Spatial structure and differential gene dispersal explains the evolutionary association of accessory genes with prophages

Sultan A. Nazir[1,*] and Bram van Dijk[1]

[1]Theoretical Biology, Utrecht University, The Netherlands
[*]Corresponding author: s.a.nazir@uu.nl

February 28, 2025

Abstract

Genes encoding for virulence factors are frequently found on prophages, yet the evolutionary forces driving this association remain unclear. The evolutionary association of mobile elements with host-beneficial genes is known to be hindered by the stability of chromosomal loci and competition with mobile elements lacking accessory genes. Using a mathematical model that incorporates these constraints, we identify two key mechanisms that help overcome these barriers, resulting in the evolutionary linkage of accessory genes, like virulence genes, to prophages. First, we show that migration in between bacterial populations favours this linkage when the gene is beneficial in an environment (e.g. virulence genes) where prophage induction is also upregulated. Second, we show that within-population spatial dynamics also promotes the association of weakly selected genes and phages. Here, virion dispersal allows phage-encoded genes to spread into patches of bacteria lacking the gene, giving them a selective advantage over immobile chromosomal genes. While both of these mechanisms are evidently applicable to phages, this is less clear for plasmids and other mobile elements. This helps to resolve the long-standing question of why phages, rather than other mobile genetic elements, appear to frequently carry virulence genes as cargo. By demonstrating how spatial heterogeneity and phage lifecycle dynamics drive gene association, our work provides new insights into the evolution of phage-encoded pathogenicity and bacterial adaptation.

1 Background

Mobile genetic elements (MGEs) are genome segments that can move within or between genomes, often replicating independently of the chromosome. Although MGEs such as bacteriophages, plasmids, and transposons have been the focus of extensive research for several decades, their functions are still being elucidated (Hall et al., 2021; Mark Osborn and Böltner, 2002). Increasingly, studies reveal the crucial roles that MGEs play in shaping the ecological and clinical functions of microbial communities (Rankin et al., 2011b; Weisberg and Chang, 2023). MGEs are especially relevant because they can carry accessory genes, which are non-essential for the MGE itself, but may substantially alter the bacterial phenotype. Different types of MGEs appear to vary in their likelihood of carrying specific accessory genes (Rankin et al., 2011b), but the mechanisms behind these associations remain poorly understood. For example, a recent analysis shows that phages are more likely to encode virulence factors than other MGEs (Takeuchi et al., 2023). This raises the question: what makes phages uniquely predisposed to turn their hosts into pathogens?

While many MGEs have mutualistic relationships with their hosts, bacteriophages lyse bacteria to produce more copies of themselves. Their integration into chromosomes, like transposons, may kill cells by

disrupting essential genes (Baltrus, 2013). Even if not for the peril of lysis and DNA damage, harbouring self-replicating MGEs pose metabolic costs to the bacterial hosts, such as the resources required for plasmid maintanence and replication (San Millan and MacLean, 2017). MGEs can compensate for these costs by carrying and expressing certain genes that benefit their bacterial host. This enables MGEs to navigate in between mutualism and parasitism (Harrison et al., 2015). However, as MGEs and their cargo are also readily lost, genes with a very high fitness benefit are more likely found on the chromosome (Bergstrom et al., 2000; Lehtinen et al., 2021). Therefore, accessory genes are typically non-essential, *e.g.* coding for functions required for local adaptation (Eberhard, 1989). In addition, they often appear to code for traits expressed extracellulary, say, in the form of public goods (Nogueira et al., 2009). Different hypotheses have been proposed to explain why non-essential and cooperative traits can be found on MGEs rather than on chromosomes (or immobile parts of the genome) (Rankin et al., 2011b). For phage-encoded pathogenicity in particular, Abedon and LeJeune have written an excellent review of different hypothesised and empirically supported mechanisms (Abedon and LeJeune, 2007).

Models have been a useful tool to unpack the conditions under which MGEs carry certain accessory genes. For example, Bergstrom et al. proposed a local adaptation hypothesis, wherein plasmid-encoded genes persist by transferring locally adapted genes to newly arriving strains (Bergstrom et al., 2000). Later, Smith proposed the infectivity hypothesis, which stated that MGEs may evolve to carry genes for public goods in order to force defectors to cooperate (Smith, 2001). More recently, van Dijk et al. have shown that HGT can rescue a beneficial gene (therewith increasing population fitness) when the benefit is not strong enough to maintain the gene against mutational loss (van Dijk et al., 2020). From these studies, we can generalise a *rescue hypothesis*: genes are MGE-encoded when horizontal gene transfer (HGT) can rescue the gene from being lost through negative selection, mutational loss or the emergence of cheaters.

While the above-mentioned models and the rescue hypothesis are insightful, they do not fully capture the biology of how MGE-driven mobilisation comes about. For an accessory gene to become mobilised by residing on an MGE, its mobility must not only be beneficial, it must also persist under competition against MGEs that do not encode the accessory gene. Competition with such "non-carrier" MGEs has indeed been shown to hinder the association of MGEs with accessory genes (Geoffroy and Uecker, 2023; Mc Ginty et al., 2011). In addition, most mechanisms of gene loss will act on the mobile locus as much as they do on the chromosomal locus, and the gene may fail to persist by mobility alone. In this case, increasing relatedness at the mobile locus by infection has been shown to result in evolution of MGE-encoded genes (Rankin et al., 2011a).

More recent work by Lehtinen et al. has shown that an accessory gene that is initially more abundant on the mobile (MGE) locus than on the immobile (chromosomal) locus will, under weak selection, stay on the mobile locus (Lehtinen et al., 2021). This "priority effect" may emerge when a species is more likely to acquire the gene via an MGE than by mutational innovation on the chromosome. While this mechanism is of likely importance to plasmids, temperate phages are generally thought to have more narrow host ranges due to their specificity during integration (Bobay et al., 2013; Bobay and Ochman, 2018; Moura de Sousa et al., 2021). Moreover, there is evidence that genes associated with prophages have a chromosomal origin due to their localisation close to the attachment site of the prophage DNA (Brüssow et al., 2004). In other words, while much research on MGEs was done with plasmids as a model MGE – implicitly assuming generalisability to other MGEs – few models have been developed to enhance our understanding of phage-encoded accessory traits.

In this article, we focus on understanding why bacteriophages carry beneficial accessory genes by means of mathematical modelling. While our model could be applied to accessory genes more generally, we built our model especially in the context of exotoxin production, a very common virulence factor encoded by prophages (Gummalla et al., 2023; Boyd and Brüssow, 2002). The benefits provided by these extracellular virulence factors may be partially enjoyed by other bacteria that also gain access to

resources of the plant or animal through tissue damage. Therefore, we model the expression of the gene as a cooperative trait with certain degree of privatisation of benefits in order to maintain a positive selection pressure. As a preliminary analysis, our mathematical model verifies that the "rescue" of certain beneficial genes by lysogenic phages is highly sensitive to competition with non-carrier phages. In competition with non-carrier phages, we find that phage-encoded genes cannot invade when the gene is already chromosomally encoded. However, we find that the gene-phage association can invade under conditions described within a framework of spatial structure. Firstly, by extending our ODE model to include a host-environment metapopulation structure, we find that when prophage induction and expression of the beneficial cargo gene are both limited to the host environment, natural selection favours the beneficial gene to be phage-encoded. Secondly, using an individual-based model, we propose phage dispersal within a spatially structured population as a mechanism that catalyses the evolution of phage-encoded cargo genes. Here, since natural selection occurs at a local scale, phage particles help the gene disperse to pockets lacking the gene, reinforce selection and replace the chromosomal gene at a global scale. Taken together, our work shines light on the exogenous factors that contribute to the evolution of phage-encoded accessory traits, especially virulence.

2 Methods

Infection cycle and genotype space: We implemented an infection model, as shown in Figure 1(A), where uninfected prokaryotic cells (denoted by U) are susceptible to infection by virion particles (denoted by V) such that the viral DNA integrates into the cell as a prophage and converts each cell into a lysogen (denoted by L). Then, the prophage is induced at a rate α which causes cell lysis and the releaase virion particles equal to its burst size r_V. Assuming they occur simultaneously, we ignore the lag in time that occurs between infection and integration and between induction and lysis. Virion particles in the external environment can etiher decay at a rate δ or infect a susceptible cell at a rate β. In this framework, we assume superinfection immunity such that a cell cannot be infected by a virion particle if it already harbours a prophage.

The viral DNA may or may not carry the gene of interest. If the virion particle is a carrier of the gene, it is marked by a subscript τ. When a non-carrier virion V infects a susceptible cell, it creates a non-carrier prophage, but if a carrier virion V_τ does, it creates a carrier prophage, and a lysogen with a carrier prophage is indicated by a subscript p to denote that the gene is phage encoded. Every cell, regardless of whether it is infected by a prophage or not, has a chromosomal locus where the gene may or may not be present. If the cell is a carrier of the gene at the chromosomal locus, it is indicated by a subscript c. Hence, as shown in Figure 1(B), there are six possible genotypes of prokaryotic cells (U, U_c, L, L_c, L_p, and L_{cp}) and two kinds of virion particles (V and V_τ).

Cooperation and privatisation: All of the cells grow at a fixed rate and compete for resources along with an intrinsic death rate of d. The effect of carrying the gene of interest on the growth dynamics is modelled as follows. We define "minimum carrying capacity" as the carrying capacity in the absence of the gene and set its value to one for non-dimensionalisation, i.e. all population sizes are expressed relative to the minimum carrying capacity. Let ϕ represent the maximum additional biomass that can be generated due to gene expression. If f_τ denotes the fraction of carriers in the population, by the definition of ϕ, the net carrying capacity becomes $1 + \phi f_\tau$. If the benefits produced by the carriers are equally shared among all cells, each cell receives a benefit of ϕf_τ. However, positive assortment and other mechanisms enable cooperators to have disproportionate access to these benefits. To model this, as shown in Figure 1(C), we introduce the parameter π, the fraction of the benefits produced by a carrier that is privatised, i.e. retained exclusively by the producer. Consequently, all cells (including carriers themselves) receive a shared benefit of $(1 - \pi)\phi f_\tau$, while carriers gain an additional private benefit of $\pi\phi$. This framework allows π to function as a "privatisation parameter", or more generally, for tuning the rate of selection acting on the gene.

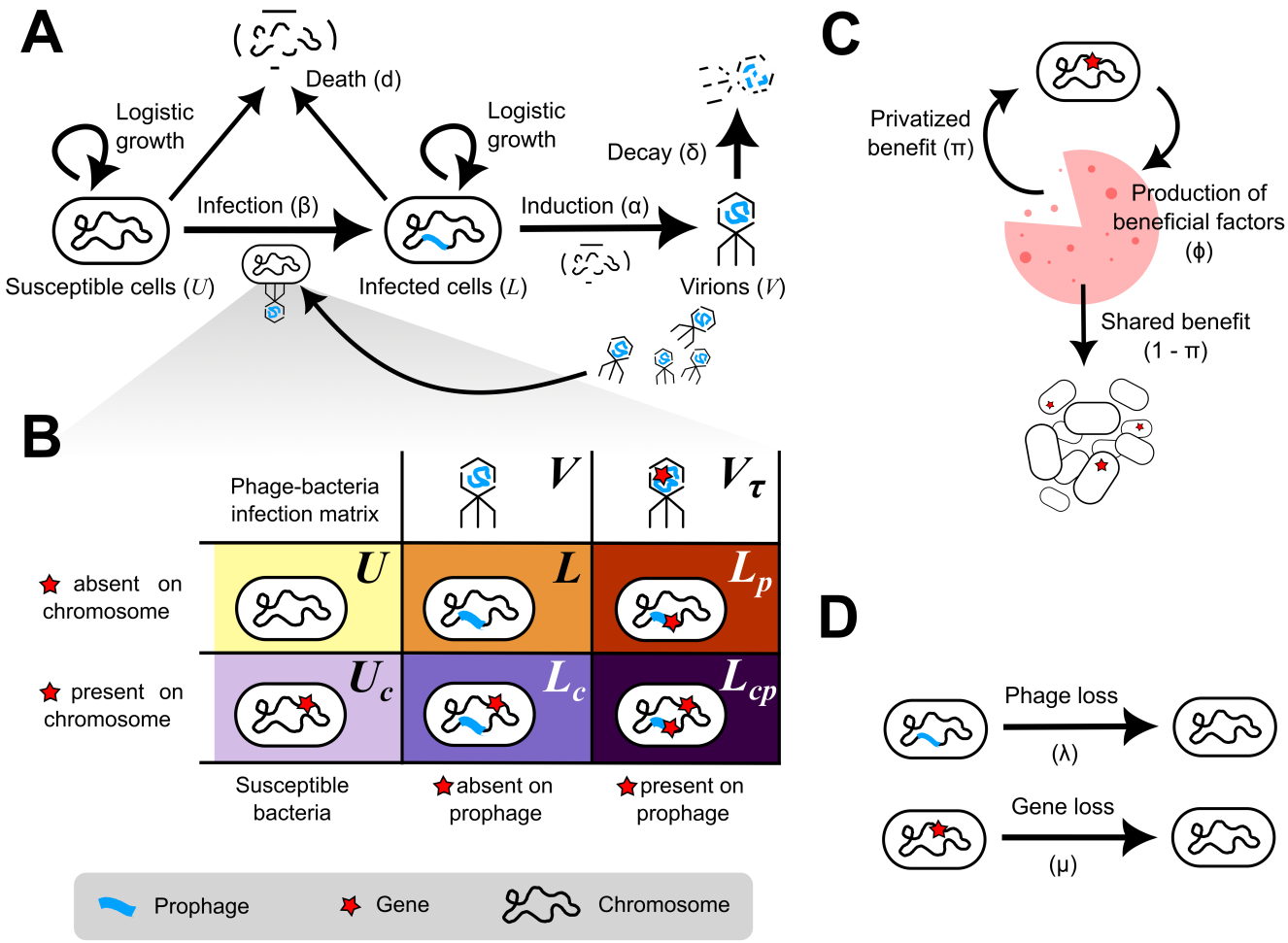


Figure 1: **An illustration of all of the model components:** (A) The lifecycle of the phage consists of a prophage stage that begins upon infecting a susceptible bacterium and a virion state which is produced upon induction of the prophage. As a virion, the phage can be lost to environmental degradation. (B) Based on whether the infecting phage carries the gene of interest or not, and whether the susceptiple cell carries the gene at the chromosomal locus, we get six cell genotypes and four of them carry the gene. The subscript denotes whether the gene is present on the chromosomal locus (c) or the phage locus (p). (C) Out of the ϕ benefits produced by a carrier, π fraction is privatised by the producer and the remaining is shared with the whole cell population. (D) Both the gene and the phage can be lost from a microbial genome due to mutations.

Mutations: In the model, we consider two kinds of deletion mutations, as shown in Figure 1(D). Firstly, every "carrier" (henceforth used to refer to a cell that has the gene of interest) can mutate into a "non-carrier" by gene loss at a rate μ. Note that this mutation occurs at the prophage locus and the chromosomal locus, at the same rate. Secondly, every lysogen can lose its prophage at a rate λ. For simplicity, we assume that this loss coincides with the loss of superimmunity, i.e. the cell will immediately be available for future phage infections. If a cargo gene resided on the prophage, this gene is lost along with the phage.

The complete set of differential equations and the parameter descriptions are given in Appendix A and B.

2.1 Spatial model

For the spatially structured version of the model, we implemented an Individual based model (IBM) of a finite population of bacteria on a square grid. The model captures the same processes described above using the ODEs while introducing spatial structure and finiteness to the cell and virus populations with the following elements.

Birth and Death: A cell can copy itself within a neighbouring empty site with a probability b as shown in Figure 2(A). We assume the rate of birth is equal for all cells at all times. The effect of cooperative interations is expressed in terms of varying death rates. We chose to vary the death rates instead of the birth rates because birth is inherently a density-dependent process on a grid. We do not want increasing competition for space to interfere with the evolutionary dynamics. In the absence of any benefit from carriers, a cell dies with probability $d_s + \Delta d$ where d_s is the intrinsic death rate in the spatial model and Δd is the rate of death that can be avoided with benefits from the gene. The maximum population size is limited by the size of the grid (total number of grid points). Therefore, to simulate dynamics comparable to the ODE, we adjust the death rate parameters such that the grid size (the maximum carrying capacity) is $1+\phi$ times the average steady population size in the absence of the gene (the minimum carrying capacity). Analogously to the ODE model, the benefits of the cooperative gene are partially privatised. A carrier privatise π fraction of its benefits to itself (lowering its own chance of death), with the remaining fraction $(1-\pi)$ of benefits shared with others. Therefore, the probability of death is

$$\mathbb{P}_{\text{death}}(\pi) = d_s + \Delta d\Big(1 - \pi \mathbb{1}_{\text{carrier}} - (1-\pi) f_\tau\Big)$$

where $\mathbb{1}_{\text{carrier}}$ is the indicator variable

$$\mathbb{1}_{\text{carrier}} = \begin{cases} 1, & \text{cell is a carrier} \\ 0, & \text{cell is not a carrier} \end{cases}$$

and f_τ is the fraction of carriers on the grid.
To ensure that the ODE is the a good approximation of the spatial model when well-mixed, all the rate parameters in the IBM are scaled by a factor of $\frac{b}{1+\phi}$ (for example, $d_s = d\frac{b}{1+\phi}$) and we set $\Delta d = \phi\frac{b}{1+\phi}$. See Appendix C for verification of the well-mixed approximation.

Phage lifecycle: Extracellular virions are also modelled as discrete entities. Therefore, at every grid point, we have an integer measure of the number of extracellular virion particles of each type - non-carrier and carrier. Upon induction and lysis, the cell releases the virion particles into the 'Virion grid' at the same grid coordinate as shown in Figure 2(B). The released virions are carriers if the induced prophage was a carrier. Irrespective of whether it is a carrier or not, at every time point, each virion particle can decay or migrate to a neighbouring grid point with fixed probabilities for each. The migration probabilty V_{diff} controls the rate of diffusion of these particles. When a susceptible cell is located at a specific coordinate on the 'Cell grid', all virion particles present at the same coordinate on the Virion grid compete to infect the cell. Each virion particle has a fixed probability of infecting the cell.

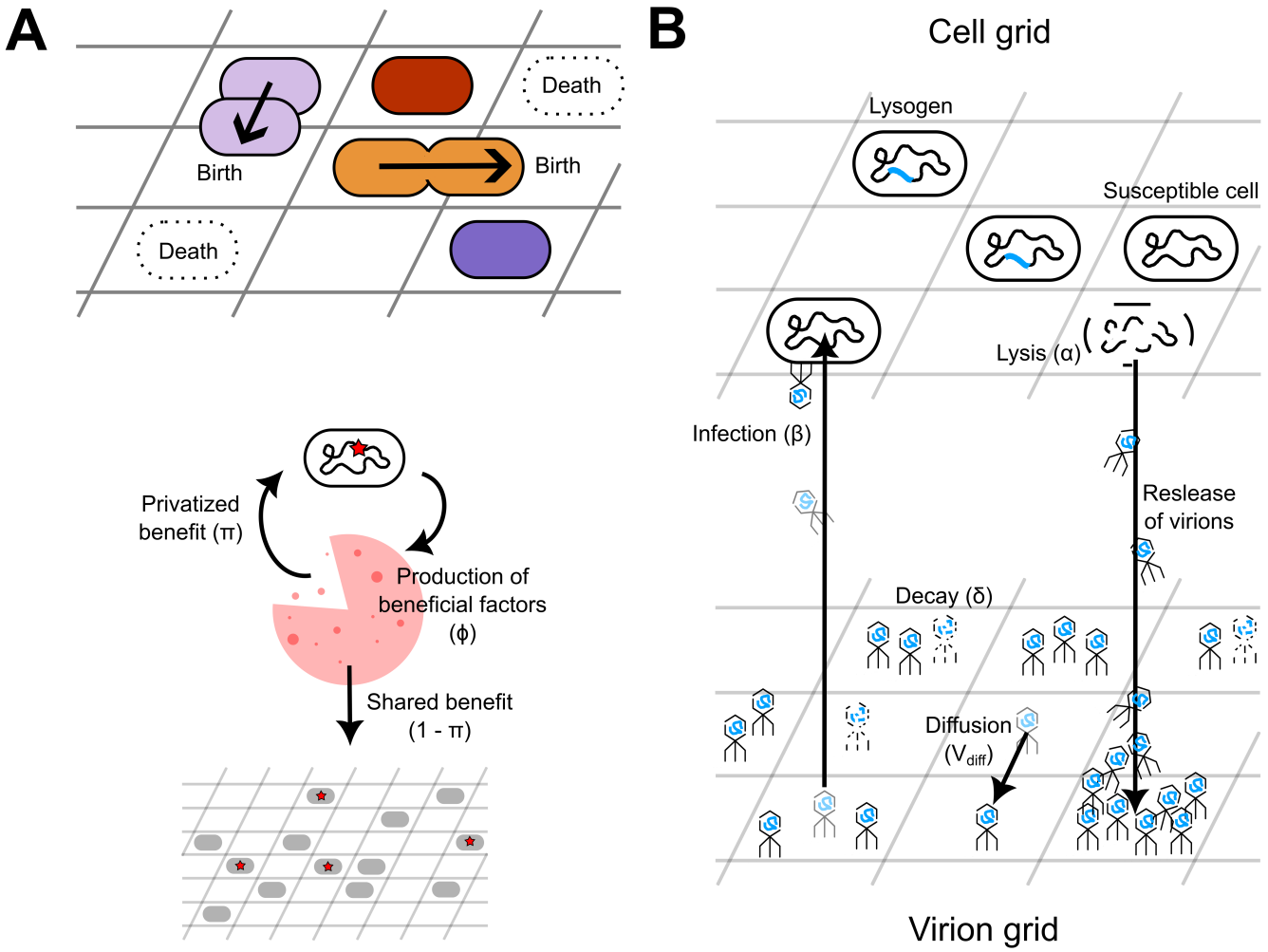


Figure 2: **Spatial model:** (A) Individual cells are placed on a square grid and the occupied grid points are coloured according to the cell genotype (identical to the colours in Figure 1(B). On the grid, the cells can reproduce into an unoccupied grid point in its Moore neighbourhood. The cells can also die, leaving an unoccupied grid point. The propensity of death reduces with the beneifts received from the gene. If a cell is a carrier, it privatises π fraction of benefits and shares the remaining $1 - \pi$ fraction with other cells on the grid. (B) Apart from the 'Cell grid', there is a 'Virion grid' where more than one virion particle can occupy a single grid point. They infect cells and undergo part of their lifecycle as prophages on the Cell grid. Upon induction and lysis, virions are released onto the Virion grid where it can diffuse and ultilmately infect a susceptible cell or undergo decay.

The outcome of the competition, including whether the cell is infected or not, is determined using a roulette wheel selection.

Software: Simulations of the analytical model were run using Jupyter notebook (Kluyver et al., 2016), and the packages Matplotlib (Hunter, 2007), NumPy (Harris et al., 2020), SciPy (Virtanen et al., 2020), and Pandas (Wes McKinney, 2010). The spatial model is implemented using Cacatoo (van Dijk, 2022), and an interactive version is available at `https://sultannazir.github.io/bacteria_prophage_codes/`.

Data availability: All code and data for the model are available online at `https://github.com/sultannazir/bacteria_prophage_codes`.

Authors' contributions: S.A.N. conceived the model with input from B.D. and designed and/or programmed the model, ran simulations, generated figures from the output data and drafted the initial manuscript. All authors gave final approval for publication and agree to be held accountable for the work performed therein.

Competing interests: We declare that we have no competing interests.

3 Results

Before exploring the conditions under which the gene evolves to be phage-associated, we first identified conditions under which the phage and/or the gene can persist in the population. As both the gene and the phage are subject to loss mutations, they will naturally disappear from the system in the absence of self-replication. They must make sufficient copies of themselves, at a rate higher than at which they are lost, in order to persist.

Ignoring the effects of the phage, the gene can persist if selection is sufficiently high:

$$\pi > \frac{\mu}{\phi}$$

Similarly, ignoring the effects of the gene, the phage can persist if the induction rate is sufficiently high:

$$\alpha > \frac{\lambda(\beta + \delta)}{\beta(r_V - 1) - \delta}$$

See Appendix D for the derivations.

When both the gene and the phage (or mobile element in general) are present in the population, studies have shown that one can rescue the other by compensating for weak selection with high transmission rate (van Dijk et al., 2020) or vice versa (Millan et al., 2014), shifting the thresholds we identified above. However, we verify that such rescue mechanisms are not successful in our model. This is because if the gene is lost due to mutation on the chromosomal locus, it is lost on the phage locus at the same rate. Hence, the phage cannot rescue the gene by transmission alone. This observation is in agreement with the findings of Mc Ginty et al. (2011). Similarly, if the phage is lost due to mutation, it cannot be rescued by linking with a beneficial gene when the gene can instead be on a more stable chromosomal locus (Bergstrom et al., 2000). Upon numerically simulating the model for different values of the parameters π and α, we observed that the thresholds for gene and phage persistence do rigidly exist, roughly partitioning the $\pi - \alpha$ parameter space into four quadrants (see Appendix E). Although there is a quadrant where both the gene and the phage persist, this does not necessarily imply that they are evolutionarily associated. For example, a population with a chromosomally encoded gene (U_c) could simply be chronically infected by a non-carrier phage (L_c).

Gene association with prophage does not evolve in a well-mixed population

To investigate eco-evolutionary outcomes where phages carry the beneficial gene as cargo from a chromosomal origin, we performed numerical simulations across the $\pi - \alpha$ parameter space. Since our model does not include gene translocation from the chromosome to the prophage, we simulate the invasion of a mutant that encodes the gene on a prophage instead of their chromosome, by setting the initial state to $L = 0.500, L_c = 0.495, L_p = 0.005$, and the remaining genotype abundances to zero. In other words: all cells have a prophage, almost all carrier cells encode the gene chromosomally, but 1% of cells have the genes encoded on a prophage instead. Hence, the question becomes whether the rare mutant encoding the gene on a prophage invades in the population. Within the persistence quadrant (see Figure 3) — where both the phage and gene can coexist — we examined the association between

the gene and the phage in terms of the percentage of phages that carry the gene and the percentage of gene copies that occur on the phage loci (shown with a 2D colour gradient).

We observed that these conditions are not sufficient for the gene-phage association, as the mutant does not establish in the population (Figure 3(A) remains fully white). This result indicates that, assuming genes with chromosomal origins, it is difficult for the gene to be genetically linked to phages as cargo. As prophage-encoded (more generally MGE-encoded) accessory genes are typically traced back to a chromosomal origin, our model cannot elegantly explain the persistence of phage-encoded cargo traits.

Why is it so difficult to evolve an association between a prophage and a beneficial gene? One reason is that the chromosomal locus offers a robustness advantage, as the phage locus is at risk of being lost through two processes: regular mutations (identical to the chromosomal locus), and when the phage is lost entirely. While the benefit of the phage locus may be to spread the gene by infecting others, this benefit is weak when most individuals carry the gene. Only with weak selection and high infectivity can the phage locus utilise this distinct advantage, as here there are sufficient non-carrier cells available for infection. However, these conditions also favour the transmission of non-carrier phages that are generated due to gene loss mutations. Therefore, the net growth rate of the gene at the phage locus can never exceed the growth rate at the chromosomal locus in a well-mixed system. With the added assumption of redundancy in multiple copies of the gene within a cell, evolution favours the chromosomal locus, wiping out the phage-encoded mutant.

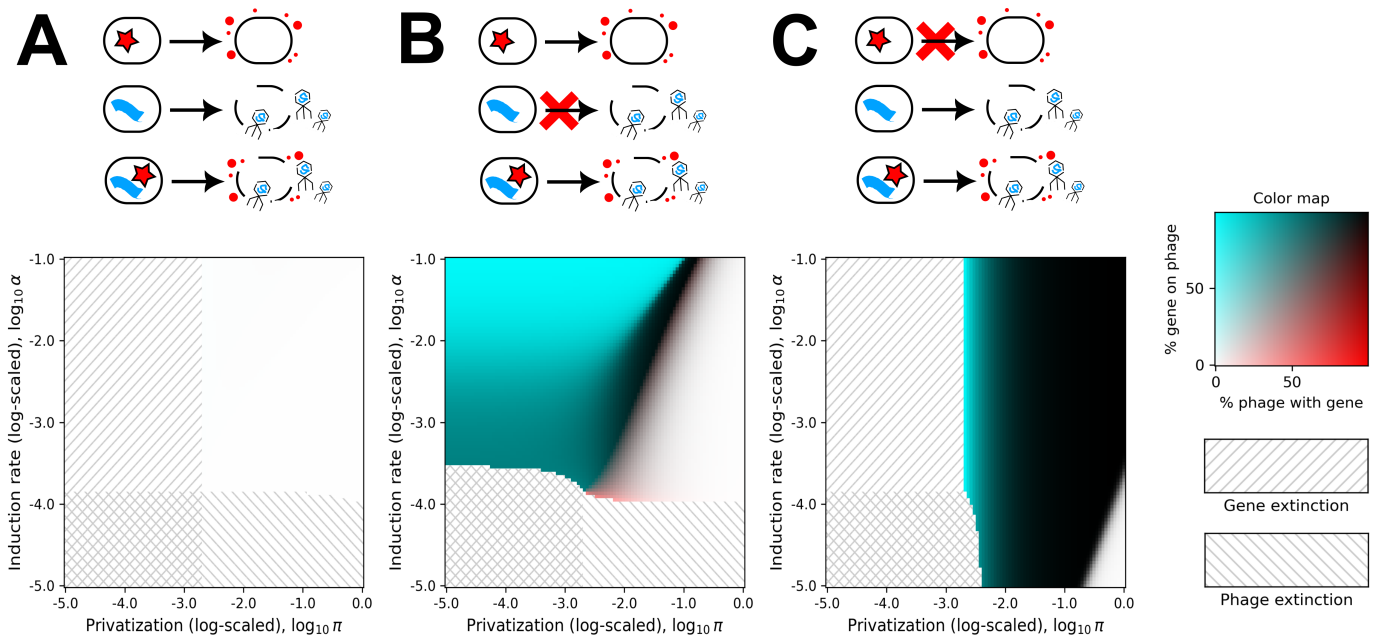


Figure 3: **Coregulation of prophage induction and gene expression leads to association of phage and gene:** Shown are gene-phage association metrics for different values of π and α at the end of 10^8 time steps. The shaded region represents the extinction of the gene, the phage or both. Initial condition: $L = 0.495, L_c = 0.500, L_p = 0.005, \text{others} = 0.000$. (A) In the absence of any coregulation mechanism, no association evolves. (B) When prophage induction is limited to carrier cells, the phage evolves to associate with the gene. (C) When gene expression is limited to cells with prophages, the gene evolves to associate with the prophage.

Coregulation of prophage induction and gene expression can evolve association

Above, we show that adding a competing non-carrier phage and a mutationally more stable chromosomal locus, gene-phage associations do not evolve. Therefore, we explore additional mechanisms that could favour their association. Abedon and LeJeune proposed that the empirically observed epistatic interactions between virulence genes and phage genes could explain why we see many virulence genes associated with prophages (Abedon and LeJeune, 2007). Prophage induction is a special feature of temperate phages that differentiates it from other MGEs and makes its horizontal transmission a strongly regulated process. This can give rise to biological scenarios of regulatory coupling between prophage induction and gene expression. For example, Shiga toxin production in *E. coli* is directly tied to prophage induction, since toxin release depends on host lysis (Wagner et al., 2002). Similar mechanisms regulate exotoxins such as diphtheria toxin and other enterotoxins, where transcriptional upregulation during prophage induction ensures effective pathogenicity (Tinsley et al., 2006; Zeaki et al., 2015). We also have evidence of phage proteins that play crucial regulatory roles in the production of exotoxins (Fortier and Sekulovic, 2013). We consider two extreme cases of coregulation of prophage induction and gene expression that can be easily implemented into the ODEs.

Case 1: Prophages cannot be induced in non-carrier cells. We first examined a scenario where prophages can only be induced in the presence of the gene (see Figure 3(B)). In the model, this is implemented by preventing regular lysogens (L) from being induced, while lysogens carrying the gene (L_c, L_p and L_{cp}) can be induced as normal (see Appendix F.1 for the ODEs). Under these conditions, phages that lack the gene incur a cost: if they infect a non-carrier cell, they fail to replicate. This introduces a new selective pressure favouring the acquisition of the gene at the phage locus. As a result, the phage helps maintain the gene even below the selection-mutation threshold, as its own propagation now depends on the presence of the gene. However, when the induction rate is low and privatisation is high, the chromosomal gene undergoes a strong selective sweep, limiting selection at the phage locus due to redundancy from the chromosomal copy of the gene.

Case 2: The gene is only expressed during prophage induction. In this inverse scenario, the gene is only expressed when a prophage is present in the cell. In the model, we implement this by preventing U_c cells from expressing the gene (see Appendix F.1 for the ODEs). This mechanism leads to a strong association between the gene and prophages when selection pressure is above the selection-mutation threshold (indicated by black regions in Figure 3(C)). Since chromosomal copies of the gene are no longer beneficial in susceptible cells, they are no longer dominated by chromosomal carriers. Consequently, associating with the phage locus provides an advantage by ensuring that the gene is always carried by a lysogen. Even below the transmission-mutation threshold, phages persist due to gene-level selection for phage association.

1. Spatial structure between populations: Environmental heterogeneity as a mechanism of coregulation

So far, we have explored endogenous factors that promote the association of beneficial genes with prophages. However, beyond within-cell biochemical interactions, the co-regulation of prophage induction and gene expression can also emerge from exogenous factors such as environmental heterogeneity. The gene may not be advantageous in all environments — for example, a pathogenicity trait is only beneficial within a host environment. Similarly, prophage induction is often influenced by extracellular conditions. Henrot and Petit identified multiple signaling molecules within the mammalian gut microbiota that can trigger prophage induction (Henrot and Petit, 2022). When prophages reside in pathogens whose virulence is expressed in the gut, pathogenicity becomes indirectly coregulated with the induction of these prophages. Additionally, gene expression itself may alter the cellular environment in ways that promote induction. For example, Pattenden et al. propose that virulent bacterial strains, due to their faster growth and increased exposure to stress, induce their prophages more frequently

(Pattenden et al., 2021). Furthermore, immune responses targeting virulent cells may inadvertently trigger prophage activation.

Motivated by these observations, we next consider a scenario in which the gene is beneficial only in a specific environment, and this same environment promotes prophage induction. To model this, we introduce a second compartment representing the "outside world" where toxin production confers no fitness benefit and prophage induction signals are weak (see Appendix F.2 for the modified equations). In the absence of migration, the system in this outside world will converge towards a population of non-carrier susceptible (U) cells. We connect this compartment with the "host environment" by introducing migration between them. We assume that the migration is symmetric and that both compartments have equal maximum cell densities.

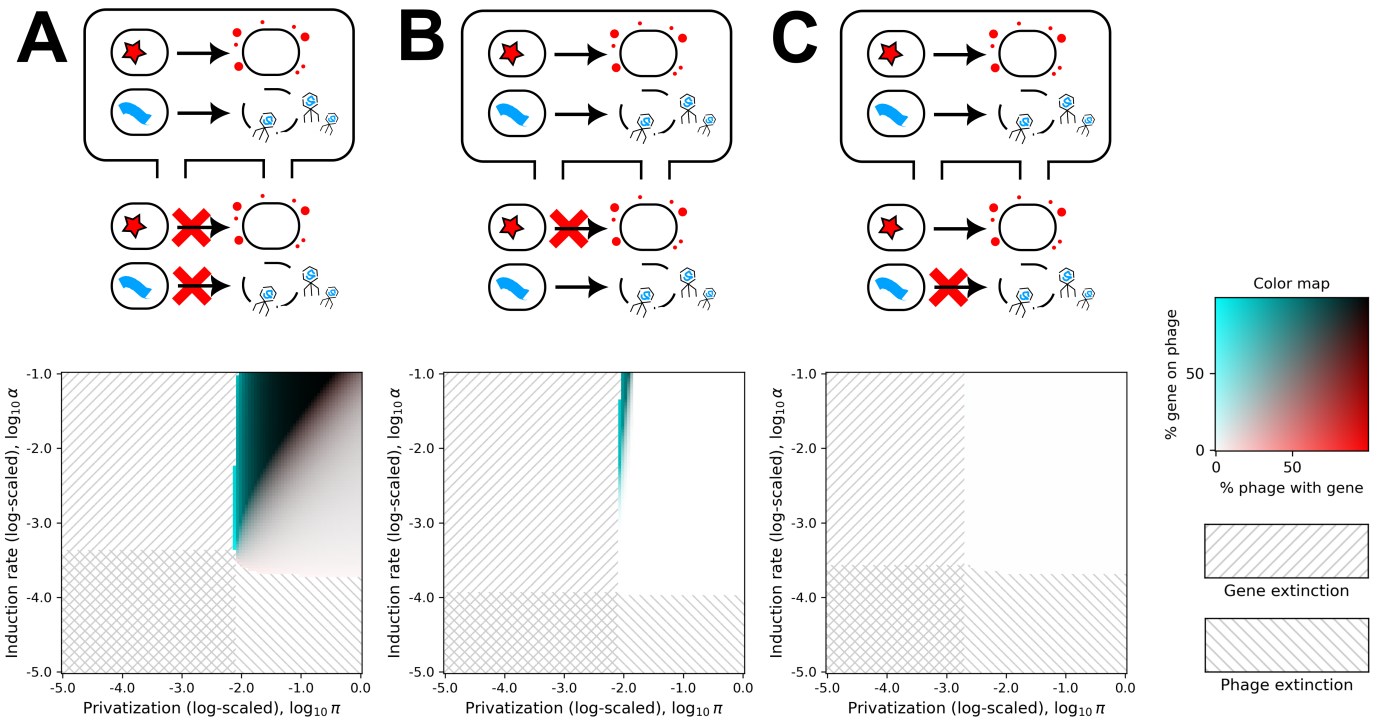


Figure 4: **Two-compartment model: Coregulation through spatial heterogeneity leads to association of phage and gene.** Shown are gene-phage association metrics within the "host compartment" for different values of π and α at the end of 10^8 time steps. The shaded region represents the extinction of the gene, the phage or both in the host compartment. Initial condition within the host compartment is always: $L = 0.495, L_c = 0.500, L_p = 0.005, \text{others} = 0.000$. (A) Prophage induction and gene expression are both limited to the host environment. Initially the outside world consists only U. (B) Only gene expression is limited to the host environment. Initially the outside world consists only L (C) Only prophage induction is limited to the host environment. Initially the outside world consists only U_c.

With this modification, we see a large portion of the parameter space permitting the invasion of phage-borne genes within the host environment (black colouring in Figure 4). This is because the cells in the outside world are inclined to lose both the gene and the phage, and the influx of these non-carrier susceptible cells increases the rate of phage-infections that successfully convert non-carriers into carriers within the host environment, hence promoting selection for mobility at the gene-level. It is interesting to note that gene-phage association is not a stable attractor when either of the compartments is considered to be a closed system. However, upon permitting migration between them, phage-association becomes a stable evolutionary attractor.

The models discussed thus far highlight the challenges in evolving MGE-gene associations de novo due to competition with non-carrier phages (or MGEs more generally) and the stability of chromosomal genes. However, most temperate phages regulate their lifecycle through prophage induction. We demonstrate that when prophage induction and gene expression become environmentally co-regulated, stable gene-phage associations can evolve. This promotes gene-level selection for mobility by ensuring a supply of non-carrier cells (U) for infection. However, natural microbial populations are also spatially structured *within* populations, exhibiting local variations in bacterial density and gene/phage distribution. In the next section, we explore whether spatial structure alone can drive gene-phage associations without requiring additional regulatory mechanisms.

2. Spatial structure within a population: Virion dispersal promotes the association of slightly beneficial genes with prophage

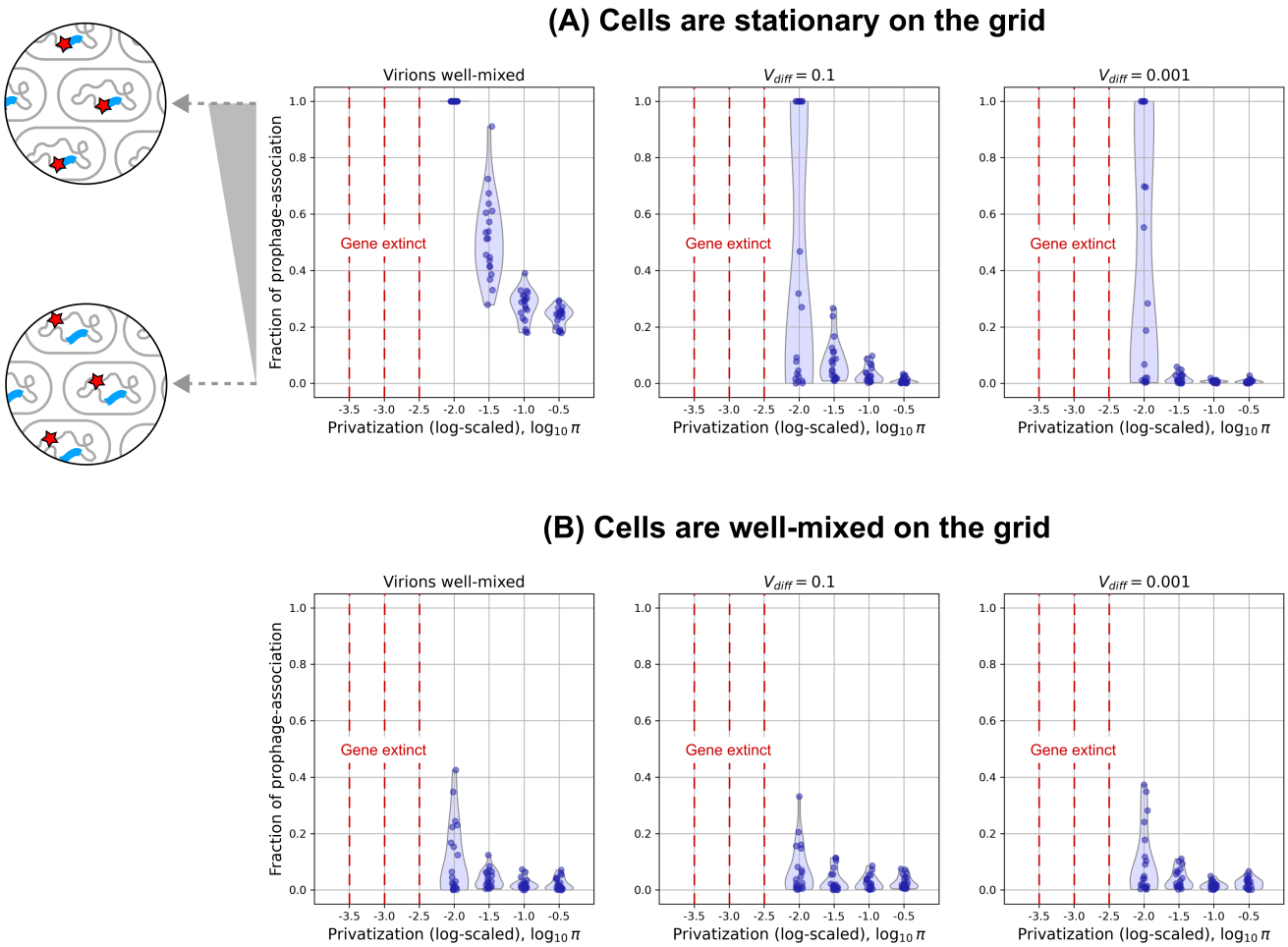


Figure 5: **Differential dispersal of virions in a spatially structured cell population leads to the evolution of gene-phage association for intermediate selection.** The plots show the distribution of final fraction of prophage-association in the gene population, excluding runs where the gene went extinct. For each value of π and V_{diff}, the distribution is plotted over 20 replicates of the stochastic simulation. (A) Cells are stationary on the grid. (B) Cells are well-mixed, i.e. the grid points on the 'Cell grid' are shuffled every time-step.

In the individual-based spatial model (described in Section 2.1), due to discrete population sizes, once a gene is lost entirely, it cannot re-emerge. To investigate whether phage-encoded cargo genes can establish, we conducted an invasion experiment. We initialized a population of L_c cells on a 120×120 grid and allowed mutant types (U, U_c, and L) to arise and stabilize their abundances. We then introduced

a one-time translocation mutation, converting 1% of L_c cells into L_p cells, effectively moving the gene from the chromosome to the prophage. If the phage-associated gene went extinct, we reintroduced the mutation. The simulation ran until either the gene was lost from both loci or the maximum time limit (5×10^6 steps) was reached. A "successful invasion" was defined as the complete replacement of the chromosomal gene by the phage-associated gene.

Running 20 replicates of the experiment across different values of privatisation (π) and virion diffusion rates (V_{diff}) revealed two key findings (Figure 5). First, phage-associated genes successfully invaded in some replicates under intermediate privatisation. While the ODE model never allowed full takeover of the phage locus, spatial simulations frequently did. Control experiments with well-mixed populations confirmed that this effect was driven by spatial structure rather than finite population size (see Figure 5(B)). Second, increasing virion dispersal significantly enhanced phage-associated gene invasion. While a change from 0.1 to 0.001 in virion diffusion rate had little impact, well-mixed virions led to successful invasion in all replicates with intermediate privatisation. However, when both cells and virions were well-mixed, virion diffusion no longer influenced gene-phage association, highlighting the role of differential dispersal in spreading the gene away from local competition with the chromosomal gene.

To better understand these dynamics, we analysed grid snapshots at different parameter values (see Figure 6). When they are spatially structured, due to local reproduction, cells of different genotypes form clusters in space. This creates niches in space where a newly introduced genotype will have different growth rates. Under weak selection, non-carrier cells (orange and yellow) persist in such clusters, creating pockets where carrier genotypes express a higher growth rate an invade. However, gene loss mutations continually regenerated these non-carrier regions. While the chromosomal gene could also invade these spaces, the phage-encoded gene possessed the advantage of discovering them via dispersing virions that happen to infect a susceptible cell within these patches. Further, we observed that lowering virion diffusion rate (Figure 6(2)) allowed phage-associated genes to reach high abundances, but their chance of outcompeting chromosomal genes was reduced because virions remained localized and unable to find distant gene-free regions. Conversely, maintainng a high diffusion rate but increasing the selection coefficient (Figure 6(3)) prevented the maintenance of gene-free regions due to strong selection sweeps. Therefore, the phage-associated gene is always surrounded by the chromosomal gene which is a better competitor due to mutational robustness. Thus, differential dispersal and persistent non-carrier populations were key to the success of phage-encoded genes.

Based on these findings, we hypothesize that accessory genes in bacteria evolved to be associated with phages because of its ability to disperse these genes to favourable niches through virion particles (see Figure 7. The occurrence of these favourable niches can be attributed to the weak selection pressure on the gene in a spatially structured cell population. This mechanism sets apart phage-encoded genes from genes associated with more locally transmitted MGEs such as plasmids and could potentially explain why phages differ from other MGEs in terms of the composition of their cargo.

4 Discussion

This study presents a suite of theoretical models to investigate the evolution of phage-associated genes originating from chromosomal loci. We show that chromosomal mutational robustness and competition with non-carrier phages create significant barriers to the evolution of phage-encoded accessory genes. However, we find that spatial structure – both between and within populations – can overcome these barriers. Previous studies have highlighted co-regulation between virulence and prophage genes as a mechanism for their association (Abedon and LeJeune, 2007). Extending this idea, we propose that environmental heterogeneity can drive phage-gene associations, while spatial structure within bacterial populations can promote association via virion-mediated gene dispersal, without any regulatory coupling.

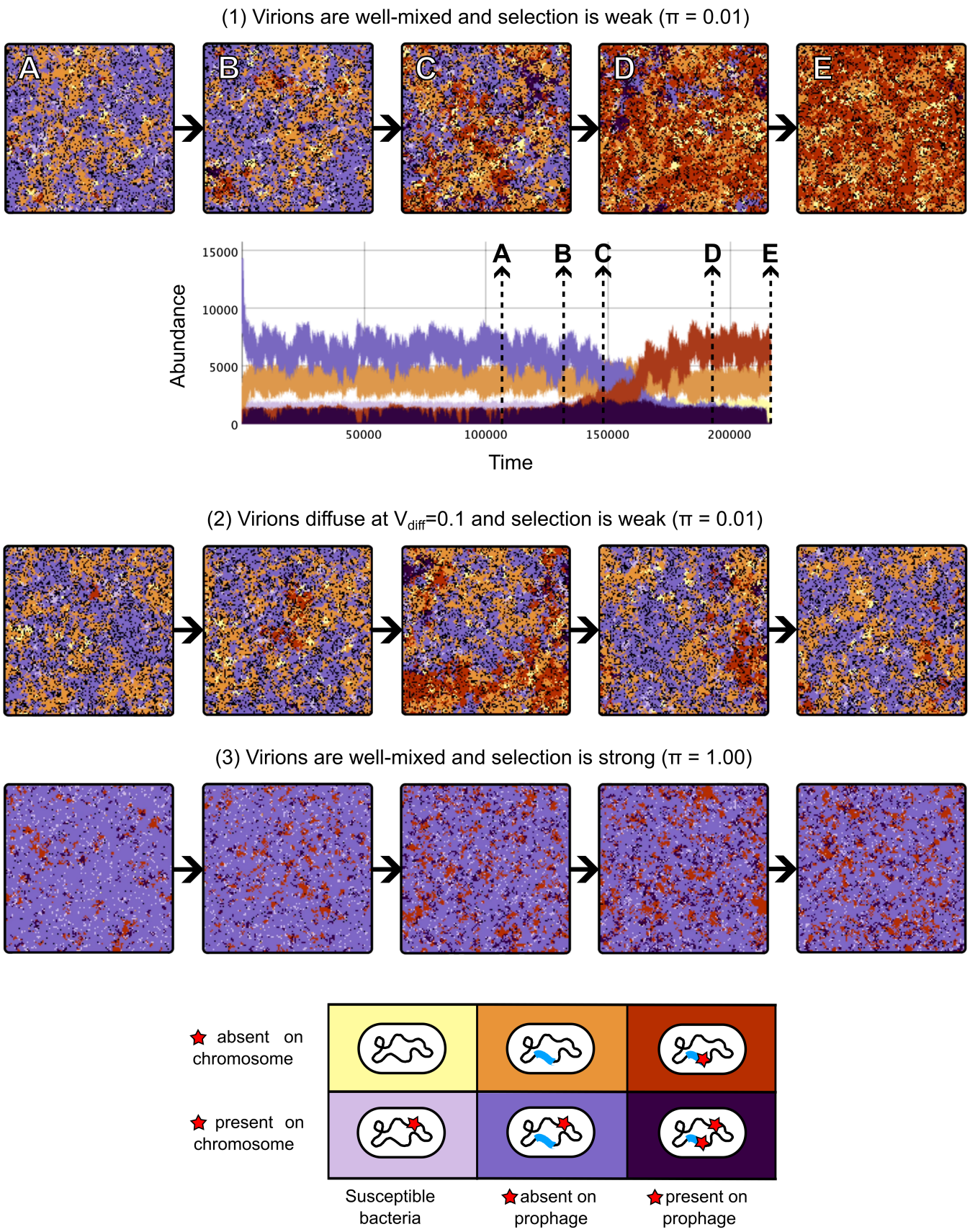


Figure 6: **Snapshots of the grid model at different time-points of a single run of the simulation.** The three cases shown here give a side-by-side comparison to illustrate the limitations of lowering virion dispersal and increasing selection coefficient in the successful invasion of the phage-encoded gene.

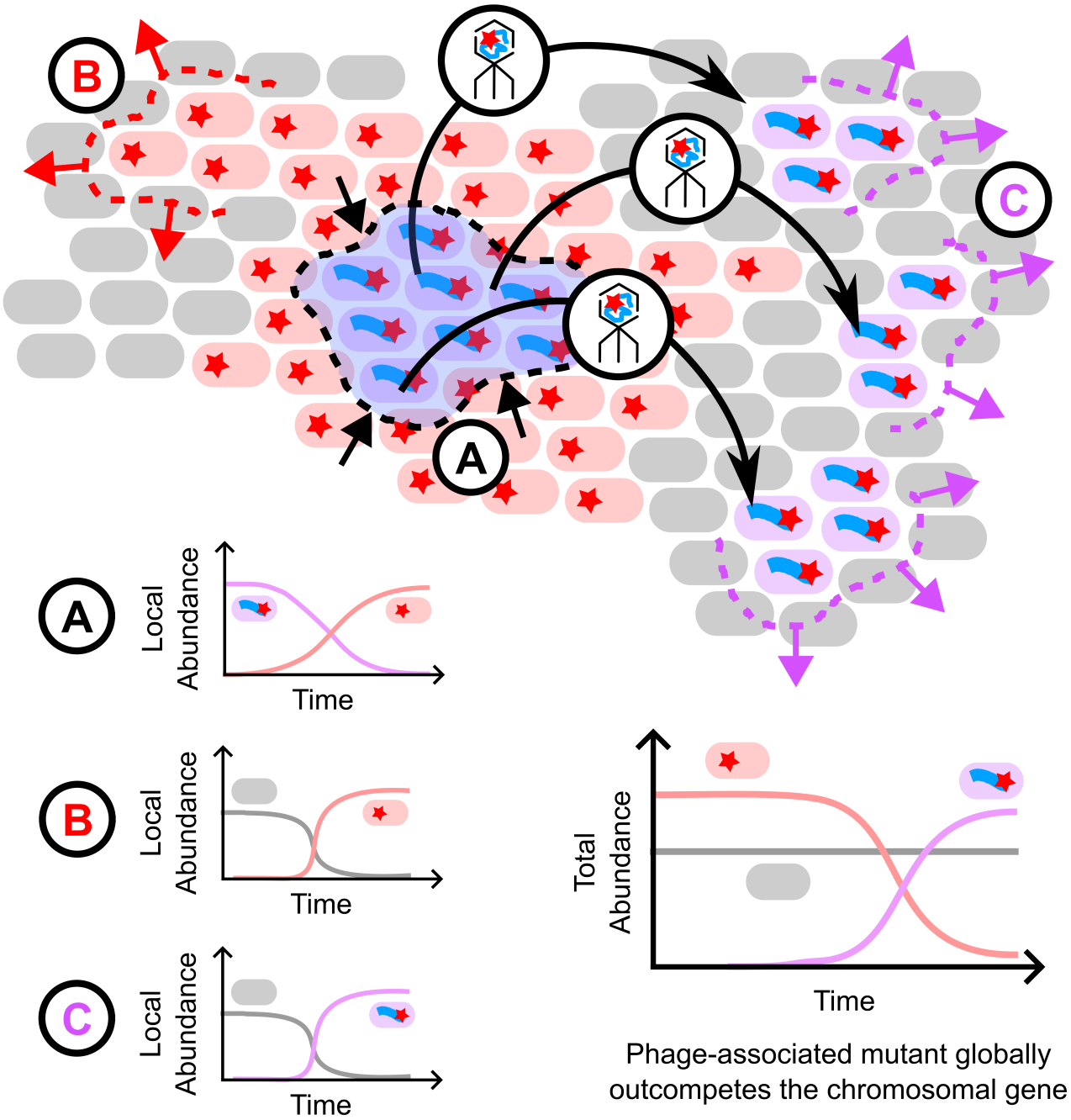


Figure 7: **Dispersal hypothesis illustrated for phage-encoded beneficial traits.** When a chromosomal gene can locally outcompete its phage-associated counterpart, but both types can locally outcompete non-carriers, the phage-associated gene gains a unique dispersal advantage. By diffusing through virion particles, it has the potential to invade new regions dominated by non-carriers. This dispersal allows for bursts of localised growth in non-carrier populations, enabling the phage-associated gene to compete on a larger spatial scale and outcompete the chromosomal gene.

Our results suggest that when prophages are induced and persist only within host environments, genes providing fitness advantages in these environments readily evolve to associate with these prophages. This process is driven by the influx of non-carrier susceptible cells from the external environment. In our analysis, we assumed symmetric cell densities and migration rates between the two environments, but in reality, these factors may vary. Exploring how different metapopulation structures influence this evolutionary process would be an interesting direction for future research. Nonetheless, our findings help explain the empirical tendency of phages to encode virulence factors (Boyd and Brüssow, 2002; Abedon and LeJeune, 2007; Schroven et al., 2021; Takeuchi et al., 2023). More broadly, our model

applies to any weakly selected beneficial trait. Virulence genes fit this criterion due to their known tendency to share the benefits of pathogenicity with surrounding bacteria (Nogueira et al., 2009). The requirement for weak selection also aligns with a scenario where increasing host immunity decreases the selective pressure on virulence genes.

Phages offer distinct advantages over plasmids in associating with virulence genes due to their lifecycle. Unlike plasmids, which rely on contact-dependent conjugation, phages achieve horizontal transfer through virion dispersal, and induction allows phages to synchronise gene transfer with adaptive stress responses. While cross-species transfer has been proposed to facilitate plasmid-associated gene evolution (Lehtinen et al., 2021), temperate phages may be more host-specific due to receptor-mediated infection and site-specific integration (Moura de Sousa et al., 2021; López-Leal et al., 2022). Furthermore, evidence suggests that phage-associated virulence genes originate from chromosomal loci via erroneous excision of prophages (Brüssow et al., 2004). While priority effect has been shown to drive plasmid-gene associations (Lehtinen et al., 2021), our models suggest that localised priority effects in spatially structured populations may instead facilitate the evolution of phage-encoded genes.

Weak selection emerges as a key condition for MGE-borne gene evolution (van Dijk and Hogeweg, 2016; van Dijk et al., 2020; Lehtinen et al., 2021). Strong selection, typically associated with essential genes, favours chromosomal association due to stability and reliability. With weak selection, frequent loss of the gene creates opportunities for lineages of non-carriers to coexist with carriers, making mobility advantageous as it enables the gene to spread independent from their host reproductive cycle. Due to weak selection, our simulations show that fixation of mobile genes occurs over long timescales, amidst significant demographic noise (see Appendix G).

Although we modelled cooperative traits, we introduced privatisation of benefits to maintain a positive selection pressure and prevent the "tragedy of the commons" (Rankin et al., 2007). This allows us to focus on whether a gene preserved by selection prefers an immobile (chromosomal) or mobile (phage-associated) state. Therefore, we distinguish this model from previous work that studied HGT as a mechanism for evolving cooperative traits in well-mixed populations (Nogueira et al., 2009; Mc Ginty et al., 2013). In a spatially structured population, however, cooperation can persist through spatial privatisation – benefits being shared primarily within local neighbourhoods. When we restrict sharing of benefits from carrier cells to a radial neighbourhood, we found that phage-gene association evolved more easily with increasing virion dispersal as well as increasing the range of cooperation (see Appendix H). This is because increasing the radius of sharing benefits increases the chance of non-carrier cells receiving the benefit and lowers the selection coefficient acting on the gene. Therefore, we highlight the generalisability of our dispersal hypothesis to any weakly selected gene in a spatially structured bacterial population.

For illustrative purposes, our models rely on several simplifying assumptions, deviating from the realities of natural systems. One notable assumption is superinfection immunity. In natural bacterial populations, particularly among pathogens, multiple prophages often coexist within a single host cell (López-Leal et al., 2022; Brüssow et al., 2004). These prophages can undergo high order interactions and genetic exchange that reflect in dynamics at higher levels of biological systems (Wendling, 2023; Moura de Sousa et al., 2021). While embedding these interactions in ODE models inflates the number of equations, it will be interesting to relax superinfection immunity in individual-based models in the future. These models may also include other processes we have thus far neglected, such as retention of partial immunity after phage loss (Arias et al., 2022). While these unmodelled complexities are potentially significant, they do not invalidate the core insights gained from our current model.

In summary, our findings highlight spatial structure as a critical mechanism for the evolution of phage-encoded accessory genes. By explicitly modelling phage-specific dynamics, rather than MGEs in the

abstract, we provide a framework to understand why phages are more likely to associate with virulence genes. Our work provides mechanistic insights into the unique evolutionary role of temperate phages in bacterial adaptability, and their role in the persistence of virulence.

References

Abedon, S. T. and LeJeune, J. T. (2007). Why Bacteriophage Encode Exotoxins and other Virulence Factors. *Evolutionary Bioinformatics Online*, 1:97–110.

Arias, C. F., Acosta, F. J., Bertocchini, F., Herrero, M. A., and Fernández-Arias, C. (2022). The coordination of anti-phage immunity mechanisms in bacterial cells. *Nature Communications*, 13(1):7412. Publisher: Nature Publishing Group.

Baltrus, D. A. (2013). Exploring the costs of horizontal gene transfer. *Trends in Ecology & Evolution*, 28(8):489–495. Publisher: Elsevier.

Bergstrom, C. T., Lipsitch, M., and Levin, B. R. (2000). Natural Selection, Infectious Transfer and the Existence Conditions for Bacterial Plasmids. *Genetics*, 155(4):1505–1519.

Bobay, L.-M. and Ochman, H. (2018). Biological species in the viral world. *Proceedings of the National Academy of Sciences*, 115(23):6040–6045. Publisher: Proceedings of the National Academy of Sciences.

Bobay, L.-M., Rocha, E. P., and Touchon, M. (2013). The Adaptation of Temperate Bacteriophages to Their Host Genomes. *Molecular Biology and Evolution*, 30(4):737–751.

Boyd, E. F. and Brüssow, H. (2002). Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. *Trends in Microbiology*, 10(11):521–529.

Brüssow, H., Canchaya, C., and Hardt, W.-D. (2004). Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. *Microbiology and Molecular Biology Reviews*, 68(3):560–602. Publisher: American Society for Microbiology.

Eberhard, W. G. (1989). Why do bacterial plasmids carry some genes and not others? *Plasmid*, 21(3):167–174.

Fortier, L.-C. and Sekulovic, O. (2013). Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence*, 4(5):354–365.

Geoffroy, F. and Uecker, H. (2023). Limits to evolutionary rescue by conjugative plasmids. *Theoretical Population Biology*, 154:102–117.

Gummalla, V. S., Zhang, Y., Liao, Y.-T., and Wu, V. C. H. (2023). The Role of Temperate Phages in Bacterial Pathogenicity. *Microorganisms*, 11(3):541. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

Hall, J. P. J., Harrison, E., and Baltrus, D. A. (2021). Introduction: the secret lives of microbial mobile genetic elements. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1842):20200460. Publisher: Royal Society.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362.

Harrison, E., Guymer, D., Spiers, A. J., Paterson, S., and Brockhurst, M. A. (2015). Parallel Compensatory Evolution Stabilizes Plasmids across the Parasitism-Mutualism Continuum. *Current Biology*, 25(15):2034–2039. Publisher: Elsevier.

Henrot, C. and Petit, M.-A. (2022). Signals triggering prophage induction in the gut microbiota. *Molecular Microbiology*, 118(5):494–502. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/mmi.14983.

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., and Willing, C. (2016). Jupyter notebooks – a publishing format for reproducible computational workflows. In Loizides, F. and Schmidt, B., editors, *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, pages 87 – 90. IOS Press.

Lehtinen, S., Huisman, J. S., and Bonhoeffer, S. (2021). Evolutionary mechanisms that determine which bacterial genes are carried on plasmids. *Evolution Letters*, 5(3):290–301.

López-Leal, G., Camelo-Valera, L. C., Hurtado-Ramírez, J. M., Verleyen, J., Castillo-Ramírez, S., and Reyes-Muñoz, A. (2022). Mining of Thousands of Prokaryotic Genomes Reveals High Abundance of Prophages with a Strictly Narrow Host Range. *mSystems*, 7(4):e00326–22. Publisher: American Society for Microbiology.

Mark Osborn, A. and Böltner, D. (2002). When phage, plasmids, and transposons collide: genomic islands, and conjugative- and mobilizable-transposons as a mosaic continuum. *Plasmid*, 48(3):202–212.

Mc Ginty, S. E., Rankin, D. J., and Brown, S. P. (2011). Horizontal gene transfer and the evolution of bacterial cooperation. *Evolution*, 65(1):21–32.

Mc Ginty, S. , Lehmann, L., Brown, S. P., and Rankin, D. J. (2013). The interplay between relatedness and horizontal gene transfer drives the evolution of plasmid-carried public goods. *Proceedings of the Royal Society B: Biological Sciences*, 280(1761):20130400.

Millan, A. S., Peña-Miller, R., Toll-Riera, M., Halbert, Z. V., McLean, A. R., Cooper, B. S., and MacLean, R. C. (2014). Positive selection and compensatory adaptation interact to stabilize non-transmissible plasmids. *Nature Communications*, 5(1):5208. Publisher: Nature Publishing Group.

Moura de Sousa, J. A., Pfeifer, E., Touchon, M., and Rocha, E. P. C. (2021). Causes and Consequences of Bacteriophage Diversification via Genetic Exchanges across Lifestyles and Bacterial Taxa. *Molecular Biology and Evolution*, 38(6):2497–2512.

Nogueira, T., Rankin, D. J., Touchon, M., Taddei, F., Brown, S. P., and Rocha, E. P. C. (2009). Horizontal Gene Transfer of the Secretome Drives the Evolution of Bacterial Cooperation and Virulence. *Current Biology*, 19(20):1683–1691. Publisher: Elsevier.

Pattenden, T., Eagles, C., and Wahl, L. M. (2021). Host life-history traits influence the distribution of prophages and the genes they carry. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1842):20200465. Publisher: Royal Society.

Rankin, D. J., Bargum, K., and Kokko, H. (2007). The tragedy of the commons in evolutionary biology. *Trends in Ecology & Evolution*, 22(12):643–651. Publisher: Elsevier.

Rankin, D. J., Ginty, S. E. M., Nogueira, T., Touchon, M., Taddei, F., Rocha, E. P. C., and Brown, S. P. (2011a). Bacterial cooperation controlled by mobile elements: kin selection and infectivity are part of the same process. *Heredity*, 107(3):279–281. Publisher: Nature Publishing Group.

Rankin, D. J., Rocha, E. P. C., and Brown, S. P. (2011b). What traits are carried on mobile genetic elements, and why? *Heredity*, 106(1):1–10. Publisher: Nature Publishing Group.

San Millan, A. and MacLean, R. C. (2017). Fitness Costs of Plasmids: a Limit to Plasmid Transmission. *Microbiology Spectrum*, 5(5):10.1128/microbiolspec.mtbp–0016–2017. Publisher: American Society for Microbiology.

Schroven, K., Aertsen, A., and Lavigne, R. (2021). Bacteriophages as drivers of bacterial virulence and their potential for biotechnological exploitation. *FEMS Microbiology Reviews*, 45(1):fuaa041.

Smith, J. (2001). The social evolution of bacterial pathogenesis. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1462):61–69. Publisher: Royal Society.

Takeuchi, N., Hamada-Zhu, S., and Suzuki, H. (2023). Prophages and plasmids can display opposite trends in the types of accessory genes they carry. *Proceedings of the Royal Society B: Biological Sciences*, 290(2001):20231088. Publisher: Royal Society.

Tinsley, C. R., Bille, E., and Nassif, X. (2006). Bacteriophages and pathogenicity: more than just providing a toxin? *Microbes and Infection*, 8(5):1365–1371.

van Dijk, B. (2022). Cacatoo: building, exploring, and sharing spatially structured models of biological systems. *Journal of Open Source Software*, 7(70):3948.

van Dijk, B. and Hogeweg, P. (2016). In silico gene-level evolution explains microbial population diversity through differential gene mobility. *Genome biology and evolution*, 8(1):176–188.

van Dijk, B., Hogeweg, P., Doekes, H. M., and Takeuchi, N. (2020). Slightly beneficial genes are retained by bacteria evolving DNA uptake despite selfish elements. *eLife*, 9:e56801. Publisher: eLife Sciences Publications, Ltd.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.

Wagner, P. L., Livny, J., Neely, M. N., Acheson, D. W. K., Friedman, D. I., and Waldor, M. K. (2002). Bacteriophage control of Shiga toxin 1 production and release by Escherichia coli. *Molecular Microbiology*, 44(4):957–970. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1365-2958.2002.02950.x.

Weisberg, A. J. and Chang, J. H. (2023). Mobile Genetic Element Flexibility as an Underlying Principle to Bacterial Evolution. *Annual Review of Microbiology*, 77(Volume 77, 2023):603–624. Publisher: Annual Reviews.

Wendling, C. C. (2023). Prophage mediated control of higher order interactions - Insights from multi-level approaches. *Current Opinion in Systems Biology*, 35:100469.

Wes McKinney (2010). Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61.

Zeaki, N., Budi Susilo, Y., Pregiel, A., Rådström, P., and Schelin, J. (2015). Prophage-Encoded Staphylococcal Enterotoxin A: Regulation of Production in Staphylococcus aureus Strains Representing Different Sea Regions. *Toxins*, 7(12):5359–5376.

A Ordinary Differential Equations

The described model can be expressed in the form of differential equations. We use the notations used for the cell genotypes as non-negative real-valued variables to measure their corresponding abundances in the population. Therefore, the total cell population size (N) and the fractional abundance of carriers (f_τ) can be expressed as

$$
\begin{aligned}
N &= U + U_c + L + L_c + L_p + L_{cp} \\
f_\tau &= (U_c + L_c + L_p + L_{cp})/N
\end{aligned}
$$

Similarly, we use V and V_τ to denote the abundances of non-carrier and carrier virion particles. Hence, we describe the temporal dynamics of the cell and phage abundances uding the following system of differential equation:

$$
\begin{aligned}
\frac{dU}{dt} &= U\left(1 - d - N + (1-\pi)\phi f_\tau\right) && - \beta U(V + V_\tau) && + \lambda(L + L_p) + \mu U_c && \\
\frac{dU_c}{dt} &= U_c\left(1 - d - N + \pi\phi + (1-\pi)\phi f_\tau\right) && - \beta U_c(V + V_\tau) && + \lambda(L_c + L_{cp}) - \mu U_c && \\
\frac{dL}{dt} &= L\left(1 - d - N + (1-\pi)\phi f_\tau\right) && + \beta UV && - \lambda L + \mu(L_c + L_p) && - \alpha L \\
\frac{dL_c}{dt} &= L_c\left(1 - d - N + \pi\phi + (1-\pi)\phi f_\tau\right) && + \beta U_c V && - (\lambda + \mu)L_c + \mu L_{cp} && - \alpha L_c \\
\frac{dL_p}{dt} &= L_p\left(1 - d - N + \pi\phi + (1-\pi)\phi f_\tau\right) && + \beta U V_\tau && - (\lambda + \mu)L_p + \mu L_{cp} && - \alpha L_p \\
\frac{dL_{cp}}{dt} &= \underbrace{L_{cp}\left(1 - d - N + \pi\phi + (1-\pi)\phi f_\tau\right)}_{\text{Growth}} && + \underbrace{\beta U_c V_\tau}_{\text{Infection}} && - \underbrace{(\lambda + 2\mu)L_{cp}}_{\text{Mutations}} && - \underbrace{\alpha L_{cp}}_{\text{Induction}}
\end{aligned}
$$

$$
\begin{aligned}
\frac{dV}{dt} &= r_V \alpha(L + L_c) && - \beta(U + U_c)V && - \delta V \\
\frac{dV_\tau}{dt} &= \underbrace{r_V \alpha(L_p + L_{cp})}_{\text{Lytic release}} && - \underbrace{\beta(U + U_c)V_\tau}_{\text{Infection}} && - \underbrace{\delta V_\tau}_{\text{Decay}}
\end{aligned}
$$

B Model parameters

Variable	Meaning	Value (ODE/IBM)
U	Abundance of non-carrier susceptible cells	$[0, \infty)$ / $\{0, 1, 2, 3...\}$
U_c	Abundance of suceptible cells carrying gene on the chromosomal locus	$[0, \infty)$ / $\{0, 1, 2, 3...\}$
L	Abundance of non-carrier lysogens	$[0, \infty)$ / $\{0, 1, 2, 3...\}$
L_c	Abundance of lysogens carrying the gene on the chromosomal locus	$[0, \infty)$ / $\{0, 1, 2, 3...\}$
L_p	Abundance of lysogens carrying the gene on the phae locus	$[0, \infty)$ / $\{0, 1, 2, 3...\}$
L_{cp}	Abundance of lysogens carrying the gene on both loci	$[0, \infty)$ / $\{0, 1, 2, 3...\}$
V	Abundance of extracellular non-carrier virions	$[0, \infty)$ / $\{0, 1, 2, 3...\}$
V_τ	Abundance of extracellular carrier virions	$[0, \infty)$ / $\{0, 1, 2, 3...\}$
Parameter	**Meaning**	**Value (ODE/IBM)**
ϕ	Ratio of maximum increase in carrying capacity to minimum carrying capacity	1 / 1
b	Intrisic birth rate	1 / 0.5
d	Intrinsic death rate	0.02 / 5×10^{-3}
π	Fraction of benefits privatised by carriers of gene	$[10^{-5}, 1]$ / $[10^{-5}, 1]$
β	Rate of phage infection	0.02 / 5×10^{-3}
α	Rate of prophage induction and lysis	$[10^{-5}, 10^{-1}]$ / 5×10^{-4}
r_V	Burst size	30 / 30
δ	Rate of virion decay	0.02 / 5×10^{-3}
μ	Rate of gene loss	2×10^{-3} / 5×10^{-4}
λ	Rate of phage loss	2×10^{-3} / 5×10^{-4}
V_{diff}	Rate of virion diffusion	$\{10^{-3}, 10^{-1}, \infty\}$ (IBM)

Table 1: Variables and parameters of the model, their meaning and the values used. The variables are non-dimensionalised in the ODEs and the rate parameters are scaled when switching between the ODE and IBM. The details of these scales are given in Appendix C

C Well-mixed approximation of the spatial model

Let x be the number of cells on the grid of a certain genotype and k be the minimum carrying capacity, i.e. the grid size is $k(1+\phi)$. Then, the probability of finding an empty grid point is

$$\mathbb{P}(\text{empty grid point}) = 1 - \frac{n}{k(1+\phi)}$$

The average death rate of a genotype will be

$$\frac{\sum_{\text{genotype}} \mathbb{P}_{\text{death}}(\pi)}{\sum_{\text{genotype}} 1} = \begin{cases} d_s + \Delta d\Big(1 - \pi - (1-\pi)f_\tau\Big), & \text{genotype is a carrier} \\ d_s + \Delta d\Big(1 - (1-\pi)f_\tau\Big), & \text{genotype is not a carrier} \end{cases}$$

The the well-mixed growth dynamics of a carrier genotype is given by

$$\begin{aligned} \frac{dx}{dt'} &= \text{Avg. \#births into empty grid points} - \text{Avg. \# deaths} \\ \frac{dx}{dt'} &= bx\Big(1 - \frac{n}{k(1+\phi)}\Big) - x\Big(d + \Delta d\Big(1 - \pi - (1-\pi)f_\tau\Big)\Big) \\ \frac{dx}{dt'} &= \frac{b}{1+\phi}x\Big(1 + \phi - \frac{n}{k} - d_s\frac{1+\phi}{b} - \Delta d\frac{1+\phi}{b}\Big(1 - \pi - (1-\pi)f_\tau\Big)\Big) \end{aligned}$$

We apply the following non-dimensionalisation

$$X = \frac{x}{k}, \quad N = \frac{n}{k}, \quad t = t'\frac{b}{1+\phi}$$

and the following substitutions

$$d_s = d\frac{b}{1+\phi}, \quad \Delta d = \frac{b\phi}{1+\phi}$$

to obtain

$$\frac{dX}{dt} = X(1 - N + \pi\phi + (1-\pi)\phi f_\tau)$$

which is the same growth dynamics we implemented in the ODEs. The same derivation can be done for the growth dynamics of a non-carrier genotype. Notice that time is scaled by a factor of $\frac{b}{1+\phi}$ in the ODE, i.e. one time-step in the IBM is equal to $\frac{1+\phi}{b}$ (which is equal to 4 for our choice of parameters) many time-steps of the ODE. Therefore, all the rate parameters, just like the intrinsic death rate, need to be adjusted to the time scale by this factor when switching between the models.

D Selection- and transmission-mutation thresholds

To derive the persistence condition of the gene in the face of mutation, we can consider a simpler version of the model consisting only of two kinds of cells: carrier (U) and non-carrier (U_c).

$$\frac{dU}{dt} = U\left(1 - U - U_c + (1-\pi)\phi\frac{U_c}{U+U_c}\right) + \mu U_c$$
$$\frac{dU_c}{dt} = U_c\left(1 - U - U_c + \pi\phi + (1-\pi)\phi\frac{U_c}{U+U_c}\right) - \mu U_c$$

We obtain the condition that for a positive abundance of the carrier type at steady-state ($U_c(t=\infty) > 0$) requires the rate of selection to exceed the rate of mutation:

$$\pi\phi > \mu$$

Proof. Consider the non-dimensionalised version of the model such that $u = U/N$ and $u_c = U_c/N$, where $N = U + U_c$.
The equations then become

$$\frac{du}{dt} = -\pi\phi u u_c + \mu u_c$$
$$\frac{du_c}{dt} = \pi\phi u u_c - \mu u_c$$

Solving for steady state, the non-trivial solutions are $u^\infty = \frac{\mu}{\pi\phi}, u_c^\infty = 1 - \frac{\mu}{\pi\phi}$, and $u^\infty = 1, u_c^\infty = 0$.

We can infer that a stable positive solution for u_c^∞ exists only when $\pi\phi > \mu$. □

Now to derive the persistence condition of the MGE, let us look at a simpler version of the model consisting only of two kinds of bacteria: uninfected (U) and infected (L). First, let us consider a plasmid.

$$\frac{dU}{dt} = U(1 - U - L) - \beta U L + \lambda L$$
$$\frac{dL}{dt} = L(1 - c - U - L) + \beta U L - \lambda L$$

We obtain the condition that the plasmid will persist in the population ($L(t=\infty) > 0$) if the conjugation rate is higher than the sum of rate of segregational loss and cost of carrying the plasmid:

$$\beta > \lambda + c$$

Proof. The plasmid will persist if it has a positive growth rate when it is rare in the population, i.e. the susceptible cells are maximally abundant ($U = 1$) and the infected cell abundance is close to zero ($L^2 \simeq 0$). Then,

$$\frac{dL}{dt} > 0 \implies \beta - \lambda - c > 0$$

□

When the MGE is a phage, we will consider an eqaution describing the virion abundance (V) as well.

$$\frac{dU}{dt} = U(1 - U - L) - \beta U V + \lambda L$$
$$\frac{dL}{dt} = L(1 - U - L) + \beta U V - \lambda L - \alpha L$$
$$\frac{dV}{dt} = r_V \alpha L - \beta U V - \delta V$$

We obtain the condition that the phage will persist in the population ($L(t = \infty) > 0$) if the lysis rate is above a certain threhsold:

$$\alpha > \frac{\lambda(\beta + \delta)}{\beta(r_V - 1) - \delta}$$

Proof. By explicitly considering the extracellular virions, we have modelled an infection cycle with two stages: the prophage stage and the virion stage. For a two-stage infection cycle, the net basic reproductive number of the infection, R_0 is given by the product of R_0 of each step. Therefore, the phage will presist if the net R_0 is greater than 1.

$$\begin{aligned} R_0 &= R_0(\text{prophage}) \times R_0(\text{virion}) \\ &= \frac{\text{max. infection rate}}{\text{phage loss rate} + \text{lysis rate}} \times \frac{\text{burst size} \times \text{lysis rate}}{\text{max. infection rate} + \text{decay rate}} \\ &= \frac{r_V \alpha \beta}{(\lambda + \alpha)(\beta + \delta)} \end{aligned}$$

Therefore,

$$R_0 > 1 \implies \alpha > \frac{\lambda(\beta + \delta)}{\beta(r_V - 1) - \delta}$$

□

E Persistence of the gene and the phage

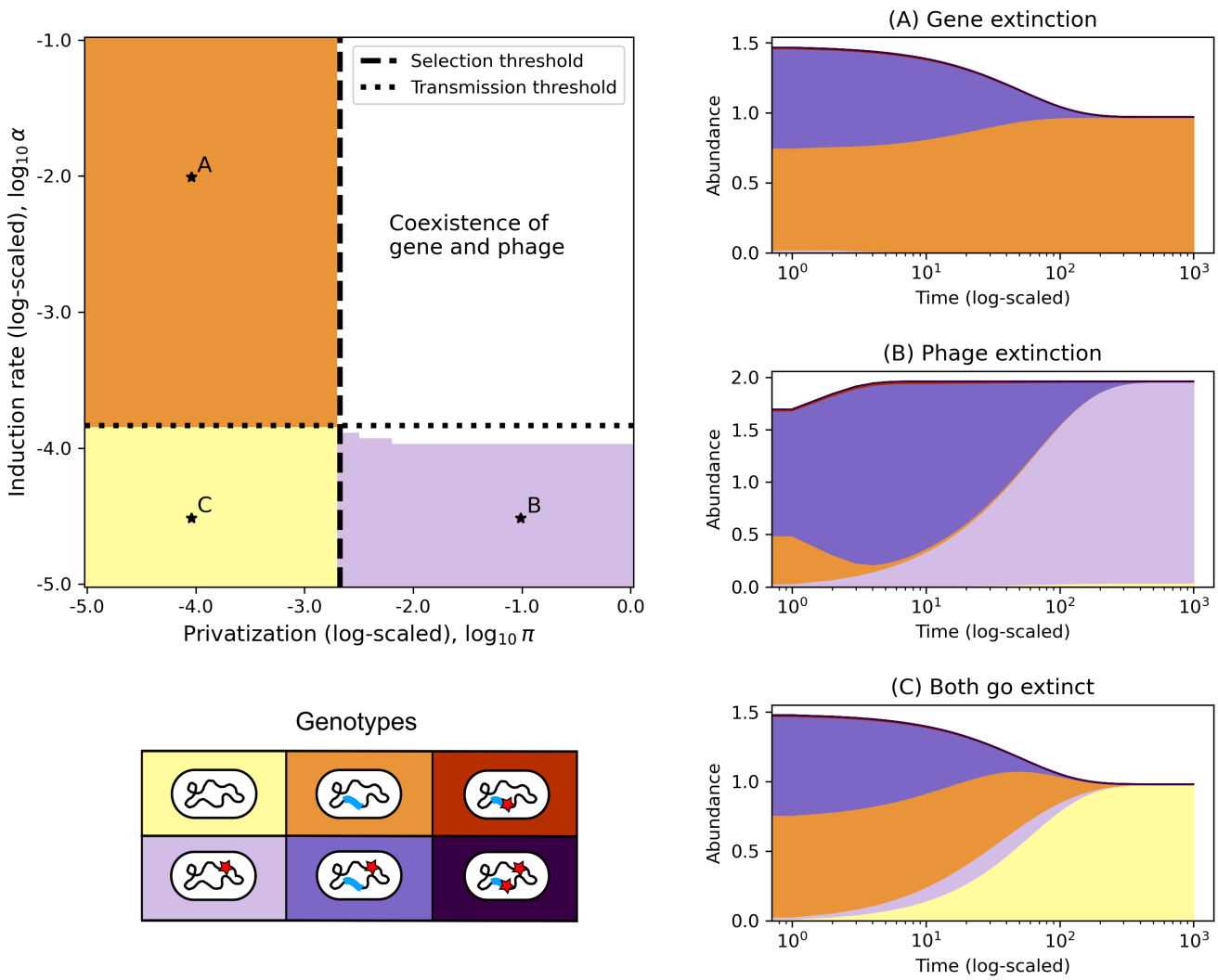


Figure 8: **Persistence conditions for the gene and the phage:** The parameter space can be roughly partitioned into four quadrants by the analytically derived thresholds (denoted by the black lines). The dashed line shows the π threshold for gene persistence and the dotted line shoes the α threshold for phage persistence. (A) This quadrant is where L genotype dominates because R_0 of the phage is greater than one but the selection is too weak to maintain the gene on either loci. (B) This quadrant is where U_c genotype dominates because R_0 of the phage is lesser than one but the selection strong enough to maintain the gene at the chromosomal locus. (C) This quadrant is where U genotype dominates because both infection and selection rates are too weak and consequently both the phage and the gene are lost. Initial condition: $L = 0.495, L_c = 0.500, L_p = 0.005, \text{others} = 0.000$.

F Modified ODEs

F.1 ODEs incorporating coregulation of prophage induction and gene expression

- Gene expression is required for phage induction.

  We remove the lysis term from the ODE of L genotype and therefore, we also remove the lytic release of phages from L in the PDE of V virions.

$$\frac{dL}{dt} = L\left(1 - d - N + (1-\pi)\phi f_\tau\right) + \beta UV - \lambda L + \mu(L_c + L_p)$$
$$\frac{dV}{dt} = r_V \alpha L_c - \beta(U + U_c)V - \delta V$$

- Prophage induction is required for gene expression.

  We remove the privatised growth term from the ODE of U_c genotype and also remove the genotype when computing the fraction of carriers in the population.

$$\frac{dU_c}{dt} = U_c\left(1 - d - N + (1-\pi)\phi f_\tau\right) - \beta U_c(V + V_\tau) + \lambda(L_c + L_{cp}) - \mu U_c$$
$$f_\tau = (L_c + L_p + L_{cp})/N$$

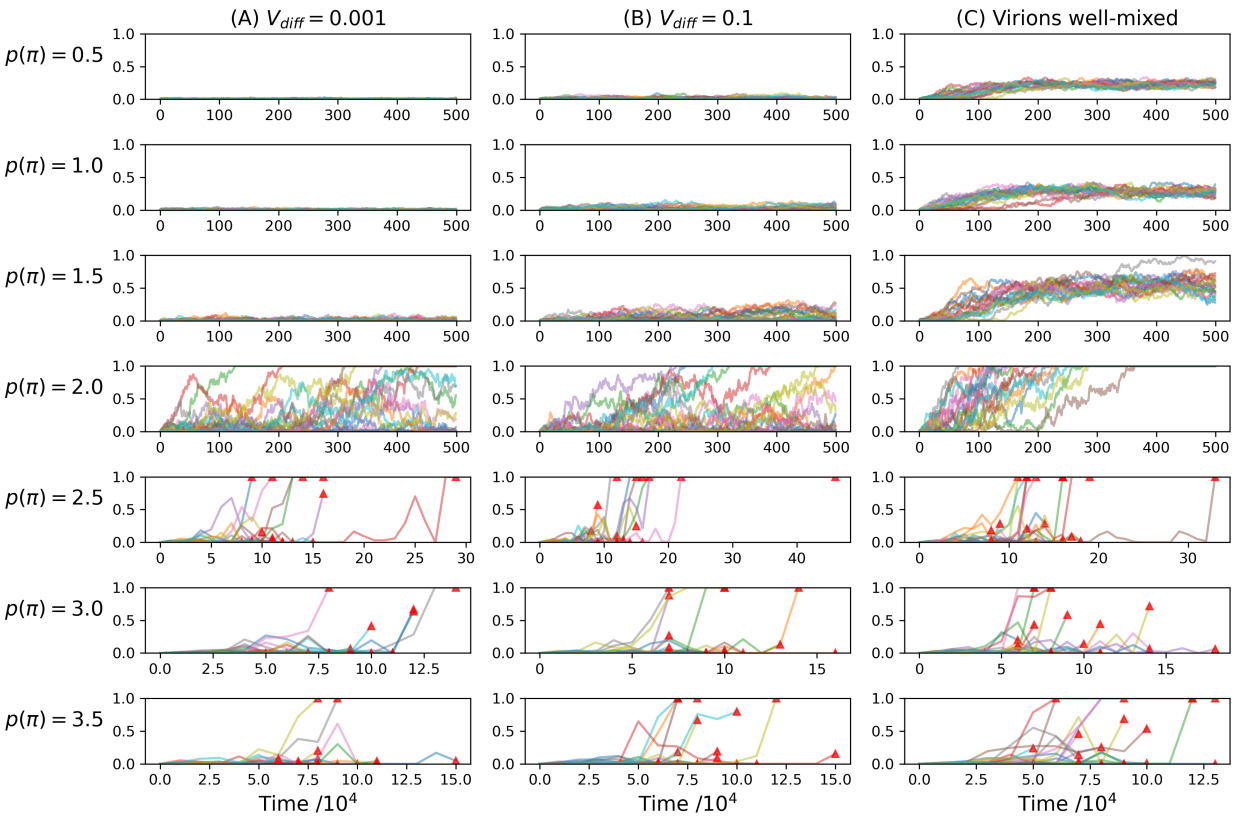
F.2 ODEs of the "outside world" compartment

Let m be the rate of migration between the 'host' and the 'outside world' compartments. Then the ODEs describing the dynamics in the outside world (variables denoted with a superscript o are

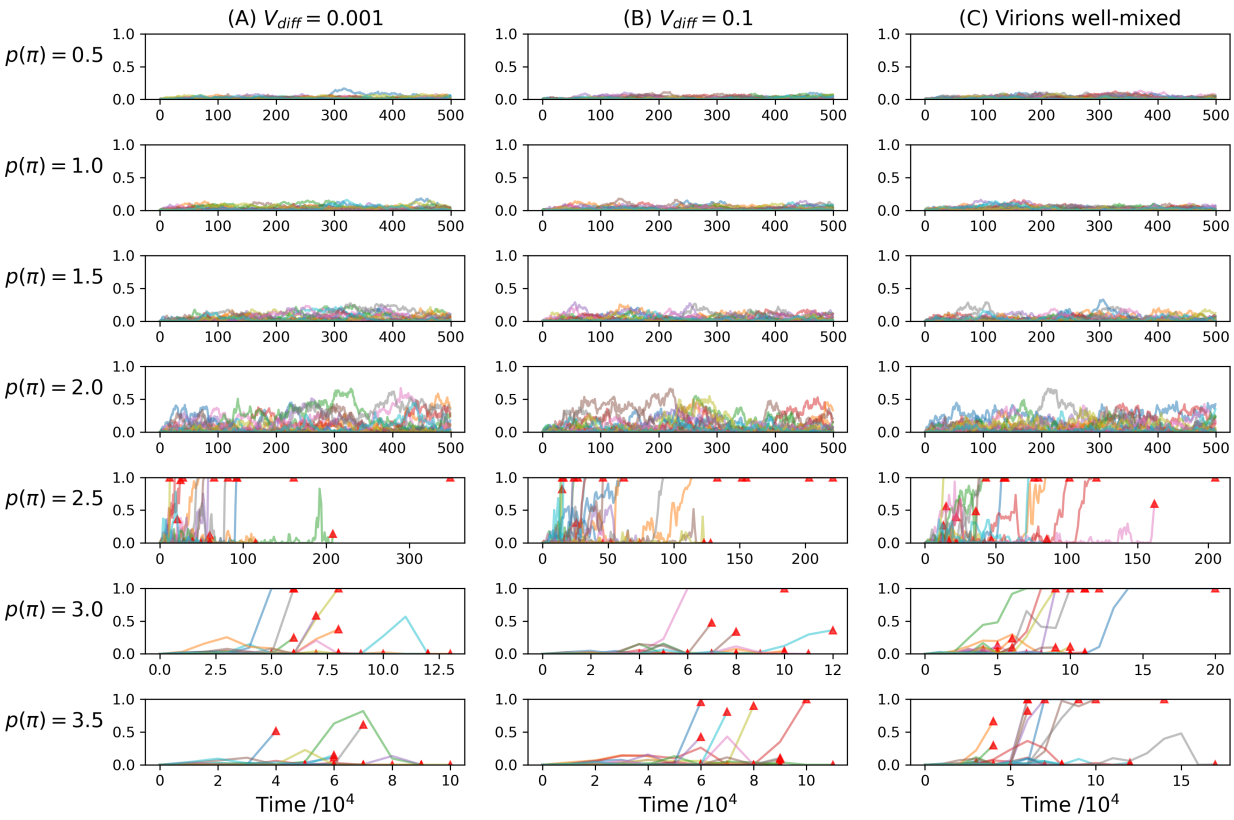
$$\frac{dU^o}{dt} = U^o(1 + \phi - d - N^o) - \beta U^o(V^o + V^o_\tau) + \lambda(L^o + L^o_p) + \mu U^o_c + m(U - U^o)$$
$$\frac{dU^o_c}{dt} = U^o_c(1 + \phi - d - N^o) - \beta U^o_c(V^o + V^o_\tau) + \lambda(L^o_c + L^o_{cp}) - \mu U^o_c + m(U_c - U^o_c)$$
$$\frac{dL^o}{dt} = L^o(1 + \phi - d - N^o) + \beta U^o V^o - \lambda L^o + \mu(L^o_c + L^o_p) + m(L - L^o)$$
$$\frac{dL^o_c}{dt} = L^o_c(1 + \phi - d - N^o) + \beta U^o_c V^o - (\lambda + \mu)L^o_c + \mu L^o_{cp} + m(L_c - L^o_c)$$
$$\frac{dL^o_p}{dt} = L^o_p(1 + \phi - d - N^o) + \beta U^o V^o_\tau - (\lambda + \mu)L^o_p + \mu L^o_{cp} + m(L_p - L^o_p)$$
$$\frac{dL^o_{cp}}{dt} = L^o_{cp}(1 + \phi - d - N^o) + \beta U^o_c V^o_\tau - (\lambda + 2\mu)L^o_p + m(L_{cp} - L^o_{cp})$$
$$\frac{dV^o}{dt} = -\beta(U^o + U^o_c)V^o - \delta V^o$$
$$\frac{dV^o_\tau}{dt} = -\beta(U^o + U^o_c)V^o_\tau - \delta V^o_\tau$$

where $N^o = U^o + U^o_c + L^o + L^o_c + L^o_p + L^o_{cp}$. The only modification to the system of ODEs for the host comparmtent is the addition of a migration term $+m(X^o - X)$ to the ODE for the corresponding X genotype abundance variable.

G Timeseries evolution prophage-association in space



(a) Cells are stationary on the grid



(b) Cells are well-mixed on the grid

Figure 9: Timeseries plots of the fraction of phage-associated genes for 20 replicates per value of privatisation (rows, $p(\pi) = -\log_{10}\pi$) and virion diffusion rate (columns). The red triangles mark points in the simulation runs where the gene went extinct.

H Spatial privatization of benefits

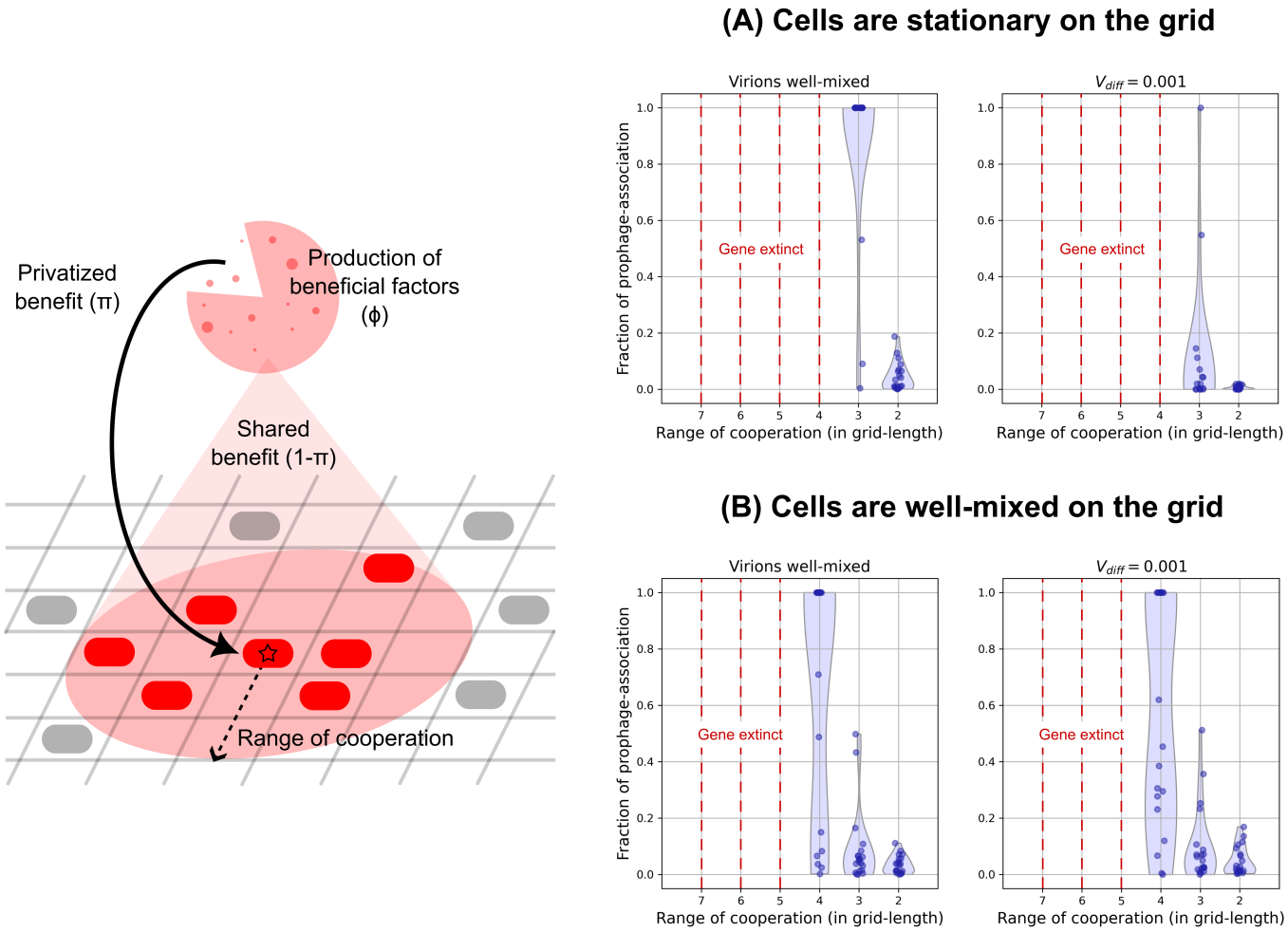


Figure 10: (Left) In a spatial model, privatization can be modelled mechanistically through spatial privatization, i.e. the benefits of expression of a cooperative trait is only experienced by cells in an immediate neighbourhood. Therefore, we introduce a parameter "Range of cooperation" which measures the maximum distance of a cell from a carrier cell in order to obtain benefits from the latter. Therefore, decreasing this radius increases the privatization. Furthermore, since a cell is likely surrounded by similar genotype, decreasing the radius also increases the selection pressure acting on the gene. (Right) With the mechanistic implementation of privatization, we yet again observe that the chance of prophage-association increases with increasing virion dispersal and decreasing selection coefficient (increasing range of cooperation). In comparison with the case where the cells are well-mixed, we can clearly see that spatial structure is required for this trend. We also see that gene extinction happens sooner with spatial structure and that the effect is stochasticity is much stronger. This is because at the local level, the carrier genotypes have the same growth rate as the non-carrier (cheater) genotypes and hence evolves towards gene extinction. Selection in this model occurs at a higher (group-) level. This sufficiently reduces the effective population size on which selection is acting - leading to increased genetic drift. To increase the effective population size, we lowered the parameter ϕ to 0.1 and adjusted other parameter values accordingly.