Why There Are No Essential Genes on Plasmids

Samuel J. Tazzyman[1] and Sebastian Bonhoeffer*[,1]

[1]Theoretical Biology, Institute of Integrative Biology (IBZ), ETH Zürich, CH 8092, Zürich, Switzerland

*Corresponding author: E-mail: sebastian.bonhoeffer@env.ethz.ch.

Associate editor: Manolo Gouy

Abstract

Mobile genetic elements such as plasmids are important for the evolution of prokaryotes. It has been suggested that there are differences between functions coded for by mobile genes and those in the "core" genome and that these differences can be seen between plasmids and chromosomes. In particular, it has been suggested that essential genes, such as those involved in the formation of structural proteins or in basic metabolic functions, are rarely located on plasmids. We model competition between genotypically varying bacteria within a single population to investigate whether selection favors a chromosomal location for essential genes. We find that in general, chromosomal locations for essential genes are indeed favored. This is because the inheritance of chromosomes is more stable than that for plasmids. We define the "degradation" rate as the rate at which chance genetic processes, for example, mutation, deletion, or translocation, render essential genes nonfunctioning. The only way in which plasmids can be a location for functioning essential genes is if chromosomal genes degrade faster than plasmid genes. If the two degradation rates are equal, or if plasmid genes degrade faster than chromosomal genes, functioning essential genes will be found only on chromosomes.

Key words: plasmids, genetic architecture, mobile genetic elements, horizontal gene transfer, antibiotic resistance.

Introduction

Mobile genetic elements play an important role in prokaryote evolution (Bergstrom et al. 2000; Ochman et al. 2000; Jain et al. 2003; Bordenstein and Reznikoff 2005; Sørensen et al. 2005; Thomas and Nielsen 2005; Young et al. 2006; Goldenfeld and Woese 2007; Koonin and Wolf 2008; Tazzyman and Bonhoeffer 2013). A distinction has been drawn between the "core" genome, consisting of genes that are not horizontally transferred, and the "accessory" genome, which contains horizontally transferable genetic elements (Hacker and Carniel 2001; Rankin et al. 2010). There are many mechanisms implicated in horizontal transfer (broadly grouped into conjugation, transduction, and transformation Hacker and Carniel 2001), but conjugative plasmids have been cited as a particularly common and successful vector of transfer (Norman et al. 2009), allowing for transfer of genes across large taxonomic distances.

It has been suggested that there are functional differences between genes that are likely to be horizontally transferred and those that are part of the core genome, both generally (Hacker and Carniel 2001) and specifically in the case of plasmid-biased genes (Eberhard 1990; Rankin et al. 2010). Plasmids replicate independently from the chromosome of their host and can be transferred horizontally and vertically: Consequently, their evolutionary interests may differ from those of the chromosome. These divergent interests can give rise to conflict or to cooperation between plasmid and chromosomal genes (Werren 2011). At the very least, the different evolutionary trajectories and interests of plasmids and chromosomes mean that the average genetic composition of each is likely to differ.

There are several examples of such differences in genetic composition. Naturally, there are a suite of phenotypic traits that plasmids will benefit from possessing, such as the ability to cause conjugation (Thomas 2000, though "small" plasmids often lack this feature). Because there are phages that can exploit the act of conjugation (Anderson 1968), it can be costly for a plasmid's bacterial host, without gaining any obvious benefit (as the donor receives nothing). Consequently, chromosomes are probably less likely than plasmids to bear genes coding for conjugation. However, there are other phenotypic traits that also seem to be coded for by plasmid-biased genes. For reasons that have not been fully explained (though see Nogueira et al. 2009), bacteriocin production (Gonzalez and Kunka 1987), gall formation (Goethals et al. 2001), plant root nodulation (Dowling and Broughton 1986), virulence factors (Evans et al. 1975), and antibiotic resistance genes (Jacob and Hobbs 1974; Tazzyman and Bonhoeffer 2014), all seem to be functions coded for by genes commonly found on plasmids (Eberhard 1990).

One particular element of plasmid genetics that has been observed is the fact that plasmids seldom carry genes that code for structural proteins or basic metabolic functions (Koch 1981; Eberhard 1990), from hereon referred to as "essential genes," However, because of rates of gene flow between bacterial chromosomes and plasmids, and the population sizes of bacterial populations, it seems likely that over evolutionary time, most or all bacterial genes have been at some time located on both plasmids and on chromosomes (Eberhard 1990). Additionally, there are examples of plasmids bearing genes that are essential for the growth of the cell they inhabit (either all the time; Young et al. 2006;

Article

Dziewit et al. 2014; or under specific, regularly encountered environments, e.g., in the dark, Nagarajan et al. 2014, or under anaerobic conditions, Ebert et al. 2013). Thus, because essential genes are in fact sometimes found on plasmids (and must have been found there at times throughout evolutionary history), why are they not found there more often?

The first possibility is that this is a semantic problem: Plasmids are often loosely defined as being replicons lacking in essential genes, and consequently, no essential genes can be found on plasmids. As we are interested in what forces make essential genes less likely to be horizontally transferred, and because (as noted above) there are in fact instances of essential genes on plasmids, we can safely disregard this semantic solution.

Another possibility is that there is only finite space on plasmids, and this space is taken up by the kinds of genes that are favored on plasmids, as noted above. For example, there have been various hypotheses to explain why plasmids would be an ideal locations for antibiotic resistance genes (Eberhard 1990; Rankin et al. 2010; Tazzyman and Bonhoeffer 2014). We feel, however, that this explanation is begging the question somewhat: Even if these other kinds of genes are favored in a plasmid context, we need to show that essential genes are not also favored. Ideally, we would like to establish some sort of advantage accruing to essential genes from a chromosomal location.

A final possibility is given by the "complexity hypothesis." This suggests that the connectivity of a protein in the cellular network, rather than the function of the protein, has a large influence over how easily it can be horizontally transferred (Jain et al. 2003; Cohen et al. 2011). Genes whose products have a high degree of protein–protein interaction will likely disrupt the metabolism of a cell into which they are transferred and consequently are unlikely candidates for horizontal transfer (Baltrus 2013). Essential genes by their very nature have been suggested to have high connectivity (Jeong et al. 2001; Zotenko et al. 2008) and so by the complexity hypothesis would be unlikely to be transferred horizontally. This is a compelling explanation. However, our model below considers a much simpler case, attempting to explain the situation without reference to the complexities of protein–protein interactions.

Understanding what forces bias essential gene location away from plasmids and on to chromosomes will help us to better understand the evolutionary processes shaping plasmid genomes. To investigate this, we here construct a competition model, incorporating essential genes on both plasmids and chromosomes and allowing for some degradation of these genes. We can therefore see under which conditions a chromosomal location is favored for such crucial genes.

The Model

The Essential Gene

Our model features a single, idealized essential gene. Individuals require at least one working copy of this gene to survive. The gene is found both on the chromosome and on plasmids. In either setting, the gene can undergo "degradation." By degradation, we mean any genetic process that makes the gene no longer functional, including, for example, mutation, deletion, and translocation events, as long as they sufficiently disrupt the gene, so that it no longer works. For simplicity, we group all such events under degradation and assume that they occur at a rate μ_c per individual for chromosomal locations and μ_p per plasmid for plasmid locations. Thus, every generation functioning essential genes on chromosomes and on plasmids lose their functionality at these rates.

Six Types of Individuals

Each individual in our population has one of two possible chromosome states: Either "wildtype" (i.e., with a functioning essential gene) or "mutant" (i.e., with a degraded essential gene). We denote these states as w and m, respectively. Wild-type individuals have a functioning essential gene on their chromosome, but mutant individuals do not. Therefore, to avoid death due to lacking the essential gene, a mutant individual must have a functioning essential gene on a plasmid.

Each individual also has one of three possible plasmid states: Either no plasmid (denoted with the "empty set" symbol $\emptyset$), a wild-type plasmid (i.e., with a functioning essential gene, denoted p), or a mutant plasmid (i.e., with a degraded essential gene, denoted q).

There are thus six types of individual in total: Wild-type chromosome with no plasmid (which we denote $w_\emptyset$), wild-type chromosome with wild-type plasmid (w_p), wild-type chromosome with mutant plasmid (w_q), mutant chromosome with no plasmid ($m_\emptyset$), mutant chromosome with wild-type plasmid (m_p), and mutant chromosome with mutant plasmid (m_q). The relative frequency of a type x_y is then denoted $f(x_y)$, and the relative frequency after t generations will be denoted $f_t(x_y)$. To simplify notation, we will denote the equilibrium relative frequency of a type x_y as x_y^* rather than the more cumbersome $f^*(x_y)$.

Fitness Scheme

All fitnesses are defined relative to the fitness of type $w_\emptyset$, which has a wild-type chromosome and no plasmid, and which we define to have fitness 1. There are then two elements to the fitness scheme. First, any individual lacking at least one functioning copy of the essential gene dies and thus has fitness zero. Therefore, the fitnesses of type $m_\emptyset$ and m_q are zero, and no individuals of these types in fact exist in our population.

The second element in our scheme is the presence of the plasmid, which confers a small fitness cost, $0 < c \ll 1$. The relative fitness of plasmid bearing types w_p, w_q, and m_p are then all $1 - c$.

Interactions between the Types

In addition to the relative fitnesses, the frequencies of each type are also determined by interactions between the types. There are three types of interactions: Mutation, conjugation, and segregation. We describe the details below; a diagram

showing all interactions between the four surviving types is shown in figure 1.

Degradation. As mentioned above, degradation occurs on both chromosomes and plasmids, at rates μ_c and μ_p, respectively. Thus, in every generation, a proportion μ_c of individuals with wild-type chromosomes become individuals with mutant chromosomes. In the case of $w_\emptyset$ and w_q individuals, the chromosomal degradation immediately results in death, because it leaves the individual without a functioning housekeeping gene. In the case of a w_p individual, however, the chromosomal degradation results in an m_p individual, which is "rescued" from death by the functioning essential gene on the plasmid. Similarly, in every generation, a proportion μ_p of individuals with wild-type plasmids become individuals with mutant plasmids. In the case of an m_p individual, the plasmid degradation immediately results in death, as it means becoming an m_q individual with no functioning essential gene. However, in the case of a w_p individual, the plasmid degradation results in a w_q individual, which still has a functioning essential gene on its chromosome.

Conjugation. Conjugation is the process through which plasmids are transferred horizontally in the population, from a plasmid bearer to a nonplasmid bearer. We assume
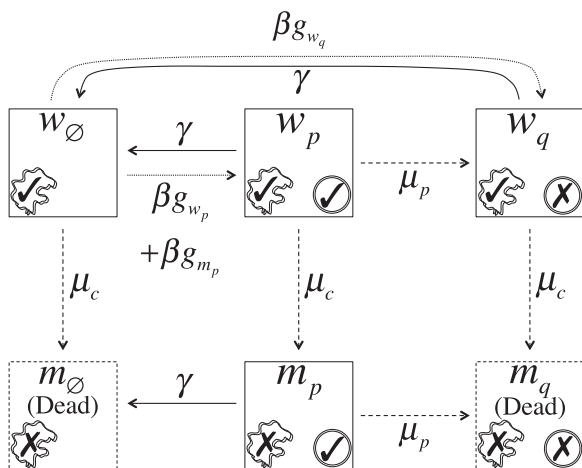

Fig. 1. Diagram of the interactions between the four surviving types. Each box represents one of the six types, with the two nonsurviving types marked with dashed lines and the label "(Dead)." Within each box, a cartoon shows the chromosomal status (squiggly line) and the plasmid status (circles), marked with a tick for functioning essential gene and a cross for degraded essential gene. Each arrow shows a transformation as a result of some interaction type. The labels for the arrows show the proportion of the type which the arrow leaves that are transformed into the type the arrow leads to. Some transformations result in the formation of the nonviable types $m_\emptyset$ or m_q; these transformations are shown here to lead to boxes marked "Dead." The lines marked βg_{w_q} and βg_{w_p} $+\beta g_{m_p}$ are conjugation reactions, where β is the conjugation rate, and the values g_{ij} are as given by equation (1). The lines marked μ_c are degradation events taking place on chromosomes, and the lines marked μ_p are degradation events taking place on plasmids. The lines marked γ are segregation events. For full details of each reaction, see the main text.

that our individuals are in a well-mixed, constant volume and that the process of horizontal transfer can therefore be reasonably well approximated by mass action-type equations (Ross 1915).

Conjugation can occur between any pair of individuals consisting of one plasmid bearer and one nonplasmid bearer. However, because in our system $m_\emptyset$ and m_q individuals immediately die, we can ignore conjugation events featuring these types. Conjugation therefore occurs between $w_\emptyset$ individuals and w_p, w_q, or m_q individuals. It transforms the nonplasmid-bearing $w_\emptyset$ into a plasmid-bearing transconjugant. Thus, conjugation between $w_\emptyset$ and w_p or m_p individuals results in the $w_\emptyset$ individual becoming the transconjugant w_p, whereas conjugation between $w_\emptyset$ and w_q results in the $w_\emptyset$ individual becoming the transconjugant w_q. Note that the chromosomal state is unchanged by conjugation.

Because we use mass action equations, the proportion of $w_\emptyset$ individuals that become transconjugants every generation is proportional to the relative frequencies of $w_\emptyset$ individuals, and of the plasmid bearers, and is controlled by a parameter β. Given frequencies $f(w_\emptyset)$, $f(w_p)$, $f(w_q)$, and $f(m_p)$ at a given generation, we then have $\beta f(w_\emptyset)(f(w_p)+f(m_p))$ transconjugant w_p individuals and $\beta f(w_\emptyset)f(w_q)$ transconjugant w_q individuals created, in that generation.

Segregation. Segregation is the process by which plasmid-bearing individuals lose their plasmids, usually during replication. It is generally believed to be rare because of a variety of phenotypic adaptations that plasmids have evolved to prevent it from occurring (e.g., postsegregational killing [Cooper and Heinemann 2000]; or methods to ensure even splitting of plasmids during cell division [Ebersbach and Gerdes 2005]). We model it by assuming that some small proportion γ of plasmid bearers lose their plasmids per generation. Thus, w_p and w_q individuals become $w_\emptyset$ individuals, whereas m_p individuals die because they become $m_\emptyset$ individuals and thus lack a functioning essential gene.

The Model

We use the above elements to construct the model as follows. We census the population at every generation. We concern ourselves with relative frequencies and assume that the total population size is a (very large) constant. Therefore at any generation, the frequencies of each type add up to one: $f(w_\emptyset)+f(w_p)+f(w_q)+f(m_p)=1$. Given frequencies $f_t(w_\emptyset)$, $f_t(w_p)$, $f_t(w_q)$, and $f_t(m_p)$ after t generations, we wish to know the frequencies after $t+1$ generations.

To simplify the mathematics, we first take fitness into account and then the interactions. The fitness of type $w_\emptyset$ is 1, whereas the fitness of types w_p, w_q, and m_p is $1-c$.

$$g_{w_\emptyset} = f_t(w_\emptyset)$$
$$g_{w_p} = (1-c)f_t(w_p)$$
$$g_{w_q} = (1-c)f_t(w_q)$$
$$g_{m_p} = (1-c)f_t(m_p)$$

(1)

Then, the frequencies of each genotype in the next generation are

$$\overline{\omega}f_{t+1}(w_\emptyset) = g_{w_\emptyset}(1 - \mu_c) - \beta g_{w_\emptyset}\big(g_{w_p} + g_{w_q} + g_{m_p}\big)$$
$$+ \gamma\big(g_{w_p} + g_{m_p}\big),$$
$$\overline{\omega}f_{t+1}(w_p) = g_{w_p}\big(1 - \mu_c - \mu_p - \gamma\big) + \beta g_{w_\emptyset}\big(g_{w_p} + g_{m_p}\big),$$
$$\overline{\omega}f_{t+1}(w_q) = g_{w_q}\big(1 - \mu_c - \gamma\big) + \beta g_{w_\emptyset}g_{w_q} + \mu_p g_{w_p},$$
$$\overline{\omega}f_{t+1}(m_p) = g_{m_p}\big(1 - \mu_p - \gamma\big) + \mu_c g_{w_p}.$$

(2)

where $\overline{\omega}$ is the average fitness for generation t, defined as

$$\overline{\omega} = f_t(w_\emptyset)(1 - \mu_c) + (1 - c)f_t\big(w_p\big)$$
$$+ (1 - c)f_t\big(w_q\big)(1 - \mu_c) \qquad (3)$$
$$+ (1 - c)f_t\big(m_p\big)\big(1 - \gamma - \mu_p\big),$$

so that the new genotype frequencies add to one. We can combine the sets of equations (1) and (2) with equation (3) to find equilibrium frequencies $w_\emptyset^*$, w_p^*, w_q^*, and m_p^*.

Results

There are four different equilibria possible. Inspection of figure 1 reveals that equilibria are possible in which there are only $w_\emptyset$ types, in which there are only m_p types, and in which there are only $w_\emptyset$ and w_q types. There is also an equilibrium in which all four types are present. We now go through the stability of these equilibria in turn

Equilibrium with Only $w_\emptyset$ Individuals

This equilibrium point corresponds to the case where the plasmid is absent from the population. Under these circumstances, because only chromosomal essential genes exist, all degradations to these genes result in immediate death. The criterion for plasmids to be able to invade the population (and thus for this equilibrium point to be unstable) is

$$c < c^* = \frac{\beta - \gamma}{1 + \beta - \gamma - \mu_c}, \qquad (4)$$

so there is an upper limit c^* on the cost that a plasmid can impose on its host and still be able to invade, depending upon the conjugation rate β, the segregation rate γ, and the degradation rate for chromosomes μ_c. As we are here interested in whether essential genes can be maintained on plasmids, we assume from now on that $c < c^*$ (though note that the reasons behind the maintenance of plasmids over evolutionary timescales in the face of low conjugation rates and nonzero costs have been debated; Stewart and Levin 1977; Bergstrom et al. 2000; Lili et al. 2007).

Equilibrium with Only m_p Individuals

This equilibrium point corresponds to the case where all functioning essential genes are plasmid borne. For reasonable parameter values, it is always unstable (see Appendix).

Equilibrium with Only $w_\emptyset$ and w_q Individuals

This equilibrium point corresponds to the case where all the essential genes on plasmids have degraded, whereas all those on chromosomes are still functional. The equilibrium frequencies are

$$w_\emptyset^* = \frac{(1 - c)\gamma}{(1 - c)\beta - c(1 - \mu_c)}, \qquad (5)$$

and $w_q^* = 1 - w_\emptyset^*$. This solution is stable if $\mu_c \leq \mu_p$, that is, if plasmid genes degrade more rapidly than chromosomal genes, and unstable in the contrary case where $\mu_c > \mu_p$, where plasmid genes degrade more slowly than chromosomal genes (Appendix).

Equilibrium with All Four Types

The exact location of the equilibrium is a lengthy expression for each genotype frequency. We have reproduced it in the Appendix, but here we instead make the substitution $\mu_p = \alpha\mu_c$, and note that μ_c is likely to be much smaller than the other expressions, and so we can let it tend to zero and see that the global equilibrium is approximately

$$w_\emptyset^* \approx \frac{(1 - c)\gamma}{(1 - c)\beta - c},$$

$$w_p^* \approx (1 - \alpha)\frac{\beta - \gamma + c(1 + \beta - \gamma)}{(1 - c)\beta - c},$$

$$\qquad\qquad (6)$$

$$w_q^* \approx \alpha\frac{\beta - \gamma + c(1 + \beta - \gamma)}{(1 - c)\beta - c},$$

$$m_p^* \approx 0.$$

(for some parameter combinations this approximation does not hold, but it is likely to be generally true for reasonable parameter values—see the Appendix for details). We can see from this that it will only be a permitted solution if $\alpha < 1$, that is, if $\mu_c > \mu_p$, since otherwise $w_p^* < 0$ (we know that the numerator in the fraction found in the expressions for w_p^* and w_q^* is positive because we assume $c > c^*$). Thus, this equilibrium is stable for values $\mu_c > \mu_p$ (Appendix).

Growth Ratios

We further illustrate the system by calculating growth ratios. The growth ratio ρ_p for the p plasmid (containing a functioning housekeeping gene) is $\rho_p = \big(f_{t+1}(w_p) + f_{t+1}(m_p)\big)/\big(f_t(w_p) + f_t(m_p)\big)$, representing the frequency of the plasmid in generation $t + 1$ compared with its frequency in generation t. If $\rho_p > 1$, then the p plasmid is growing in frequency, if $\rho_p < 1$, it is declining in frequency, and if $\rho_p = 1$, its frequency is unchanging. Similar calculations can be made for the q plasmid, as well as for the chromosomes w and m.

To make analysis easier, we multiply all the growth ratios by the average fitness in generation t, $\overline{\omega}$.

$$\overline{\omega}\rho_p = (1-c)(1 + \beta f_t(w_\emptyset) - \gamma - \mu_p)$$

$$\overline{\omega}\rho_q = (1-c)\left(1 + \beta f_t(w_\emptyset) - \gamma - \mu_c + \mu_p \frac{f_t(w_p)}{f_t(w_q)}\right),$$

$$\overline{\omega}\rho_w = (1-c)(1 - \mu_c) + \frac{c(1-\mu_c)f_t(w_\emptyset)}{f_t(w_\emptyset) + f_t(w_p) + f_t(w_q)}$$

$$\overline{\omega}\rho_m = (1-c)\left(1 - \gamma - \mu_p + \mu_c \frac{f_t(w_p)}{f_t(m_p)}\right).$$

We can now see more clearly why it is that the relative sizes of μ_c and μ_p have such an effect on the competition between the p and q plasmids. If $\mu_p > \mu_c$, then $\rho_q > \rho_p$, and so the p plasmid will always be growing faster than the q plasmid. Thus, either plasmids go extinct entirely (if c is large enough relative to the other parameters, so that $\rho_q < 1$), or the q plasmid will fix. If, on the other hand, $\mu_c \geq \mu_p$, the values $f_t(w_p)$ and $f_t(w_q)$ will determine which is the larger out of ρ_p and ρ_q, with the two growth ratios being equal at the equilibrium with all four types present given above.

We can rewrite the expression containing ρ_w in terms of the expression containing ρ_m to give

$$\overline{\omega}\rho_w = \begin{cases} \overline{\omega}\rho_m \\ + \\ \gamma(1-c) + (\mu_p - \mu_c)(1-c) \\ + \\ c(1-\mu_c)\frac{f_t(w_\emptyset)}{1 - f_t(m_p)} - \mu_c(1-c)\frac{f_t(w_p)}{f_t(m_p)} \end{cases} \tag{7}$$

From this, we see that a higher degradation rate γ improves the fitness of the w chromosome over the m chromosome, as does a greater positive difference $\mu_p - \mu_c$. The effects of parameters c and μ_c will depend on the genotype frequencies in quite a complicated way.

Overview of Results

Our results are summarized in figure 2. There are two key elements, the cost c of bearing a plasmid, and the ratio α of the degradation rate of plasmids to that of chromosomes.

If $c > c^*$, then plasmids are too costly to persist in the population and will go extinct. The only genotype in the population is then $w_\emptyset$, and consequently, all chromosomal genes will bear functioning essential genes. If $c < c^*$, then plasmids can persist in the population. Then, the equilibrium will depend on α.

If $\alpha \geq 1$, plasmid genes degrade more rapidly than chromosomal genes. At equilibrium, there will be only $w_\emptyset$ and w_q types. This means that all chromosomes will bear functioning essential genes, whereas no plasmids will. If $\alpha < 1$, chromosomal genes degrade at a higher rate than plasmid genes. Then, all four types will be present in the population at equilibrium, although type m_p will be present at a vanishingly
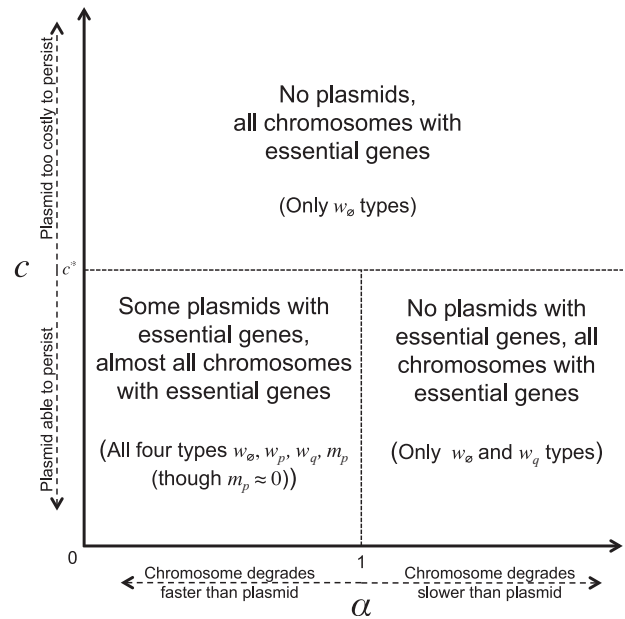

Fig. 2. Qualitative results of the model as a result of two parameters, α, ratio of the degradation rate of plasmids to that of chromosomes, and c, the cost of bearing the plasmid. For $c > c^*$ (eq. 4), no plasmids exist at equilibrium, and the equilibrium point is $w_\emptyset^* = 1$, with all other frequencies zero. For $c < c^*$, the results depend on $\alpha = \mu_p / \mu_c$. For $\alpha \geq 1$, there are only $w_\emptyset$ and w_q types at equilibrium, and their frequencies are given by equation (5). For $\alpha < 1$, all four types are present at equilibrium, and their frequencies are given by equation (6).

small frequency, $m_p^* \approx 0$. A fraction $1 - m_p^* \approx 1$ of chromosomes will bear functioning essential genes, whereas a fraction $1 - \alpha$ of plasmids will bear functioning essential genes.

In particular, note that if $\mu_c = \mu_p$, we will have only $w_\emptyset$ and w_q types, and even if $\mu_c \approx \mu_p$ (i.e., $\alpha \approx 1$), we will be very close to this equilibrium. Thus, if the degradation rates on plasmids and chromosomes are similar, there will be very few essential genes on plasmids.

Discussion

Our results theoretically validate the empirical observation that essential genes are mainly found on chromosomes, rather than on plasmids. At all possible equilibrium states, all (or nearly all) chromosomes in the population will bear functioning essential genes. This will not necessarily be the case for plasmids. If the degradation rate on plasmids is equal to or greater than the degradation rate for chromosomes (i.e., $\alpha > 1$), then no plasmids will have functioning housekeeping genes. Alternatively, if the degradation rate on plasmids is less than the degradation rate on chromosomes, there will be a mix of plasmids in the population, some which have functioning housekeeping genes and some which do not. The proportion of plasmids with degraded, nonfunctioning essential genes will be equal to the ratio of the degradation rate on chromosomes with that on plasmids.

The intuition behind these results is that the key element is the stable inheritance of essential genes. Offspring lacking essential genes die; consequently, the fitness of the different genotypes in our model is determined to a large degree by
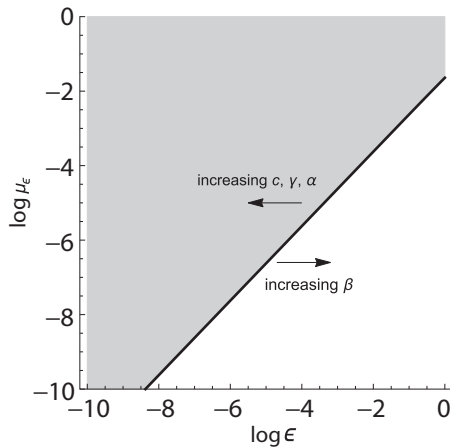
FIG. 3. Threshold value μ_ϵ as a function of ϵ. To have $m_p^* > \epsilon$, we require $\mu_c > \mu_\epsilon$ (i.e., to be in the gray area). Increasing parameter values generally affects the threshold as shown by the arrows above. Here, the relationship is plotted on a log–log scale, and the parameter values are $\beta = 0.4$, $\alpha = 0.5$, $\gamma = 0.01$, $c = 0.05$.

FIG. 4. Log values of μ_c required to give $m_p^* > 10^{-2}$ for given values of β, γ, where $\alpha = 0$ and $c = 0$. For example, for values of β and γ along the line marked "-4," we would require $\mu_c > 10^{-4}$ in order for $m_p^* > 10^{-2}$. Looking at the plot, one can see that mutation rates below $\mu_c = 10^{-5}$ will require extremely small values of γ to lead to appreciable quantities of m_p individuals at equilibrium, even where $c = 0$ and $\alpha = 0$.

how likely their offspring are to stably inherit the essential genes. The m chromosome relies on being paired with a p plasmid to make up for its own lacking functioning essential gene. This plasmid can fail to be inherited due to segregation or due to mutation (rates γ and μ_p, respectively). On the other hand, the w chromosome will fail to pass a functioning essential gene to its offspring only in the case where it mutates (rate μ_c) and is not paired with a p plasmid. Thus, higher values of γ and μ_p benefit the w chromosome, whereas higher values of μ_c benefit the m chromosome. The w chromosome also benefits from the extra fitness of the $w_\emptyset$ types. Although the exact relative fitnesses of each chromosome depend upon the frequencies of the four types (eq. 7), in general the w chromosome is usually fitter than the m chromosome and rises to fixation (or near fixation) in all cases. The fact that the w chromosome is prevalent in turn affects the competition between the p and q plasmids. The competition between them is largely determined by the rates of mutation. The q plasmid will lose out when the w chromosome with which it is associated mutates (at a rate μ_c), because the resulting m_q individual lacks a functioning essential gene and dies. The same chromosomal mutation does not immediately damage the p plasmid, as the m_p individual created survives. So high rates of μ_c favor the p plasmid over the q plasmid by reducing the fitness of the latter. However, when the p plasmid mutates (at rate μ_p), it becomes a q plasmid, meaning that higher values of μ_p doubly favor the q plasmid over the p plasmid by increasing the fitness of the former and reducing that of the latter. Consequently, the result of the competition between the two is largely determined by the relationship between μ_c and μ_p, with the q plasmid having an inherent advantage, as described above.

Note that in reality, μ_p and μ_c are not necessarily different. In fact, we are unaware of any a priori reason why there should be a large difference between the two. Perhaps, they do generally differ for some reason, but our work is not
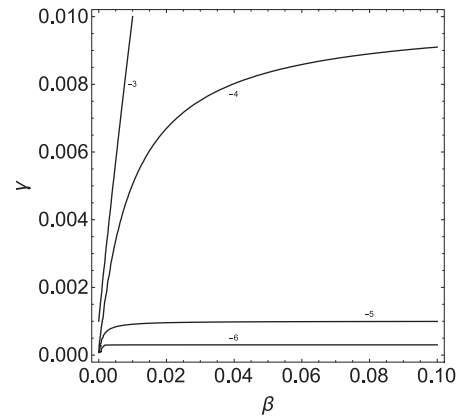
contingent on this. We simply aim to show the consequences of these rates, rather than claiming anything specific about how they differ in reality. Our results show that if the two values are similar, then there will be almost no plasmids bearing functioning housekeeping genes (assuming plasmids are stably maintained in the population). The only way that plasmids will bear functioning housekeeping genes in our model is if they degrade less rapidly than chromosomes and are thus able to make up for their lack of stable inheritance.

In addition to the effect of the relationship between μ_p and μ_c, there are also effects of the parameters β, γ, and c. However, supposing that μ_c and μ_p are such that $m_p^* \approx 0$ at equilibrium (i.e., the two degradation parameters are sufficiently small and/or $\mu_p > \mu_c$, see Appendix), we will be approximately close to the equilibrium positions given by equation (6), and these other parameters largely affect only the frequency of plasmid-bearing individuals in the population, rather than the outcome of the competition between the two types of plasmid.

There are some potential caveats to our conclusions. We have treated plasmids here as a single entity, with a binary status of either "functioning" or "degraded." Thus, our model implicitly treats the situation in which there is only a single plasmid location in each plasmid-bearing cell. In reality, however, plasmids generally exist at multiple copy numbers, and there is the possibility that some of the plasmids within a cell will carry degraded copies of the essential gene, whereas others will carry functioning copies. The inheritance, both vertically and by conjugation, of such a multiple copy number system would require a model of greatly increased complexity. We expect, however, that our qualitative conclusions, that chromosomal locations are favored for essential genes, and that the frequency of these genes on plasmids strongly depends on the degradation ratio α, will still hold.

Our work can be compared with the complexity hypothesis, which would also predict a lack of essential genes on plasmids but for different reasons; namely, that the high connectivity of the essential proteins means that they are not good candidates for horizontal transfer as they cause too much disruption (Jain et al. 2003; Cohen et al. 2011; Baltrus 2013). We could capture this disruption in our framework by introducing an additional cost k that represents the problem caused by having two copies of the essential gene, so that the fitness of type w_p is $1 - c - k$, whereas the fitnesses of types m_p and w_q are both $1 - c$. Unfortunately, this additional parameter causes the complexity of the model to increase vastly, and so for space reasons, we have elected not to follow this line of research here but rather to leave it for a later work. The complexity hypothesis is not mutually exclusive with our hypothesis that the fidelity of inheritance of the genes is what is important; both could be contributory factors.

Our findings are particularly interesting because they show that, even at equilibrium, it is possible that essential genes could be found on plasmids. This is the case if the degradation rate of these genes is higher on the chromosome than it is on a plasmid (fig. 2). This may help to explain the discovery of plasmid-borne essential genes (e.g., Young et al. 2006; Ebert et al. 2013; Dziewit et al. 2014; Nagarajan et al. 2014); alternatively, such genes may be an example of a population that is yet to reach equilibrium. It might be that in the long term, these genes will migrate to the chromosome.

On a related note, one kind of mechanism that is essential under certain environmental conditions and is often encoded by genes found on plasmids is antibiotic resistance (Jacob and Hobbs 1974; Eberhard 1990), for reasons that are imperfectly understood (Tazzyman and Bonhoeffer 2014). Our findings here support the assertion that, over long timescales, antibiotic resistance genes are likely to move to chromosomal locations (Bergstrom et al. 2000), indirectly supporting the thesis that short-term effects are responsible for the presence of plasmid-borne antibiotic resistance (Bergstrom et al. 2000; Tazzyman and Bonhoeffer 2014).

In conclusion, we have shown that chromosomes are indeed favored locations for functioning essential genes, due to their stable inheritance. Whether plasmids will also bear such genes, however, will depend on the rate at which these genes degrade by chance events in each location. If plasmids very rarely bear functioning essential genes, it is perhaps an indication that the rate of degradation in plasmid locations is greater than that on the chromosome.

Acknowledgments

Appendix

Stability of Equilibria

We first note that, since the frequencies $f(w_\emptyset) + f(w_p) + f(w_q) + f(m_p) = 1$, there are only three independent variables, so we can use the substitution $m_p^* = w_\emptyset^* + w_p^* + w_q^*$. For a given type x_y, we define $J_{x_y} = f_{t+1}(x_y) - f_t(x_y)$, the change in frequency per generation. Then, we define the Jacobian matrix $\mathbf{J}$,

$$\mathbf{J} = \begin{pmatrix} \dfrac{\partial J_{w_\emptyset}}{\partial w_\emptyset} & \dfrac{\partial J_{w_\emptyset}}{\partial w_p} & \dfrac{\partial J_{w_\emptyset}}{\partial w_q} \\[2ex] \dfrac{\partial J_{w_p}}{\partial w_\emptyset} & \dfrac{\partial J_{w_p}}{\partial w_p} & \dfrac{\partial J_{w_p}}{\partial w_q} \\[2ex] \dfrac{\partial J_{w_q}}{\partial w_\emptyset} & \dfrac{\partial J_{w_q}}{\partial w_p} & \dfrac{\partial J_{w_q}}{\partial w_q} \end{pmatrix}.$$

The stability of an equilibrium point $(w_\emptyset^*, w_p^*, w_q^*)$ is then determined by the three eigenvalues of $\mathbf{J}$, denoted λ_1, λ_2, and λ_3, evaluated at this point, with the equilibrium being stable if all eigenvalues have negative real part.

$$w_\emptyset^* = 1$$

Evaluated at the equilibrium point for which there are just $w_\emptyset$ individuals, the matrix $\mathbf{J}$ has eigenvalues

$$\lambda_1 = \frac{\beta - \gamma - c(1 + \beta - \gamma - \mu_c)}{1 - \mu_c},$$

$$\lambda_2 = \frac{\beta - \gamma - \mu_p - c(1 + \beta - \gamma - \mu_c - \mu_p)}{1 - \mu_c},$$

$$\lambda_3 = -1,$$

and so the equilibrium will be unstable where $\beta - \gamma - c(1 + \beta - \gamma - \mu_c) > 0$. This gives the condition on c given in the text above.

$$m_p^* = 1$$

This equilibrium point corresponds to the case where all functioning essential genes are plasmid borne. Our eigenvalues are then

$$\lambda_1 = -\frac{\mu_c}{1 - \gamma - \mu_p},$$

$$\lambda_2 = \frac{-\left(\psi + (1-c)(\mu_c - \mu_p)\right) - \sqrt{\left(\psi + (1-c)(\mu_c - \mu_p)\right)^2 + \eta}}{2(1-c)(1 - \gamma - \mu_p)},$$

$$\lambda_3 = \frac{-\left(\psi + (1-c)(\mu_c - \mu_p)\right) + \sqrt{\left(\psi + (1-c)(\mu_c - \mu_p)\right)^2 + \eta}}{2(1-c)(1 - \gamma - \mu_p)},$$

where

$$\psi = \beta - \gamma + \mu_c - \mu_p - c(1 + \beta - \gamma - \mu_p),$$

$$\eta = 4(1-c)\left((1-c)\beta\gamma - (\mu_c - \mu_p)\psi\right).$$

We can have stability only where $\psi + (1-c)(\mu_c - \mu_p) \geq 0$ and $\eta < 0$, since otherwise $\lambda_3 > 0$. $\eta > 0$ implies that

$$\mu_c - \mu_p > \frac{(1-c)\beta\gamma}{\psi} > 0, \qquad (8)$$

and since $\psi > 0$, if $\mu_c - \mu_p > 0$, we have that this equilibrium is stable if and only if criterion (8) is true. As ψ depends on μ_c and μ_p, this is not a closed form solution, but we make the substitution $\mu_p = \alpha\mu_c$, we can show that equation (8) will be true where

$$\mu_c(1-\alpha) - \frac{(1-c)\beta\gamma}{\beta - \gamma + \mu_c - c(1+\beta-\gamma) - \alpha(1-c)\mu_c} > 0.$$

This expression gives us a quadratic equation to solve for α, which shows that there is no solution in the region $0 < \alpha < 1$ using our parameter ranges. So this solution is never stable.

Just $w_\emptyset$ and w_q Types

This equilibrium point corresponds to the case where all essential genes on plasmids are degraded, whereas all those on chromosomes are still functional. The equilibrium frequencies are $w_q^* = 1 - w_\emptyset^*$ and

$$w_\emptyset^* = \frac{(1-c)\gamma}{(1-c)\beta - c(1-\mu_c)}.$$

The corresponding eigenvalues of $\mathbf{J}$ evaluated at this point are

$$\lambda_1 = -\frac{(\beta - c(1+\beta-\mu_c))(\beta - \gamma - c(1+\beta-\gamma-\mu_c))}{(1-c)(1-\mu_c)(\beta - c(1+\beta-\gamma-\mu_c))},$$

$$\lambda_2 = \frac{(\beta - c(1+\beta-\mu_c))(\mu_c - \mu_p)}{(1-\mu_c)(\beta - c(1+\beta-\gamma-\mu_c))},$$

$$\lambda_3 = \frac{c(1-\mu_c)\mu_p - (1-c)\beta(\gamma+\mu_p)}{(1-\mu_c)(\beta - c(1+\beta-\gamma-\mu_c))},$$

As $c < c^*$, we know that $\beta - \gamma - c(1+\beta-\gamma-\mu_c) > 0$, and so the denominator in all three cases is greater than 0. It follows that also $\beta - c(1+\beta-\mu_c) > \gamma(1-\mu_c)/(1+\beta-\gamma-\mu_c) > 0$, so $\lambda_1 < 0$, and λ_2 has the same sign as $\mu_c - \mu_p$. Finally, after some rearrangement, we have that $\lambda_3 < 0$ if $c < \hat{c}$, where

$$\hat{c} = \frac{\beta(\gamma+\mu_p)}{(1-\mu_c)\mu_p + \beta(\gamma+\mu_p)},$$

but

$$\hat{c} - c^* = \frac{\gamma(1-\mu_c)(\beta+\mu_p)}{(1+\beta-\gamma-\mu_c)((1-\mu_c)\mu_p + \beta(\gamma+\mu_p))} > 0,$$

so $\hat{c} > c^* > c$, and $\lambda_3 < 0$.

Thus, for our region of parameter space, $\lambda_1 < 0$ and $\lambda_3 < 0$, and so the equilibrium is unstable if and only if

$\lambda_2 > 0$, which is true if and only if $\mu_c > \mu_p$, that is, if chromosomes degrade at a faster rate than plasmids.

All Four Types

The equilibrium point is at

$$w_\emptyset^* = \frac{(\beta + \mu_c - \mu_p - c(1+\beta-\mu_p))(\gamma + \mu_p - \mu_c) + c\gamma(1-\mu_c)}{\beta(\beta + \mu_c - \mu_p - c(1+\beta-\mu_p))},$$

$$w_p^* = (\mu_c - \mu_p)\frac{((1-c)\beta\gamma - (\mu_c - \mu_p)\psi)\psi}{(1-c)\beta\gamma\mu_c(\beta + \mu_c - \mu_p - c(1+\beta-\mu_p))},$$

$$w_q^* = \mu_p\frac{((1-c)\beta\gamma - (\mu_c - \mu_p)\psi)\psi}{(1-c)\beta\gamma\mu_c(\beta + \mu_c - \mu_p - c(1+\beta-\mu_p))},$$

$$m_p^* = \frac{(\mu_c - \mu_p)\psi}{(1-c)\beta\gamma},$$

where

$$\psi = \beta - \gamma + \mu_c - \mu_p - c(1+\beta-\gamma-\mu_p).$$

As $w_q^* = w_p^*\mu_p/(\mu_c - \mu_p)$, the two will be of opposite sign if $\mu_p > \mu_c$, and consequently, this equilibrium will be outside the region we are interested in. Thus, this solution, in which all four types feature, will only occur if $\mu_c > \mu_p$: Exactly the region for which the above solution with only types $w_\emptyset$ and w_q is unstable. In addition, as $\mu_c \to \mu_p$ from above, the solution with all four types tends to exactly the solution with only $w_\emptyset$ and w_q given above. It remains to show that this solution with four types is stable where $\mu_c > \mu_p$. Unfortunately, the eigenvalues do not admit a simple analytical closed form solution, so we instead consider the characteristic equation of the matrix $\mathbf{J}$,

$$x^3 - \text{Tr}[\mathbf{J}]x^2 + (|\mathbf{J}_{11}| + |\mathbf{J}_{22}| + |\mathbf{J}_{33}|)x - |\mathbf{J}| = 0,$$

where $\text{Tr}[\cdot]$ is the trace of a matrix, $|\cdot|$ is the determinant of a matrix, and $\mathbf{J}_{ii}$ is the two-by-two matrix formed from $\mathbf{J}$ by deleting the ith row and column. Then, from the Routh–Hurwitz stability criteria, the solution is stable if, when $\mathbf{J}$ is evaluated at the appropriate fixed point,

$$|\mathbf{J}| < 0,$$

$$\text{Tr}[\mathbf{J}] < 0,$$

$$|\mathbf{J}_{11}| + |\mathbf{J}_{22}| + |\mathbf{J}_{33}| > 0,$$

and

$$-\text{Tr}[\mathbf{J}](|\mathbf{J}_{11}| + |\mathbf{J}_{22}| + |\mathbf{J}_{33}|) > -|\mathbf{J}|$$

Define

$$c_\psi = \frac{\beta - \gamma + \mu_c - \mu_p}{1 + \beta - \gamma - \mu_p}.$$

Then if $c < c_\psi$, $\psi > 0$. But we know $c < c^*$ and

$$c_\psi - c^* = \frac{(1-\mu_c)(\mu_c - \mu_p)}{(1+\beta-\gamma-\mu_c)(1+\beta-\gamma-\mu_p)},$$

so if $\mu_c > \mu_p$, we have $c_\psi > c^* > c$ and so $\psi > 0$. After some calculation, it can be shown that

$$|\mathbf{J}| = \frac{((\mu_c - \mu_p)\psi - (1-c)\beta\gamma) \times (\mu_c - \mu_p)\psi(\psi + (1-c)\gamma)^2((1-c)(\beta\gamma + \gamma\mu_c) + \psi\mu_p)}{(1-c)^2\beta\gamma(1-\mu_c)^3(\psi+\gamma)^3}.$$

If $\mu_c > \mu_p$, the fractional part of this expression has positive denominator and numerator. Thus, $|\mathbf{J}| < 0$ if $\mu_c - \mu_p < (1-c)\beta\gamma/\psi$, which we showed above must be the case for our parameter range.

We can also show that

$$\mathrm{Tr}[\mathbf{J}] = -\left(\frac{(\psi + \gamma + (1-c)\mu_c)(\psi + \gamma)}{(1-c)(1-\mu_c)(\psi+\gamma)} + \frac{c\gamma(1 - 2\beta - 4\mu_c + 2\mu_p + c(1 + 2\beta - \gamma + 2(\mu_c - \mu_p)))}{(1-c)(1-\mu_c)(\psi+\gamma)} \right).$$

As $\psi > 0$, the denominator of both expressions in brackets on the right-hand side is positive, and the first expression is also positive. Then, because we are assuming that β, c, μ_p, and μ_c are all considerably less than 1, the second expression is also positive. So $\mathrm{Tr}[\mathbf{J}] < 0$.

The expression for $|\mathbf{J}_{11}| + |\mathbf{J}_{22}| + |\mathbf{J}_{33}|$ is considerably less tractable. We were able to evaluate it numerically for a wide range of parameter values (c varying between 0.01 and 0.05, β varying between 0.01 and 0.1, γ varying between 10^{-4} and 10^{-3}, μ_c varying between 10^{-6} and 10^{-5}, and $\alpha = \mu_p/\mu_c$ varying between e^{-5} and e^5). For this region of parameter space, we found that as long as $\alpha < 1$ (i.e., $\mu_c > \mu_p$), $|\mathbf{J}_{11}| + |\mathbf{J}_{22}| + |\mathbf{J}_{33}| > 0$ as required.

We investigated the difference $-\mathrm{Tr}[\mathbf{J}](|\mathbf{J}_{11}| + |\mathbf{J}_{22}| + |\mathbf{J}_{33}|) + |\mathbf{J}|$ similarly and found that it was always positive, as required. The Mathematica (Wolfram Research, Inc. 2010) file in which these calculations were performed is available for download on Dryad at (insert URL).

Size of m_p^* at the Equilibrium with All Four Types

Making the substitution $\mu_p = \alpha\mu_c$, we have from above that at the equilibrium with all four types,

$$m_p^* = \frac{\mu_c(1-\alpha)(\beta - \gamma + \mu_c(1-\alpha) - c(1 + \beta - \gamma - \alpha\mu_c))}{(1-c)\beta\gamma}, \tag{9}$$

where $0 < \alpha < 1$ since otherwise the four type equilibrium is unstable, and we revert to the two-type equilibrium. We claim that with reasonable parameter values, this equilibrium value $m_p^* \approx 0$.

To support our claim, note that we can rearrange equation (9) to derive a value μ_ϵ,

$$\mu_\epsilon = \frac{1}{2}\left(\frac{c - (\beta - \gamma)(1-c)}{1 - (1-c)\alpha} + \frac{1}{\sqrt{\frac{(c - (\beta - \gamma)(1-c))^2}{(1 - (1-c)\alpha)^2} + \frac{4(1-c)\beta\gamma}{(1-\alpha)(1 - (1-c)\alpha)}\epsilon}}} \right),$$

such that for any $0 < \epsilon < 1$, $m_p^* > \epsilon$ if and only if $\mu_c > \mu_\epsilon$. The expression for μ_ϵ is complicated, but we can investigate it numerically, plotting μ_ϵ as a function of ε to see the required mutation rate to sustain a proportion ε of m_p types at equilibrium (see fig. 3).

This investigation was carried out using Mathematica (Wolfram Research, Inc. 2010), and the files are stored on Dryad at http://dx.doi.org/10.5061/dryad.m7t63, allowing interested readers to investigate the parameter space themselves.

To see that $m_p^* \approx 0$ for reasonable parameter values is to investigate the case where $c = 0$ and $\alpha = 0$, a best-case scenario in terms of increasing the size of m_p^*. This corresponds to the case where plasmid-borne essential genes do not degrade and plasmids cost nothing. Under these circumstances, we can determine the required value of μ_c to give $m_p^* = 10^{-2}$. As can be seen from figure 4, even under these circumstances, which are very favorable to high m_p^* values at equilibrium, we would require either (relatively) high values of μ_c or (relatively) low values of γ to get $m_p^* > 10^{-2}$.

Finally, we numerically calculated values of m_p^* across the parameter space that we used to numerically investigate the stability of the four-type equilibrium above and found that across this range of parameter space, $m_p^* < 0.09$, and that even this relatively high value is obtained at the edge of the space with a combination of parameters $\mu_c = 10^{-5}$, $c = 0.01$, $\beta = 0.1$, $\gamma = 10^{-4}$, $\alpha = 0$, and $\mu_c = 10^{-5}$.

In other words, in order for m_p types to be frequent at equilibrium, we would need an unrealistic combination of low degradation rates on plasmids, high degradation rates on chromosomes, and/or low segregation rates. Under these circumstances, plasmids essentially become back-up chromosomes, as they are inherited almost as well as chromosomes. If they also degrade at a much slower rate, then we can see appreciable quantities of m_p types. These circumstances are somewhat peculiar, however, so that we feel justified in the assertion in the main text that $m_p^* \approx 0$ for reasonable parameter values.

References

Anderson E. 1968. The ecology of transferable drug resistance in the enterobacteria. *Annu Rev Microbiol.* 22(1):131–180.

Baltrus DA. 2013. Exploring the costs of horizontal gene transfer. *Trends Ecol Evol.* 28(8):489–495.

Bergstrom C, Lipsitch M, Levin B. 2000. Natural selection, infectious transfer and the existence conditions for bacterial plasmids. *Genetics* 155(4):1505–1519.

Bordenstein SR, Reznikoff WS. 2005. Mobile DNA in obligate intracellular bacteria. *Nat Rev Microbiol.* 3(9):688–699.

Cohen O, Gophna U, Pupko T. 2011. The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. *Mol Biol Evol.* 28(4):1481–1489.

Cooper TF, Heinemann JA. 2000. Postsegregational killing does not increase plasmid stability but acts to mediate the exclusion of competing plasmids. *Proc Natl Acad Sci U S A.* 97(23): 12643–12648.

Dowling D, Broughton W. 1986. Competition for nodulation of legumes. *Annu Rev Microbiol.* 40(1):131–157.

Dziewit L, Czarnecki J, Wibberg D, Radlinska M, Mrozek P, Szymczak M, Schlüter A, Pühler A, Bartosik D. 2014. Architecture and functions of a multipartite genome of the methylotrophic bacterium *Paracoccus aminophilus* jcm 7686, containing primary and secondary chromids. *BMC Genomics* 15(1):124.

Eberhard W. 1990. Evolution in bacterial plasmids and levels of selection. *Q Rev Biol.* 65(1):3–22.

Ebersbach G, Gerdes K. 2005. Plasmid segregation mechanisms. *Annu Rev Genet.* 39:453–479.

Ebert M, Laaß S, Burghartz M, Petersen J, Koßmehl S, Wöhlbrand L, Rabus R, Wittmann C, Tielen P, Jahn D. 2013. Transposon mutagenesis identified chromosomal and plasmid genes essential for adaptation of the marine bacterium *Dinoroseobacter shibae* to anaerobic conditions. *J Bacteriol.* 195(20):4769–4777.

Evans DG, Silver R, Evans D, Chase D, Gorbach S. 1975. Plasmid-controlled colonization factor associated with virulence in *Escherichia coli* enterotoxigenic for humans. *Infect Immun.* 12(3): 656–667.

Goethals K, Vereecke D, Jaziri M, Van Montagu M, Holsters M. 2001. Leafy gall formation by *Rhodococcus fascians*. *Annu Rev Phytopathol.* 39(1):27–52.

Goldenfeld N, Woese C. 2007. Biology's next revolution. *Nature* 445(7126):369–369.

Gonzalez CF, Kunka BS. 1987. Plasmid-associated bacteriocin production and sucrose fermentation in *Pediococcus acidilactici*. *Appl Environ Microbiol.* 53(10):2534–2538.

Hacker J, Carniel E. 2001. Ecological fitness, genomic islands and bacterial pathogenicity. *EMBO Rep.* 2(5):376–381.

Jacob AE, Hobbs SJ. 1974. Conjugal transfer of plasmid-borne multiple antibiotic resistance in *Streptococcus faecalis* var. zymogenes. *J Bacteriol.* 117(2):360–372.

Jain R, Rivera MC, Moore JE, Lake JA. 2003. Horizontal gene transfer accelerates genome innovation and evolution. *Mol Biol Evol.* 20(10):1598–1602.

Jeong H, Mason SP, Barabási A-L, Oltvai ZN. 2001. Lethality and centrality in protein networks. *Nature* 411(6833):41–42.

Koch A. 1981. Evolution of antibiotic-resistance gene-function. *Microbiol Rev.* 45(2):355–378.

Koonin EV, Wolf YI. 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* 36(21):6688–6719.

Lili LN, Britton NF, Feil EJ. 2007. The persistence of parasitic plasmids. *Genetics* 177(1):399–405.

Nagarajan S, Srivastava S, Sherman LA. 2014. Essential role of the plasmid hik31 operon in regulating central metabolism in the dark in *Synechocystis* sp. pcc. 6803. *Mol Microbiol.* 91(1):79–97.

Nogueira T, Rankin DJ, Touchon M, Taddei F, Brown SP, Rocha EP. 2009. Horizontal gene transfer of the secretome drives the evolution of bacterial cooperation and virulence. *Curr Biol.* 19(20):1683–1691.

Norman A, Hansen LH, Sørensen SJ. 2009. Conjugative plasmids: vessels of the communal gene pool. *Philos Trans R Soc B Biol Sci.* 364(1527):2275–2289.

Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* 405(6784):299–304.

Rankin D, Rocha E, Brown S. 2010. What traits are carried on mobile genetic elements, and why? *Heredity* 106(1):1–10.

Ross R. 1915. Some a priori pathometric equations. *Br Med J.* 1(2830):546–547.

Sørensen SJ, Bailey M, Hansen LH, Kroer N, Wuertz S. 2005. Studying plasmid horizontal transfer in situ: a critical review. *Nat Rev Microbiol.* 3(9):700–710.

Stewart FM, Levin BR. 1977. The population biology of bacterial plasmids: a priori conditions for the existence of conjugationally transmitted factors. *Genetics* 87(2):209–228.

Tazzyman SJ, Bonhoeffer S. 2013. Fixation probability of mobile genetic elements such as plasmids. *Theor Popul Biol.* 90:49–55.

Tazzyman SJ, Bonhoeffer S. 2014. Plasmids and evolutionary rescue by drug resistance. *Evolution* 68(7):2066–2078.

Thomas CM. 2000. Paradigms of plasmid organization. *Mol Microbiol.* 37(3):485–491.

Thomas CM, Nielsen KM. 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol.* 3(9):711–721.

Werren JH. 2011. Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc Natl Acad Sci U S A.* 108(Suppl 2), 10863–10870.

Wolfram Research, Inc. 2010. Mathematica, version 8.0 edition. Champaign (IL): Wolfram Research, Inc.

Young JP, Crossman LC, Johnston AW, Thomson NR, Ghazoui ZF, Hull KH, Wexler M, Curson AR, Todd JD, Poole PS, et al. 2006. The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biol.* 7(4):R34.

Zotenko E, Mestre J, O'Leary DP, Przytycka TM. 2008. Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality. *PLoS Comput Biol.* 4(8):e1000140.