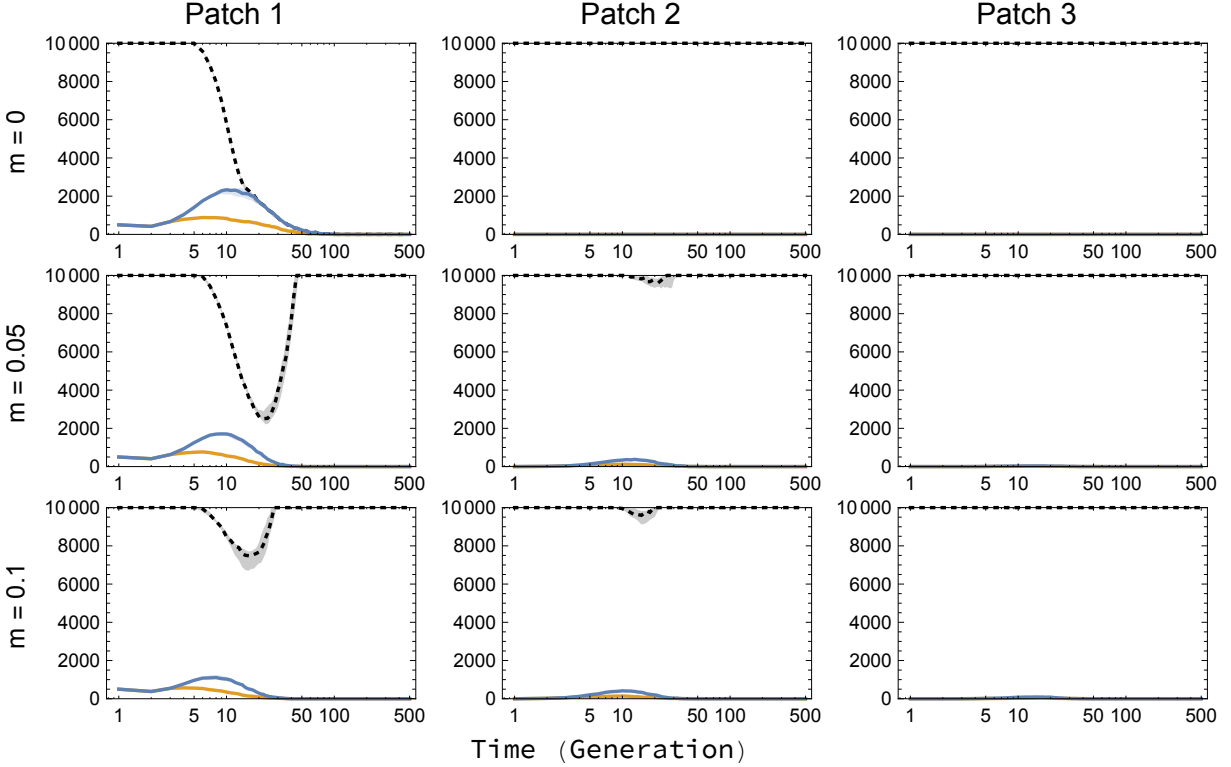


Figure S2: Population suppression with daisy quorum drive in finite populations. Identical to Figure S1 but with fertility of $F = 1.05$, which is too low for a subpopulation to replace itself once the underdominant construct has fixed.

with only wildtype individuals. Thus, daisy quorum drive is only transiently effective when the goal is population suppression in the face of gene flow.

5.C Numerical simulations in continuous space

5.C.1 Relationship between discrete and continuous space

To consider the dynamics of daisy quorum drive in a continuous spatial environment, we first relate the migration rate among discrete patches to the diffusion rate, by decreasing the distance between patches. From section 5.3.4, we note that the frequency of gametes of type i after migration for a non-boundary population j is:

$$p_i^{j'} = (1 - m) p_i^j + \frac{1}{2} m p_i^{j-1} + \frac{1}{2} m p_i^{j+1}. \quad (5.10)$$

By dividing space into small patches of size Δx and then tracking changes due to migration, reproduction, and death in time units of Δt , we derive a continuous model in space and time by allowing Δx and Δt to approach zero. From equation (5.10), the frequency of gametes of type i in patch x at time t changes according to:

$$p_i(t + \Delta t, x) = (1 - m) p_i(t, x) + \frac{1}{2} m p_i(t, x - \Delta x) + \frac{1}{2} m p_i(t, x + \Delta x) \quad (5.11)$$

$$\iff \frac{p_i(t + \Delta t, x) - p_i(t, x)}{\Delta t} = \frac{m(\Delta x)^2}{2\Delta t} \frac{p_i(t, x - \Delta x) - 2p_i(t, x) + p_i(t, x + \Delta x)}{(\Delta x)^2}. \quad (5.12)$$

The last equation clarifies the relationship between the migration rate and the diffusion rate:

$$\mathcal{D} = \frac{m(\Delta x)^2}{2\Delta t}.$$

In the continuous limit, $(\Delta x)^2/\Delta t$ tends to a finite value meaning that for a fixed value of the migration rate m , the resulting diffusion rate \mathcal{D} is constant. When comparing discrete and continuous environments (i.e., varying Δx), we fix the diffusion rate across space, so that the migration rate between patches declines as the patches move further apart (increased spatial steps). We use Neumann boundary conditions as in section 5.3.4; that is, the boundary populations only give and receive migrants from the interior.

5.C.2 Numerical simulations across a homogeneous landscape

We simulate the propagation of alleles A , B , C , and D over a one-or two-dimensional continuous homogeneous domain with two different initial conditions. We consider the introduction of either fully modified individuals ($ABCD$, light blue) or partially modified individuals ($abCD$, dark blue), replacing all individuals in a small central region (locations 90-110 over a domain of length 200). The rest of the domain is full of wild-type individuals ($abcd$), and all the other genotypes are absent at $t = 0$.

As time passes, the cargo wave (measuring either X_C or X_D , which remain equal in frequency) approaches a constant speed. The wave either continues to move outward with a positive asymptotic speed (top row of Figure S1), or the wave collapses with a negative asymptotic speed (bottom row). In our simulations, increasing the payload fitness cost decreases the asymptotic wave speed for a given toxin load.

Importantly, the asymptotic wave speed does not depend on the daisy chain (loci **A** and **B**), which stops driving the cargo as time passes (as seen in panels B and F of Figure S1, where the drive alleles are more restricted in space than the cargo). In fact, the same asymptotic speed is reached in simulations started without a daisy chain and only alleles C and D (darker curves in right column of Figure S1).

We investigated the hypothesis that the asymptotic behaviour transitions from a spreading wave (positive speed) to a collapsing wave (negative speed) when the unstable equilibrium of the two-locus underdominant system crosses the midpoint where $p_C = p_c = p_D = p_d = 1/2$ (see Equation 5.3), as occurs in the one-locus underdominant system [15], but we found an imperfect match. For example, in Figure 5.6 with $s_t = 0.9$ and $r = 1/2$, the unstable equilibrium crosses the midpoint when $s_p = 0.44$, but the threshold for the payload fitness cost needed for a negative asymptotic wave speed occurs slightly below $s_p = 0.3$. Thus Equation (5.3) does not serve to determine the threshold for negative wave speeds, presumably because of the asymmetries in how the toxin load and payload fitness cost interact (Table 5.2).

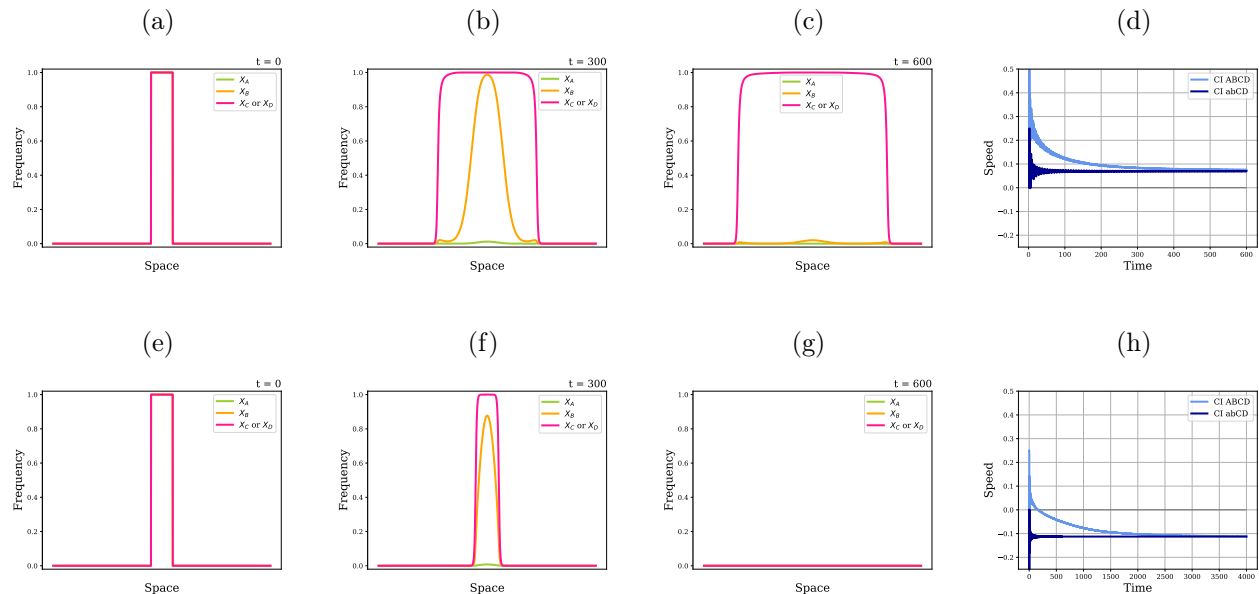


Figure S1: **Spread of introduced genotypes in a one-dimensional continuous space.** The first three panels in each row represent allele frequencies in space at times $t = 0, 300, 600$. The last panel is the speed of the X_C and X_D wave as a function of time. The rows illustrate simulations with a low payload fitness cost (top row, $s_p = 0.1$) or a high payload fitness cost (bottom row, $s_p = 0.5$). The simulations introduced fully modified individuals ($ABCD$) in the centre of the range, except that the last column compares the wave speed for this case (light blue) to the case where only the underdominant component ($abCD$) is introduced. Parameters are as follows: $r = 0.5$ (recombination rate), $s_d = 0.02$ (drive load), $s_t = 0.9$ (toxin load), $\mathcal{D} = 0.2$ (diffusion rate), $\delta = 0.9$ (drive rate), $T = 600$ (final time), and $L = 200$ (length of the spatial domain). Time and space were subdivided into ten steps each to approximate continuous time and space (i.e., using a spatial step size of 0.1).

5.C.3 Numerical simulations across a heterogeneous landscape

While daisy quorum drive cannot be stably maintained in a restricted region when space is homogeneous, spatial heterogeneity, such as variation in the viscosity of the environment to movement, can stabilize the system.

First, consider an example where the landscape is easy to traverse in the centre of the range (spatial steps of size 1) but challenging to traverse outside of this region (spatial steps of size 2). With $s_p = 0.1$, the critical step size for an expanding wave is around 1.7 (Figure 5.6). Thus, in this case, the underdominant construct propagates outwards in a wave only until reaching the area where dispersal between sites becomes challenging (Figure S2).

In the main text, we consider random spatial heterogeneity, where dispersal rates between neighboring sites were determined by randomly placing 400 boundaries across a spatial domain of length 200 (mean step size of one). The step size separating adjacent regions then determines local dispersal rates, with higher dispersal between neighboring sites when the boundaries are closer. As seen in Figure 5.6 (bottom row), spatial heterogeneity can also stabilize waves that are collapsing once drive has been exhausted. In this case, regions of low dispersal prevent wildtype alleles from displacing the construct further, and the underdominant construct can be stably maintained in a localized region.

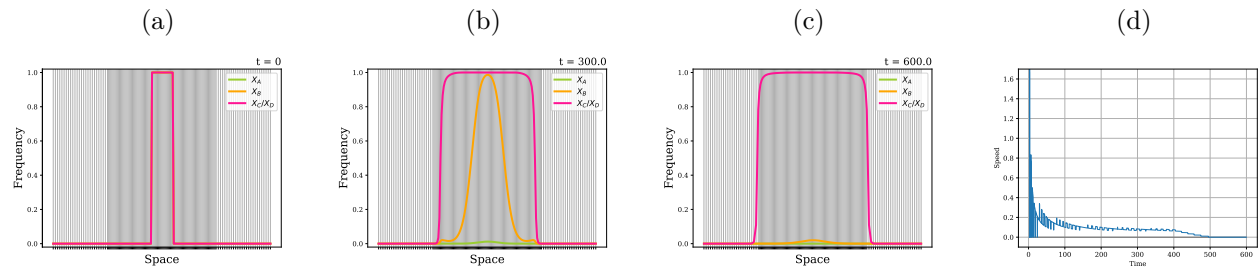


Figure S2: The first three panels A, B and C represent allele frequencies in space at time $t = 0, 300, 600$. The last panel D describes the speed of the cargo wave, measuring the frequency of either C or D , i.e. the speed of the pink curve in the first three graphs. The domain is heterogeneous, with steps of size 1 (middle half) and steps of size 2 (outer half). Parameters are otherwise the same as Figure S1.

Bibliography

- [1] Zach Adelman et al. “Rules of the Road for Insect Gene Drive Research and Testing”. In: *Nature Biotechnology* 35.8 (Aug. 2017), pp. 716–718. ISSN: 1546-1696. DOI: 10.1038/nbt.3926.
- [2] Omar S. Akbari et al. “Novel Synthetic Medea Selfish Genetic Elements Drive Population Replacement in *Drosophila*; a Theoretical Exploration of Medea-Dependent Population Suppression”. In: *ACS Synthetic Biology* 3.12 (Dec. 2014), pp. 915–928. DOI: 10.1021/sb300079h.
- [3] Omar S. Akbari et al. “Safeguarding Gene Drive Experiments in the Laboratory”. In: *Science* 349.6251 (Aug. 2015), pp. 927–929. DOI: 10.1126/science.aac7932.
- [4] Luke S. Alphey et al. “Standardizing the Definition of Gene Drive”. In: *Proceedings of the National Academy of Sciences* 117.49 (Dec. 2020), pp. 30864–30867. DOI: 10.1073/pnas.2020417117.
- [5] Nina Alphey and Michael B. Bonsall. “Interplay of Population Genetics and Dynamics in the Genetic Control of Mosquitoes”. In: *Journal of The Royal Society Interface* 11.93 (Apr. 2014), p. 20131071. ISSN: 1742-5689, 1742-5662. DOI: 10.1098/rsif.2013.1071.
- [6] Philipp M. Altrock, Arne Traulsen, and Floyd A. Reed. “Stability Properties of Underdominance in Finite Subdivided Populations”. In: *PLoS computational biology* 7.11 (Nov. 2011), e1002260. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1002260.
- [7] Philipp M. Altrock et al. “Using Underdominance to Bi-Stably Transform Local Populations”. In: *Journal of Theoretical Biology* 267.1 (Nov. 2010), pp. 62–75. ISSN: 1095-8541. DOI: 10.1016/j.jtbi.2010.08.004.
- [8] Jing An, Christopher Henderson, and Lenya Ryzhik. *Pushed, Pulled and Pushmi-Pullyu Fronts of the Burgers-FKPP Equation*. Aug. 2021. DOI: 10.48550/arXiv.2108.07861. arXiv: 2108.07861 [math].
- [9] Jing An, Christopher Henderson, and Lenya Ryzhik. *Quantitative Steepness, Semi-FKPP Reactions, and Pushmi-Pullyu Fronts*. Aug. 2022. DOI: 10.48550/arXiv.2208.02880. arXiv: 2208.02880 [math].
- [10] Peter Armbruster, Robert A. Hutchinson, and Trevor Linvell. “Equivalent Inbreeding Depression under Laboratory and Field Conditions in a Tree-Hole-Breeding Mosquito”. In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 267.1456 (Oct. 2000), pp. 1939–1945. DOI: 10.1098/rspb.2000.1233.
- [11] D. G. Aronson and H. F. Weinberger. “Nonlinear Diffusion in Population Genetics, Combustion, and Nerve Pulse Propagation”. In: *Partial Differential Equations and Related Topics*. Ed. by Jerome A. Goldstein. Lecture Notes in Mathematics. Berlin, Heidelberg: Springer, 1975, pp. 5–49. ISBN: 978-3-540-37440-4. DOI: 10.1007/BFb0070595.

- [12] D. G. Aronson and Hans F Weinberger. “Multidimensional Nonlinear Diffusion Arising in Population Genetics”. In: *Advances in Mathematics* 30.1 (Oct. 1978), pp. 33–76. ISSN: 0001-8708. DOI: 10.1016/0001-8708(78)90130-5.
- [13] Montie Avery, Matt Holzer, and Arnd Scheel. *Pushed-to-Pulled Front Transitions: Continuation, Speed Scalings, and Hidden Monotonicity*. June 2022. DOI: 10.48550/arXiv.2206.09989. arXiv: 2206.09989 [nlin].
- [14] Rowida Baeshen et al. “Differential Effects of Inbreeding and Selection on Male Reproductive Phenotype Associated with the Colonization and Laboratory Maintenance of *Anopheles Gambiae*”. In: *Malaria Journal* 13.1 (Jan. 2014), p. 19. ISSN: 1475-2875. DOI: 10.1186/1475-2875-13-19.
- [15] N. H. Barton. “The Dynamics of Hybrid Zones”. In: *Heredity* 43.3 (Dec. 1979), pp. 341–359. ISSN: 1365-2540. DOI: 10.1038/hdy.1979.87.
- [16] N. H. Barton and Michael Turelli. “Spatial Waves of Advance with Bistable Dynamics: Cytoplasmic and Genetic Analogues of Allee Effects.” In: *The American Naturalist* 178.3 (Sept. 2011), E48–E75. ISSN: 0003-0147. DOI: 10.1086/661246.
- [17] Andrea Beaghton, Pantelis John Beaghton, and Austin Burt. “Gene Drive through a Landscape: Reaction–Diffusion Models of Population Suppression and Elimination by a Sex Ratio Distorter”. In: *Theoretical Population Biology* 108 (Apr. 2016), pp. 51–69. ISSN: 0040-5809. DOI: 10.1016/j.tpb.2015.11.005.
- [18] Andrea K. Beaghton et al. “Gene Drive for Population Genetic Control: Non-Functional Resistance and Parental Effects”. In: *Proceedings of the Royal Society B: Biological Sciences* 286.1914 (Oct. 2019), p. 20191586. DOI: 10.1098/rspb.2019.1586.
- [19] Céline Bellard, Phillip Cassey, and Tim M. Blackburn. “Alien Species as a Driver of Recent Extinctions”. In: *Biology Letters* 12.2 (Feb. 2016), p. 20150623. ISSN: 1744-957X. DOI: 10.1098/rsbl.2015.0623.
- [20] Jean Bérard and Jean-Baptiste Gouéré. “Brunet-Derrida Behavior of Branching-Selection Particle Systems on the Line”. In: *Communications in Mathematical Physics* 298.2 (Sept. 2010), pp. 323–342. ISSN: 1432-0916. DOI: 10.1007/s00220-010-1067-y.
- [21] Federica Bernardini et al. “Site-Specific Genetic Engineering of the *Anopheles Gambiae* Y Chromosome”. In: *Proceedings of the National Academy of Sciences* 111.21 (May 2014), pp. 7600–7605. DOI: 10.1073/pnas.1404996111.
- [22] Daniela Bertacchi and Fabio Zucca. *Recent Results on Branching Random Walks*. May 2018. DOI: 10.48550/arXiv.1104.5085. arXiv: 1104.5085 [math].
- [23] Guowu Bian et al. “Wolbachia Invades *Anopheles Stephensi* Populations and Induces Refractoriness to Plasmodium Infection”. In: *Science* 340.6133 (May 2013), pp. 748–751. DOI: 10.1126/science.1236192.
- [24] Ethan Bier. “Gene Drives Gaining Speed”. In: *Nature Reviews Genetics* 23.1 (Jan. 2022), pp. 5–22. ISSN: 1471-0064. DOI: 10.1038/s41576-021-00386-0.
- [25] Aysegul Birand et al. “Gene Drives for Vertebrate Pest Control: Realistic Spatial Modelling of Eradication Probabilities and Times for Island Mouse Populations”. In: *Molecular Ecology* 31.6 (Mar. 2022), pp. 1907–1923. ISSN: 0962-1083. DOI: 10.1111/mec.16361.
- [26] Gabriel Birzu, Oskar Hallatschek, and Kirill S. Korolev. “Fluctuations Uncover a Distinct Class of Traveling Waves”. In: *Proceedings of the National Academy of Sciences* 115.16 (Apr. 2018), E3645–E3654. DOI: 10.1073/pnas.1715737115.

- [27] Norwegian Biotechnology Advisory Board. “Statement on Gene Drives”. In: (2017).
- [28] Eric Brunet and Bernard Derrida. “Shift in the Velocity of a Front Due to a Cutoff”. In: *Physical Review E* 56.3 (Sept. 1997), pp. 2597–2604. DOI: 10.1103/PhysRevE.56.2597.
- [29] Éric Brunet and Bernard Derrida. “Effect of Microscopic Noise on Front Propagation”. In: *Journal of Statistical Physics* 103.1 (Apr. 2001), pp. 269–282. ISSN: 1572-9613. DOI: 10.1023/A:1004875804376.
- [30] Anna Buchman et al. “Broad Dengue Neutralization in Mosquitoes Expressing an Engineered Antibody”. In: *PLOS Pathogens* 16.1 (Jan. 2020), e1008103. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1008103.
- [31] Anna Buchman et al. “Synthetically Engineered Medea Gene Drive System in the Worldwide Crop Pest *Drosophila Suzukii*”. In: *Proceedings of the National Academy of Sciences* 115.18 (May 2018), pp. 4725–4730. DOI: 10.1073/pnas.1713139115.
- [32] Anna B. Buchman et al. “Engineered Reciprocal Chromosome Translocations Drive High Threshold, Reversible Population Replacement in *Drosophila*”. In: *ACS synthetic biology* 7.5 (May 2018), pp. 1359–1370. ISSN: 2161-5063. DOI: 10.1021/acssynbio.7b00451.
- [33] James J Bull, Christopher H Remien, and Stephen M Krone. “Gene-Drive-Mediated Extinction Is Thwarted by Population Structure and Evolution of Sib Mating”. In: *Evolution, Medicine, and Public Health* 2019.1 (Jan. 2019), pp. 66–81. ISSN: 2050-6201. DOI: 10.1093/emph/eoz014.
- [34] Austin Burt. “Site-Specific Selfish Genes as Tools for the Control and Genetic Engineering of Natural Populations.” In: *Proceedings of the Royal Society B: Biological Sciences* 270.1518 (May 2003), pp. 921–928. ISSN: 0962-8452. DOI: 10.1098/rspb.2002.2319.
- [35] Austin Burt and Andrea Crisanti. “Gene Drive: Evolved and Synthetic”. In: *ACS Chemical Biology* 13.2 (Feb. 2018), pp. 343–346. ISSN: 1554-8929. DOI: 10.1021/acscchembio.7b01031.
- [36] Ewen Callaway. “‘Gene Drive’ Moratorium Shot down at UN Biodiversity Meeting”. In: *Nature* (Dec. 2016). ISSN: 1476-4687. DOI: 10.1038/nature.2016.21216.
- [37] Ewen Callaway. “Controversial CRISPR ‘Gene Drives’ Tested in Mammals for the First Time”. In: *Nature* 559.7713 (July 2018), pp. 164–164. DOI: 10.1038/d41586-018-05665-1.
- [38] Karl J. Campbell et al. “The next Generation of Rodent Eradications: Innovative Technologies and Tools to Improve Species Specificity and Increase Their Feasibility on Islands”. In: *Biological Conservation*. Special Issue on Tropical Island Conservation: Rat Eradication for Species Recovery 185 (May 2015), pp. 47–58. ISSN: 0006-3207. DOI: 10.1016/j.biocon.2014.10.016.
- [39] Jackson Champer, Anna Buchman, and Omar S. Akbari. “Cheating Evolution: Engineering Gene Drives to Manipulate the Fate of Wild Populations”. In: *Nature Reviews Genetics* 17.3 (Mar. 2016), pp. 146–159. ISSN: 1471-0064. DOI: 10.1038/nrg.2015.34.
- [40] Jackson Champer et al. “A CRISPR Homing Gene Drive Targeting a Haplolethal Gene Removes Resistance Alleles and Successfully Spreads through a Cage Population”. In: *Proceedings of the National Academy of Sciences* 117.39 (Sept. 2020), pp. 24377–24383. DOI: 10.1073/pnas.2004373117.
- [41] Jackson Champer et al. “Design and Analysis of CRISPR-based Underdominance Toxin-antidote Gene Drives”. In: *Evolutionary Applications* 14.4 (Dec. 2020), pp. 1052–1069. ISSN: 1752-4571. DOI: 10.1111/eva.13180.
- [42] Jackson Champer et al. “Molecular Safeguarding of CRISPR Gene Drive Experiments”. In: *eLife* 8 (Jan. 2019). Ed. by Kevin M Esvelt and Patricia J Wittkopp, e41439. ISSN: 2050-084X. DOI: 10.7554/eLife.41439.

- [43] Jackson Champer et al. “Novel CRISPR/Cas9 Gene Drive Constructs Reveal Insights into Mechanisms of Resistance Allele Formation and Drive Efficiency in Genetically Diverse Populations”. In: *PLOS Genetics* 13.7 (July 2017), e1006796. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1006796.
- [44] Jackson Champer et al. “Performance Analysis of Novel Toxin-Antidote CRISPR Gene Drive Systems”. In: *BMC Biology* 18.1 (Mar. 2020), p. 27. ISSN: 1741-7007. DOI: 10.1186/s12915-020-0761-2.
- [45] Jackson Champer et al. “Reducing Resistance Allele Formation in CRISPR Gene Drive”. In: *Proceedings of the National Academy of Sciences* 115.21 (May 2018), pp. 5522–5527. DOI: 10.1073/pnas.1720354115.
- [46] Jackson Champer et al. “Suppression Gene Drive in Continuous Space Can Result in Unstable Persistence of Both Drive and Wild-Type Alleles”. In: *Molecular Ecology* 30.4 (Feb. 2021), pp. 1086–1101. ISSN: 1365-294X. DOI: 10.1111/mec.15788.
- [47] Samuel E Champer et al. “Anopheles Homing Suppression Drive Candidates Exhibit Unexpected Performance Differences in Simulations with Spatial Structure”. In: *eLife* 11 (Oct. 2022). Ed. by George H Perry, Sebald ANR Verkuijl, and Jim Bull, e79121. ISSN: 2050-084X. DOI: 10.7554/eLife.79121.
- [48] Samuel E. Champer et al. *Finding the Strongest Gene Drive: Simulations Reveal Unexpected Performance Differences between Anopheles Homing Suppression Drive Candidates*. Mar. 2022. DOI: 10.1101/2022.03.28.486009.
- [49] Jason W. Chapman, V. Alistair Drake, and Don R. Reynolds. “Recent Insights from Radar Studies of Insect Flight”. In: *Annual Review of Entomology* 56.1 (2011), pp. 337–356. DOI: 10.1146/annurev-ento-120709-144820.
- [50] Chun-Hong Chen et al. “A Synthetic Maternal-Effect Selfish Genetic Element Drives Population Replacement in *Drosophila*”. In: *Science* 316.5824 (Apr. 2007), pp. 597–600. DOI: 10.1126/science.1138595.
- [51] Brett S. Chevalier and Barry L. Stoddard. “Homing Endonucleases: Structural and Functional Insight into the Catalysts of Intron/Intein Mobility”. In: *Nucleic Acids Research* 29.18 (Sept. 2001), pp. 3757–3774. ISSN: 0305-1048. DOI: 10.1093/nar/29.18.3757.
- [52] Michelle Christian et al. “Targeting DNA Double-Strand Breaks with TAL Effector Nucleases”. In: *Genetics* 186.2 (Oct. 2010), pp. 757–761. ISSN: 1943-2631. DOI: 10.1534/genetics.110.120717.
- [53] Phillip A. Cleves et al. “Reduced Thermal Tolerance in a Coral Carrying CRISPR-induced Mutations in the Gene for a Heat-Shock Transcription Factor”. In: *Proceedings of the National Academy of Sciences* 117.46 (Nov. 2020), pp. 28899–28905. DOI: 10.1073/pnas.1920779117.
- [54] Stanley N. Cohen et al. “Construction of Biologically Functional Bacterial Plasmids In Vitro”. In: *Proceedings of the National Academy of Sciences of the United States of America* 70.11 (Nov. 1973), pp. 3240–3244. ISSN: 0027-8424.
- [55] Matthew A. Combs et al. “Leveraging Eco-Evolutionary Models for Gene Drive Risk Assessment”. In: *Trends in Genetics* 39.8 (Aug. 2023), pp. 609–623. ISSN: 0168-9525. DOI: 10.1016/j.tig.2023.04.004.
- [56] John B. Connolly et al. “Gene Drive in Species Complexes: Defining Target Organisms”. In: *Trends in Biotechnology* 41.2 (Feb. 2023), pp. 154–164. ISSN: 0167-7799. DOI: 10.1016/j.tibtech.2022.06.013.

- [57] John B. Connolly et al. “Recommendations for Environmental Risk Assessment of Gene Drive Applications for Malaria Vector Control”. In: *Malaria Journal* 21.1 (May 2022), p. 152. ISSN: 1475-2875. DOI: 10.1186/s12936-022-04183-w.
- [58] John B. Connolly et al. “Systematic Identification of Plausible Pathways to Potential Harm via Problem Formulation for Investigational Releases of a Population Suppression Gene Drive to Control the Human Malaria Vector *Anopheles Gambiae* in West Africa”. In: *Malaria Journal* 20.1 (Mar. 2021), p. 170. ISSN: 1475-2875. DOI: 10.1186/s12936-021-03674-6.
- [59] Convention on Biological Diversity. *The Cartagena Protocol on Biosafety*. Tech. rep. 2016.
- [60] Franck Courchamp, Ludek Berec, and Joanna Gascoigne. *Allee Effects in Ecology and Conservation*. OUP Oxford, Feb. 2008. ISBN: 978-0-19-152466-0.
- [61] Virginie Courtier-Orgogozo, Baptiste Morizot, and Christophe Boëte. “Agricultural Pest Control with CRISPR-based Gene Drive: Time for Public Debate”. In: *EMBO reports* 18.6 (June 2017), pp. 878–880. ISSN: 1469-221X. DOI: 10.15252/embr.201744205.
- [62] Virginie Courtier-Orgogozo et al. “Evaluating the Probability of CRISPR-based Gene Drive Contaminating Another Species”. In: *Evolutionary Applications* 13.8 (2020), pp. 1888–1905. ISSN: 1752-4571. DOI: 10.1111/eva.12939.
- [63] G. B. Craig, W. A. Hickey, and R. C. VandeHey. “An Inherited Male-Producing Factor in *Aedes Aegypti*”. In: *Science* 132.3443 (Dec. 1960), pp. 1887–1889. DOI: 10.1126/science.132.3443.1887.
- [64] C. F. Curtis. “Possible Use of Translocations to Fix Desirable Genes in Insect Pest Populations”. In: *Nature* 218.5139 (Apr. 1968), pp. 368–369. ISSN: 0028-0836. DOI: 10.1038/218368a0.
- [65] Maxime Dahirel et al. “Shifts from Pulled to Pushed Range Expansions Caused by Reduction of Landscape Connectivity”. In: *Oikos* 130.5 (2021), pp. 708–724. ISSN: 1600-0706. DOI: 10.1111/oik.08278.
- [66] STEPHEN Davis, NICHOLAS Bax, and PETER Grewe. “Engineered Underdominance Allows Efficient and Economical Introgression of Traits into Pest Populations”. In: *Journal of Theoretical Biology* 212.1 (Sept. 2001), pp. 83–98. ISSN: 0022-5193. DOI: 10.1006/jtbi.2001.2357.
- [67] Tom J. de Jong. “Gene Drives Do Not Always Increase in Frequency: From Genetic Models to Risk Assessment”. In: *Journal of Consumer Protection and Food Safety* 12.4 (Dec. 2017), pp. 299–307. ISSN: 1661-5867. DOI: 10.1007/s00003-017-1131-z.
- [68] Peter K. Dearden et al. “The Potential for the Use of Gene Drives for Pest Control in New Zealand: A Perspective”. In: *Journal of the Royal Society of New Zealand* 48.4 (Oct. 2018), pp. 225–244. ISSN: 0303-6758. DOI: 10.1080/03036758.2017.1385030.
- [69] Anne Deredec, Austin Burt, and H. C. J. Godfray. “The Population Genetics of Using Homing Endonuclease Genes in Vector and Pest Management”. In: *Genetics* 179.4 (Aug. 2008), pp. 2013–2026. ISSN: 0016-6731. DOI: 10.1534/genetics.108.089037.
- [70] Sumit Dhole, Alun L. Lloyd, and Fred Gould. “Gene Drive Dynamics in Natural Populations: The Importance of Density Dependence, Space, and Sex”. In: *Annual Review of Ecology, Evolution, and Systematics* 51.1 (Nov. 2020), pp. 505–531. ISSN: 1543-592X. DOI: 10.1146/annurev-ecolsys-031120-101013.
- [71] Sumit Dhole, Alun L. Lloyd, and Fred Gould. “Tethered Homing Gene Drives: A New Design for Spatially Restricted Population Replacement and Suppression”. In: *Evolutionary Applications* 12.8 (2019), pp. 1688–1702. ISSN: 1752-4571. DOI: 10.1111/eva.12827.

- [72] Sumit Dhole et al. “Invasion and Migration of Spatially Self-limiting Gene Drives: A Comparative Analysis”. In: *Evolutionary Applications* 11.5 (Jan. 2018), pp. 794–808. ISSN: 1752-4571. DOI: 10.1111/eva.12583.
- [73] Sandra Myrna Díaz et al. *The Global Assessment Report on Biodiversity and Ecosystem Services: Summary for Policy Makers*. Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, 2019. ISBN: 978-3-947851-13-3.
- [74] James E. DiCarlo et al. “Safeguarding CRISPR-Cas9 Gene Drives in Yeast”. In: *Nature Biotechnology* 33.12 (Dec. 2015), pp. 1250–1255. ISSN: 1546-1696. DOI: 10.1038/nbt.3412.
- [75] Heverton Leandro Carneiro Dutra et al. “Wolbachia Blocks Currently Circulating Zika Virus Isolates in Brazilian *Aedes Aegypti* Mosquitoes”. In: *Cell Host & Microbe* 19.6 (June 2016), pp. 771–774. ISSN: 1931-3128. DOI: 10.1016/j.chom.2016.04.021.
- [76] Philip A. Eckhoff et al. “Impact of Mosquito Gene Drive on Malaria Elimination in a Computational Model with Explicit Spatial and Temporal Dynamics”. In: *Proceedings of the National Academy of Sciences* 114.2 (Jan. 2017), E255–E264. DOI: 10.1073/pnas.1611064114.
- [77] Matthew P. Edgington and Luke S. Alphey. “Population Dynamics of Engineered Underdominance and Killer-Rescue Gene Drives in the Control of Disease Vectors”. In: *PLOS Computational Biology* 14.3 (Mar. 2018), e1006059. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1006059.
- [78] Matthew P. Edgington, Tim Harvey-Samuel, and Luke Alphey. “Split Drive Killer-Rescue Provides a Novel Threshold-Dependent Gene Drive”. In: *Scientific Reports* 10.1 (Nov. 2020), p. 20520. ISSN: 2045-2322. DOI: 10.1038/s41598-020-77544-7.
- [79] Andrea Egizi et al. “The Hitchhiker’s Guide to Becoming Invasive: Exotic Mosquitoes Spread across a US State by Human Transport Not Autonomous Flight”. In: *Molecular Ecology* 25.13 (2016), pp. 3033–3047. ISSN: 1365-294X. DOI: 10.1111/mec.13653.
- [80] Roger Eritja et al. “Direct Evidence of Adult *Aedes Albopictus* Dispersal by Car”. In: *Scientific Reports* 7.1 (Oct. 2017), p. 14399. ISSN: 2045-2322. DOI: 10.1038/s41598-017-12652-5.
- [81] Kevin M Esvelt et al. “Concerning RNA-guided Gene Drives for the Alteration of Wild Populations”. In: *eLife* 3 (July 2014). Ed. by Diethard Tautz, e03401. ISSN: 2050-084X. DOI: 10.7554/eLife.03401.
- [82] Kevin M. Esvelt and Neil J. Gemmill. “Conservation Demands Safe Gene Drive”. In: *PLOS Biology* 15.11 (Nov. 2017), e2003850. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.2003850.
- [83] ETC group. *Call for a Global Moratorium on Genetically-Engineered Gene Drives*. Tech. rep. 2016.
- [84] ETC Group and Heinrich Böll Foundation. *Forcing the Farm*. <https://www.etcgroup.org/content/forcing-farm>. Oct. 2018.
- [85] Nicky R. Faber et al. “Novel Combination of CRISPR-based Gene Drives Eliminates Resistance and Localises Spread”. In: *Scientific Reports* 11.1 (Mar. 2021), p. 3719. ISSN: 2045-2322. DOI: 10.1038/s41598-021-83239-4.
- [86] José María Fernández-Palacios et al. “Scientists’ Warning – The Outstanding Biodiversity of Islands Is in Peril”. In: *Global Ecology and Conservation* 31 (Nov. 2021), e01847. ISSN: 2351-9894. DOI: 10.1016/j.gecco.2021.e01847.
- [87] R. A. Fisher. “The Wave of Advance of Advantageous Genes”. In: *Annals of Eugenics* 7.4 (1937), pp. 355–369. ISSN: 2050-1439. DOI: 10.1111/j.1469-1809.1937.tb02153.x.

- [88] Silke Fuchs et al. “Resistance to a CRISPR-based Gene Drive at an Evolutionarily Conserved Site Is Revealed by Mimicking Genotype Fixation”. In: *PLOS Genetics* 17.10 (Oct. 2021), e1009740. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1009740.
- [89] Roberto Galizi et al. “A CRISPR-Cas9 Sex-Ratio Distortion System for Genetic Control”. In: *Scientific Reports* 6.1 (Aug. 2016), p. 31139. ISSN: 2045-2322. DOI: 10.1038/srep31139.
- [90] Roberto Galizi et al. “A Synthetic Sex Ratio Distortion System for the Control of the Human Malaria Mosquito”. In: *Nature Communications* 5.1 (June 2014), p. 3977. ISSN: 2041-1723. DOI: 10.1038/ncomms4977.
- [91] Valentino Gantz et al. “Highly Efficient Cas9-mediated Gene Drive for Population Modification of the Malaria Vector Mosquito *Anopheles Stephensi*”. In: *Proceedings of the National Academy of Sciences* 112 (Nov. 2015). DOI: 10.1073/pnas.1521077112.
- [92] Valentino M. Gantz and Ethan Bier. “The Dawn of Active Genetics”. In: *BioEssays* 38.1 (2016), pp. 50–63. ISSN: 1521-1878. DOI: 10.1002/bies.201500102.
- [93] Valentino M. Gantz and Ethan Bier. “The Mutagenic Chain Reaction: A Method for Converting Heterozygous to Homozygous Mutations”. In: *Science* 348.6233 (Apr. 2015), pp. 442–444. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aaa5945.
- [94] S. Gershenson. “A New Sex-Ratio Abnormality in *DROSOPHILA OBSCURA*”. In: *Genetics* 13.6 (Nov. 1928), pp. 488–507. ISSN: 0016-6731.
- [95] Léo Girardin and Florence Débarre. “Demographic Feedbacks Can Hamper the Spatial Spread of a Gene Drive”. In: *Journal of Mathematical Biology* 83.6 (Dec. 2021), p. 67. ISSN: 1432-1416. DOI: 10.1007/s00285-021-01702-2.
- [96] Matthew R. Goddard, Duncan Greig, and Austin Burt. “Outcrossed Sex Allows a Selfish Gene to Invade Yeast Populations”. In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 268.1485 (Dec. 2001), pp. 2537–2542. DOI: 10.1098/rspb.2001.1830.
- [97] H. Charles J. Godfray, Ace North, and Austin Burt. “How Driving Endonuclease Genes Can Be Used to Combat Pests and Disease Vectors”. In: *BMC Biology* 15.1 (Sept. 2017), p. 81. ISSN: 1741-7007. DOI: 10.1186/s12915-017-0420-4.
- [98] John Godwin et al. “Rodent Gene Drives for Conservation: Opportunities and Data Needs”. In: *Proceedings of the Royal Society B: Biological Sciences* 286.1914 (Nov. 2019), p. 20191606. DOI: 10.1098/rspb.2019.1606.
- [99] Hannah A. Grunwald, Alexander J. Weitzel, and Kimberly L. Cooper. “Applications of and Considerations for Using CRISPR–Cas9-mediated Gene Conversion Systems in Rodents”. In: *Nature Protocols* 17.1 (Jan. 2022), pp. 3–14. ISSN: 1750-2799. DOI: 10.1038/s41596-021-00646-7.
- [100] Hannah A. Grunwald et al. “Super-Mendelian Inheritance Mediated by CRISPR-Cas9 in the Female Mouse Germline”. In: *Nature* 566.7742 (Feb. 2019), pp. 105–109. ISSN: 1476-4687. DOI: 10.1038/s41586-019-0875-2.
- [101] K. P. Hadeler and F. Rothe. “Travelling Fronts in Nonlinear Diffusion Equations”. In: *Journal of Mathematical Biology* 2.3 (Sept. 1975), pp. 251–263. ISSN: 1432-1416. DOI: 10.1007/BF00277154.
- [102] Andrew Brantley Hall et al. “Radical Remodeling of the Y Chromosome in a Recent Radiation of Malaria Mosquitoes”. In: *Proceedings of the National Academy of Sciences* 113.15 (Apr. 2016), E2114–E2123. DOI: 10.1073/pnas.1525164113.

- [103] Andrew Hammond et al. “A CRISPR-Cas9 Gene Drive System Targeting Female Reproduction in the Malaria Mosquito Vector *Anopheles Gambiae*”. In: *Nature biotechnology* 34 (Dec. 2015). DOI: 10.1038/nbt.3439.
- [104] Andrew Hammond et al. “Gene-Drive Suppression of Mosquito Populations in Large Cages as a Bridge between Lab and Field”. In: *Nature Communications* 12.1 (July 2021), p. 4589. ISSN: 2041-1723. DOI: 10.1038/s41467-021-24790-6.
- [105] Andrew Hammond et al. “Gene-Drive Suppression of Mosquito Populations in Large Cages as a Bridge between Lab and Field”. In: *Nature Communications* 12.1 (July 2021), p. 4589. ISSN: 2041-1723. DOI: 10.1038/s41467-021-24790-6.
- [106] Andrew M. Hammond et al. “The Creation and Selection of Mutations Resistant to a Gene Drive over Multiple Generations in the Malaria Mosquito”. In: *PLOS Genetics* 13.10 (Oct. 2017), e1007039. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1007039.
- [107] Daniel L. Hartl and Andrew G. Clark. *Principles of Population Genetics*. Fourth Edition, Fourth Edition. Oxford, New York: Oxford University Press, Dec. 2006. ISBN: 978-0-87893-308-2.
- [108] Sarah Hartley, Riley Taitingfong, and Pedro Fidelman. “The Principles Driving Gene Drives for Conservation”. In: *Environmental Science & Policy* 135 (Sept. 2022), pp. 36–45. ISSN: 1462-9011. DOI: 10.1016/j.envsci.2022.04.021.
- [109] Bruce A. Hay, Georg Oberhofer, and Ming Guo. “Engineering the Composition and Fate of Wild Populations with Gene Drive”. In: *Annual Review of Entomology* 66.1 (2021), pp. 407–434. DOI: 10.1146/annurev-ento-020117-043154.
- [110] Keith R. Hayes et al. “Identifying and Detecting Potentially Adverse Ecological Outcomes Associated with the Release of Gene-Drive Modified Organisms”. In: *Journal of Responsible Innovation* 5.sup1 (Jan. 2018), S139–S158. ISSN: 2329-9460. DOI: 10.1080/23299460.2017.1415585.
- [111] Volker Henn and Mareike Imken. *Gene Drive Report: A New Dimension of Genetic Engineering*. May 2021.
- [112] Mark Hochstrasser. “Cytoplasmic Incompatibility: A Wolbachia Toxin–Antidote Mechanism Comes into View”. In: *Current Biology* 32.6 (Mar. 2022), R287–R289. ISSN: 0960-9822. DOI: 10.1016/j.cub.2022.02.014.
- [113] A. A. Hoffmann et al. “Successful Establishment of Wolbachia in *Aedes* Populations to Suppress Dengue Transmission”. In: *Nature* 476.7361 (Aug. 2011), pp. 454–457. ISSN: 1476-4687. DOI: 10.1038/nature10356.
- [114] Matt Holzer. *Personal Communication, Conference "Parabolic and Kinetic Models in Population Dynamics", Labex CIMI, Toulouse*. <https://indico.math.cnrs.fr/event/7589/>. 2022.
- [115] Gao Hu et al. “Mass Seasonal Bioflows of High-Flying Insect Migrants”. In: *Science* 354.6319 (Dec. 2016), pp. 1584–1587. DOI: 10.1126/science.aah4379.
- [116] Diana L. Huestis et al. “Windborne Long-Distance Migration of Malaria Mosquitoes in the Sahel”. In: *Nature* 574.7778 (Oct. 2019), pp. 404–408. ISSN: 1476-4687. DOI: 10.1038/s41586-019-1622-4.
- [117] Grant L. Hughes et al. “Wolbachia Infections Are Virulent and Inhibit the Human Malaria Parasite *Plasmodium Falciparum* in *Anopheles Gambiae*”. In: *PLOS Pathogens* 7.5 (May 2011), e1002043. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1002043.
- [118] Alison T. Isaacs et al. “Transgenic *Anopheles Stephensi* Coexpressing Single-Chain Antibodies Resist *Plasmodium Falciparum* Development”. In: *Proceedings of the National Academy of Sciences* 109.28 (July 2012). ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1207738109.

- [119] IUCN. *Invasive Alien Species and Climate Change*. <https://www.iucn.org/resources/issues-brief/invasive-alien-species-and-climate-change>. Resource. Feb. 2021.
- [120] Alain Jacquier and Bernard Dujon. “An Intron-Encoded Protein Is Active in a Gene Conversion Process That Spreads an Intron into a Mitochondrial Gene”. In: *Cell* 41.2 (June 1985), pp. 383–394. ISSN: 0092-8674. DOI: 10.1016/S0092-8674(85)80011-8.
- [121] Martin Jinek et al. “A Programmable Dual RNA-guided DNA Endonuclease in Adaptive Bacterial Immunity”. In: *Science (New York, N.Y.)* 337.6096 (Aug. 2012), pp. 816–821. ISSN: 0036-8075. DOI: 10.1126/science.1225829.
- [122] Dukas Jurėnas et al. “Biology and Evolution of Bacterial Toxin–Antitoxin Systems”. In: *Nature Reviews Microbiology* 20.6 (June 2022), pp. 335–350. ISSN: 1740-1534. DOI: 10.1038/s41579-021-00661-1.
- [123] Bhagyashree Kaduskar et al. “Reversing Insecticide Resistance with Allelic-Drive in *Drosophila Melanogaster*”. In: *Nature Communications* 13.1 (Jan. 2022), p. 291. ISSN: 2041-1723. DOI: 10.1038/s41467-021-27654-1.
- [124] P. E. Kaiser et al. “Homozygous Translocations in *Anopheles Albimanus*”. In: *Theoretical and Applied Genetics* 65.3 (May 1983), pp. 207–211. ISSN: 1432-2242. DOI: 10.1007/BF00308067.
- [125] P. E. Kaiser et al. “Isolation of a Sex-Linked Homozygous Translocation in *Anopheles Albimanus*”. In: *Journal of Heredity* 73.4 (July 1982), pp. 273–276. ISSN: 0022-1503. DOI: 10.1093/oxfordjournals.jhered.a109638.
- [126] P. E. Kaiser et al. “Radiation Induced Reciprocal Translocations and Inversions in *Anopheles Albimanus*”. In: *Canadian Journal of Genetics and Cytology. Journal Canadien De Genetique Et De Cytologie* 24.2 (1982), pp. 177–188. ISSN: 0008-4093. DOI: 10.1139/g82-017.
- [127] Nicholas G. Karavolias et al. “Application of Gene Editing for Climate Change in Agriculture”. In: *Frontiers in Sustainable Food Systems* 5 (2021). ISSN: 2571-581X.
- [128] Katharina Kawall, Janet Cotter, and Christoph Then. “Broadening the GMO Risk Assessment in the EU for Genome Editing Technologies in Agriculture”. In: *Environmental Sciences Europe* 32.1 (Aug. 2020), p. 106. ISSN: 2190-4715. DOI: 10.1186/s12302-020-00361-2.
- [129] Léna Kläy et al. “Pulled, Pushed or Failed: The Demographic Impact of a Gene Drive Can Change the Nature of Its Spatial Spread”. In: *Journal of Mathematical Biology* 87.2 (July 2023), p. 30. ISSN: 1432-1416. DOI: 10.1007/s00285-023-01926-4.
- [130] A.N. Kolmogorov, I.G. Petrovsky, and N.S. Piskunov. “Étude de l’équation de La Diffusion Avec Croissance de La Quantité de Matière et Son Application à Un Problème Biologique”. In: *Moscow Univ. Bull. Ser. Internat. Sect. A* 1 (1937), p. 1.
- [131] Amanda M Koltz and Lauren E Culler. “Biting Insects in a Rapidly Changing Arctic”. In: *Current Opinion in Insect Science*. Global Change Biology. Molecular Physiology. October (2021) 47 (Oct. 2021), pp. 75–81. ISSN: 2214-5745. DOI: 10.1016/j.cois.2021.04.009.
- [132] Michael Kosicki, Kärt Tomberg, and Allan Bradley. “Repair of Double-Strand Breaks Induced by CRISPR-Cas9 Leads to Large Deletions and Complex Rearrangements”. In: *Nature Biotechnology* 36.8 (Sept. 2018), pp. 765–771. ISSN: 1546-1696. DOI: 10.1038/nbt.4192.
- [133] Kyros Kyrou et al. “A CRISPR–Cas9 Gene Drive Targeting Doublesex Causes Complete Population Suppression in Caged *Anopheles Gambiae* Mosquitoes”. In: *Nature Biotechnology* 36.11 (Nov. 2018), pp. 1062–1066. ISSN: 1546-1696. DOI: 10.1038/nbt.4245.

- [134] Philip T. Leftwich et al. “Recent Advances in Threshold-Dependent Gene Drives for Mosquitoes”. In: *Biochemical Society Transactions* 46.5 (Sept. 2018), pp. 1203–1212. ISSN: 0300-5127. DOI: 10.1042/BST20180076.
- [135] Caroline M. Leitschuh et al. “Developing Gene Drive Technologies to Eradicate Invasive Rodents from Islands”. In: *Journal of Responsible Innovation* 5.sup1 (Jan. 2018), S121–S138. ISSN: 2329-9460. DOI: 10.1080/23299460.2017.1365232.
- [136] Philip J. Lester, David O’Sullivan, and George L. W. Perry. “Gene Drives for Invasive Wasp Control: Extinction Is Unlikely, with Suppression Dependent on Dispersal and Growth Rates”. In: *Ecological Applications* (2023). ISSN: 1939-5582. DOI: 10.1002/eap.2912.
- [137] Philip J. Lester et al. “The Potential for a CRISPR Gene Drive to Eradicate or Suppress Globally Invasive Social Wasps”. In: *Scientific Reports* 10.1 (July 2020), p. 12398. ISSN: 2045-2322. DOI: 10.1038/s41598-020-69259-6.
- [138] Jun Li et al. “Can CRISPR Gene Drive Work in Pest and Beneficial Haplodiploid Species?”. In: *Evolutionary Applications* 13.9 (2020), pp. 2392–2403. ISSN: 1752-4571. DOI: 10.1111/eva.13032.
- [139] Ming Li et al. “Development of a Confinable Gene Drive System in the Human Disease Vector *Aedes Aegypti*”. In: *eLife* 9 (Jan. 2020). Ed. by Aleksandra M Walczak and Hugo J Bellen, e51701. ISSN: 2050-084X. DOI: 10.7554/eLife.51701.
- [140] Anna K. Lindholm et al. “The Ecology and Evolutionary Dynamics of Meiotic Drive”. In: *Trends in Ecology & Evolution* 31.4 (Apr. 2016), pp. 315–326. ISSN: 0169-5347. DOI: 10.1016/j.tree.2016.02.001.
- [141] Yang Liu et al. *Overriding Mendelian Inheritance in Arabidopsis with a CRISPR Toxin-Antidote Gene Drive That Impairs Pollen Germination*. Oct. 2023. DOI: 10.1101/2023.10.10.561637.
- [142] Yiran Liu and Jackson Champer. “Modelling Homing Suppression Gene Drive in Haplodiploid Organisms”. In: *Proceedings of the Royal Society B: Biological Sciences* 289.1972 (Apr. 2022), p. 20220320. DOI: 10.1098/rspb.2022.0320.
- [143] Yiran Liu et al. “Adversarial Interspecies Relationships Facilitate Population Suppression by Gene Drive in Spatially Explicit Models”. In: *Ecology Letters* 26.7 (2023), pp. 1174–1185. ISSN: 1461-0248. DOI: 10.1111/ele.14232.
- [144] Nicolas Loeuille. “Protection de l’environnement : « Le concept d’espèce utile ou nuisible est scientifiquement dépassé »”. In: *Le Monde.fr* (Aug. 2023).
- [145] Nancy Lorimer, E. Hallinan, and K. S. Rai. “Translocation Homozygotes in the Yellow Fever Mosquito, *Aedes Aegypti*”. In: *Journal of Heredity* 63.4 (July 1972), pp. 159–166. ISSN: 0022-1503. DOI: 10.1093/oxfordjournals.jhered.a108261.
- [146] Gloria M. Luque et al. “The Genetic Allee Effect: A Unified Framework for the Genetics and Demography of Small Populations”. In: *Ecosphere* 7.7 (2016), e01413. ISSN: 2150-8925. DOI: 10.1002/ecs2.1413.
- [147] *Male Bias and Female Fertility*.
- [148] Andri Manser, Barbara König, and Anna K. Lindholm. “Polyandry Blocks Gene Drive in a Wild House Mouse Population”. In: *Nature Communications* 11.1 (Nov. 2020), p. 5590. ISSN: 2041-1723. DOI: 10.1038/s41467-020-18967-8.
- [149] Andri Manser et al. “Controlling Invasive Rodents via Synthetic Gene Drive and the Role of Polyandry”. In: *Proceedings of the Royal Society B: Biological Sciences* 286.1909 (Aug. 2019), p. 20190852. DOI: 10.1098/rspb.2019.0852.

- [150] Yanfei Mao et al. “Application of the CRISPR–Cas System for Efficient Genome Engineering in Plants”. In: *Molecular Plant* 6.6 (Nov. 2013), pp. 2008–2011. ISSN: 1674-2052. DOI: 10.1093/mp/sst121.
- [151] John M. Marshall. “The Effect of Gene Drive on Containment of Transgenic Mosquitoes”. In: *Journal of Theoretical Biology* 258.2 (May 2009), pp. 250–265. ISSN: 0022-5193. DOI: 10.1016/j.jtbi.2009.01.031.
- [152] John M. Marshall and Bruce A. Hay. “Medusa: A Novel Gene Drive System for Confined Suppression of Insect Populations”. In: *PLOS ONE* 9.7 (July 2014), e102694. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0102694.
- [153] Leo Marx. “Are Science and Society Going in the Same Direction?” In: *Science, Technology, & Human Values* 8.4 (1983), pp. 6–9. ISSN: 0162-2439. JSTOR: 689243.
- [154] Raul F. Medina. “Gene Drives and the Management of Agricultural Pests”. In: *Journal of Responsible Innovation* 5.sup1 (Jan. 2018), S255–S262. ISSN: 2329-9460. DOI: 10.1080/23299460.2017.1407913.
- [155] Matthew Metzloff et al. *Experimental Demonstration of Tethered Gene Drive Systems for Confined Population Modification or Suppression*. Apr. 2022. DOI: 10.1101/2021.05.29.446308.
- [156] J Miller, A D McLachlan, and A Klug. “Repetitive Zinc-Binding Domains in the Protein Transcription Factor IIIA from *Xenopus* Oocytes.” In: *The EMBO Journal* 4.6 (June 1985), pp. 1609–1614. ISSN: 0261-4189.
- [157] John Min et al. *Daisy Quorum Drives for the Genetic Restoration of Wild Populations*. Mar. 2017. DOI: 10.1101/115618.
- [158] Catherine L. Moyes et al. “Contemporary Status of Insecticide Resistance in the Major *Aedes* Vectors of Arboviruses Infecting Humans”. In: *PLOS Neglected Tropical Diseases* 11.7 (July 2017), e0005625. ISSN: 1935-2735. DOI: 10.1371/journal.pntd.0005625.
- [159] Carl Mueller, Leonid Mytnik, and Jeremy Quastel. “Effect of Noise on Front Propagation in Reaction-Diffusion Equations of KPP Type”. In: *Inventiones mathematicae* 184.2 (May 2011), pp. 405–453. ISSN: 1432-1297. DOI: 10.1007/s00222-010-0292-5.
- [160] Grégoire Nadin, Martin Strugarek, and Nicolas Vauchelet. “Hindrances to Bistable Front Propagation: Application to *Wolbachia* Invasion”. In: *Journal of Mathematical Biology* 76.6 (May 2018), pp. 1489–1533. ISSN: 1432-1416. DOI: 10.1007/s00285-017-1181-y.
- [161] Alice Namias et al. “From *Wolbachia* Genomics to Phenotype: Molecular Models of Cytoplasmic Incompatibility Must Account for the Multiplicity of Compatibility Types”. In: *Current Opinion in Insect Science* 49 (Feb. 2022), pp. 78–84. ISSN: 2214-5745. DOI: 10.1016/j.cois.2021.12.005.
- [162] National Academies of Sciences, Engineering, and Medicine. *Gene Drives on the Horizon: Advancing Science, Navigating Uncertainty, and Aligning Research with Public Values*. Washington (DC): National Academies Press (US), 2016. ISBN: 978-0-309-43787-5.
- [163] Pierre Netter et al. “Mitochondrial Genetics VII. Allelism and Mapping Studies of Ribosomal Mutants Resistant to Chloramphenicol, Erythromycin and Spiramycin in *S. CEREVISIAE*”. In: *Genetics* 78.4 (Dec. 1974), pp. 1063–1100. ISSN: 0016-6731.
- [164] Paul Neve. “Gene Drive Systems: Do They Have a Place in Agricultural Weed Management?” In: *Pest Management Science* 74.12 (2018), pp. 2671–2679. ISSN: 1526-4998. DOI: 10.1002/ps.5137.

- [165] Charleston Noble et al. “Current CRISPR Gene Drive Systems Are Likely to Be Highly Invasive in Wild Populations”. In: *eLife* 7 (June 2018). Ed. by Michael Doebeli, e33423. ISSN: 2050-084X. DOI: 10.7554/eLife.33423.
- [166] Charleston Noble et al. “Daisy-Chain Gene Drives for the Alteration of Local Populations”. In: *Proceedings of the National Academy of Sciences* 116.17 (Apr. 2019), pp. 8275–8282. DOI: 10.1073/pnas.1716358116.
- [167] Charleston Noble et al. “Evolutionary Dynamics of CRISPR Gene Drives”. In: *Science Advances* 3.4 (Apr. 2017), e1601964. DOI: 10.1126/sciadv.1601964.
- [168] Tony Nolan. “Control of Malaria-Transmitting Mosquitoes Using Gene Drives”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 376.1818 (Dec. 2020), p. 20190803. DOI: 10.1098/rstb.2019.0803.
- [169] Ace North, Austin Burt, and H. Charles J. Godfray. “Modelling the Spatial Spread of a Homing Endonuclease Gene in a Mosquito Population”. In: *Journal of Applied Ecology* 50.5 (2013), pp. 1216–1225. ISSN: 1365-2664. DOI: 10.1111/1365-2664.12133.
- [170] Ace R. North, Austin Burt, and H. Charles J. Godfray. “Modelling the Potential of Genetic Control of Malaria Mosquitoes at National Scale”. In: *BMC Biology* 17.1 (Mar. 2019), p. 26. ISSN: 1741-7007. DOI: 10.1186/s12915-019-0645-5.
- [171] Ace R. North, Austin Burt, and H. Charles J. Godfray. “Modelling the Suppression of a Malaria Vector Using a CRISPR-Cas9 Gene Drive to Reduce Female Fertility”. In: *BMC Biology* 18 (Aug. 2020), p. 98. ISSN: 1741-7007. DOI: 10.1186/s12915-020-00834-z.
- [172] David Norton et al. “How Do We Restore New Zealand’s Biological Heritage by 2050?” In: *Ecological Management & Restoration* 17 (Sept. 2016), pp. 170–179. DOI: 10.1111/emr.12230.
- [173] Akira Okubo and Simon A. Levin. *Diffusion and Ecological Problems: Modern Perspectives*. Ed. by S. S. Antman et al. Vol. 14. Interdisciplinary Applied Mathematics. New York, NY: Springer, 2001. ISBN: 978-1-4419-3151-1 978-1-4757-4978-6. DOI: 10.1007/978-1-4757-4978-6.
- [174] World Health Organization. “Global Vector Control Response 2017-2030.” In: *Global vector control response 2017-2030*. (2017).
- [175] James Orsborne et al. “Investigating the Blood-Host Plasticity and Dispersal of *Anopheles Coluzzii* Using a Novel Field-Based Methodology”. In: *Parasites & Vectors* 12.1 (Mar. 2019), p. 143. ISSN: 1756-3305. DOI: 10.1186/s13071-019-3401-3.
- [176] Sarah P. Otto and Troy Day. *A Biologist’s Guide to Mathematical Modeling in Ecology and Evolution*. Princeton University Press, Sept. 2011. ISBN: 978-1-4008-4091-5. DOI: 10.1515/9781400840915.
- [177] Jeff F. Paril and Ben L. Phillips. “Slow and Steady Wins the Race: Spatial and Stochastic Processes and the Failure of Suppression Gene Drives”. In: *Molecular Ecology* (July 2022), mec.16598. ISSN: 0962-1083, 1365-294X. DOI: 10.1111/mec.16598.
- [178] S. Peischl et al. “On the Accumulation of Deleterious Mutations during Range Expansions”. In: *Molecular Ecology* 22.24 (2013), pp. 5972–5982. ISSN: 1365-294X. DOI: 10.1111/mec.12524.
- [179] Thai Binh Pham et al. “Experimental Population Modification of the Malaria Vector Mosquito, *Anopheles Stephensi*”. In: *PLOS Genetics* 15.12 (Dec. 2019), e1008440. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1008440.
- [180] Antoinette J. Piaggio et al. “Is It Time for Synthetic Biodiversity Conservation?” In: *Trends in Ecology & Evolution* 32.2 (Feb. 2017), pp. 97–107. ISSN: 0169-5347. DOI: 10.1016/j.tree.2016.10.016.

- [181] Tom A. R. Price et al. “Resistance to Natural and Synthetic Gene Drive Systems”. In: *Journal of Evolutionary Biology* 33.10 (Oct. 2020), pp. 1345–1360. ISSN: 1010-061X. DOI: 10.1111/jeb.13693.
- [182] Petr Pyšek et al. “Scientists’ Warning on Invasive Alien Species”. In: *Biological Reviews* 95.6 (2020), pp. 1511–1534. ISSN: 1469-185X. DOI: 10.1111/brv.12627.
- [183] Floyd A Reed. “CRISPR/Cas9 Gene Drive: Growing Pains for a New Technology”. In: *Genetics* 205.3 (Mar. 2017), pp. 1037–1039. ISSN: 1943-2631. DOI: 10.1534/genetics.116.198887.
- [184] Floyd A. Reed et al. *RPM-Drive: A Robust, Safe, and Reversible Gene Drive System That Remains Functional after 200+ Generations*. July 2018. DOI: 10.48550/arXiv.1806.05304. arXiv: 1806.05304 [q-bio].
- [185] R. Guy Reeves et al. “First Steps towards Underdominant Genetic Transformation of Insect Populations”. In: *PLOS ONE* 9.5 (May 2014), e97557. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0097557.
- [186] Nicolas O Rode, Virginie Courtier-Orgogozo, and Florence Débarre. “Can a Population Targeted by a CRISPR-Based Homing Gene Drive Be Rescued?” In: *G3 Genes/Genomes/Genetics* 10.9 (Sept. 2020), pp. 3403–3415. ISSN: 2160-1836. DOI: 10.1534/g3.120.401484.
- [187] Nicolas O. Rode et al. “Population Management Using Gene Drive: Molecular Design, Models of Spread Dynamics and Assessment of Ecological Risks”. In: *Conservation Genetics* 20.4 (Aug. 2019), pp. 671–690. ISSN: 1572-9737. DOI: 10.1007/s10592-019-01165-5.
- [188] Lionel Roques et al. “Allee Effect Promotes Diversity in Traveling Waves of Colonization”. In: *Proceedings of the National Academy of Sciences* 109.23 (June 2012), pp. 8828–8833. DOI: 10.1073/pnas.1201695109.
- [189] Perran A. Ross, Nancy M. Endersby-Harshman, and Ary A. Hoffmann. “A Comprehensive Assessment of Inbreeding and Laboratory Adaptation in *Aedes Aegypti* Mosquitoes”. In: *Evolutionary Applications* 12.3 (2019), pp. 572–586. ISSN: 1752-4571. DOI: 10.1111/eva.12740.
- [190] Claus Rueffler, Martijn Egas, and Johan A. J. Metz. “Evolutionary Predictions Should Be Based on Individual-Level Traits.” In: *The American Naturalist* 168.5 (Nov. 2006), E148–E162. ISSN: 0003-0147. DOI: 10.1086/508618.
- [191] C. Sánchez et al. “Modeling Confinement and Reversibility of Threshold-Dependent Gene Drive Systems in Spatially-Explicit *Aedes Aegypti* Populations”. In: *BMC Biology* 18.1 (May 2020), p. 50. ISSN: 1741-7007. DOI: 10.1186/s12915-020-0759-9.
- [192] Héctor M. Sánchez C. et al. “MGDrivE: A Modular Simulation Framework for the Spread of Gene Drives through Spatially Explicit Mosquito Populations”. In: *Methods in Ecology and Evolution* 11.2 (2020), pp. 229–239. ISSN: 2041-210X. DOI: 10.1111/2041-210X.13318.
- [193] Dane Scott. “The Technological Fix Criticisms and the Agricultural Biotechnology Debate”. In: *Journal of Agricultural and Environmental Ethics* 24.3 (June 2011), pp. 207–226. ISSN: 1187-7863, 1573-322X. DOI: 10.1007/s10806-010-9253-7.
- [194] Maxwell J. Scott et al. “Agricultural Production: Assessment of the Potential Use of Cas9-mediated Gene Drive Systems for Agricultural Pest Control”. In: *Journal of Responsible Innovation* 5.sup1 (Jan. 2018), S98–S120. ISSN: 2329-9460, 2329-9037. DOI: 10.1080/23299460.2017.1410343.
- [195] Megan Scudellari. “Self-Destructing Mosquitoes and Sterilized Rodents: The Promise of Gene Drives”. In: *Nature* 571.7764 (July 2019), pp. 160–162. DOI: 10.1038/d41586-019-02087-5.

- [196] V. P. Sharma. “Health Hazards of Mosquito Repellents and Safe Alternatives”. In: *Current Science* 80.3 (2001), pp. 341–343. ISSN: 0011-3891. JSTOR: 24105693.
- [197] J. Dylan Shropshire and Seth R. Bordenstein. “Two-By-One Model of Cytoplasmic Incompatibility: Synthetic Recapitulation by Transgenic Expression of *cifA* and *cifB* in *Drosophila*”. In: *PLOS Genetics* 15.6 (June 2019), e1008221. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1008221.
- [198] Alekos Simoni et al. “A Male-Biased Sex-Distorter Gene Drive for the Human Malaria Vector *Anopheles Gambiae*”. In: *Nature Biotechnology* 38.9 (Sept. 2020), pp. 1054–1060. ISSN: 1546-1696. DOI: 10.1038/s41587-020-0508-1.
- [199] Alekos Simoni et al. “Development of Synthetic Selfish Elements Based on Modular Nucleases in *Drosophila Melanogaster*”. In: *Nucleic Acids Research* 42.11 (June 2014), pp. 7461–7472. ISSN: 0305-1048. DOI: 10.1093/nar/gku387.
- [200] Steven P. Sinkins and Fred Gould. “Gene Drive Systems for Insect Disease Vectors”. In: *Nature Reviews Genetics* 7.6 (June 2006), pp. 427–435. ISSN: 1471-0064. DOI: 10.1038/nrg1870.
- [201] Martin Strugarek and Nicolas Vauchelet. “Reduction to a Single Closed Equation for 2-by-2 Reaction-Diffusion Systems of Lotka–Volterra Type”. In: *SIAM Journal on Applied Mathematics* 76.5 (Jan. 2016), pp. 2060–2080. ISSN: 0036-1399. DOI: 10.1137/16M1059217.
- [202] Jaye Sudweeks et al. “Locally Fixed Alleles: A Method to Localize Gene Drive to Island Populations”. In: *Scientific Reports* 9.1 (Nov. 2019), p. 15821. ISSN: 2045-2322. DOI: 10.1038/s41598-019-51994-0.
- [203] SynBioWatch. *A Call for Conservation with a Conscience: No Place for Gene Drives in Conservation*. Tech. rep. 2016.
- [204] Hidenori Tanaka, Howard A. Stone, and David R. Nelson. “Spatial Gene Drives and Pushed Genetic Waves”. In: *Proceedings of the National Academy of Sciences of the United States of America* 114.32 (Aug. 2017), pp. 8452–8457. ISSN: 1091-6490. DOI: 10.1073/pnas.1705868114.
- [205] Gerard Terradas et al. “Inherently Confinable Split-Drive Systems in *Drosophila*”. In: *Nature Communications* 12.1 (Mar. 2021), p. 1480. ISSN: 2041-1723. DOI: 10.1038/s41467-021-21771-7.
- [206] Paul B. Thompson. *Food Biotechnology in Ethical Perspective*. Vol. 10. The International Library of Environmental, Agricultural and Food Ethics. Dordrecht: Springer Netherlands, 2007. ISBN: 978-1-4020-5790-8. DOI: 10.1007/1-4020-5791-1.
- [207] David R. Towns, Ian A. E. Atkinson, and Charles H. Daugherty. “Have the Harmful Effects of Introduced Rats on Islands Been Exaggerated?” In: *Biological Invasions* 8.4 (June 2006), pp. 863–891. ISSN: 1573-1464. DOI: 10.1007/s10530-005-0421-z.
- [208] A. Tsoularis and J. Wallace. “Analysis of Logistic Growth Models”. In: *Mathematical Biosciences* 179.1 (July 2002), pp. 21–55. ISSN: 0025-5564. DOI: 10.1016/S0025-5564(02)00096-2.
- [209] Michael Turelli and Nicholas H. Barton. “Deploying Dengue-Suppressing *Wolbachia* : Robust Models Predict Slow but Effective Spatial Spread in *Aedes Aegypti*”. In: *Theoretical Population Biology* 115 (June 2017), pp. 45–60. ISSN: 0040-5809. DOI: 10.1016/j.tpb.2017.03.003.
- [210] Michael Turelli and Nicholas H. Barton. “Why Did the *Wolbachia* Transinfection Cross the Road? Drift, Deterministic Dynamics, and Disease Control”. In: *Evolution Letters* 6.1 (Feb. 2022), pp. 92–105. ISSN: 2056-3744, 2056-3744. DOI: 10.1002/evl3.270.
- [211] James M. A. Turner. “Meiotic Sex Chromosome Inactivation”. In: *Development* 134.10 (May 2007), pp. 1823–1831. ISSN: 0950-1991. DOI: 10.1242/dev.000018.

- [212] Robert L Unckless, Andrew G Clark, and Philipp W Messer. “Evolution of Resistance Against CRISPR/Cas9 Gene Drive”. In: *Genetics* 205.2 (Feb. 2017), pp. 827–841. ISSN: 1943-2631. DOI: 10.1534/genetics.116.197285.
- [213] Robert L Unckless et al. “Modeling the Manipulation of Natural Populations by the Mutagenic Chain Reaction”. In: *Genetics* 201.2 (Oct. 2015), pp. 425–431. ISSN: 1943-2631. DOI: 10.1534/genetics.115.177592.
- [214] Simon J Unterholzner, Brigitte Poppenberger, and Wilfried Rozhon. “Toxin–Antitoxin Systems”. In: *Mobile Genetic Elements* 3.5 (Sept. 2013), e26219. ISSN: 2159-2543. DOI: 10.4161/mge.26219.
- [215] Laurence Van Melderen. “Toxin–Antitoxin Systems: Why so Many, What For?” In: *Current Opinion in Microbiology*. Growth and Development: Eukaryotes/Prokaryotes 13.6 (Dec. 2010), pp. 781–785. ISSN: 1369-5274. DOI: 10.1016/j.mib.2010.10.006.
- [216] George W. Ware. “Effects of Pesticides on Nontarget Organisms”. In: *Residue Reviews*. Ed. by Francis A. Gunther and Jane Davies Gunther. Residue Reviews. New York, NY: Springer, 1980, pp. 173–201. ISBN: 978-1-4612-6107-0. DOI: 10.1007/978-1-4612-6107-0_9.
- [217] J. D. Watson and F. H. C. Crick. “Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid”. In: *Nature* 171.4356 (Apr. 1953), pp. 737–738. ISSN: 1476-4687. DOI: 10.1038/171737a0.
- [218] N. Wedell, T. a. R. Price, and A. K. Lindholm. “Gene Drive: Progress and Prospects”. In: *Proceedings of the Royal Society B: Biological Sciences* 286.1917 (Dec. 2019), p. 20192709. DOI: 10.1098/rspb.2019.2709.
- [219] Lynn White. “The Historical Roots of Our Ecologic Crisis”. In: *Science* 155.3767 (Mar. 1967), pp. 1203–1207. DOI: 10.1126/science.155.3767.1203.
- [220] Katie Willis and Austin Burt. “Double Drives and Private Alleles for Localised Population Genetic Control”. In: *PLOS Genetics* 17.3 (Mar. 2021), e1009333. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1009333.
- [221] Ernst A. Wimmer. “Insect Biotechnology: Controllable Replacement of Disease Vectors”. In: *Current Biology* 23.10 (May 2013), R453–R456. ISSN: 0960-9822. DOI: 10.1016/j.cub.2013.03.058.
- [222] Roger J. Wood and Martha E. Newton. “Sex-Ratio Distortion Caused by Meiotic Drive in Mosquitoes”. In: *The American Naturalist* 137.3 (Mar. 1991), pp. 379–391. ISSN: 0003-0147. DOI: 10.1086/285171.
- [223] Bing Wu, Liqun Luo, and Xiaojing J. Gao. “Cas9-Triggered Chain Ablation of Cas9 as a Gene Drive Brake”. In: *Nature Biotechnology* 34.2 (Feb. 2016), pp. 137–138. ISSN: 1546-1696. DOI: 10.1038/nbt.3444.
- [224] Emily Yang et al. “A Homing Suppression Gene Drive with Multiplexed gRNAs Maintains High Drive Conversion Efficiency and Avoids Functional Resistance Alleles”. In: *G3 Genes/Genomes/Genetics* 12.6 (June 2022), jkac081. ISSN: 2160-1836. DOI: 10.1093/g3journal/jkac081.
- [225] Jiangbo Zhou et al. “Critical Traveling Waves in a Diffusive Disease Model”. In: *Journal of Mathematical Analysis and Applications* 476.2 (Aug. 2019), pp. 522–538. ISSN: 0022-247X. DOI: 10.1016/j.jmaa.2019.03.066.
- [226] Yutong Zhu and Jackson Champer. “Simulations Reveal High Efficiency and Confinement of a Population Suppression CRISPR Toxin-Antidote Gene Drive”. In: *ACS Synthetic Biology* 12.3 (Mar. 2023), pp. 809–819. DOI: 10.1021/acssynbio.2c00611.

Mathematical models of gene drive for population management

Léna KLÄY

Abstract

Artificial gene drive is a genetic engineering technology that could be used for the control of natural populations. Gene drive alleles bias their own transmission and can therefore spread in a population within a relatively small number of generations, even if they are deleterious. Understanding the potential outcomes of this technology, including the modification and/or the eradication of a natural population, is essential before real-world applications are considered. In this PhD thesis, I study the spatial spread of gene drive alleles through modelling and more specifically, I focus on the following questions: i) will the drive alleles spread? ii) at which speed? iii) how will the density and/or the genetic composition of the population be affected over time? And iv) is there a possibility to spatially confine this spread? In a first part, I use a deterministic approach to study the influence of demography over the first three questions. Among other results, I mathematically demonstrate how the intrinsic growth rate can make the difference between a drive invasion (large values) and the decay of the drive allele uniformly in space leaving only wild-type individuals in the environment (small values). In a second part, I generalise the results I previously obtained by taking into account several other biological assumptions on population dynamics. I show how an Allee effect might help eradicate or reduce the density of the targeted population. In a context of resources rarefaction, I also show how a logistic density-dependence increasing the death rate instead of decreasing the birth rate might accelerate the drive invasion. In a third part, I explore stochastic dynamics when the population size gets small. In particular, I study wild-type recolonising events (*chasing events*) that might prevent the eradication of the population by the drive. I demonstrate that chasing dynamics are very unlikely for a large enough carrying capacity and a small enough drive fitness cost. In a fourth part, I investigate necessary conditions for a drive underdominant construct to spread inside a targeted area but fail to spread outside. I demonstrate how this strategy is inefficient to spatially confine the drive in continuous environments, and how this result can also be extended to discrete environments with close enough spatial sites. Overall this thesis contributes to develop a clear understanding of the spatial spread of artificial gene drive and assess several modelling choices to provide more relevant insights on real-world dynamics.

Modèles mathématiques de forçage génétique pour la gestion de populations

Léna KLÄY

Résumé

Le forçage génétique artificiel est une technologie qui pourrait permettre de modifier génétiquement des populations sauvages, notamment pour réduire leur taille. Les allèles forcés génétiquement ont un taux de transmission plus élevé que le taux mendélien classique et peuvent ainsi se fixer dans une population en un nombre de générations relativement faible, même s'ils sont délétères. Avant d'envisager leur introduction dans la nature, il est essentiel d'appréhender les conséquences d'un tel lâcher. Dans cette thèse, j'étudie la propagation spatiale des allèles forcés génétiquement grâce à des modèles mathématiques. Je m'intéresse plus particulièrement aux questions suivantes : i) les allèles forcés génétiquement se propagent-ils ? ii) si oui, à quelle vitesse ? iii) comment la densité et/ou la composition génétique de la population varie-t-elle au cours du temps ? et iv) est-il possible de limiter la propagation de ces allèles dans l'espace ? Dans une première partie, j'utilise une approche déterministe pour étudier l'influence de la démographie sur les trois premières questions. Entre autres, je montre comment un fort taux de croissance intrinsèque peut mener à la fixation des allèles introduits, alors qu'un faible taux de croissance intrinsèque conduit à la disparition de ces allèles uniformément dans l'espace, ne laissant que des individus sauvages dans l'environnement. Dans une deuxième partie, je généralise les résultats obtenus dans la partie précédente en prenant en compte plusieurs hypothèses biologiques sur la dynamique des populations. Je démontre qu'un effet Allee peut contribuer à éradiquer ou à réduire en densité la population sauvage ciblée, et dans un contexte de raréfaction des ressources, je démontre également que la vitesse de propagation des allèles forcés génétiquement varie en fonction de la composante de fitness (natalité ou mortalité) affectée par la densité-dépendance. Dans une troisième partie, j'explore les dynamiques stochastiques d'une population de taille restreinte. J'étudie en particulier les événements de recolonisation par des individus sauvages (*chasing events*) qui pourraient empêcher l'éradication d'une population. Je démontre que ces dynamiques sont très peu probables pour une capacité de charge suffisamment grande et des individus forcés génétiquement peu désavantagés. Dans une quatrième partie, j'étudie les conditions nécessaires pour qu'un allèle forcé génétiquement et sous-dominant se propage dans une zone géographique, mais y reste confiné. Je montre que ce confinement n'est envisageable ni dans des environnements continus, ni dans des environnements discrets dont les sites spatiaux sont suffisamment proches. L'ensemble de ces travaux de thèse contribuent à une plus fine compréhension des processus spatiaux tout en améliorant les modèles prédictifs, dans le but final d'éclairer le débat public.