

# **Models of evolutionary rescue of bacteria and yeast by extrachromosomal genetic elements**

**Dissertation**

zur Erlangung des Doktorgrades

*Doctor rerum naturalium*

der Mathematisch-Naturwissenschaftlichen Fakultät  
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

**Ian Alexander Dewan**

Plön, November 2024

Erste Gutachterin:  
Zweiter Gutachter:

Dr Hildegard Uecker  
Prof. Dr Sören Christensen

Tag der mündlichen Prüfung: 12. Dezember 2024

## ABSTRACT

Many organisms contain extrachromosomal genetic elements, additional components of the genome outside of the main chromosomal component. Mathematical modelling has already made an extensive contribution to the theoretical population genetics of normal, chromosomal genes, but extrachromosomal genetic elements are less well studied. In this thesis, I develop mathematical models of the contribution of two extrachromosomal genetic elements, plasmids in bacteria and eccDNA in yeast, to the evolutionary rescue of their hosts. In both cases, the unique extrachromosomal properties of the genetic system facilitate their contribution to the rapid adaptation of the host to an environmental change.

First, however, I set out, in the first two chapters, an overview of plasmid biology suitable for an audience of modellers, and introduce both the existing body of empirical knowledge of plasmids and the existing modelling work and open questions. This introduction will hopefully help to expand the existing body of plasmid theory by easing the process of starting out for new plasmid modellers. I also address a confusion that has arisen in the plasmid literature between the distribution of plasmid lengths and the distribution of log-lengths. This clarification will avoid misleading inferences.

In the third chapter, I develop a model of evolutionary rescue of a bacterial population in a scenario of heterozygote advantage. Multicopy plasmids can provide heterozygosity and allow a population to be rescued in a heterozygote advantage context. Heterozygosity of plasmids is subject to constant loss, however, due to random segregation of plasmid copies on cell division. I determine establishment and rescue probabilities, and derive analytical expressions for the threshold fitness of heterozygotes required to overcome loss to segregation, among other results.

In the final chapter, I develop and analyze a model of evolutionary rescue of a yeast population by gene amplification on eccDNA. In particular, the model shows the role of biased eccDNA segregation in enabling rescue.



## ZUSAMMENFASSUNG

Viele Organismen beinhalten extrachromosomale genetische Elemente, zusätzliche Komponenten des Genoms, die außerhalb der chromosomalen Hauptkomponente liegen. Mathematische Modelle leisten schon einen umfangreichen Beitrag zur theoretischen Populationsgenetik der normalen, chromosomalen Gene; extrachromosomale genetische Elemente aber sind nicht so gründlich erforscht. In dieser Arbeit stelle ich mathematische Modelle des Beitrags zweier extrachromosomale genetische Elemente, die Plasmiden in den Bakterien und die eccDNA in den Hefen, zur evolutionären Rettung ihrer Wirte vor. In den beiden Fällen erleichtern die einzigartigen extrachromosomalen Eigenschaften dieser genetischen Systeme ihren Beitrag zur schnellen Adaptation des Wirtes an einem Umgebungswandel.

Zuerst aber nehme ich vor einen zu einem Publikum der Modellierer angemessen Überblick der Biologie der Plasmide zu geben und sowohl das bestehende Korpus des empirischen Wissens der Plasmide als auch die bestehenden Modelle und offenen Fragen. Hoffentlich wird diese Einführung das bestehende Korpus der Plasmidentheorie vergrößern durch die Erleichterung des Einstiegsprozesses. Ich gehe auch eine Verwechslung an, die in der Plasmidenliteratur aufgetreten ist: die Verwechslung zwischen der Verteilung der Längen der Plasmiden und der Verteilung der Logarithmen der Längen. Diese Verdeutlichung wird missverständliche Folgerungen vermeiden.

Im dritten Kapitel, entwickle ich ein Modell der evolutionäre Rettung einer Bakterienpopulation in einem Szenario des heterozygoten Vorteil. Multikopie-Plasmiden können Heterozygotie erzeugen und eine Population in einem Heterozygoten-Vorteil-Szenario gerettet werden lassen. Die Heterozygotie der Plasmide unterliegt aber stetigen Verlusten wegen der zufälligen Segregation der Plasmidkopien während der Zellteilung. Ich bestimme Etablierungs- und Rettungswahrscheinlichkeiten, und leite analytische Ausdrücke für die Heterozygote-Fitnessschwelle, die erforderlich ist, um Segregationsverlust zu überwinden, her.

Im letzten Kapitel, entwickle und analysiere ich ein Modell der evolutionäre Rettung einer Hefenpopulation durch Amplifikation der Gene auf eccDNA. Insbesondere, das Model zeigt die Rolle der verzerrten eccDNA-Segregation bei der Ermöglichung der Rettung.



## ACKNOWLEDGEMENTS

Firstly and most importantly, I have to thank Hildegard Uecker for all the ways she has supported and guided me through the last four years. Working with Hildegard has been a great experience, and I am grateful to have had the chance to have an advisor who is so dedicated to doing excellent science, so gracious, and so helpful as a mentor. I also want to thank Sally Otto for her guidance on my final project, and for her and her group so kindly welcoming me when I visited them last summer.

I'm grateful to the members of my thesis advisory committee, Arne Traulsen, Tal Dagan, and Félix Geoffroy, for their all of their advice and assistance. I'd like to thank everyone with whom I have discussed my various projects over the course of my doctorate, in particular Álvaro San Millán, Dustin Hanke, Johannes Effe, Tal Dagan, Félix Geoffroy, Mario Santer, Nikhil Sharma, and Lindi Wahl. I would especially like to thank everyone who has read one of my manuscripts and whose helpful and valuable comments have greatly improved my work, including Román Zapién Campos, Ernesto Berríos Caro, Christin Nyhoegen, Arne Traulsen, and several editors and anonymous reviewers.

Working with the Rescue Team has been a wonderful experience and has made the last four years fly by. I'd like to thank all the current and former members of the group for making it such a welcoming, fun, and scientifically exciting environment: Hildegard Uecker, Mario Santer, Christin Nyhoegen, Félix Geoffroy, Puneeth Deraje, Ernesto Berríos Caro, and Janika Elfert. I'm also indebted to all the members of the Theoretical Biology department and the wider institute in Plön who make our isolated institute in a small town such a vibrant and exciting place to work. I should like particularly to thank Jun Ishigohoka, Jule Neumann, Mayo Ambekar, and Alec Jacobsen. Lastly, but in no sense least, I'd like to extend my gratitude to my family.



# CONTENTS

<b>Abstract</b>	<b>iii</b>
<b>Zusammenfassung</b>	<b>v</b>
<b>Acknowledgements</b>	<b>vii</b>
<b>Contents</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 A Mathematician's guide to plasmids: An introduction to plasmid biology for modellers</b>	<b>5</b>
2.1 Introduction . . . . .	5
2.2 The natural history of plasmids . . . . .	9
2.3 The anatomy of plasmids . . . . .	12
2.3.1 Genes carried on plasmids . . . . .	12
2.3.2 Genetic structure . . . . .	15
2.3.3 Plasmid multimers . . . . .	15
2.4 The physiology of plasmids . . . . .	17
2.4.1 Replication, plasmid copy number, and their control .	17
2.4.2 Segregation of plasmids at host cell division . . . . .	20
2.4.3 Horizontal transfer by conjugation . . . . .	21
2.4.4 Postsegregational killing . . . . .	24
2.5 The ecology of plasmids . . . . .	25
2.5.1 Interactions with the individual host . . . . .	25
2.5.2 Interactions with the host community . . . . .	29
2.5.3 Interactions with other plasmids and mobile genetic elements . . . . .	30
2.6 Final remarks . . . . .	32
<b>3 Is the distribution of plasmid lengths bimodal?</b>	<b>37</b>
3.1 Introduction . . . . .	37
3.2 Analysis and results . . . . .	38
3.2.1 The distribution of log-lengths is not the distribution of lengths . . . . .	38
3.2.2 Within bacterial families, the plasmid length distri- bution is often bimodal, but the modes and the valley are incorrectly predicted by the log-length distribution	41
3.3 Discussion . . . . .	43

<b>4</b>	<b>Evolutionary rescue of bacterial populations by heterozygosity on multicopy plasmids</b>	<b>47</b>
4.1	Introduction . . . . .	47
4.2	The model . . . . .	51
4.3	The establishment probability . . . . .	54
4.3.1	Analysis . . . . .	54
4.3.2	Results . . . . .	56
4.4	The rescue probability . . . . .	58
4.4.1	Analysis . . . . .	58
4.4.2	Results . . . . .	58
4.5	Critical values of $n$ and $s_{\max}$ . . . . .	59
4.6	The final distribution of cell types . . . . .	67
4.7	Cointegration . . . . .	67
4.7.1	The cointegration model . . . . .	69
4.7.2	Analysis and results . . . . .	70
4.8	Discussion . . . . .	71
<b>5</b>	<b>Rapid adaptation through gene amplification on eccDNA in yeast</b>	<b>77</b>
5.1	Introduction . . . . .	77
5.2	Asynchronous reproduction . . . . .	80
5.2.1	The model . . . . .	80
5.2.2	Analysis: neutral conditions . . . . .	81
5.2.3	Results: copy number distribution . . . . .	83
5.2.4	Results: population persistence . . . . .	85
5.3	Synchronous reproduction . . . . .	85
5.3.1	The model . . . . .	85
5.3.2	Analysis: neutral conditions . . . . .	87
5.3.3	Results: copy number distribution and population persistence . . . . .	87
5.4	Discussion . . . . .	88
<b>6</b>	<b>Conclusion</b>	<b>93</b>
	<b>References</b>	<b>97</b>
<b>A</b>	<b>Supplementary figures to chapter 3</b>	<b>131</b>
<b>B</b>	<b>Mathematical notes</b>	<b>137</b>
<b>C</b>	<b>Supplementary figures to chapter 4</b>	<b>143</b>
	<b>Declaration</b>	<b>149</b>

## INTRODUCTION

The genomes of all organisms are mostly composed of one or more chromosomes, which contain the vast majority of their genetic information. There are also, however, a variety of extrachromosomal genetic elements found in different organisms which exist independently from the chromosomal component of the genome. Well-known examples include the plasmids found in prokaryotes (and some fungi) and the organelle (mitochondrial and plastid) genomes of eukaryotes (Rush and Misra 1985); other types include extrachromosomal circular DNA, found in a wide variety of eukaryotes, and accessory or supernumerary chromosomes found in many organisms (Camaracho et al. 2000).

There are few properties that are common to all the various types of extrachromosomal element. They are often nonessential, or accessory, genetic elements which are not found in all individuals of a species, and therefore must not carry any essential genes (indeed, in the case of plasmids, this is often taken as a defining characteristic). However, the genomes of mitochondria and chloroplasts provide a counterexample, as they are absolutely vital to their host cells. They often exist in more copies, often many more copies, than the chromosomal DNA; but some plasmids exist in only one copy per cell. Occasionally they are horizontally transferable, but this is by no means universal. The one universal constant is that they are somehow separated from the chromosome or chromosomes, which exempts them from the strict control of chromosomal replication or segregation. Some extrachromosomal elements replace this with their own regulation, while others are simply less strictly regulated; in either case, the result is that some component of the life cycle of the extrachromosomal element occurs independently of the chromosome. The inheritance of traits carried on extrachromosomal elements is therefore distinct from chromosomal inheritance.

The first extrachromosomal genetic elements were discovered on the basis of the non-Mendelian inheritance of the traits they carried. Variegation of leaves in plants played an important role here: the first work on genetics of extrachromosomal genetic elements was the discovery that the plastids of plants had heritable genetic information, which was responsible for variegation in some plant species (Baur 1908; Correns 1908; Hagemann 2000). Later, McClintock (1932) found the first examples of extrachromosomal circles in variegated corn leaves. Bacterial plasmids entered the scene a few decades later, with the discovery of horizontally transmissible drug resistance factors in bacteria leading to an explosion of interest (Lederberg 1952;

Falkow 1975; Helinski 2022).

Around the same time that the first extrachromosomal elements were being discovered, the foundations of population genetics were being laid by the work of Fisher (1923) and Haldane (1927). Over the subsequent century, the theoretical population genetics of chromosomal genes has developed extensively, but perhaps unsurprisingly, extrachromosomal genetic elements have been less deeply explored. The extrachromosomal nature of these elements provides unique considerations for their inheritance, which need to be taken into account in the development of models of their population genetics. The wide diversity of different types of extrachromosomal element mean that these considerations will be very different for different types of element, and models will generally be quite specific to a particular genetic system and corresponding set of mechanisms.

\* \* \*

Evolutionary rescue occurs when a population which would have gone extinct due to maladaptation is instead rescued by an evolutionary change that leaves the population better adapted to its environment and able to grow and thrive (Carlson et al. 2014). Rescue therefore lies at the interface between ecology and evolution: the central process of interest is an evolutionary change, but in a context where the demography of the population cannot be ignored. Once an environmental change to which a population is sufficiently maladapted to have absolute fitness below one occurs, a race between the demographic decline of the population and the processes of adaptive evolution begins. The demography of the initial population as it goes extinct plays an important role, as the declining population makes adaptive evolution less likely as it provides less and less mutational input and loses standing genetic variation. But the demography of the new, adapted population also has an important role to play: this population will initially consist of a small number of individuals (if the allele that rescue the population arises as a novel mutation, the population will initially consist of a single individual), and in the early stages of rescue it will be in danger of loss to demographic stochasticity despite its higher absolute fitness.

The name evolutionary rescue was introduced, and some of the first models of evolutionary rescue were developed, by Gomulkiewicz and Holt (1995). The most common modelling approach today, where a branching process is used to model the establishment of the allele that rescues the population, was introduced by Orr and Unckless (2008). Today, there is a panoply of different models of evolutionary rescue looking at a variety of different ecological contexts and underlying genetic bases for rescue (e.g. Uecker and Hermisson 2011; Orr and Unckless 2014; Tazzyman and Bonhoeffer 2014; Santer and Uecker 2020; Geoffroy and Uecker 2023; Madgwick et al. 2024). The study of evolutionary rescue has been largely driven by theorists, and models make up an important part of its literature. This may be because it is difficult to show definitively that any particular pro-

---

posed natural example of rescue really is evolutionary rescue, since it requires matched genetic and demographic data from the population (Carlson et al. 2014). Nonetheless, there are some fairly well-confirmed examples, and an empirical literature on rescue in experimental evolution (Martin et al. 2013). To see the importance of evolutionary rescue, however, one need only think of the example of antibiotic resistant bacteria. The development of antibiotic resistance is a very serious problem in the clinical context, which threatens our ability to treat bacterial infections (Bennett 2008; Carattoli 2013). The emergence of resistance in a population subjected to antibiotic treatment is also a classic evolutionary rescue scenario (provided of course that the mechanism of resistance has a genetic basis, but this is usually the case).

\* \* \*

In this thesis, I bring evolutionary rescue and extrachromosomal genetic elements together. I develop and present models of evolutionary rescue scenarios in which the underlying genetic basis for rescue is a gene carried on an extrachromosomal element. These models will involve two different types of extrachromosomal element, plasmids in bacteria and extrachromosomal circular DNA (eccDNA) in yeast. As mentioned above, the specific biology of the type of extrachromosomal element is of central importance to the population genetics of genes on that element. I will not go into more detail here about the biology of plasmids, however, because Chapter 2 provides an extensive overview of all aspects of plasmids. Aimed at an audience of modellers—who are likely to have basic biological knowledge, but not to be familiar with plasmids specifically—the chapter covers a wide variety of topics, including the abundance and diversity of plasmids, the genes they carry, their regulation of their own replication, horizontal gene transfer, and their interactions with other plasmids, their hosts, and other genetic elements. The goal of the project that lead to this chapter was to provide an introduction that would let modellers interested in plasmids get up to speed quickly on the background they need; to that end, the chapter also includes discussion of common methods and open problems in plasmid modelling.

In the process of writing Chapter 2, I wished to make a figure showing the distribution of the lengths of plasmids observed in bacteria (what is now Figure 2.1). In the process, I noticed that the shape of the distribution in my figure was qualitatively different than the shape in comparable figures I had seen in other papers. After a bit of puzzlement, I realized that it was customary to use a logarithmic scale for the plasmid length axis, as opposed to my linear scale, and this meant that most of the figures in the literature were actually showing the distribution of the log-length of plasmids, which has a notably different shape. Although that fact that the shape of the distribution of a quantity and its logarithm differ may seem a technical mathematical point, the difference in shape can affect biological

inferences we draw from the distributions: for example, the distribution of log-lengths of plasmids is obviously bimodal, while the distribution of lengths is not. Chapter 3 is a letter written to point out this consideration to plasmid biologists, and to argue for being careful to make biologically motivated choices for the transformations used in presenting data.

In Chapter 4, I finally get down to the business of evolutionary rescue. In this chapter, I develop and analyze a model of evolutionary rescue by plasmid-mediated heterozygosity. In this scenario, not only is the gene responsible for rescuing the population located on a multicopy plasmid and not on the chromosome, but also there is not a single allele that can rescue the population: instead, two distinct alleles must be maintained together in a cell to provide the phenotype needed for rescue. In this heterozygote-advantage context, the survival of the population requires the growth of heterozygote cells to exceed the constant loss of heterozygosity caused by the random segregation of plasmid copies at host cell division. I determine the establishment probabilities of new rescue mutations and rescue probabilities of the population. These are both zero (that is, establishment and rescue become impossible) when a condition on the heterozygote fitness and plasmid copy number is not satisfied. I derive an exact analytic expression for this threshold condition. I present some smaller results on the rescue process, and also explore the effect of plasmid fusion on establishment.

In Chapter 5, I turn away from plasmids to examine extrachromosomal circular DNA (eccDNA), small circular DNA molecules found in the nuclei of a wide variety of eukaryotes. I again develop an evolutionary rescue model, this time with yeast hosts carrying eccDNA. Like the plasmid model, there is an additional aspect to the evolutionary rescue scenario beyond just the presence of the rescue allele on an extrachromosomal genetic element. Since an eccDNA copy does not carry its “own” genes, but is a copy of a region of the chromosome, the evolutionary change leading to rescue in this case will not be the arrival and establishment of a novel mutation, but a change in copy number of an existing gene, which will be amplified on the eccDNA. I present analytic and simulation results showing the distribution of eccDNA copy numbers and the probability of rescue of the population. In particular, I examine the effect of our assumptions about eccDNA replication and the degree of bias in the segregation of eccDNA copies on rescue.

---

## A MATHEMATICIAN'S GUIDE TO PLASMIDS: AN INTRODUCTION TO PLASMID BIOLOGY FOR MODELLERS

**Abstract.** *Plasmids, extrachromosomal DNA molecules commonly found in bacterial and archaeal cells, play an important role in bacterial genetics and evolution. Our understanding of plasmid biology has been greatly furthered by the development of mathematical models, and there are many questions about plasmids which models would be useful in answering. In this review, we present an introductory, yet comprehensive, overview of the biology of plasmids suitable for modellers unfamiliar with plasmids who want to get up to speed and to begin working on plasmid-related models. In addition to reviewing the diversity of plasmids and the genes they carry, their key physiological functions, and interactions between plasmid and host, we also highlight selected plasmid topics which may be of particular interest to modellers and areas where there is a particular need for theoretical development. The world of plasmids holds a great variety of subjects that will interest mathematical biologists, and introducing new modellers to the subject will help to expand the existing body of plasmid theory.*

### 2.1 Introduction

Plasmids are extrachromosomal DNA molecules common in many bacteria (Summers 1996). They replicate independently from the chromosome (and from other DNA molecules in the cell), and often exist in the cell in multiple copies. They can be transmitted vertically to daughter cells on host cell division and in some cases horizontally to other bacteria. The simplest plasmids are effectively parasites of their hosts: they colonize the host and use its cellular machinery to reproduce themselves. But they also form part of the bacterial genome, and genes located on plasmids have effects on the metabolic processes of their hosts, on the host phenotype, and therefore on host fitness. The most studied of these plasmid-borne genes are antibiotic resistance genes, which are a serious threat to the continuing effectiveness of antibiotics in clinical use (Falkow 1975; Carattoli 2013).

The term “plasmid” was coined by Lederberg (1952) to refer to any genetic determinant outside the chromosome, including chloroplast and mitochondrial genomes and certain viruses, but is now restricted to the simple extrachromosomal DNA molecules found in bacteria, archaea, and some

---

This chapter has been published as: Ian Dewan and Hildegard Uecker (2023). “A mathematician’s guide to plasmids: an introduction to plasmid biology for modellers”. In: *Microbiology* 169.7, DOI: 10.1099/mic.0.001362

eukaryotes. Historically, plasmids were of interest for two primary reasons: horizontal transfer between cells and antibiotic resistance. The earliest plasmids discovered were capable of transferring themselves between bacterial cells (Falkow 1975, p. 9), and this novel method of horizontal gene transfer was of great interest to bacterial geneticists. Antibiotic resistance plasmids were also discovered early on: this greatly increased interest in plasmids as a clinically important contributor to the spread of antibiotic resistance (Helinski 2022). Today, a much greater variety of plasmids are known and studied by biologists. As components of the bacterial genome, plasmids play an important role in the evolution of bacterial populations; therefore, a clear understanding of all aspects of plasmid biology is necessary to a full understanding of bacterial genetics. Plasmids are also of interest in their own right, as evolving biological entities.

Mathematical modelling can make, and has already made, an important contribution to understanding the biology of plasmids and their role in the ecology and evolution of bacteria. Existing literature (reviewed in Hernández-Beltrán et al. 2021) includes models focusing both on fundamental plasmid biology and on particular contributions of plasmids to bacterial populations, especially to antibiotic resistance. Mathematical approaches are diverse and depend on the question at hand: we briefly review them in Box 2.1. While a large number of plasmid models exist, this number is still small in the light of their biological interest. For those interested in learning about plasmids, there are a large variety of existing reviews covering plasmids and their evolution generally (Rodríguez-Beltrán et al. 2021a; Hall et al. 2021b; Helinski 2022), the diversity of plasmids (Fernandez-Lopez et al. 2017; Garcillán-Barcia et al. 2011; Hülter et al. 2017), antibiotic resistance plasmids (Bennett 2008; Carattoli 2013; Gama et al. 2018; Zwanzig 2021), the physiological functions of plasmids (Bahl et al. 2009; Bingle and Thomas 2001; Solar and Espinosa 2002; Solar et al. 1998; Novick 1987; Zielenkiewicz and Cegłowski 2001), and interactions between plasmids and between hosts and plasmids (Dionisio et al. 2019; Gama et al. 2018; Garcillán-Barcia and Cruz 2008; Getino and Cruz 2018; San Millan and MacLean 2019; Vial and Hommais 2019). Why therefore another review article? The diversity and complexity of plasmids can easily be overwhelming for those without prior knowledge, which includes many theoretical biologists who originally have a background in a subject other than biology such as mathematics, physics, or computer science. In particular, as soon as a novice sets up a model, many questions about meaningful biological assumptions appear. To lower the hurdle to start working on plasmids, we here present an overview of plasmid biology explicitly targeted at modellers new to the field. An introduction to plasmids for modellers could certainly have various levels of biological detail and complexity: here, we want to go beyond a basic and largely conceptual level. From our own experience, we mostly see the need for a review that, while

introductory, does not oversimplify the biology. While no prior knowledge about plasmids is required to read the article, some background about bacteria and microbiology is assumed. In the interest of accessibility to modellers with differing knowledge of microbiology, we provide a glossary of terms used: terms in the glossary are marked with an asterisk on first use.

---

*Box 2.1: Modelling approaches used in plasmid models.*

A variety of methods have been applied to modelling plasmids and their hosts. The oldest and most widely used are deterministic differential equation models that divide the host population into groups based on plasmid content, in the simplest case just plasmid-carrying and plasmid-free bacteria. These models track the population sizes of the groups, coupling the population dynamics with the dynamics of the plasmid. This technique has been applied to exploring the effects of conjugation on plasmid dynamics (Levin et al. 1979; Lundquist and Levin 1986) and to the “plasmid paradox” (Stewart and Levin 1977; Levin and Stewart 1980; Bergstrom et al. 2000; Lili et al. 2007) (discussed further in Boxes 2.4 and 2.5 respectively), as well as many other questions (Condit and Levin 1990; Mongold 1992; Rankin et al. 2012; Lehtinen et al. 2021). Some models use difference equations in a similar fashion (Hoeven 1985; Ponciano et al. 2007). Stochastic simulations of such models, usually with stochastic rates paralleling the transition rates of the deterministic models, are also common (Ponciano et al. 2007; Tepekule et al. 2019). Other models have used analytical stochastic approaches (Novick and Hoppensteadt 1978); branching process models are particularly useful when modelling the invasion of a plasmid or the early spread of a novel plasmid variant (Tazzyman and Bonhoeffer 2013; Tazzyman and Bonhoeffer 2014; Santer and Uecker 2020). While most plasmid models include population dynamics, some models, in the tradition of population genetics, fix the population size. Based on extensions of the classical Moran model (Moran 1958), these models consider only changes in the relative frequencies of cell types: these can be deterministic (Rankin et al. 2011; Santer et al. 2022) or stochastic (Novozhilov et al. 2005; Santer et al. 2022). The final common approach is the use of individual-based simulations (Summers et al. 1993; Krone et al. 2007; Werisch et al. 2017; Domingues et al. 2021), which allow a realistic incorporation of many biological processes at the cost of little analytical tractability. Although most models are of a host population, there are some models of plasmids within a single cell, both deterministic (Lee and Bailey 1984; Brendel and Perelson 1993) and stochastic (Seneta and Tavaré 1983). Of course, many studies combine multiple modelling methods: particularly common are the combination of an analytically tractable model of some sort with a simulation (Ponciano et al. 2007; Santer and Uecker 2020; Domingues et al. 2021).

The biology, the specific question, and the goal of the model determine which processes and features are taken into account. Models of a host population usually incorporate a few common biological processes. The most basic of these is the host population dynamics, which includes a growth model (often either exponential or Monod growth), competition between cells (often via a Lotka-Volterra model or explicit inclusion of a common resource), washout of cells (for chemostat models), and fitness costs or benefits to carrying particular plasmids. Horizontal transfer is frequently included: see Box 2.4 for a discussion of modelling horizontal transfer. Finally, loss of the plasmid during segregation is also a common model component: this can be modelled explicitly as a component of a stochastic model, but in ODE models it is usually modelled as a flux of cells from a plasmid-carrying to a plasmid-free compartment at a constant per capita rate. Other model features include physiological changes in the cells or age structure in the population (Lundquist and Levin 1986; Summers et al. 1993), migration (Rankin et al. 2011), explicit modelling of the segregation of plasmids on host cell division (Hoeven 1985; Santer and Uecker 2020) and plasmid replication during the cell cycle (Münch et al. 2019), and many others.

The parameters associated with these models usually need to be assigned values, although for some models it may be possible to obtain analytical results (e.g. Bergstrom et al. 2000). Many studies incorporate experimental and modelling work into the same study (Levin et al. 1979; Krone et al. 2007; Garoña et al. 2023), and then the parameters can be estimated from the accompanying experiments. In the absence of codesigned experiments, parameter values or plausible parameter ranges usually need to be obtained from the literature, either directly or by additional parameter estimation from published data. For models with one or few parameters, it may be possible to explicitly explore the plausible parameter space. Otherwise, depending on the goals and scope of the study, a parameter sensitivity analysis might be required. Parameter estimates are usually obtained from *in vitro* experimental studies. Parameter estimation is often non-trivial (see also Box 4 for the estimation of transfer rates). The parameterization of *in vivo* models is notoriously difficult. An example of thoughtful and sophisticated parameter estimation for an *in vivo* model is the method developed by Tepekule et al. (2019), who make use of various kinds of data and observations from multiple different sources, including microbiome time series data, to parameterize their model.

---

In the next three sections, we focus on plasmids themselves, discussing the variety of plasmids, their gene content, and the important physiological functions they carry out; in the fifth section, we explicitly turn to the relationship between plasmids and their hosts, considering plasmid-host interactions, as well as interactions between plasmids. In four boxes, we

highlight selected key topics related to plasmids and their hosts that are especially captivating for modellers, and provide a few references to relevant modelling work as a starting point for further reading. At the end of the article, we discuss several areas where we see particular scope (and need) for more modelling. Hopefully, the review will enable readers to begin working on models that will expand our understanding of plasmids and their hosts.

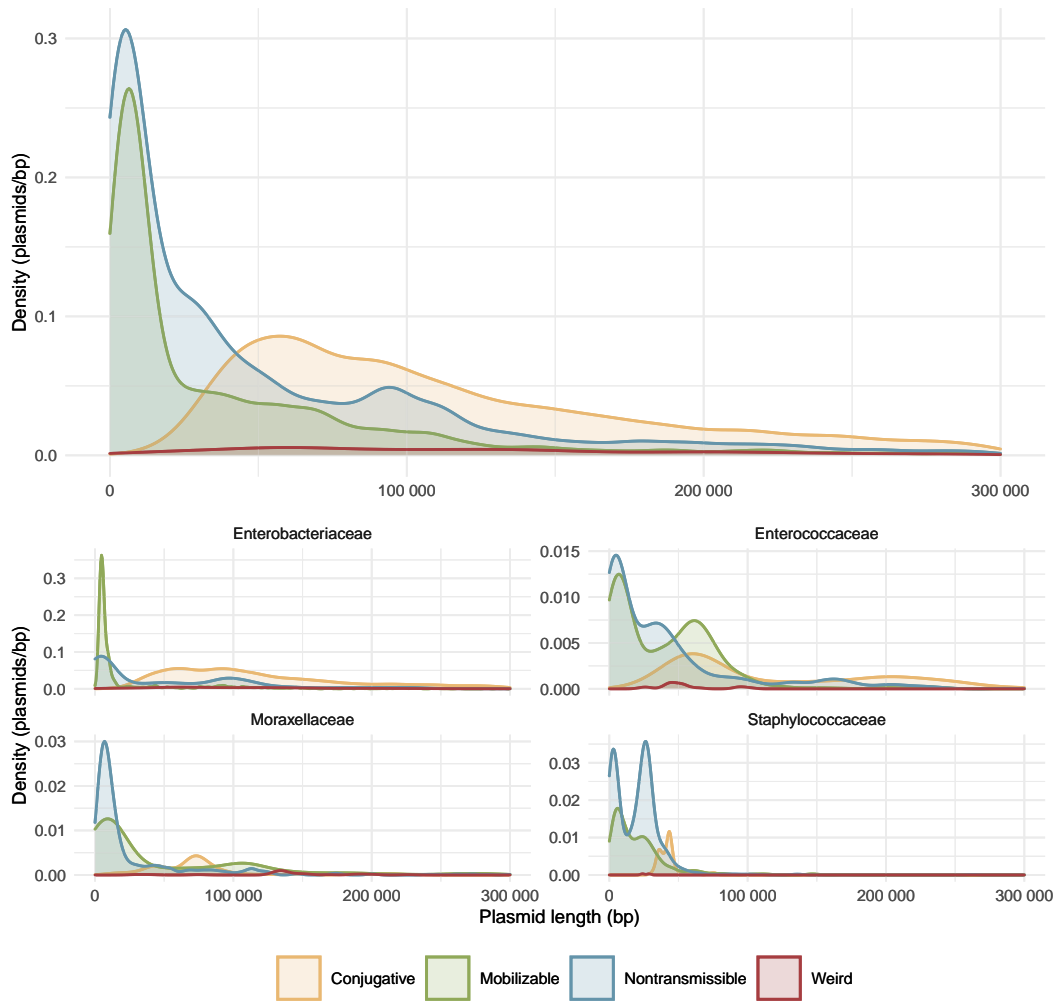
### 2.2 The natural history of plasmids

Plasmids are extremely widespread in the wild: bacteria and archaea from almost all taxonomic groups have been found to contain them (Rodríguez-Beltrán et al. 2021a). They also exist in eukaryotes, particularly fungi (see e.g. Kiyosawa et al. 1993; Qin et al. 1993), and even in mitochondria within eukaryotic cells (Silliker and Cummings 1990); but we shall focus our attention on bacterial (and archaeal) plasmids. Precise quantification of the frequency of plasmids in different taxa is complicated by variable search effort: plasmids from bacteria of clinical importance or from model species are overrepresented among sequenced plasmids, particularly those from the Gammaproteobacteria, where antibiotic resistance plasmids have been the subject of extensive study (Shintani et al. 2015).

Plasmids themselves may be naturally classified on the basis of biologically relevant properties. Perhaps the most important of these is the capacity for horizontal transmission\* between bacteria by conjugation (the details of conjugation are discussed below in section 2.4.3). Conjugative plasmids carry all of the genes necessary for conjugation, and are therefore self-transmissible; mobilizable plasmids do not carry the full transfer machinery, but have a sufficient subset to be able to undergo conjugation in the presence of a conjugative plasmid which supplies the rest; the remaining plasmids are nontransmissible. It has been estimated that about half of all plasmids are nontransmissible, with the remaining half approximately equally divided between conjugative and mobilizable (Smillie et al. 2010; Coluzzi et al. 2022); however, recent results suggest that a large fraction of plasmids that are traditionally classified as nontransmissible might in fact be mobilizable, and mobilizable plasmids might make up the majority of all plasmids (Ares-Arroyo et al. 2022). The capacity for conjugation is connected to two other biologically relevant properties: plasmid size and the number of copies of the plasmid maintained in a host cell. In general, conjugative plasmids are large and have a low copy number (typically one or a few copies per cell), while small plasmids, which tend to have a high copy number, are more often mobilizable or nontransmissible (Rodríguez-Beltrán et al. 2021a): see Figure 2.1. Nontransmissible plasmids exhibit a much larger range of sizes than transmissible plasmids, and there are some nontransmissible plasmids larger even than typical conjugative plasmids,

which may be in the process of becoming accessory chromosomes\* (Smillie et al. 2010; Hall et al. 2021a). Plasmids may also be classified as broad or narrow host range plasmids, depending on whether they are capable of becoming established in a large variety of hosts, or are reliant on a particular group of bacteria (Bahl et al. 2009). Plasmids are also classified based on topology, which has effects particularly on the biology of replication: the majority of plasmids are circular, but some linear plasmids have been found (Ravin 2011; Chen et al. 1993; Casjens et al. 2000; Ventura et al. 2007).

Beyond this division of plasmids based on broad properties, it is natural to develop a systematic biological classification of plasmid types. The classical approach is to divide plasmids into *incompatibility groups*. We say that two plasmids are in the same incompatibility group if they are incompatible; that is, if they cannot be stably maintained together in the same cell line (Novick et al. 1976). This concept is perhaps a bit counterintuitive—plasmids are in the same group if they are incompatible with each other—but it produces natural kinds because the cause of incompatibility is usually interference between common regulatory systems on the incompatible plasmids (Novick 1987): this mechanism of incompatibility is shown in Figure 2.2A, and discussed further in Section 2.4.1. Because incompatible plasmids share fundamental genes, incompatibility groups may be looked on as “species” of plasmids, and the production of new incompatibility groups as plasmid speciation (Chabbert and Roussel 1977; Sýkora 1992). It has been suggested that in the modern world of abundant plasmid sequences, the concept of incompatibility groups could be replaced altogether in favour of directly comparing sequences of fundamental plasmid genes; this has been argued to address perceived conceptual limitations such as the ability of single point mutations to create new incompatibility groups (Petersen 2011; Garcillán-Barcia et al. 2009). These methods have used the sequences of the replication proteins (REP classification), the conjugative transfer proteins (MOB classification), or the entire plasmid to reconstruct phylogenetic relationships between plasmids (Francia et al. 2004; Fernandez-Lopez et al. 2017; Redondo-Salvo et al. 2020; Redondo-Salvo et al. 2021; Garcillán-Barcia et al. 2023). We shall see below that plasmids can fuse with each other and separate again, and they gain or lose segments by the movement of mobile genetic elements or by recombination: therefore the concepts of a plasmid type and the identity of a plasmid over time can be quite fluid in practice, with plasmids that belong to multiple incompatibility groups, or that significantly change their structure or gene content while remaining in the same incompatibility group. For the purposes of modelling, whether different plasmids count as being of the same type will generally simply have to be imposed by the modeller.



**Figure 2.1:** Distribution of plasmid lengths among mobility classes of plasmids. Plasmid length data is taken from PLSDB (Schmartz et al. 2021, retrieved 1 March 2023), and the distribution derived by kernel density estimation in R (R Core Team 2020) with the `ggplot2` library (Wickham 2016). The top panel shows all plasmids in the database; the bottom four panels show plasmids from the four bacterial families with the most sequences in the database. The graph was cut off at 300 000 bp for reasons of scale: 1844 plasmids in the database (5.34 % of the total) are longer than the limit; the longest plasmid in the database is 4 605 385 bp. The plasmid sequences in PLSDB have been annotated with MOB-typer (Robertson and Nash 2018), to identify putative relaxases\* or conjugative genes (see the explanation of these terms in section 2.4.3). Those plasmids with neither are nontransmissible, those with both conjugative, and those with only the relaxase mobilizable; note that this means that those mobilizable plasmids with an *oriT* but no relaxase are not recognized as mobilizable. The “weird” class includes those plasmids which had conjugative genes but no relaxase: these would be nontransmissible, but nonetheless have all the genes for a secretion system and mating pair formation. It is possible that these are misidentified conjugative plasmids, which have an unknown relaxase, or misidentified mobilizable or nontransmissible plasmids, which do not actually have conjugative genes.

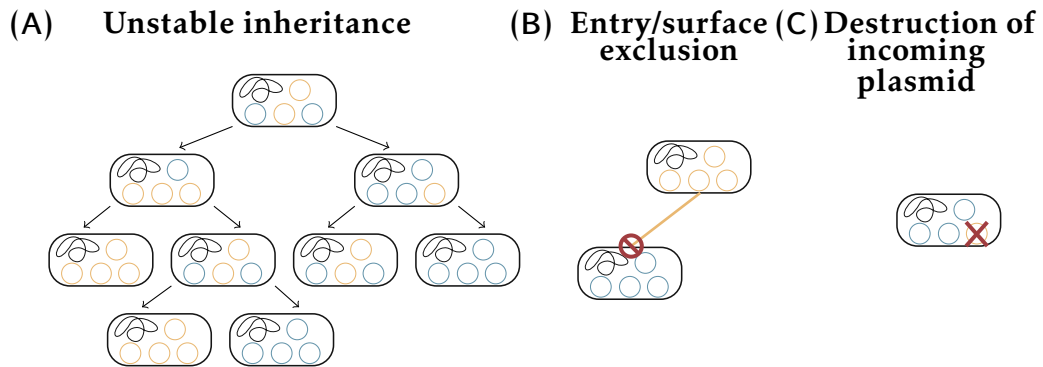


Figure 2.2: Mechanisms limiting coinfection of bacteria by multiple plasmids. (A) Unstable inheritance. Because the two plasmid types share a common copy number, random segregation at host cell division eventually ensures that they are separated into distinct hosts carrying only one plasmid type. (B) Surface and entry exclusion. The plasmids encode proteins which prevent the host cell from being a recipient in conjugation. (C) Destruction of the incoming plasmid. Novel plasmids are degraded by plasmid-encoded immune systems, such as CRISPR-Cas\* or restriction enzymes\* once in the host cell.

### 2.3 The anatomy of plasmids

The understanding of plasmids must of course start with the plasmid itself, as a molecule, a sequence of nucleotides, and a collection of genes.

#### 2.3.1 Genes carried on plasmids

The genes carried on a plasmid may be broadly divided into two groups based on their function: those with plasmid-specific functions and those which primarily affect the host phenotype. The genes responsible for plasmids' own housekeeping functions make up the plasmid backbone, while the other genes carried by a plasmid are called payload or accessory genes (Bahl et al. 2009).

The housekeeping genes of the plasmid typically fall into a few classes: there are the genes responsible for the replication of the plasmid, a stability system that ensures the plasmid is stably inherited across host generations, and, if the plasmid is transmissible, the genes responsible for conjugation or mobilization and overcoming host defences to establish the plasmid in the recipient cell (Garcillán-Barcia et al. 2011). The most important of these are the genes controlling replication. They will include the *oriV* (origin of vegetative replication, the sequence at which plasmid replication begins) as well as the genes for any proteins necessary for replication and for the components of a system to control replication of the plasmid (Garcillán-Barcia et al. 2011). Typically, the origin of replication and a few genes involved in

replication from the minimum subsequence of the plasmid that is capable of replication in a host, called the “basic replicon\*”, which constitutes an absolutely minimal plasmid (Kollek et al. 1978).

The simplest plasmids—often called cryptic plasmids—may have no payload genes (Burian et al. 1997; Zaleski et al. 2006; Moran and Hall 2019); they therefore do nothing but hang around in the cell. But many plasmids, and for obvious reasons the plasmids of greatest interest, carry genes which contribute to the phenotype of the host cell. The most studied plasmid payload genes are antibiotic resistance genes. These are found extremely frequently, in a wide variety of hosts, and encoding resistance to a wide variety of antibiotics (Carattoli 2013; Falkow 1975; Wang et al. 2021); some aspects of plasmid-borne antibiotic resistance and models thereof are discussed in Box 2.2. Plasmids may also encode resistance to other environmental dangers, such as heavy metals or other toxins (Silver 1992; Kothari et al. 2019; Mukhtar et al. 2019). A second large class of payload genes provide some new metabolic process (Palomino et al. 2022; Yu et al. 1996); this might include, for example, the ability to metabolize a new substrate for growth (Chakrabarty 1974; Meer 1994; Pandeeti et al. 2012). Virulence factors are also often found on plasmids (Portnoy and Martinez 1985). Plasmids also carry genes responsible for social interactions between bacteria, discussed in Section 2.5.2. Particularly interesting examples of plasmid-borne traits include the gall-forming properties of the plant pathogen *Agrobacterium tumefaciens*, which are caused by the horizontal transfer of a portion of a plasmid to the host plant (Beijersbergen et al. 1994), and the symbiosis between nitrogen-fixing rhizobial bacteria and leguminous plants, in which both the genes responsible for nitrogen fixation and for interactions with the plant host are found on plasmids (Wardell et al. 2021). An extensive list of functions observed on plasmids may be found in Summers 1996, 4–5. Sometimes, even essential genes are found on plasmids (Anda et al. 2015; Wein et al. 2021), although extrachromosomal replicons with the unique copy of an essential gene are sometimes categorized as secondary chromosomes\* or chromids\* instead of plasmids.

---

*Box 2.2: Plasmids and antibiotic resistance.*

Antibiotic resistance genes are frequently located on plasmids (Doumith et al. 2011; Ramirez et al. 2019), and the serious threat of antibiotic resistance means that understanding the contribution of plasmids to the evolution of antibiotic resistance is one of the most important applications of plasmid biology. The reasons antibiotic resistance genes are so often found on plasmids may include that plasmids allow bacteria to temporarily acquire resistance genes that are not needed in nonselective environments (Wang et al. 2021), or that the presence of plasmids in multiple copies increases either the dosage of resistance genes (San Millan et al. 2015; Sandegren and Andersson 2009), or the

rate of emergence of new resistance alleles (Rodriguez-Beltran et al. 2018; San Millan et al. 2016), or both. Horizontal transmission of plasmids enables antibiotic resistance to spread very quickly in bacterial communities (Bennett 2008; Carattoli 2013; Lopatkin et al. 2017).

The human gut is a site of substantial plasmid transfer. Resistance plasmids within the gut continue to evolve (DelaFuente et al. 2022), and are transferred between members of the microbiota within individuals and spread across individuals, as shown by León-Sampedro et al. (2021) for the plasmid pOXA-48 in the gut of hospital patients. During antibiotic treatment, not only the pathogen population but also the patient's normal microbiome is exposed to antibiotic pressure, inadvertently selecting for resistance (so-called bystander selection). However, resistance plasmids are also found in human populations without strong exposure to antibiotics (Conteville and Vicente 2022). Resistance in commensal bacteria can be problematic for two reasons. Some commensal species such as *Escherichia coli*, *Klebsiella pneumoniae*, or *Staphylococcus aureus* can turn pathogenic and cause infections of the urinary tract, the lung, or wounds and are a major cause of nosocomial infections. Second, resistance plasmids may spread from commensal bacteria to obligate pathogens (Schwalbe et al. 1990; Archambaud et al. 1991; Carattoli 2003; Boekhoud et al. 2020).

Despite plasmids' immense clinical relevance, modelling studies are biased towards chromosomal resistance. Nonetheless, a body of theoretical literature on plasmid-mediated resistance has accumulated (Leclerc et al. 2019; Zwanzig 2021), and plasmids are increasingly attracting the attention of modellers. For example, Svava and Rankin (2011) developed an ODE model to determine under which treatment conditions—antibiotic dose and intervals between administrations—resistance on a conjugative plasmid would be favored over resistance on the chromosome. Their model also includes an incompatible plasmid without the resistance gene, leading to six different cell types that compete with each other. Other models have, for example, addressed how patterns of drug use determine the prevalence of resistance plasmids within the microbiota at the level of the individual (Tepekule et al. 2019) and the level of the population (Austin et al. 1997) and examined the interaction of plasmid-borne resistance with specific modes of drug action (Willms et al. 2006). Models furthermore often complement *in vitro* experiments to help explain and interpret the observed dynamics of plasmid-mediated antibiotic resistance (e.g. De Gelder et al. 2004; Rodriguez-Beltran et al. 2018).

---

Why particular genes are found on plasmids rather than on chromosomes is a longstanding question. Numerous hypotheses and models have been developed to explain the distribution of genes (see e.g. Eberhard 1989; Eberhard 1990; Rankin et al. 2011; Lehtinen et al. 2021; Wardell et al. 2021). Possible explanations include the advantages of mobility for local

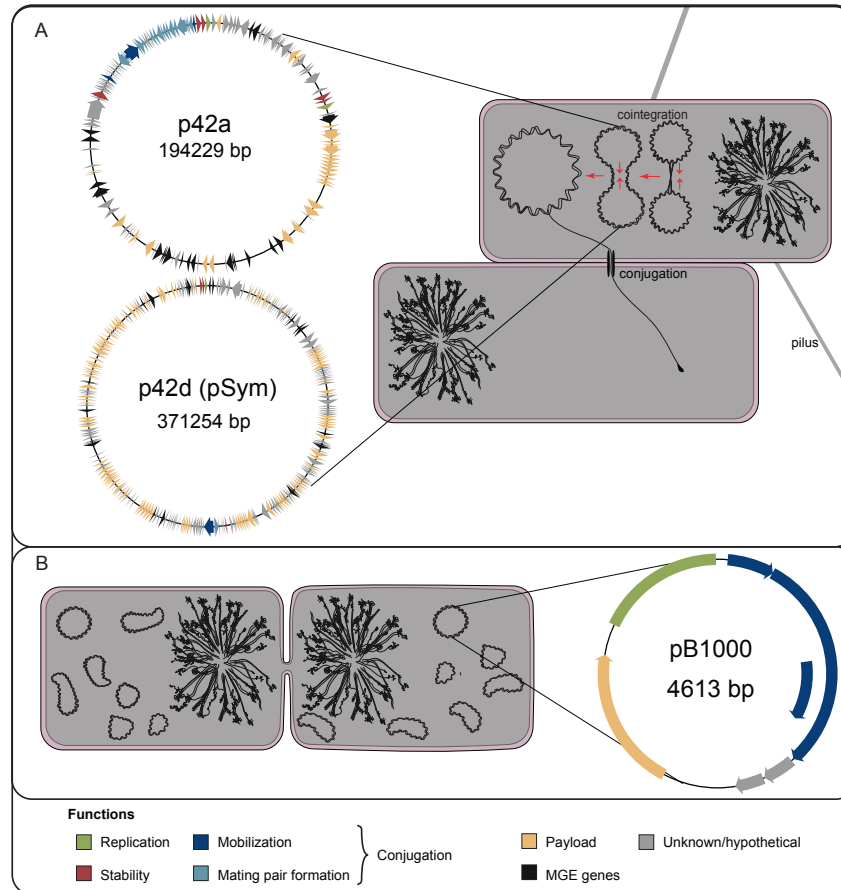
adaptation when there is patchy or temporally varying selection (Eberhard 1990; Wardell et al. 2021). In different conditions the same genes may be located on different replicons: it has been hypothesized that the majority (if not all) bacterial genes have spent time on plasmids and on chromosomes over evolutionary time (Eberhard 1990).

### 2.3.2 Genetic structure

The genes on plasmids are not arbitrarily distributed, but tend to have an organized structure. This structure is often modular, with genes with related functions located together on the plasmid; there are then segments of the plasmid for replication, conjugation, different payload genes, and so on (Garcillán-Barcia et al. 2011): see for example the plasmids depicted in Figure 2.3. In addition, the genes carried on plasmids are often located within other, nested mobile genetic elements\*: a plasmid-borne gene may be in a gene cassette\*, which is integrated into an integron\*, which is carried by a transposon\*, which is located on the plasmid (Bahl et al. 2009). The modular structure suggests that plasmids may frequently evolve by the gain or loss of entire modules, whether by recombination\* with chromosomes or other plasmids or by the movement of mobile genetic elements. This also means that plasmids are frequently genetic mosaics, which may combine components of several mobile genetic elements together with pieces of multiple original plasmids and chunks of chromosomes (Boyd et al. 1996; Osborn et al. 2000; Froehlich et al. 2005; Fecskeová et al. 2012; Spaková et al. 2013; Maida et al. 2014). Large plasmids may contain multiple basic replicons (e.g. F, see Gibbs et al. 1993); this means that they can replicate starting from any one of their several *oriV* sequences using the replication mechanism encoded by that basic replicon. Usually, only one will be active at a time, since otherwise there would be conflicts between the replication processes; often different replicons will be active in different hosts, extending the plasmid host range (Hülter et al. 2017).

### 2.3.3 Plasmid multimers

In addition to the formation of mosaics by accumulation of portions of plasmids and other replicons, several plasmid copies may fuse into one molecule, forming a plasmid multimer; plasmids frequently exist in hosts as a mixture of monomers and multimers of various sizes (Bedbrook and Ausubel 1976; Projan et al. 1987; Wein et al. 2020; Abe et al. 2021). Multimerization occurs not only between plasmids of the same type, but also of distinct types, forming plasmid cointegrates which exhibit the properties of both their components (Milliken and Clowes 1973; Nugent and Hedges 1979; Froehlich et al. 2005; Hülter et al. 2020; Taylor et al. 1981): this is depicted in Figure 2.3A. The formation of plasmid multimers and cointegrates is driven by homologous recombination\* or recombination of trans-



**Figure 2.3:** Illustration of important processes in plasmid biology with examples of three different plasmids. (A) The low copy number (approx. 1 copy per cell) plasmids p42a (conjugative) and p42d (nontransmissible) of *Rhizobium etli* CFN42 (González et al. 2006; González et al. 2003, GenBank accession numbers CP000134.1 and U80928.5) fuse into a cointegrate and then are transferred by conjugation from a host to a new recipient. The two cells have been pulled together by a pilus, which has then been retracted. A single strand of a plasmid copy is transferred, while the other strand remains within the cell in circular form. Afterwards, a second strand will be synthesized in both cells to generate double-stranded DNA. p42d is also referred to as pSym, since it contains most of the genes responsible for rhizobial symbiosis in this strain; it is normally horizontally transferred by cointegration with p42a (Brom et al. 2004; Tun-Garrido et al. 2003). (B) The copies of the small high copy number (approx. 14 copies per cell) resistance plasmid pB1000 (San Millan et al. 2009, GenBank accession number GU080070.1) have been segregated between daughter cells at host division; cell division is almost complete. Inside each bacterial cell, the plasmids and chromosome are depicted. The diagrams of the plasmids show the open reading frame\* of each gene on the plasmid as an arrow coloured by the function of the encoded protein. The region of pB1000 marked in the replication colour contains the plasmid origin of replication and the coding sequences for the RNAs involved in regulation of replication. The origins of replication of p42a and p42d are located inside the RepA gene; the origin of transfer of p42a is marked with a dot in the mobilization colour.

posable elements located on one or the other plasmid (Ohtsubo et al. 1981; Machida et al. 1982; Liu et al. 2021), and sometimes by more exotic processes (reviewed in Dionisio et al. 2019); to counteract multimerization, plasmids frequently encode multimer resolution systems.

## 2.4 The physiology of plasmids

The backbone genes of plasmids have functions related to the “life” of the plasmid itself, and are responsible for its maintenance in the host, its propagation (vertically or horizontally), and similar housekeeping. These functions constitute the physiology of the plasmid, and their presence is what distinguishes a plasmid from a simple fragment of DNA that might be picked up by transformation\*.

### 2.4.1 Replication, plasmid copy number, and their control

It is a defining attribute of plasmids that they replicate in their host cells autonomously from the chromosome. Although most plasmids rely partially on host proteins for replication, they also carry genes essential to their replication and responsible for its regulation. The regulation of replication ensures that the plasmid is present in the host population at a fixed number of copies per cell. Control of copy number is necessary to ensure the stable maintenance of the plasmid: if the copy number falls too low, there is a greater risk of producing plasmid free segregants, while if the copy number is allowed to grow without limit, the plasmid will impose a heavy fitness cost on its hosts; Box 2.3 discusses the multilevel selection acting on traits like plasmid copy number. We have already seen that there are low and high copy number plasmids; but there is much copy number variation among high copy number plasmids, from around ten copies to hundreds (e.g. Projan et al. 1987; Santos-Lopez et al. 2016). The copy number is not exactly identical from cell to cell, but is subject to some noise, which can be reduced by a partitioning system (Ng et al. 2010). The plasmid copy number may of course be subject to evolution, even over short time scales when the plasmid carries an important host function (San Millan et al. 2016; Mei et al. 2019; Hernandez-Beltran et al. 2022).

The replication control system ensures that the replication rate per plasmid copy per host generation is greater than one when the copy number is below the target and below one when the copy number exceeds the target (Summers 1996, p. 31), so that the copy number is increased by replication or reduced as host cell divisions dilute the plasmids faster than they replicate to reach the target. The rate of replication is regulated by negative feedback: one common mechanism is by control of the production of a plasmid-encoded Rep protein which is responsible for initiating the replication of the plasmid by binding to the *oriV* and recruiting the necessary enzymes for replication. Expression of Rep is repressed by some *trans-*

acting\* product encoded by the plasmid: a separate, specialized repressor\*, Rep itself, or another protein cotranscribed with Rep. As the plasmid copy number increases, more of this repressor is produced, and it prevents the production of Rep protein, creating a negative feedback loop (Bingle and Thomas 2001; Solar and Espinosa 2002; Burian et al. 1999; Nordström 2006; Rivera-Urbalejo et al. 2015; Summers 1996, 36–40).

This mechanism explains the origin of plasmid incompatibility due to a common basic replicon. Since the repressors are *trans*-acting, two distinct plasmids with the same basic replicon will contribute to repressing each other's replication, and therefore will share a common copy number. In the absence of selection for both plasmid types, this means that after a few generations, random segregation\* at host cell division will have separated the plasmid types into distinct host cells, and thus the two plasmids cannot be stably maintained together (Novick 1987); see Figure 2.2A.

When replication is permitted by the control system, the selection of the plasmid copies to be replicated seems to be at random in at least some cases (Reyes-Lamothe et al. 2013; Nordström and Gerdes 2003), although it is difficult to definitively confirm this experimentally, and it may not be the case for all plasmids (Garoña et al. 2023). The mechanisms of replication are reviewed in Solar et al. 1998.

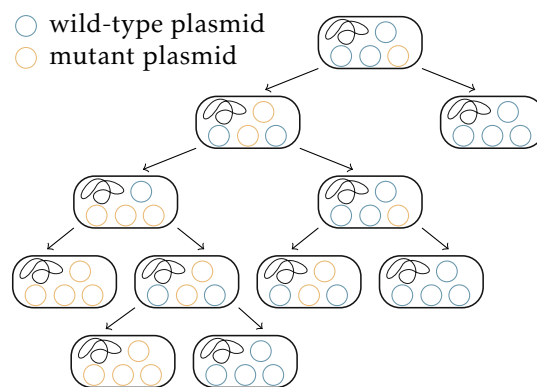
---

*Box 2.3: Plasmids and multilevel evolutionary dynamics.*

**Multilevel selection.** Because plasmids depend on a bacterial host to survive and propagate, they offer an excellent example of multilevel selection: the fitness of plasmids depends both on their own properties (of replication, maintenance, etc.) and indirectly on the fitness of their hosts (Eberhard 1990). As we have seen, plasmid-borne genes frequently affect host fitness as well as the fitness of the plasmid itself: this includes both direct effects on host phenotype and the indirect fitness effects discussed in section 2.5.1. Interactions between these two levels have important implications for the evolution of plasmids and for their contribution to the evolution of their hosts. For example, there is a trade-off between rates of vertical and horizontal transmission of plasmids (Bethke et al. 2022; Zwanzig 2021; Guynet et al. 2011): the costs to the host of conjugation mean that increased horizontal transmission reduces vertical transmission. A similar trade-off also leads to conflicting selection pressures on the plasmid copy number: a plasmid variant with a higher copy number outcompetes a variant with a lower copy number at the intracellular level, but if the copy number increases too much, the burden on the host cell may become too high (Paulsson 2002; Watve et al. 2010).

**Allele dynamics on multicopy plasmids.** An effect of multicopy plasmids on bacterial evolution is the possibility for loci on multi-

copy plasmids to be heterozygous (reviewed by Rodríguez-Beltrán et al. 2021a). The dynamics of alleles on multicopy plasmids depend on processes at two levels—intracellular plasmid replication and segregation and population dynamics at the cellular level. If the alleles have fitness effects, the host fitness depends on the plasmid composition within the cell, and various forms of dominance or heterozygote advantage are possible. As illustrated below, random segregation of plasmid copies at host cell division changes the plasmid composition from mother to daughter cells, which is termed segregational drift (Ilhan et al. 2018). In the long term, this leads to the loss of heterozygosity on plasmids unless there is selection for maintaining the two plasmid types together. Segregational drift does not alter the allele frequency in the bacterial population *per se*, but since it leads to the generation of wild-type homozygous cells, it reduces the number of cells carrying the mutant allele, which increases the strength of genetic drift. This means novel mutations on plasmids have a higher chance of being lost than those on a monoploid chromosome, at least in the absence of gene dosage effects (Santer and Uecker 2020; Garoña et al. 2021; Garoña et al. 2023). Replication of plasmids may also be a source of drift: selection of plasmids to replicate, if at least partially random, creates a “rich-get-richer” effect which leads to more extreme biases in plasmid content in offspring (Santer et al. 2022). The evolutionary dynamics on multicopy plasmids have recently received substantial attention from both experimentalists (San Millan et al. 2016; Garoña et al. 2021), theoreticians (Santer and Uecker 2020; Santer et al. 2022) and from both together (Ilhan et al. 2018; Rodríguez-Beltrán et al. 2018; Rodríguez-Beltrán et al. 2020; Garoña et al. 2023). Early models have studied the dynamics of incompatible multicopy plasmids, which is a closely related problem (Ishii et al. 1978; Novick and Hoppensteadt 1978; Cullum and Broda 1979).



*Schema of the effect of segregational drift on a novel allele located on a plasmid (figure adapted from San Millan et al. 2016; Rodríguez-Beltrán et al. 2018). A novel allele originally arises on one copy of the plasmid, and by random segregation wild-type and mutant plasmids eventually end up in different, homozygous cells. We here describe one way*

of modelling the process mathematically (Santer and Uecker 2020). In the simplest case, cell dynamics are modelled by a birth-death process with per capita birth and death rates  $b_i$  and  $d_i$ , which may depend on the number of mutant plasmids  $i$  in a given cell. Prior to cell division, plasmids are replicated such that the cell contains twice the original number. Two possible models for plasmid replication are 'regular replication', where each copy is replicated exactly once, and 'random replication', where copies are randomly picked for replication one by one (cf. Novick and Hoppensteadt 1978). Each daughter cell then receives half of those plasmids with segregation being random with respect to mutant and wild-type variants. With  $n$  total plasmid copies and  $x$  mutant copies after plasmid replication, the probability that one daughter cell receives  $j$  and the other one  $x - j$  mutant copies is then given by  $(2 - \delta_{j,x/2}) \binom{x}{j} \binom{2n-x}{n-j} / \binom{2n}{n}$ , where  $\delta_{i,j}$  denotes Kronecker's delta.

---

#### 2.4.2 Segregation of plasmids at host cell division

When a host cell divides, the plasmid copies present end up in one or the other daughter cell (see Figure 2.3B). In the absence of any intervention by the plasmid, this will happen essentially at random: the plasmids in one portion of the cell when it divides go to one daughter, and the plasmids elsewhere go to the other. For plasmids with a high copy number, this may be sufficient to stably vertically transmit the plasmid: with  $n$  copies, the probability of producing a plasmid-free daughter by random segregation is  $2^{1-n}$ , which may be sufficiently small to be negligible. Indeed, some high copy number plasmids do not seem to have any other partitioning mechanism (Reyes-Lamothe et al. 2013). In practice, this is quite successful in reducing plasmid loss to a very low level (Lau et al. 2013).

For low copy number plasmids a further active partitioning system that ensures both daughter cells receive a copy of the plasmid is necessary. This can be done by localizing plasmids in separate regions of the cell before cell division (e.g. Guynet et al. 2011; Yao et al. 2007). A common class of such systems functions by encoding a protein complex and a *cis*-acting\* site on the plasmid. The protein complex binds a pair of plasmids together at their respective copies of *cis*-acting site, then physically separates the pair at cell division, possibly by binding to a point on the cell membrane or another host structure (Zielenkiewicz and Ceglowski 2001).

Even high copy number plasmids need to ensure that they are not made unstable by multimerization. The replication control systems we have discussed above control the number of copies of *the basic replicon* in the cell; this means that if multimers of the plasmid form, the number of physical molecules containing the plasmid falls without a derepression of replication (as the number of copies of the basic replicon remains the same). Since the number of physical molecules is what determines the probability of

producing a plasmid-free daughter cell, multimerization reduces the stability of the plasmid. Moreover, larger plasmid multimers are replicated more often, since more copies of the basic replicon means more *oriVs* at which to initiate replication. A cell with multimers is thus likely to get more multimers, and reduce the plasmid's stability further, in the so-called "dimer catastrophe" (Summers et al. 1993). To counteract this effect, plasmids generally encode a multimer resolution system, which converts multimers back to monomers by site-specific recombination\* (Croizat et al. 2014; Summers and Sherratt 1984; Zielenkiewicz and Cegłowski 2001).

Another group of plasmid functions, variously called postsegregational killing, host killing, toxin-antitoxin systems, or plasmid addiction, are often included with stability systems, but we shall consider them separately in Section 2.4.4 below.

### 2.4.3 Horizontal transfer by conjugation

Perhaps the most famous physiological function of plasmids is the ability to transfer themselves from one host to another. Transfer by conjugation requires physical contact between the donor and recipient cell, which is usually, at least in gram-negative\* bacteria, achieved using a proteinaceous structure called a pilus: see Figure 2.3A. Usually, a single strand of the plasmid is transferred into the recipient cell, and the second strand is then resynthesized in both cells, meaning that conjugation also involves duplication of the plasmid; however in some plasmids of *Streptomyces*, the entire double-stranded plasmid is translocated from the donor to recipient (Possoz et al. 2008; Thoma and Muth 2012). In some cases, there is the possibility of retrotransfer, the transfer of genetic material from the recipient back to the host, during conjugation (Szpirer et al. 1999). Some plasmids can undergo horizontal transfer through more exotic mechanisms (Erdmann et al. 2017), sometimes involving cooperation between donor and recipient (Blesa et al. 2017), but here we focus on common conjugation.

In addition to the pili, which bring the donor and recipient into contact so that conjugation can occur, the typical conjugation machinery consists of four parts. The first is the *oriT*, or origin of transfer, the sequence on the plasmid at which transfer begins. One strand of the plasmid is nicked\* here by the second component, a relaxase\*, which binds to the free end of the strand. The relaxase interacts with a type IV coupling protein (T<sub>4</sub>CP, the third component), and is transferred, together with the attached plasmid strand, by a type IV secretion system (T<sub>4</sub>SS, the fourth component) into the recipient cell. Once the entire strand has passed into the recipient cell, the relaxase catalyses the ligation\* of the transferred strand back into a circular form (Smillie et al. 2010). This mechanism explains the distinction between conjugative and mobilizable plasmids discussed above: conjugative plasmids carry all of these components, while mobilizable plasmids encode only the *oriT* and corresponding relaxase, but not the expensive T<sub>4</sub>SS (and

may or may not carry the T<sub>4</sub>CP), and therefore need a conjugative plasmid to provide the missing components. Some mobilizable plasmids have only the *oriT*, and rely on another plasmid even for a matching relaxase; these generally have a reduced rate of transfer (Ramsay et al. 2016; Dionisio et al. 2019; Ares-Arroyo et al. 2022).

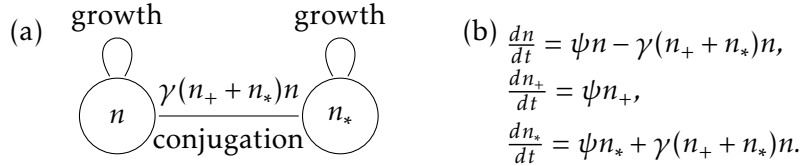
The T<sub>4</sub>SS, T<sub>4</sub>CP, and pili are each large protein complexes, which are costly to express, and the process of conjugation itself is energetically costly (San Millan and MacLean 2019). Moreover, some bacteriophages\* bind to the proteins of the pilus (so-called “male-specific phages”), so bacterial cells expressing pili are at greater risk of infection. Therefore the conjugative phenotype is tightly regulated, to avoid imposing a large fitness cost on the host cell, and is almost always unexpressed (Koraimann and Wagner 2014; Getino and Cruz 2018). Signals that lead to expression of conjugation may include the detection of suitable recipients or specific environmental conditions (Koraimann and Wagner 2014; Bingle and Thomas 2001); also, a temporary period of expression typically occurs immediately after a novel plasmid is first transferred to a recipient, before the repression becomes effective (Lundquist and Levin 1986). The modelling of conjugation in general, and conjugation rates in particular, is discussed in Box 2.4.

---

*Box 2.4: Conjugative plasmid spread.*

Dating back to Stewart and Levin (1977), the majority of modelling studies consider a well-mixed population (such as a liquid culture) and assume that conjugative plasmid transfer follows the principles of mass-action kinetics (Leclerc et al. 2019). Transfer is thus proportional to the densities of plasmid-free and plasmid-carrying cells and a conjugation coefficient  $\gamma$ . The framework is similar to classical epidemiological models with infected and uninfected patients. Based on the model shown below, Levin et al. (1979) showed that mass-action kinetics with a constant  $\gamma$  captures the plasmid dynamics well in populations with constantly dividing cells (but not during lag phase or close to stationary phase). Models based on this approach have been used to address a large range of questions on plasmid dynamics such as the plasmid paradox described in Box 2.5 (Bergstrom et al. 2000), the question of which genes are carried on plasmids (Tazzyman and Bonhoeffer 2015; Lehtinen et al. 2021), the dynamics of cooperative genes and cheating (Mc Ginty et al. 2010), the evolution of antibiotic resistance (Svara and Rankin 2011; Tepekule et al. 2019, see also Box 2.2), and the role of conjugation in the rate of adaptation (Tazzyman and Bonhoeffer 2013; Geoffroy and Uecker 2023). A simplifying assumption in most of these models is that the transfer coefficient  $\gamma$  is constant in time and independent of, for example, resource availability (Leclerc et al. 2019). Dependency of  $\gamma$  on a resource  $C$  can be modeled by a Monod function, i.e.  $\gamma(c) = \gamma_{\max} \frac{C}{Q+C}$  (see e.g. Simonsen

et al. 1990). Spatial models of plasmid transfer are the minority. One of the first examples of a spatial model is the individual-based lattice model by Krone et al. (2007), in which zero, one, or two cells may reside on the sites of a lattice. Plasmid-free cells receive plasmids at stochastic rates that depend on the number of donors and transconjugants\* in a local neighbourhood and the amount of nutrients in a ‘nutrient neighbourhood’. The model is combined with experiments on agar surfaces, capturing the experimental observations well.



Flow diagram (A) and system of ODES (B) for the model of Levin et al. (1979). Three types of cells are distinguished: plasmid-free (with density  $n$ ), an original plasmid-carrying population (with density  $n_+$ , not shown in the flow diagram), and transconjugants (with density  $n_*$ ). Conjugation is modelled through a mass-action kinetics term with conjugation coefficient  $\gamma$ , and the bacterial populations grow exponentially at rate  $\psi$ .

How much do plasmids conjugate? The donor and recipient species, the plasmid, environmental factors, and coinfecting plasmids influence the rate of plasmid transfer (Sheppard et al. 2020). There is, surprisingly, no standard method for estimating conjugation rates (Kosterlitz et al. 2022; Huisman et al. 2022; Kosterlitz and Huisman 2023). Many studies do not estimate the conjugation coefficient  $\gamma$  but report quantities such as the number of transconjugants per donor or per recipient. Methods to estimate the conjugation coefficient  $\gamma$  are mostly derived from models that describe plasmid spread by ordinary differential equations (Levin et al. 1979; Simonsen et al. 1990; Huisman et al. 2022). For example, Levin et al. (1979) solved their ODEs to obtain the expression

$$\gamma = \frac{\psi}{N(b) - N(a)} \ln \frac{R + \rho(b)}{R + \rho(a)},$$

where  $N(b)$  and  $N(a)$  are the total population sizes at time points  $a$  and  $b$ ,  $\rho(t) = \frac{n_+(t)}{n(t)}$ , and  $R = \frac{n_+(t)}{n(t)}$  (which is a constant in the model). All quantities on the right-hand side can be estimated from experimental data to obtain an estimate for  $\gamma$ . Recently, Kosterlitz et al. (2022) proposed a fluctuation test similar to the fluctuation assays for mutation rate estimation, which go back to Luria and Delbrück (1943). Estimates for  $\gamma$  range over several orders of magnitude, ranging from around  $10^{-20} \text{ mL cell}^{-1} \text{ h}^{-1}$  to  $5 \times 10^{-7} \text{ mL cell}^{-1} \text{ h}^{-1}$  (Sheppard et al. 2020). It seems that conjugation rates can be higher *in vivo* than *in vitro* (Gottig et al. 2015; Fischer et al. 2019; Ding et al. 2022).

Like most—if not all—traits, the conjugation rate is itself subject to evolution (Sheppard et al. 2021; Dimitriu 2022). As the conjugation rate can be controlled by the host, the recipient cell, or the plasmid, evolution of both bacteria and plasmids can change how much transfer occurs. Selection pressures on bacteria and plasmids might thereby diverge (recently formalized in a model by Sheppard et al. (2021)). As explained in the main text, conjugation is costly. Unless the plasmid carries genes that are beneficial to its host and outweigh the costs, selection drives the conjugation rate down. The plasmid, on the other hand, needs to strike a balance between horizontal transfer and vertical transmission (Sheppard et al. 2021; Dimitriu 2022). Higher conjugation rates are beneficial but only to an extent that the host is not harmed too much, similar to the transmission-virulence trade-off in pathogen evolution.

---

Transmissible plasmids often also encode genes which assist the plasmid in establishing in newly infected cells, particularly if the cells are genetically distinct from the previous host. The region containing these genes is usually close to the *oriT* and is transferred first during conjugation, so that genes which suppress restriction enzymes\* or the SOS response\* in the newly infected cell can be expressed early on (Garcillán-Barcia et al. 2011; Samuel and Burstein 2023). These genes are responsible for overcoming the various components of the bacterial immune system (Fecskeová et al. 2012; Samuel and Burstein 2023), which we will discuss in Section 2.5.1.

Conjugative and some mobilizable plasmids prevent closely related plasmids from transferring into their hosts by conjugation (see Figure 2.2B). There are two mechanisms for this: surface exclusion, which prevents mating pair formation, and entry exclusion, which is more common and prevents plasmid DNA from entering the host during conjugation (reviewed in Garcillán-Barcia and Cruz 2008; Getino and Cruz 2018).

#### 2.4.4 *Postsegregational killing*

Many plasmids exhibit a spiteful behaviour: they kill plasmid-free daughter cells of their hosts. This phenomenon is variously called postsegregational killing, host killing, or plasmid addiction. It was briefly mentioned above together with stability mechanisms, because this is the classical explanation of postsegregational killing: it ensures the stable inheritance of the plasmid (see e.g. Gerdes et al. 1986; Zielenkiewicz and Cegłowski 2001; Tsilibaris et al. 2007; Summers 1996, 88–91).

The postsegregational killing system involves two components (generally proteins or RNAs), one which kills (or at least suppresses growth in) the host, and a second which suppresses the first: for this reason these systems are also called toxin-antitoxin systems. Both are expressed constitutively\* in the presence of the plasmid, and the antitoxin prevents the death

of the host. But the antitoxin has a much smaller half-life, so in plasmid-free daughters it degrades faster than the toxin, and the toxin kills the cell. (Gerdes et al. 1986; Jensen and Gerdes 1995; Zielenkiewicz and Ceglowski 2001). A slight variant is found in restriction-modification systems, in which there is no need for a difference in half-life: the antitoxin (methylase) needs to be present in a higher concentration to suppress killing than the toxin (restriction enzyme) needs to kill, and as the concentrations fall in the absence of the plasmid the antitoxin becomes ineffective first (Naito et al. 1995; Kusano et al. 1995).

A plasmid-free daughter cell is plasmid-free whether it is alive or dead; postsegregational killing can maintain relative stability of a plasmid by ensuring that the *proportion* of hosts carrying a plasmid does not decrease, but that does not affect the *absolute number* of hosts carrying the plasmid, which continues to decrease. This is confirmed both empirically (Cooper and Heinemann 2000) and by modelling (Mongold 1992). In the exponential phase of bacterial growth therefore, postsegregational killing would have no stabilizing effect. In the stationary phase, however, where competition between hosts and plasmid-free cells for resources becomes more important, or when there is significant spatial structure (Rankin et al. 2012), postsegregational killing may have a more important role in plasmid stability. Postsegregational killing may also serve other functions besides stability, including interplasmid competition (Cooper and Heinemann 2000; Cooper et al. 2010). Toxin-antitoxin systems are also frequently found on bacterial chromosomes, where their function is somewhat obscure (Jensen and Gerdes 1995; Tsilibaris et al. 2007; Magnuson 2007; Song and Wood 2018).

## 2.5 The ecology of plasmids

Thus far, we have focused on the biology of the plasmid itself, largely in isolation; however, plasmids exist in a larger ecological context which includes their hosts, the larger population of hosts (or potential hosts) and the community it is part of, and other plasmids and nonplasmid mobile genetic elements in their host or the host population. (Pilosof (2023) provides a brief review of relevant concerns in modelling such complex communities.) In this section, we explicitly consider the influence of interactions with all of these entities on the biology of the plasmid and host.

### 2.5.1 Interactions with the individual host

The plasmid only exists inside the host, and depends on the host to carry out all of its physiological processes, in particular replication. As we have already seen in Box 2.3, this aspect of the plasmid lifestyle means that multilevel selection is of great importance to plasmids. Plasmid fitness can be divided into two components corresponding to the two ways a plasmid can

propagate to new hosts: vertical, from vertical transmission to daughter cells of their current host, and horizontal, from infection of new hosts by conjugation. Vertical fitness of plasmids is directly dependent on the host fitness, while horizontal fitness is more independent (although plasmid spread by horizontal transmission still requires persistence of the host). In general there may be strong trade-offs between vertical and horizontal transmission (Paulsson 2002; Bethke et al. 2022), for example due to the fitness costs of conjugation to the host. Since both contribute to the total plasmid fitness, there will often be different possible successful life-history strategies having a different emphasis on horizontal versus vertical transmission (Zwanzig 2021) (the extreme case being nontransmissible plasmids that rely entirely on vertical transmission).

The importance of the plasmid–host relationship is not unilateral. Although hosts are not dependent on plasmids (with some exceptions where essential genes are found on plasmids), plasmids form a part of the host genome, and are therefore intrinsically linked to the host phenotype and fitness. We have already discussed the wide range of payload genes found on many plasmids which provide important host phenotypes, including antibiotic resistance, virulence, metabolism of novel substrates, and rhizobial symbiosis. Payload genes on plasmids may be particularly important for phenotypes that are involved in local adaptation (Eberhard 1989; Wardell et al. 2021). They often have context-dependent fitness effects, providing fitness benefits in some environments and conditions and causing fitness costs in others (Platt et al. 2014; Hall et al. 2015). In addition to carrying important genes, plasmids also frequently influence the regulation and expression of chromosomal genes, which similarly affects the host phenotype (reviewed in Billane et al. 2021). In some cases this can take the form of interference between plasmid and host gene regulation, producing a reduction in host fitness (Hall et al. 2021c; Ricci and Hernández 2000), particularly for plasmids that are new to the host and not well adapted. However, changes in the expression of chromosomal genes can also be adaptive, affecting traits like motility, biofilm formation, antibiotic resistance, virulence, or aspects of the bacterial metabolism (Vial and Hommais 2019; Billane et al. 2021).

Nevertheless, in the absence of genes with important host fitness effects, plasmids are often costly to the bacteria which carry them, and this cost increases with copy number (San Millan et al. 2016; Valle et al. 2021); the fitness costs of plasmids lead to a so-called “plasmid paradox” discussed in Box 2.5. In addition to interference between gene regulation or physiological processes of plasmid and chromosome (San Millan and MacLean 2019; Gama et al. 2018; Ricci and Hernández 2000; Harr and Schlötterer 2006; Yano et al. 2016; Hall et al. 2021c; Zhang et al. 2022a; Rajer and Sandegren 2022), causes of this fitness cost include bacterial stress responses to foreign DNA (San Millan and MacLean 2019; Gama et al. 2018), the cost of

plasmid replication and gene expression (San Millan and MacLean 2019; Gama et al. 2018; Lamberte et al. 2017; Rodríguez-Beltrán et al. 2021b), and the possibility of disruption of host genes by the plasmid recombining with the chromosome (San Millan and MacLean 2019). These costs can be reduced or eliminated by compensatory evolution of the plasmid (Heuer et al. 2007; Porse et al. 2016; Yano et al. 2016; Dorado-Morales et al. 2021), the host chromosome (Bouma and Lenski 1988; San Millan et al. 2014; Loftie-Eaton et al. 2015; Harrison et al. 2015; Hall et al. 2020; Zhang et al. 2022b), or both (Dahlberg and Chao 2003; Dionisio et al. 2005; Sota et al. 2010; Zwanzig et al. 2019; Zhang et al. 2022a), and evolution experiments have shown this occurs regularly, especially if selection for the plasmid ensures it is maintained before compensation is complete (Bouma and Lenski 1988; Dahlberg and Chao 2003; Dionisio et al. 2005; Heuer et al. 2007; Sota et al. 2010; San Millan et al. 2014; Harrison et al. 2015; Porse et al. 2016; Hall et al. 2020; Dorado-Morales et al. 2021; Carrilero et al. 2021; Zhang et al. 2022a; Given et al. 2022); there is also some observational evidence of compensatory evolution in plasmids in the wild (Malaka De Silva et al. 2022; Takeda et al. 2011). The coevolved plasmid and chromosome produced by compensatory evolution are specifically adapted to each other, and to existence as components of a single genome (Vial and Hommais 2019) (although some compensatory changes on the plasmid are general enough to also reduce the fitness cost in other hosts Dionisio et al. 2005), even to the extent that coevolved cells cured of the plasmid have lower fitness than wild-type cells (Bouma and Lenski 1988; Harrison and Brockhurst 2012; Brockhurst and Harrison 2022). Gama et al. (2018) argue that this coevolution—or rather its loss—makes estimated fitness costs of plasmids in laboratory conditions overestimates of those that are common in nature: the reported *in vitro* fitness cost of a plasmid seems to increase with time since plasmid isolation, suggesting that fitness costs seen in experiments may be at least partially due to maladaptive plasmid evolution under laboratory conditions, leading to the loss of evolved cost mitigation (see also Valle et al. 2021). As already pointed out above, the fitness cost of the plasmid may also strongly depend on the environmental context (Hall et al. 2015), further reducing the generalizability of experimental measurements.

---

*Box 2.5: The plasmid paradox.*

One of the great questions of plasmid biology is the “plasmid paradox”: why are there plasmids at all? A simple cryptic plasmid, which does nothing but replicate itself inside the host, imposes a fitness cost. Therefore, over the long term, bacteria carrying plasmids should be outcompeted by plasmid-free bacteria and go extinct, taking their plasmids with them (Brockhurst and Harrison 2022; Harrison and Brock-

hurst 2012). In some cases, plasmids may transfer horizontally fast enough to be maintained despite fitness costs, analogously to how parasites survive (Brockhurst and Harrison 2022), although reliable estimates for transfer rates in nature are needed to determine whether conjugation rates are in fact high enough. It is difficult to see, however, how horizontal transfer could maintain mobilizable plasmids, and of course it is impossible that horizontal transmission explains the maintenance of nontransmissible plasmids.

Plasmids could also be maintained by carrying advantageous genes that would make the net fitness effect of the plasmid positive: but these genes could move by recombination from the plasmid to another replicon such as the chromosome, allowing the bacterium to lose the plasmid and have its cake and eat it too. Bergstrom et al. (2000) analyzed an ODE model of plasmid persistence, starting from the assumption that the rate of horizontal transfer is not high enough to maintain the plasmids as parasites: this assumption is formally given by the relation

$$\psi\alpha + \tau > \gamma N^+,$$

where  $\psi$  is the per capita growth rate of plasmid-free bacteria,  $\alpha$  is the proportional fitness cost of carrying the plasmid,  $\tau$  is the per capita rate of segregative loss,  $\gamma$  is the horizontal transfer rate, and  $N^+$  is the maximum total bacterial population size obtainable. They show this implies the eventual extinction of plasmid-carrying cells, even if there is a beneficial gene on the plasmid (or fluctuating selection), provided that the beneficial gene can also be located on the chromosome.

Bergstrom et al. (2000) do show, however, that plasmids can persist in two circumstances. The first is if the plasmids have a net fitness benefit *and* the population is subject to regular strong selective sweeps of adaptive variants immigrating into the population: in this case, bacteria that have incorporated the beneficial gene onto the chromosome are eliminated by the selective sweep, while the plasmid can transfer to the new strain. The second is if there are multiple species or ecotypes of bacteria present and the beneficial gene is occasionally lost in each ecotype (due to drift, demographic stochasticity, selective sweeps, etc.): the plasmid is maintained by the ability to transfer between ecotypes.

There has been extensive modelling of other possible resolutions of the plasmid paradox (see e.g. Stewart and Levin 1977; Levin and Stewart 1980; Simonsen 1991; Bergstrom et al. 2000; Lili et al. 2007; MacLean and San Millan 2015; Carroll and Wong 2018). Possible answers include selective pressures encouraging maintenance of some genes on disposable or transmissible replicons (Slater et al. 2008; Wardell et al. 2021; Rankin et al. 2011; Eberhard 1989; Eberhard 1990), nonconstant or variable rates of horizontal transfer (Lundquist and Levin 1986; Dionisio et al. 2002), the maintenance of nonconjugative plasmids by constant replenishment by conjugative plasmids that lose transfer genes (Coluzzi et al. 2022), fitness effects of plasmid-plasmid

interactions (Werisch et al. 2017), and, particularly in the case of large nontransmissible plasmids close to being accessory chromosomes, to evolutionary constraints of genome organization (diCenzo and Finan 2017; Hülter et al. 2017)—after all no one calls the presence of more than one chromosome in an organism the “chromosome paradox.” Future modelling work will hopefully help to tie the theories to the circumstances of specific plasmids and bacteria and determine which mechanisms are responsible for maintenance in particular systems.

---

Despite the wonders of plasmid biology, some bacteria are just not interested in hosting a plasmid. They instead prevent the establishment of plasmids in general or of particular plasmids using defence systems that destroy plasmids (and other mobile genetic elements) upon arrival in the cell (e.g. Price et al. 2016; Pursey et al. 2021; Jaskólska et al. 2022; Miyakoshi et al. 2012). The best known bacterial immune systems are CRISPR-Cas\*, which targets specific mobile genetic elements based on recorded DNA sequences, and restriction enzymes\*, which target foreign DNA in general (Price et al. 2016; Pursey et al. 2021).

### 2.5.2 Interactions with the host community

The individual host of a plasmid will be part of a larger population of host or potential host bacteria, which in turn makes up part of a larger ecological community. Since plasmids are associated with their hosts, they are inherently influenced by all community interactions. Through their effects on host phenotypes they in turn alter these interactions, and may thus be subject to eco-evolutionary feedbacks. Moreover, social traits are frequently encoded by payload genes on plasmids: for example, plasmids have a higher density of genes for secreted proteins than chromosomes (Nogueira et al. 2009), which includes proteins involved in both cooperative and antagonistic interactions between bacteria (Rankin et al. 2011). Cooperative interactions between plasmid hosts and other bacteria include antibiotic resistance, if the mechanism of resistance is the degradation of the antibiotic (Kelsic et al. 2015), while antagonistic interactions include production of bacteriocins, excreted toxins which kill other bacteria and are often found on plasmids (indeed, one of the earliest discovered families of plasmids codes for colicins, bacteriocins produced by and targeting *Escherichia coli* Falkow 1975, 40–42). The plasmid codes for both the bacteriocin, which is excreted (in some cases killing the host in the process), and an immunity gene, which protects against the bacteriocin (Riley and Wertz 2002); thus bacteriocins behave like an extracellular version of a postsegregational killing system, killing those cells which do not carry the plasmid. Plasmid-borne genes can also play a role in interactions with other cells without actively affecting them: for example, some plasmids of *Pseudomo-*

*nas* spp. sensitize their hosts to growth-inhibiting compounds excreted by other bacteria (Smith et al. 2021).

Bacterial interactions strongly depend on whether the bacteria are free-living or in a biofilm\*. Planktonic\* bacteria in liquid culture can potentially encounter many different cells, but are usually not in close proximity to each other. In biofilms, bacteria are nonmotile and in close contact with their direct neighbours. Biofilms are moreover highly spatially heterogeneous, with different conditions on the surface and the interior of the biofilm, affecting the physiology and growth of bacteria. Conjugative plasmids can increase the propensity of their hosts to form biofilms (this may in part be due to the pilus serving to attach planktonic cells to the growing biofilm) (Ghigo 2001; Madsen et al. 2012; Cook and Dunny 2014; Abe et al. 2020; Gama et al. 2020). How the aggregation of bacteria in biofilms in turn affects conjugative plasmid spread is complicated. Conjugation rates are often higher in biofilms than in planktonic bacteria due to stable cell-to-cell contact and stable mating pair formation (Abe et al. 2020), yet the spread of plasmids through biofilms is nonetheless often limited, as discussed by Stalder and Top (2016). While plasmids may not spread through the entire biofilm, the biofilm lifestyle may aid the maintenance of conjugative plasmids and prevent the loss of transfer genes (Metzger et al. 2022).

Plasmids do not only influence ecological interactions, but also cause genetic interactions between individuals of the host population by horizontally transferring chromosomal genes: conjugative plasmids that integrate into the host chromosome or that pick up pieces of the chromosome by homologous recombination can then transfer chromosomal DNA to a recipient on conjugation (Lee et al. 2011), such as in the classic Hfr strains of *Escherichia coli* (Falkow 1975, 10–14). There is also the possibility of retrotransfer of chromosomal genes (Szpirer et al. 1999).

Since horizontal transfer of plasmids between distinct bacterial species is common, multispecies communities enable new plasmid dynamics, including the maintenance of plasmids in less-than-ideal hosts by transfer from a species to which they are better adapted (Hall et al. 2016), and improved compensatory amelioration of fitness costs through transfer between multiple species (Heuer et al. 2007; Li et al. 2020). Microbial communities also include species that are not bacteria, and plasmids can also contribute to or be affected by interactions of their hosts with these organisms. For example, Cairns et al. (2018) showed that exposure to a virus that targets conjugation machinery to infect a cell selected for reduced conjugative ability in a plasmid, but predation by a ciliate increased conjugation rate.

### 2.5.3 *Interactions with other plasmids and mobile genetic elements*

Multiple plasmids often coexist in the same population, and even a single bacterium often carries multiple plasmids of distinct types (see e.g. Burian

et al. 1997; Casjens et al. 2000; San Millan et al. 2009; Smith et al. 1998; Ventura et al. 2007). There is thus much opportunity for interactions between plasmids. The simplest such interaction is preventing coinfection of the same cell by excluding other plasmids. We have already seen mechanisms which can prevent coinfection: plasmid incompatibility and entry and surface exclusion (Hoeven 1985). Postsegregational killing can also lead to loss of a coinfecting plasmid over time (Cooper and Heinemann 2000). Some plasmids also actively destroy other potentially coinfecting plasmids with plasmid-borne CRISPR-Cas (Pinilla-Redondo et al. 2019; Kamruzzaman and Iredell 2020; Pinilla-Redondo et al. 2021) or restriction-modification systems (Naito et al. 1995; Kusano et al. 1995; Iglar et al. 2021a) (see Figure 2.2C), which work analogously to their chromosomal counterparts.

If multiple plasmids successfully establish in a common host, they may impose a high fitness cost. However, compensatory evolution that eliminates the fitness costs of one plasmid can sometimes allow for the acquisition of other related plasmids at no additional cost (Santos-Lopez et al. 2017). Similarly, some plasmids exhibit positive fitness epistasis\*, meaning the total fitness cost of carrying multiple plasmids is less than the sum of the fitness costs of each individually (San Millan et al. 2013; Platt et al. 2014; Lopez et al. 2021). Environmental conditions can affect the maintenance of coinfection: if coinfecting plasmids carry similar beneficial payload genes, the less advantageous might be lost in the presence of selection, even if coexistence is possible in the absence of selection (Carrilero et al. 2021). Coinfection will have important effects on compensatory evolution to reduce plasmid fitness costs, as in coinfecting hosts there will have to be evolution not only to reduce plasmid-chromosome conflicts, but also plasmid-plasmid conflicts.

The effects of plasmid interactions on horizontal transfer can be varied (Horne et al. 2023). We have already seen that mobilizable plasmids rely on a coinfecting conjugative plasmid to be transferred; cooperative horizontal transfer can also occur through a process called conduction, in which a plasmid forms a cointegrate with a conjugative plasmid and is transferred by conjugation (Clark and Adelberg 1962; Xiao et al. 1994; Brom et al. 2004); this is shown in Figure 2.3A. Between conjugative plasmids, coinfection can produce both an increase in conjugation rate (called facilitation Dionisio et al. 2019), and a decrease (Gama et al. 2017a). Similar interactions between plasmids can alter their ability to promote biofilm formation (Gama et al. 2020).

So far we have implicitly talked about plasmid-plasmid interactions from the perspective of pairs of plasmids. When there are more than two types of plasmids present in a host cell, there will be higher-order interactions between plasmids that will need to be taken into account (Gama et al. 2017b; Given et al. 2022).

Plasmids also coexist in their hosts with other mobile genetic elements. Some of these exist in the host in the form of a plasmid, and therefore may exhibit interactions with plasmids similar to plasmid-plasmid interactions: for example, some bacteriophages (Ravin 2011; Falkow 1975, 50–51), and integrative conjugative elements\* (ICEs) (Burrus et al. 2002). Many phages target plasmid-encoded structures to infect their hosts (particularly structures related to conjugation) (Ojala et al. 2013; Ojala et al. 2016); other plasmid-phage interactions include the horizontal transmission of plasmids by transduction\* in phages (Dionisio et al. 2019), and the ability of some plasmids to inhibit the infection of their hosts by phages (Igler et al. 2021b).

### 2.6 Final remarks

In this review, we provided an overview of the biology of plasmids, aiming to make the topic accessible to modellers unfamiliar with plasmids. Mathematical modelling provides a powerful tool to understand the biology of plasmids and to make predictions about their contributions to bacterial evolution (which is especially important with respect to the evolution of clinically important traits like antibiotic resistance). The great diversity of plasmids means modelling is needed both to seek out general unifying patterns and to determine properties specific to particular plasmid systems. We conclude by putting forward a few areas of plasmid biology that have so far scarcely been addressed by modellers.

The majority of plasmid models consider plasmids of only one type and distinguish only between plasmid-carrying and plasmid-free cells. Models that account for the plasmid copy number, despite appearing early on (Novick and Hoppensteadt 1978; Ishii et al. 1978; Cullum and Broda 1979; Seneta and Tavaré 1983; Nordström and Aagaard-Hansen 1984; Paulsson 2002), have remained overall rare. Recently the topic started to receive growing attention, mostly (but not exclusively) in the context of allele dynamics on multicopy plasmids (see Box 2.3, Hernandez-Beltran et al. 2022). Several of the early models of multicopy plasmids consider segregation of two incompatible plasmids and thus include coinfection of cells by multiple plasmid types. Generally, however, there are still few models that consider the coexistence of plasmids of different types in the same cell (for two examples, see Condit and Levin 1990; Verweij and Griswold 2023). For a guide to designing such models, see the recent article by Igler et al. (2021b). Studies allowing for multiple plasmid types in the population (but not within the same cell) are more common, but also rare compared to those with a single plasmid type; they mostly focus on competition between different plasmids for host cells (Svara and Rankin 2011; Tazyman and Bonhoeffer 2014; Werisch et al. 2017; Lehtinen et al. 2021; Geoffroy and Uecker 2023). Coinfection and interactions between different plas-

mids such as cotransfer of nonconjugative with conjugative plasmids are common and likely have a great influence on the evolution and ecology of bacteria. To understand the evolutionary dynamics of plasmids and hosts, it is thus crucial to develop more theory that takes multiple plasmid types and coinfection into account.

A related question is why we see such great diversity of plasmids. The huge number of bacterial species and environments they inhabit implies a huge number of niches for plasmids. Yet it remains puzzling why some plasmids have a high and others a low copy number, why some plasmids are very large and others tiny, why some plasmids are conjugative and others not even transmissible, etc. Moreover, plasmids of different types coexisting in a bacterial population share the same niche, at least to a first approximation. Modeling has shown that two types of plasmids can coexist if they follow different strategies regarding their transfer rate and costs (Hoeven 1984; Hoeven 1986; Werisch et al. 2017). Similarly, the evolution of plasmid host range (Loftie-Eaton et al. 2015; Sota et al. 2010; Yano et al. 2013) and rapid speciation of plasmids (Santos-Lopez et al. 2016) may contribute to development and spread of new plasmid types in the wild. Future models could combine ecological considerations at the level of plasmids (such as differential strategies or interactions between coinfecting plasmids) with evolutionary dynamics (such as the evolution of conjugation rates or plasmid speciation) to provide further insights into the processes that shape plasmid diversity.

As pointed out in Box 2.2, plasmid-mediated antibiotic resistance is less studied in models than chromosomal resistance. Many models of antibiotic resistance evolution consider the effects of treatments such as combination therapies or drug dosing regimes on the evolution of resistance within hosts and strategies to manage existing resistance at the population level (for a few examples of models including plasmids, see (D'Agata et al. 2008; Bergstrom et al. 2004; Webb et al. 2005; Willms et al. 2006)). Given the clinical relevance of plasmid-mediated resistance, it seems of great importance to develop further models to test how treatment regimes and population-wide protocols affect resistance on plasmids. Many infections are polymicrobial, in which case plasmids might spread from one species to others during treatment (Orazi and O'Toole 2019), influencing the treatment outcome. The effects of treatment protocols on bacterial communities harboring plasmids are an important future direction of modelling. Resistance plasmids in commensal bacteria are also important since they may be transferred to pathogenic bacteria or the commensals themselves may become opportunistic pathogens. During treatment, commensal bacteria are often inevitably exposed to antibiotic pressure, selecting for resistance (so-called bystander selection). Moreover, plasmids carried by commensal bacteria may evolve both during treatment and in the absence of antibiotics and become less costly, for example. Transmission of commensal bacteria

between individuals can spread resistance plasmids in a host population. Plasmids in commensals have already received some attention from modellers but there is much more to explore and understand (see Box 2.2 and also Volkova et al. 2012; Volkova et al. 2013). Topics of concern where we see need for further modelling include the evolution of plasmids within commensals, their maintenance in the absence of treatment, their potential transfer to pathogens, and their maintenance and spread at the population level. Of special interest are how these processes interact and how they are influenced by treatment protocols and frequencies.

Lastly more models of bacterial biofilms are needed, both in the context of general plasmid biology as well as in a clinical context, taking into account a three-dimensional spatial structure with varying nutrient and antibiotic concentrations and potentially the attachment and detachment of planktonic bacteria (for an example of a biofilm model, see Fox et al. 2008).

Plasmids are very widespread, diverse, and of great relevance for bacterial ecology and evolution (and as a consequence for human health). While a large body of theory on the evolution of plasmids and their hosts exists, there remain many puzzles and unsolved questions. We hope that our review provides interested mathematical biologists with the necessary background to start to develop novel theory on plasmid dynamics.

### Glossary

**accessory chromosome** See secondary chromosome.

**bacteriophage** Virus which infects bacteria.

**biofilm** An aggregation of bacteria which adhere to each other and a surface, and excrete an extracellular polymer which surrounds them.

**bp (base pairs)** The unit of length of a DNA molecule; a base pair consists of a pair of nucleotides from each strand of the molecule.

**chromid** See secondary chromosome.

***cis*-acting** Of a gene or regulatory sequence: only affecting the same replicon on which it is located; the opposite of *trans*-acting.

**constitutively expressed** Of a gene: constantly transcribed and expressed (as opposed to being regulated so it is only transcribed in specific circumstances where it is needed).

**CRISPR-Cas** A bacterial adaptive immune system: it consists of specific sequences from mobile genetic elements copied into the bacterial chromosome (CRISPR) and a Cas enzyme which recognizes and cuts DNA molecules with those sequences.

**epistasis** The interaction of different genes contributing to the same phenotype: when the collective effect of two (or more) genes on the phenotype is not just the sum of the effects to the two genes alone, there is epistasis between the two loci.

**gene cassette** A small mobile genetic element consisting of a gene and a site where recombination into a replicon can occur; see integron.

**gram positive/negative** A major classification of bacteria, based on whether they are stained by the dye in a Gram stain procedure (named after the Danish bacteriologist Hans Christian Gram); gram negative bacteria have a second outer membrane surrounding their cell wall.

**horizontal transmission/horizontal (gene) transfer** The transfer of genetic material between two cells, as opposed to vertical inheritance of genes from mother to daughter cell.

**integrative conjugative element** A type of mobile genetic element: they normally exist integrated into the host chromosome, but can excise from the chromosome, take on a temporary plasmid-like form, and transfer to another cell by conjugation.

**integron** The place where gene cassettes are integrated into a replicon: consists of a gene coding for a site-specific recombinase and a recombination site where that recombinase can integrate gene cassettes.

**ligation** The process of joining the backbone of a strand of DNA together, sealing a break which previously existed between nucleotides; the opposite of nicking.

**mobile genetic element** Any genetic material that can be transferred between locations in the genome of an organism or between organisms.

**nicking** The process of disconnecting the backbone of a strand of DNA by breaking the bonds between adjacent nucleotides; the opposite of ligation.

**open reading frame** The sequence of DNA coding for a protein.

**phage** See bacteriophage.

**planktonic** Floating freely in the water column, as opposed to being associated with a surface or able to swim against the current.

**recombination** The process of “crossing over” between two DNA molecules: the molecules are broken at the target site, and the new ends are exchanged between molecules and resealed; when it occurs between two circular DNA molecules, it fuses them into one circle, and

when it occurs between two parts of the same circular DNA molecule it separates it into two circular molecules; has site-specific and homologous forms.

**recombination, homologous** Recombination between two double-stranded DNA molecules occurring at any location where the two molecules have a homology (that is, where their sequence is the same); most bacteria have recombinases which catalyze this process.

**recombination, site-specific** Recombination between two DNA molecules catalyzed by an enzyme that recognizes a specific sequence on the molecules to be recombined.

**relaxase** An enzyme that catalyzes nicking.

**replicon** An independently replicating DNA molecule; strictly speaking, everything that replicates from a particular origin of replication.

**repressor** A protein that prevents the transcription (and therefore expression) of a gene.

**restriction enzyme** An enzyme which catalyzes the breaking of a DNA molecule at a specific sequence.

**secondary chromosome** A large replicon in some bacteria that contains essential genes, but is smaller than the main chromosome; something between an ordinary chromosome and an ordinary plasmid.

**segregation** The division of copies of a replicon between daughter cells on cell division.

**SOS response** A bacterial stress response.

**trans-acting** Of a gene or regulatory sequence: affecting both the replicon it is located on and any homologous genes or regions on other replicons; the opposite of *cis-acting*.

**transconjugant** A cell that has received a plasmid through conjugation from another cell

**transduction** A mechanism of horizontal gene transfer in which nonviral genetic material is accidentally packaged into a bacteriophage, and transferred by the phage to other bacteria.

**transformation** A mechanism of horizontal gene transfer in which free DNA molecules in the environment are taken up by a bacterium.

**transposon/transposable element** A mobile genetic element that can move itself from one location to another in the genome, either by copying or by excision and reinsertion.

## IS THE DISTRIBUTION OF PLASMID LENGTHS BIMODAL?

**Abstract.** *The length of a plasmid is a key property which is linked to many aspects of plasmid biology. When distributions of plasmid lengths are shown in the literature, they are usually plotted with length on a logarithmic scale. However, a quantity and its logarithm have distinct distributions which may differ considerably in shape. Mistaking the distribution of log-lengths for the distribution of lengths can therefore lead to distorted conclusions about the distribution; in particular, the distribution of log-lengths may be bimodal when the distribution of lengths is only unimodal. This particular confusion has arisen in the literature where the length distribution is often claimed to be bimodal based on examination of what is in fact the log-length distribution. While the length distribution is indeed bimodal within many bacterial families, it is not across the ensemble of all plasmids. We suggest that authors should be careful to show the plasmid length distribution, or to distinguish the two distributions, to avoid misleading inferences.*

### 3.1 Introduction

Bacterial plasmids have a wide variety of sizes, from tiny cryptic plasmids less than a kilobase long, to multiple-megabase megaplasmids that rival the size of chromosomes (Ciok et al. 2016; Hall et al. 2021a). The length is closely linked to other aspects of a plasmid's biology, since smaller plasmids contain fewer genes and may therefore be limited in their functions (plasmid-specific or related to the host phenotype), while very large plasmids may contain a wide variety of accessory genes.

It has generally been assumed in the literature that plasmids can be broadly divided into two groups on the basis of length, and that this division corresponds to a biologically meaningful distinction between two plasmid lifestyles. The shorter group, with a modal length less than 10 kbp (Hall et al. 2021a), is said to comprise plasmids with high copy numbers which are mostly mobilizable or nontransmissible, while the longer group, with a modal length around 100 kbp, comprises low copy number plasmids which are more likely to be conjugative. Part of the evidence for the existence of two length-based groups is the distribution of plasmid lengths estimated using standard techniques (kernel density estimation or histograms): the resulting density estimates clearly have a bimodal shape (this is done for all plasmids together by Smillie et al. 2010; Garcillán-

---

This chapter has been published as: Ian Dewan and Hildegard Uecker (2024). "Is the distribution of plasmid lengths bimodal?" In: *Plasmid* 129–130, doi: 10.1016/j.plasmid.2024.102721

Barcia et al. 2023 and for plasmids divided by host family by Hall et al. 2021a; Rodríguez-Beltrán et al. 2021a).

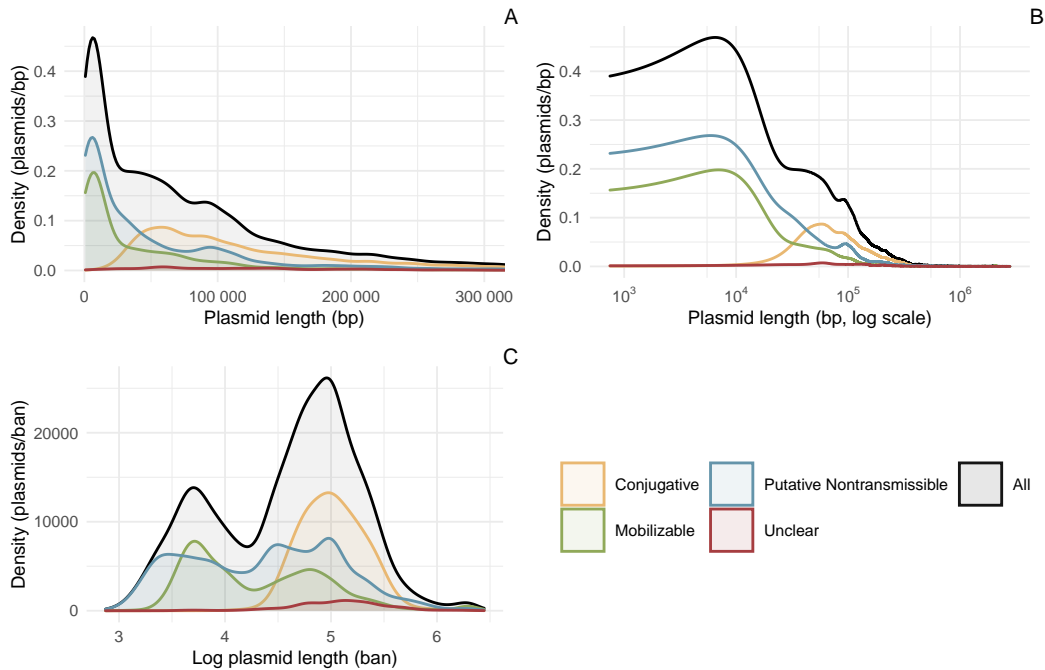
These graphs show densities with a logarithmic scale on the  $x$ -axis, a natural choice given that the lengths of plasmids extend over four orders of magnitude: however, when we plotted plasmid length data (pooled across bacterial families) on a linear scale (in Chapter 2), the bimodal distribution disappeared (see also diCenzo and Finan 2017 who also use a linear scale and find a unimodal distribution). In this letter, we explain the reason for this discrepancy, and what implications, if any, it has for our understanding of plasmids.

## 3.2 Analysis and results

### 3.2.1 *The distribution of log-lengths is not the distribution of lengths*

Why are there two types of figure that make very different statements about the plasmid length distribution in the literature? They are in fact showing two different distributions: the actual distribution of lengths (shown in Figure 3.1A) and what, despite often being called the distribution of lengths, is in fact the distribution of *log-lengths* (shown in Figure 3.1C). In other words, the graphs with a log-scaled  $x$ -axis do not simply plot the same numerical values from the distribution of plasmid length but on a log scale (which would instead produce Figure 3.1B), but a distinct distribution. The distributions of plasmid lengths and log-lengths may have quite different shapes; in general, the shape of a distribution changes, possibly drastically, when a transformation of the random variable is applied. Intuitively, the reason for this change is that bins of equal sizes on the log scale correspond to bins of different sizes on the linear scale: a bin at small plasmid log-lengths contains a much narrower range of plasmid lengths than a bin at large plasmid log-lengths, and the latter might contain more sequences than the former for this reason alone. As we show in the next paragraph, to obtain the density of log-lengths, we need to multiply the distribution of lengths by an exponentially increasing function. This can convert a heavy tail of the distribution on the linear scale into an additional mode on the log scale. This is illustrated in Figure 3.2 by a hypothetical example of a distribution of a random variable and the corresponding distribution of its logarithm. The distribution of the variable on the linear scale is unimodal, but the distribution of its logarithm is bimodal, precisely what happens with plasmid lengths.

We can show the change in the shape of the distribution formally by examining the effect of a log transformation on the density function of a distribution. Suppose we have a (positive) random variable  $X$  with a probability density function  $f$ . The defining characteristic of the probability density function is that it can be integrated over an interval to find the



*Figure 3.1:* The distributions of plasmid lengths and log-lengths for all plasmids together and for each mobility class individually. The data is taken from PLSDB (Schmartz et al. 2021, retrieved 1 March 2023), and the distribution estimate made by kernel density estimation in R (R Core Team 2020). Panel A shows the estimated distribution of plasmid lengths (with length on a linear scale). Only plasmids of length less than 300 000 bp are shown for reasons of scale: 1844 plasmids in the database (5.34 % of the total) are longer than the limit. Based on Figure 2.1. Panel B shows the plasmid length distribution, estimated as in panel A, but with a log-scaled  $x$ -axis; i.e. kernel density estimation is performed on a linear scale, and the logarithm is only applied for the plot itself. Panel C shows the estimated distribution of plasmid log-lengths (‘ban’ are the units of the base-10 logarithm); i.e. the logarithm is applied before performing the density kernel estimation. To classify the plasmids by mobility, the annotations in PLSDB from MOB-typer (Robertson and Nash 2018), indicating putative relaxases or conjugative genes, have been used. Those plasmids with both are conjugative, those with only a relaxase are mobilizable, and those with neither are putatively nontransmissible (although Ares-Arroyo et al. 2022 have shown that many plasmids lacking a relaxase still have an *oriT*, and are therefore in fact mobilizable). The ‘unclear’ class includes those plasmids which had conjugative genes but no relaxase. It is possible that these are misidentified conjugative plasmids, which have an unknown relaxase, or misidentified mobilizable or nontransmissible plasmids, which do not actually have conjugative genes.

### 3. IS THE DISTRIBUTION OF PLASMID LENGTHS BIMODAL?

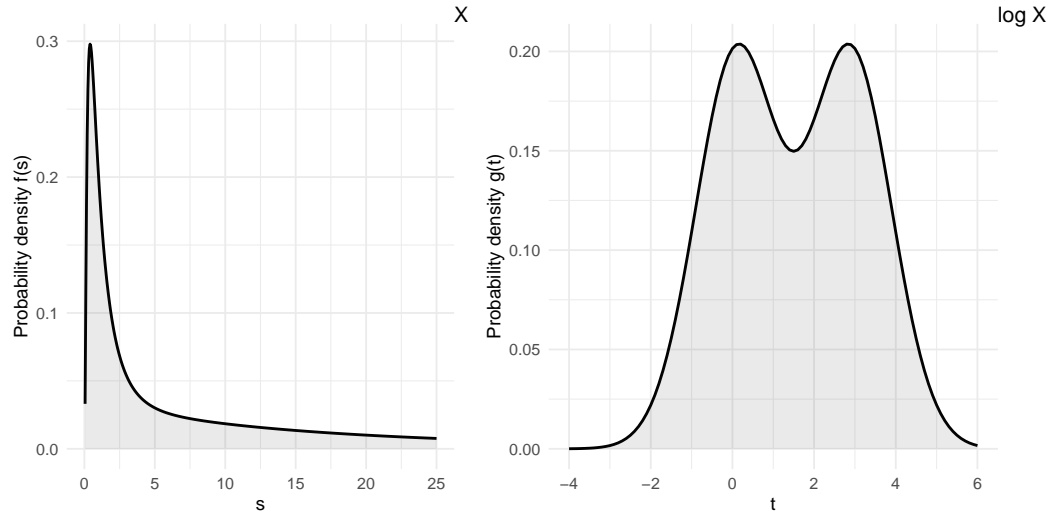


Figure 3.2: The probability density function of a distribution (left) and its (natural) logarithm (right). The distribution of  $\log X$  is an equal mixture of two normal distributions, with means 0.1 and 2.9 and standard deviation 1; the distribution of  $X$  is therefore a mixture of two log-normal distributions.

probability that  $X$  lies in that interval; that is for any numbers  $a \leq b$ ,

$$P(a \leq X \leq b) = \int_a^b f(s) ds.$$

If we know  $f$ , we can find the probability density function of  $\log X$ :

$$\begin{aligned} P(a \leq \log X \leq b) &= P(e^a \leq X \leq e^b) \\ &= \int_{e^a}^{e^b} f(s) ds \\ &\stackrel{s=e^t}{=} \int_a^b \underbrace{f(e^t)e^t}_{=:g(t)} dt, \end{aligned}$$

so the probability density function of  $\log X$  is  $g(t) = f(e^t)e^t$ . (Here we have used the natural logarithm, but the same argument applies to the “common” or base-10 logarithm usually used in log scales.) When we apply the change of variable  $s = e^t$  in the above integral to pass from integrating on a linear scale to integrating on a log scale, we multiply  $f$  by a nonconstant—indeed exponentially increasing—function  $e^t$ , changing the shape of the function. If the probability density function  $f$  decreases too slowly (i.e. if it has a heavy tail), this creates a second mode.

The difference in the number of modes between Figure 3.1A and C is therefore simply a mathematical consequence of the log scale. Changes

in the shape of the distribution of course also occur in subsets of plasmids as well, such as the distributions within mobility classes shown in Figure 3.1. Note, for example, that looking at the log-length distributions in Figure 3.1C would suggest that mobilizable plasmids have a bimodal distribution and nontransmissible plasmids are roughly equally abundant over a wide range of lengths; but the length distributions in Figure 3.1A show that both groups have a unimodal distribution with a long right tail.

To verify that these changes in distribution shape are indeed caused by the log transformation and are not instead artifacts of the mechanism of density estimation, we have confirmed that the general shape of the length and log-length distributions is consistent under different estimation methods: different bandwidth selections for kernel density estimation (results shown in Figures A.1 and A.4), the empirical cumulative distribution function (shown in Figures A.2 and A.5), and equal-area histograms (shown in Figure A.3). These do indeed all give similar results.

### 3.2.2 *Within bacterial families, the plasmid length distribution is often bimodal, but the modes and the valley are incorrectly predicted by the log-length distribution*

Above we considered the plasmid length distribution for all plasmids together, irrespective of their bacterial hosts. However, the size of plasmids differs between groups of bacteria, for example because bacteria with larger genome sizes in general have larger plasmids (Slater et al. 2008; Smillie et al. 2010). If the length distribution is really bimodal within groups, but the modes (and the antimode, or local minimum density, between them) have different locations in different groups because of a general difference in size, then in the pooled distribution for all plasmids the modes and antimodes of different groups might obscure each other. Looking at the length distribution by family, as shown in Figure 3.3, shows that some bacterial families indeed show a pronounced bimodal distribution on the linear scale (plasmids from Enterobacteriaceae, Enterococcaceae, or Staphylococcaceae), while for others there is no clear second mode (plasmids from Bacillaceae or Moraxcellaceae), or it is very weak (plasmids from Lactobacillaceae). A caveat to the within family density estimates is that they are based on much fewer sequences than the estimate for all plasmids. We would like to point out that we have identified modes by eye, only counting what we visually identify as major modes and not taking any smaller secondary modes into account (which may or may not be due to sampling effects or artifacts of the kernel density estimation). In particular, we do not perform any statistical analysis to determine which of the small modes are “real” (using, e.g. Silverman 1981), which is not necessary to make the point of the article.

Nevertheless, even if both distributions have the same number of modes, looking at the log-length distribution can be misleading, since taking logarithms can change not only the number but also the location of modes

### 3. IS THE DISTRIBUTION OF PLASMID LENGTHS BIMODAL?

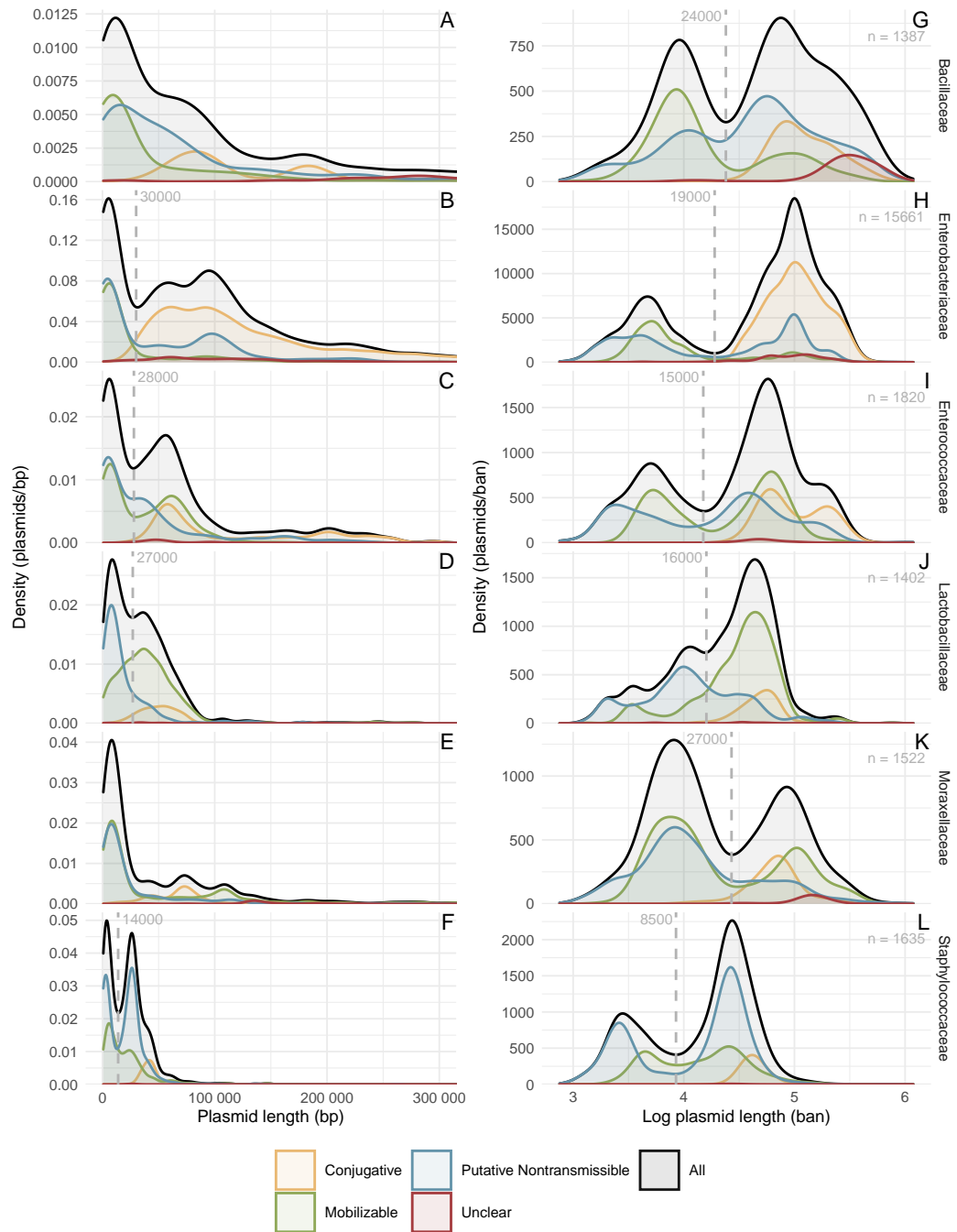


Figure 3.3: Distributions of plasmid length (left column, panels A–F) and log-length (right column, panels G–L) for plasmids from the six bacterial families with at least 1000 sequences in the database. Note that the length distributions are cut off at 300 000 bp for reasons of scale. The vertical lines show the location of the antimode of each distribution to the nearest kilobase. “Ban” are the units of the base-10 logarithm. The sample size for each family is given by  $n$  in the upper right corner of each row. For the data source and details on the figure construction see Figure 3.1.

(and of antimodes). This shift in the mode can, for example, be seen for conjugative plasmids in Figure 3.1. The modal log-length of a conjugative plasmid appears to be about 5 (corresponding to a length of  $10^5$  bp), while the modal length is smaller (closer to 50 000 bp). Similarly, within bacterial families where both the length and log-length distributions are bimodal, the antimode dividing small and large plasmids is not the same for the length and the log-length distributions (Figure 3.3). Thus even if there are distinct groups of “large” and “small” plasmids, trying to read their typical sizes or the dividing line between them off of an estimate of the distribution of log-lengths will be misleading. The accuracy of the location of modes and antimodes is, of course, in either case affected by the quality of the identification of plasmids in the database.

### 3.3 Discussion

We have shown that estimates of the distribution of plasmid lengths from available plasmid sequence data are not bimodal, at the global level at least, even though estimates of the distribution of log-lengths are, and that this kind of discrepancy is an ordinary feature of probability distributions. Within bacterial families, such a bimodal length distribution often exists, but the modes and antimode of the length distribution are incorrectly predicted by the log-length distribution. It is important to note that this does not mean that considering the distribution of log-lengths is wrong. There may well be reasons to be interested in the distribution of log-lengths rather than lengths. But the distribution of log-lengths is different from the distribution of lengths, and information about the distribution of lengths cannot be directly read off the graph of the distribution of log-lengths.

How did the confusion of the length and log-length distributions arise? We suspect that it is due to combining density estimation and a log scale automatically. It is common and reasonable, when plotting a graph with plasmid length on one axis, to use a log scale, because of the large range of plasmid lengths. However, “plotting a density on a log scale” could mean one of two very different things depending on the order that the density estimation and the log transformation are done. If the density estimation is done first, on the raw lengths, and then this estimate is plotted with a log-scaled axis, the result is like Figure 3.1B, which is a graph of the distribution of lengths (the numerical values of the plotted functions are the densities of plasmid lengths), although an unconventional one. But if the log transformation is done first, and then the density estimated from the log-transformed data, one gets an estimate of the distribution of log-lengths. Since Figure 3.1B is almost never what is actually wanted (it is hard to interpret and puts excessive emphasis on the lower tail of the distribution), graphing software asked to estimate a density and use a log scale will almost always choose to do the log transformation first.

### 3. IS THE DISTRIBUTION OF PLASMID LENGTHS BIMODAL?

---

Irrespective of the distribution of plasmid lengths, plasmid length and the biology of plasmids are closely entangled. For example, it is fairly clear from Figure 3.1 that there is a lower bound on the size of a conjugative plasmid (presumably set by the size of the genes needed to code for the complete conjugation machinery), and below this bound there are only mobilizable and nontransmissible plasmids (see also Coluzzi et al. 2022), while above it there are plasmids of all mobility classes. Another example is the decrease in the rate of transformation with plasmid size (Hanahan 1983). Plasmid size also influences the routes to adaptation: small plasmids are more likely to escape from bacterial restriction-modification systems through mutations in the target sequence than large plasmids, which due to their length contain more copies of the target sequence and need to resort to other means to escape restriction-modification systems (Shaw et al. 2023). Conversely, plasmid size may itself be the result of adaptation. Humphrey et al. (2021) hypothesize that the two modes in the size distribution of plasmids in *Staphylococcus aureus* (observed there on a linear scale; see also our Figure 3.3F) have evolved because those sizes are optimal for transduction by phages and *Staphylococcus aureus* pathogenicity islands (SaPis). It has further been found that small and large plasmids in *Klebsiella*, *Escherichia*, and *Salmonella*, as distinguished by the antimode of the log-length distribution, belong to different plasmid taxonomic units (Wang and Dagan 2024). It is possible that there is a fundamental reason why plasmids would cluster into plasmid taxonomic units by log-length, but it is also possible that the antimodes of the length and the log-length distributions are similar enough that the choice of distribution does not make a meaningful difference, i.e., both lead to nearly identical sets of small and large plasmids. Generally, since plasmid sizes and their distribution vary across bacterial families, within-family distributions of plasmid length are for many questions more informative than the distribution pooled across all plasmids.

The division of plasmids into large and small is said to be associated with a corresponding division into low and high copy number plasmids. It would be interesting to see if the distribution of plasmid copy numbers is bimodal, and to measure the extent to which it is associated with length. Unfortunately, it would be difficult to perform a similar analysis of the copy number distribution: plasmid copy number is not as frequently measured and reported as length, and is not included in databases of plasmid sequences. Moreover, the copy number is inherently more plastic than length, and may vary at different points of the host cell cycle or in different growth conditions, so that it may become impossible to speak of “the” copy number or to compare copy numbers of plasmids measured in different conditions.

To conclude, a clear understanding of the biology of plasmid length would be best served by clearly distinguishing between the distribution of plasmid lengths and that of log-lengths when discussing plasmid length,

and presenting whichever is relevant to the question being asked. Since the fact that the two distributions have different shapes is not immediately apparent, confounding the two distributions can cause confusion or be misleading.



## EVOLUTIONARY RESCUE OF BACTERIAL POPULATIONS BY HETEROZYGOSITY ON MULTICOPY PLASMIDS

**Abstract.** *Bacterial plasmids and other extrachromosomal DNA elements frequently carry genes with important fitness effects for their hosts. Multicopy plasmids can additionally carry distinct alleles of host-fitness-relevant genes on different plasmid copies, allowing for heterozygosity not possible for loci on haploid chromosomes. Plasmid-mediated heterozygosity may increase the fitness of bacterial cells in circumstances where there is an advantage to having multiple distinct alleles (heterozygote advantage); however, plasmid-mediated heterozygosity is also subject to constant loss due to random segregation of plasmid copies on cell division. We analyze a multitype branching process model to study the evolution and maintenance of plasmid-mediated heterozygosity under a heterozygote advantage. We focus on an evolutionary rescue scenario in which a novel mutant allele on a plasmid must be maintained together with the wild-type allele to allow population persistence (although our results apply more generally to the maintenance of heterozygosity due to heterozygote advantage). We determine the probability of rescue and derive an analytical expression for the threshold on the fitness of heterozygotes required to overcome segregation and make rescue possible; this threshold decreases with increasing plasmids copy number. We further show that the formation of cointegrates from the fusion of plasmid copies increases the probability of rescue. Overall, our results provide a rigorous quantitative assessment of the conditions under which bacterial populations can adapt to multiple stressors through plasmid-mediated heterozygosity. Many of the results are furthermore applicable to the related problem of the maintenance of incompatible plasmids in the same cell under selection for both.*

### 4.1 Introduction

Many bacteria carry, in addition to the bacterial chromosome, extrachromosomal genetic elements called plasmids, which replicate independently of the chromosome and often exist in multiple copies. The simplest plasmids have only the core backbone genes required to ensure their stable maintenance in the bacterial host, but many plasmids also carry genes that have important effects on the host phenotype (Garcillán-Barcia et al. 2011). Perhaps the most famous of these are antibiotic resistance genes, which pose a serious threat to the effectiveness of the clinical treatment of bacterial infections (Carattoli 2013), but phenotypes contributed by plasmid-borne

---

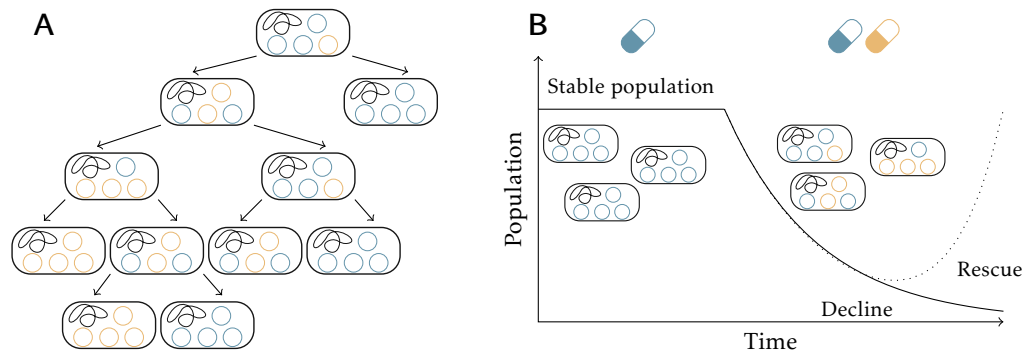
This chapter has been accepted for publication as: Ian Dewan and Hildegard Uecker (2025). "Evolutionary rescue of bacterial populations by heterozygosity on multicopy plasmids". In: *Journal of Mathematical Biology*. DOI: 10.1007/s00285-025-02182-4. Forthcoming

genes also include heavy-metal resistance, virulence, metabolism of novel carbon sources, and symbiotic interactions with other organisms, among many others (see e.g. Portnoy and Martinez 1985; Silver 1992; Beijersbergen et al. 1994; Yu et al. 1996; Anda et al. 2015; Wardell et al. 2021). The plasmids carried by a bacterium form an important part of the bacterial genome, and the contribution of plasmids must be considered in explaining and predicting bacterial evolution. In this context, much attention has been given to the ability of many plasmids to transfer horizontally between bacteria, infecting new hosts, in a process called conjugation (Falkow 1975; Smillie et al. 2010); however, very many plasmids are transmitted only vertically, or depend on other plasmids for horizontal transfer (Coluzzi et al. 2022), and purely vertically transmitted plasmids can play an important role in bacterial adaptation.

The backbone genes of a plasmid are responsible for ensuring the stable maintenance of the plasmid by vertical transmission. This requires ensuring that both daughter cells receive copies of the plasmid at host cell division and limiting the possible reduction in host growth imposed by fitness costs of plasmid carriage. To this end, the number of copies of a plasmid per host cell is regulated by the plasmid itself to remain approximately constant; the plasmid copy number is therefore an intrinsic property of the plasmid-host system which plays an important role in the biology of the plasmid (Rodríguez-Beltrán et al. 2021a). Although for some plasmids the copy number is kept quite low, at only one or a few copies (e.g. Gustafsson and Nordström 1980; Casjens et al. 2000; Nordström 2006; Ismail et al. 2014; Brovedan et al. 2019; Smith et al. 2021), many plasmids are found in their hosts in high numbers of copies, often in tens or sometimes even hundreds (e.g. Kiyosawa et al. 1993; Burian et al. 1997; Projan et al. 1987; Prangishvili et al. 1998; Anda et al. 2015; San Millan et al. 2016; Santos-Lopez et al. 2016). Interest in the evolution of these multicopy plasmids and their contribution to bacterial adaptation has recently increased, particularly due to their contribution to antibiotic resistance (San Millan et al. 2009; Gama et al. 2018).

Existence in multiple copies per cell has multiple effects on the contribution of multicopy plasmids to bacterial evolution. The copy numbers of genes carried on a multicopy plasmid are higher than those of genes on the (for many species haploid) bacterial chromosome, so genes on plasmids have both higher mutational input and dosage (these effects have been shown in experiments by, e.g., San Millan et al. 2016). The presence of the plasmid in multiple copies also provides the possibility of heterozygosity for loci on a multicopy plasmid which is not possible for loci on a haploid chromosome.

Once there are multiple variants of the plasmid in the cell, the process of plasmid segregation at host cell division will have important effects on the distribution of plasmid variants among cells in the population (San



**Figure 4.1:** A model of the rescue of a bacterial population by plasmid-mediated heterozygosity. (A) The loss of plasmid-mediated heterozygosity due to segregation. From an initial heterozygote cell, random segregation results in the two plasmid types ending up in different, homozygous cells. (B) The evolutionary rescue scenario. An initial stable population carrying a plasmid with an adaptive gene on it is exposed to a sudden environmental change, which leads the population to decline. A mutation on the plasmid (orange plasmids) can rescue the population, but only if it is maintained together with the wild-type (blue) plasmid. Based on Figure 1 of Santer and Uecker (2020).

Millan et al. 2016; Ilhan et al. 2018; Santer and Uecker 2020; Rodríguez-Beltrán et al. 2021a). At host cell division, the copies of a plasmid in the host are distributed between the two daughter cells: for low copy number plasmids this is often aided by an active partitioning system to ensure each daughter gets at least one copy of the plasmid, while high copy number plasmids may simply rely on random distribution of the plasmids between daughter cells (Zielenkiewicz and Cegłowski 2001). Random assortment of the plasmid copies at host cell division (i.e. segregation independent of the allele at the heterozygous locus) tends to eliminate heterozygosity, since every cell division has some nonzero probability of producing a homozygote daughter cell by chance, and the homozygotes are absorbing states of this process (see Figure 4.1A). An analogous situation appears when two different (multicopy) plasmids that share the same replication system reside within the same cell—the plasmid variants are said to be *incompatible*, since they will not be maintained together in the same cell line indefinitely, and over time separate lines containing either one or the other will emerge (Novick 1987). Such segregative loss of one or the other plasmid was already observed and quantified in early studies on plasmid incompatibility; this loss can be counteracted by selection for maintenance of the plasmid variants together (see e.g. Uhlin and Nordström 1975; Cullum and Broda 1979; Rodríguez-Beltrán et al. 2018).

Driven by these empirical observations, early models were developed to describe the kinetics of the loss of heterozygosity in multicopy plasmids (or loss of incompatible plasmids) in the absence of selection under varying as-

assumptions about the mechanisms of segregation and replication (Ishii et al. 1978; Novick and Hoppensteadt 1978; Cullum and Broda 1979). Recently, interest in the evolution of multicopy plasmids and their hosts has gained momentum (e.g. San Millan et al. 2016; Rodríguez-Beltrán et al. 2018; Santer and Uecker 2020; Rodríguez-Beltrán et al. 2021a; Garoña et al. 2021; Hernandez-Beltrán et al. 2022; Santer et al. 2022; Garoña et al. 2023). This includes experimental studies of multicopy plasmids with accompanying models: some of these focus on the fixation process of novel beneficial alleles on plasmids (Ilhan et al. 2018; Garoña et al. 2023), while others focus on the effect of selection for heterozygotes on maintaining plasmid variants together despite segregation (Rodríguez-Beltrán et al. 2018). Recent modelling studies have also examined the effects of plasmid replication and segregation on the fate of novel beneficial alleles that appear on a single copy of a multicopy plasmid (Halleran et al. 2019; Santer and Uecker 2020).

We here focus on scenarios of heterozygote advantage, where the optimal bacterial fitness is obtained by having multiple variants of a single plasmid carrying distinct alleles. This could be because, for example, the two alleles confer resistance to distinct antibiotics present in the environment. The models we present are of an *evolutionary rescue* scenario, of the kind introduced by Gomulkiewicz and Holt (1995) and shown in Figure 4.1B, extending previous models of rescue on plasmids (Tazyman and Bonhoeffer 2014; Santer and Uecker 2020). In such a scenario, an environmental change exposes a population to new conditions to which it is maladapted, and the population therefore enters a demographic decline which would lead to extinction. However, if a novel mutation emerges which adapts individuals to the new conditions, and this mutation survives to spread in the population, the population can be rescued from extinction. Since this rescue depends on the mutation occurring and surviving the initial period in which it is rare in the population, rescue is inherently a stochastic process, and the key question is the probability that rescue will occur in a given population. When maintenance of both plasmid variants is required for rescue, rescue requires not only the establishment of the new mutation against the force of genetic drift, but also the persistence of both variants against the force of segregation at all times. It is intuitively clear that such persistence is only possible if the fitness of heterozygote cells is large enough to bear the production of unfit homozygous cells. Based on a multitype branching process model, we determine the probability of evolutionary rescue and derive analytical conditions on the plasmid copy number and the fitness of heterozygous cells that need to be fulfilled for maintenance of heterozygosity, and thus rescue, to be possible at all. While we derive this condition in the context of evolutionary rescue, it holds more broadly for the maintenance of heterozygosity or of incompatible plasmids within the same cell.

## 4.2 The model

Consider a demographically stable population of bacteria which carries a plasmid present at a fixed copy number  $n$  in each bacterial cell. This population is then exposed to novel environmental conditions to which it is maladapted, and the population begins a demographic decline which will eventually lead to its extinction. Evolution, however, may rescue this population from extinction. A mutation might appear at a locus on the plasmid which, *if maintained together with the wild-type allele at that locus*, will adapt the population to the novel environmental conditions; that is, mutant *heterozygotes* can survive in the novel environmental conditions, but mutant and wild-type homozygotes are both maladapted. We assume that the population is homozygous for this particular plasmid before the environmental change; that is, that there is only one variant of the plasmid in the population, at least with respect to the locus of interest.

These conditions constitute the evolutionary-rescue-due-to-heterozygote-advantage scenario we wish to describe in the model. We expect that heterozygote advantage of plasmid-borne loci could occur in many circumstances, but as a motivating example we consider a bacterial population resistant to antibiotic A due to a resistance gene located on a plasmid. If the population is exposed to antibiotics A and B simultaneously, then lack of resistance to antibiotic B will cause the population to decline. But if a mutation occurs on a plasmid that converts the A-resistance allele into a B-resistance allele, its host is now resistant to antibiotic B, and (provided that  $n > 1$ ) the remaining wild-type plasmids continue to confer resistance to antibiotic A: therefore the host can survive and grow in the novel conditions. Its descendants are resistant to both antibiotics and the population might be rescued from extinction, provided that both plasmid types, and therefore both resistances, are maintained. Rodriguez-Beltran et al. (2018) found very similar situation arose in an evolution experiment where a mutation in a plasmid-borne TEM-1  $\beta$ -lactamase shifted it from being mostly effective against ampicillin to mostly effective against ceftazidime. More generally, this scenario could be particularly likely with  $\beta$ -lactam antibiotics, as common small mutations can alter the affinity of  $\beta$ -lactamases for different drugs (Farr et al. 2023). Such a combination of two closely-related antibiotics would usually not be used in a clinical context to treat a patient (although recent research has suggested that sequential treatment with different  $\beta$ -lactams—including both penicillins, to which the wild-type TEM allele provide resistance, and cephalosporins, to which mutant extended-spectrum TEM alleles provide resistance—may be more effective than previously thought: Batra et al. (2021)). However, our scenario might arise in accidental environmental exposure of bacterial populations to antibiotics, for example from agricultural sources, where nothing prevents two related antibiotics from cooccurring.

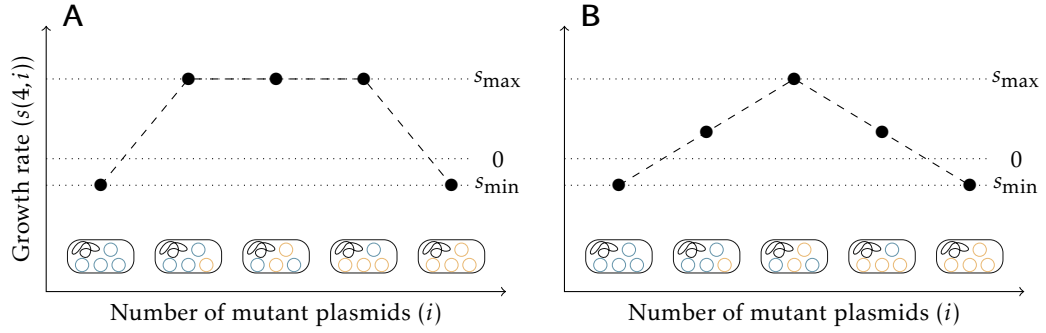


Figure 4.2: Examples of the two fitness functions considered, for  $n = 4$ : (A) the dominant fitness function; (B) the peaked fitness function.

Our model extends that of Santer and Uecker (2020) to the case of heterozygote advantage. Suppose that cells with  $i$  mutant plasmids (and therefore  $n - i$  wild-type plasmids) reproduce at a rate  $\lambda_i$  and die at a rate  $\mu_i$ . We assume throughout that the plasmid composition only affects the replication rate but not the death rate. Since the unit of time is arbitrary, we may fix  $\mu_i = 1$ . If we then set  $\lambda_i = 1 + s(n, i)$ , the function  $s(n, i)$  gives the net growth rate (or Malthusian fitness) of cells with  $i$  mutant plasmids out of  $n$ . Since maintenance of both wild-type and mutant plasmids is necessary for the population to persist, we have that  $s(n, 0) < 0$  and  $s(n, n) < 0$ ; beyond this, the choice of fitness function is free.

We will consider two different possible fitness functions. The simpler is

$$s(n, i) = \begin{cases} s_{\min} & i = 0 \text{ or } i = n, \\ s_{\max} & \text{otherwise,} \end{cases}$$

where  $s_{\min} < 0$  and  $s_{\max} > 0$ . We shall call this the *dominant fitness function* (shown in Figure 4.2A), since a single copy of either allele confers the full fitness effect of that allele. It is also possible to imagine that dominance is intermediate or that there is a gene dosage effect, in which additional copies of a gene increase fitness. Combined with the necessity to maintain both alleles, this might produce a fitness function like

$$s(n, i) = s_{\max} - \frac{2(s_{\max} - s_{\min})}{n} \left| i - \frac{n}{2} \right|,$$

with  $s_{\min} < 0$  and  $s_{\max} > 0$  as before; this function linearly interpolates between a maximum fitness at an exact half-and-half mixture of wild-type and mutant plasmids, and a minimum fitness for the homozygotes (shown in Figure 4.2B). We shall call this the *peaked fitness function*. If there are gene dosage effects,  $s_{\max}$  would increase with  $n$ ; we do not consider this

in our numerical examples, but all results for a given  $n$  also apply to this scenario.

We have assumed above that every cell has  $n$  copies of the plasmid of interest. To maintain this number, we assume the cell replicates its plasmids to a copy number of  $2n$  before reproduction, and then segregates equal numbers into each daughter cell. At replication of wild-type plasmids, mutations happen with probability  $u$ , but we disregard this for the moment. We denote by  $P(i \rightarrow \{j, k\})$  the probability that when a cell with  $i$  mutant plasmids (out of  $n$  total plasmids) divides it produces daughters with  $j$  and  $k$  mutant plasmids (again out of  $n$  total). We will consider two models of replication, called *regular* and *random replication*, after the taxonomy of Novick and Hoppensteadt (1978).

In regular replication, every plasmid copy in the parent cell is replicated once, and then the copies are segregated randomly, but maintaining the copy number, into each daughter cell. The number of mutant plasmids in the daughter cell then has a hypergeometric distribution, giving a segregation probability function

$$P(i \rightarrow \{j, k\}) = \begin{cases} 0 & j + k \neq 2i, \\ \frac{\binom{2i}{j} \binom{2(n-i)}{n-j}}{\binom{2n}{n}} & j = k = i, \\ 2 \frac{\binom{2i}{j} \binom{2(n-i)}{n-j}}{\binom{2n}{n}} & \text{otherwise.} \end{cases} \quad (4.1)$$

In random replication, each new plasmid to be replicated is randomly chosen from the pool of the initial plasmids together with the products of previous replications. This proceeds until there are  $2n$  total plasmids, when the number of mutants will be distributed according to a Pólya urn scheme. These  $2n$  plasmids are then randomly segregated into the daughter cells as before. The number of mutant plasmids in a daughter cell is then distributed as

$$P(i \rightarrow \{j, k\}) = \begin{cases} \frac{\binom{j+k}{j} \binom{2n-j-k}{n-j}}{\binom{2n}{n}} \cdot \frac{\binom{2n-j-k-1}{n-j-k+i} \binom{j+k-1}{j+k-i}}{\binom{2n-1}{n}} & j = k, \\ 2 \frac{\binom{j+k}{j} \binom{2n-j-k}{n-j}}{\binom{2n}{n}} \cdot \frac{\binom{2n-j-k-1}{n-j-k+i} \binom{j+k-1}{j+k-i}}{\binom{2n-1}{n}} & \text{otherwise.} \end{cases} \quad (4.2)$$

The first fraction in the expression is the same hypergeometric distribution for segregation as in the regular replication model, representing the probability that the  $j+k$  mutant plasmids produced by replication are divided into  $j$  in one daughter cell and  $k$  in the other. The second fraction is the probability of producing  $j+k$  mutant plasmids from  $i$  mutant plasmids during replication under the Pólya urn process model: for the derivation see appendix B.

For rescue to happen, a mutation must occur on the plasmid and be maintained together with the wild-type plasmid in the same cell indefinitely. To determine the probability of this occurring in a given population,

we split the process into two parts. In the first, we look at the descendants of a single novel mutant, and determine the probability that these descendants will survive indefinitely rather than suffering stochastic loss. This is called the *establishment probability* of the mutant. We ignore additional mutations that might recurrently generate the mutant from the wild-type plasmid, which is a rare event in a small cell line. In the second part, we can then estimate the number of mutants which will occur in wild-type homozygous cells before the initial population goes extinct, and determine the probability that at least one of them will establish; this is called the *rescue probability*.

We have described our birth-death model thus far as a continuous-time multitype branching process, in which cells divide and die independently of each other, excluding in particular resource competition. Of course, a population cannot grow indefinitely. However, we are here considering a population that is, at least initially, declining due to maladaptation. Mutant cells are rare in the early stages of establishment in which we are interested and therefore independent from each other to a good approximation; this approximation is commonly made even in more complicated models and dates back to the very early calculations of establishment probabilities of beneficial alleles (Haldane 1927). We further assume that we are far enough away from carrying capacity that we can also neglect competition with the declining wild-type population.

### 4.3 The establishment probability

#### 4.3.1 Analysis

To determine the establishment probability for the descendants of a given bacterial cell, we first determine the extinction probability, the probability that the cell's descendants will at some point go extinct. While our model was described above in continuous time, we transition to a discrete-time branching process for the analysis of the establishment probability. We can do this because we are only interested in the final outcome of the process—extinction or survival—and not in the timing of events.

A given cell will have either zero daughter cells (if it dies before reproducing), with probability  $\mu_i/(\lambda_i + \mu_i)$ , or two daughter cells (if it reproduces), with probability  $\lambda_i/(\lambda_i + \mu_i)$ . Since the survival of the cell depends on its complement of plasmids, we will need to track the number of mutant and wild-type plasmids in each daughter cell. The extinction probabilities  $Q_0, Q_1, \dots, Q_n$  of the descendants of cells with, respectively,  $0, 1, \dots, n$  mutant and  $n, n-1, \dots, 0$  wild-type plasmids are given by the least fixed point of the generating function of the distribution of daughter cells of a single cell (Sewastjanow 1975, Folgerung V.1.1). This function is given by  $f(z_0, \dots, z_n) = (f_0(z_0, \dots, z_n), \dots, f_n(z_0, \dots, z_n))$ , where each component  $f_i$  is

given by

$$f_i(z_0, \dots, z_n) = \frac{\mu_i}{\lambda_i + \mu_i} + \frac{\lambda_i}{\lambda_i + \mu_i} \sum_{\{j,k\}} P(i \rightarrow \{j, k\}) z_j z_k, \quad (4.3)$$

the sum being taken over all unordered pairs  $\{j, k\}$  of numbers of plasmids. The first term of this expression corresponds to the probability of immediate cell death, while the sum gives the probabilities of each possible pair of daughter cell types upon reproduction. The fixed point of the generating function is then the solution to the system of  $n+1$  equations in (4.3) in  $n+1$  unknowns  $z_i = Q_i$ , which can be solved numerically. This is also intuitive: the initial cell has either zero or two daughter cells. In the case that the initial cell has zero daughter cells, its descendants go extinct immediately (corresponding to the first term in the fixed point equation); in the case it has two, its descendants go extinct if and only if the descendants of both daughter cells go extinct (corresponding to the second term).

The fixed point gives us the probability that the descendants of a bacterial cell eventually go extinct, but what we really want is the probability that the descendants of a novel mutant never go extinct: the establishment probability. Under the regular replication model, this is very simple to determine: a novel mutant allele appears at first on a single plasmid in a single host cell, so the establishment of the mutant allele occurs only when the descendants of that cell do not go extinct, and

$$P_{\text{est}} = 1 - Q_1. \quad (4.4)$$

In the random replication model the situation is slightly more complex. When a mutation occurs during plasmid replication, the novel mutant plasmid is available to be replicated during the same bacterial generation. If it is replicated, there will be multiple mutant plasmids to segregate between daughter cells, and each daughter cell might get one or more mutant plasmids. Thus the establishment process of the mutation gets a head start, possibly starting from multiple mutant plasmids which may possibly be in two separate cells. The establishment probability becomes a quadratic form in the extinction probabilities of the individual types

$$P_{\text{est}} = 1 - \sum_{0 \leq j, k \leq n} \underbrace{\frac{(2n + (n-1)(j+k))}{n(j+k)(j+k+1)} \frac{\binom{n}{j+k}}{\binom{2n-1}{j+k}}}_{\text{replication}} \underbrace{\frac{\binom{j+k}{j} \binom{2n-j-k}{n-j}}{\binom{2n}{n}}}_{\text{segregation}} Q_j Q_k, \quad (4.5)$$

where the first factor inside the sum is the probability of going from no mutant to  $j+k$  mutant plasmids during replication, conditional on exactly one mutation occurring (for the derivation of this probability and its assumptions, see appendix B), and the second factor is the probability of the  $j+k$  mutant plasmids being segregated into daughter cells with  $j$  and  $k$  mutant plasmids.

### 4.3.2 Results

The establishment probabilities for the regular replication model with a dominant fitness function are shown in Figure 4.3A. Perhaps unsurprisingly, increasing the fitness of heterozygotes ( $s_{\max}$ ) increases the establishment probability. More interesting is the effect of the copy number of the plasmid. At low copy number, the establishment probability increases with copy number: this is due to the loss of heterozygosity to segregation, which is most pronounced for small copy numbers. For larger copy numbers, the establishment probability stabilizes—indeed, a close examination shows that it begins to decrease for larger copy numbers, probably because at high copy numbers it takes more generations to go from a single mutant plasmid copy to an approximately equal number of wild-type and mutant plasmids, where the probability of producing a homozygote daughter cell is lowest (a similar effect was observed in the non-heterozygote-advantage case by Santer and Uecker (2020)).

The establishment probabilities for the regular replication model with a peaked fitness function are shown in Figure 4.3B. The establishment probabilities are lower than with the dominant fitness function, and decline more precipitously for high copy numbers, since the fitness of most cell types, other than the perfectly balanced heterozygotes, has been reduced. In fact, for large enough copy numbers, the fitness of cells with positive but small numbers of one or the other plasmid type becomes negative.

The establishment probabilities for the random replication model are shown in Figures 4.3C (with the dominant fitness function) and 4.3D (with the peaked fitness function). The general trend is the same as under the regular replication model, but the establishment probabilities are all reduced. This is because random replication increases the rate of loss of heterozygosity to segregation relative to regular replication: the “rich-get-richer” behaviour of the Pólya urn scheme means that the proportion of mutant plasmids is likely to be more unbalanced after replication, and the probability of a homozygote daughter cell is increased.

For comparison, we performed stochastic computer simulations, which are in excellent agreement with equations (4.4) and (4.5) (see Appendix C).

Note that although the establishment probabilities appear to level off for the random replication model with dominant fitness function, in fact the establishment probability goes to zero for arbitrarily large copy number under random replication. This is because the mutant allele appears initially in a single copy: as the copy number increases, the probability of that mutant plasmid being selected for replication decreases. In the limit, the mutant plasmid is never replicated at all, and eventually the cell line containing it is lost to demographic stochasticity, and the mutant goes extinct.

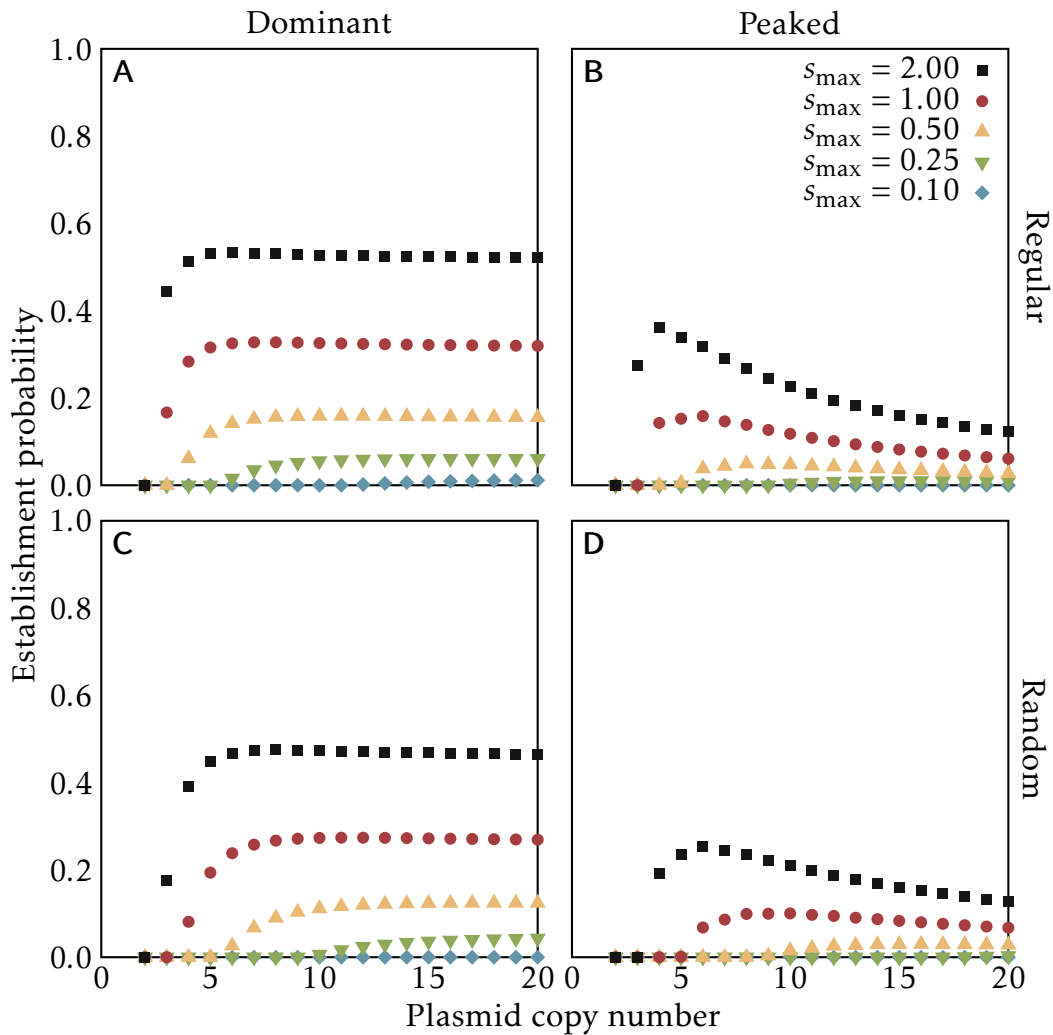


Figure 4.3: Establishment probabilities of rescue mutations on a plasmid of a given copy number in the heterozygote advantage scenario, under the regular (panels A and B, from equation (4.4)) or random (panels C and D, from equation (4.5)) replication assumption. Colours of points indicate the fitness  $s_{\max}$  of all heterozygotes (with the dominant fitness function, panels A and C) or the maximum fitness of heterozygotes (with the peaked fitness function, panels B and D). For all cases,  $s_{\min} = -0.1$ . The code for numerical calculation of establishment probabilities is available as supplementary material.

#### 4.4 The rescue probability

##### 4.4.1 Analysis

Now let us consider the overall probability of rescue for the entire population. We follow the typical approach for approximating rescue probabilities from branching process models (used by, e.g, Orr and Unckless 2008; Alexander and Bonhoeffer 2012; Martin et al. 2013; Tazzyman and Bonhoeffer 2014; Santer and Uecker 2020): we assume that the initial population of homozygote wild-type cells is large, and that mutations are sufficiently rare that their occurrence and establishment can be treated as being independent. The first assumption allows us to treat the initial population of homozygote wild-type cells deterministically; it has an initial size  $N_0$  and growth rate  $\lambda_0 - \mu_0 = s(n, 0) < 0$ , so its size at time  $t$  is given by

$$N(t) = N_0 e^{s(n,0)t}.$$

If the per locus mutation rate is  $u$ , then at each cell division the probability that a mutation occurs and will successfully establish is  $unP_{\text{est}}$  (we assume that at most one mutation occurs per cell division). Thus the instantaneous rate of occurrence at time  $t$  of mutations that will successfully establish is  $unP_{\text{est}}\lambda_0 N(t)$ . Our second assumption, of independence of mutations, means that successfully establishing mutations form an inhomogeneous Poisson process with intensity function  $unP_{\text{est}}\lambda_0 N(t)$ . The rescue probability is just the probability that this process has at least one event (i.e., at least one mutation occurs and successfully establishes), and is given by

$$P_{\text{rescue}} = 1 - e^{-\int_0^\infty unP_{\text{est}}\lambda_0 N(t) dt} = 1 - e^{-unN_0 \frac{1+s(n,0)}{|s(n,0)|} P_{\text{est}}}. \quad (4.6)$$

The Poisson process approximation also allows us to find the distribution of the time until the first successful rescue mutation occurs (conditional on one occurring at all): the time of occurrence  $T$  is distributed as

$$P(T \leq t) = \frac{1 - e^{-\int_0^t unP_{\text{est}}\lambda_0 N(\tau) d\tau}}{P_{\text{rescue}}}. \quad (4.7)$$

Our choice to set  $\mu_i = 1$  means that the unit of time here is the mean time to death of host cells. Note, however, that visible recovery of the population will in general occur much later: even after a successful mutation has occurred, the total population size still keeps decreasing initially, since the majority of cells are still maladapted wild-type cells.

##### 4.4.2 Results

The rescue probabilities are shown in Figure 4.4. The major difference to the trends in the establishment probabilities in Figure 4.3 is that the rescue

probability increases with copy number even when the establishment probability is stable or declining; this is because the increased mutational input with a greater copy number results in more mutants arising before wild-type population extinction, and thus a higher probability that at least one establishes; this effect turns out to be stronger than that of the decrease in the establishment probability. As for the establishment probability, we additionally performed stochastic computer simulations, which confirm the results (see supplementary material).

Exemplary trajectories of the population size over the course of the rescue process from stochastic simulations following the Gillespie algorithm (Gillespie 1976) are shown in Figure 4.5. Given rescue occurs, population recovery occurs earlier and is faster for higher copy numbers: population growth turns from negative to positive at earlier times, the population size at the minimum is larger, and subsequent growth is more rapid. We calculate the (long-run) growth rate of the rescued population for the dominant fitness function at the end of the next section.

#### 4.5 Critical values of $n$ and $s_{\max}$

Examining Figures 4.3 and 4.4 shows that not only does the effect of loss of heterozygosity during segregation cause the establishment and rescue probabilities to decline sharply at low copy numbers, but below a certain threshold (which depends on  $s_{\max}$ ) establishment and therefore rescue becomes impossible ( $P_{\text{rescue}} = P_{\text{est}} = 0$ ). We can, in fact, describe analytically the relation between the heterozygote advantage  $s_{\max}$  and the copy number  $n$  which is required to hold in order for establishment to be possible with the dominant fitness function, in both the regular and random replication models. We take the opportunity here to give names to the probabilities that a single bacterial cell reproduces before it dies, which will be important in deriving the analytical relations. With the dominant fitness function, there are only two distinct values of this probability,

$$\rho_{\text{hom}} = \frac{\lambda_0}{\lambda_0 + \mu_0} = \frac{\lambda_n}{\lambda_n + \mu_n} = \frac{1 + s_{\min}}{2 + s_{\min}} \quad (4.8)$$

for homozygote cells and

$$\rho_{\text{het}} = \frac{\lambda_1}{\lambda_1 + \mu_1} = \dots = \frac{\lambda_{n-1}}{\lambda_{n-1} + \mu_{n-1}} = \frac{1 + s_{\max}}{2 + s_{\max}} \quad (4.9)$$

for heterozygotes.

**Theorem 1.** *In the model with regular replication and a dominant fitness function, the establishment probability (and therefore the rescue probability) is non-zero if and only if  $n \geq 2$  and*

$$s_{\max} > \frac{2}{2n - 3}.$$

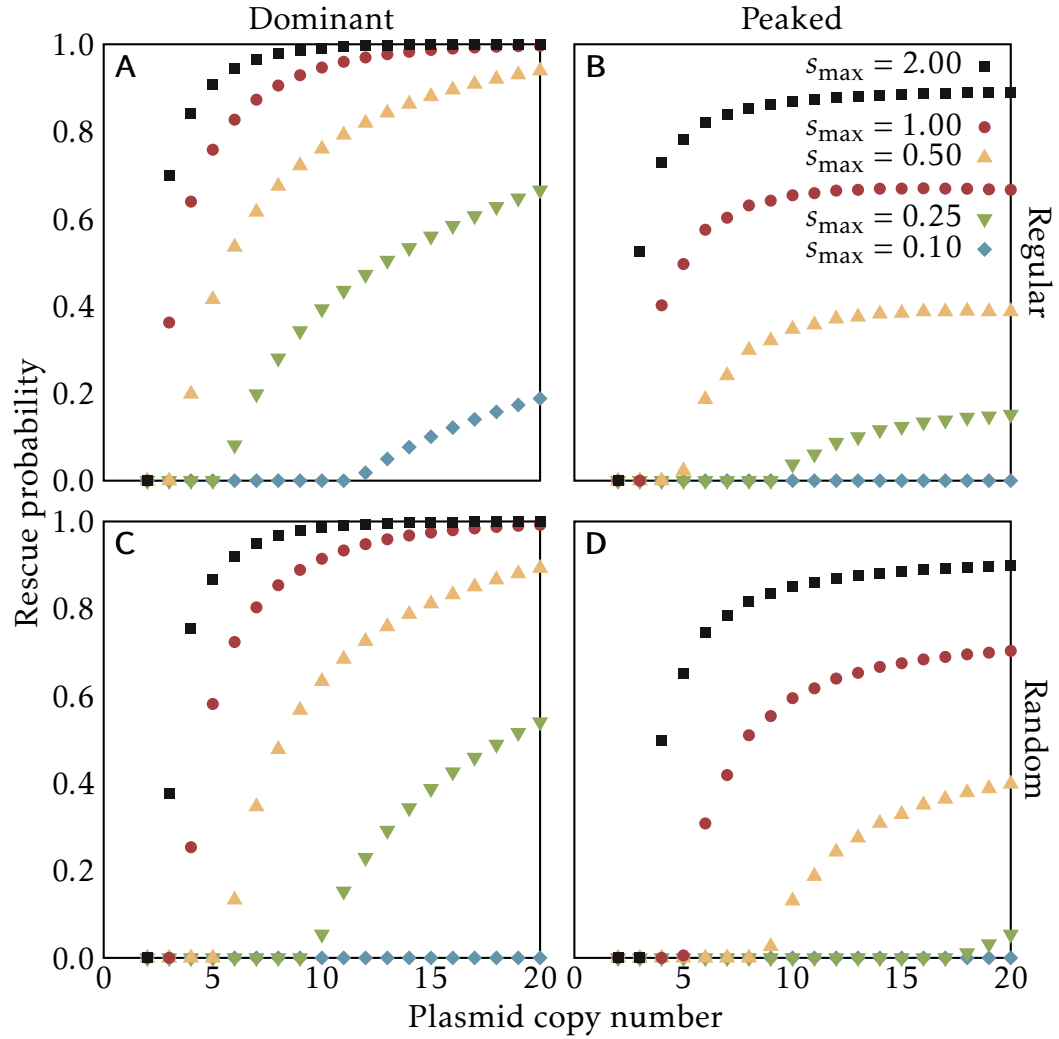
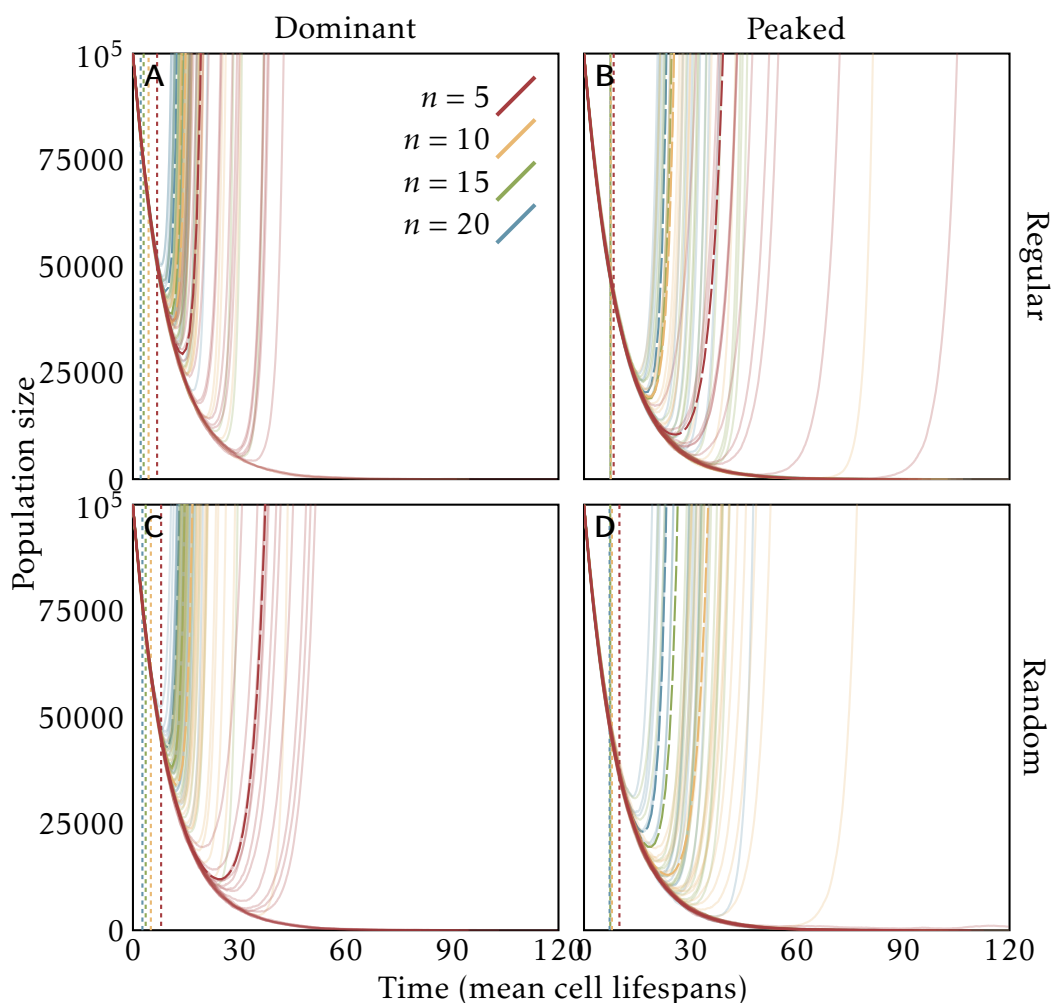


Figure 4.4: Probabilities (from equation (4.6)) of a bacterial population being rescued by a mutation on a plasmid of a given copy number in the heterozygote advantage scenario, under the regular (panels A and B) or random (panels C and D) replication assumption. Colours of points indicate the fitness  $s_{\max}$  of all heterozygotes (with the dominant fitness function, panels A and C) or the maximum fitness of heterozygotes (with the peaked fitness function, panels B and D). For all cases,  $s_{\min} = -0.1$  and  $uN_0 = 0.1$ . The code for numerical calculation of rescue probabilities is available as supplementary material.



*Figure 4.5:* Simulated population size trajectories of a bacterial population being rescued by a mutation on a plasmid in the heterozygote advantage scenario, under the regular (panels A and B) or random (panels C and D) replication assumption and dominant (panels A and C) or peaked (panels B and D) fitness function. Time (on the  $x$ -axis) is measured in units of the mean lifespan of host cells  $1/\mu_i$ . Colours of lines indicate the plasmid copy number: for each copy number, there are 20 simulated trajectories. The dashed curves show the mean of the simulations which did not go extinct; the dotted vertical lines indicate the mean time of occurrence of the first successful rescue mutation conditioned on rescue (from equation (4.7)). For all cases,  $s_{\max} = 1.0$ ,  $s_{\min} = -0.1$ , and  $uN_0 = 0.1$ . For the peaked fitness function and random replication (panel D), the establishment probability with  $n = 5$  is very small (although non-zero), and none of the 20 replicate populations survived. The code for the stochastic simulations is available as supplementary material.

*Proof.* As in the derivation of the establishment probability, we consider the branching process in discrete time. Let  $M$  be the matrix the  $i, j$  element of which is the expected number of daughter cells with  $j$  mutant plasmids from a cell with  $i$  mutant plasmids. In the model with regular replication and a dominant fitness function, the elements of this matrix have the form

$$m_{ij} = \frac{\lambda_i}{\lambda_i + \mu_i} \sum_{k=0}^n 2^{\delta_{jk}} P(i \rightarrow \{j, k\}) = 2 \frac{\lambda_i}{\lambda_i + \mu_i} \frac{\binom{2i}{j} \binom{2(n-i)}{n-j}}{\binom{2n}{n}},$$

where we have substituted the definition of  $P(i \rightarrow \{j, k\})$  for regular replication given in equation (4.1). By a general result of branching process theory, a discrete-time multitype branching process goes extinct with probability one if and only if it has no final classes and the largest eigenvalue of  $M$  is less than or equal to one (Sewastjanow 1975, Satz V.2.5). The first condition is a technical one, to exclude the degenerate case where there is no stochasticity in the branching process and the population stays at a fixed size: a final class is a subset of the types in the multitype branching process such that an individual of one of those types will have, with probability one, exactly one offspring and that offspring will be of one of the types in the class. In our case, there can be no final classes because every individual has a nonzero probability of having no offspring. Therefore, there is a nonzero probability of establishment if and only if  $M$  has an eigenvalue greater than one.

To find the eigenvalues of  $M$ , we factor it as a product  $M = 2RP$ , where  $P$  is the matrix with  $i, j$  element  $\binom{2i}{j} \binom{2(n-i)}{n-j} / \binom{2n}{n}$  and  $R$  a diagonal matrix containing the probabilities of reproduction, then find the eigenvalues of each factor individually—it will turn out, luckily, that all the eigenvalues of  $M$  are products of the eigenvalues of these factors. Finding the eigenvalues of  $R$  is easy, since it is diagonal: in fact, it has only two distinct eigenvalues, which are given by the two possible probabilities  $\rho_{\text{hom}}$  and  $\rho_{\text{het}}$  of reproduction for a cell with the dominant fitness function given in equations (4.8) and (4.9). The eigenvalue  $\rho_{\text{hom}}$  has an eigenspace consisting of all of the “pure homozygote” vectors, that is those with nonzero entries only in the first and last component, and the eigenvalue  $\rho_{\text{het}}$  has an eigenspace consisting of the “pure heterozygote” vectors (those with zero first and last components).

The eigenvalues of  $P$  were calculated by Schensted (1958); we present here an elaboration of her argument. The trick will be to find a basis of  $\mathbf{R}^{n+1}$  in which the matrix that represents the same linear transformation represented by  $P$  in the standard basis is upper triangular, which makes reading off the eigenvalues trivial (in other words, we show that  $P$  is similar to certain upper triangular matrix; Strang (2009, § 6.6)). This new basis will arise from considering the vector space  $\mathbf{R}[X]$  of polynomials in an arbitrary unknown  $X$  with real coefficients. The map  $T: \mathbf{R}[X] \rightarrow \mathbf{R}^{n+1}$  that takes a

polynomial  $p(X)$  to the vector

$$Tp = \begin{bmatrix} p(0) \\ p(1) \\ \vdots \\ p(n) \end{bmatrix}.$$

is clearly linear, and if we restrict the domain to polynomials of degree at most  $n$ , it becomes an isomorphism (since any  $n + 1$  points define a polynomial of degree at most  $n$ ). For any integer  $\beta$ ,  $\binom{X}{\beta} = X(X - 1)\cdots(X - \beta + 1)/\beta!$  is a polynomial in  $X$  of degree  $\beta$ , and the sequence  $\{\binom{X}{0}, \binom{X}{1}, \binom{X}{2}, \dots\}$  forms a basis for  $\mathbf{R}[X]$ ; moreover, if we truncate the sequence at  $\binom{X}{n}$ , we get a basis for the subspace of polynomials of degree at most  $n$ , which we can pass through the isomorphism  $T$  to obtain a basis for  $\mathbf{R}^{n+1}$ .

Looking carefully at the binomial coefficient identity

$$\sum_{\alpha=0}^n \underbrace{\frac{\binom{2i}{\alpha}\binom{2n-2i}{n-\alpha}}{\binom{2n}{n}}}_{P_{i\alpha}} \underbrace{\binom{\alpha}{\beta}}_{T\binom{X}{\beta}_\alpha} = \frac{\binom{n}{\beta}}{\binom{2n}{\beta}} \underbrace{\binom{2i}{\beta}}_{T\binom{2X}{\beta}_i} \quad (\text{integers } n > 0, 0 \leq \beta \leq n)$$

(proven in appendix B), we see that it is the  $i$ -th component of the equation

$$\begin{aligned} PT\binom{X}{\beta} &= \frac{\binom{n}{\beta}}{\binom{2n}{\beta}} T\binom{2X}{\beta} \\ &= \frac{\binom{n}{\beta}}{\binom{2n}{\beta}} \sum_{\alpha=0}^{\beta} b_{\beta\alpha} T\binom{X}{\alpha} \end{aligned} \tag{4.10}$$

where the second equality follows from the fact that  $\binom{2X}{\beta}$  is a polynomial of degree  $\beta$ , and so must itself be a linear combination of  $\binom{X}{\alpha}$  with  $\alpha \leq \beta$ .

From equation (4.10), we can see that  $\{T\binom{X}{0}, T\binom{X}{1}, \dots, T\binom{X}{n}\}$  is the basis we are looking for: the linear transformation represented by  $P$  in the standard basis is represented in this basis by an upper triangular matrix with diagonal elements  $b_{ii}\binom{n}{i}/\binom{2n}{i}$ . Since the eigenvalues are a property of the linear transformation, not its matrix representation in a particular basis, the eigenvalues  $\chi_0, \dots, \chi_n$  of  $P$  are exactly these diagonal elements. To calculate  $b_{ii}$ , we note that the coefficient of  $X^i$  in the polynomial  $\binom{X}{i}$  is  $1/i!$ , while in the polynomial  $\binom{2X}{i}$  it is  $2^i/i!$ . None of the  $\binom{X}{\alpha}$  for  $\alpha < i$  can contribute an  $X^i$  term, so it must be the case that  $b_{ii} = 2^i$ . Thus we have that

$$\chi_i = 2^i \frac{\binom{n}{i}}{\binom{2n}{i}}.$$

The ratio  $\chi_{i+1}/\chi_i$  is less than or equal to 1, showing the sequence of eigenvalues is nonincreasing. The first three values are  $\chi_0 = 1$ ,  $\chi_1 = 1$ , and  $\chi_2 = \frac{2(n-1)}{2n-1}$ .

Finally we get to the eigenvalues of  $M$  itself. In general, of course, it is not the case that the eigenvalues of a product of matrices are the products of the eigenvalues of the factors. However, if a pair of eigenvalues of the two factors share a common eigenspace, then their product is an eigenvalue of the product. This condition is not true for all the pairs of eigenvalues of  $R$  and  $P$ , but it *is* true for  $\rho_{\text{hom}}$  (an eigenvalue of  $R$ ) and 1 (an eigenvalue of  $P$ ). Recall that the  $\rho_{\text{hom}}$ -eigenspace of  $R$  consists of the “pure homozygote” vectors: these are also exactly the *left* eigenvectors of  $P$  corresponding to the eigenvalue 1 (this can be seen from the fact that the first and last rows of  $P$  have a one on the diagonal and zeros everywhere else). Therefore  $2\rho_{\text{hom}} = 2\frac{1+s_{\text{min}}}{2+s_{\text{min}}}$  is an eigenvalue of  $M = 2RP$  (with multiplicity 2). But since  $s_{\text{min}} < 0$ , this eigenvalue is always less than 1, and we can ignore it: either it is not the dominant eigenvalue, in which case our conclusion does not depend on it, or it is the dominant eigenvalue, in which case the next largest eigenvalue will also be less than 1, and our conclusion will be the same as if that were the largest eigenvalue.

We now turn to the remaining eigenvalues of  $M$ . Let  $\chi \neq 1$  be another eigenvalue of  $P$  and  $x$  a corresponding *left* eigenvector. We now show that  $2\rho_{\text{het}}\chi$  is an eigenvalue of  $M$ . Unfortunately, it will not be so easy this time:  $x$  does not have to lie in the  $\rho_{\text{het}}$ -eigenspace of  $R$ . However, we can split  $x$  into a pure homozygote component  $x_{\text{hom}}$ , which is the component of  $x$  in the  $\rho_{\text{hom}}$ -eigenspace of  $R$ , and a pure heterozygote component  $x - x_{\text{hom}}$ , which lies in the  $\rho_{\text{het}}$ -eigenspace. Of these three vectors,  $x$  is a (left) eigenvector of  $P$  (by definition), and  $x - x_{\text{hom}}$  is an eigenvector of  $R$  (by construction), but  $x_{\text{hom}}$  is an eigenvector of *both*: this enables us to construct a vector that will be a left eigenvector of  $RP$  corresponding to  $\rho_{\text{het}}\chi$ , namely  $\rho_{\text{het}}(1 - \chi)x_{\text{hom}} + (\rho_{\text{hom}} - \chi\rho_{\text{het}})(x - x_{\text{hom}})$ . To see it is an eigenvector of  $RP$ , note that

$$\begin{aligned} P^T R(\rho_{\text{het}}(1 - \chi)x_{\text{hom}} + (\rho_{\text{hom}} - \chi\rho_{\text{het}})(x - x_{\text{hom}})) \\ &= P^T (\rho_{\text{het}}\rho_{\text{hom}}(1 - \chi)x_{\text{hom}} + (\rho_{\text{hom}}\rho_{\text{het}} - \chi\rho_{\text{het}}^2)(x - x_{\text{hom}})) \\ &= P^T (\rho_{\text{het}}\chi(\rho_{\text{het}} - \rho_{\text{hom}})x_{\text{hom}} + (\rho_{\text{hom}}\rho_{\text{het}} - \chi\rho_{\text{het}}^2)x) \\ &= \rho_{\text{het}}\chi(\rho_{\text{het}} - \rho_{\text{hom}})x_{\text{hom}} + (\rho_{\text{hom}}\rho_{\text{het}}\chi - \chi^2\rho_{\text{het}}^2)x \\ &= \rho_{\text{het}}\chi((\rho_{\text{het}} - \rho_{\text{hom}})x_{\text{hom}} + (\rho_{\text{hom}} - \chi\rho_{\text{het}})x) \\ &= \rho_{\text{het}}\chi(\rho_{\text{het}}(1 - \chi)x_{\text{hom}} + (\rho_{\text{hom}} - \chi\rho_{\text{het}})(x - x_{\text{hom}})), \end{aligned}$$

where the first equality uses that  $x_{\text{hom}}$  and  $x - x_{\text{hom}}$  are eigenvectors of  $R$  and the third uses that  $x$  and  $x_{\text{hom}}$  are left eigenvectors of  $P$ . From the eigenvalues  $\chi_2, \chi_3, \dots, \chi_n$  of  $P$  we obtain in this way the remaining  $n - 1$  eigenvalues of  $M = 2RP$ , and have thus found all eigenvalues of  $M$ .

Since  $\chi_2 = \frac{2(n-1)}{2n-1}$  is the largest of the remaining eigenvalues of  $P$ , to check if there is at least one eigenvalue of  $M$  greater than 1 we need only check if  $2\rho_{\text{het}}\chi_2 > 1$ . We have that

$$2\rho_{\text{het}}\chi_2 = 2\frac{(1+s_{\max})2(n-1)}{(2+s_{\max})(2n-1)},$$

which is greater than one if and only if  $s_{\max} > \frac{2}{2n-3}$ .  $\square$

**Theorem 2.** *In the model with random replication and a dominant fitness function, the establishment probability (and therefore the rescue probability) is non-zero if and only if  $n \geq 2$  and*

$$s_{\max} > \frac{4n}{2n^2 - 3n - 1}.$$

*In addition, in the long term limit, the heterozygote types are all equally abundant.*

*Proof.* Let  $M$  once again be the matrix the  $i, j$  element of which is the expected number of daughter cells with  $j$  mutant plasmids from a cell with  $i$  mutant plasmids. In the model with random replication and a dominant fitness function, the elements of this matrix have the form

$$\begin{aligned} m_{ij} &= \frac{\lambda_i}{\lambda_i + \mu_i} \sum_{k=0}^n 2^{\delta_{jk}} P(i \rightarrow \{j, k\}) \\ &= 2 \frac{\lambda_i}{\lambda_i + \mu_i} \sum_{k=0}^n \frac{\binom{j+k}{j} \binom{2n-j-k}{n-j}}{\binom{2n}{n}} \frac{\binom{2n-j-k-1}{n-j-k+i} \binom{j+k-1}{j+k-i}}{\binom{2n-1}{n}}, \end{aligned}$$

where we have substituted the definition of  $P(i \rightarrow \{j, k\})$  for random replication from equation (4.2). As in the proof for the regular replication case, we use general branching process theory (Sewastjanow 1975, Satz V.2.5) to determine that there is a nonzero probability of establishment if and only if  $M$  has an eigenvalue greater than one.

The elements of the first and last rows of  $M$ , the numbers  $m_{0j}$  and  $m_{nj}$ , must all be zero except the diagonal elements  $m_{00} = m_{nn} = 2\rho_{\text{hom}}$ . This can be seen either from biological considerations—a homozygote cell will have only homozygote daughters—or by substitution into the definition of  $m_{ij}$ —if  $i = 0$ , then  $\binom{j+k-1}{j+k-i}$  will be zero unless  $j = k = 0$ , in which case it is  $\binom{-1}{0} = 1$ , and if  $i = n$ , then the same is true of  $\binom{2n-j-k-1}{n-j-k+i}$ . This means  $M$  has the structure

$$\begin{bmatrix} 2\rho_{\text{hom}} & 0 & 0 \\ 2\rho_{\text{het}A} & 2\rho_{\text{het}B} & 2\rho_{\text{het}C} \\ 0 & 0 & 2\rho_{\text{hom}} \end{bmatrix},$$

where  $B$  is an  $(n-1) \times (n-1)$  matrix and  $A$  and  $C$  are  $n-1$  component column vectors. Then  $2\rho_{\text{hom}}$  is an eigenvalue of  $M$ ; this eigenvalue, however, is always less than one, since  $s_{\text{min}} < 0$ . Every other eigenvalue of  $M$  will be  $2\rho_{\text{het}}$  times an eigenvalue of  $B$ , and its corresponding left eigenvectors are of the form  $[x_0 \ v \ x_1]$ , where  $v$  is a left eigenvector of  $B$ .

We now show that  $(n-1)(2n+1)/(n+1)(2n-1)$  is such an eigenvalue of  $B$  with a corresponding left eigenvector  $v$  with all components equal to 1 (the identification of this eigenvalue is due to Novick and Hoppensteadt 1978). The  $j$ th element of  $vB$  is

$$\sum_{i=1}^{n-1} \sum_{k=0}^n \frac{\binom{j+k}{j} \binom{2n-j-k}{n-j} \binom{2n-j-k-1}{n-j-k+i} \binom{j+k-1}{j+k-i}}{\binom{2n}{n} \binom{2n-1}{n}},$$

which by equation (B.2) in appendix B is equal to  $(n-1)(2n+1)/(n+1)(2n-1)$ ; so  $vB = (n-1)(2n+1)/(n+1)(2n-1)v$ . This must also be the largest eigenvalue of  $B$ :  $B$  is a nonnegative irreducible matrix (since a cell with  $1 \leq i \leq n-1$  mutant plasmids can produce a descendant cell with  $1 \leq k \leq n-1$  mutant plasmids after at most  $|i-k|$  generations,  $B^{n-1}$  is a positive matrix), so by the Perron-Frobenius Theorem (Frobenius 1912; Sewastjanow 1975, Satz IV.5.2) its only eigenvalue with a strictly positive eigenvector is the largest eigenvalue.

Thus  $2\rho_{\text{het}}(n-1)(2n+1)/(n+1)(2n-1)$  is the largest eigenvalue of  $M$  (other than possibly  $2\rho_{\text{hom}}$ ) and its corresponding left eigenvector has all heterozygote types equally abundant. By simple rearrangement, we have that

$$2\rho_{\text{het}} \frac{(n-1)(2n+1)}{(n+1)(2n-1)} = 2 \frac{(1+s_{\text{max}})(n-1)(2n+1)}{(2+s_{\text{max}})(n+1)(2n-1)}$$

is greater than one if and only if  $s_{\text{max}} > \frac{4n}{2n^2-3n-1}$ .  $\square$

Although we have given the conditions in the statements of the theorems as thresholds on  $s_{\text{max}}$ , they can be equivalently restated as thresholds on  $n$  necessary for establishment: for regular replication, the threshold is  $n > \frac{1}{s_{\text{max}}} + \frac{3}{2}$ , and for random replication it is  $n > \frac{1}{s_{\text{max}}} + \frac{3}{4} + \sqrt{\left(\frac{1}{s_{\text{max}}} + \frac{3}{4}\right)^2 + 1/2}$ . Note that as a direct consequence of these two theorems, if establishment is impossible in the regular case for a given  $n$  and  $s_{\text{max}}$ , then is also impossible in the random case (the threshold  $s_{\text{max}}$  is always larger in the regular case, since  $n-3/2 > \frac{n-3/2}{2} - \frac{1}{4n}$ ). We have derived the threshold condition only for the dominant fitness function: we relied on having one common reproduction probability  $\rho_{\text{het}}$  for all heterozygote types, which made the interaction between host reproduction and the combinatorics of plasmid replication and segregation particularly simple.

With the insights obtained in this section, we can now also determine the exponential growth rate of the population, once rescued. For this, we switch back to the birth-death process in continuous time. At long enough

times, growth is well described deterministically by a system of ODEs. This system will be linear, with a matrix of coefficients given by  $\Lambda(2P - I) - D$ , where  $P$  has the same meaning as in the proof of Theorem 1, the combinatorial part of the expected progeny matrix,  $\Lambda$  is a diagonal matrix of birth rates  $\lambda_i$ ,  $I$  is the identity matrix, and  $D$  is the diagonal matrix of death rates, which by assumption are the same for all cell types and given by  $\mu_i = 1$ . By the same argument as in the proof of Theorem 1, the eigenvalues of  $\Lambda(2P - I)$  are products of an eigenvalue of  $\Lambda$  and an eigenvalue of  $2P - I$  (the arguments in the proof of Theorem 1 also apply to random replication). The largest eigenvalue of  $\Lambda(2P - I) - D$ , which determines the growth rate of the rescued population, is therefore  $(1 + s_{\max})(2\chi - 1) - 1$ , where  $\chi$  is the largest eigenvalue of  $P$  other than 1: this is  $\frac{2(n-1)}{2n-1}$  for regular replication and  $\frac{(n-1)(2n+1)}{(n+1)(2n-1)}$  for random replication, where the latter result is obtained with the same reasoning as applied to the matrix  $M$  in the proof of Theorem 2.

#### 4.6 The final distribution of cell types

If establishment occurs, we can ask what the relative abundances of cells with different numbers of mutant plasmids will be. The distribution of the number of mutant plasmids in a cell in the infinite time limit can be found for the discrete-time branching process from the expected progeny matrix  $M$ , with the expected number of daughters with  $j$  mutant plasmids from a cell with  $i$  mutant plasmids as its  $i, j$  element. If establishment is possible ( $P_{\text{est}} > 0$ ) then  $M$  has a dominant eigenvalue, and a corresponding left eigenvector to that eigenvalue gives the distribution of cell types in the long run (Mode 1971, Theorem 4.1). This eigenvector can be found numerically.

Examples of the final stable distribution of the number of mutant plasmids per cell under the four models considered (regular and random replication and dominant and peaked fitness functions) are shown in Figure 4.6. Even after establishment, considerable numbers of homozygote cells are present in the population, because of their constant replenishment by segregation. Heterozygotes are more abundant under regular replication, because there is a lower rate of loss of heterozygosity to segregation than with random replication. Note that in the model with random replication and dominant fitness function, all heterozygote cell types are equally abundant in the long-term limit (as shown by Theorem 2).

#### 4.7 Cointegration

As we have seen, segregation plays an important role in limiting the possibility of the population being rescued. One process which alters the rate of segregation is the fusion of plasmids to form cointegrates. When plasmids

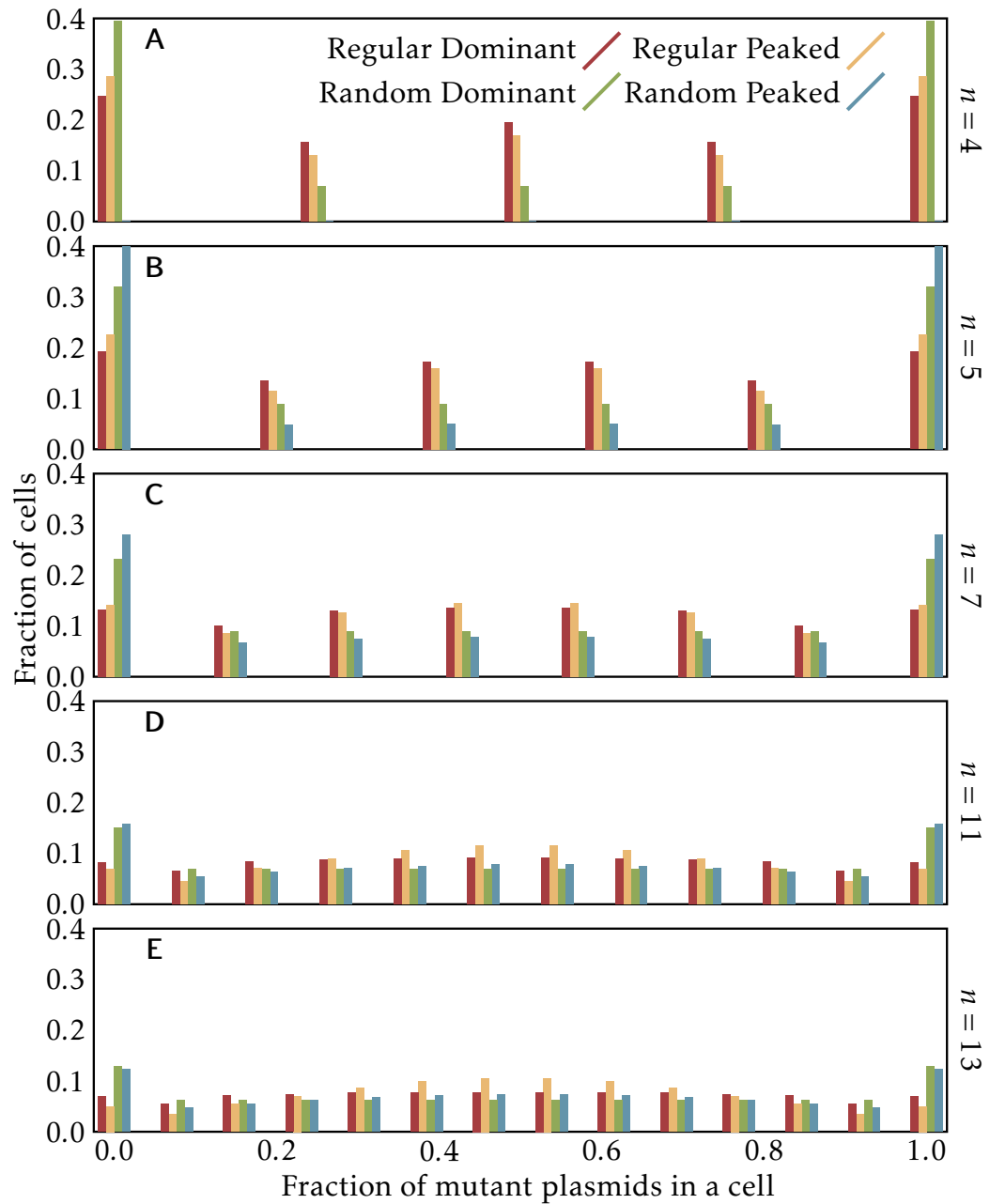


Figure 4.6: The final distribution of the proportion of mutant plasmids in a cell for plasmids of given copy numbers, under each of the four combinations of replication assumption and fitness function considered, and with  $s_{\min} = -0.1$  and  $s_{\max} = 1.0$ . The eigenvector corresponding to the leading eigenvalue of  $M$  was calculated numerically (code available as supplementary material); note that for  $n = 4$  with random replication and the peaked fitness function, the population goes extinct with probability one, and there is no well-defined final distribution.

fuse, the resulting multimer has multiple copies of the plasmid backbone, including multiple copies of the copy number control system, so we expect that an  $m$ -mer still “counts against” the copy number  $m$  times, even though it is only one molecule (Chiang and Bremer 1991; Summers and Sherratt 1984; Summers et al. 1993). Cointegration can eliminate the possibility of losing heterozygosity to segregation in the descendants of cells in which it occurs (experimentally observed by Hülter et al. 2020; Garoña et al. 2021): if plasmids with distinct alleles fuse, the resulting cointegrate provides the heterozygote phenotype on its own and can no longer be separated by segregation (unless the cointegrate is resolved into independent plasmids again). But conversely, if plasmids with identical alleles fuse, the number of independently segregating units in the cell is reduced, and the rate of loss will increase. We now present an extension of the model to include cointegration. We have chosen to use a very simplified cointegration model, with no resolution of multimers, only regular replication and the dominant fitness function, and novel mutations only appearing in cells with all monomers, to provide a comparison to the model without cointegration without a great deal of complexity.

#### 4.7.1 *The cointegration model*

In a model which incorporates cointegration, there are a much larger number of possible cell types, since not only are there multiple loci that may have one allele or the other, but also these loci may be distributed among plasmids in a variety of ways. In the following, a cell type will be written in the form (AA) (B) (AB), where each letter represents a locus carrying one of the two alleles, A and B, that must be maintained for the population to survive, loci within a pair of parentheses are on the same plasmid, and the total number of loci is fixed at the copy number  $n$ . In the example, the cell thus contains a cointegrate of two plasmids with allele A, a plasmid with allele B, and a cointegrate of a plasmid with allele A and a plasmid with allele B. Each cell type  $a$  then has a death rate  $\mu_a$ , which will be fixed to 1 as before, and a reproduction rate  $\lambda_a$ , which is determined by the number of mutant and wild-type alleles in the cell (irrespective of their distribution on different plasmids). Incorporating cointegration into our model requires slight changes to the reproduction probabilities to account for the greater number of cell types, and adds a new process of cointegration which needs to be modelled.

#### Reproduction

The reproduction probabilities are determined by the regular replication model used previously. The probability that a dividing cell of type  $a$  has daughters of types  $b$  and  $c$  will be denoted  $P(a \rightarrow \{b, c\})$ . All plasmids are duplicated and then divided into two subsets having the same copy number

of the copy number control system. This last part is important because now plasmids contribute differently to the total copy number. This means that some assortments of plasmids into daughter cells are no longer possible: for example, an (AAA)(B) cell can only have daughters identical to itself, since a (B)(B) cell has too few and an (AAA)(AAA) cell too many copies of the copy number control system. The probabilities  $P(\bar{a} \rightarrow \{b, c\})$  are calculated by enumeration of the possible pairs of daughter cells and the number of ways to produce each one.

### Cointegration

The cointegration rate is divided into two parts: the overall probability of a cointegration event occurring in a given cell, which is the biological part, and the probability, conditional on cointegration occurring, that two particular plasmids fuse, which is purely combinatorial.

The probability, conditional on a fusion occurring, that it takes the cell from type  $\bar{a}$  to type  $\bar{a}'$  is denoted  $P(\bar{a} \Rightarrow \bar{a}')$ . This probability is the fraction of all pairs of plasmids in the  $\bar{a}$ -cell that are pairs of the two types that need to fuse for this transition to occur. For example,  $P((A)(A)(B) \Rightarrow (A)(AB)) = \frac{2}{3}$  (there are 3 pairs of plasmids in the original cell, of which two are  $\{(A), (B)\}$  pairs). We only allow for fusion of plasmids but not for the resolution of cointegrates, therefore  $P((AA)(A)(B) \Rightarrow (A)(A)(AB)) = 0$ .

The probability of cointegration occurring in one generation in cells of type  $\bar{a}$  is denoted  $\kappa_{\bar{a}}$ . In the model presented here, this will be simply a constant  $\kappa$ , except in cells that have only a single large multimer, where it is 0 because no further cointegration can take place. We assume that at most one cointegration event occurs in each cell cycle.

#### 4.7.2 Analysis and results

Combining the reproduction and resolution models, the extinction probability  $Q_{\bar{a}}$  of type  $\bar{a}$  is given by

$$Q_{\bar{a}} = \frac{\mu_{\bar{a}}}{\lambda_{\bar{a}} + \mu_{\bar{a}}} + \frac{\lambda_{\bar{a}}}{\lambda_{\bar{a}} + \mu_{\bar{a}}} \sum_{\{b, c\}} \left( \kappa_{\bar{a}} \sum_{\bar{a}'} P(\bar{a} \Rightarrow \bar{a}') P(\bar{a}' \rightarrow \{b, c\}) Q_b Q_c + (1 - \kappa_{\bar{a}}) P(\bar{a} \rightarrow \{b, c\}) Q_b Q_c \right),$$

where the sums are taken over all types or unordered pairs of types. Note that this expression is equally compatible with selection happening before or after the fusion of plasmid copies, since the values of  $\lambda_{\bar{a}}$  and  $\mu_{\bar{a}}$  are determined only by the alleles present and not their distribution between plasmid copies. Since we are using the regular replication model and make the assumption that the mutation first appears in a cell with only monomers

present, the establishment probability is then given by

$$P_{\text{est}} = 1 - Q_{(A)^{n-1}(B)}.$$

Establishment probabilities for the cointegrate model are shown in Figure 4.7. An important qualitative difference is that for non-zero  $\kappa$ , the establishment probability no longer exhibits a threshold effect: establishment is theoretically possible for any value of  $s_{\text{max}}$ .

## 4.8 Discussion

Multicopy plasmids are a frequent component of bacterial genomes, where they may be cryptic passengers or play an important role in bacterial adaptation and evolution. One way in which they can play such a role is through their inherently polyploid nature: loci on a multicopy plasmid can exhibit heterozygosity which is not usually possible for loci on the (for many species) haploid chromosome. Here we have examined one scenario in which heterozygosity of plasmids may provide an adaptive advantage to their hosts, by enabling rapid evolution in scenarios of heterozygote advantage.

It is apparent that a key factor in determining the fate of heterozygosity on plasmids is plasmid segregation at host cell division. The effects of segregation on the establishment of novel alleles located on multicopy plasmids have already been explored both in models (Santer and Uecker 2020; Garoña et al. 2023) and experiments (Ilhan et al. 2018; Garoña et al. 2023), showing that segregation reduces the establishment probability of adaptive alleles on multicopy plasmids. Here, we have shown that the effect of segregation is even more drastic in scenarios of heterozygote advantage: the constant loss of heterozygosity to segregation puts strong limits on the probability of successful adaptation unless the fitness of heterozygotes is sufficiently large. The condition that we derived for the required fitness of heterozygotes more generally holds for the maintenance of heterozygosity on multicopy plasmids, even outside a rescue scenario. While we have phrased our model in terms of wild-type and mutant plasmids (i.e. variants of the same plasmid), the results carry over to the establishment and the maintenance of two incompatible plasmids variants sharing the same replication mechanism.

The importance of plasmid-mediated heterozygosity in allowing populations to escape from fitness trade-offs was previously explored in both experiments and models by Rodriguez-Beltran et al. (2018). In a system where different plasmid alleles provide resistance to different antibiotics, they find that fluctuating selection is capable of maintaining heterozygosity, and intermediate antibiotic concentrations maintained heterozygosity the longest (note that unlike in our model, there is always selection for only one plasmid variant at a time). Their model is simpler than ours in

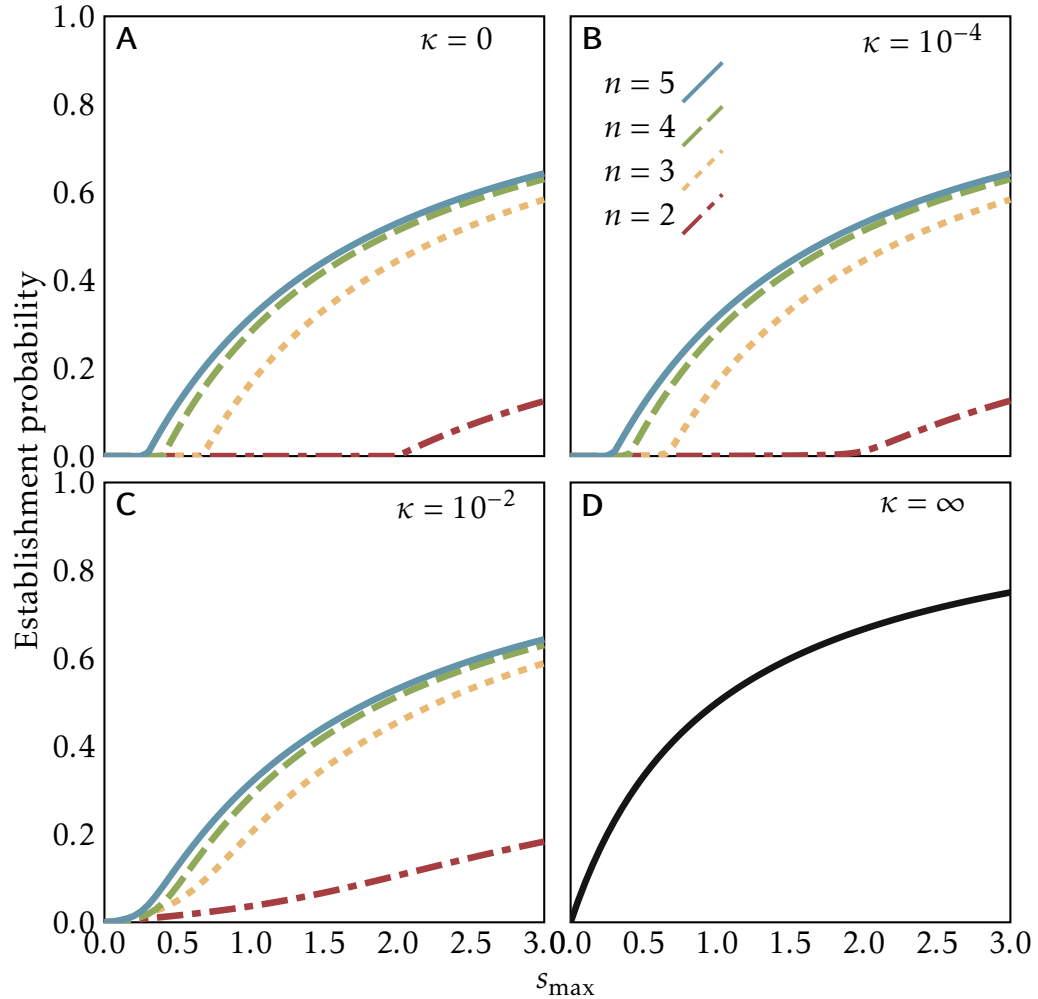


Figure 4.7: Establishment probabilities of rescue mutations on possibly-cointegrating plasmids of a given copy number, for varying values of the heterozygote fitness  $s_{\max}$  and the cointegration probability  $\kappa$  (for all cases,  $s_{\min} = -0.1$ ). The  $\kappa = \infty$  panel corresponds to a model in which the mutation is immediately followed by the fusion of all plasmids into a single multimer, and thus is the theoretical maximum establishment probability. The code for numerical calculation of establishment probabilities is available as supplementary material.

one important respect: they bundle all heterozygotes into a single compartment of their ODE model, ignoring the precise number of mutant and wild-type plasmids. To estimate the rate of segregation, Rodriguez-Beltran et al. (2018) assume that all heterozygote cells have half-and-half mutant and wild-type composition, so that the probability of having a homozygote daughter cell is  $2^{1-n}$ . This means that the expected number of heterozygote daughters of a heterozygote cell is  $2(1 - 2^{1-n})(1 + s_{\max})/(2 + s_{\max})$ , and therefore the branching process founded by a heterozygote cell is supercritical, and the establishment probability of a mutant on a plasmid is positive, if and only if

$$s_{\max} > \frac{1}{2^{n-2} - 1},$$

which is a much weaker condition than our Theorem 1 (provided that  $n > 3$ ): ignoring the unbalanced heterozygotes overestimates heterozygote stability. The extension of our model to fluctuating selection remains for future work.

The important role of segregation on multicopy plasmids strongly contrasts with the state of affairs on a haploid bacterial chromosome, where true heterozygosity is impossible, so that our two focal alleles would have to either replace each other completely or be located at different loci and not subject to segregation. Cells in rapidly dividing populations will contain multiple copies of the chromosome and thus be effectively polyploid; although this exposes them to the effects of dominance (Sun et al. 2018) their segregation is not random and heterozygosity cannot be maintained long-term. There is also a strong contrast to the role of segregation in the system one normally thinks of when considering heterozygosity, sexually reproducing diploid organisms. Homologous chromosomes in diploids segregate at meiosis, but heterozygosity is maintained by sexual reproduction; unlike with multicopy plasmids, no selection for heterozygotes is required to maintain heterozygosity at Hardy-Weinberg proportions in the population indefinitely, provided homozygous individuals can thrive and both homozygous types are equally fit. However, if homozygous individuals have fitness less than one as in our rescue scenario and heterozygotes are required for population persistence, there is also a threshold on the heterozygote fitness below which the population cannot survive; unlike for multicopy plasmids, this threshold depends on the fitness of homozygotes. The closest analogue in diploids to the role of segregation in multicopy plasmids is inbreeding, which also reduces heterozygosity in the population without altering allele frequencies. Segregation is in a sense a stronger force in reducing heterozygosity than any degree of partial selfing, since the equilibrium heterozygosity is zero in the absence of selection for the heterozygote (unlike partial selfing, which produces a nonzero equilibrium), but less strong than complete selfing; this can be seen by comparing the  $n = 2$  case (when segregation has the largest effect on multicopy plasmids, and

also the case most directly comparable to diploids), where 2/3 of the daughter cells of heterozygotes are heterozygotes, to a selfing diploid, where only 1/2 of the offspring of heterozygotes are heterozygotes. Nevertheless, constraints on heterozygote fitness are stronger for plasmid-mediated rescue of bacterial populations than for rescue of selfing populations: division of a heterozygote cell into two homozygote cells leads to loss of a heterozygote from the population, while offspring reproduction by a selfing heterozygote leaves the heterozygote parent individual intact. This means that the rate of reproduction  $1 + s_{\max}$  only needs to be greater than two for the selfing individual (or for a random mating population in which homozygotes are lethal or infertile) rather than three as in the bacterial population. The nonplasmid systems that most closely resemble the genetics of multicopy plasmids are the organelle (mitochondrial and plastid) genomes of eukaryotes (Birky 1983) and the macronuclei of ciliate protists (Allen and Nanney 1958)—indeed, it was in the latter context that Schensted (1958) developed a model of chromosomal segregation which is equivalent our regular replication model for plasmids and first determined the eigenvalues of the matrix we refer to as  $P$  in the proof of Theorem 1.

We have also seen that the mechanism or mode of replication has a substantial effect on the fate of a novel mutation in heterozygote advantage scenarios. In the regular replication model, there are no additional stochastic effects in replication, while in the random replication model the rich-get-richer property of the replication process is an additional mechanism reducing heterozygosity in the population. Our understanding of the biological mechanisms of plasmid replication would seem to suggest that random replication is a more realistic model of the replication process of most plasmids; however, some empirical studies have found the regular replication model to be a better fit to the data (Garoña et al. 2023). Here we have followed Novick and Hoppensteadt (1978) in presenting both models to permit a comparison.

We have kept properties of the plasmid such as copy number fixed over the course of rescue. While this may be a realistic assumption over the short timescale of the rescue process, over the longer term (for example in the rescued population) these properties are also subject to evolution. Given the important role of copy number in the fate of bacterial populations depending on plasmid-mediated heterozygosity, the copy number might well be under selection. Some experimental studies have shown rapid evolution of plasmid copy number in response to selection on plasmid-borne traits in other contexts (Dimitriu et al. 2021; San Millan et al. 2015; San Millan et al. 2016). A mutation in the plasmid replication control system that produces a new plasmid type compatible with the existing plasmid—a plasmid speciation event—would also confer a selective advantage if it allowed the mutant and wild-type alleles to exist on plasmids of distinct compatible types, no longer sharing a common copy number between them, and no longer

subject to loss of heterozygosity by segregation. Rapid plasmid speciation has also been observed in experiments (Santos-Lopez et al. 2016).

Another change in the plasmid that would affect the maintenance of plasmid-mediated heterozygosity, which we have modelled, is the formation of plasmid cointegrates (Hülter et al. 2020; Garoña et al. 2021). Our results showed that stable multimerization of plasmids can increase the probability of successful establishment of the novel allele. Our model of cointegration is deliberately simple and limited, as it is intended as a short demonstration of the importance of cointegration in the heterozygote advantage scenario and not a deep exploration of the cointegration process. Many aspects of the cointegration process not explored here, particularly the precise behaviour of multimers during replication and segregation and the resolution of cointegrates into smaller multimers or monomers, are no doubt important to the fate of alleles on cointegrates and merit further investigation. We especially have not included consequences of multimerization on the stability of plasmid inheritance, which has been examined in a model by Summers et al. (1993). The possibility of evolution of a multidrug resistance plasmid from two incompatible single-resistance plasmids has been previously demonstrated by Condit and Levin (1990) both experimentally and in a model; it is left open in their model whether the plasmid carrying both resistance genes has been formed through plasmid fusion or exchange of DNA between plasmids, and they do not account for the plasmid copy number. A detailed model that combines elements of their and our models and allows for dependence of recombination on the plasmid copy number or the ratio of variant frequencies could be interesting to further understand the evolution of multidrug resistance on plasmids.

Broadly speaking, whether rescue occurs or not depends on the availability of rescue mutations and their probability to escape stochastic loss while rare, which is reflected in equation (4.6) by the product of the mutational input and the establishment probability. In our model, all rescue mutations are assumed to appear after the population decline begins. However, if the mutation were to occur before the environmental change that causes the population decline, then there could already be heterozygotes present in the population at the beginning of the evolutionary rescue process. Their contribution relative to that of *de novo* mutations strongly depends on the fitness effect of the mutation prior to the environmental change and the rate of decay of the wild-type population (Orr and Unckless 2014). Rescue from the standing genetic variation has been explored in many models of evolutionary rescue (e.g. Orr and Unckless 2008), including rescue on plasmids (Santer and Uecker 2020). Plasmids also create avenues for the arrival of rescue mutations through horizontal gene transfer: a mutant plasmid could be introduced into the host population by transformation, or by conjugation (where heterozygosity can only be generated if transfer of the mutant plasmid is not excluded by entry or surface exclu-

sion from the wild-type plasmid variant). While horizontal transfer from a source population to the target population increases the availability of the rescue mutation, horizontal transfer within the target population affects its establishment probability. There has been work on modelling rescue from genes on plasmids arriving and spreading by horizontal transfer, splitting the problem up into the rate of appearance and the establishment probability of the (single-copy) rescue plasmid analogous to our equation (4.6) (Tazzyman and Bonhoeffer 2014; Geoffroy and Uecker 2023).

Plasmids are pervasive in natural populations of bacteria, and play an important role in adaptation and evolution of bacteria. Understanding bacterial evolution therefore involves a good understanding of the population genetics of plasmids. The properties of plasmids as independently replicating units within their hosts make them an intriguing genetic system, with complexities not present for haploid or diploid chromosomes. We have explored one aspect of plasmid genetics—heterozygosity on multi-copy plasmids—in the context of evolutionary rescue. We have shown that, as intuitively expected, the maintenance of heterozygosity and thus rescue is impossible below a threshold copy number for a given heterozygote fitness. Going beyond this intuition, our concise criterion in terms of the plasmid copy number and the fitness of heterozygous cells quantifies the conditions for the persistence of plasmid-mediated heterozygosity through heterozygote advantage, and thus contributes to a population genetic theory of bacterial evolution.

## RAPID ADAPTATION THROUGH GENE AMPLIFICATION ON ECCDNA IN YEAST

**Abstract.** *Small circular DNA molecules found in the nuclei of eukaryotic cells, extrachromosomal circular DNA, exist in multiple copies. Since they consist of copies of chromosomal DNA, this means that eccDNA can amplify chromosomal genes. In circumstances where gene amplification has important fitness benefits, this can affect population persistence, and even rescue populations from extinction. We develop and analyze stochastic process models of the amplification of genes on eccDNA in yeast, and the contribution of eccDNA to rescuing yeast populations. We show that the abundance of eccDNA is driven not only by intrinsic eccDNA processes like replication and production of eccDNA copies, but also by the replication of hosts; in particular, the between-cell variation in eccDNA copy number is affected by the degree of bias in the segregation of eccDNA copies at host cell division. This in turn affects the probability of successful rescue of a yeast population by gene amplification on eccDNA. Our results show the unique role of extrachromosomal genetic elements in the population genetics of rapid adaptation.*

### 5.1 Introduction

The nuclei of many eukaryotes contain, in addition to the chromosomes, extrachromosomal circular DNA (eccDNA), circular DNA molecules of various sizes that are distinct and independent from the chromosome, but which consist of copies of regions of chromosomal DNA (Gaubatz 1990; Hotta and Bassel 1965). Extrachromosomal circular DNA is often associated with cancers, where it has been extensively studied, and is particularly common, perhaps because of the general fragmentation and polyploidy of the genetic material common in cancer cells (Cox et al. 1965; Windle et al. 1991; Curtis et al. 2024). But eccDNA is not (at least exclusively) a pathological symptom of cancer: eccDNA is found in the nuclei of normal, noncancerous cells from all eukaryotic kingdoms (Beverley et al. 1984; Kinoshita et al. 1985; Cohen et al. 2003; Møller et al. 2015; Shoura et al. 2017; Møller et al. 2018), and therefore appears to be a normal component of eukaryote genomes.

Since it is both extrachromosomal and circular, eccDNA superficially resembles the better-known plasmids present in prokaryotes (and some fungi): eccDNA differs, however, in that it contains only copies of regions of the chromosomal DNA. The eccDNA copies therefore have no genes of their own, and no capacity for independent control or regulation of their

---

This chapter is in preparation for publication.

physiological processes by *cis*-acting genes. This means that the life cycle of an eccDNA is entirely determined by the host biology. Following Arrey et al. (2022), we consider the key processes of eccDNA biology to be production from the chromosome, replication, elimination, and the effects of host cell division. The production of new eccDNA from the chromosome occurs by mechanisms that have not been completely explained, and may vary among host taxa; however, they generally seem to be produced as byproducts of replication or transcription on the chromosome (Paulsen et al. 2018; Hull et al. 2019), and the frequency of eccDNA production is not uniform between different chromosomal loci, but varies, resulting in certain eccDNA hotspots (Arrey et al. 2022). Existing eccDNA copies can be replicated in the host: frequently, the eccDNA contains an origin of replication from the chromosome, and then eccDNA copies can be replicated by the same mechanism as chromosomes (Møller et al. 2015), making their replication synchronized with the host cell cycle (Barker et al. 1980). This is by far the most commonly observed mechanism of replication, but in some cases, eccDNA copies have also been observed to replicate independently of the host cell cycle by rolling circle replication (Cohen et al. 2005). Copies of the eccDNA may also be degraded and destroyed by the host. Finally, when the host cell divides, eccDNA copies must be segregated between the two resulting cells. Since eccDNA copies are generally acentric (lacking a centromere), the host segregation machinery cannot divide them evenly between daughter cells, and we therefore expect in general that eccDNA segregation will be random (confirmed experimentally by Lange et al. 2022). In yeast (*Saccharomyces cerevisiae*), however, there is an additional complication to segregation: in the budding division of yeast, there is a clear mother and daughter cell (Mortimer and Johnston 1959), and the mother can selectively retain eccDNA (Murray and Szostak 1983; Sinclair and Guarente 1997). Accumulation of eccDNA in older yeast cells is therefore common, and it has been proposed both as a measure of, and a possible cause of, ageing in yeast (Hull and Houseley 2020).

Because it exists in multiple copies, eccDNA can be a source of copy number variation of genes; if these copy number changes have important fitness consequences for the cell (for example because production of an important gene product is transcription-limited), amplification by eccDNA can serve an adaptive function. Adaptive amplification of genes on eccDNA has been observed in several distinct experimental systems. In cancer cells, oncogenes are frequently found to be amplified on eccDNA (Turner et al. 2017; Curtis et al. 2024). In normal cells, eccDNA has been found to be associated with gene amplifications responsible for rapid adaptation to environmental change (e.g. Libuda and Winston 2006; Gresham et al. 2010; Møller et al. 2013; Demeke et al. 2015). Adaptation to environmental toxins, such as antibiotics, cancer drugs, or heavy metals, is a particularly common form of adaptation facilitated by eccDNA (e.g. Baskin et al. 1981; Bev-

erley et al. 1984; Ubeda et al. 2014; Hull et al. 2019), perhaps because this is circumstance in which increasing the production of a particular gene product by increasing the rate of transcription can have important effects (Hull and Houseley 2020). In addition to increasing copy number, the presence of genes on eccDNA can also increase their expression by making them more physically accessible for transcription (Wu et al. 2019; Hung et al. 2022). Beyond amplification of particular genes, eccDNA can also alter the regulation of genes on eccDNA or on the chromosome (Zhu et al. 2021; Hung et al. 2022). The flexibility and transient nature of amplification on eccDNA means that it can be a particularly important mechanism of bet-hedging in fluctuating or unpredictable environments (Møller et al. 2013); when environmental change is not fluctuating, eccDNA can also reintegrate into the chromosome, enabling the stabilization of amplification for long-term adaptation (Holt et al. 2023).

There is a small modelling literature on eccDNA, mostly motivated by eccDNA in cancers, and largely focused on predicting the copy numbers of eccDNA. The use of various stochastic process models is a natural fit to this system, as the less-regulated and inherently stochastic nature of many eccDNA processes, especially segregation, means stochasticity plays an important role in eccDNA dynamics. Early models used branching processes to predict copy number (Kimmel and Axelrod 1990; Kimmel et al. 1992); more recently Lange et al. (2022) made similar predictions from simulations of eccDNA segregation. Turner et al. (2017) used simulations of a branching process model to show that amplified genes on eccDNA were able to reach much higher copy numbers than amplification by repeats on a chromosome, and that this also resulted in a much higher degree of between-cell variability in the copy number of the amplified gene.

Here we present models of *evolutionary rescue* of a population of budding yeast by gene amplification on eccDNA. The concept of evolutionary rescue was introduced by Gomulkiewicz and Holt (1995), and it has been the subject of extensive modelling work (e.g Orr and Unckless 2008; Orr and Unckless 2014; Tazzyman and Bonhoeffer 2014; Geoffroy and Uecker 2023; Madgwick et al. 2024). In an evolutionary rescue scenario, a population is exposed to a sudden environmental change to which it is maladapted. In the new unfavourable environment, the population begins a demographic decline which would lead to extinction. An evolutionary change (classically, usually either a novel mutation or selection for an existing allele present in the standing genetic variation) which adapts individuals to the new conditions occurs; if it overcomes demographic stochasticity to produce a new, better adapted subpopulation, the population can be rescued from extinction. In our case, the evolutionary change will be a change in copy number of a gene due to amplification on eccDNA: our modelled population is subjected to environmental conditions under which cells can only survive if they have a sufficient number of copies of a certain

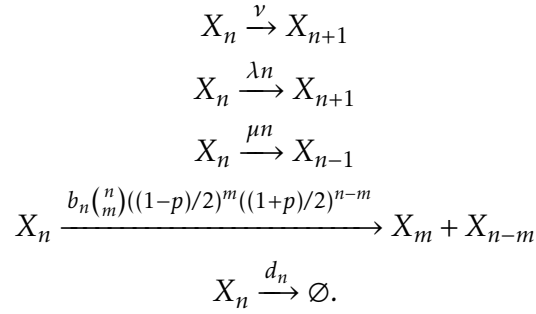
gene, more copies than are present on the chromosome. Accumulation of eccDNA copies carrying the gene therefore provides a mechanism for population persistence. Our motivating example is the rescue of a *S. cerevisiae* population subjected to heavy-metal toxicity from a high environmental concentration of copper by amplification of the *CUP1* copper-resistance (metallothionein) gene, which has been observed in evolution experiments by Hull et al. (2019), but the model is of course more widely applicable. In the following, we organize the presentation of our models on the basis of our assumptions about the replication of eccDNA copies. First, we consider a model where eccDNA copies replicate continuously, without any synchronization with the host cell cycle (as seen in, e.g., Cohen et al. 2005). Second, we consider a model where eccDNA replication occurs by the same mechanism as chromosomal replication, and is therefore synchronized to the host cell cycle (as seen in, e.g., Barker et al. 1980). In both cases, we use analytic methods and stochastic simulations to predict the copy number of the eccDNA and the probability of eccDNA-facilitated rescue of the yeast population.

## 5.2 Asynchronous reproduction

### 5.2.1 The model

We wish to describe a population of yeast cells hosting eccDNA copies with a particular possibly-fitness-relevant gene on them (what we shall call the *focal eccDNA*). In our first model, we assume that eccDNA replication is not synchronized with the host cell cycle (for example because it occurs by rolling-circle replication independent of whether the eccDNA copy has a chromosomal origin of replication): we shall call this assumption *asynchronous reproduction*. We describe the population of eccDNA-carrying yeast cells with a stochastic process, in which host cells are categorized into (countably) infinitely many types  $X_n$  based on the number  $n$  of copies of our particular focal eccDNA they carry. Within each host nucleus, copies of the focal eccDNA are produced from the chromosome at a rate  $\nu$ , they replicate (asynchronously) at a *per capita* rate  $\lambda$ , and they are degraded and lost at a *per capita* rate  $\mu$ . The host cells reproduce at a *per capita* rate  $b_n$  and die at a *per capita* rate  $d_n$ ; both of these rates may in general depend on the number of eccDNA copies. These processes collectively describe a

stochastic process with the elementary reactions



The first three reactions describe the intranuclear dynamics of the eccDNA copies, which constitute a birth-death-immigration process. The latter two reactions describe the population dynamics of the host population, which constitutes of a birth-death process with eccDNA-copy-number-dependent birth and death rates. In the birth reaction, we also need to take into account segregation of eccDNA copies: we assume that segregation is random and independent for each eccDNA copy, which results in the binomial distribution appearing in the rate for the birth reaction. To account for the bias towards retaining eccDNA copies in the mother cell observed in budding yeast, however, the probability of an eccDNA copy ending up in the daughter cell is not  $1/2$ , but  $(1-p)/2$  (meaning that the probability of being retained in the mother cell is  $(1+p)/2$ ), where the parameter  $p$  measures the degree of bias in segregation: at  $p = 0$ , segregation is unbiased (the probabilities are in fact both  $1/2$ ), and the degree of bias increases with  $p$  until at  $p = 1$  all eccDNA copies are retained by the mother cell.

This model is easy to describe, but solving it to find the distribution of numbers of host cells of each type at a particular time is more difficult; it seems quite analytically intractable. Below, we instead present stochastic simulations of the process and analytical results for some special cases.

### 5.2.2 Analysis: neutral conditions

We start by looking at the dynamics under neutral conditions, where the number of eccDNA copies has no fitness effect on host cells. In this case  $b_n = b$  and  $d_n = d$  are constant. The host population dynamics then become uninteresting, both because they are not affected by the eccDNA dynamics, and because they are merely a simple time-homogeneous birth-death process: the yeast population will go extinct with probability  $\min\{1, d/b\}$ , and otherwise increase without bound. We therefore look only at the distribution of eccDNA copy number within cells in the population, by following a single, random cell lineage in the population. Between cell divisions, the eccDNA copy number follows a birth-death-immigration process with rates  $\lambda$ ,  $\mu$ , and  $\nu$ ; at cell division, the eccDNA copies are segregated between mother and daughter cells, and we pick which cell to follow in the

next generation at random. Since the single lineage of cells we follow is chosen randomly, the distribution of eccDNA copy number in our focal lineage is the same as the distribution of eccDNA copy number in the entire population. We derive master equations for the probability  $p_n(t)$  that there are  $n$  eccDNA copies in our cell lineage at time  $t$ : joining all the probabilities together into a probability generating function  $f(x, t) = \sum_{n=0}^{\infty} p_n(t)x^n$ , we obtain a partial differential equation

$$\begin{aligned} \frac{\partial f}{\partial t} = & (\nu(x-1) - b)f + (\lambda x - \mu)(x-1) \frac{\partial f}{\partial x} \\ & + \frac{b}{2} \left( f \left( \frac{(1-p)x + (1+p)}{2}, t \right) + f \left( \frac{(1+p)x + (1-p)}{2}, t \right) \right). \end{aligned}$$

This is still difficult to solve, but we can extract from it a system of ordinary differential equations for the moments of the number  $N_t$  of eccDNA copies at time  $t$ :

$$\begin{aligned} \frac{d}{dt} \mathbf{E}[N_t] &= \nu + (\lambda - \mu - b/2) \mathbf{E}[N_t] \\ \frac{d}{dt} \mathbf{E}[N_t(N_t - 1)] &= 2(\nu + \lambda) \mathbf{E}[N_t] + 2(\lambda - \mu - b(3 - p^2)/8) \mathbf{E}[N_t(N_t - 1)], \end{aligned}$$

and so

$$\begin{aligned} \frac{d}{dt} \mathbf{Var}[N_t] &= \nu + (\lambda + \mu + b(1 - p^2)/4) \mathbf{E}[N_t] + (b(1 + p^2)/4) \mathbf{E}[N_t]^2 \\ &\quad + 2(\lambda - \mu - b(3 - p^2)/8) \mathbf{Var}[N_t]. \end{aligned}$$

From these equations, we can immediately see the long-term behaviour of the mean eccDNA copy number. When

$$\lambda - \mu - b/2 \geq 0,$$

the mean diverges to infinity, meaning that eccDNA copies will accumulate to higher and higher copy numbers indefinitely. Otherwise, the mean converges in the long term to  $\nu/(\mu - \lambda + b/2)$ . Similarly, the variance increases without bound if

$$\lambda - \mu - \frac{3 - p^2}{8} b \geq 0,$$

and otherwise converges to a finite quantity. Since we have that  $\lambda - \mu - b(3 - p^2)/8 > \lambda - \mu - b/2$ , the parameter space divides into three regions: if  $\lambda - \mu - b(3 - p^2)/8 > \lambda - \mu - b/2 \geq 0$ , both mean and variance become increase without bound, if  $0 > \lambda - \mu - b(3 - p^2)/8 > \lambda - \mu - b/2$ , they both converge to finite values, and if  $\lambda - \mu - b(3 - p^2)/8 \geq 0 > \lambda - \mu - b/2$  the long-run mean is finite but the variance still becomes arbitrarily large.

## 5.2.3 Results: copy number distribution

Our analytic results allow us to describe the behaviour of the moments of the eccDNA copy number distribution under neutral conditions. To examine the full distribution, we performed stochastic simulations. The complete stochastic process model described in subsection 5.2.1 was simulated using the Gillespie algorithm (Gillespie 1976). The birth rate of host cells was fixed at  $b = 1$ , allowing us to use the average length of a host generation as a unit of time. To estimate the stable limiting distribution of eccDNA copy number, simulations were run for 20 generations, by which point the copy number distribution had stabilized. Example copy number distributions under neutral conditions are shown in Figure 5.1. They confirm what we found from the analytic results in the previous section: the abundance of eccDNA copies increases with increasing net growth rate of the eccDNA copy population, and increasing the bias of segregation increases the variance in the copy number distribution.

To compare the copy number distribution under selective conditions, we need to choose functional form for the fitness effect of the eccDNA. We will assume that all of the fitness effect of the eccDNA is on the death rate of host cells (a reasonable assumption if, for example, the fitness effect is in reducing the effect of an environmental toxin): the birth rate will remain a constant  $b_n = b = 1$ . The death rate will be given by

$$d_n = d_{\min} + \frac{d_{\max} - d_{\min}}{1 + cn}; \quad (5.1)$$

this decreases hyperbolically from a maximum death rate  $d_{\max} > 1$  at  $n = 0$  towards a lower asymptote  $d_{\min} < 1$  (representing the background death rate from causes not affected by the gene on the focal eccDNA) at a rate determined by  $c$ . When  $n < (d_{\max} - 1)/(1 - d_{\min})/c$ , the death rate is greater than the birth rate  $b = 1$ , and the host cell is more likely to die than to reproduce; when  $n$  is greater than this threshold, the death rate is lower than the birth rate, and the host cell is viable. This choice of fitness function is motivated the example of *CUP1*: producing more of the metallothionein enables binding and sequestering more copper, but with diminishing returns as the concentration of the enzyme becomes larger relative to the concentration of its copper substrate.

Stable copy number distributions were determined by simulation, as in the neutral case. Figure 5.1 shows the copy number distributions under one possible fitness function. Compared to each other, the selective distributions show the same general trends as under neutral conditions. Compared to the corresponding distributions under neutral conditions, the typical copy number of the eccDNA increases, although only slightly.

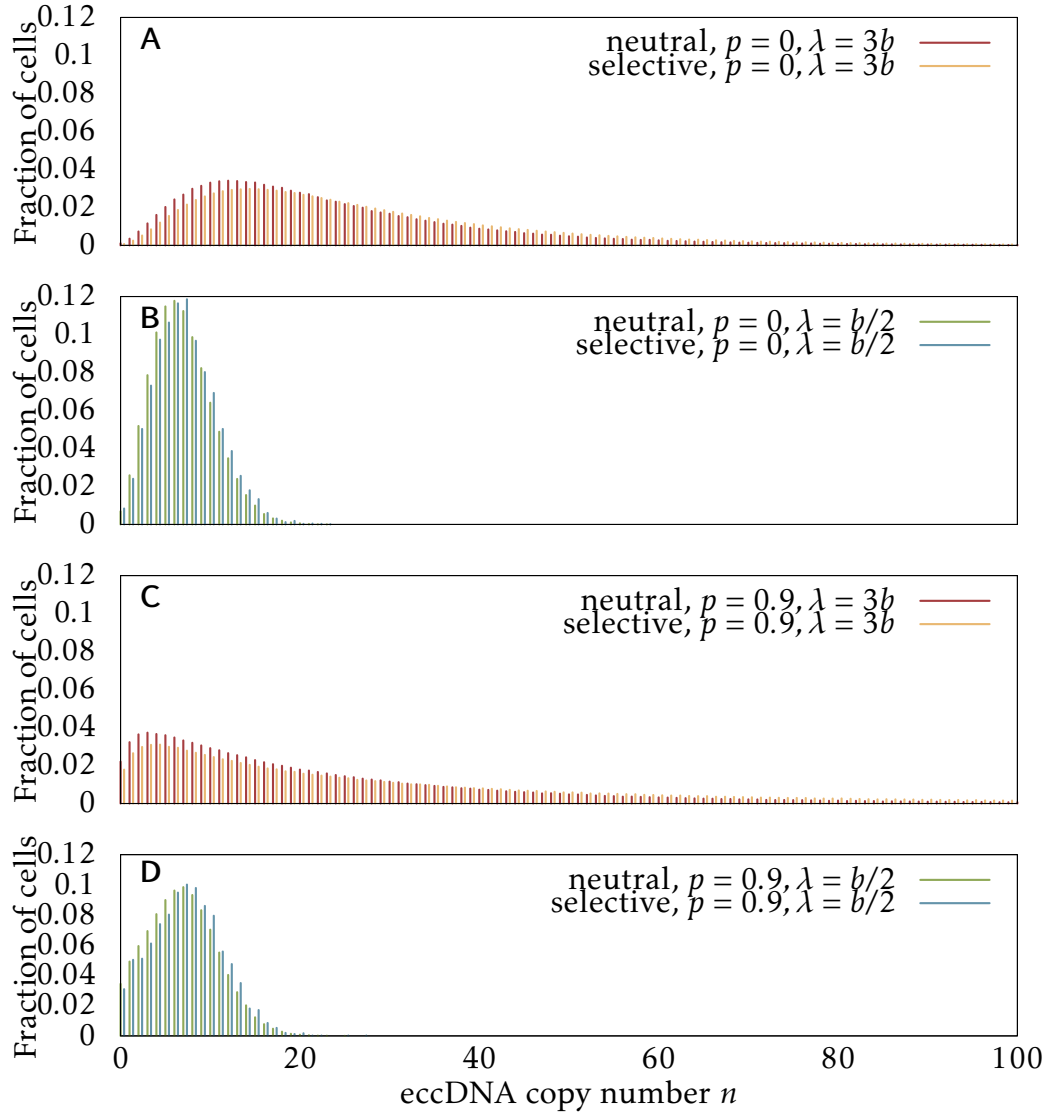


Figure 5.1: Stable distributions of the eccDNA copy number with asynchronous replication under neutral and selective conditions, calculated from stochastic simulations. (A) Unbiased segregation of eccDNA copies ( $p = 0$ ) and fast eccDNA replication  $\lambda = 3b$ . (B) Unbiased segregation of eccDNA copies ( $p = 0$ ) and slow eccDNA replication  $\lambda = b/2$ . (C) Biased segregation of eccDNA copies ( $p = 0.9$ ) and fast eccDNA replication  $\lambda = 3b$ . (D) Biased segregation of eccDNA copies ( $p = 0.9$ ) and slow eccDNA replication  $\lambda = b/2$ . In all cases  $b = 1$ ,  $\nu = 24b$ , and  $\mu = 3b$ . In the neutral cases,  $d_n = 0.25$ ; in the selective conditions,  $d_n$  is given by equation (5.1), with  $d_{\max} = 2$ ,  $d_{\min} = 0.25$ , and  $c = 1/6$ , which gives a threshold of positive selection  $d_n = b_n$  at  $n = 8$ .

#### 5.2.4 Results: population persistence

We are interested not only in the eccDNA copy number, but also in whether it can rescue a yeast population from extinction. Our fitness function in equation 5.1 is chosen to produce a rescue scenario: at low eccDNA copy numbers,  $d_n > b_n$ , and the population is doomed to eventual extinction. Only if there are a sufficient number of cells with high enough eccDNA copy numbers to have positive net growth rates can the population persist. To determine the probability of population persistence, we used further stochastic simulations to find the single-cell establishment probability, the probability that the lineage founded by an individual cell persists arbitrarily far into the future (for the simulation results, a lineage was considered to have established if it had not gone extinct after 20 generations). The single-cell establishment probability can be extended to a larger founding population: if  $P_{\text{est}}$  is the single-cell establishment probability, the probability a population of initial size  $N$  persists is  $1 - (1 - P_{\text{est}})^N$ , assuming the cells in the initial population can be treated independently (for example because the population density is low). Establishment probabilities from stochastic simulations are shown in Figure 5.2. Unsurprisingly, increasing  $\lambda$  makes establishment more likely, because it increases the typical eccDNA copy number. The bias of segregation also has an effect on establishment: increased bias makes establishment less probable, because it increases the variance in copy number, and means there are more cells with copy numbers below the persistence threshold.

### 5.3 Synchronous reproduction

#### 5.3.1 The model

In this second model, we incorporate the same eccDNA processes as we did in the first, except that we assume that the replication of eccDNA copies is synchronized with chromosomal replication, so that each copy is replicated at most once per cell cycle: “at most” because we assume that eccDNA replication is less successful than chromosomal replication, and the eccDNA origins of replication may fail to recruit replication machinery. For this synchronous reproduction model, we change to a discrete-time process and model the yeast population generation by generations. Host cells still reproduce at a rate  $b_n$  and die at a rate  $d_n$ , which may depend on the number  $n$  of copies of the focal eccDNA present at cell birth: now these two rates affect the process only together via the probability that a cell survives to reproduce in its generation,  $b_n/(b_n + d_n)$ . Note that, since the eccDNA copy number will still change over the course of the generation, we are making an implicit approximation here. During each generation, the eccDNA copies undergo the same processes. A number of eccDNA copies are produced from the chromosome and added to the eccDNA pool according to

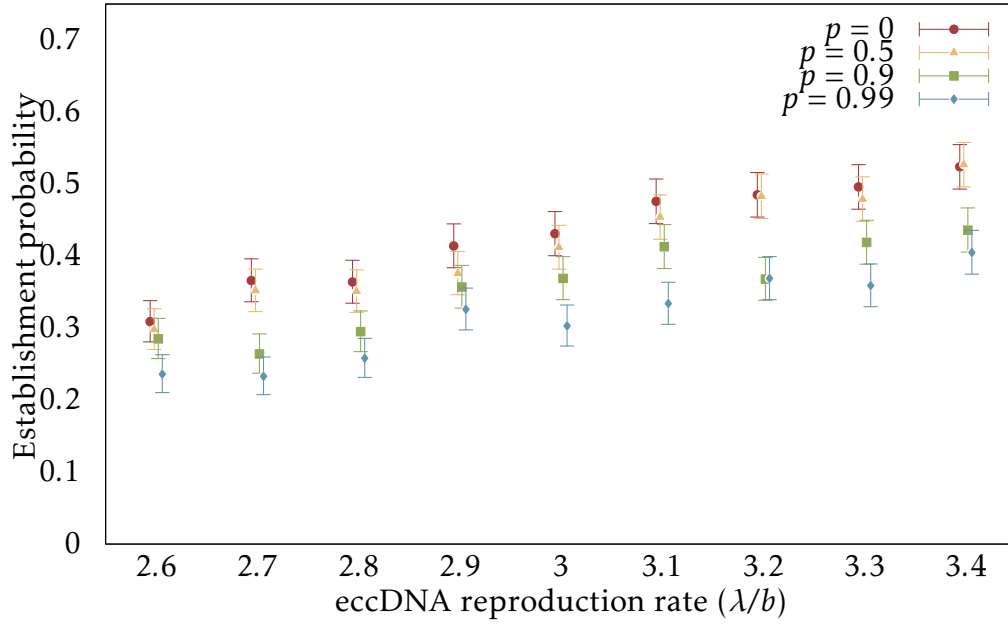


Figure 5.2: Establishment probabilities of a cell line founded by a single yeast cell under selection for the eccDNA, together with 95% confidence intervals, from stochastic simulations. Each simulation started from a single cell with 26 eccDNA copies (the mean number of copies for a cell with the parameters used for this figure and  $\lambda = 2.6b$ ). In all cases,  $b = 1$ ,  $\nu = 24b$ ,  $\mu = 3b$ , and  $d_n$  is given by equation (5.1), with  $d_{\max} = 2$ ,  $d_{\min} = 0.25$ , and  $c = 1/6$ , which gives a threshold of positive selection  $d_n = b_n$  at  $n = 8$ . Confidence intervals calculated by the likelihood-ratio test.

some distribution. Each eccDNA copy in the cell can then be independently degraded with probability  $m$ . We assume that all eccDNA production and degradation occurs in the earliest portion of the cell cycle, the  $G_1$  phase. Afterwards, each eccDNA copy present in the cell during the S phase of the cell cycle is replicated with probability  $r < 1$ . After replication, the host cell divides, and the eccDNA copies segregate in a random but possibly biased fashion, with bias parameter  $p$ , as in the first model.

Suppose a host cell is born with  $N_i$  copies of a focal eccDNA: what is the distribution of the eccDNA copy number  $N_{i+1}$  in a future daughter of this cell? During the cell's life time,  $P_i$  new eccDNA copies will be produced from the chromosome: the  $P_i$  for different cells and generations will all be independent and identically distributed and independent from  $N_i$ . After some of the eccDNA copies are possibly degraded, there will be

$$S_i \sim \text{Bin}(N_i + P_i, 1 - m)$$

surviving copies remaining. The eccDNA copies may then be independently replicated, and the number of new copies produced during repli-

cation is

$$R_i \sim \text{Bin}(S_i, r).$$

When the host cell divides, the  $S_i + R_i$  eccDNA copies present are then segregated between the mother and daughter cell; we randomly pick one or the other to follow into the next generation, and our chosen cell will have

$$N_{i+1} \sim \frac{1}{2} \text{Bin}(S_i + R_i, (1+p)/2) + \frac{1}{2} \text{Bin}(S_i + R_i, (1-p)/2)$$

eccDNA copies.

### 5.3.2 Analysis: neutral conditions

If we again look at the neutral case, where there is no copy-number-dependent selection, we can obtain recurrence relations for the mean copy number of a cell in generation  $i$ ,

$$\mathbf{E}[N_{i+1}] = \frac{1+r}{2}(1-m)(\mathbf{E}[N_i] + \mathbf{E}[P_i]),$$

and the variance of the copy number,

$$\begin{aligned} \mathbf{Var}[N_{i+1}] &= (1+p^2) \left( \frac{1+r}{2} \right)^2 (1-m)^2 (\mathbf{Var}[N_i] + \mathbf{Var}[P_i]) + p^2 \mathbf{E}[N_{i+1}]^2 \\ &\quad + \frac{r(1-r)(1+p^2) + (1-p^2)(1+r) + (1+p^2)(1+r)^2 m}{2(1+r)} \mathbf{E}[N_{i+1}]. \end{aligned}$$

The solution for the mean is then

$$\mathbf{E}[N_i] = N_0 \left( \frac{(1+r)(1-m)}{2} \right)^i + \frac{\mathbf{E}[P_i]}{(1-r)(1+m)} \left( 1 - \left( \frac{(1+r)(1-m)}{2} \right)^i \right),$$

which, since  $(1+r)(1-m)/2 < 1$ , converges to a finite quantity in the long-term limit. Whether the variance converges to a finite quantity in the long-term limit depends on the particular values of  $p$ ,  $r$ , and  $m$ : in the absence of biased segregation, the variance remains bounded, but if the bias is sufficiently large,  $p \geq \sqrt{(2/(1+r)/(1-m))^2 - 1}$ , it increases without bound (this further entails that if  $(1+r)(1-m) < \sqrt{2}$ , the variance is bounded for any segregational bias).

### 5.3.3 Results: copy number distribution and population persistence

As with the asynchronous model, we have again performed stochastic simulations to obtain an estimate of the stable copy number distribution. The complete stochastic process model described in subsection 5.3.1 was simulated as a discrete-time branching process. Each cell survived to reproduce with probability  $b_n/(b_n + d_n)$ , and if it did reproduce, its eccDNA content

was updated for the events occurring during the cell life time and segregated into the two progeny cells according to the model description. The birth rate of host cells was fixed again at  $b = 1$ , and the death rate  $d_n$  once again followed equation (5.1). The distribution of  $P_i$ , the number of new eccDNA copies produced in a generation, was chosen to be geometric, with its mean as a parameter. To estimate the stable limiting distribution of eccDNA copy number, simulations were run for 20 generations, by which point the copy number distribution had stabilized. Example copy number distributions under neutral and selective conditions are shown in Figure 5.3. The general patterns are similar to those for the asynchronous model, but the eccDNA is generally less abundant.

We also calculated single-cell establishment probabilities from the stochastic simulations, shown in Figure 5.4. The general trends are similar to the asynchronous replication model, although the effect of biased segregation is not so clearly monotonic.

#### 5.4 Discussion

Extrachromosomal circular DNA is a common component of eukaryote genomes. Although eccDNA copies only contain copies of chromosomal DNA, they can still have important effects on the fitness of their hosts when they amplify chromosomal genes. Here, we have shown that the amplification of genes on eccDNA can contribute to rescuing populations of yeast from environmental degradation, due, for example, to heavy metal toxicity, by providing a nonzero probability of population persistence in an environment that would drive a population without eccDNA to extinction. We have also shown the role of the eccDNA and host life cycle in determining the abundance of eccDNA copies: in particular, the intrinsic properties of host and eccDNA seem to be much more influential on the copy number of eccDNA than selection for eccDNA, at least for the parameter ranges we considered. In yeast especially, the properties of yeast cell division play an important role in the distribution of eccDNA copies between cells, and biased segregation of eccDNA copies at host division reduces the probability of rescue due to amplification on eccDNA.

We considered two different models of the replication of eccDNA copies, an asynchronous model in which eccDNA copies are continually able to be replicated, and a synchronous model in which eccDNA replication occurs together with chromosomal replication. Although there is evidence for replication adhering to both models in different host systems (Cohen et al. 2005; Barker et al. 1980), the synchronous model seems to be more common in general. Our results, both analytic and from simulations, show that the synchronous model is much more stringent, resulting generally in lower abundance of eccDNA copies and lower probabilities of rescue at comparable parameter values. In particular, synchronous replication excludes the

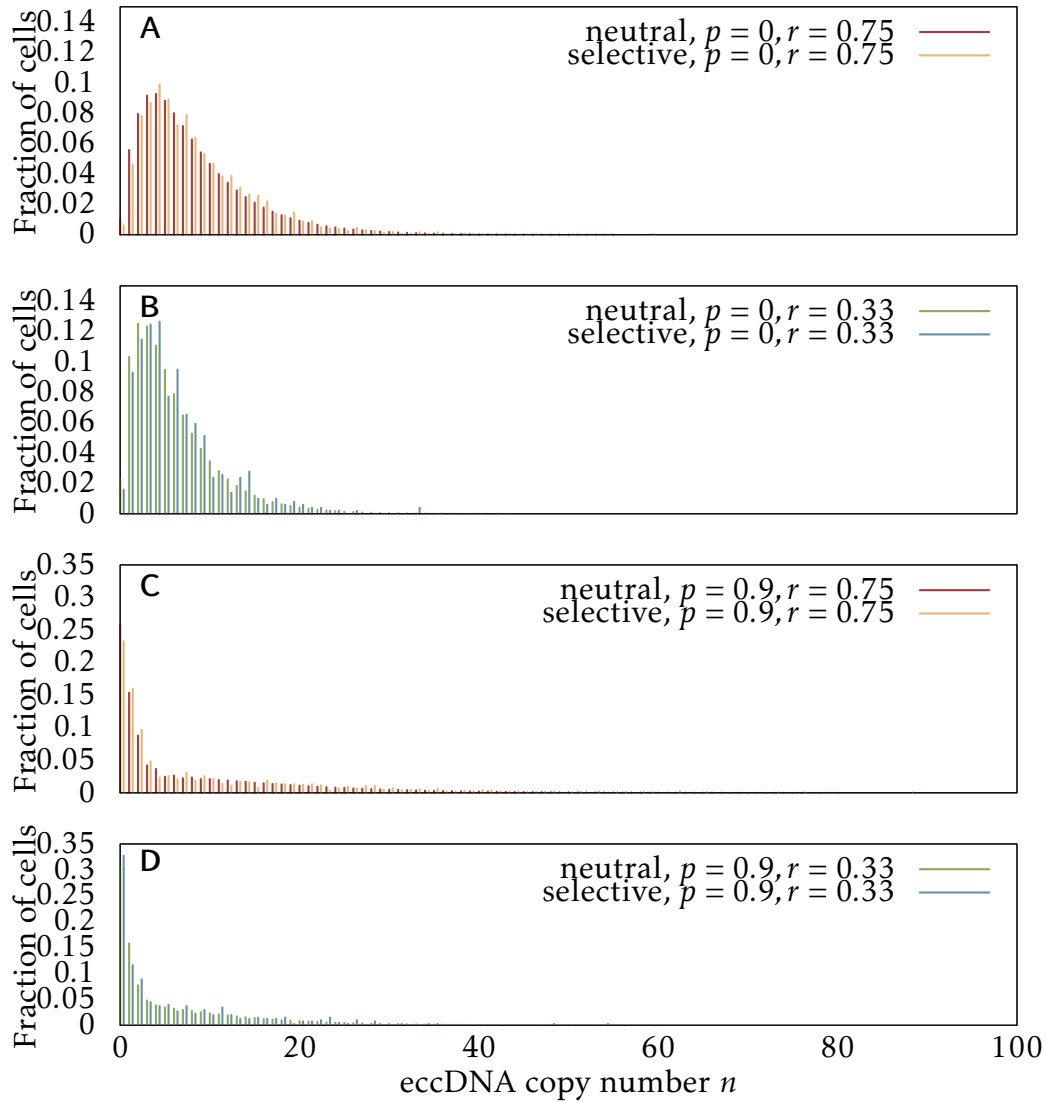


Figure 5.3: Stable distributions of the eccDNA copy number with synchronous replication under neutral and selective conditions, calculated from stochastic simulations. (A) Unbiased segregation of eccDNA copies ( $p = 0$ ) and high probability of eccDNA replication  $r = 0.75$ . (B) Unbiased segregation of eccDNA copies ( $p = 0$ ) and low probability of eccDNA replication  $r = 0.33$ . (C) Biased segregation of eccDNA copies ( $p = 0.9$ ) and high probability of eccDNA replication  $r = 0.75$ . (D) Biased segregation of eccDNA copies ( $p = 0.9$ ) and low probability of eccDNA replication  $r = 0.33$ . In all cases  $b = 1$ ,  $\mathbf{E}[P_i] = 24$ , and  $m = 0.75$ . In the neutral cases,  $d_n = 0.5$ ; in the selective conditions,  $d_n$  is given by equation (5.1), with  $d_{\max} = 2$ ,  $d_{\min} = 0.25$ , and  $c = 1/6$ , which gives a threshold of positive selection  $d_n = b_n$  at  $n = 8$ .

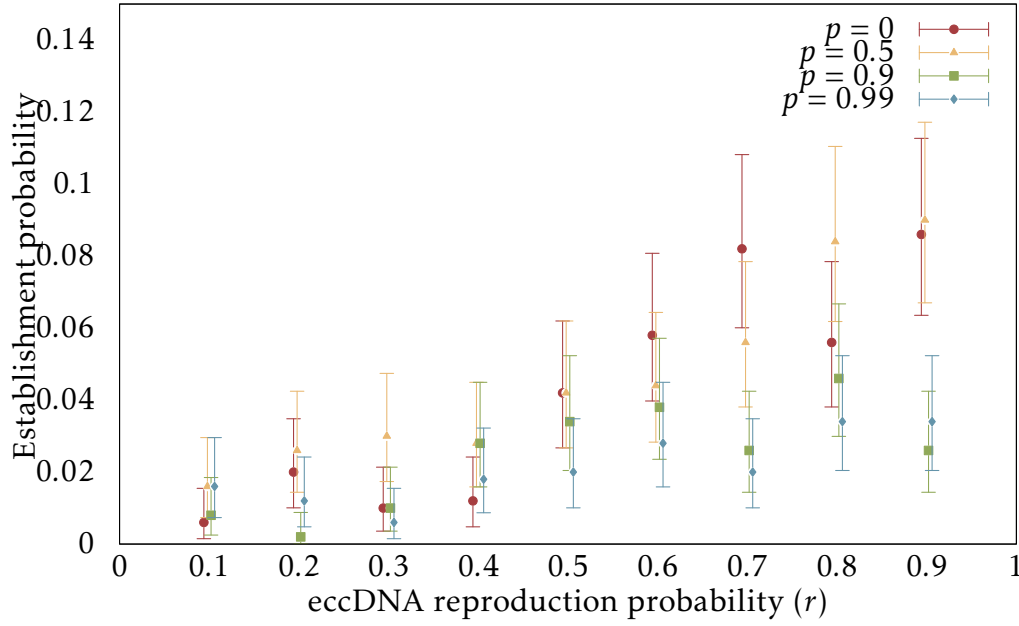


Figure 5.4: Establishment probabilities of a cell line founded by a single yeast cell under selection for the eccDNA, together with 95% confidence intervals, from stochastic simulations. Each simulation started from a single cell with 21 eccDNA copies (the mean number of copies for a cell with the parameters used for this figure and  $r = 0.1$ ). In all cases,  $b = 1$ ,  $\mathbf{E}[P_i] = 24$ ,  $m = 0.75$ , and  $d_n$  is given by equation (5.1), with  $d_{\max} = 2$ ,  $d_{\min} = 0.25$ , and  $c = 1/6$ , which gives a threshold of positive selection  $d_n = b_n$  at  $n = 8$ . Confidence intervals calculated by the likelihood-ratio test.

possibility of the mean eccDNA copy number in a population increasing indefinitely.

The degree of bias in the segregation of eccDNA copies plays an important role in the biology of eccDNA. It has no effect on the average abundance of eccDNA copies (since both the mother and daughter cell, no matter how similar or different their eccDNA copy number, contribute equally to the population mean), but is a key contributor to variance within the population. This generally accords with empirical studies of eccDNA abundance in yeast, which find that there is considerable variability in eccDNA content between subpopulations (Prada-Luengo et al. 2020). Our analytical results showed that there are regions of our parameter space where, while the mean eccDNA copy number is bounded, the variance increases without bound (this is true for both synchronous and asynchronous replication, although with synchronous replication this region only exists if the degree of segregational bias is sufficiently high): this may be because biased segregation allows a subpopulation to accumulate eccDNA copies to a large

number, while they are lost from the rest of the population quickly, keeping the typical copy number low. There is some evidence of a phenomenon like this in yeast with *CUP1* eccDNA: although this eccDNA has a very low replication rate (it has a chromosomal origin of replication, but seemingly not a very effective one), some cells are still able to accumulate very large quantities of the eccDNA, due to a high rate of production and a strong bias in segregation (Hull et al. 2019; Hull and Houseley 2020).

Our model considered a single, unidirectional change in the environment which made the gene carried by our focal eccDNA essential to population persistence. This is a reasonably good model for scenarios like the contamination of the environment of a population with heavy metals or other toxins. Other selective pressures, however, may be intermittent or fluctuating, so that the population is alternately exposed to harsh and permissive environments, or to different selective pressures in alternation. In this scenario, the transient nature of gene amplification on eccDNA may be particularly advantageous, allowing eccDNA copies to accumulate when needed and dissipate when not. A natural extension of our model would be to explore this scenario by allowing the fitness effect of eccDNA to vary over time.

Extrachromosomal circular DNA seems particularly common—or at least particularly well-studied—in yeast, and we have incorporated one component of the asymmetrical budding division of yeast into the model by allowing for biased segregation of eccDNA copies between mother and daughter cells. Other than this, however, we treat the mother and daughter cells symmetrically in the model. A more biologically realistic model could also incorporate the effects of aging: yeast mother and daughter cells are not identical; in particular yeast cells can only produce a finite number of buds during their lifespan (Mortimer and Johnston 1959; Sinclair and Guarente 1997), so mothers are distinct from daughters in having a higher reproductive age. An extended model could add an age structure to the yeast population, at the cost of greatly increasing the analytic complexity of the model.

We have focused on short-term, rapid adaptation to an environmental change, but if the selective conditions persist over a longer period, there may be further evolutionary changes in the yeast population. One of the simplest of these is reincorporation of eccDNA copies into the chromosome, which could serve to stabilize the amplified phenotype when the conditions that require it persist indefinitely (this was modelled in the context of cancers by Kimmel et al. (1992)). The reverse phenomenon can also occur: in some cases the formation of new eccDNA may involve the deletion of the chromosomal copy of a gene (Møller et al. 2018; Paulsen et al. 2018). Over the longer term, parameters that we have taken to be fixed on the time scale of our model could also evolve. Cells could adapt to increase the production rate of new eccDNA copies, increasing the abundance of the eccDNA

when it might be needed. Ideally, this increase in production could be limited to circumstances when the eccDNA is particularly needed; there is some evidence that this has happened in some chromosomal regions, which appear to be particularly prone to producing eccDNA when heavily transcribed (Hull and Houseley 2020; Arrey et al. 2022). The degree of bias in segregation has been found to vary in response to environmental conditions (Baldi et al. 2017; Prada-Luengo et al. 2020), a plastic response, and over longer time scales it could also evolve. Indeed, this is one of the major questions our model suggests: why, if abundant eccDNA copies are beneficial, would there be a bias to retain eccDNA copies in the mother cell, and leave the daughter relatively unprotected? It is possible that the answer to this question lies in a conflict of interest between mother and daughter; another possible explanation is that selective retention is optimized for normal circumstances when eccDNA may be neutral or even harmful for the cell, and not the exceptional circumstances when it is needed to rescue the population. Since yeast are capable of preventing the transmission of eccDNA to both progeny of a cell division, and do so during meiosis (King et al. 2019), it is plausible that there is an adaptive advantage to biased segregation over elimination. Hull and Houseley (2020) propose that there is a trade-off between fast growth (impeded by eccDNA) and capacity for adaptation (facilitated by eccDNA), and that biased segregation allows the yeast to change from one side of this trade-off to the other over the course of their life span.

This also points out another limitation of our model, which is that we have not chosen to incorporate a cost to the host cell from carrying the eccDNA copies. We have assumed that this cost is relatively negligible at the copy numbers we treat in our model, but at sufficiently large numbers we expect there would eventually be a cost effect. This places limits on the plausible values of the parameters describing the eccDNA life cycle, since indefinite accumulation (which we predicted when  $\lambda - \mu - b/2 \geq 0$ ) would eventually lead to death of the host cells.

There are many circumstances in which rapid adaptation of populations is crucial to their persistence. The population genetics of rapid adaptation, and especially of evolutionary rescue, are of great importance to understanding the history of biological populations. Here, we have examined the role of one special genetic system, an extrachromosomal genetic element, to rapid adaptation in yeast populations. We have shown that gene amplification on eccDNA copies can facilitate rescue of yeast populations, and that the ability of eccDNA to do so is modulated by special properties of the system particular to their extrachromosomal nature, especially the role of biased segregation. These results should contribute to furthering the population genetics of extrachromosomal elements in a novel system.

## CONCLUSION

Extrachromosomal genetic elements make up a relatively small proportion of the genome of the organisms in which they appear, but they can carry genes that have important fitness effects on their hosts. The pattern of inheritance of these extrachromosomal elements is often very distinct from that of the chromosomal DNA, which means that there are unique factors to be taken into account in the population genetics of genes on extrachromosomal elements. The theoretical population genetics of these elements has also been less thoroughly explored than that of chromosomes. The diversity of extrachromosomal elements means that there is still a largely untapped well of interesting problems in the genetics of extrachromosomal elements for modellers to attack.

In this thesis, I have attacked two of those problems, modelling the population genetics of two particular extrachromosomal systems, plasmids in bacteria and eccDNA in yeast, in the special context of evolutionary rescue. In the scenarios we chose to model, multicopy plasmids and eccDNA are uniquely positioned to make an important contribution to the fitness of their hosts: the multicopy plasmid in a heterozygote advantage scenario, and eccDNA in a scenario where gene amplification contributes to rescue. In both cases, we predicted rescue probabilities and relative abundances of the extrachromosomal elements. In the plasmid case, it was possible to go further and find analytic conditions on the possibility of rescue. Hopefully, our results will contribute to the continued development of the theoretical population genetics of extrachromosomal elements, as will our introduction to plasmid biology.

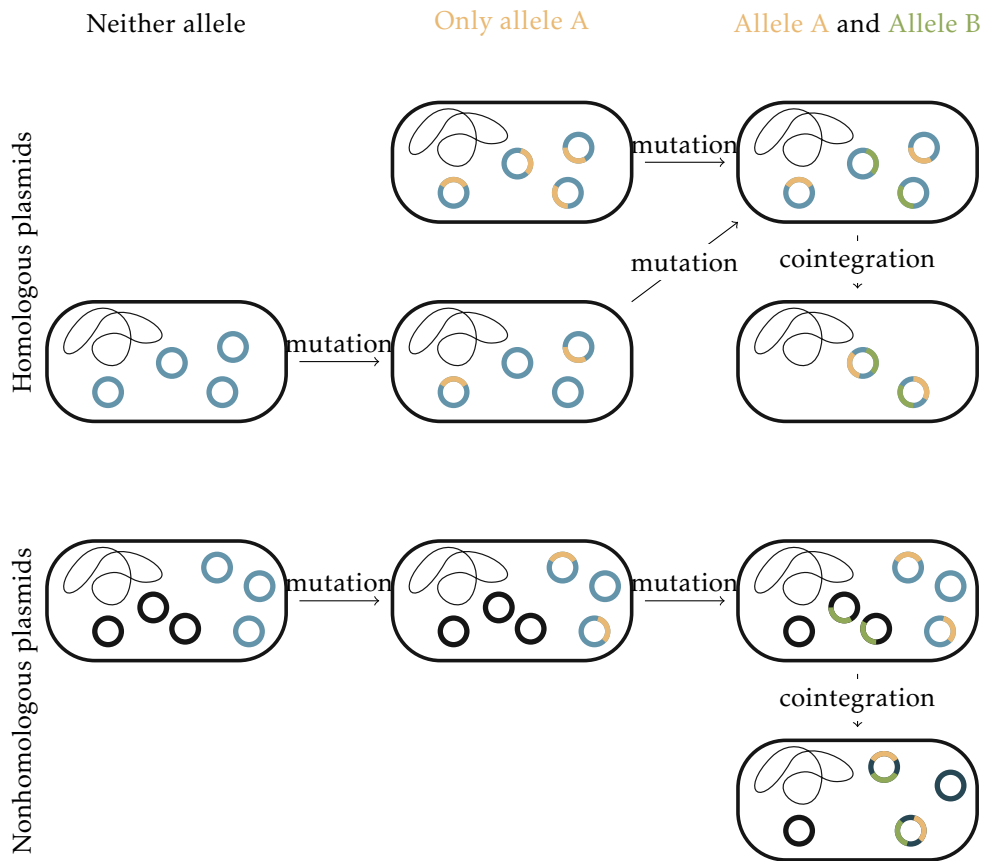
The multicopy plasmids and eccDNA we considered here are superficially similar—they are both circular and extrachromosomal—but in many respects lie at opposite ends of a spectrum of possible extrachromosomal elements. A plasmid contains many plasmid-specific genes, forming its backbone, which regulate its physiology, and in particular control its copy number and replication. The eccDNA, on the contrary, has no genes of its own, and relies completely on its host, resulting in a copy number that is essentially completely controlled by external factors. They are similar also in one respect that we did not explore: they both exist in multiple copies, which means that the advantages of the multicopy state that we considered in our plasmid model—especially the possibility of heterozygosity and the increased mutational input—should also apply to eccDNA. This was probably not relevant to the scenario that motivated our eccDNA model, but future models of eccDNA could consider this: mutations have been ob-

served to occur on eccDNA (Xu et al. 2018). Even more generally, multiple types of eccDNA from different chromosomal origins coexist in a single cell (Prada-Luengo et al. 2020), but we have focused only on a single focal type at a time; since interactions between distinct types of eccDNA have been observed (Hung et al. 2023), other eccDNA models could consider this aspect.

In the plasmid rescue model, the maintenance of multiple alleles together was required to rescue the bacterial population. We explored one mechanism by which this could occur—one allele already being present in the existing population and the other arriving by mutation during rescue—but there are other possible routes. Some of these are shown in Figure 6.1 (adapted from my initial thesis proposal). What we have covered (the top right corner of the diagram) is probably the most interesting portion of the diagram, because it is where the loss of heterozygosity to segregation is most important. Nonetheless, future work on the persistence of multiple alleles on plasmids during a rescue scenario (motivated perhaps by the example of resistance to multiple antibiotics) could consider the acquisition of both alleles during the rescue process or the maintenance of the alleles on compatible plasmids (or fusions thereof).

Developing more mechanistic models of plasmid replication and segregation that describe the process in a more biologically realistic manner would be a useful extension to our work. Our concession to realistic modelling of the mechanisms of plasmid reproduction and segregation was limited to comparing regular and random models of replication, and although we believe that the random model is more biologically realistic, the regular model has sometimes been found to be a better fit to experimental data (Garoña et al. 2023). By looking into the literature on the mechanisms of control of plasmid replication, it might be possible to develop a mechanistic model of plasmid replication that could supersede both regular and random replication.

Our model for segregation could also be made more realistic. We adopted a fairly strict assumption that every daughter cell receives exactly the same number of copies of the plasmid, meaning that every cell in the model is born with exactly  $n$  total copies. For those low-copy-number plasmids which rely on an active partitioning system to separate plasmids at host cell division, this is a fairly reasonable assumption. But plasmids with higher copy numbers which don't rely on active partitioning cannot guarantee exact partitioning of plasmid copies (Zielenkiewicz and Cegłowski 2001). For these, a model with random segregation, in which the number of plasmids received by a daughter cell is not fixed (it would be given by a binomial distribution) and in the daughter cell plasmids are replicated from however many copies the cell starts with up to  $2n$  for the next host generation, might be more suitable. However, because of clustering and other non-random spatial distribution of plasmids within the cell, the random model



*Figure 6.1:* Stages in the acquisition of two plasmid-borne alleles. Cells in the left-most column have neither allele, those in the middle column have only allele A (indicated in orange), and those in the right column have both alleles A and B (indicated in green). Cells in the upper half of the diagram have only one plasmid type carrying both alleles, either on distinct plasmids of the common type (homologous plasmids), or on a cointegrate formed by fusion of plasmids. Cells in the lower half have two plasmid types, and distinct alleles are located on plasmids of distinct types (nonhomologous plasmids), or both on a cointegrate formed by hybridization of the two plasmid types.

might not reflect the full biological reality either. Indeed, clustering could make the segregation process not neutral between mutant and wild-type plasmids. The development of mechanistic models of segregation could allow us to drop such unrealistic assumptions.

It would also be helpful to have a better model of replication and segregation when modelling cointegration. Our cointegration model was deliberately simple, but the constraints of our strict assumptions are already felt in such aspects as the fact that a cell with  $n = 3$  which contains a dimer can have only daughters identical with itself, since if two dimers segregated to the same daughter, it would have four copies of the plasmid monomer. This seems implausible: it is difficult to imagine a biological mechanism that would achieve this effect given that for multicopy plasmids copy number control principally works by regulating plasmid replication, not segregation (Summers 1996, chap. 4). A more realistic model for segregation would hopefully give better guidance on how cointegrates should behave. Generally, a more elaborate model of cointegration and resolution of plasmid multimers would be a desirable goal for future modelling work: right now, both the empirical and theoretical literature on the subject are a little thin, making it difficult to, for example, choose appropriate parameter values.

In the discussion of Chapter 3, we mention that it would be interesting to perform a similar analysis of the distribution of plasmid copy numbers, to see if it is bimodal and to examine the relationship between copy number and length. At the time, we thought this was not a feasible proposal because of a lack of available data on copy number and because it is questionable whether *the* copy number of a plasmid is even a well-defined concept, given that copy number control is not exact and plasmid copy numbers may vary substantially between different hosts and in different growth conditions (Bazaraal and Helinski 1970; Gustafsson and Nordström 1980). At least one of these limitations has been overcome by the power of bioinformatics, however, and recently two groups have reported data on the distribution of plasmid copy numbers (Maddamsetti et al. 2024; Ramiro-Martínez et al. 2024). Both report a power law relationship between length and copy number. They also both report that the distribution of copy numbers is bimodal—but on a log scale! It would be interesting to compare the distribution of copy numbers to that of log copy numbers; perhaps a sequel to Chapter 3 is in order.

## REFERENCES

- Abe, Kimihiro, Nobuhiko Nomura, and Satoru Suzuki (2020). "Biofilms: hot spots of horizontal gene transfer (HGT) in aquatic environments, with a focus on a new HGT mechanism." In: *FEMS Microbiology Ecology* 96. DOI: 10.1093/femsec/fiaa031.
- Abe, Ryuichiro, Yukihiro Akeda, Yo Sugawara, Yuki Matsumoto, Daisuke Motooka, Ryuji Kawahara, Norihisa Yamamoto, Kazunori Tomono, Tetsuya Iida, and Shigeyuki Hamada (2021). "Enhanced Carbapenem Resistance through Multimerization of Plasmids Carrying Carbapenemase Genes." In: *mBio* 12.3, DOI: 10.1128/mbio.00186-21.
- Alexander, Helen K. and Sebastian Bonhoeffer (2012). "Pre-existence and emergence of drug resistance in a generalized model of intra-host viral dynamics." In: *Epidemics* 4.4, pp. 187–202. DOI: 10.1016/j.epidem.2012.10.001.
- Allen, Sally Lyman and David L. Nanney (1958). "An Analysis of Nuclear Differentiation in the Selfers of *Tetrahymena*." In: *The American Naturalist* 92.864, pp. 139–160. DOI: 10.1086/282022.
- Anda, Mizue, Yoshiyuki Ohtsubo, Takashi Okubo, Masayuki Sugawara, Yuji Nagata, Masataka Tsuda, Kiwamu Minamisawa, and Hisayuki Mitsui (2015). "Bacterial clade with the ribosomal RNA operon on a small plasmid rather than the chromosome." In: *Proceedings of the National Academy of Sciences* 112.46, pp. 14343–14347. DOI: 10.1073/pnas.1514326112.
- Archambaud, M., G. Gerbaud, E. Labau, N. Marty, and P. Courvalin (1991). "Possible in-vivo transfer of  $\beta$ -lactamase TEM-3 from *Klebsiella pneumoniae* to *Salmonella kedougou*." In: *Journal of Antimicrobial Chemotherapy* 27.4, pp. 427–436. DOI: 10.1093/jac/27.4.427.
- Ares-Arroyo, Manuel, Charles Coluzzi, and Eduardo P.C. Rocha (2022). "Origins of transfer establish networks of functional dependencies for plasmid transfer by conjugation." In: *Nucleic Acids Research*. DOI: 10.1093/nar/gkac1079.
- Arrey, Gerard, S. Keating, and Birgitte Regenber (2022). "A unifying model for extrachromosomal circular DNA load in eukaryotic cells." In: *Seminars in Cell and Developmental Biology* 128, pp. 40–50. DOI: 10.1016/j.semcdb.2022.03.002.
- Austin, D. J., M. Kakehashi, and R. M. Anderson (1997). "The transmission dynamics of antibiotic-resistant bacteria: the relationship between resistance in commensal organisms and antibiotic consumption." In: *Proceedings of the Royal Society B* 264.1388, pp. 1629–1638. DOI: 10.1098/rspb.1997.0227.

- Bahl, Martin Iain, Lars Hestbjerg Hansen, and Søren J. Sørensen (2009). "Persistence Mechanisms of Conjugative Plasmids." In: *Horizontal Gene Transfer*. Ed. by Maria Boekels Gogarten, Johann Peter Gogarten, and Lorraine Olendzenski. *Methods in Molecular Biology* 532. Chap. 5, pp. 73–102. ISBN: 978-1-60327-852-2. DOI: 10.1007/978-1-60327-853-9\_5.
- Baldi, Sandro, Alessio Bolognesi, Anne Cornelis Meinema, and Yves Barral (2017). "Heat stress promotes longevity in budding yeast by relaxing the confinement of age-promoting factors in the mother cell." In: *eLife* 6, DOI: 10.7554/eLife.28329.
- Barker, P., H. Drwinga, W. Hittelman, and A. Maddox (1980). "Double minutes replicate once during S phase of the cell cycle." In: *Experimental Cell Research* 130.2, pp. 353–360. DOI: 10.1016/0014-4827(80)90012-9.
- Baskin, F., R. N. Rosenberg, and V. Dev (1981). "Correlation of double-minute chromosomes with unstable multidrug cross-resistance in uptake mutants of neuroblastoma cells." In: *Proceedings of the National Academy of Sciences* 78.6, pp. 3654–3658. DOI: 10.1073/pnas.78.6.3654.
- Batra, Aditi, Roderich Roemhild, Emilie Rousseau, Sören Franzenburg, Stefan Niemann, and Hinrich Schulenburg (2021). "High potency of sequential therapy with only  $\beta$ -lactam antibiotics." In: *eLife* 10, DOI: 10.7554/eLife.68876.
- Baur, Erwin (1908). "Das Wesen und die Erblchkeitsverhältnisse der „Varietates albomarginatae hort.“ von *Pelargonium zonale*." In: *Zeitschrift für Induktive Abstammungs- und Vererbungslehre* 1.1, pp. 330–351. DOI: 10.1007/bf01990603.
- Bazara, Michael and Donald R. Helinski (1970). "Replication of a Bacterial Plasmid and an Episome in *Escherichia coli*." In: *Biochemistry* 9.2, pp. 399–406. DOI: 10.1021/bi00804a029.
- Bedbrook, John R. and Frederick M. Ausubel (1976). "Recombination between bacterial plasmids leading to the formation of plasmid multimers." In: *Cell* 9.4, pp. 707–716. DOI: 10.1016/0092-8674(76)90134-3.
- Beijersbergen, Alice, Sander J. Smith, and Paul J.J. Hooykaas (1994). "Localization and Topology of VirB Proteins of *Agrobacterium tumefaciens*." In: *Plasmid* 32.2, pp. 212–218. DOI: 10.1006/plas.1994.1057.
- Bennett, P. M. (2008). "Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria." In: *British Journal of Pharmacology* 153.S1, S347–S357. DOI: 10.1038/sj.bjpp.0707607.
- Bergstrom, Carl T., Marc Lipsitch, and Bruce R. Levin (2000). "Natural Selection, Infectious Transfer and the Existence Conditions for Bacterial Plasmids." In: *Genetics* 155.4, pp. 1505–1519. DOI: 10.1093/genetics/155.4.1505.
- Bergstrom, Carl T., Monique Lo, and Marc Lipsitch (2004). "Ecological theory suggests that antimicrobial cycling will not reduce antimicrobial re-

- sistance in hospitals.” In: *Proceedings of the National Academy of Sciences* 101.36, pp. 13285–13290. DOI: 10.1073/pnas.0402298101.
- Bethke, Jonathan H., Helena R. Ma, Ryan Tsoi, Li Cheng, Minfeng Xiao, and Lingchong You (2022). “Vertical and horizontal gene transfer trade-offs direct plasmid fitness.” In: *Molecular Systems Biology* 19.2, DOI: 10.15252/msb.202211300.
- Beverly, Stephen M., Jeffrey A. Coderre, Daniel V. Santi, and Robert T. Schimke (1984). “Unstable DNA amplifications in methotrexate resistant *Leishmania* consist of extrachromosomal circles which relocalize during stabilization.” In: *Cell* 38.2, pp. 431–439. DOI: 10.1016/0092-8674(84)90498-7.
- Billane, Kathryn, Ellie Harrison, Duncan Cameron, and Michael A. Brockhurst (2021). “Why do plasmids manipulate the expression of bacterial phenotypes?” In: *Philosophical Transactions of the Royal Society B* 377, DOI: 10.1098/rstb.2020.0461.
- Bingle, Lewis E. H. and Christopher M. Thomas (2001). “Regulatory circuits for plasmid survival.” In: *Current Opinion in Microbiology* 4.2, pp. 194–200. DOI: 10.1016/S1369-5274(00)00188-0.
- Birky Jr., C. William (1983). “Relaxed Cellular Controls and Organelle Heredity.” In: *Science* 222.4623, pp. 468–475. DOI: 10.1126/science.6353578.
- Blesa, Alba, Ignacio Baquedano, Nieves G. Quintáns, Carlos P. Mata, José R. Castón, and José Berenguer (2017). “The transjugation machinery of *Thermus thermophilus*: Identification of TdtA, an ATPase involved in DNA donation.” In: *PLOS Genetics* 13.3. DOI: 10.1371/journal.pgen.1006669.
- Boekhoud, Ilse M., Bastian V. H. Hornung, Eloisa Sevilla, Céline Harmanus, Ingrid M. J. G. Bos-Sanders, Elisabeth M. Terveer, Rosa Bolea, Jeroen Corver, Ed J. Kuijper, and Wiep Klaas Smits (2020). “Plasmid-mediated metronidazole resistance in *Clostridioides difficile*.” In: *Nature Communications* 11.1, DOI: 10.1038/s41467-020-14382-1.
- Bouma, Judith E. and Richard E. Lenski (1988). “Evolution of a bacteria/plasmid association.” In: *Nature* 335.6188, pp. 351–352. DOI: 10.1038/335351a0.
- Boyd, E. Fidelma, Charles W. Hill, Stephen M. Rich, and Daniel L. Hartl (1996). “Mosaic Structure of Plasmids From Natural Populations of *Escherichia coli*.” In: *Genetics* 143.3, pp. 1091–1100. DOI: 10.1093/genetics/143.3.1091.
- Brendel, Volker and Alan S. Perelson (1993). “Quantitative Model of ColE1 Plasmid Copy Number Control.” In: *Journal of Molecular Biology* 229.4, pp. 860–872. DOI: 10.1006/jmbi.1993.1092.
- Brockhurst, Michael A. and Ellie Harrison (2022). “Ecological and evolutionary solutions to the plasmid paradox.” In: *Trends in Microbiology* 30.6, pp. 534–543. DOI: 10.1016/j.tim.2021.11.001.

- Brom, Susana, Lourdes Girard, Cristina Tun-Garrido, Alejandro García-de los Santos, Patricia Bustos, Víctor González, and David Romero (2004). "Transfer of the Symbiotic Plasmid of *Rhizobium etli* CFN42 Requires Cointegration with p42a, Which May Be Mediated by Site-Specific Recombination." In: *Journal of Bacteriology* 186.22, pp. 7538–7548. DOI: 10.1128/jb.186.22.7538-7548.2004.
- Brovedan, Marco, Guillermo D. Repizo, Patricia Marchiaro, Alejandro M. Viale, and Adriana Limansky (2019). "Characterization of the diverse plasmid pool harbored by the bla<sub>N<sub>DM-1</sub></sub>-containing *Acinetobacter bereziniae* HPC229 clinical strain." In: *PLOS ONE* 14.11, DOI: 10.1371/journal.pone.0220584.
- Burian, Ján, Ladislav Guller, Marián Mavčor, and William W. Kay (1997). "Small Cryptic Plasmids of Multiplasmid, Clinical *Escherichia coli*." In: *Plasmid* 37.1, pp. 2–14. DOI: 10.1006/plas.1996.1273.
- Burian, Ján, Stanislav Stuchlík, and William W. Kay (1999). "Replication control of a small cryptic plasmid of *Escherichia coli*." In: *Journal of Molecular Biology* 294.1, pp. 49–65. DOI: 10.1006/jmbi.1999.3266.
- Burrus, Vincent, Guillaume Pavlovic, Bernard Decaris, and Gérard Guédon (2002). "Conjugative transposons: the tip of the iceberg." In: *Molecular Microbiology* 46.3, pp. 601–610. DOI: 10.1046/j.1365-2958.2002.03191.x.
- Cairns, Johannes et al. (2018). "Black Queen Evolution and Trophic Interactions Determine Plasmid Survival after the Disruption of the Conjugation Network." In: *mSystems* 3.5, DOI: 10.1128/mSystems.00104-18.
- Camacho, Juan Pedro M., Timothy F. Sharbel, and Leo W. Beukeboom (2000). "B-chromosome evolution." In: *Philosophical Transactions of the Royal Society B* 355.1394, pp. 163–178. DOI: 10.1098/rstb.2000.0556.
- Carattoli, Alessandra (2003). "Plasmid-Mediated Antimicrobial Resistance in *Salmonella enterica*." In: *Current Issues in Molecular Biology* 5, pp. 113–122. DOI: 10.21775/cimb.005.113.
- (2013). "Plasmids and the spread of resistance." In: *International Journal of Medical Microbiology* 303, pp. 298–304. DOI: 10.1016/j.ijmm.2013.02.001.
- Carlson, Stephanie M., Curry J. Cunningham, and Peter A.H. Westley (Sept. 2014). "Evolutionary rescue in a changing world." In: *Trends in Ecology & Evolution* 29.9, pp. 521–530. DOI: 10.1016/j.tree.2014.06.005.
- Carrilero, Laura, Anastasia Kottara, David Guymmer, Ellie Harrison, James P. J. Hall, and Michael A. Brockhurst (2021). "Positive Selection Inhibits Plasmid Coexistence in Bacterial Genomes." In: *mBio* 12.3. DOI: 10.1128/mbio.00558-21.
- Carroll, Amanda C. and Alex Wong (2018). "Plasmid persistence: costs, benefits, and the plasmid paradox." In: *Canadian Journal of Microbiology* 64.5, pp. 293–304. DOI: 10.1139/cjm-2017-0609.

- Casjens, Sherwood et al. (2000). "A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*." In: *Molecular Microbiology* 35.3, pp. 490–516.
- Chabbert, Y. A. and A. Roussel (1977). "Taxonomy and epidemiology of R plasmids as molecular species." In: *Journal of Antimicrobial Chemotherapy* 3 (suppl C), pp. 25–33. DOI: 10.1093/jac/3.suppl\_c.25.
- Chakrabarty, A. M. (1974). "Dissociation of a Degradative Plasmid Aggregate in *Pseudomonas*." In: *Journal of Bacteriology* 118.3, pp. 815–820. DOI: 10.1128/jb.118.3.815-820.1974.
- Chen, Carton W., Tin-Wein Yu, Yi-Shing Lin, Helen M. Kieser, and David A. Hopwood (1993). "The conjugative plasmid SLP<sub>2</sub> of *Streptomyces lividans* is a 50 kb linear molecule." In: *Molecular Microbiology* 7.6, pp. 925–932. DOI: 10.1111/j.1365-2958.1993.tb01183.x.
- Chiang, C.-S. and H. Bremer (1991). "Maintenance of pBR322-derived plasmids without functional RNAI." In: *Plasmid* 26.3, pp. 186–200. DOI: 10.1016/0147-619x(91)90042-u.
- Ciok, Anna, Lukasz Dziewit, Jakub Grzesiak, Karol Budzik, Dorota Gorniak, Marek K. Zdanowski, and Dariusz Bartosik (2016). "Identification of miniature plasmids in psychrophilic Arctic bacteria of the genus *Variovorax*." In: *FEMS Microbiology Ecology* 92.4, DOI: 10.1093/femsec/fiw043.
- Clark, Alvin J. and Edward A. Adelberg (1962). "Bacterial Conjugation." In: *Annual Review of Microbiology* 16.1, pp. 289–319. DOI: 10.1146/annurev.mi.16.100162.001445.
- Cohen, Sarit, Neta Agmon, Keren Yacobi, Margarita Mislovati, and D. Segal (2005). "Evidence for rolling circle replication of tandem genes in *Drosophila*." In: *Nucleic Acids Research* 33.14, pp. 4519–4526. DOI: 10.1093/nar/gki764.
- Cohen, Sarit, Keren Yacobi, and Daniel Segal (2003). "Extrachromosomal Circular DNA of Tandemly Repeated Genomic Sequences in *Drosophila*." In: *Genome Research* 13, pp. 1133–1145. DOI: 10.1101/gr.907603.
- Coluzzi, Charles, Maria Pilar Garcillán-Barcia, Fernando de la Cruz, and Eduardo P.C. Rocha (2022). "Evolution of Plasmid Mobility: Origin and Fate of Conjugative and Nonconjugative Plasmids." In: *Molecular Biology and Evolution* 39.6, DOI: 10.1093/molbev/msac115.
- Condit, Richard and Bruce R. Levin (1990). "The Evolution of Plasmids Carrying Multiple Resistance Genes: The Role of Segregation, Transposition, and Homologous Recombination." In: *The American Naturalist* 135.4, pp. 573–596. DOI: 10.1086/285063.
- Conteville, Liliane Costa and Ana Carolina Paulo Vicente (2022). "A plasmid network from the gut microbiome of semi-isolated human groups reveals unique and shared metabolic and virulence traits." In: *Scientific Reports* 12.1, DOI: 10.1038/s41598-022-16392-z.

- Cook, Laura C. C. and Gary M. Dunny (2014). "The Influence of Biofilms in the Biology of Plasmids." In: *Microbiology Spectrum* 2.5. DOI: 10.1128/microbiolspec.plas-0012-2013.
- Cooper, Tim F. and Jack A. Heinemann (2000). "Postsegregational killing does not increase plasmid stability but acts to mediate the exclusion of competing plasmids." In: *Proceedings of the National Academy of Sciences* 97.23, pp. 12643–12648. DOI: 10.1073/pnas.220077897.
- Cooper, Tim F., Tiago Paixão, and Jack A. Heinemann (2010). "Within-host competition selects for plasmid-encoded toxin-antitoxin systems." In: *Proceedings of the Royal Society B* 277.1697, pp. 3149–3155. DOI: 10.1098/rspb.2010.0831.
- Correns, C. (1908). "Vererbungsversuche mit blass(gelb)grünen und buntblättrigen Sippen bei *Mirabilis jalapa*, *Urtica pilulifera*, und *Lunaria annua*." In: *Zeitschrift für Induktive Abstammungs- und Vererbungslehre* 1.1, pp. 291–329. DOI: 10.1007/bf01990602.
- Cox, David, Catherine Yuncken, and Arthur I. Spriggs (1965). "Minute Chromatin Bodies in Malignant Tumours of Childhood." In: *The Lancet* 286, pp. 55–58. DOI: 10.1016/S0140-6736(65)90131-5.
- Crozat, Estelle, Florian Fournes, François Cornet, Bernard Hallet, and Philippe Rousseau (2014). "Resolution of Multimeric Forms of Circular Plasmids and Chromosomes." In: *Microbiology Spectrum* 2.5, DOI: 10.1128/microbiolspec.plas-0025-2014.
- Cullum, John and Paul Broda (1979). "Rate of segregation due to plasmid incompatibility." In: *Genetical Research* 33.1, pp. 61–79. DOI: 10.1017/S0016672300018176.
- Curtis, Ellis J., John C. Rose, Paul S. Mischel, and Howard Y. Chang (2024). "Extrachromosomal DNA: Biogenesis and Functions in Cancer." In: *Annual Review of Cancer Biology* 8.1, pp. 135–153. DOI: 10.1146/annurev-cancerbio-070620-092730.
- D'Agata, Erika M. C., Myrielle Dupont-Rouzeyrol, Pierre Magal, Damien Olivier, and Shigui Ruan (2008). "The Impact of Different Antibiotic Regimens on the Emergence of Antimicrobial-Resistant Bacteria." In: *PLoS ONE* 3.12, DOI: 10.1371/journal.pone.0004036.
- Dahlberg, Cecilia and Lin Chao (2003). "Amelioration of the Cost of Conjugative Plasmid Carriage in *Escherichia coli* K12." In: *Genetics* 165.4, pp. 1641–1649. DOI: 10.1093/genetics/165.4.1641.
- De Gelder, Leen, José M. Ponciano, Zaid Abdo, Paul Joyce, Larry J Forney, and Eva M. Top (2004). "Combining Mathematical Models and Statistical Methods to Understand and Predict the Dynamics of Antibiotic-Sensitive Mutants in a Population of Resistant Bacteria During Experimental Evolution." In: *Genetics* 168.3, pp. 1131–1144. DOI: 10.1534/genetics.104.033431.

- DelaFuente, Javier et al. (2022). “Within-patient evolution of plasmid-mediated antimicrobial resistance.” In: *Nature Ecology & Evolution* 6.12, pp. 1980–1991. DOI: 10.1038/s41559-022-01908-7.
- Demeke, Mekonnen M., María R. Foulquié-Moreno, Françoise Dumortier, and Johan M. Thevelein (2015). “Rapid Evolution of Recombinant *Saccharomyces cerevisiae* for Xylose Fermentation through Formation of Extrachromosomal Circular DNA.” In: *PLOS Genetics* 11.3, DOI: 10.1371/journal.pgen.1005010.
- Dewan, Ian and Hildegard Uecker (2023). “A mathematician’s guide to plasmids: an introduction to plasmid biology for modellers.” In: *Microbiology* 169.7, DOI: 10.1099/mic.0.001362.
- (2024). “Is the distribution of plasmid lengths bimodal?” In: *Plasmid* 129–130, DOI: 10.1016/j.plasmid.2024.102721.
- (2025). “Evolutionary rescue of bacterial populations by heterozygosity on multicopy plasmids.” In: *Journal of Mathematical Biology*. DOI: 10.1007/s00285-025-02182-4. Forthcoming.
- diCenzo, George C. and Turlough M. Finan (2017). “The Divided Bacterial Genome: Structure, Function, and Evolution.” In: *Microbiology and Molecular Biology Reviews* 81.3, DOI: 10.1128/mmbR.00019-17.
- Dimitriu, Tatiana (2022). “Evolution of horizontal transmission in antimicrobial resistance plasmids.” In: *Microbiology* 168.7, DOI: 10.1099/mic.0.001214.
- Dimitriu, Tatiana, Andrew C. Matthews, and Angus Buckling (2021). “Increased copy number couples the evolution of plasmid horizontal transmission and plasmid-encoded antibiotic resistance.” In: *Proceedings of the National Academy of Sciences* 118.31. DOI: 10.1073/pnas.2107818118.
- Ding, Manlin et al. (2022). “Subinhibitory antibiotic concentrations promote the horizontal transfer of plasmid-borne resistance genes from *Klebsiellae pneumoniae* to *Escherichia coli*.” In: *Frontiers in Microbiology* 13. DOI: 10.3389/fmicb.2022.1017092.
- Dionisio, F., I.C. Conceição, A.C.R. Marques, L. Fernandes, and I. Gordo (2005). “The evolution of a conjugative plasmid and its ability to increase bacterial fitness.” In: *Biology Letters* 1.2, pp. 250–252. DOI: 10.1098/rsbl.2004.0275.
- Dionisio, Francisco, Ivan Matic, Miroslav Radman, Olivia R. Rodrigues, and François Taddei (2002). “Plasmids Spread Very Fast in Heterogeneous Bacterial Communities.” In: *Genetics* 162.4, pp. 1525–1532. DOI: 10.1093/genetics/162.4.1525.
- Dionisio, Francisco, Rita Zilhão, and João Alves Gama (2019). “Interactions between plasmids and other mobile genetic elements affect their transmission and persistence.” In: *Plasmid* 102, pp. 29–36. DOI: 10.1016/j.plasmid.2019.01.003.
- Domingues, Célia P. F., João S. Rebelo, Francisca Monteiro, Teresa Nogueira, and Francisco Dionisio (2021). “Harmful behaviour through plasmid

- transfer: a successful evolutionary strategy of bacteria harbouring conjugative plasmids." In: *Philosophical Transactions of the Royal Society B* 377.1842, DOI: 10.1098/rstb.2020.0473.
- Dorado-Morales, Pedro, M. Pilar Garcillán-Barcia, Iñigo Lasa, and Cristina Solano (2021). "Fitness Cost Evolution of Natural Plasmids of *Staphylococcus aureus*." In: *mBio* 12.1, DOI: 10.1128/mbio.03094-20.
- Doumith, M., H. Dhanji, M. J. Ellington, P. Hawkey, and N. Woodford (2011). "Characterization of plasmids encoding extended-spectrum  $\beta$ -lactamases and their addiction systems circulating among *Escherichia coli* clinical isolates in the UK." In: *Journal of Antimicrobial Chemotherapy* 67.4, pp. 878–885. DOI: 10.1093/jac/dkr553.
- Eberhard, William G. (1989). "Why do bacterial plasmids carry some genes and not others?" In: *Plasmid* 21.3, pp. 167–174. DOI: 10.1016/0147-619x(89)90040-1.
- (1990). "Evolution in Bacterial Plasmids and Levels of Selection." In: *The Quarterly Review of Biology* 65.1, pp. 3–22. DOI: 10.1086/416582.
- Eggenberger, F. and G. Pólya (1923). "Über die Statistik verketteter Vorgänge." In: *Zeitschrift für Angewandte Mathematik und Mechanik* 3.4, pp. 279–289. DOI: 10.1002/zamm.19230030407.
- Erdmann, Susanne, Bernhard Tschitschko, Ling Zhong, Mark J. Raftery, and Ricardo Cavicchioli (2017). "A plasmid from an Antarctic haloarchaeon uses specialized membrane vesicles to disseminate and infect plasmid-free cells." In: *Nature Microbiology* 2.10, pp. 1446–1455. DOI: 10.1038/s41564-017-0009-2.
- Falkow, S. (1975). *Infectious Multiple Drug Resistance*. London: Pion. ISBN: 0-85086-049-4.
- Farr, Andrew D., Diego Pesce, Suman G. Das, Mark P. Zwart, and J. Arjan G. M. de Visser (2023). "The Fitness of Beta-Lactamase Mutants Depends Nonlinearly on Resistance Level at Sublethal Antibiotic Concentrations." In: *mBio* 14.3. DOI: 10.1128/mbio.00098-23.
- Fecskeová, L., M. Kovařík, P. Javorský, and P. Pristaš (2012). "Mosaic structure of the small cryptic plasmid pKST<sub>23</sub> from *Escherichia coli*." In: *Folia Microbiologica* 57.4, pp. 277–279. DOI: 10.1007/s12223-012-0124-7.
- Fernandez-Lopez, Raul, Santiago Redondo, M. Pilar Garcillan-Barcia, and Fernando de la Cruz (2017). "Towards a taxonomy of conjugative plasmids." In: *Current Opinion in Microbiology* 38, pp. 106–113. DOI: 10.1016/j.mib.2017.05.005.
- Fischer, Egil A. J., Cindy M. Dierikx, Alieda van Essen-Zandbergen, Dik Mevius, Arjan Stegeman, Francisca C. Velkers, and Don Klinkenberg (2019). "Competition between *Escherichia coli* Populations with and without Plasmids Carrying a Gene Encoding Extended-Spectrum Beta-Lactamase in the Broiler Chicken Gut." In: *Applied and Environmental Microbiology* 85.17, DOI: 10.1128/aem.00892-19.

- Fisher, R. A. (1923). "On the Dominance Ratio." In: *Proceedings of the Royal Society of Edinburgh* 42, pp. 321–341. DOI: 10.1017/S0370164600023993.
- Fox, Randal E., Xue Zhong, Stephen M. Krone, and Eva M. Top (2008). "Spatial structure and nutrients promote invasion of IncP-1 plasmids in bacterial populations." In: *The ISME Journal* 2.10, pp. 1024–1039. DOI: 10.1038/ismej.2008.53.
- Francia, M. Victoria, Athanasia Varsaki, M. Pilar Garcillán-Barcia, Amparo Latorre, Constantin Drainas, and Fernando de la Cruz (2004). "A classification scheme for mobilization regions of bacterial plasmids." In: *FEMS Microbiology Reviews* 28.1, pp. 79–100. DOI: 10.1016/j.femsre.2003.09.001.
- Frobenius, G. (1912). "Über Matrizen aus nicht negativen Elementen." In: *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, pp. 456–477.
- Froehlich, Barbara, Julian Parkhill, Mandy Sanders, Michael A. Quail, and June R. Scott (2005). "The pCoo Plasmid of Enterotoxigenic *Escherichia coli* Is a Mosaic Cointegrate." In: *Journal of Bacteriology* 187.18, pp. 6509–6516. DOI: 10.1128/jb.187.18.6509-6516.2005.
- Gama, João Alves, Elizabeth G. Aarag Fredheim, François Cléon, Ana Maria Reis, Rita Zilhão, and Francisco Dionisio (2020). "Dominance Between Plasmids Determines the Extent of Biofilm Formation." In: *Frontiers in Microbiology* 11, DOI: 10.3389/fmicb.2020.02070.
- Gama, João Alves, Rita Zilhão, and Francisco Dionisio (2017a). "Conjugation efficiency depends on intra and intercellular interactions between distinct plasmids: Plasmids promote the immigration of other plasmids but repress co-colonizing plasmids." In: *Plasmid* 93, pp. 6–16. DOI: 10.1016/j.plasmid.2017.08.003.
- (2017b). "Multiple plasmid interference — Pledging allegiance to my enemy's enemy." In: *Plasmid* 93, pp. 17–23. DOI: 10.1016/j.plasmid.2017.08.002.
- (2018). "Impact of plasmid interactions with the chromosome and other plasmids on the spread of antibiotic resistance." In: *Plasmid* 99, pp. 82–88. DOI: 10.1016/j.plasmid.2018.09.009.
- Garcillán-Barcia, M. Pilar, Santiago Redondo-Salvo, and Fernando de la Cruz (2023). "Plasmid classifications." In: *Plasmid* 126, DOI: 10.1016/j.plasmid.2023.102684.
- Garcillán-Barcia, Maria Pilar, Andrés Alvarado, and Fernando de la Cruz (2011). "Identification of bacterial plasmids based on mobility and plasmid population biology." In: *FEMS Microbiology Reviews* 35.5, pp. 936–956. DOI: 10.1111/j.1574-6976.2011.00291.x.
- Garcillán-Barcia, Maria Pilar and Fernando de la Cruz (2008). "Why is entry exclusion an essential feature of conjugative plasmids?" In: *Plasmid* 60.1, pp. 1–18. DOI: 10.1016/j.plasmid.2008.03.002.

- Garcillán-Barcia, María Pilar, María Victoria Francia, and Fernando de la Cruz (2009). "The diversity of conjugative relaxases and its application in plasmid classification." In: *FEMS Microbiology Reviews* 33.3, pp. 657–687. DOI: 10.1111/j.1574-6976.2009.00168.x.
- Garoña, Ana, Nils F. Hülter, Devani Romero Picazo, and Tal Dagan (2021). "Segregational Drift Constrains the Evolutionary Rate of Prokaryotic Plasmids." In: *Molecular Biology and Evolution*. DOI: 10.1093/molbev/msab283.
- Garoña, Ana, Mario Santer, Nils F. Hülter, Hildegard Uecker, and Tal Dagan (2023). "Segregational drift hinders the evolution of antibiotic resistance on polyploid replicons." In: *PLOS Genetics* 19.8, DOI: 10.1371/journal.pgen.1010829.
- Gaubatz, James W. (1990). "Extrachromosomal circular DNAs and genomic sequence plasticity in eukaryotic cells." In: *Mutation Research* 237, pp. 271–292. DOI: 10.1016/0921-8734(90)90009-g.
- Geoffroy, Félix and Hildegard Uecker (2023). "Limits to evolutionary rescue by conjugative plasmids." In: *Theoretical Population Biology* 154, pp. 102–117. DOI: 10.1016/j.tpb.2023.10.001.
- Gerdes, Kenn, Poul B. Rasmussen, and Søren Molin (1986). "Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells." In: *Proceedings of the National Academy of Sciences* 83.10, pp. 3116–3120. DOI: 10.1073/pnas.83.10.3116.
- Getino, María and Fernando de la Cruz (2018). "Natural and Artificial Strategies To Control the Conjugative Transmission of Plasmids." In: *Microbiology Spectrum* 6.1. DOI: 10.1128/microbiolspec.mtbp-0015-2016.
- Ghigo, Jean-Marc (2001). "Natural conjugative plasmids induce bacterial biofilm development." In: *Nature* 412.6845, pp. 442–445. DOI: 10.1038/35086581.
- Gibbs, M. D., A. J. Spiers, and P. L. Bergquist (1993). "RepFIB: A Basic Replicon of Large Plasmids." In: *Plasmid* 29.3, pp. 165–179. DOI: 10.1006/plas.1993.1020.
- Gillespie, Daniel T. (1976). "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions." In: *Journal of Computational Physics* 22.4, pp. 403–434. DOI: 10.1016/0021-9991(76)90041-3.
- Given, Cindy, Reetta Penttinen, and Matti Jalasvuori (2022). "Plasmid Viability Depends on the Ecological Setting of Hosts within a Multiplasmid Community." In: *Microbiology Spectrum* 10.2. DOI: 10.1128/spectrum.00133-22.
- Gomulkiewicz, Richard and Robert D. Holt (1995). "When Does Evolution by Natural Selection Prevent Extinction?" In: *Evolution* 49.1, pp. 201–207. DOI: 10.1111/j.1558-5646.1995.tb05971.x.
- González, Víctor et al. (2003). "The mosaic structure of the symbiotic plasmid of *Rhizobium etli* CFN42 and its relation to other symbiotic genome

- compartments.” In: *Genome Biology* 4.6, R36. DOI: 10.1186/gb-2003-4-6-r36.
- González, Víctor et al. (2006). “The partitioned *Rhizobium etli* genome: Genetic and metabolic redundancy in seven interacting replicons.” In: *Proceedings of the National Academy of Sciences* 103.10, pp. 3834–3839. DOI: 10.1073/pnas.0508502103.
- Gottig, S., T. M. Gruber, B. Stecher, T. A. Wichelhaus, and V. A. J. Kempf (2015). “In Vivo Horizontal Gene Transfer of the Carbapenemase OXA-48 During a Nosocomial Outbreak.” In: *Clinical Infectious Diseases* 60.12, pp. 1808–1815. DOI: 10.1093/cid/civ191.
- Graham, Ronald L., Donald E. Knuth, and Oren Patashnik (1994). *Concrete Mathematics. A Foundation for Computer Science*. 2nd ed. Upper Saddle River, NJ: Addison-Wesley. ISBN: 978-0-201-55802-9.
- Gresham, David, Renata Usaite, Susanne Manuela Germann, Michael Lisby, David Botstein, and Birgitte Regenberg (2010). “Adaptation to diverse nitrogen-limited environments by deletion or extrachromosomal element formation of the *GAP1* locus.” In: *Proceedings of the National Academy of Sciences* 107.43, pp. 18551–18556. DOI: 10.1073/pnas.1014023107.
- Gustafsson, Petter and Kurt Nordström (1980). “Control of Plasmid R1 Replication: Kinetics of Replication in Shifts Between Different Copy Number Levels.” In: *Journal of Bacteriology* 141.1, pp. 106–110. DOI: 10.1128/jb.141.1.106-110.1980.
- Guynet, Catherine, Ana Cuevas, Gabriel Moncalián, and Fernando de la Cruz (2011). “The *stb* Operon Balances the Requirements for Vegetative Stability and Conjugative Transfer of Plasmid R388.” In: *PLoS Genetics* 7.5, DOI: 10.1371/journal.pgen.1002073.
- Hagemann, R. (2000). “Erwin Baur or Carl Correns: who really created the theory of plastid inheritance?” In: *Journal of Heredity* 91.6, pp. 435–440. DOI: 10.1093/jhered/91.6.435.
- Haldane, J. B. S. (1927). “A Mathematical Theory of Natural and Artificial Selection, Part V: Selection and Mutation.” In: *Mathematical Proceedings of the Cambridge Philosophical Society* 23.7, pp. 838–844. DOI: 10.1017/S0305004100015644.
- Hall, James P. J., João Botelho, Adrian Cazares, and David A. Baltrus (2021a). “What makes a megaplasmid?” In: *Philosophical Transactions of the Royal Society B* 377.1842, DOI: 10.1098/rstb.2020.0472.
- Hall, James P. J., Ellie Harrison, and David A. Baltrus (2021b). “Introduction: the secret lives of microbial mobile genetic elements.” In: *Philosophical Transactions of the Royal Society B* 377.1842, DOI: 10.1098/rstb.2020.0460.
- Hall, James P. J., A. Jamie Wood, Ellie Harrison, and Michael A. Brockhurst (2016). “Source–sink plasmid transfer dynamics maintain gene mobility in soil bacterial communities.” In: *Proceedings of the National Academy of Sciences* 113.29, pp. 8260–8265. DOI: 10.1073/pnas.1600974113.

- Hall, James P. J., Rosanna C. T. Wright, David Guymer, Ellie Harrison, and Michael A. Brockhurst (2020). "Extremely fast amelioration of plasmid fitness costs by multiple functionally diverse pathways." In: *Microbiology* 166.1, pp. 56–62. DOI: 10.1099/mic.0.000862.
- Hall, James P. J., Rosanna C. T. Wright, Ellie Harrison, Katie J. Muddiman, A. Jamie Wood, Steve Paterson, and Michael A. Brockhurst (2021c). "Plasmid fitness costs are caused by specific genetic conflicts enabling resolution by compensatory mutation." In: *PLOS Biology* 19.10, DOI: 10.1371/journal.pbio.3001225.
- Hall, James P.J., Ellie Harrison, Andrew K. Lilley, Steve Paterson, Andrew J. Spiers, and Michael A. Brockhurst (2015). "Environmentally co-occurring mercury resistance plasmids are genetically and phenotypically diverse and confer variable context-dependent fitness effects." In: *Environmental Microbiology* 17.12, pp. 5008–5022. DOI: 10.1111/1462-2920.12901.
- Halleran, Andrew D., Emanuel Flores-Bautista, and Richard M. Murray (2019). *Quantitative characterization of random partitioning in the evolution of plasmid-encoded traits*. BioRxiv preprint. DOI: 10.1101/594879.
- Hanahan, Douglas (1983). "Studies on transformation of *Escherichia coli* with plasmids." In: *Journal of Molecular Biology* 166.4, pp. 557–580. DOI: 10.1016/S0022-2836(83)80284-8.
- Harr, Bettina and Christian Schlötterer (2006). "Gene Expression Analysis Indicates Extensive Genotype-specific Crosstalk between the Conjugative F-plasmid and the *E. coli* Chromosome." In: *BMC Microbiology* 6, DOI: 10.1186/1471-2180-6-80.
- Harrison, Ellie and Michael A. Brockhurst (2012). "Plasmid-mediated horizontal gene transfer is a coevolutionary process." In: *Trends in Microbiology* 20.6, pp. 262–267. DOI: 10.1016/j.tim.2012.04.003.
- Harrison, Ellie, David Guymer, Andrew J. Spiers, Steve Paterson, and Michael A. Brockhurst (2015). "Parallel Compensatory Evolution Stabilizes Plasmids across the Parasitism-Mutualism Continuum." In: *Current Biology* 25.15, pp. 2034–2039. DOI: 10.1016/j.cub.2015.06.024.
- Helinski, Donald R. (2022). "A Brief History of Plasmids." In: *EcoSal Plus* 10.1. DOI: 10.1128/ecosalplus.esp-0028-2021.
- Hernandez-Beltran, J. Carlos R., Verónica Miró Pina, Arno Siri-Jégousse, Sandra Palau, Rafael Peña-Miller, and Adrián González Casanova (2022). "Segregational instability of multicopy plasmids: A population genetics approach." In: *Ecology and Evolution* 12.12. DOI: 10.1002/ece3.9469.
- Hernández-Beltrán, José Carlos Ramón, Alvaro San Millán, Ayari Fuentes-Hernández, and Rafael Peña-Miller (2021). "Mathematical Models of Plasmid Population Dynamics." In: *Frontiers in Microbiology* 12. DOI: 10.3389/fmicb.2021.606396.
- Heuer, Holger, Randal E. Fox, and Eva M. Top (2007). "Frequent conjugative transfer accelerates adaptation of a broad-host-range plasmid to an

- unfavorable *Pseudomonas putida* host.” In: *FEMS Microbiology Ecology* 59.3, pp. 738–748. DOI: 10.1111/j.1574-6941.2006.00223.x.
- Hoeven, Nelly van der (1984). “A mathematical model for the co-existence of incompatible, conjugative plasmids in individual bacteria of a bacterial population.” In: *Journal of Theoretical Biology* 110.3, pp. 411–423. DOI: 10.1016/S0022-5193(84)80183-6.
- (1985). “Evolution of bacterial surface exclusion against incompatible plasmids.” In: *Journal of Theoretical Biology* 117.3, pp. 431–452. DOI: 10.1016/S0022-5193(85)80153-3.
- (1986). “Coexistence of incompatible plasmids in a bacterial population living under a feast and famine regime.” In: *Journal of Mathematical Biology* 24.3, pp. 313–325. DOI: 10.1007/BF00275640.
- Holt, Sylvester, Gerard Arrey, and Birgitte Regenberg (2023). “Did circular DNA shape the evolution of mammalian genomes?” In: *Trends in Biochemical Sciences* 48.4, pp. 317–320. DOI: 10.1016/j.tibs.2022.09.010.
- Horne, Tanya, Victoria T. Orr, and James P. J. Hall (2023). “How do interactions between mobile genetic elements affect horizontal gene transfer?” In: *Current Opinion in Microbiology* 73, DOI: 10.1016/j.mib.2023.102282.
- Hotta, Yasuo and Alix Bassel (1965). “Molecular Size and Circularity of DNA in Cells of Mammals and Higher Plants.” In: *Proceedings of the National Academy of Sciences* 53.2, pp. 356–362. DOI: 10.1073/pnas.53.2.356.
- Huisman, Jana S., Fabienne Benz, Sarah J.N. Duxbury, J. Arjan G.M. de Visser, Alex R. Hall, Egil A.J. Fischer, and Sebastian Bonhoeffer (2022). “Estimating plasmid conjugation rates: A new computational tool and a critical comparison of methods.” In: *Plasmid* 121, DOI: 10.1016/j.plasmid.2022.102627.
- Hull, Ryan M. and Jonathan Houseley (2020). “The adaptive potential of circular DNA accumulation in ageing cells.” In: *Current Genetics* 66.5, pp. 889–894. DOI: 10.1007/s00294-020-01069-9.
- Hull, Ryan M., Michelle King, Grazia Pizza, Felix Krueger, Xabier Vergara, and Jonathan Houseley (2019). “Transcription-induced formation of extrachromosomal DNA during yeast ageing.” In: *PLOS Biology* 17.12. DOI: 10.1371/journal.pbio.3000471.
- Hülter, Nils F., Judith Ilhan, Tanita Wein, A Samer Kadibalban, Katrin Hamerschmidt, and Tal Dagan (2017). “An evolutionary perspective on plasmid lifestyle modes.” In: *Current Opinion in Microbiology* 38, pp. 74–80. DOI: 10.1016/j.mib.2017.05.001.
- Hülter, Nils F., Tanita Wein, Johannes Effe, Ana Garoña, and Tal Dagan (2020). “Intracellular Competitions Reveal Determinants of Plasmid Evolutionary Success.” In: *Frontiers in Microbiology* 11. DOI: 10.3389/fmicb.2020.02062.

- Humphrey, Suzanne, Álvaro San Millán, Macarena Toll-Riera, John Connolly, Alejandra Flor-Duro, John Chen, Carles Ubeda, R. Craig MacLean, and José R. Penadés (2021). “Staphylococcal phages and pathogenicity islands drive plasmid evolution.” In: *Nature Communications* 12.1. DOI: 10.1038/s41467-021-26101-5.
- Hung, King L., Paul S. Mischel, and Howard Y. Chang (2022). “Gene regulation on extrachromosomal DNA.” In: *Nature Structural & Molecular Biology* 29.8, pp. 736–744. DOI: 10.1038/s41594-022-00806-7.
- Hung, King L. et al. (July 2023). *Coordinated inheritance of extrachromosomal DNA species in human cancer cells*. BioRxiv preprint. DOI: 10.1101/2023.07.18.549597.
- Igler, Claudia, Jana S. Huisman, Berit Siedentop, Sebastian Bonhoeffer, and Sonja Lehtinen (2021a). “Plasmid co-infection: linking biological mechanisms to ecological and evolutionary dynamics.” In: *Philosophical Transactions of the Royal Society B* 377.1842, DOI: 10.1098/rstb.2020.0478.
- Igler, Claudia, Lukas Schwyter, Daniel Gehrig, and Carolin Charlotte Wendling (2021b). “Conjugative plasmid transfer is limited by prophages but can be overcome by high conjugation rates.” In: *Philosophical Transactions of the Royal Society B* 377.1842, DOI: 10.1098/rstb.2020.0470.
- Ilhan, Judith, Anne Kupczok, Christian Woehle, Tanita Wein, Nils F. Hülter, Philip Rosenstiel, Giddy Landan, Itzhak Mizrahi, and Tal Dagan (2018). “Segregational Drift and the Interplay between Plasmid Copy Number and Evolvability.” In: *Molecular Biology and Evolution* 36.3, pp. 472–486. DOI: 10.1093/molbev/msy225.
- Ishii, Kazushige, Tamotsu Hashimoto-Gotoh, and Kenichi Matsubara (1978). “Random replication and random assortment model for plasmid incompatibility in bacteria.” In: *Plasmid* 1.4, pp. 435–445. DOI: 10.1016/0147-619x(78)90002-1.
- Ismail, Emadeldeen et al. (2014). “A novel plasmid pEA68 of *Erwinia amylovora* and the description of a new family of plasmids.” In: *Archives of Microbiology* 196.12, pp. 891–899. DOI: 10.1007/s00203-014-1028-5.
- Jaskólska, Milena, David W. Adams, and Melanie Blokesch (2022). “Two defence systems eliminate plasmids from seventh pandemic *Vibrio cholerae*.” In: *Nature* 604, pp. 323–329. DOI: 10.1038/s41586-022-04546-y.
- Jensen, Rasmus Bugge and Kenn Gerdes (1995). “Programmed cell death in bacteria: proteic plasmid stabilization systems.” In: *Molecular Microbiology* 17.2, pp. 205–210. DOI: 10.1111/j.1365-2958.1995.mmi\_17020205.x.
- Kamruzzaman, Muhammad and Jonathan R. Iredell (2020). “CRISPR-Cas System in Antibiotic Resistance Plasmids in *Klebsiella pneumoniae*.” In: *Frontiers in Microbiology* 10. DOI: 10.3389/fmicb.2019.02934.
- Kelsic, Eric D., Jeffrey Zhao, Kalin Vetsigian, and Roy Kishony (2015). “Counteraction of antibiotic production and degradation stabilizes microbial communities.” In: *Nature* 521, pp. 516–519. DOI: 10.1038/nature14485.

- Kimmel, M. and D. E. Axelrod (1990). "Mathematical models of gene amplification with applications to cellular drug resistance and tumorigenicity." In: *Genetics* 125.3, pp. 633–644. DOI: 10.1093/genetics/125.3.633.
- Kimmel, Marek, David E. Axelrod, and Geoffrey M. Wahl (1992). "A branching process model of gene amplification following chromosome breakage." In: *Mutation Research* 276.3, pp. 225–239. DOI: 10.1016/0165-1110(92)90010-7.
- King, Grant A., Jay S. Goodman, Jennifer G. Schick, Keerthana Chetlapalli, Danielle M. Jorgens, Kent L. McDonald, and Elçin Ünal (2019). "Meiotic cellular rejuvenation is coupled to nuclear remodeling in budding yeast." In: *eLife* 8. DOI: 10.7554/eLife.47156.
- Kinoshita, Yasuhiro, Noboru Ohnishi, Yasuyuki Yamada, Takahiro Kunisada, and Hideo Yamagishi (1985). "Extrachromosomal Circular DNA from Nuclear Fraction of Higher Plants." In: *Plant and Cell Physiology* 26.7, pp. 1401–1409. DOI: 10.1093/oxfordjournals.pcp.a077040.
- Kiyosawa, Hidenori, Joanne E. Hughes, Gregory J. Podgorski, and Dennis L. Welker (1993). "Small Circular Plasmids of the Eukaryote *Dicystostelium purpureum* Define Two Novel Plasmid Families." In: *Plasmid* 30.2, pp. 106–118. DOI: 10.1006/plas.1993.1038.
- Kollek, Regine, Wolfgang Oertel, and Werner Goebel (1978). "Isolation and characterization of the minimal fragment required for autonomous replication ("Basic replicon") of a copy mutant (pKN102) of the antibiotic resistance factor R1." In: *Molecular and General Genetics* 162.1, pp. 51–57. DOI: 10.1007/bf00333850.
- Koraimann, Günther and Maria A. Wagner (2014). "Social behavior and decision making in bacterial conjugation." In: *Frontiers in Cellular and Infection Microbiology* 4. DOI: 10.3389/fcimb.2014.00054.
- Kosterlitz, Olivia and Jana S. Huisman (2023). "Guidelines for the estimation and reporting of plasmid conjugation rates." In: *Plasmid* 126, DOI: 10.1016/j.plasmid.2023.102685.
- Kosterlitz, Olivia, Adamaris Muñoz Tirado, Claire Wate, Clint Elg, Ivana Bozic, Eva M. Top, and Benjamin Kerr (2022). "Estimating the transfer rates of bacterial plasmids with an adapted Luria-Delbrück fluctuation analysis." In: *PLOS Biology* 20.7, DOI: 10.1371/journal.pbio.3001732.
- Kothari, Ankita, Yu-Wei Wu, John-Marc Chandonia, Marimikel Charrier, Lara Rajeev, Andrea M. Rocha, Dominique C. Joyner, Terry C. Hazen, Steven W. Singer, and Aindrila Mukhopadhyay (2019). "Large Circular Plasmids from Groundwater Plasmidomes Span Multiple Incompatibility Groups and Are Enriched in Multimetal Resistance Genes." In: *mBio* 10.1, DOI: 10.1128/mbio.02899-18.
- Krone, Stephen M., Ruinan Lu, Randal Fox, Haruo Suzuki, and Eva M. Top (2007). "Modelling the spatial dynamics of plasmid transfer and persis-

- tence." In: *Microbiology* 153.8, pp. 2803–2816. DOI: 10.1099/mic.0.2006/004531-0.
- Kusano, K., T. Naito, N. Handa, and I. Kobayashi (1995). "Restriction-modification systems as genomic parasites in competition for specific sequences." In: *Proceedings of the National Academy of Sciences* 92.24, pp. 11095–11099. DOI: 10.1073/pnas.92.24.11095.
- Lamberte, Lisa E., Gabriele Baniulyte, Shivani S. Singh, Anne M. Stringer, Richard P. Bonocora, Mathew Stracy, Achillefs N. Kapanidis, Joseph T. Wade, and David C. Grainger (2017). "Horizontally acquired AT-rich genes in *Escherichia coli* cause toxicity by sequestering RNA polymerase." In: *Nature Microbiology* 2.3. DOI: 10.1038/nmicrobiol.2016.249.
- Lange, Joshua T. et al. (2022). "The evolutionary dynamics of extrachromosomal DNA in human cancers." In: *Nature Genetics* 54.10, pp. 1527–1533. DOI: 10.1038/s41588-022-01177-x.
- Lau, Billy T.C., Per Malkus, and Johan Paulsson (2013). "New quantitative methods for measuring plasmid loss rates reveal unexpected stability." In: *Plasmid* 70.3, pp. 353–361. DOI: 10.1016/j.plasmid.2013.07.007.
- Leclerc, Quentin J., Jodi A. Lindsay, and Gwenan M. Knight (2019). "Mathematical modelling to study the horizontal transfer of antimicrobial resistance genes in bacteria: current state of the field and recommendations." In: *Journal of The Royal Society Interface* 16.157, p. 20190260. DOI: 10.1098/rsif.2019.0260.
- Lederberg, Joshua (1952). "Cell Genetics and Hereditary Symbiosis." In: *Physiological Reviews* 32.4, pp. 403–430. DOI: 10.1152/physrev.1952.32.4.403.
- Lee, Hsuan-Hsuan, Chin-Chen Hsu, Yen-Ling Lin, and Carton W. Chen (2011). "Linear plasmids mobilize linear but not circular chromosomes in *Streptomyces*: support for the 'end first' model of conjugal transfer." In: *Microbiology* 157.9, pp. 2556–2568. DOI: 10.1099/mic.0.051441-0.
- Lee, Sun Bok and James E. Bailey (1984). "A mathematical model for  $\lambda$ dv plasmid replication: Analysis of wild-type plasmid." In: *Plasmid* 11.2, pp. 151–165. DOI: 10.1016/0147-619x(84)90020-9.
- Lehtinen, Sonja, Jana S. Huisman, and Sebastian Bonhoeffer (2021). "Evolutionary mechanisms that determine which bacterial genes are carried on plasmids." In: *Evolution Letters* 5.3, pp. 290–301. DOI: 10.1002/evl3.226.
- León-Sampedro, Ricardo et al. (2021). "Pervasive transmission of a carbapenem resistance plasmid in the gut microbiota of hospitalized patients." In: *Nature Microbiology* 6.5, pp. 606–616. DOI: 10.1038/s41564-021-00879-y.
- Levin, Bruce R. and Frank M. Stewart (1980). "The Population Biology of Bacterial Plasmids: *a priori* Conditions for the Existence of Mobilizable Nonconjugative Factors." In: *Genetics* 94.2, pp. 425–443. DOI: 10.1093/genetics/94.2.425.

- Levin, Bruce R., Frank M. Stewart, and Virginia A. Rice (1979). "The kinetics of conjugative plasmid transmission: Fit of a simple mass action model." In: *Plasmid* 2.2, pp. 247–260. DOI: 10.1016/0147-619x(79)90043-x.
- Li, Liguang, Arnaud Dechesne, Jonas Stenlørkke Madsen, Joseph Nesme, Søren J. Sørensen, and Barth F. Smets (2020). "Plasmids persist in a microbial community by providing fitness benefit to multiple phylotypes." In: *The ISME Journal* 14.5, pp. 1170–1181. DOI: 10.1038/s41396-020-0596-4.
- Libuda, Diana E. and Fred Winston (2006). "Amplification of histone genes by circular chromosome formation in *Saccharomyces cerevisiae*." In: *Nature* 443.7114, pp. 1003–1007. DOI: 10.1038/nature05205.
- Lili, Loukia N., Nicholas F. Britton, and Edward J. Feil (2007). "The Persistence of Parasitic Plasmids." In: *Genetics* 177.1, pp. 399–405. DOI: 10.1534/genetics.107.077420.
- Liu, Ying-ying, Dan-dan He, Meng-ke Zhang, Yu-shan Pan, Hua Wu, Li Yuan, Jian-hua Liu, and Gong-zheng Hu (2021). "The Formation of Two Hybrid Plasmids Mediated by IS26 and Tn6952 in *Salmonella enterica* Serotype *Enteritidis*." In: *Frontiers in Microbiology* 12. DOI: 10.3389/fmicb.2021.676574.
- Loftie-Eaton, Wesley et al. (2015). "Evolutionary Paths That Expand Plasmid Host-Range: Implications for Spread of Antibiotic Resistance." In: *Molecular Biology and Evolution* 33.4, pp. 885–897. DOI: 10.1093/molbev/msv339.
- Lopatkin, Allison J., Hannah R. Meredith, Jaydeep K. Srimani, Connor Pfeiffer, Rick Durrett, and Lingchong You (2017). "Persistence and reversal of plasmid-mediated antibiotic resistance." In: *Nature Communications* 8.1. DOI: 10.1038/s41467-017-01532-1.
- Lopez, Jaime G., Mohamed S. Donia, and Ned S. Wingreen (2021). "Modeling the ecology of parasitic plasmids." In: *The ISME Journal* 15.10, pp. 2843–2852. DOI: 10.1038/s41396-021-00954-6.
- Lundquist, Peter D. and Bruce R. Levin (1986). "Transitory Derepression and the Maintenance of Conjugative Plasmids." In: *Genetics* 113.3, pp. 483–497. DOI: 10.1093/genetics/113.3.483.
- Luria, S. E. and M. Delbrück (1943). "Mutations of Bacteria from Virus Sensitivity to Virus Resistance." In: *Genetics* 28.6, pp. 491–511. DOI: 10.1093/genetics/28.6.491.
- Machida, Yasunori, Chiyoko Machida, Hisako Ohtsubo, and Eiichi Ohtsubo (1982). "Factors determining frequency of plasmid cointegration mediated by insertion sequence IS1." In: *Proceedings of the National Academy of Sciences* 79.2, pp. 277–281. DOI: 10.1073/pnas.79.2.277.
- MacLean, R. Craig and Alvaro San Millan (2015). "Microbial Evolution: Towards Resolving the Plasmid Paradox." In: *Current Biology* 25.17, R764–R767. DOI: 10.1016/j.cub.2015.07.006.

- Maddamsetti, Rohan, Maggie L. Wilson, Hye-In Son, Zhengqing Zhou, Jia Lu, and Lingchong You (2024). *Scaling laws of plasmids across the microbial tree of life*. BioRxiv preprint. DOI: 10.1101/2024.10.04.616653.
- Madgwick, Philip G., Thomas Tunstall, and Ricardo Kanitz (2024). "Evolutionary rescue in resistance to pesticides." In: *Proceedings of the Royal Society B* 291.2025. DOI: 10.1098/rspb.2024.0805.
- Madsen, Jonas Stenlørkke, Mette Burmølle, Lars Hestbjerg Hansen, and Søren Johannes Sørensen (2012). "The interconnection between biofilm formation and horizontal gene transfer." In: *FEMS Immunology & Medical Microbiology* 65.2, pp. 183–195. DOI: 10.1111/j.1574-695x.2012.00960.x.
- Magnuson, Roy David (2007). "Hypothetical Functions of Toxin-Antitoxin Systems." In: *Journal of Bacteriology* 189.17, pp. 6089–6092. DOI: 10.1128/jb.00958-07.
- Maida, Isabel, Marco Fondi, Valerio Orlandini, Giovanni Emiliani, Maria Cristiana Papaleo, Elena Perrin, and Renato Fani (2014). "Origin, duplication and reshuffling of plasmid genes: Insights from *Burkholderia vietnamiensis* G4 genome." In: *Genomics* 103.2-3, pp. 229–238. DOI: 10.1016/j.ygeno.2014.02.004.
- Malaka De Silva, P., George E. Stenhouse, Grace A. Blackwell, Rebecca J. Bengtsson, Claire Jenkins, James P. J. Hall, and Kate S. Baker (2022). "A tale of two plasmids: contributions of plasmid associated phenotypes to epidemiological success among *Shigella*." In: *Proceedings of the Royal Society B* 289.1980, DOI: 10.1098/rspb.2022.0581.
- Martin, Guillaume, Robin Aguilée, Johan Ramsayer, Oliver Kaltz, and Ophélie Ronce (2013). "The probability of evolutionary rescue: towards a quantitative comparison between theory and evolution experiments." In: *Philosophical Transactions of the Royal Society B* 368.1610, DOI: 10.1098/rstb.2012.0088.
- Mc Ginty, Sorchia E., Daniel J. Rankin, and Sam P. Brown (2010). "Horizontal Gene Transfer and the Evolution of Bacterial Cooperation." In: *Evolution* 65.1, pp. 21–32. DOI: 10.1111/j.1558-5646.2010.01121.x.
- McClintock, Barbara (1932). "A Correlation of Ring-Shaped Chromosomes with Variegation in *Zea mays*." In: *Proceedings of the National Academy of Sciences* 18.12, pp. 677–681. DOI: 10.1073/pnas.18.12.677.
- Meer, Jan Roelof van der (1994). "Genetic adaptation of bacteria to chlorinated aromatic compounds." In: *FEMS Microbiology Reviews* 15.2-3, pp. 239–249. DOI: 10.1111/j.1574-6976.1994.tb00137.x.
- Mei, Han, Barbara Arbehther, Marzia A Cremona, Michael DeGiorgio, and Anton Nekrutenko (2019). "A High-Resolution View of Adaptive Event Dynamics in a Plasmid." In: *Genome Biology and Evolution* 11.10, pp. 3022–3034. DOI: 10.1093/gbe/evz197.
- Metzger, Genevieve A., Benjamin J. Ridenhour, Michael France, Karol Gliniewicz, Jack Millstein, Matthew L. Settles, Larry J. Forney, Thibault Stal-

- der, and Eva M. Top (2022). "Biofilms preserve the transmissibility of a multi-drug resistance plasmid." In: *npj Biofilms and Microbiomes* 8.1. DOI: 10.1038/s41522-022-00357-1.
- Milliken, Christine E. and R. C. Clowes (1973). "Molecular Structure of an R Factor, Its Component Drug-Resistance Determinants and Transfer Factor." In: *Journal of Bacteriology* 113.2, pp. 1026–1033. DOI: 10.1128/jb.113.2.1026-1033.1973.
- Miyakoshi, Masatoshi, Masaki Shintani, Kei Inoue, Tsuguno Terabayashi, Fumiya Sai, Moriya Ohkuma, Hideaki Nojiri, Yuji Nagata, and Masataka Tsuda (2012). "ParI, an orphan ParA family protein from *Pseudomonas putida* KT2440-specific genomic island, interferes with the partition system of IncP-7 plasmids." In: *Environmental Microbiology* 14.11, pp. 2946–2959. DOI: 10.1111/j.1462-2920.2012.02861.x.
- Mode, Charles J. (1971). *Multitype Branching Processes. Theory and Applications*. Ed. by Richard Bellman. Modern Analytic and Computational Methods in Science and Mathematics. New York: American Elsevier Publishing Company, Inc. ISBN: 0-444-00086-0.
- Møller, Henrik D., Kaj S. Andersen, and Birgitte Regenberg (2013). "A model for generating several adaptive phenotypes from a single genetic event. *Saccharomyces cerevisiae* GAP1 as a potential bet-hedging switch." In: *Communicative & Integrative Biology* 6.3, DOI: 10.4161/cib.23933.
- Møller, Henrik Devitt, Lance Parsons, Tue S. Jørgensen, David Botstein, and Birgitte Regenberg (2015). "Extrachromosomal circular DNA is common in yeast." In: *Proceedings of the National Academy of Sciences* 112.24. DOI: 10.1073/pnas.1508825112.
- Møller, Henrik Devitt et al. (2018). "Circular DNA elements of chromosomal origin are common in healthy human somatic tissue." In: *Nature Communications* 9.1. DOI: 10.1038/s41467-018-03369-8.
- Mongold, Judith A. (1992). "Theoretical Implications for the Evolution of Postsegregational Killing by Bacterial Plasmids." In: *The American Naturalist* 139.4, pp. 677–689. DOI: 10.1086/285352.
- Moran, P. A. P. (1958). "Random processes in genetics." In: *Mathematical Proceedings of the Cambridge Philosophical Society* 54.1, pp. 60–71. DOI: 10.1017/S0305004100033193.
- Moran, Robert A. and Ruth M. Hall (2019). "pBuzz: A cryptic rolling-circle plasmid from a commensal *Escherichia coli* has two inversely oriented *oriTs* and is mobilised by a B/O plasmid." In: *Plasmid* 101, pp. 10–19. DOI: 10.1016/j.plasmid.2018.11.001.
- Mortimer, Robert K. and John R. Johnston (1959). "Life Span of Individual Yeast Cells." In: *Nature* 183.4677, pp. 1751–1752. DOI: 10.1038/1831751a0.
- Mukhtar, Salma, Samia Ahmad, Aftab Bashir, Samina Mehnaz, Muhammad Sajjad Mirza, and Kauser Abdulla Malik (2019). "Identification of plasmid encoded osmoregulatory genes from halophilic bacteria isolated

- from the rhizosphere of halophytes." In: *Microbiological Research* 228, DOI: 10.1016/j.micres.2019.126307.
- Münch, Karin, Richard Münch, Rebekka Biedendieck, Dieter Jahn, and Johannes Müller (2019). "Evolutionary model for the unequal segregation of high copy plasmids." In: *PLOS Computational Biology* 15.3, DOI: 10.1371/journal.pcbi.1006724.
- Murray, Andrew W. and Jack W. Szostak (1983). "Pedigree analysis of plasmid segregation in yeast." In: *Cell* 34.3, pp. 961–970. DOI: 10.1016/0092-8674(83)90553-6.
- Naito, Taku, Kohji Kusano, and Ichizo Kobayashi (1995). "Selfish Behavior of Restriction-Modification Systems." In: *Science* 267.5199, pp. 897–899. DOI: 10.1126/science.7846533.
- Ng, Jérôme Wong, Didier Chatenay, Jérôme Robert, and Michael Guy Poirier (2010). "Plasmid copy number noise in monoclonal populations of bacteria." In: *Physical Review E* 81.1, p. 011909. DOI: 10.1103/physreve.81.011909.
- Nogueira, Teresa, Daniel J. Rankin, Marie Touchon, François Taddei, Sam P. Brown, and Eduardo P.C. Rocha (2009). "Horizontal Gene Transfer of the Secretome Drives the Evolution of Bacterial Cooperation and Virulence." In: *Current Biology* 19.20, pp. 1683–1691. DOI: 10.1016/j.cub.2009.08.056.
- Nordström, Kurt (2006). "Plasmid R<sub>1</sub>—Replication and its control." In: *Plasmid* 55.1, pp. 1–26. DOI: 10.1016/j.plasmid.2005.07.002.
- Nordström, Kurt and Helle Aagaard-Hansen (1984). "Maintenance of bacterial plasmids: Comparison of theoretical calculations and experiments with plasmid R<sub>1</sub> in *Escherichia coli*." In: *Molecular and General Genetics MGG* 197.1, pp. 1–7. DOI: 10.1007/bf00327915.
- Nordström, Kurt and Kenn Gerdes (2003). "Clustering versus random segregation of plasmids lacking a partitioning function: a plasmid paradox?" In: *Plasmid* 50.2, pp. 95–101. DOI: 10.1016/S0147-619X(03)00056-8.
- Novick, R. P. (1987). "Plasmid incompatibility." In: *Microbiological Reviews* 51.4, pp. 381–395. DOI: 10.1128/mr.51.4.381-395.1987.
- Novick, R. P., R. C. Clowes, S. N. Cohen, R. Curtiss, N. Datta, and S. Falkow (1976). "Uniform nomenclature for bacterial plasmids: a proposal." In: *Bacteriological Reviews* 40.1, pp. 168–189. DOI: 10.1128/br.40.1.168-189.1976.
- Novick, Richard P. and F.C. Hoppensteadt (1978). "On plasmid incompatibility." In: *Plasmid* 1.4, pp. 421–434. DOI: 10.1016/0147-619X(78)90001-X.
- Novozhilov, Artem S., Georgy P. Karev, and Eugene V. Koonin (2005). "Mathematical Modeling of Evolution of Horizontally Transferred Genes." In: *Molecular Biology and Evolution* 22.8, pp. 1721–1732. DOI: 10.1093/molbev/msi167.

- Nugent, M. E. and R. W. Hedges (1979). "Recombinant Plasmids Formed in vivo Carrying and Expressing Two Incompatibility Regions." In: *Journal of General Microbiology* 114.2, pp. 467–470. DOI: 10.1099/00221287-114-2-467.
- Ohtsubo, E., M. Zenilman, H. Ohtsubo, M. McCormick, C. Machida, and Y. Machida (1981). "Mechanism of Insertion and Cointegration Mediated by IS<sub>1</sub> and Tn<sub>3</sub>." In: *Cold Spring Harbor Symposia on Quantitative Biology*. Vol. 45, pp. 283–295. DOI: 10.1101/sqb.1981.045.01.041.
- Ojala, Ville, Jarkko Laitalainen, and Matti Jalasvuori (2013). "Fight evolution with evolution: plasmid-dependent phages with a wide host range prevent the spread of antibiotic resistance." In: *Evolutionary Applications* 6.6, pp. 925–932. DOI: 10.1111/eva.12076.
- Ojala, Ville, Sari Mattila, Ville Hoikkala, Jaana KH Bamford, Teppo Hiltunen, and Matti Jalasvuori (Aug. 2016). "Scoping the effectiveness and evolutionary obstacles in using plasmid-dependent phages to fight antibiotic resistance." In: *Future Microbiology* 11.8, pp. 999–1009. DOI: 10.2217/fmb-2016-0038.
- Orazi, Giulia and George A. O'Toole (2019). "'It Takes a Village': Mechanisms Underlying Antimicrobial Recalcitrance of Polymicrobial Biofilms." In: *Journal of Bacteriology* 202.1, DOI: 10.1128/jb.00530-19.
- Orr, H. Allen and Robert L. Unckless (2008). "Population Extinction and the Genetics of Adaptation." In: *The American Naturalist* 172.2, pp. 160–169. DOI: 10.1086/589460.
- (2014). "The Population Genetics of Evolutionary Rescue." In: *PLoS Genetics* 10.8, DOI: 10.1371/journal.pgen.1004551.
- Osborn, A. Mark, Fernanda M. da Silva Tatley, Lafras M. Steyn, Roger W. Pickup, and Jon R. Saunders (2000). "Mosaic plasmids and mosaic replicons: evolutionary lessons from the analysis of genetic diversity in IncFII-related replicons." In: *Microbiology* 146.9, pp. 2267–2275. DOI: 10.1099/00221287-146-9-2267.
- Palomino, Alana, Danya Gewurz, Lela DeVine, Ujana Zajmi, Jenifer Moralez, Fatima Abu-Rumman, Robert P. Smith, and Allison J. Lopatkin (2022). "Metabolic genes on conjugative plasmids are highly prevalent in *Escherichia coli* and can protect against antibiotic treatment." In: *The ISME Journal*. DOI: 10.1038/s41396-022-01329-1.
- Pandeeti, Emmanuel Vijay Paul et al. (2012). "Multiple Mechanisms Contribute to Lateral Transfer of an Organophosphate Degradation (*opd*) Island in *Sphingobium fuliginis* ATCC 27551." In: *G3 Genes|Genomes|Genetics* 2.12, pp. 1541–1554. DOI: 10.1534/g3.112.004051.
- Paulsen, Teresa, Pankaj Kumar, M. Murat Koseoglu, and Anindya Dutta (2018). "Discoveries of Extrachromosomal Circles of DNA in Normal and Tumor Cells." In: *Trends in Genetics* 34.4, pp. 270–278. DOI: 10.1016/j.tig.2017.12.010.

- Paulsson, Johan (2002). "Multileveled Selection on Plasmid Replication." In: *Genetics* 161.4, pp. 1373–1384. DOI: 10.1093/genetics/161.4.1373.
- Petersen, Jörn (2011). "Phylogeny and compatibility: plasmid classification in the genomics era." In: *Archives of Microbiology* 193, pp. 313–321. DOI: 10.1007/s00203-011-0686-9.
- Pilosof, Shai (2023). "Conceptualizing microbe–plasmid communities as complex adaptive systems." In: *Trends in Microbiology* 31.7, pp. 672–680. DOI: 10.1016/j.tim.2023.01.007.
- Pinilla-Redondo, Rafael, David Mayo-Muñoz, Jakob Russel, Roger A. Garrett, Lennart Randau, Søren J. Sørensen, and Shiraz A. Shah (2019). "Type IV CRISPR-Cas systems are highly diverse and involved in competition between plasmids." In: *Nucleic Acids Research* 48.4, pp. 2000–2012. DOI: 10.1093/nar/gkz1197.
- Pinilla-Redondo, Rafael, Jakob Russel, David Mayo-Muñoz, Shiraz A. Shah, Roger A. Garrett, Joseph Nesme, Jonas S. Madsen, Peter C. Fineran, and Søren J. Sørensen (2021). "CRISPR-Cas systems are widespread accessory elements across bacterial and archaeal plasmids." In: *Nucleic Acids Research* 50.8, pp. 4315–4328. DOI: 10.1093/nar/gkab859.
- Platt, Thomas G., Elise R. Morton, Ian S. Barton, James D. Bever, and Clay Fuqua (2014). "Ecological dynamics and complex interactions of *Agrobacterium* megaplasmids." In: *Frontiers in Plant Science* 5, DOI: 10.3389/fpls.2014.00635.
- Ponciano, José M., Leen De Gelder, Eva M. Top, and Paul Joyce (2007). "The Population Biology of Bacterial Plasmids: A Hidden Markov Model Approach." In: *Genetics* 176.2, pp. 957–968. DOI: 10.1534/genetics.106.061937.
- Porse, Andreas, Kristian Schønning, Christian Munck, and Morten O. A. Sommer (2016). "Survival and Evolution of a Large Multidrug Resistance Plasmid in New Clinical Bacterial Hosts." In: *Molecular Biology and Evolution* 33.11, pp. 2860–2873. DOI: 10.1093/molbev/msw163.
- Portnoy, D. A. and R. J. Martinez (1985). "Role of a Plasmid in the Pathogenicity of *Yersinia* Species." In: *Genetic Approaches to Microbial Pathogenicity*. Ed. by W. Goebel. Current Topics in Microbiology and Immunology 118. Berlin: Springer-Verlag, pp. 29–51. ISBN: 978-3-642-70588-5. DOI: 10.1007/978-3-642-70586-1\_3.
- Possoz, Christophe, Carin Ribard, Josette Gagnat, Jean-Luc Pernodet, and Michel Guérineau (2008). "The integrative element pSAM2 from *Streptomyces*: kinetics and mode of conjugal transfer." In: *Molecular Microbiology* 42.1, pp. 159–166. DOI: 10.1046/j.1365-2958.2001.02618.x.
- Prada-Luengo, Iñigo, Henrik D. Møller, Rasmus A. Henriksen, Qian Gao, Camilla Eggert Larsen, Sefa Alizadeh, Lasse Maretty, Jonathan Houseley, and Birgitte Regenber (2020). "Replicative aging is associated with loss of genetic heterogeneity from extrachromosomal circular DNA in

- Saccharomyces cerevisiae*." In: *Nucleic Acids Research* 48.14, pp. 7883–7898. DOI: 10.1093/nar/gkaa545.
- Prangishvili, David et al. (1998). "Conjugation in Archaea: Frequent Occurrence of Conjugative Plasmids in *Sulfolobus*." In: *Plasmid* 40.3, pp. 190–202. DOI: 10.1006/plas.1998.1363.
- Price, Valerie J., Wenwen Huo, Ardalan Sharifi, and Kelli L. Palmer (2016). "CRISPR-Cas and Restriction-Modification Act Additively against Conjugative Antibiotic Resistance Plasmid Transfer in *Enterococcus faecalis*." In: *mSphere* 1.3, DOI: 10.1128/msphere.00064-16.
- Projan, Steven J., Michel Monod, C. S. Narayanan, and David Dubnau (1987). "Replication properties of pIM13, a naturally occurring plasmid found in *Bacillus subtilis*, and of its close relative pE5, a plasmid native to *Staphylococcus aureus*." In: *Journal of Bacteriology* 169.11, pp. 5131–5139. DOI: 10.1128/jb.169.11.5131-5139.1987.
- Purseley, Elizabeth, Tatiana Dimitriu, Fernanda L. Paganelli, Edze R. Westra, and Stineke van Houte (2021). "CRISPR-Cas is associated with fewer antibiotic resistance genes in bacterial pathogens." In: *Philosophical Transactions of the Royal Society B* 377.1842, DOI: 10.1098/rstb.2020.0464.
- Qin, Haiying, Dennis L. Welker, and Nabil N. Youssef (1993). "Isolation and Characterization of a Linear Plasmid from the Entomopathogenic Fungus *Ascospaera apis*." In: *Plasmid* 29.1, pp. 19–30. DOI: 10.1006/plas.1993.1003.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. URL: <https://www.R-project.org/>.
- Rajer, Fredrika and Linus Sandegren (2022). "The Role of Antibiotic Resistance Genes in the Fitness Cost of Multiresistance Plasmids." In: *mBio* 13.1, DOI: 10.1128/mbio.03552-21.
- Ramirez, Maria S., Andrés Iriarte, Rodrigo Reyes-Lamothe, David J. Sherratt, and Marcelo E. Tolmasky (2019). "Small *Klebsiella pneumoniae* Plasmids: Neglected Contributors to Antibiotic Resistance." In: *Frontiers in Microbiology* 10, DOI: 10.3389/fmicb.2019.02182.
- Ramiro-Martínez, Paula, Ignacio de Quinto, João Alves Gama, and Jerónimo Rodríguez-Beltrán (2024). *Universal rules govern plasmid copy number*. BioRxiv preprint. DOI: 10.1101/2024.10.04.616648.
- Ramsay, Joshua P., Stephen M. Kwong, Riley J. T. Murphy, Karina Yui Eto, Karina J. Price, Quang T. Nguyen, Frances G. O'Brien, Warren B. Grubb, Geoffrey W. Coombs, and Neville Firth (2016). "An updated view of plasmid conjugation and mobilization in *Staphylococcus*." In: *Mobile Genetic Elements* 6.4, DOI: 10.1080/2159256x.2016.1208317.
- Rankin, Daniel J., E.P.C. Rocha, and S.P. Brown (2011). "What traits are carried on mobile genetic elements, and why?" In: *Heredity* 106, pp. 1–10. DOI: 10.1038/hdy.2010.24.

- Rankin, Daniel J., Leighton A. Turner, Jack A. Heinemann, and Sam P. Brown (2012). "The coevolution of toxin and antitoxin genes drives the dynamics of bacterial addiction complexes and intragenomic conflict." In: *Proceedings of the Royal Society B* 279.1743, pp. 3706–3715. DOI: 10.1098/rspb.2012.0942.
- Ravin, Nikolai V. (2011). "N15: The linear phage-plasmid." In: *Plasmid* 65.2, pp. 102–109. DOI: 10.1016/j.plasmid.2010.12.004.
- Redondo-Salvo, Santiago, Roger Bartomeus-Peñalver, Luis Vielva, Kaitlin A. Tagg, Hattie E. Webb, Raúl Fernández-López, and Fernando de la Cruz (2021). "COPLA, a taxonomic classifier of plasmids." In: *BMC Bioinformatics* 22.1. DOI: 10.1186/s12859-021-04299-x.
- Redondo-Salvo, Santiago, Raúl Fernández-López, Raúl Ruiz, Luis Vielva, María de Toro, Eduardo P. C. Rocha, M. Pilar Garcillán-Barcia, and Fernando de la Cruz (2020). "Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids." In: *Nature Communications* 11.1. DOI: 10.1038/s41467-020-17278-2.
- Reyes-Lamothe, Rodrigo, Tung Tran, Diane Meas, Laura Lee, Alice M. Li, David J. Sherratt, and Marcelo E. Tolmasky (2013). "High-copy bacterial plasmids diffuse in the nucleoid-free space, replicate stochastically and are randomly partitioned at cell division." In: *Nucleic Acids Research* 42.2, pp. 1042–1051. DOI: 10.1093/nar/gkt918.
- Ricci, Juan C. Diaz and Murría Eugenia Hernández (2000). "Plasmid Effects on *Escherichia coli* Metabolism." In: *Critical Reviews in Biotechnology* 20.2, pp. 79–108. DOI: 10.1080/07388550008984167.
- Riley, Margaret A. and John E. Wertz (2002). "Bacteriocins: Evolution, Ecology, and Application." In: *Annual Review of Microbiology* 56.1, pp. 117–137. DOI: 10.1146/annurev.micro.56.012302.161024.
- Rivera-Urbalejo, América, Ángeles Pérez-Oseguera, Ofelia E. Carreón-Rodríguez, and Miguel A. Cevallos (2015). "Mutations in an antisense RNA, involved in the replication control of a repABC plasmid, that disrupt plasmid incompatibility and mediate plasmid speciation." In: *Plasmid* 78, pp. 48–58. DOI: 10.1016/j.plasmid.2015.01.004.
- Robertson, James and John H. E. Nash (2018). "MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies." In: *Microbial Genomics* 4.8. DOI: 10.1099/mgen.0.000206.
- Rodríguez-Beltrán, Jerónimo, J. Carlos R. Hernández-Beltrán, Javier DelaFuente, Jose A. Escudero, Ayari Fuentes-Hernandez, R. Craig MacLean, Rafael Peña-Miller, and Alvaro San Millán (2018). "Multicopy plasmids allow bacteria to escape from fitness trade-offs during evolutionary innovation." In: *Nature Ecology & Evolution* 2.5, pp. 873–881. DOI: 10.1038/s41559-018-0529-z.
- Rodríguez-Beltrán, Jerónimo, Javier DelaFuente, Ricardo León-Sampedro, R. Craig MacLean, and Álvaro San Millán (2021a). "Beyond horizontal gene transfer: the role of plasmids in bacterial evolution." In: *Nature*

- Reviews Microbiology* 19.6, pp. 347–359. DOI: 10.1038/s41579-020-00497-1.
- Rodríguez-Beltrán, Jerónimo, Ricardo León-Sampedro, Paula Ramiro-Martínez, Carmen de la Vega, Fernando Baquero, Bruce R. Levin, and Álvaro San Millán (2021b). “Translational demand is not a major source of plasmid-associated fitness costs.” In: *Philosophical Transactions of the Royal Society B* 377.1842, DOI: 10.1098/rstb.2020.0463.
- Rodríguez-Beltrán, Jerónimo, Vidar Sørum, Macarena Toll-Riera, Carmen de la Vega, Rafael Peña-Miller, and Álvaro San Millán (2020). “Genetic dominance governs the evolution and spread of mobile genetic elements in bacteria.” In: *Proceedings of the National Academy of Sciences* 117.27, pp. 15755–15762. DOI: 10.1073/pnas.2001240117.
- Rush, Mark G. and Ravi Misra (1985). “Extrachromosomal DNA in eucaryotes.” In: *Plasmid* 14.3, pp. 177–191. DOI: 10.1016/0147-619x(85)90001-0.
- Samuel, Bruria and David Burstein (Feb. 2023). *A diverse repertoire of anti-defense systems is encoded in the leading region of plasmids*. BioRxiv preprint. DOI: 10.1101/2023.02.15.528439.
- San Millan, A., R. Peña-Miller, M. Toll-Riera, Z. V. Halbert, A. R. McLean, B. S. Cooper, and R. C. MacLean (2014). “Positive selection and compensatory adaptation interact to stabilize non-transmissible plasmids.” In: *Nature Communications* 5.1. DOI: 10.1038/ncomms6208.
- San Millan, Alvaro, Jose Antonio Escudero, Danna R. Gifford, Didier Mazel, and R. Craig MacLean (2016). “Multicopy plasmids potentiate the evolution of antibiotic resistance in bacteria.” In: *Nature Ecology & Evolution* 1.1. DOI: 10.1038/s41559-016-0010.
- San Millan, Alvaro, Jose Antonio Escudero, Belen Gutierrez, Laura Hidalgo, Nerea Garcia, Montserrat Llagostera, Lucas Dominguez, and Bruno Gonzalez-Zorn (2009). “Multiresistance in *Pasteurella multocida* Is Mediated by Coexistence of Small Plasmids.” In: *Antimicrobial Agents and Chemotherapy* 53.8, pp. 3399–3404. DOI: 10.1128/aac.01522-08.
- San Millan, Alvaro, Karl Heilbron, and R. Craig MacLean (2013). “Positive epistasis between co-infecting plasmids promotes plasmid survival in bacterial populations.” In: *The ISME Journal* 8.3, pp. 601–612. DOI: 10.1038/ismej.2013.182.
- San Millan, Alvaro and R. Craig MacLean (2019). “Fitness Costs of Plasmids: A Limit to Plasmid Transmission.” In: *Microbial Transmission*. Ed. by Fernando Baquero, Emilio Bouza, J.A. Gutiérrez-Fuentes, and Teresa M. Coque. Washington, DC: ASM Press. Chap. 4, pp. 65–79. DOI: 10.1128/9781555819743.ch4.
- San Millan, Alvaro, Alfonso Santos-Lopez, Rafael Ortega-Huedo, Cristina Bernabe-Balas, Sean P. Kennedy, and Bruno Gonzalez-Zorn (2015). “Small-Plasmid-Mediated Antibiotic Resistance Is Enhanced by Increases in

- Plasmid Copy Number and Bacterial Fitness.” In: *Antimicrobial Agents and Chemotherapy* 59.6, pp. 3335–3341. DOI: 10.1128/aac.00235-15.
- Sandegren, Linus and Dan I. Andersson (2009). “Bacterial gene amplification: implications for the evolution of antibiotic resistance.” In: *Nature Reviews Microbiology* 7.8, pp. 578–588. DOI: 10.1038/nrmicro2174.
- Santer, Mario, Anne Kupczok, Tal Dagan, and Hildegard Uecker (2022). “Fixation dynamics of beneficial alleles in prokaryotic polyploid chromosomes and plasmids.” In: *Genetics* 222.2. DOI: 10.1093/genetics/iyac121.
- Santer, Mario and Hildegard Uecker (2020). “Evolutionary Rescue and Drug Resistance on Multicopy Plasmids.” In: *Genetics* 215.3, pp. 847–868. DOI: 10.1534/genetics.119.303012.
- Santos-Lopez, Alfonso, Cristina Bernabe-Balas, Manuel Ares-Arroyo, Rafael Ortega-Huedo, Andreas Hoefler, Alvaro San Millan, and Bruno Gonzalez-Zorn (2016). “A Naturally Occurring Single Nucleotide Polymorphism in a Multicopy Plasmid Produces a Reversible Increase in Antibiotic Resistance.” In: *Antimicrobial Agents and Chemotherapy* 61.2. DOI: 10.1128/aac.01735-16.
- Santos-Lopez, Alfonso, Cristina Bernabe-Balas, Alvaro San Millan, Rafael Ortega-Huedo, Andreas Hoefler, Manuel Ares-Arroyo, and Bruno Gonzalez-Zorn (2017). *Compensatory evolution facilitates the acquisition of multiple plasmids in bacteria*. BioRxiv preprint. DOI: 10.1101/187070.
- Schensted, Irene V. (1958). “Model of Subnuclear Segregation in the Macronucleus of Ciliates.” In: *The American Naturalist* 92.864, pp. 161–170. DOI: 10.1086/282023.
- Schmartz, Georges P., Anna Hartung, Pascal Hirsch, Fabian Kern, Tobias Fehlmann, Rolf Müller, and Andreas Keller (2021). “PLSDB: advancing a comprehensive database of bacterial plasmids.” In: *Nucleic Acids Research* 50, pp. D273–D278. DOI: 10.1093/nar/gkab1111.
- Schwalbe, R. S., C. W. Hoge, J. G. Morris, P. N. O’Hanlon, R. A. Crawford, and P. H. Gilligan (1990). “In vivo selection for transmissible drug resistance in *Salmonella typhi* during antimicrobial therapy.” In: *Antimicrobial Agents and Chemotherapy* 34.1, pp. 161–163. DOI: 10.1128/aac.34.1.161.
- Seneta, Eugene and Simon Tavaré (1983). “Some stochastic models for plasmid copy number.” In: *Theoretical Population Biology* 23.2, pp. 241–256. DOI: 10.1016/0040-5809(83)90016-3.
- Sewastjanow, B. A. (1975). *Verzweigungsprozesse*. Trans. from the Russian by Uwe Prehn. Munich: R. Oldenbourg Verlag. ISBN: 3-486-39861-X.
- Shaw, Liam P., Eduardo P. C. Rocha, and R. Craig MacLean (2023). “Restriction-modification systems have shaped the evolution and distribution of plasmids across bacteria.” In: *Nucleic Acids Research* 51.13, pp. 6806–6818. DOI: 10.1093/nar/gkad452.

- Sheppard, Richard J., Timothy G. Barraclough, and Vincent A. A. Jansen (2021). "The Evolution of Plasmid Transfer Rate in Bacteria and Its Effect on Plasmid Persistence." In: *The American Naturalist* 198.4, pp. 473–488. DOI: 10.1086/716063.
- Sheppard, Richard J., Alice E. Beddis, and Timothy G. Barraclough (2020). "The role of hosts, plasmids and environment in determining plasmid transfer rates: A meta-analysis." In: *Plasmid* 108, DOI: 10.1016/j.plasmid.2020.102489.
- Shintani, Masaki, Zoe K. Sanchez, and Kazuhide Kimbara (2015). "Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy." In: *Frontiers in Microbiology* 6. DOI: 10.3389/fmicb.2015.00242.
- Shoura, Massa J., Idan Gabdank, Loren Hansen, Jason Merker, Jason Gotlib, Stephen D. Levene, and Andrew Z. Fire (2017). "Intricate and Cell Type-Specific Populations of Endogenous Circular DNA (eccDNA) in *Caenorhabditis elegans* and *Homo sapiens*." In: *G3 Genes|Genomes|Genetics* 7.10, pp. 3295–3303. DOI: 10.1534/g3.117.300141.
- Silliker, Margaret E. and Donald J. Cummings (1990). "A mitochondrial DNA rearrangement and three new mitochondrial plasmids from long-lived strains of *Podospora anserina*." In: *Plasmid* 24.1, pp. 37–44. DOI: 10.1016/0147-619x(90)90023-6.
- Silver, Simon (1992). "Plasmid-determined metal resistance mechanisms: Range and overview." In: *Plasmid* 27.1, pp. 1–3. DOI: 10.1016/0147-619x(92)90001-q.
- Silverman, B. W. (1981). "Using Kernel Density Estimates to Investigate Multimodality." In: *Journal of the Royal Statistical Society: Series B* 43.1, pp. 97–99. DOI: 10.1111/j.2517-6161.1981.tb01155.x.
- Simonsen, L., D. M. Gordon, F. M. Stewart, and B. R. Levin (1990). "Estimating the rate of plasmid transfer: an end-point method." In: *Journal of General Microbiology* 136.11, pp. 2319–2325. DOI: 10.1099/00221287-136-11-2319.
- Simonsen, Lone (1991). "The existence conditions for bacterial plasmids: Theory and reality." In: *Microbial Ecology* 1, pp. 187–205. DOI: 10.1007/bf02540223.
- Sinclair, David A. and Leonard Guarente (1997). "Extrachromosomal rDNA Circles — A Cause of Aging in Yeast." In: *Cell* 91.7, pp. 1033–1042. DOI: 10.1016/s0092-8674(00)80493-6.
- Slater, Frances R., Mark J. Bailey, Adrian J. Tett, and Sarah L. Turner (2008). "Progress towards understanding the fate of plasmids in bacterial communities." In: *FEMS Microbiology Ecology* 66.1, pp. 3–13. DOI: 10.1111/j.1574-6941.2008.00505.x.
- Smillie, Chris, M. Pilar Garcillán-Barcia, M. Victoria Francia, Eduardo P. C. Rocha, and Fernando de la Cruz (2010). "Mobility of Plasmids." In: *Mi-*

- crobiology and Molecular Biology Reviews* 74.3, pp. 434–452. DOI: 10.1128/mmb.00020-10.
- Smith, Brian A., Kevin Dougherty, Meara Clark, and David A. Baltrus (2021). “Experimental evolution of the megaplasmid pMPPla107 in *Pseudomonas stutzeri* enables identification of genes contributing to sensitivity to an inhibitory agent.” In: *Philosophical Transactions of the Royal Society B* 377.1842, DOI: 10.1098/rstb.2020.0474.
- Smith, C.J., G.D. Tribble, and D.P. Bayley (1998). “Genetic Elements of *Bacteroides* Species: A Moving Story.” In: *Plasmid* 40.1, pp. 12–29. DOI: 10.1006/plas.1998.1347.
- Solar, Gloria del and Manuel Espinosa (2002). “Plasmid copy number control: an ever-growing story.” In: *Molecular Microbiology* 37.3, pp. 492–500. DOI: 10.1046/j.1365-2958.2000.02005.x.
- Solar, Gloria del, Rafael Giraldo, Mariá Jesús Ruiz-Echevarría, Manuel Espinosa, and Ramón Díaz-Orejás (1998). “Replication and Control of Circular Bacterial Plasmids.” In: *Microbiology and Molecular Biology Reviews* 62.2, pp. 434–464. DOI: 10.1128/mmb.62.2.434-464.1998.
- Song, Sooyeon and Thomas K. Wood (2018). “Post-segregational Killing and Phage Inhibition Are Not Mediated by Cell Death Through Toxin/Antitoxin Systems.” In: *Frontiers in Microbiology* 9. DOI: 10.3389/fmicb.2018.00814.
- Sota, Masahiro, Hirokazu Yano, Julie M. Hughes, Gary W. Daughdrill, Zaid Abdo, Larry J. Forney, and Eva M. Top (2010). “Shifts in the host range of a promiscuous plasmid through parallel evolution of its replication initiation protein.” In: *The ISME Journal* 4.12, pp. 1568–1580. DOI: 10.1038/ismej.2010.72.
- Spaková, Tímea, Livia Kolesar Fecskeová, Peter Javorský, and Peter Pristas (2013). “Two Rep Genes in Small Cryptic Plasmid pKST21 of *Escherichia coli*.” In: *Current Microbiology* 67.4, pp. 437–441. DOI: 10.1007/s00284-013-0386-6.
- Stalder, Thibault and Eva Top (2016). “Plasmid transfer in biofilms: a perspective on limitations and opportunities.” In: *npj Biofilms and Microbiomes* 2.1. DOI: 10.1038/npjbiofilms.2016.22.
- Stewart, Frank M. and Bruce R. Levin (1977). “The Population Biology of Bacterial Plasmids: *a priori* Conditions for the Existence of Conjugationally Transmitted Factors.” In: *Genetics* 87.2, pp. 209–228. DOI: 10.1093/genetics/87.2.209.
- Strang, Gilbert (2009). *Introduction to Linear Algebra*. 4th ed. Wellesley, Mass.: Wellesley-Cambridge Press. ISBN: 978-0-9802327-1-4.
- Summers, David K. (1996). *The Biology of Plasmids*. Oxford: Blackwell Science. ISBN: 0-632-03436-X.
- Summers, David K., Christopher W. H. Beton, and Helen L. Withers (1993). “Multicopy plasmid instability: the dimer catastrophe hypothesis.” In:

- Molecular Microbiology* 8.6, pp. 1031–1038. DOI: 10.1111/j.1365-2958.1993.tb01648.x.
- Summers, David K. and David J. Sherratt (1984). “Multimerization of high copy number plasmids causes instability: ColE1 encodes a determinant essential for plasmid monomerization and stability.” In: *Cell* 36.4, pp. 1097–1103. DOI: 10.1016/0092-8674(84)90060-6.
- Sun, Lei, Helen K. Alexander, Balazs Bogos, Daniel J. Kiviet, Martin Ackermann, and Sebastian Bonhoeffer (2018). “Effective polyploidy causes phenotypic delay and influences bacterial evolvability.” In: *PLOS Biology* 16.2, DOI: 10.1371/journal.pbio.2004644.
- Svara, Fabian and Daniel J. Rankin (2011). “The evolution of plasmid-carried antibiotic resistance.” In: *BMC Evolutionary Biology* 11.1, DOI: 10.1186/1471-2148-11-130.
- Sýkora, Peter (1992). “Macroevolution of plasmids: A model for plasmid speciation.” In: *Journal of Theoretical Biology* 159.1, pp. 53–65. DOI: 10.1016/S0022-5193(05)80767-2.
- Szpirer, Cedric, Eva Top, Martine Couturier, and Max Mergeay (1999). “Retrotransfer or gene capture: a feature of conjugative plasmids, with ecological and evolutionary significance.” In: *Microbiology* 145.12, pp. 3321–3329. DOI: 10.1099/00221287-145-12-3321.
- Takeda, Toshiharu, Choong-Soo Yun, Masaki Shintani, Hisakazu Yamane, and Hideaki Nojiri (2011). “Distribution of Genes Encoding Nucleoid-Associated Protein Homologs in Plasmids.” In: *International Journal of Evolutionary Biology* 2011, DOI: 10.4061/2011/685015.
- Taylor, Diane E., Jessie G. Levine, and David E. Bradley (1981). “In vivo formation of a plasmid cointegrate expressing two incompatibility phenotypes.” In: *Plasmid* 5.3, pp. 233–244. DOI: 10.1016/0147-619x(81)90001-9.
- Tazyman, Samuel J. and Sebastian Bonhoeffer (2013). “Fixation probability of mobile genetic elements such as plasmids.” In: *Theoretical Population Biology* 90, pp. 49–55. DOI: 10.1016/j.tpb.2013.09.012.
- (2014). “Plasmids and Evolutionary Rescue by Drug Resistance.” In: *Evolution* 68.7, pp. 2066–2078. DOI: 10.1111/evo.12423.
- (2015). “Why There Are No Essential Genes on Plasmids.” In: *Molecular Biology and Evolution* 32.12, pp. 3079–3088. DOI: 10.1093/molbev/msu293.
- Tepekule, Burcu, Pia Abel zur Wiesch, Roger D. Kouyos, and Sebastian Bonhoeffer (2019). “Quantifying the impact of treatment history on plasmid-mediated resistance evolution in human gut microbiota.” In: *Proceedings of the National Academy of Sciences* 116.46, pp. 23106–23116. DOI: 10.1073/pnas.1912188116.
- Thoma, Lina and Günther Muth (2012). “Conjugative DNA transfer in *Streptomyces* by TraB: is one protein enough?” In: *FEMS Microbiology Letters* 337.2, pp. 81–88. DOI: 10.1111/1574-6968.12031.

- Tsilibaris, Virginie, Geneviève Maenhaut-Michel, Natacha Mine, and Laurence Van Melderen (2007). "What Is the Benefit to *Escherichia coli* of Having Multiple Toxin-Antitoxin Systems in Its Genome?" In: *Journal of Bacteriology* 189.17, pp. 6101–6108. DOI: 10.1128/jb.00527-07.
- Tun-Garrido, Cristina, Patricia Bustos, Víctor González, and Susana Brom (2003). "Conjugative Transfer of p42a from *Rhizobium etli* CFN42, Which Is Required for Mobilization of the Symbiotic Plasmid, Is Regulated by Quorum Sensing." In: *Journal of Bacteriology* 185.5, pp. 1681–1692. DOI: 10.1128/jb.185.5.1681-1692.2003.
- Turner, Kristen M. et al. (2017). "Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity." In: *Nature* 543, pp. 122–125. DOI: 10.1038/nature21356.
- Ubeda, Jean-Michel et al. (2014). "Genome-Wide Stochastic Adaptive DNA Amplification at Direct and Inverted DNA Repeats in the Parasite *Leishmania*." In: *PLoS Biology* 12.5, DOI: 10.1371/journal.pbio.1001868.
- Uecker, Hildegard and Joachim Hermisson (2011). "On the Fixation Process of a Beneficial Mutation in a Variable Environment." In: *Genetics* 188.4, pp. 915–930. DOI: 10.1534/genetics.110.124297.
- Uhlin, B. E. and K. Nordström (1975). "Plasmid incompatibility and control of replication: copy mutants of the R-factor R1 in *Escherichia coli* K-12." In: *Journal of Bacteriology* 124.2, pp. 641–649. DOI: 10.1128/jb.124.2.641-649.1975.
- Valle, Aida Alonso-del, Ricardo León-Sampedro, Jerónimo Rodríguez-Beltrán, Javier DelaFuente, Marta Hernández-García, Patricia Ruiz-Garbajosa, Rafael Cantón, Rafael Peña-Miller, and Alvaro San Millán (2021). "Variability of plasmid fitness effects contributes to plasmid persistence in bacterial communities." In: *Nature Communications* 12.1, DOI: 10.1038/s41467-021-22849-y.
- Ventura, Marco, Carlos Canchaya, Andreas Tauch, Govind Chandra, Gerald F. Fitzgerald, Keith F. Chater, and Douwe van Sinderen (2007). "Genomics of *Actinobacteria*: Tracing the Evolutionary History of an Ancient Phylum." In: *Microbiology and Molecular Biology Reviews* 71.3, pp. 495–548. DOI: 10.1128/mmb.00005-07.
- Verweij, Wilco and Cortland K. Griswold (2023). "Spatial structure and benefits to hosts allow plasmids with and without post-segregational killing systems to coexist." In: *Biology Letters* 19.3. DOI: 10.1098/rsbl.2022.0376.
- Vial, Ludovic and Florence Hommais (2019). "Plasmid-chromosome cross-talks." In: *Environmental Microbiology* 22.2, pp. 540–556. DOI: 10.1111/1462-2920.14880.
- Volkova, Victoriya V., Cristina Lanzas, Zhao Lu, and Yrjö Tapio Gröhn (2012). "Mathematical Model of Plasmid-Mediated Resistance to Ceftiofur in Commensal Enteric *Escherichia coli* of Cattle." In: *PLoS ONE* 7.5, DOI: 10.1371/journal.pone.0036738.

- Volkova, Victoriya V., Zhao Lu, Cristina Lanzas, H. Morgan Scott, and Yrjö T. Gröhn (2013). “Modelling dynamics of plasmid-gene mediated antimicrobial resistance in enteric bacteria using stochastic differential equations.” In: *Scientific Reports* 3.1. DOI: 10.1038/srep02463.
- Wang, Yiqing, Aditi Batra, Hinrich Schulenburg, and Tal Dagan (2021). “Gene sharing among plasmids and chromosomes reveals barriers for antibiotic resistance gene transfer.” In: *Philosophical Transactions of the Royal Society B* 377.1842, DOI: 10.1098/rstb.2020.0467.
- Wang, Yiqing and Tal Dagan (2024). “The evolution of antibiotic resistance islands occurs within the framework of plasmid lineages.” In: *Nature Communications* 15.1. DOI: 10.1038/s41467-024-48352-8.
- Wardell, Grace E., Michael F. Hynes, Peter J. Young, and Ellie Harrison (2021). “Why are rhizobial symbiosis genes mobile?” In: *Philosophical Transactions of the Royal Society B* 377.1842, DOI: 10.1098/rstb.2020.0471.
- Watve, Mukta M., Neelesh Dahanukar, and Milind G. Watve (2010). “Socio-biological Control of Plasmid Copy Number in Bacteria.” In: *PLoS ONE* 5.2, DOI: 10.1371/journal.pone.0009328.
- Webb, Glenn F., Erika M. C. D’Agata, Pierre Magal, and Shigui Ruan (2005). “A model of antibiotic-resistant bacterial epidemics in hospitals.” In: *Proceedings of the National Academy of Sciences* 102.37, pp. 13343–13348. DOI: 10.1073/pnas.0504053102.
- Wein, Tanita, Yiqing Wang, Myriam Barz, Fenna T. Stücker, Katrin Hammerschmidt, and Tal Dagan (2021). “Essential gene acquisition destabilizes plasmid inheritance.” In: *PLOS Genetics* 17.7, DOI: 10.1371/journal.pgen.1009656.
- Wein, Tanita, Yiqing Wang, Nils F. Hülter, Katrin Hammerschmidt, and Tal Dagan (2020). “Antibiotics Interfere with the Evolution of Plasmid Stability.” In: *Current Biology* 30.19, 3841–3847.e4. DOI: 10.1016/j.cub.2020.07.019.
- Werisch, Martin, Uta Berger, and Thomas U. Berendonk (2017). “Conjugative plasmids enable the maintenance of low cost non-transmissible plasmids.” In: *Plasmid* 91, pp. 96–104. DOI: 10.1016/j.plasmid.2017.04.004.
- Wickham, Hadley (2016). *ggplot2. Elegant Graphics for Data Analysis*. 2nd ed. Use R! Springer. ISBN: 978-3-319-24275-0.
- Willms, Allan R., Paul D. Roughan, and Jack A. Heinemann (2006). “Static recipient cells as reservoirs of antibiotic resistance during antibiotic therapy.” In: *Theoretical Population Biology* 70.4, pp. 436–451. DOI: 10.1016/j.tpb.2006.04.001.
- Windle, B., B. W. Draper, Y. X. Yin, S. O’Gorman, and G. M. Wahl (1991). “A central role for chromosome breakage in gene amplification, deletion formation, and amplicon integration.” In: *Genes & Development* 5.2, pp. 160–174. DOI: 10.1101/gad.5.2.160.

- Wu, Sihan et al. (2019). "Circular ecDNA promotes accessible chromatin and high oncogene expression." In: *Nature* 575.7784, pp. 699–703. DOI: 10.1038/s41586-019-1763-5.
- Xiao, J., R. E. Melton, and T. Kieser (1994). "High-frequency homologous plasmid-plasmid recombination coupled with conjugation of plasmid SCP2\* in *Streptomyces*." In: *Molecular Microbiology* 14.3, pp. 547–555. DOI: 10.1111/j.1365-2958.1994.tb02188.x.
- Xu, Ke et al. (2018). "Structure and evolution of double minutes in diagnosis and relapse brain tumors." In: *Acta Neuropathologica* 137.1, pp. 123–137. DOI: 10.1007/s00401-018-1912-1.
- Yano, Hirokazu, Linda M. Rogers, Molly G. Knox, Holger Heuer, Kornelia Smalla, Celeste J. Brown, and Eva M. Top (2013). "Host range diversification within the IncP-1 plasmid group." In: *Microbiology* 159, pp. 2303–2315. DOI: 10.1099/mic.0.068387-0.
- Yano, Hirokazu, Katarzyna Wegrzyn, Wesley Loftie-Eaton, Jenny Johnson, Gail E. Deckert, Linda M. Rogers, Igor Konieczny, and Eva M. Top (June 2016). "Evolved plasmid-host interactions reduce plasmid interference cost." In: *Molecular Microbiology* 101.5, pp. 743–756. DOI: 10.1111/mmi.13407.
- Yao, Shiyin, Donald R. Helinski, and Aresa Toukdarian (2007). "Localization of the Naturally Occurring Plasmid ColE1 at the Cell Pole." In: *Journal of Bacteriology* 189.5, pp. 1946–1953. DOI: 10.1128/jb.01451-06.
- Yu, Weizhu, Kevin Gillies, Jeffery K. Kondo, Jeffery R. Broadbent, and Larry L. McKay (1996). "Loss of Plasmid-Mediated Oligopeptide Transport System in Lactococci: Another Reason for Slow Milk Coagulation." In: *Plasmid* 35.3, pp. 145–155. DOI: 10.1006/plas.1996.0017.
- Zaleski, Piotr, Renata Wolinowska, Krystyna Strzezek, Agata Lakomy, and Andrzej Plucienniczak (2006). "The complete sequence and segregational stability analysis of a new cryptic plasmid pIGWZ12 from a clinical strain of *Escherichia coli*." In: *Plasmid* 56.3, pp. 228–232. DOI: 10.1016/j.plasmid.2006.05.004.
- Zhang, Jin-Fei et al. (2022a). "A Trade-Off for Maintenance of Multidrug-Resistant IncHI2 Plasmids in *Salmonella enterica* Serovar Typhimurium through Adaptive Evolution." In: *mSystems* 7.5. DOI: 10.1128/msystems.00248-22.
- Zhang, Linyue et al. (2022b). "Co-evolutionary adaptations of *Acinetobacter baumannii* and a clinical carbapenemase-encoding plasmid during carbapenem exposure." In: *Evolutionary Applications* 15.7, pp. 1045–1061. DOI: 10.1111/eva.13441.
- Zhu, Yanfen et al. (2021). "Oncogenic extrachromosomal DNA functions as mobile enhancers to globally amplify chromosomal transcription." In: *Cancer Cell* 39.5, 694–707.e7. DOI: 10.1016/j.ccell.2021.03.006.
- Zielenkiewicz, Urszula and Piotr Ceglowski (2001). "Mechanisms of plasmid stable maintenance with special focus on plasmid addiction sys-

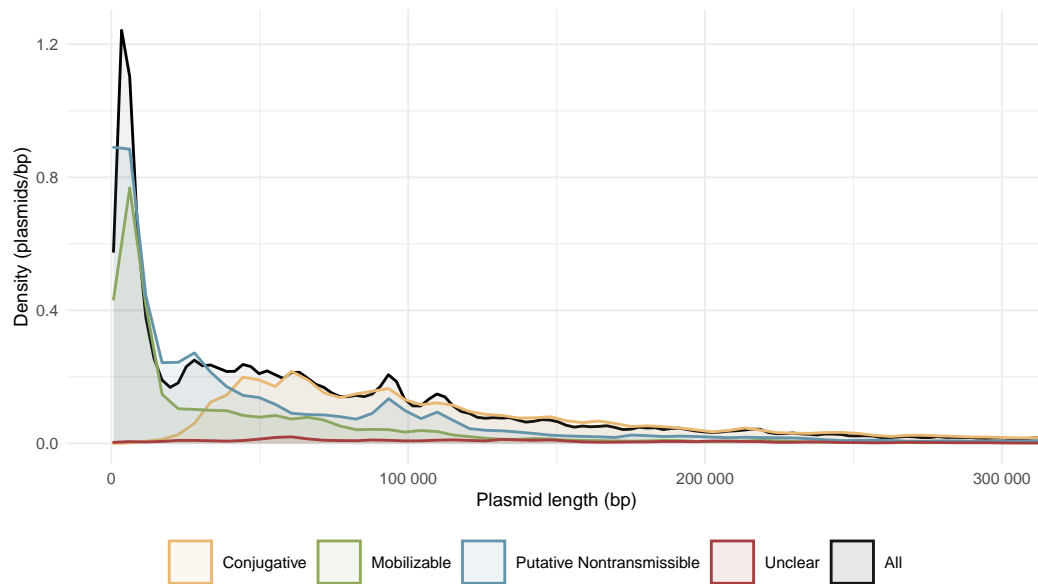
- tems." In: *Acta Biochimica Polonica* 48.4, pp. 1003–1023. DOI: 10.18388/abp.2001\_3863.
- Zwanzig, Martin (2021). "The ecology of plasmid-coded antibiotic resistance: a basic framework for experimental research and modeling." In: *Computational and Structural Biotechnology Journal* 19, pp. 586–599. DOI: 10.1016/j.csbj.2020.12.027.
- Zwanzig, Martin, Ellie Harrison, Michael A. Brockhurst, James P. J. Hall, Thomas U. Berendonk, and Uta Berger (2019). "Mobile Compensatory Mutations Promote Plasmid Survival." In: *mSystems* 4.1. DOI: 10.1128/mSystems.00186-18.



APPENDIX A

---

SUPPLEMENTARY FIGURES TO CHAPTER 3



*Figure A.1:* The distribution of plasmid lengths for plasmids shorter than 300 000 bp: in black, all plasmids; in colour, plasmids separated by mobility. The data sources and figure preparation are described in the caption to Figure 3.1. The kernel density estimates are made with a bandwidth of 1000 bp, considerably smaller than the value of 6822.532 bp used for Figure 3.1A, to exclude the possibility of oversmoothing (the value used in Figure 3.1A is the median of the values chosen for each of the conjugative, mobilizable, and nontransmissible classes by R's default density estimation: this was done to ensure that all the densities in the figure were estimated with a common bandwidth). This estimate is, as expected, notably less smooth than the estimate in Chapter 3, but it retains the same general features; in particular, it resembles the log-length distribution even less if anything, and it does not show particular signs of being bimodal.

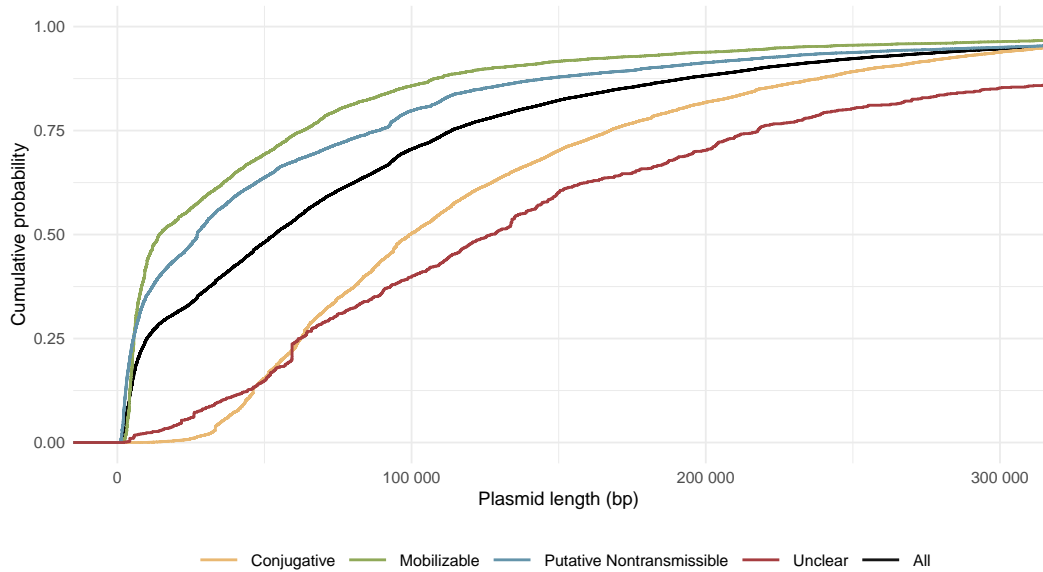
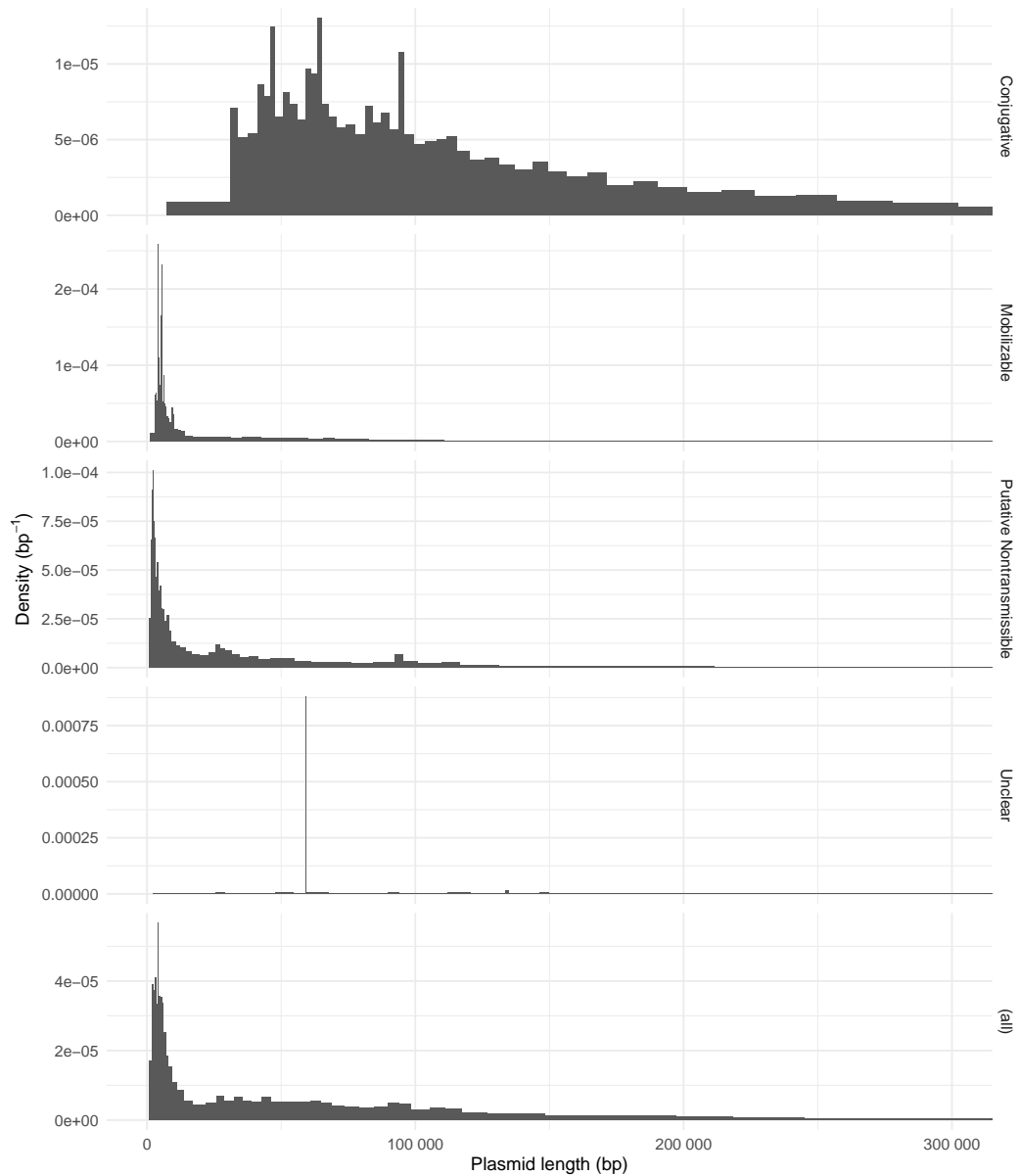


Figure A.2: Empirical cumulative distribution function of the distribution of plasmid lengths: in black, all plasmids; in colour, plasmids separated by mobility. The data sources are described in the caption to Figure 3.1. The cumulative distribution function has the advantage that there is no parameter choice involved in its estimation (unlike the need to choose a bandwidth for kernel density estimation or a bin width for a histogram), but is more difficult to interpret. A bimodal distribution should show two regions of steep slope separated by a region of shallow slope: there is no notably bimodal signs in the overall distribution. Note that unlike the kernel density estimates, each cumulative distribution function is normalized independently to a maximum of one.



*Figure A.3:* Equal-area histograms (percentograms) of the lengths of plasmids shorter than 300 000 bp. The data sources and figure preparation are described in the caption to Figure 3.1. Each bin of the histogram holds 2% of the total number of plasmids (i.e. the borders between bins are at the second, fourth, sixth, ... percentiles). Yet again, the distribution looks notably different from the log-length distribution and the overall distribution is not particularly bimodal. Note that unlike the kernel density estimates, each percentogram is individually normalized to a total area of one.

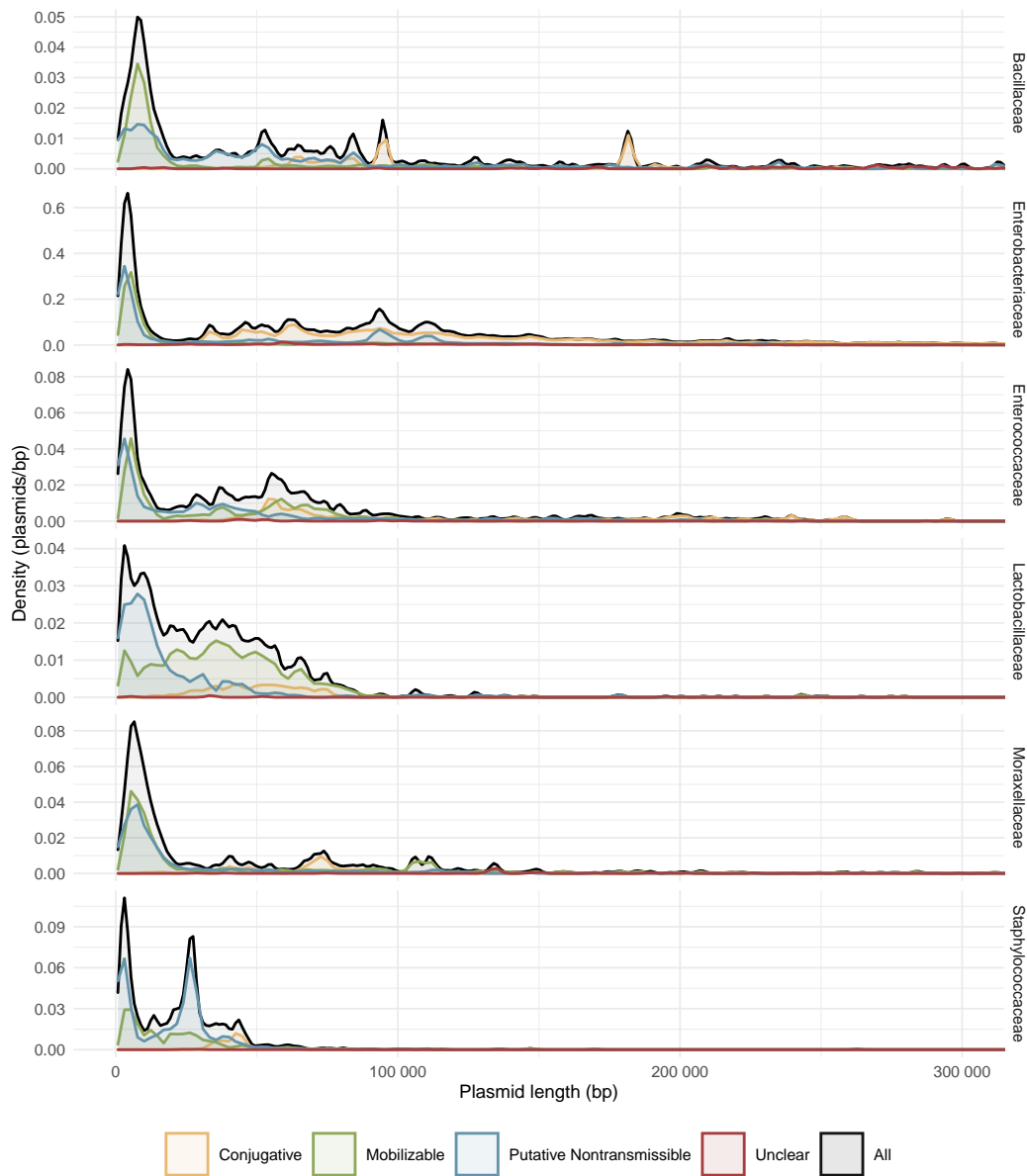


Figure A.4: The distributions of plasmid lengths for plasmids shorter than 300000 bp from the six bacterial families with at least 1000 plasmid sequences in the database: in black, all plasmids; in colour, plasmids separated by mobility. The data sources and figure preparation are described in the caption to Figure 3.1. The kernel density estimates are made with a bandwidth of 1000 bp, considerably smaller than the values used in Figure 3.3 (which range from 4018.741 bp to 13284.687 bp), to exclude the possibility of oversmoothing (the value used in each panel of Figure 3.3 is the median of the values chosen for each of the conjugative, mobilizable, and nontransmissible classes for that panel by R's default density estimation: this was done to ensure that all the densities in each panel were estimated with a common bandwidth).

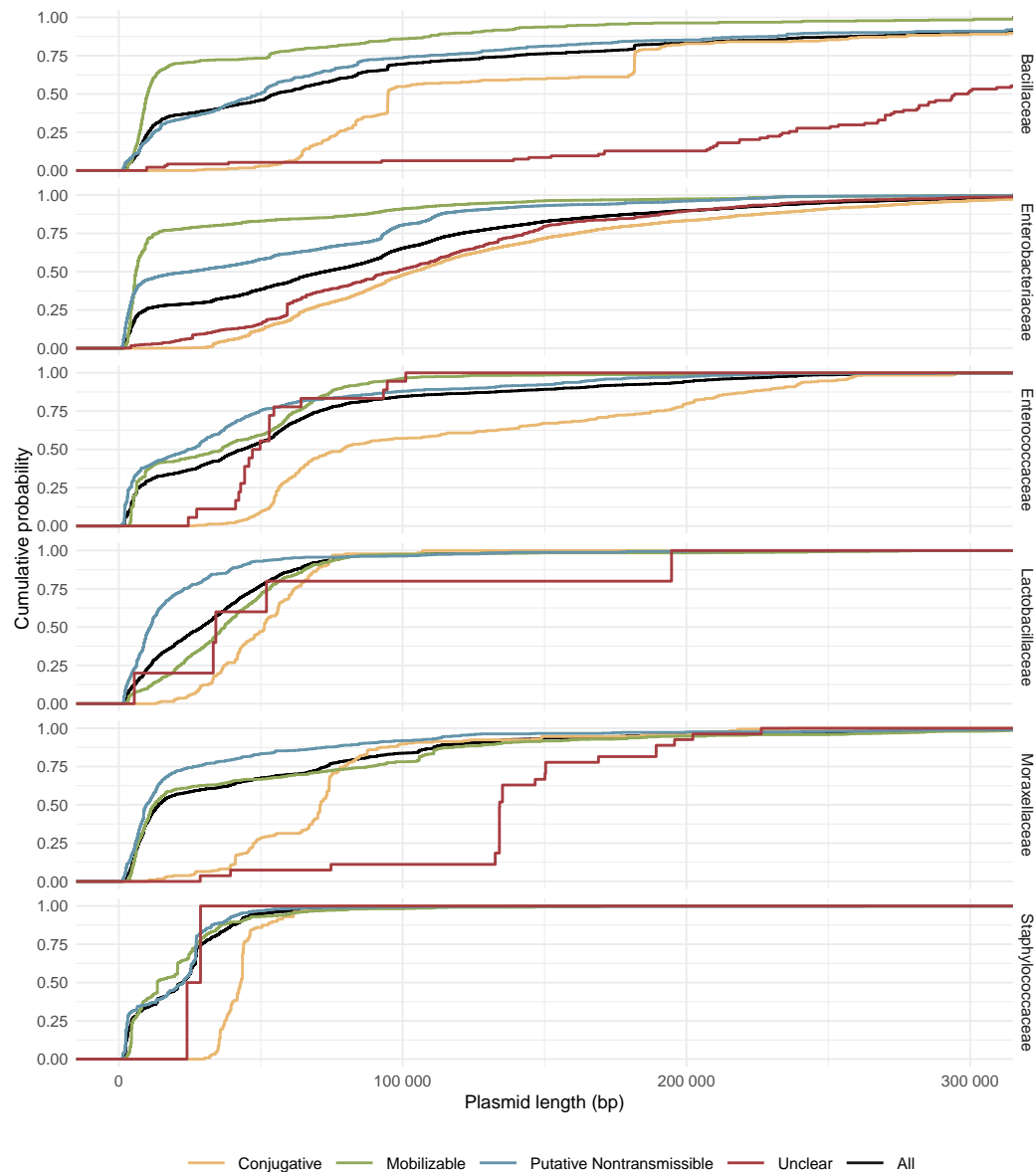


Figure A.5: Empirical cumulative distribution functions of the distributions of plasmid lengths for plasmids from the six bacterial families with at least 1000 plasmid sequences in the database: in black, all plasmids; in colour, plasmids separated by mobility. The data sources are described in the caption to Figure 3.1. A bimodal distribution should show two regions of steep slope separated by a region of shallow slope: many but not all families show a clear bimodal distribution. Note that unlike the kernel density estimates, each cumulative distribution function is normalized individually to a maximum of one.

---

 MATHEMATICAL NOTES

In this appendix, we derive several combinatorial results needed in the main text of the paper.

First, for completeness, we derive the distribution of outcomes of a Pólya urn process (Eggenberger and Pólya 1923). Suppose we start with an urn of  $n$  balls,  $i$  blue and  $n - i$  white. We pick a ball out of the urn, then replace it together with another ball of the same colour. After repeating this process  $m - n$  times, there are  $m$  balls in the urn: what is the probability that  $k$  of them are blue?

**Proposition 1.** *Denote the probability described above by  $P\left(\begin{smallmatrix} n \rightarrow m \\ i \rightarrow k \end{smallmatrix}\right)$ . Then for  $m \geq n > 0$  and  $k \geq i \geq 0$ ,*

$$P\left(\begin{smallmatrix} n \rightarrow m \\ i \rightarrow k \end{smallmatrix}\right) = \frac{\binom{m-k-1}{(m-k)-(n-i)} \binom{k-1}{k-i}}{\binom{m-1}{m-n}}.$$

*Proof.* First we prove the special case

$$P\left(\begin{smallmatrix} n \rightarrow n \\ i \rightarrow k \end{smallmatrix}\right) = \frac{\binom{n-k-1}{i-k} \binom{k-1}{k-i}}{\binom{n-1}{0}};$$

since no balls are drawn in this case, the left-hand side is 1 if  $i = k$  and 0 otherwise. The lower indices of the binomial coefficients in the numerator are negatives of each other, so the coefficients can only both be nonzero if  $k - i = 0$ , and in that case all three coefficients are equal to 1, satisfying the equality.

Next we prove another special case

$$P\left(\begin{smallmatrix} n \rightarrow m \\ i \rightarrow i \end{smallmatrix}\right) = \frac{\binom{m-i-1}{m-n} \binom{i-1}{0}}{\binom{m-1}{m-n}}.$$

Suppose that this equality is established for  $P\left(\begin{smallmatrix} n \rightarrow m-1 \\ i \rightarrow i \end{smallmatrix}\right)$ . In order to reach  $m$  balls in the urn without increasing the number of blue balls, we must first reach  $m - 1$  balls, then draw a white ball on the final draw. Therefore

$$\begin{aligned} P\left(\begin{smallmatrix} n \rightarrow m \\ i \rightarrow i \end{smallmatrix}\right) &= P\left(\begin{smallmatrix} n \rightarrow m-1 \\ i \rightarrow i \end{smallmatrix}\right) \frac{m-i-1}{m-1} \\ &= \frac{\binom{m-i-2}{m-n-1} \binom{i-1}{0} (m-i-1)}{\binom{m-2}{m-n-1} (m-1)} \\ &= \frac{\binom{m-i-1}{m-n} \binom{i-1}{0} (m-n)}{\binom{m-1}{m-n} (m-n)}, \end{aligned}$$

and by induction on  $m$  (with our previous special case as the base case), the equality is established.

Finally, we prove the general case. Suppose we know the equality holds for  $P\left(\begin{smallmatrix} n \rightarrow m-1 \\ i \rightarrow k \end{smallmatrix}\right)$  and  $P\left(\begin{smallmatrix} n \rightarrow m-1 \\ i \rightarrow k-1 \end{smallmatrix}\right)$ . In order for there to be  $k$  blue balls in the urn when we get to the point of having  $m$  total balls, then either the last ball drawn was white, and there were  $k$  blue out of  $m-1$  total balls before, or the last ball drawn was blue, and there were  $k-1$  blue out of  $m-1$  total balls before. Thus

$$\begin{aligned} P\left(\begin{smallmatrix} n \rightarrow m \\ i \rightarrow k \end{smallmatrix}\right) &= P\left(\begin{smallmatrix} n \rightarrow m-1 \\ i \rightarrow k \end{smallmatrix}\right) \frac{m-k-1}{m-1} + P\left(\begin{smallmatrix} n \rightarrow m-1 \\ i \rightarrow k-1 \end{smallmatrix}\right) \frac{k-1}{m-1} \\ &= \frac{\binom{m-k-2}{(m-k-1)-(n-i)} \binom{k-1}{k-i} (m-k-1)}{\binom{m-2}{m-n-1} (m-1)} + \frac{\binom{m-k-1}{(m-k)-(n-i)} \binom{k-2}{k-i-1} (k-1)}{\binom{m-2}{m-n-1} (m-1)} \\ &= \frac{\binom{m-k-1}{(m-k)-(n-i)} \binom{k-1}{k-i}}{\binom{m-1}{m-n}} \left( \frac{((m-k)-(n-i)) + (k-i)}{m-n} \right), \end{aligned}$$

and by induction on  $m$  and  $k$ , the equality is established.  $\square$

Obviously if we replace blue and white balls in the urn with mutant and wild-type plasmids in a cell, this process describes the replication of plasmids under the random replication model. The probability that starting with a cell with  $i$  mutant plasmids there are  $j+k$  mutant plasmids after replication is then  $P\left(\begin{smallmatrix} n \rightarrow 2n \\ i \rightarrow j+k \end{smallmatrix}\right)$ .

When a mutation occurs during plasmid replication in the random replication model, it is possible for the mutant plasmid to be replicated again the same generation, and so the initial mutant cell can end up with multiple mutant plasmids. Suppose that a mutation occurs during each replication of a plasmid with probability  $u$ ; then the probability that in a certain generation exactly one mutation occurs, it occurs during the  $(\ell+1)$ -th replication, and the cell ends up with  $i$  mutant plasmids after all of the replication has occurred but before cell division is

$$A_{\ell i} = (1-u)^\ell u P\left(\begin{smallmatrix} n+\ell+1 \rightarrow 2n \\ 1 \rightarrow i \end{smallmatrix}\right) (1-u)^{n-i-\ell}.$$

We assume that  $u$  is small, so that  $1-u \approx 1$ ; then the probability that a

---

mutation occurs in a given generation is

$$\begin{aligned}
\sum_{i=1}^n \sum_{\ell=0}^{n-1} A_{\ell i} &\approx \sum_{i=1}^n \sum_{\ell=0}^{n-1} u P \binom{n+\ell+1}{1} \binom{2n}{i} \\
&= u \sum_{i=1}^n \sum_{\ell=0}^{n-1} \frac{\binom{2n-i-1}{n-\ell-i} \binom{i-1}{i-1}}{\binom{2n-1}{n-\ell-1}} \\
&= u \sum_{\ell=0}^{n-1} \frac{1}{\binom{2n-1}{n-\ell-1}} \sum_{i=1}^n \binom{2n-i-1}{n+\ell-1} \\
&= u \sum_{\ell=0}^{n-1} \frac{1}{\binom{2n-1}{n+\ell}} \sum_{i=0}^{2n-2} \binom{i}{n+\ell-1} \\
&= u \sum_{\ell=0}^{n-1} \frac{\binom{2n-1}{n+\ell}}{\binom{2n-1}{n+\ell}} \\
&= un.
\end{aligned}$$

The probability of ending up with  $i$  mutant plasmids conditional on a mu-

tation occurring is then

$$\begin{aligned}
 \frac{\sum_{\ell=0}^{n-1} A_{\ell i}}{\sum_{i=1}^n \sum_{\ell=0}^{n-1} A_{\ell i}} &\approx \frac{\sum_{\ell=0}^{n-1} uP\left(\begin{matrix} n+\ell+1 \\ 1 \end{matrix} \rightarrow \begin{matrix} 2n \\ i \end{matrix}\right)}{un} \\
 &= \frac{1}{n} \sum_{\ell=0}^{n-1} \frac{\binom{2n-i-1}{n-\ell-i} \binom{i-1}{i-1}}{\binom{2n-1}{n-\ell-1}} \\
 &= \frac{1}{n} \sum_{\ell=0}^{n-1} \frac{n+\ell}{n-\ell} \frac{\binom{n-\ell}{i}}{\binom{2n-1}{i}} \\
 &= \frac{1}{ni \binom{2n-1}{i}} \sum_{\ell=0}^{n-1} (n+\ell) \binom{n-\ell-1}{i-1} \\
 &= \frac{1}{ni \binom{2n-1}{i}} \left( \sum_{\ell=0}^{n-1} 2n \binom{n-\ell-1}{i-1} - \sum_{\ell=0}^{n-1} (n-\ell) \binom{n-\ell-1}{i-1} \right) \\
 &= \frac{1}{ni \binom{2n-1}{i}} \left( 2n \sum_{\ell=0}^{n-1} \binom{n-\ell-1}{i-1} - i \sum_{\ell=0}^{n-1} \binom{n-\ell}{i} \right) \\
 &= \frac{1}{ni \binom{2n-1}{i}} \left( 2n \sum_{\iota=0}^{n-1} \binom{\iota}{i-1} - i \sum_{\iota=0}^n \binom{\iota}{i} \right) \\
 &= \frac{1}{ni \binom{2n-1}{i}} \left( 2n \binom{n}{i} - i \binom{n+1}{i+1} \right) \\
 &= \frac{1}{ni \binom{2n-1}{i}} \left( 2n - \frac{i(n+1)}{i+1} \right) \binom{n}{i} \\
 &= \frac{(2n+(n-1)i)}{ni(i+1)} \frac{\binom{n}{i}}{\binom{2n-1}{i}};
 \end{aligned}$$

by substituting  $i = j + k$ , we obtain the replication part of equation (4.5).

**Lemma 1.** For all integers  $n > 0$ ,  $0 \leq \beta \leq n$ , and  $\gamma > 0$ ,

$$\sum_{\alpha=0}^n \frac{\binom{2i}{\alpha} \binom{2n-2i}{n-\alpha}}{\binom{2n}{n}} \binom{\alpha}{\beta} = \frac{\binom{n}{\beta}}{\binom{2n}{\beta}} \binom{2i}{\beta} \tag{B.1}$$

and

$$\sum_{\alpha=1}^{n-1} \sum_{\delta=\gamma}^{n+\gamma} \frac{\binom{\delta}{\gamma} \binom{2n-\delta}{n-\gamma}}{\binom{2n}{n}} \frac{\binom{2n-\delta-1}{n-\delta+\alpha} \binom{\delta-1}{\delta-\alpha}}{\binom{2n-1}{n}} = \frac{(n-1)(2n+1)}{(n+1)(2n-1)}. \tag{B.2}$$

*Proof.* For (B.1), we have that

$$\begin{aligned}
\sum_{\alpha=0}^n \frac{\binom{2i}{\alpha} \binom{2n-2i}{n-\alpha}}{\binom{2n}{n}} \binom{\alpha}{\beta} &= \frac{\binom{2i}{\beta}}{\binom{2n}{n}} \sum_{\alpha=0}^n \binom{2i-\beta}{\alpha-\beta} \binom{2n-2i}{n-\alpha} \\
&= \binom{2i}{\beta} \binom{2n-\beta}{n-\beta} \Big/ \binom{2n}{n} \\
&= \binom{2i}{\beta} \binom{2n-\beta}{n-\beta} \binom{n}{\beta} \Big/ \binom{2n}{n} \binom{n}{\beta} \\
&= \binom{2i}{\beta} \binom{2n-\beta}{n-\beta} \binom{n}{\beta} \Big/ \binom{2n}{\beta} \binom{2n-\beta}{n-\beta} \\
&= \frac{\binom{n}{\beta}}{\binom{2n}{\beta}} \binom{2i}{\beta},
\end{aligned}$$

where the first and fourth equalities use trinomial revision and the second uses Vandermonde convolution (with an implicit change of index to  $\alpha - \beta$ ). For (B.2), we have that

$$\begin{aligned}
\sum_{\alpha=1}^{n-1} \sum_{\delta=\gamma}^{n+\gamma} \frac{\binom{\delta}{\gamma} \binom{2n-\delta}{n-\gamma} \binom{2n-\delta-1}{n-\delta+\alpha} \binom{\delta-1}{\delta-\alpha}}{\binom{2n}{n} \binom{2n-1}{n}} \\
&= \frac{1}{\binom{2n}{n} \binom{2n-1}{n}} \sum_{\delta=\gamma}^{n+\gamma} \binom{\delta}{\gamma} \binom{2n-\delta}{n-\gamma} \sum_{\alpha=-\infty}^{\infty} \binom{2n-\delta-1}{n-\delta+\alpha} \binom{\delta-1}{\delta-\alpha} \\
&= \frac{\binom{2n-2}{n}}{\binom{2n}{n} \binom{2n-1}{n}} \sum_{\delta=0}^{2n} \binom{\delta}{\gamma} \binom{2n-\delta}{n-\gamma} \\
&= \frac{\binom{2n-2}{n} \binom{2n+1}{n+1}}{\binom{2n}{n} \binom{2n-1}{n}} = \frac{(n-1)(2n+1)}{(2n-1)(n+1)},
\end{aligned}$$

where the second equality uses Vandermonde convolution (with an implicit change of index to  $\delta - \alpha$ ) and the third uses equation (5.26) of Graham et al. (1994).  $\square$



## SUPPLEMENTARY FIGURES TO CHAPTER 4

Stochastic simulations were performed to verify the results of the numerical analysis. The branching process was simulated in discrete time, with the number and type of daughters of each cell in every generation being sampled from the same progeny distribution as in the analytic model, except that the assumptions on mutation were relaxed. Mutations were allowed to occur at every plasmid replication in every cell division, as opposed to the analytic model where we assumed that there is at most one mutation per host generation, and never multiple mutation events in the same lineage. The simulation programs are available as a further supplement.

Establishment probability was calculated from simulations of a population starting from a pair of cells sampled from the distribution of pairs of daughters of a homozygote wild-type cell conditional on a mutation occurring; the population was considered to have gone extinct if the total number of cells reached zero, and the mutation to have established if the number of heterozygote cells reached 100000. The use of this threshold was validated after the estimated establishment probability was calculated by verifying that the probability that at least one of the simulations that resulted in an establishment was a false positive (given the calculated establishment probability) was less than 0.1 %. The calculated establishment probabilities are shown in Figure C.1; the 95 % confidence intervals for the estimated probabilities are narrow enough that no error bars are shown in the figure, as they would be obscured by the symbols. The difference between the simulations and the numerical approximations is shown in Figure C.2; in general, the numerical results are very close to the simulation results, within the natural stochastic variation of the simulation results.

Rescue probability was calculated from simulations starting from a homozygote wild-type population of a given initial size  $N_0 = 100000$ . The population was considered to have gone extinct if the total number of cells reached zero, and to have been rescued if the number of heterozygote cells reached  $N_0$ . The calculated rescue probabilities are shown in Figure C.3; the 95 % confidence intervals for the estimated probabilities are narrow enough that no error bars are shown in the figure, as they would be obscured by the symbols. The difference between the simulations and the numerical approximations is shown in Figure C.4; again, the differences are within the stochastic variability of the simulations.

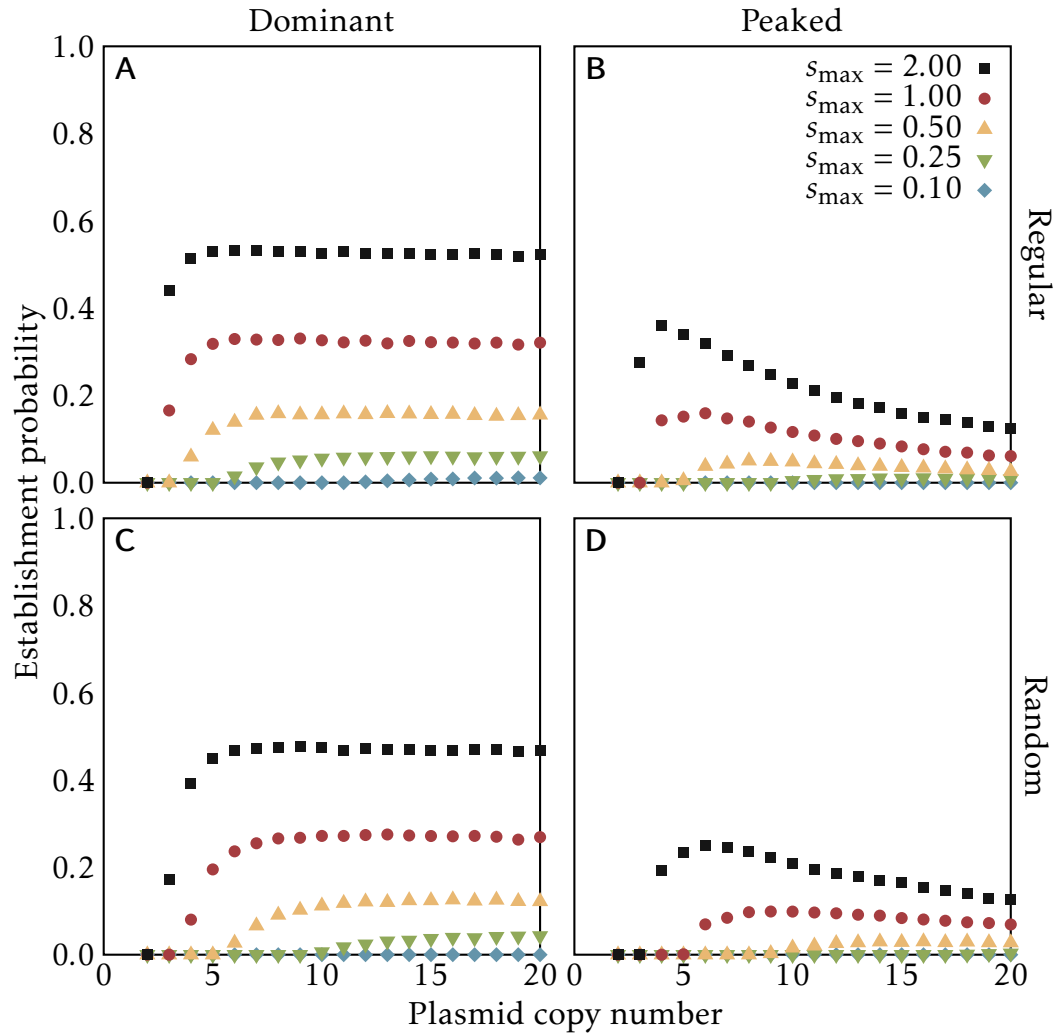


Figure C.1: Establishment probabilities of rescue mutations on a plasmid of a given copy number in the heterozygote advantage scenario, under the regular (top row) or random (bottom row) replication assumption, calculated from stochastic simulations. Colours of points indicate the fitness  $s_{\max}$  of all heterozygotes (with the dominant fitness function, left column) or the maximum fitness of heterozygotes (with the peaked fitness function, right column). For all cases,  $s_{\min} = -0.1$  and  $u = 10^{-6}$ .

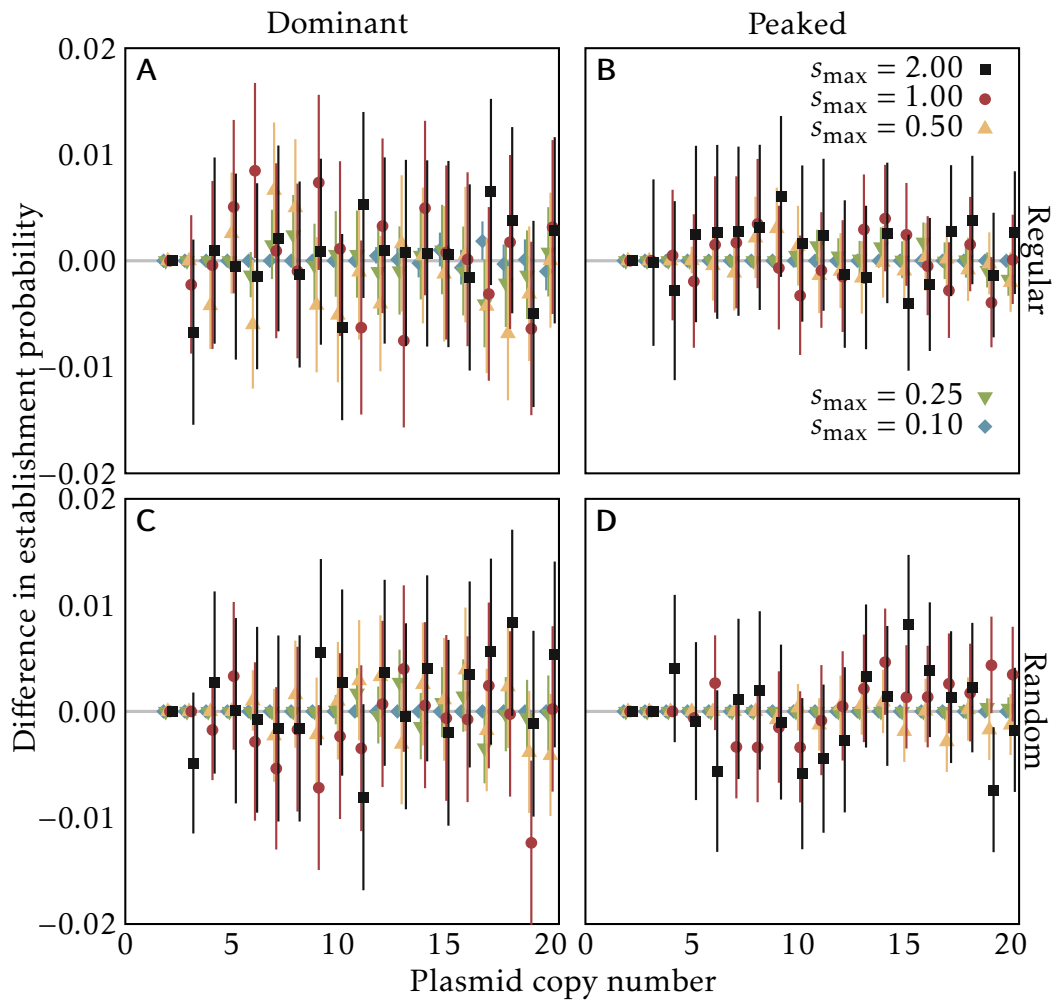


Figure C.2: Difference between the stochastic simulation and numerically derived establishment probabilities, with 95 % confidence intervals from the likelihood-ratio test. Simulation parameters described in the caption to Figure C.1.

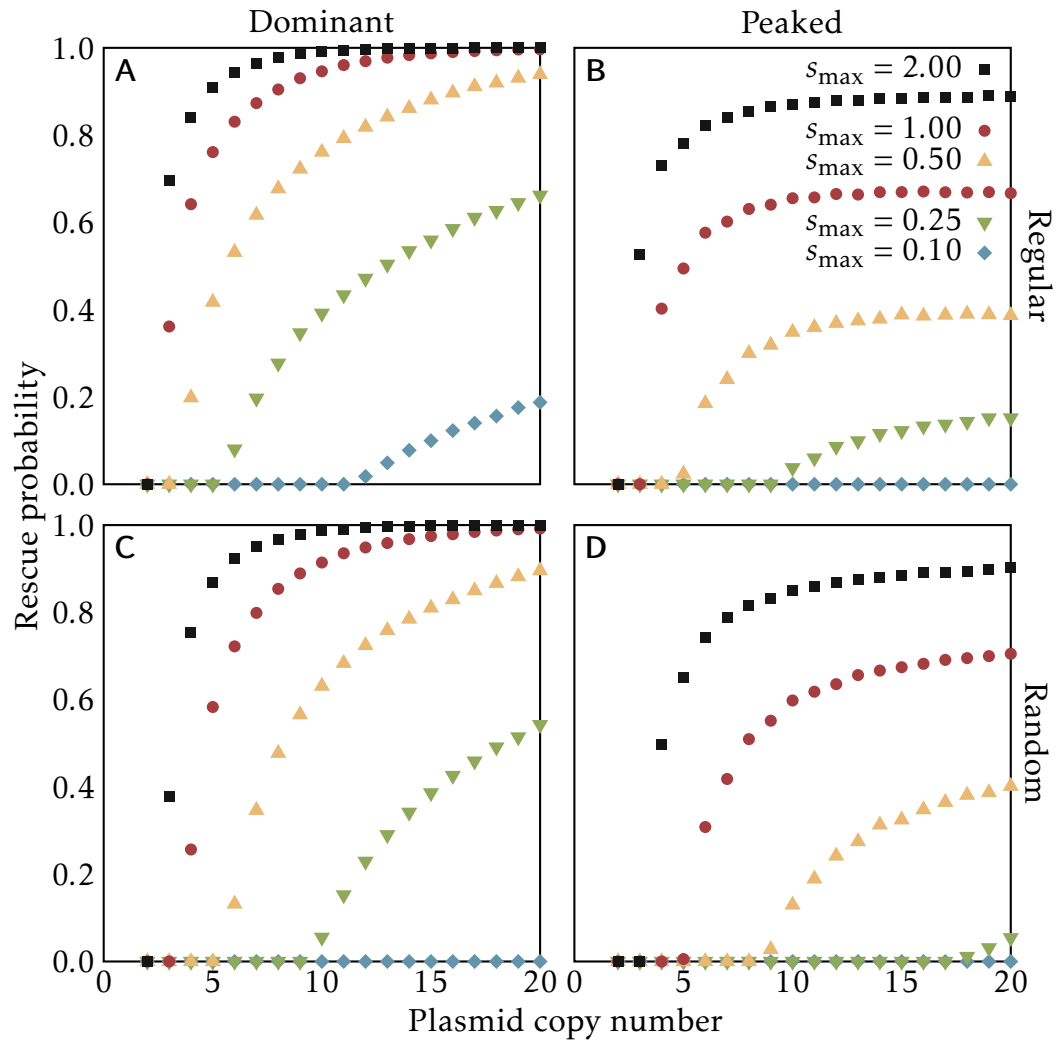


Figure C.3: Probabilities of a bacterial population being rescued by a mutation on a plasmid of a given copy number in the heterozygote advantage scenario, under the regular (top row) or random (bottom row) replication assumption, calculated from stochastic simulations. Colours of points indicate the fitness  $s_{\max}$  of all heterozygotes (with the dominant fitness function, left column) or the maximum fitness of heterozygotes (with the peaked fitness function, right column). For all cases,  $s_{\min} = -0.1$ ,  $u = 10^{-6}$ , and  $N_0 = 10^5$ .

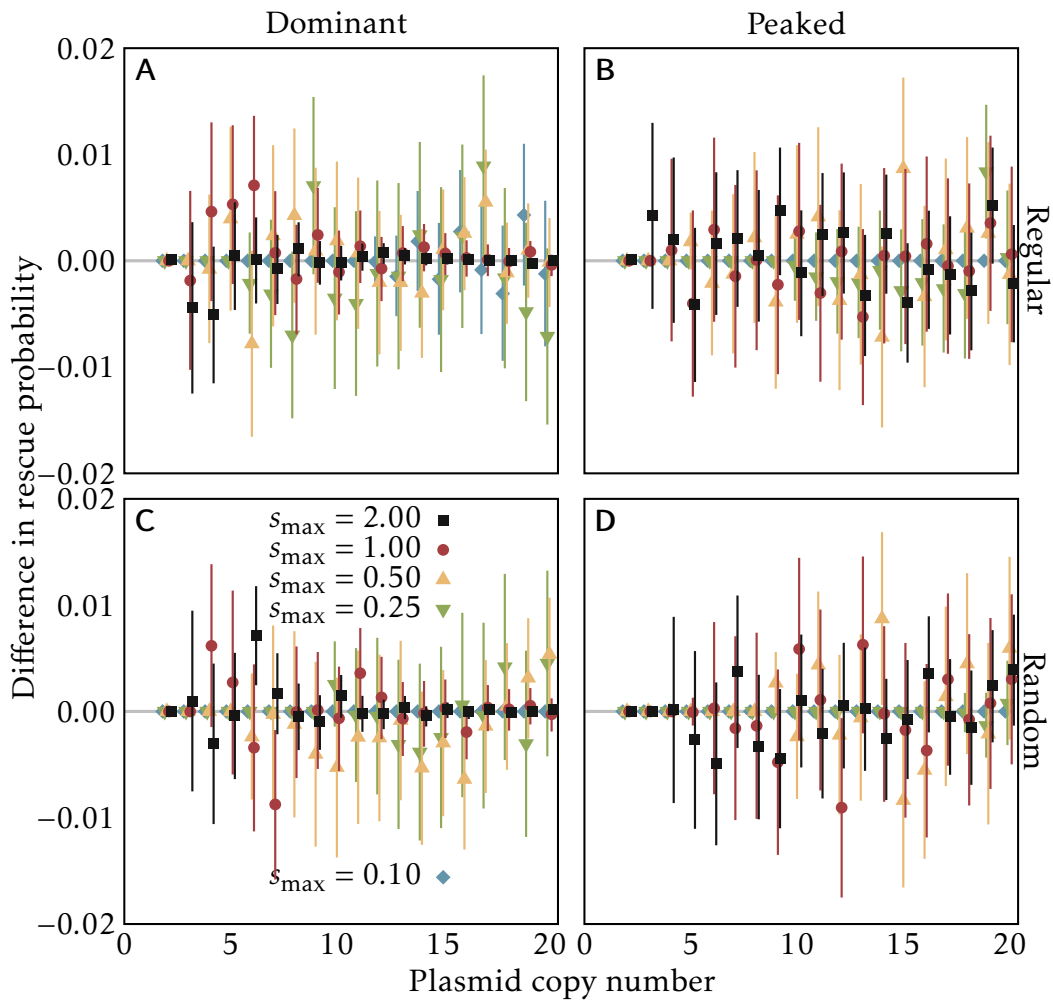


Figure C.4: Difference between the stochastic simulation and numerically derived rescue probabilities, with 95 % confidence intervals from the likelihood-ratio test. Simulation parameters described in the caption to Figure C.3.



## DECLARATION

I declare and affirm that the thesis entitled *Models of evolutionary rescue of bacteria and yeast by extrachromosomal genetic elements* was written by me, Ian Dewan, under the supervision of Dr Hildegard Uecker at the Max Planck Institute for Evolutionary Biology, Plön; apart from my supervisor's guidance, all of the content of the thesis is my own work. This thesis has not been submitted, in whole or in part, for a degree at any other university. The work herein was completed in accordance with the rules of good scientific practice of the German Science Foundation. I further declare that I have not had any academic degree withdrawn.

---

Ian Dewan

Kiel, \_\_\_\_\_