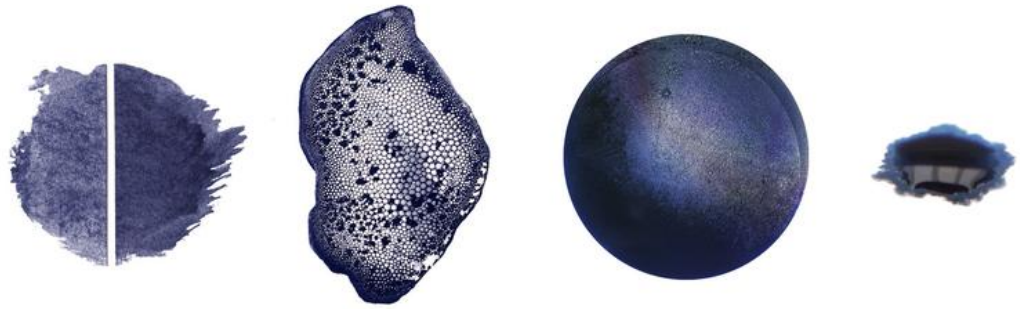
Plasmid dynamics in bacterial communities

A. N. Salazar de Dios

Supervised by M. G. J. de Vos, G. S. van Doorn & F. J. Weissing



MSc Programme in Ecology and Evolution
Erasmus Mundus Master Programme in Evolutionary Biology
February 2021

Abstract

Plasmids are considered a major driving force of bacterial ecology and evolution. Unlike bacterial chromosomes, plasmids frequently move horizontally from one cell to another. Although plasmid horizontal transmission is usually intraspecific, it can often occur between phylogenetically distant organisms. The collection of distinct organisms that a single plasmid can infect is called the host range. Despite many of the mechanistic barriers limiting host ranges are well known, it is not clear how plasmid host ranges evolve. Evolutionary studies on plasmid host ranges have shown these ranges are very dynamic. Here, I ask what is the expected change in plasmid host range assuming the plasmid's main objective is to increase its fitness. To answer this question, first it is necessary to identify the components of plasmid fitness on which natural selection can act upon. To derive such components, I develop an eco-evolutionary model of plasmid population dynamics. Specifically, I perform an evolutionary invasion analysis of a trait encoded on a plasmid that affects host growth, competition, conjugation, and plasmid loss. From this analysis, I obtain a fitness function to predict the invasion prospects of a plasmid mutant in an existing plasmid population. My results indicate that there are multiple ways in which plasmids can increase their fitness. I further my analysis by considering trade-offs between different life history traits of the plasmid and discuss possible implications of the plasmid fitness components on host range dynamics. In other words, I discuss how plasmid host ranges would be expected to evolve given the maximization of the plasmid fitness components under scrutiny.

---

1 Introduction

Plasmids are autonomously replicating mobile extra-chromosomal DNA molecules widely distributed across bacteria (Box 1). Typically, plasmids carry genes encoding plasmid-specific functions, such as self-replication, partitioning, and transfer (Bennett, 2008). In addition, plasmids may encode a myriad of mechanisms for bacterial local adaptation: resistance to antibiotics (Martínez, 2008) and heavy metals (Hall et al., 2017), metabolism of rare carbon compounds (Brinkmann et al., 2018), and production of bacteriocins and

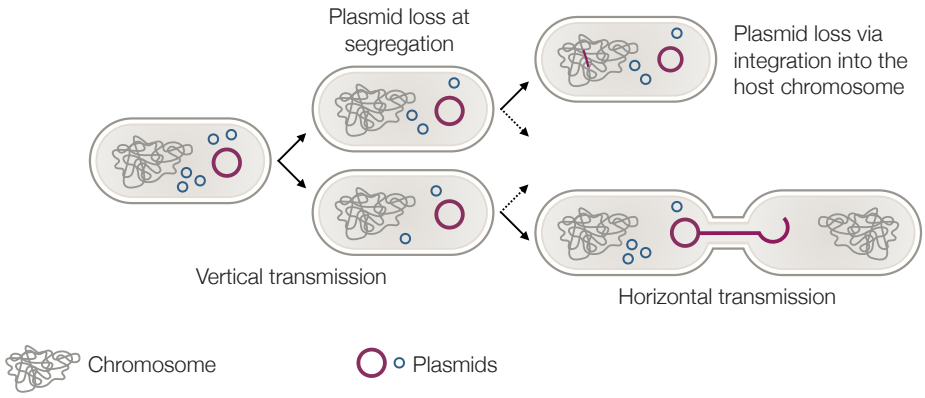
virulence factors (Buchrieser et al., 2000; Brown et al., 2006), enabling their bacterial hosts to colonize and compete in natural communities (Levin, 1993; Rankin et al., 2011).

Plasmids are among the main drivers of horizontal gene transfer (HTG) in bacterial populations. HGT is mediated by four main mechanisms: conjugation (Lederberg and Tatum, 1946), transduction (Zinder and Lederberg, 1952), transformation (Griffith, 1928), and vesiduction (Erdmann et al., 2017). All of these mechanisms give plasmids and other mobile genetic elements control over their transmission, independent of the transmission of the bacterial chromosome. This difference in transmission patterns creates a potential conflict between the fitness interests of the host chromosome and the fitness interests of the plasmid. Importantly, plasmid horizontal transmission can occur between phylogenetically distant organisms. The collection of different organisms a single plasmid can infect is called the *host range* of the plasmid (Norman et al., 2009). Host ranges can vary from a few members of a bacterial class to members of different domains of life (White et al., 1985). Mechanisms limiting host ranges include replication barriers, such as the number and structure of replication origins, mobilization barriers such as surface exclusion and host defense systems (Wheatley and MacLean, 2020), and other structural barriers such as the number of sites for restriction enzymes (Thomas and Nielsen, 2005; Jain and Srivastava, 2013).

Studies on plasmid host range evolution have shown host ranges are very dynamic (Figure 1). For example, evolutionary experiments of the plasmid pMS0506 have shown that, when evolved in *Shewanella oneidensis* during 500 generations, pMS0506 exhibits a shift in host range compared to the ancestral pMS0506. That is, after specialization in *S. oneidensis*, the evolved pMS0506 is no longer stable in hosts where the ancestral pMS0506 is (Sota et al., 2010). In contrast, when evolved in *Pseudomonas moraviensis*, pMS0506 expands its host range compared to the ancestral plasmid (Loftie-Eaton et al., 2016). Notably, most genomic changes during specialization observed in both experiments occurred in the plasmid. These results highlight the importance of plasmid evolution in shaping plasmid host ranges, as well as the difficulties to formulate general conclusions from single experiments.

In this master project, I focus on plasmid evolution and its consequences on host range dynamics. I assume there is a trait x encoded on a plasmid that affects host growth, competition, conjugation, and plasmid loss. To study plasmid evolution via the evolution of trait x, I develop an eco-evolutionary model of plasmid population dynamics and perform an evolutionary invasion analysis. The result of an invasion analysis is a fitness expression that predicts the fate of an invading plasmid mutant in a resident plasmid population. For this project, such expression is a function of plasmid fitness that depends on the trait x. I discern this function to answer two questions: *What are the components of plasmid fitness?* and whether *Are those components context depend?*. Understanding the components of plasmid fitness on which natural selection can act upon provides insights to understand when and why would plasmids evolve broader or narrower host ranges under specific conditions. I focus my discussion in the possible implications of the plasmid fitness components one general question: *When do plasmid host ranges are expected to be maintained or expanded?*

BOX 1 | Key features of plasmids



- Plasmids coexist with other plasmids and the host chromosome within bacterial cells
- Plasmids vary in size and copy number
- Plasmids impose a metabolic cost on their hosts
- (Some) plasmids encode beneficial genes for the host
- Plasmids are evolutionary individuals with fitness interests of their own

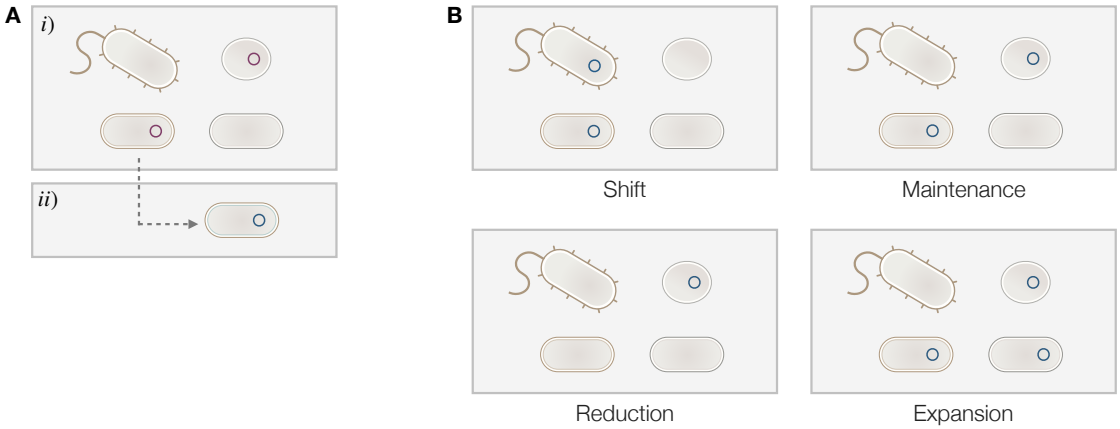


Figure 1 Host range dynamics. **A** | (i) An ancestral bacterial community in which a plasmid (purple) can infect two different species. (ii) Plasmid-host co-evolution results in an evolved plasmid (blue) and an evolved host. **B** | Possible changes in host range after introducing the evolved plasmid (blue) into the ancestral bacterial community.

2 An eco-evolutionary model of plasmid population dynamics

The ecological model

My model is an extension of a model described in Daras, 2018. I consider a large bacterial host population where cells can belong to two distinct states or sub-populations: plasmid-free, N_0, and plasmid-bearing, N_1, (Figure 2). N_0 and N_1 cells are born at rates r_0 and r_1, respectively. Plasmid-free cells can become plasmid-bearing cells as a result of plasmid horizontal transmission. I assume horizontal transmission relies on conjugation, which occurs at a conjugation rate constant c_1. Similarly, plasmid-bearing cells can become plasmid-free cells as a result of plasmid loss, which occurs at a rate constant λ_1. Competition is characterized by the Lotka-Volterra equations (Lotka, 1924; Volterra, 1931), which describe an additive effect of competition within and between host sub-populations. The following system of ordinary differential equations describes the ecological model. The notation $\dot{X}$ stands for a time derivative. All model variants are listed and explained in Table 1.

$$\dot{N}_0 = N_0 (r_0 - \alpha_{00} N_0 - \alpha_{01} N_1) + \lambda_1 N1 - N_0 c_{01} N_1 \tag{1a}$$

$$\dot{N}_1 = N_1 (r_1 - \alpha_{10} N_0 - \alpha_{11} N_1) + c_{01} N_0 N_1 - \lambda_1 N_1 \tag{1b}$$

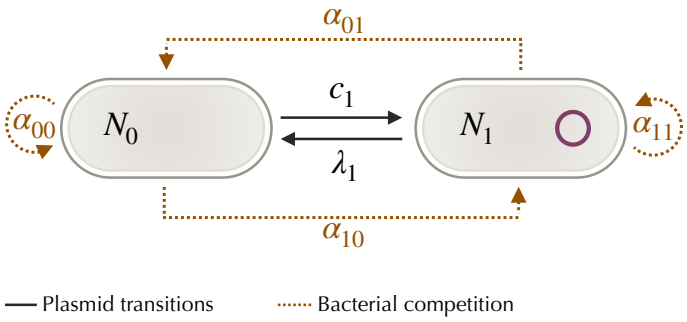


Figure 2 A Lotka-Volterra competition model for plasmid population dynamics. Host states, plasmid-free (N_0) and plasmid-bearing (N_1), are indicated by ellipsoids. A bacterial plasmid is illustrated in purple. Gray arrows correspond to transitions between host states and their associated letters correspond to their rates: N_0 cells become N_1 cells at a conjugation rate c_1, while N_1 cells become N_0 cells at a loss rate λ_1. Host competition is depicted in brown arrows. α_{ij} describes the effects of competition by host j over host i.

The eco-evolutionary model

Here, I assume there exist a plasmid-encoded allele, x, that affects host growth, competition, horizontal transmission and plasmid loss. In other words, all rate constants described in equation 1 are now functions of x. Hence, the system of equations 1 can be rewritten as:

$$\dot{N}_0 = N_0 (r_0 - \alpha_{00} N_0 - \alpha_{01}(x) N_1) + \lambda_1(x) N1 - N_0 c_{01}(x) N_1 \tag{2a}$$

$$\dot{N}_1 = N_1 (r_1(x) - \alpha_{10}(x) N_0 - \alpha_{11}(x,x) N_1) + c_{01}(x) N_0 N_1 - \lambda_1(x) N_1 \tag{2b}$$

To consider the evolution of the allele x, I assume x can mutate, giving rise to a mutant allele, x_m, and a new host state, N_m (Figure 3). Now, N_0, N_1 and N_m are born at

rates r_0, $r_1(x)$ and $r_1(x_m)$, respectively. Events of plasmid horizontal transmission, plasmid loss and host competition in N_1 and N_m differ only in their dependency on x or x_m, respectively. Importantly, all traits involving two plasmid-bearing cells, such as plasmid conjugation or host competition between N_1 and N_m cells are assumed to be influenced by the plasmid-encoded allele, x or x_m, in *each* host cell. For this reason, all traits involving two plasmid-bearing cells are functions of two parameters. For instance, $c_{11}(x_m, x)$ describes the conjugation rate between N_1 and N_m cells for the horizontal transfer of a mutant plasmid into a host carrying the resident plasmid. Likewise, $c_{11}(x, x_m)$ describes the horizontal transfer of a resident plasmid into a host carrying the mutant plasmid — notice order of the arguments on each function changes. Notably, I assume co-infections are not possible. In other words, different plasmids cannot coexist within the same host cell. This is supported by assuming that the plasmids encoding for x and x_m belong to the same incompatibility group (Del Solar et al., 1998). To model plasmid incompatibility, I assume that, immediately after conjugation between two plasmid-bearing cells (*i.e.* between N_1 and N_m cells), one of the plasmids prevails within the cell with certain probability. For the mutant allele, x_m, this probability is given by $p(x_m, x)$, while for the resident allele, x, this probability is given by $p(x, x_m)$. Therefore, the effective conjugation rate between two plasmid-bearing cells is given by $\varphi(x_m, x) = c_{11}(x_m, x)p(x_m, x)$ for the transition from N_1 to N_m and by $\varphi(x, x_m) = c_{11}(x, x_m)p(x, x_m)$ for the transition from N_m to N_1. The full eco-evolutionary model is described by the following system of equations. The notation $\dot{X}$ stands for a time derivative. All model variants are listed in Table 1.

$$\begin{aligned}\dot{N}_0 =& N_0\left(r_0 - \alpha_{00}N_0 - \alpha_{01}(x)N_1 - \alpha_{01}(x_m)N_m\right) + \lambda_1(x)N1 + \lambda_1(x_m)N_m \\ & - N_0\left(c_{01}(x)N_1 + c_{01}(x_m)N_m\right)\end{aligned} \tag{3a}$$

$$\begin{aligned}\dot{N}_1 =& N_1\left(r_1(x) - \alpha_{10}(x)N_0 - \alpha_{11}(x,x)N_1 - \alpha_{11}(x,x_m)N_m\right) \\ & + \left(c_{01}(x)N_0 + \varphi(x,x_m)N_m - \lambda_1(x)\right)N_1\end{aligned} \tag{3b}$$

$$\begin{aligned}\dot{N}_m =& N_m\left(r_1(x_m) - \alpha_{10}(x_m)N_0 - \alpha_{11}(x_m,x)N_1 - \alpha_{11}(x_m,x_m)N_m\right) \\ & + \left(c_{01}(x_m)N_0 + \varphi(x_m,x)N_1 - \lambda_1(x_m)\right)N_m\end{aligned} \tag{3c}$$

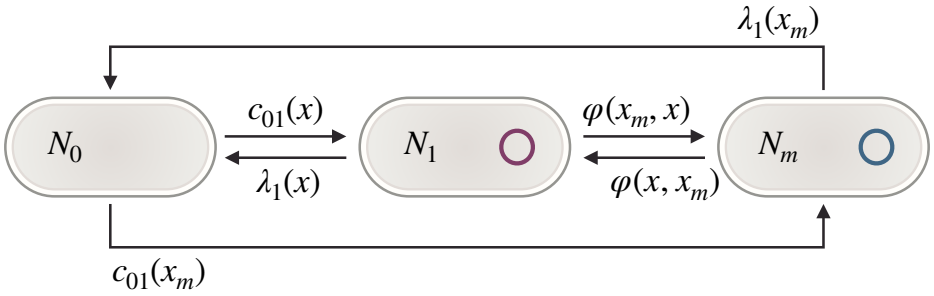


Figure 3 An eco-evolutionary model for plasmid population dynamics. Host states, plasmid-free, N_0, plasmid- N_1 and N_m, are indicated in ellipsoids. The resident plasmid is shown in purple and the mutant plasmid is shown in blue. Arrows depict transitions between host states and their associated letters correspond to the rates of those transitions. Bacterial competition is not shown.

Table 1 Notation of variables and parameters of all models

Variables	
N_0	Density of plasmid-free cells
N_1	Density of cells carrying a *resident* plasmid
N_m	Density of cells carrying a *mutant* plasmid
f_A	Frequency of cells carrying the *resident* plasmid
f_B	Frequency of cells carrying the *mutant* plasmid
n	Total population density
Parameters	
r_0	Plasmid-free cells maximal per-capita growth rate
$r_1(u)$	Plasmid-bearing cells maximal per-capita growth rate, where $u \in \{x, x_m\}$
α_{00}	Plasmid-free intra-specific competition coefficient
$\alpha_{ij}(u)$	Plasmid-bearing and plasmid-free cells competition where $i, j \in \{0, 1\}$ and $i \neq j$ and $u \in \{x, x_m\}$

$\alpha_{11}(u,u)$	Plasmid-bearing cells competition coefficient, where $u \in \{x, x_m\}$
$\alpha_{11}(u,v)$	Plasmid-bearing cells competition coefficient, where $u, v \in \{x, x_m\}$ and $u \neq v$
$\lambda_1(x_k)$	Plasmid loss rate, where $k \in \{1, m\}$
$c_{ij}(x_k)$	Plasmid conjugation rate, where $i \neq j$ and $i, j \in \{0, 1\}$ and $k, l \in \{1, m\}$
$\rho(x_k, x_l)$	Probability of plasmid k of winning plasmid-plasmid competition where $k \neq l$ and $k, l \in \{1, m\}$
$\varphi(x_i, x_j)$	Effective conjugation rate between plasmid-bearing cells where $i, j \in \{1, m\}$

3 Evolution analysis

To study the evolution of the plasmid-encoded allele, x, I performed an evolutionary invasion analysis (Metz et al., 1992). In an evolutionary invasion analysis, the population is assumed to be initially *monomorphic* for the allele of interest. In my model, that means that all plasmids encode for the same allele x. This so-called *resident* population can potentially be invaded by a plasmid encoding a *mutant* allele x_m. The direction of evolution can be inferred by determining the alleles that can successfully invade the population. Importantly, I assumed the bacterial host does not co-evolve with the plasmid.

Studying the dynamics of the resident population

For my analysis, I assumed mutations to occur rarely enough such that the resident population reaches its dynamic equilibrium (attractor) before a new mutant arises. At equilibrium, the growth rate of the resident population equals zero. In this way, the resident equilibrium sets the environmental conditions that must be met for the mutant to invade. In the case of my model, the invasion condition could not be computed by directly solving for the dynamic equilibrium of equation (1) (*i.e.* setting equation (1) = 0). For this reason, I assumed selection to have weak effects on the phenotype. More specifically, I assumed that the effect of the plasmid-encoded allele on the phenotype is small. Mathematically, that means that all the effects of the plasmid on the different phenotypes of the host will be of order ϵ with $\epsilon \approx 0$. Importantly, the assumption of weak selection creates

a change of variables (Table 2). For instance, the loss rate of the resident plasmid, $\lambda_1(x)$, is now described by a baseline loss rate, λ, in addition to a small deviation in this baseline value caused by the plasmid, $\epsilon\sigma(x)$. In the following equations, all letters without subscripts are baseline parameters and all functions of x are the deviations in the value of those parameters caused by the plasmid trait. Selection can only act over those small deviations.

Table 2 Change of variables assuming weak selection

Variables	
r	Baseline bacterial replication rate
a	Baseline bacterial competition coefficient
c	Baseline plasmid conjugation rate
λ	Baseline plasmid loss rate
β	Advantage of a mutant plasmid in plasmid-plasmid competition
$\rho_1(x)$	Deviation in host growth rate (cost) caused by the plasmid-encoded trait x
$\sigma_1(x)$	Deviation in plasmid loss caused by the plasmid-encoded trait x
$\gamma_{0,1}(x)$	Deviation in conjugation rate between plasmid-free and plasmid-bearing cells driven by the plasmid-encoded trait x
$\gamma_{1,1}(x,x_m)$	Deviation in donor/recipient preference at conjugation between two plasmid-bearing cells caused by the plasmid-encoded traits x and x_m
$\alpha_{1,0}(x)$	Deviation in competition coefficient between plasmid-free and plasmid-bearing cells caused by the plasmid-encoded trait x
$\alpha_{1,1}(x,x_m)$	Deviation in competition coefficient between plasmid-bearing cells driven by the plasmid-encoded traits x and x_m

Under the new set of variables, the equilibrium of (1) is given by:

$$n^\star = \frac{r}{a}, \quad f_A^\star = 1 - \frac{\lambda}{cn^\star} \tag{4}$$

Where, n is the total population size and f_A is the frequency of plasmid-bearing cells. r is the baseline replication rate, a is the baseline host competition coefficient, λ is the basal loss rate and c is the baseline conjugation rate.

To quantify the deviation of the equilibrium by means of weak selection, I add sub-correction terms, $\epsilon\delta n$ and $\epsilon\delta f_A$ into equation (4):

$$n^\star = \frac{r}{a} + \epsilon\delta n, \quad f_A^\star = 1 - \frac{\lambda}{cn^\star} + \epsilon\delta f_A \tag{5}$$

Equation (5) indicates that the total population density at equilibrium depends on the baseline bacterial replication rate and the strength of bacterial competition. In other words, the total population density at equilibrium does not depend on the dynamics of the plasmid. On the other hand, the the frequency of plasmid-bearing cells depends on their growth dynamics as well as on the dynamics of plasmids (*i.e.* loss and horizontal transfer rates).

Stability analysis for the resident equilibrium

To determine the *local* stability of the *resident* equilibrium, equation (5), I derived the Jacobian matrix of the system, $\mathcal{J}_{res}$:

$$\mathcal{J}_{res} = \begin{pmatrix} -r & 0 \\ \frac{a\lambda}{cr^2}(cr - a\lambda) & \lambda - \frac{r}{a}c \end{pmatrix}$$

If the real part of the dominant eigenvalue of $\mathcal{J}_{res}$, $s(\mathcal{J}_{res})$, is strictly negative (*i.e.* < 0), then the system is *stable* at the resident equilibrium. Thus, the stability condition is given by $s(\mathcal{J}_{res}) = \lambda - \frac{r}{a}c < 0$. That is, as long as the plasmid loss rate is less than its conjugation rate, escalated by host grow, plasmid-free and plasmid-bearing can be stably maintained in the population.

Evaluating the gradient of selection

Now I introduce the plasmid-encoded mutant allele, x_m, into the resident population at equilibrium (equation (3) and Figure 3). The mutant will invade the resident population if its *invasion fitness*, or growth rate when rare, is greater than zero. Given that my model is fairly complicated, it is not possible to obtain a complete expression of the invasion fitness. Instead, I will focus on its first order approximation for values of x_m close to x, to derive the so-called *gradient of selection*, $G(x)$. The sign of $G(x)$ determines the direction of evolution. Positive selection gradients imply that the trait will increase, and negative selection gradients imply the trait will decrease. In other words, if $G(x) > 0$, the mutant allele x_m will invade if $x_m > x$, and if $G(x) < 0$, the mutant allele x_m will invade if $x_m < x$.

To derive $G(x)$, first, I assume plasmid-bearing cells and plasmid-free cells have reproductive values that are 'close' to one another because the effect of the evolving trait on the phenotype is small. In other words, plasmid-bearing cells are only slightly different from plasmid-free cells, such that selection on the plasmid is *weak*. Second, I assume the mutant and resident plasmid are only slightly different, so that the fitness difference between mutant and resident plasmids is small as well. For further details in how to obtain $G(x)$ see the Appendix.

For my eco-evolutionary model, the gradient of selection is given by:

$$\begin{aligned} G(x) = \epsilon \Bigg[2\beta \left(\frac{cr}{a} - \lambda \right) + \rho_1'(x) - \sigma_1'(x) + \frac{1}{2} \left(\frac{r}{a} - \frac{\lambda}{c} \right) \left(\gamma_{\{1,1\}}^{(0,1)}(x,x) - \gamma_{\{1,1\}}^{(1,0)}(x,x) \right) \\ + \frac{\lambda}{c} \gamma_{\{0,1\}}'(x) + \frac{\lambda}{c} \left(\alpha_{\{1,1\}}^{(1,0)}(x,x) - \alpha_{\{1,0\}}'(x) \right) - \frac{r}{a} \alpha_{\{1,1\}}^{(1,0)}(x,x) \Bigg] \end{aligned} \tag{6}$$

The notation $\zeta^{(1,0)}_{\{1,1\}}(y,x)$ refers to the partial derivative of ζ with respect to the first argument, y, and $\zeta^{(0,1)}_{\{1,1\}}(y,x)$ refers to the partial derivative of ζ with respect to the second argument, x. $\zeta'(x)$ is the derivative of ζ with respect to its unique argument x. Descriptions of all variables can be found in Table 2.

Equation (6) is the single most important equation of my results and it can be decomposed into the following *fitness components*:

$\beta\left(\frac{cr}{a}-\lambda\right)$	*Plasmid-plasmid competition*
$\rho_1'(x)$	*Plasmid cost*
$\sigma_1'(x)$	*Plasmid loss*
$\frac{1}{2}\left(\frac{r}{a}-\frac{\lambda}{c}\right)\left(\gamma^{(0,1)}_{\{1,1\}}(x,x)-\gamma^{(1,0)}_{\{1,1\}}(x,x)\right)$	*Plasmid horizontal transmission*
$\frac{\lambda}{c}\gamma'_{\{0,1\}}(x)+\frac{\lambda}{c}\left(\alpha^{(1,0)}_{\{1,1\}}(x,x)-\alpha'_{\{1,0\}}(x)\right)-\frac{r}{a}\alpha^{(1,0)}_{\{1,1\}}(x,x)$	*Host-host competition*

Here, the baseline parameters can be seen as properties of the host, while those parameters depending on x are the deviations in the values of those properties under control of the plasmid. It is worth mentioning that this decomposition of the selection gradient is not unique and it is subject to my own interpretation. The first component, for instance, is called *plasmid-plasmid competition* because it involves the advantage a mutant plasmid has in winning intracellular plasmid-plasmid competition, β. The term β appears exclusively in this factor, and it is multiplied by the rate of conjugation, c normalized by host growth r/a, minus the loss rate λ. Next, *plasmid cost* and *plasmid loss*, respectively, involve the deviations in host growth rate and plasmid loss caused by the plasmid-encoded trait, x. The fourth component, *plasmid horizontal transmission* features deviations in the preference of a host cell to be a donor or a recipient at conjugation events, $\gamma_{\{1,1\}}(x,x)$ multiplied by half the difference between growth minus the loss/conjugation ratio. Finally, I called the fifth component *host-host competition* because it contains all terms related to deviations in host competition coefficients caused by x. In the next sub-section, I will analyse in more detail plasmid cost, plasmid loss and plasmid horizontal transmission.

Trade-offs

Looking at the gradient of selection and its components (equation 6), it can be seen that there are several ways in which plasmids could increase their fitness. Let's assume, for instance, that the fitness of the plasmid is determined only by the component of plasmid loss, $\rho_1'(x)$. In such case, equation (6) would look like $G(x) = -\sigma_1'(x)$. This would imply that mutants with lower loss rates would always invade the population. Therefore, in the long run, loss rates would evolve towards zero. All other components can be isolated in a similar fashion. By only considering individual components, however, one would find plasmids would evolve towards infecting all host cells at an infinitely fast rate, causing no metabolic cost and having a perfect system that ensures it would not, by any means, get lost at cell division. Yet, in nature, such *Plasmid Demons* do not exist (Law, 1979). Hence, it is reasonable to assume there are constraints limiting the evolution of plasmids.

In biology, one of the most important types of constraints are trade-offs. Trade-offs stem from the negative correlations between the fitness components of the plasmid (or any evolving individual), such that increasing fitness by means of one trait, decreases fitness by the modification of a correlated trait (Stearns, 1992). For the purpose of this analysis, I will focus on pairs of traits. Specifically, I will focus in the trade-off between plasmid metabolic cost and plasmid loss, and plasmid metabolic cost and plasmid conjugation or horizontal transmission. Nevertheless, it is important to mention that trade-offs can involve multiple traits. Understanding these pair-wise trade-offs can allow me to interpret cases involving more than two traits or fitness components.

In the next two sections, I will provide justification for each trade-off, followed by its mathematical analysis and interpretation. When studying a specific trade-off, I assume all other phenotypic traits are *independent* of the evolving allele x. Therefore, all the derivatives not involved in the trade-off under scrutiny become zero.

(a) Bacterial growth and plasmid loss

There are several ways in which plasmids can decrease their loss rate. For small plasmids lacking active partitioning systems, for instance, an effective way to decrease their loss rate is to increase their copy number. In this way, small plasmids reduce the probability of producing plasmid-free cells when segregating plasmids at cell division (Friehs, 2004). With an increase in plasmid copy number, an increase in metabolic cost is expected, and therefore a decrease in host growth rate. Other mechanisms to reduce plasmid loss, such as active partitioning systems and dimerization resolution, are expected to increase metabolic cost as well (Brendler et al., 2004).

Assuming the allele x do not have an effect on any other trait, the selection gradient, $G_a(x)$, is given by:

$$G_a(x) = \epsilon \left(\rho_1'(x) - \sigma_1'(x)\right) \tag{7}$$

Here, $\rho_1'(x)$ and $\sigma_1'(x)$ exhibit a linear relationship, suggesting that for the plasmid there is no difference in investing in reducing the metabolic burden it imposes on its host versus investing in reducing its loss rate. Notably, all terms involved in this trade-off are under control of the plasmid-encoded trait x.

(b) Bacterial growth and plasmid conjugation

For the analysis of the trade-off between bacterial growth and plasmid conjugation, I assumed conjugation to be costly and the cost of conjugation to increase proportionally to its rate. Indeed, evolutionary experiments in *Escherichia coli* showed that plasmids that evolved lower conjugation rates imposed lower fitness costs in the ancestral bacterial background, whereas plasmids that evolved an increased conjugation rates imposed greater costs (Turner et al., 1998).

Assuming the allele x do not have an effect on any other trait, the selection gradient, $G_b(x)$, is given by:

$$G_b(x) = \epsilon \left(\rho_1'(x) + \lambda \frac{\gamma'_{\{0,1\}}(x)}{c} + \frac{cr - a\lambda}{2a} \frac{\gamma^{(0,1)}_{\{1,1\}}(x,x) - \gamma^{(1,0)}_{\{1,1\}}(x,x)}{c} \right) \tag{8}$$

Here, it is important to notice the presence of baseline parameters. As previously mentioned, the baseline parameters can be seen as properties of the host and those parameters depending on x as the deviations in control of the plasmid. The third factor in equation (8) suggests that, if there is a mutation causing asymmetry in conjugation, the plasmid will have a fitness advantage if this mutation provides the plasmid with the ability to participate in a conjugation event where its host cell acts as a donor more often than as a recipient. The importance of this fitness effect scales with the opportunity for the plasmid spread by conjugation, which depends on properties of the host population, such as its density at equilibrium (equation 4).

Generally, equation (8) suggests that conjugation symmetry, or the tendency of a plasmid-bearing cell to act as a donor or a recipient, together with plasmid loss rate, will define whether a plasmid invests in reducing its cost, thus increasing its vertical transmission, or in increasing its horizontal transmission.

4 Discussion

I presented an evolutionary invasion analysis of a trait x encoded on a plasmid that affects host growth, competition, horizontal transmission, and plasmid loss. I separated the plasmid fitness into five components: plasmid-plasmid competition, plasmid effects on host growth (cost), plasmid loss, plasmid horizontal transmission, and host-host competition. From these components, I found that plasmid-plasmid competition depends exclusively on host properties, while plasmid effects on host growth (cost) and plasmid loss depend exclusively on the plasmid traits. In contrast, plasmid horizontal transmission and host-host competition depend on both the plasmid evolving trait and the host properties. To understand the evolutionary fate of x, I considered and analyzed pairwise trade-offs between different fitness components of the plasmid. From this analysis, I found that there

exists a linear relationship between plasmid cost and plasmid loss. That is, the fitness advantage that a plasmid can obtain from ameliorating its cost is linearly proportional to the fitness advantage that it can obtain from reducing its loss. Importantly, both plasmid loss and plasmid cost only affect the plasmid's vertical transmission. For the trade-off between plasmid cost and horizontal transmission (*i.e.* conjugation), my model suggests the baseline loss rate in a given host and the host donor/recipient preference at conjugation play an important role in determining plasmid evolution. This last trade-off can be interpreted as a compromise between vertical and horizontal modes of transmission. Investing in vertical transmission can lead to plasmid specialization and the plasmid's decision to specialize, as opposed to investing in horizontal transmission (that is, generalize) is weighted by how stable the plasmid is in a given host population. Looking back at my general question *'When do plasmid host ranges are expected to be maintained or expanded?'*, I expect plasmid host ranges to be reduced or maintained when it is more advantageous for the plasmid to specialize. On the other hand, I expect host ranges to expand when it is more advantageous for the plasmid to invest in horizontal transmission.

It is now worth stressing the limitations of my conclusions. First, my model relies on the assumption that changes in the value of x can be achieved via small mutations and that selection can only have small effects on the phenotype. Nonetheless, it has been seen that small mutations do not have significant effects on the fitness of plasmids (San Millan et al., 2014). In addition, several studies had reported plasmids increase their stability through large mutations (Loftie-Eaton et al., 2016; Wein et al., 2019). However, these assumptions are generally unavoidable if one wants to arrive at analytical results of complex models. One way to find out if my conclusions are robust to the weak-selection assumption is to develop a simulation model that allows for strong selection, and compare its predictions against my analytical results. Second, I described bacterial interactions using the Lotka-Volterra competition model, which has been shown to fail to capture pairwise interactions in bacterial communities because these communities violate the assumptions of additive effects of interactions and universality (Momeni et al., 2017). Likewise, it is important to address the constraints and consequences of only considering the evolution of the plasmid. In real host-plasmid systems, the plasmid *and* the host are always co-

evolving. Thus, to do a comprehensive analysis of the components of plasmid fitness, it would be necessary to include the host cell perspective. In this way, both the host and the plasmid would have shared control over the processes I assumed to be in exclusive control of the plasmid. In other words, instead of having parameters depending exclusively on the plasmid-encoded trait, x, I would have parameters depending on both a plasmid-encoded trait, x, and a host- or chromosome-encoded trait, y. Therefore, co-evolution can potentially alter my conclusions. Co-culture studies have highlighted the importance of co-adaptation in explaining plasmid rates of horizontal transfer and plasmid stability (Harrison and Brockhurst, 2012). For instance, in a study examining the cost of the conjugative plasmids R1 and RP4 in *E. coli*, the authors showed that a reduction in plasmid cost occurred by adaptations both on the plasmid alone and on the bacterial host alone (Dahlberg and Chao, 2003). The authors suggest there exists a trade-off between horizontal and vertical transmission (cost) of plasmids, and that the consequences of this trade-off are driven by host-directed suppression of plasmid transfer. In a different study, plasmid-host co-evolution facilitated the acquisition of a second plasmid by increasing plasmid permissiveness via chromosomal mutations, favouring the emergence of multidrug resistant cells (Jordt et al., 2020).

Predicting plasmid host range dynamics requires a comprehensive understanding of plasmid fitness at different levels of selection. The examples mentioned in the previous paragraph highlight how intricate this task is, as well as its importance. Gaining insight into the drivers of plasmid host ranges would allow tackling a myriad of unanswered questions; from practical questions regarding the evolution and spread of drug resistance, to more fundamental questions such as why are plasmids widely spread in bacteria but rarely found in other organisms.

5 Epilogue: advice to a follow-up student

Hello future student! First of all, best of luck with choosing a research topic and a question for your project. That is, without any doubt, one of the hardest parts of any project. If you are planning to follow-up on my project, here are some ideas and thoughts I can offer to guide you.

It is a truth universally acknowledged that projects are never *finished*, only *abandoned*. I always have felt I've abandoned my projects when things were getting the most interesting. That might have to do with the fact that I've only worked in small projects, but perhaps it is a shared feeling for longer projects too. In any case, in the following paragraphs I'll list, in order of feasibility, possible directions my project could have taken if I had more time.

Explore the unstudied fitness components

There are two fitness components I've left unexplored: plasmid-plasmid competition and host-host competition. An important reason why I decided to focus on the other three components is that there is more literature on them and because it is easier to reason about them. Microbial interactions, and in particular plasmid interactions, are largely understudied. However, there is a comprehensive review of the known interactions between plasmids published recently (Dionisio et al., 2019). Here, I would choose a particular system and develop a more mechanistic model for those components with the hope of drawing some generalizations from it. For the component on host-host competition I would follow the same approach.

Co-evolution

As I mentioned in my discussion, co-adaptation plays a major role in the evolution of plasmids and their bacterial hosts (Harrison and Brockhurst, 2012). For this reason, as well as from a technical standpoint, exploring the dynamics of two co-evolving traits is a rather logical next step. However, as you may have noticed, my model is already quite

complicated and I had to make use of some *advanced* techniques, such as assuming weak selection. Interpreting the results of analytical models can go from hard to impossible very quickly, and a co-evolutionary model is probably at the limit of tractability, at least for a master project. Fortunately, there are ways out, such as doing *numerical simulations*. Although easier to implement, numerical simulation can be a lot of work because there are a lot of dimensions to explore. Thus, here I would again recommend to focus on a particular system to limit the parameter space.

Including a second host species

To understand the dynamics of plasmids in bacterial communities, a useful thing to do would be to look at an actual bacterial community. That is, to look at an assembly of two (or more) bacterial species sharing the same environment. I strongly believe that analyzing an eco-evolutionary model of two bacterial species and a plasmid through analytical methods is virtually an impossible task for a master project. Therefore, I would recommend taking a look at individual-based simulations. However, bear in mind that individual-based simulations require a more mechanistic view of the processes driving plasmid population dynamics. Hence, once again, you'll need to limit yourself to more specific systems.

Acknowledgements

First, I'd like to thank Marjon, as it was because of her commitment and enthusiasm this project was ever born. I would also like to thank her for her contagious and encouraging fascination for microbiology, which translated beyond screens. Second, I would like to thank Sander, who became my main support as we moved towards a fully theoretical project. I met Sander during my Evolutionary Theory course and his lectures were my favorites in the entire program. Sander is an incredibly talented teacher, always clear and concise, always highlighting the most relevant things. Always insightful. Having the opportunity to learn again from him was a huge pleasure and an honor. Third, I would like to thank Franjo for sharing some of his wisdom with me and caring. Next, I would

like to thank Misshelle, with whom I shared my learning experience and from whom I also learned a lot. I wish her the best for her Ph.D.

Finally, I would like to thank The Flying Sheepman: Ben, Niccole and Celia for the many meals, movies, walks and conversations. Life as it is would not be possible without their friendship; and the Café con Ciencia, specially Rafael, Charly and Ayari, for making science fun and exiting and being a fantastic group of ñoños. You made of Thursdays my favorite day of the week. Also, I would like to thank Alberto, because I want to send him a copy of this report and he would complain if his name is not in here.

References

Bennett, P. (2008). Plasmid encoded antibiotic resistance: Acquisition and transfer of antibiotic resistance genes in bacteria. *British journal of pharmacology*, *153*(S1), S347–S357.

Brendler, T., Reaves, L., & Austin, S. (2004). Interplay between plasmid partition and postsegregational killing systems. *Journal of bacteriology*, *186*(8), 2504–2507.

Brinkmann, H., Göker, M., Koblížek, M., Wagner-Döbler, I., & Petersen, J. (2018). Horizontal operon transfer, plasmids, and the evolution of photosynthesis in rhodobacteraceae. *The ISME journal*, *12*(8), 1994–2010.

Brown, S. P., Le Chat, L., De Paepe, M., & Taddei, F. (2006). Ecology of microbial invasions: Amplification allows virus carriers to invade more rapidly when rare. *Current Biology*, *16*(20), 2048–2052.

Buchrieser, C., Glaser, P., Rusniok, C., Nedjari, H., d'Hauteville, H., Kunst, F., Sansonetti, P., & Parsot, C. (2000). The virulence plasmid pwr100 and the repertoire of proteins secreted by the type iii secretion apparatus of shigella flexneri. *Molecular microbiology*, *38*(4), 760–771.

Dahlberg, C., & Chao, L. (2003). Amelioration of the cost of conjugative plasmid carriage in *Eschericha coli* K12. *Genetics*, *165*(4), 1641–1649.

Daras, I. M. (2018). *The Persistence of Bacterial Plasmids* (Master's thesis). Rijksuniversiteit Groningen. The Netherlands.

Del Solar, G., Giraldo, R., Ruiz-Echevarrıa, M. J., Espinosa, M., & Dıaz-Orejas, R. (1998). Replication and control of circular bacterial plasmids. *Microbiology and molecular biology reviews*, *62*(2), 434–464.

Dionisio, F., Zilhão, R., & Gama, J. A. (2019). Interactions between plasmids and other mobile genetic elements affect their transmission and persistence. *Plasmid*, *102*, 29–36.

Erdmann, S., Tschitschko, B., Zhong, L., Raftery, M. J., & Cavicchioli, R. (2017). A plasmid from an antarctic haloarchaeon uses specialized membrane vesicles to disseminate and infect plasmid-free cells. *Nature microbiology*, *2*(10), 1446–1455.

Friehs, K. (2004). Plasmid copy number and plasmid stability. *New trends and developments in biochemical engineering*, 47–82.

Griffith, F. (1928). The significance of pneumococcal types. *Epidemiology & Infection*, *27*(2), 113–159.

Hall, J. P., Williams, D., Paterson, S., Harrison, E., & Brockhurst, M. A. (2017). Positive selection inhibits gene mobilization and transfer in soil bacterial communities. *Nature ecology & evolution*, *1*(9), 1348–1353.

Harrison, E., & Brockhurst, M. A. (2012). Plasmid-mediated horizontal gene transfer is a coevolutionary process. *Trends in microbiology*, *20*(6), 262–267.

Jain, A., & Srivastava, P. (2013). Broad host range plasmids. *FEMS microbiology letters*, *348*(2), 87–96.

Jordt, H., Stalder, T., Kosterlitz, O., Ponciano, J. M., Top, E. M., & Kerr, B. (2020). Coevolution of host–plasmid pairs facilitates the emergence of novel multidrug resistance. *Nature ecology & evolution*, *4*(6), 863–869.

Law, R. (1979). Optimal life histories under age-specific predation. *The American Naturalist*, *114*(3), 399–417.

Lederberg, J., & Tatum, E. L. (1946). Gene recombination in escherichia coli. *Nature*, *158*(4016), 558–558.

Levin, B. R. (1993). The accessory genetic elements of bacteria: Existence conditions and (co) evolution. *Current opinion in genetics & development*, *3*(6), 849–854.

Loftie-Eaton, W., Yano, H., Burleigh, S., Simmons, R. S., Hughes, J. M., Rogers, L. M., Hunter, S. S., Settles, M. L., Forney, L. J., Ponciano, J. M., et al. (2016). Evolutionary paths that expand plasmid host-range: Implications for spread of antibiotic resistance. *Molecular biology and evolution*, *33*(4), 885–897.

Lotka, A. J. (1924). *Elements of physical biology*. Williams & Wilkins.

Martínez, J. L. (2008). Antibiotics and antibiotic resistance genes in natural environments. *Science*, *321*(5887), 365–367.

Metz, J. A., Nisbet, R. M., & Geritz, S. A. (1992). How should we define 'fitness' for general ecological scenarios? *Trends in ecology & evolution*, *7*(6), 198–202.

Momeni, B., Xie, L., & Shou, W. (2017). Lotka-volterra pairwise modeling fails to capture diverse pairwise microbial interactions. *Elife*, *6*, e25051.

Norman, A., Hansen, L. H., & Sørensen, S. J. (2009). Conjugative plasmids: Vessels of the communal gene pool. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1527), 2275–2289.

Rankin, D. J., Rocha, E. P., & Brown, S. P. (2011). What traits are carried on mobile genetic elements, and why? *Heredity*, *106*(1), 1–10.

San Millan, A., Peña-Miller, R., Toll-Riera, M., Halbert, Z., McLean, A., Cooper, B., & MacLean, R. (2014). Positive selection and compensatory adaptation interact to stabilize non-transmissible plasmids. *Nature communications*, *5*(1), 1–11.

Sota, M., Yano, H., Hughes, J. M., Daughdrill, G. W., Abdo, Z., Forney, L. J., & Top, E. M. (2010). Shifts in the host range of a promiscuous plasmid through parallel evolution of its replication initiation protein. *The ISME journal*, *4*(12), 1568–1580.

Stearns, S. C. (1992). *The evolution of life histories*. Oxford: Oxford University Press.

Thomas, C. M., & Nielsen, K. M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature reviews microbiology*, *3*(9), 711–721.

Turner, P. E., Cooper, V. S., & Lenski, R. E. (1998). Tradeoff between horizontal and vertical modes of transmission in bacterial plasmids. *Evolution*, *52*(2), 315–329.

Volterra, V. (1931). Leçons sur la théorie mathématique de la lutte pour la vie. 1931. *Gauthier-Villars, Paris*.

Wein, T., Hülter, N. F., Mizrahi, I., & Dagan, T. (2019). Emergence of plasmid stability under non-selective conditions maintains antibiotic resistance. *Nature communications*, *10*(1), 1–13.

Wheatley, R. M., & MacLean, R. C. (2020). Crispr-cas systems restrict horizontal gene transfer in pseudomonas aeruginosa. *The ISME Journal*, 1–14.

White, F. F., Taylor, B. H., Huffman, G. A., Gordon, M. P., & Nester, E. W. (1985). Molecular and genetic analysis of the transferred dna regions of the root-inducing plasmid of agrobacterium rhizogenes. *Journal of bacteriology*, *164*(1), 33–44.

Zinder, N. D., & Lederberg, J. (1952). Genetic exchange in salmonella. *Journal of bacteriology*, *64*(5), 679.