

In Silico Gene-Level Evolution Explains Microbial Population Diversity through Differential Gene Mobility

Bram van Dijk^{1,*} and Paulien Hogeweg¹

¹Department of Theoretical Biology and Bioinformatics, Utrecht University, The Netherlands

*Corresponding author: E-mail: b.vandijk@uu.nl.

Accepted: December 15, 2015

Abstract

Microbial communities can show astonishing ecological and phylogenetic diversity. What is the role of pervasive horizontal gene transfer (HGT) in shaping this diversity in the presence of clonally expanding “killer strains”? Does HGT of antibiotic production and resistance genes erase phylogenetic structure? To answer these questions, we study a spatial eco-evolutionary model of prokaryotes, inspired by recent findings on antagonistic interactions in *Vibrionaceae* populations. We find toxin genes evolve to be highly mobile, whereas resistance genes minimize mobility. This differential gene mobility is a requirement to maintain a diverse and dynamic ecosystem. The resistance gene repertoire acts as a core genome that corresponds to the phylogeny of cells, whereas toxin genes do not follow this phylogeny and have a patchy distribution. We also show that interstrain HGT makes the emergent phylogenetic structure robust to selective sweeps. Finally, in this evolved ecosystem we observe antagonistic interactions between, rather than within, spatially structure subpopulations, as has been previously observed for prokaryotes in soils and oceans. In contrast to ascribing the diversification and evolution of microbial communities to clonal dynamics, we show that multilevel evolution can elegantly explain the observed phylogenetic structure and ecosystem diversity.

Key words: horizontal gene transfer, HGT, microbial evolution, gene mobility, antagonistic interactions, toxin, resistance, diversity, core genome, accessory genome, flexible genomes, multilevel evolution, eco-evolutionary dynamics, gene-level selection.

Introduction

Recent metagenomics studies have led to the realization that the commonly observed phenotypic diversity within microbial ecosystems is just the tip of the iceberg. High-throughput sequencing has revealed an immense variation at the level of DNA sequences (Ochman et al. 2000) and ecological adaptation (Preheim et al. 2011), even in narrow taxonomic groups. How this diversity comes about, and moreover how it is maintained, is still an open question.

An important source of diversity in prokaryotes is the remarkable flexibility of their genome content (Kolstø 1997; Snel et al. 2002; Koonin and Wolf 2008). This flexibility is not equal for all parts of the genome, as some genes are more dispensable than others. The well-conserved core genes generally encode basal cellular functions, and are complemented with a variable and more niche-specific set of accessory genes. The core genome tends to have a phylogeny that is consistent with that of the species. On the other hand, the accessory genome has a phylogeny that differs from this vertical inheritance pattern (e.g., Shapiro et al. 2012), showing

that horizontal gene transfer (HGT) is an important process in shaping the flexibility and diversity of prokaryotes (Fraser et al. 2009). Indeed, recent results have shown that the rate at which HGT happens is up to 2 orders of magnitude higher than per-gene point mutations (Puigbò et al. 2014). These results indicate that the gene content of microbes can change in relatively short timescales.

Populations of bacteria do not only display high diversity in gene repertoires but also at the phylogenetic level, as was recently shown for *Vibrionaceae* (Preheim et al. 2011) and *Streptomyces* sp. (Vetsigian et al. 2011). Strikingly, this phylogenetic diversity is maintained despite the presence of “superkiller” strains (Vetsigian et al. 2011; Cordero et al. 2012). Vetsigian et al. (2011) explained this superkiller paradox in a model assuming high rates of de novo gene gain (e.g., HGT from a distant source), which creates an evolutionary arms race. However, the resulting clonal diversity and short evolutionary life span of lineages appear to fall short in explaining the observed phylogenetic diversity in wild

populations. Surprisingly, the studies also show that killing happens less frequently between related individuals, that is, close kin are spared from killing. The frequency of killing is low up to a certain genetic threshold, after which it sharply increases. This observation once again suggests high levels of phylogenetic structure. The combination of high phenotypic diversity and high phylogenetic diversity suggests that adaptive genes sweep the population on their own, rather than resulting in genome-wide, clonal sweeps (Shapiro et al. 2012). Furthermore, as de novo gene discovery fails to reproduce these patterns, internal HGT of locally present DNA is a likely candidate for the flexibility and diversity of prokaryotic populations.

In *Vibrio ordalii*, the toxin production gene is part of a mobile genetic element (MGE) (Cordero et al. 2012). Surprisingly, this toxin-production gene was not linked to the corresponding resistance factor, and can thus only transfer to individuals that already have the resistance factor in their genetic background. The resistance of *V. ordalii* was indeed shown to be ancestral, and other resistant-only individuals show no trace of the MGE. These results indicate that the resistance gene is part of the core genome, whereas the toxin gene is a mobile accessory gene which is frequently transferred within and potentially between populations of Vibrionaceae. This genome structure appears to prevent clonal dynamics, and impacts the observed diversity and interaction dynamics by means of gene-level selection.

We here ask the question how local HGT impacts population diversity and the differential gene mobility as observed for the toxin and resistance genes in *Vibrio*'s. Furthermore, we wish to understand what selection pressures underlie this process, and how it shapes core-, accessory-, and pan-genomes.

We investigate these questions by using a minimal model of microbial evolution, where cells undergo HGT of locally present DNA. The model includes three different levels: Genes, cells, and spatial patterns. Cells compete locally for space by means of antagonistic interactions of toxin and resistance genes. Cells stochastically take up genes from the local extracellular DNA (eDNA) pool. Genes encode their own mobility parameter, which represents the likelihood of integration into the genome after uptake. Antagonistic interactions take place between spatially patterned subpopulations. We focus on the evolution of gene mobility, and how this influences diversity and population structure. The aim is to reveal the intrinsic evolutionary pressure of gene mobility, abstracting away from the mechanisms underlying this process (e.g., uptake sequences [Mell and Redfield 2014] or illegitimate recombination [de Vries and Wackernagel 2002]). Our minimal model shows how HGT and the resulting interplay of gene- and cell-level selection can explain the observed ecosystem diversity and the different phylogenetic patterns of core and accessory genomes of natural prokaryotic populations.

Materials and Methods

General Description

We model a simple in silico microbial population that is subject to local HGT of toxin and resistance genes (for an overview, see fig. 1). A spatial grid is implemented on which a single species is competing for reproduction space through local interactions by means of 20 sets of predefined toxin- and resistance genes. We assume that the costs for producing toxins or resistance factors are equal, and focus our observations on the evolution of gene mobility. The gene mobility is a gene-

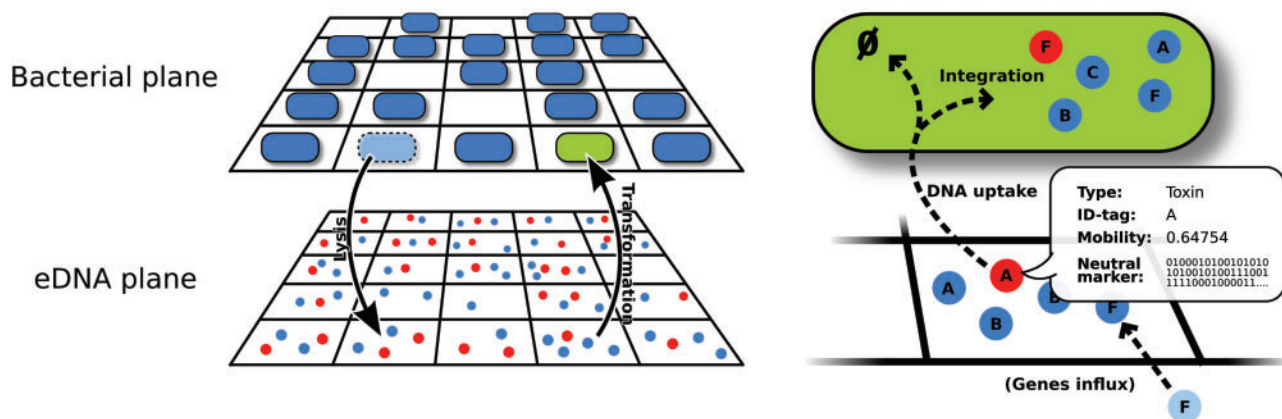


FIG. 1.—Graphical representation of the modelled microbial ecosystem. Microbe-like individuals compete on a grid, with fitness-proportional reproduction (R_i and B_i , eqs. 1 and 2) and stochastic death (D_i , eq. 3). Fitness is decreased for every gene in the genome of an individual linearly (parameter c). D_i is increased for each toxin produced in the Moore (8) neighborhood for which the individual does not produce resistance factors (parameter t). Dying individuals leave their genes behind (blue and red circles for resistance and toxin genes, respectively), creating an eDNA pool that is diffused using Toffoli–Margolus Diffusion (Toffoli and Margolus 1987). Genes can then be taken up with a fixed probability (u), and get integrated into the genome with a probability that is evolved per gene (m).

specific, evolvable feature which determines the chance of successful integration after uptake (i.e., abstracting away from mechanisms such as flanking sequence homology, MGEs, etc.). The genes available for uptake originate from dying individuals, leaving their genetic information behind. This pool of eDNA undergoes diffusion and decay, whereas individuals in the proximity of this genetic material can stochastically take up genes.

Population Dynamics

Our model is a stochastic cellular automaton, where grid points represent reproduction space. They are either empty or contain one individual bacterium. Every time step, all grid points are synchronously updated based on their own state, and the state of adjacent grid points in the Moore Neighborhood (eight adjacent sites). If the grid point is empty, local neighbors can compete for reproduction (R_i) into the empty spot with chances that are dependent on their relative birth rate (B_i) with respect to the birth rate of the other competitors (see eqs. 1 and 2). B_i is linearly decreasing for the total number of toxin- and resistance genes produced by that individual (G_{t+r}), with a fixed per-gene cost (c). It is also possible that none of the individuals reproduces, for example, when there are very few or unfit individuals surrounding the empty grid point (this chance is scaled by parameter e in eq. 2).

Individual bacteria die with a constant probability (d), increased by the toxicity parameter (t) for every direct neighbor that is capable of killing the focal cell (eq. 3). Note that killing is self-inclusive, meaning that individuals need both the toxin gene and the resistance gene to survive.

$$B_i = \max\left(0, 1 - c \cdot \sum_{j=1}^{n_{\text{genes}}} (G_j)\right), \quad (1)$$

$$R_i = \frac{B_i}{\sum_j B_j + e}, \quad (2)$$

$$D_i = \min\left(1, d + t \cdot \sum_{j=1}^{n_{\text{neigh}}} \sum_{k=1}^{n_{\text{genes}}} (G_{i_{k(\text{res})}} == 0 \ \&\& \ G_{j_{k(\text{tox})}} > 0)\right). \quad (3)$$

Gene-Level Dynamics

The genomes of bacteria are implemented as simple collections of toxin- and resistance genes, where the order of genes is not taken into account (i.e., excluding the possibility of physical coupling of genes). In this model, 20 different pairs of toxin- and resistance genes are taken into account (a total of 40 unique genes). The genes have an identity-tag (letters A–T in fig. 1), a mobility (m), and a neutral bit-string marker used for phylogenetic

reconstruction of gene trees (see details on phylogenetic reconstruction below). Upon reproduction, an individual copies its collection of genes allowing for mutations on both the mobility parameter (m), and bit flip mutations on the bit-string marker. There is also a fixed per-gene probability of loss (l). De novo gene discovery and gene duplication do not happen as a result of replicating the genome for reproduction. However, gene duplications and gene discovery can both be the result of the simplified form of HGT in the model.

HGT, and the resulting gene-level dynamics, is included by modelling a simple process of uptake of eDNA (as shown in fig. 1). Individuals that die leave their genes behind, which diffuse and decay with a fixed rate. Genes that reside on the same spot as any bacterium have a fixed chance (u) to be taken up by the bacterium. Subsequently, the gene can be integrated into the genome of the individual with a chance $P(m)$ which is an evolvable property of that toxin or resistance gene. If that cell already had a copy of that gene, it pays the cost without getting any ecological benefit. Thus, in this model, having multiple copies of a gene is always detrimental to the direct fitness, and can only be advantageous in an evolutionary context (e.g., redundancy when one of the copies is lost). If the gene is not integrated into the genome after uptake, it is degraded ($P(1 - m) \rightarrow \emptyset$).

For example, if toxin A in figure 1 (colored red) gets taken up by the green cell, it gets incorporated into the genome with a chance (in the figure indicated as ~65%), which is unique for that instance of gene A. Because of mutations, many other instances of toxin gene A are present with different values for m , resulting in a process of natural selection on the gene-level. Note that in this example, the cell indeed has the resistance gene for compound A, meaning it could give a selective advantage to the cell if it would get integrated into the genome, indirectly also benefitting the toxin gene.

Finally, every time step a fixed number of random toxin and resistance genes flux into the eDNA-pool (f). These genes flux in at a random position. They can have a novel and unique identity, but the gene mobility is sampled from the existing ecosystem, under the assumption that there is influx from a relatively distant, but otherwise similar ecosystem. This parameter is used to study the effect of closed ecosystems, where the majority of genes comes from the population itself, and gradually more open ecosystems.

Visualizing Fitness Landscapes

Fitness landscapes are generated to understand when the loss/gain of a gene is beneficial and gives an immediate benefit for a cell. Here we define fitness as $(B_i - D_i)$ and calculate the average difference in fitness with respect to the eight adjacent neighbors, before and after the mutation. The difference between before and after (E_i) gives the immediate benefit gain/loss of a gene would give the cell (eq. 4). Note that long-term,

indirect effects (e.g., toxin gain enabling to locally overgrow more than the direct neighbors) are not included in this measure. We also measure E_i with eight randomly chosen individuals, which helps to understand the importance of local interactions in the ecosystem:

$$E_i = \frac{\sum_{j=1}^{n_{\text{neighbours}}} ((B^*_i - D^*_i) - (B_j - D^*_j))}{8} - \frac{\sum_{j=1}^{n_{\text{neighbours}}} ((B_i - D_i) - (B_j - D_j))}{8}, \quad (4)$$

where B^*_i , D^*_i , and D^*_j are the parameters after the gain/loss of the gene.

Strain Diversity and Phylogenetic Reconstruction

In this model, we distinguish between phenotypic strain diversity and phylogenetic diversity. Strain diversity is measured at the phenotypic level, by the presence or absence of resistance and toxin-production (i.e., strains with redundant genes do not contribute to strain diversity). To exclude transient mutants, only strains consisting of more than 20 individuals are included.

In addition to strain diversity, phylogenetic trees are constructed to study the phylogenetic diversity and the balance between vertical and horizontal transfer of genes. We added neutral bit-string markers to both the cells and to each of the genes. Note however that these markers are strictly for observational purposes, and have no meaning other than to reconstruct the evolutionary history. All individuals carry a neutral marker that is only inherited vertically, but the marker for genes can both be inherited vertically and horizontally (through the transformation-like process described above). Upon inheritance, the 1,000-bit-long binary strings mutate with a chance of 10^{-6} per bit. Distance matrices are made of the populations of genes and individuals, allowing us to make species trees, gene trees, and most importantly, to map closely related genes onto the species tree. All trees were generated in R version 3.0.2 using hclust with an average similarity method.

Results

We first focus on some key results of the model under parameter conditions given by table 1. Both gene costs and loss were set as low as was possible without resulting in the loss of too many genes. Although in larger grids lower values can be used without genes being lost, we chose a relatively small grid (250×250) for computational reasons. In the section “Evolution of the Fitness Landscape,” we discuss how too low a gene loss/cost eventually leads to the loss of most genes.

Table 1

Default Parameters for the Model

Parameter	Value
Grid size	250×250
Base death (d)	0.05 per time step
Toxicity (t)	0.3 per toxin
Gene cost (c)	0.05 per gene
Gene loss (l)	0.0005 per gene per replication
Rate of uptake (u)	0.01 per cell per gene
Initial HGT rates	0.1
Gene influx No. (f)	3 per time step
Decay genes	0.1
eDNA diffusion No.	3 per time step
Initial gene mobility	0.1
Mobility mutation rate	0.01 per replication
Mobility mutation size	Uniform 0.05
Neutral marker mutation rate	10^{-6} per bit

Evolution of Differential Gene Mobility

Individuals are initialized by randomly assigning relatively immobile toxin and resistance genes ($m \sim 0.1$) to a subset of the population, yielding a population where individuals mutually kill one another (see fig. 2B and [supplementary movie S1, Supplementary Material](#) online). During this initial phase, the average mobility of both toxin and resistance genes, respectively, called T_{mobility} and R_{mobility} henceforward, is increasing (fig. 2A). The flux of genes favors the spread of resistance, and after some time, the pattern of mutual killing virtually disappears (fig. 2C). From this point, the mobility of resistance genes (R_{mobility}) no longer increases, but minimizes to approximately 0.10, whereas T_{mobility} evolves to a high value (~ 0.95). Surprisingly, diversity is maintained even after long-term evolution, despite the high levels of resistance. In this equilibrium phase, long periods of static coexistence (fig. 2C) are interspersed with localized killing (fig. 2D and [supplementary movie S1, Supplementary Material](#) online).

We next wished to understand how the divergence of T_{mobility} and R_{mobility} comes about, and how this process maintains a diverse population. Figure 3A displays how the average genome composition of individuals changes over time. First, there is a strong increase in individual resistance (gray area), as was already suggested by the disappearing pattern of mutual killing. The benefit of producing toxins should be decreasing, as individuals become resistant to a variety of toxins. Indeed, the decline in the number of toxin genes per individual (red area) reflects that producing a toxin is, on average, costly. Some toxin genes are completely lost from the population, as can be seen in the declining number of toxins in the pan genome (dark blue line). However, toxins genes that evolve higher mobility than their corresponding resistance factor appear to be maintained (fig. 3B, gray lines), even after long-term evolution. Note that toxin and resistance genes with less diverged gene mobility, eventually go extinct (fig.

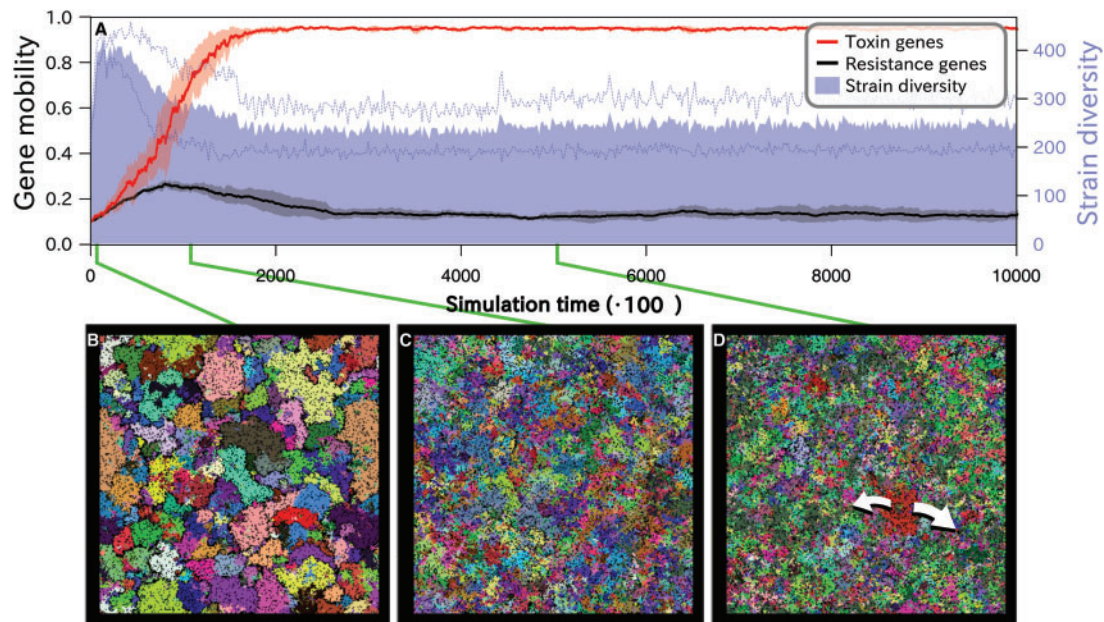


Fig. 2.—(A) Evolution toward differential gene mobility. T_{mobility} (red line), R_{mobility} (black line), and strain diversity (blue area) are plotted against time. As the results are averaged for five independent simulations, the dotted lines represent the standard deviations. Only strains that were prevalent (>20 individuals) were accounted for the strain diversity. Panels (B)–(D) depict snapshots of the field-dynamics, where black depicts empty cells, and all other colors are unique strains. The white arrows in panel (D) depict how the red strain locally overgrows sensitive individuals.

3B, red lines). We conclude from figure 3B that toxin genes can survive prolonged periods of unrewarding toxin production, by shifting selection toward the gene-level. The persistence of these genes in the population is rewarded when a strain that has lost resistance to the toxin is locally overgrown (as shown in fig. 2D). These results explain how only toxin and resistance genes with strongly differential gene mobility are maintained after long-term evolution, which in turn helps to uphold strain diversity.

Differential Gene Mobility Is Required for Phenotypic Diversity

Spatial patterns allow for many strains to coexist, because there is no all-against-all competition. Indeed, mixing the population results in the extinction of almost all toxin and resistance genes (supplementary fig. S1, Supplementary Material online). Maintaining diversity depends mostly on maintaining standing variation, as the influx of new genes (f) is low (table 1). Does the observed diversity depend on the evolution of the differential gene mobility, or is the process of HGT alone enough to explain the maintained diversity? To answer this question, simulations are done without allowing gene mobility to evolve. Figure 4 shows that diversity always decreases to very low values when T_{mobility} and R_{mobility} are fixed to the same value, even when T_{mobility} and R_{mobility} are both equally high (green line). As a reference, it is shown that the diversity of previously discussed simulations (blue area) is much higher. When the mobility of toxin and resistance genes is diverged

from the start of the simulation (gray area), the population retains very high diversity within the timespan of the simulation, showing that differential gene mobility indeed help in preventing the decline of diversity. Concluding, in our closed ecosystem with low influx, diversity can only be maintained through differential gene mobility.

Evolution of the Fitness Landscape

Why does R_{mobility} remain low, if higher mobility could help prevent local sweeps? Why are there no all-resistant individuals, outcompeting all toxin producing individuals? To further understand the evolved genome composition, and appearance of differential gene mobility, we constructed the fitness landscape for gain and loss of genes both early and late in the evolution. Figure 5A shows that gaining a resistance gene is almost always detrimental for the cell. This peak at -0.05 reflects the per-gene costs (c parameter) the host pays for a resistance gene, which yields no direct benefit. In contrast, figure 5B shows that the effect of gaining a toxin gene is much more varied, and although it is often detrimental because of self-killing or the gene cost, it can be very advantageous as well. Interestingly, long-term evolution leading to differential gene mobility increases both neutrality and the beneficial effects of toxin gain. In other words, the fairly weak initial bias leads to evolution of higher T_{mobility} , which in turn enhances the advantage of gaining genes.

Loss of genes (i.e., a nonevolvable parameter) shows an opposite trend. Losing resistance genes has a varied effect

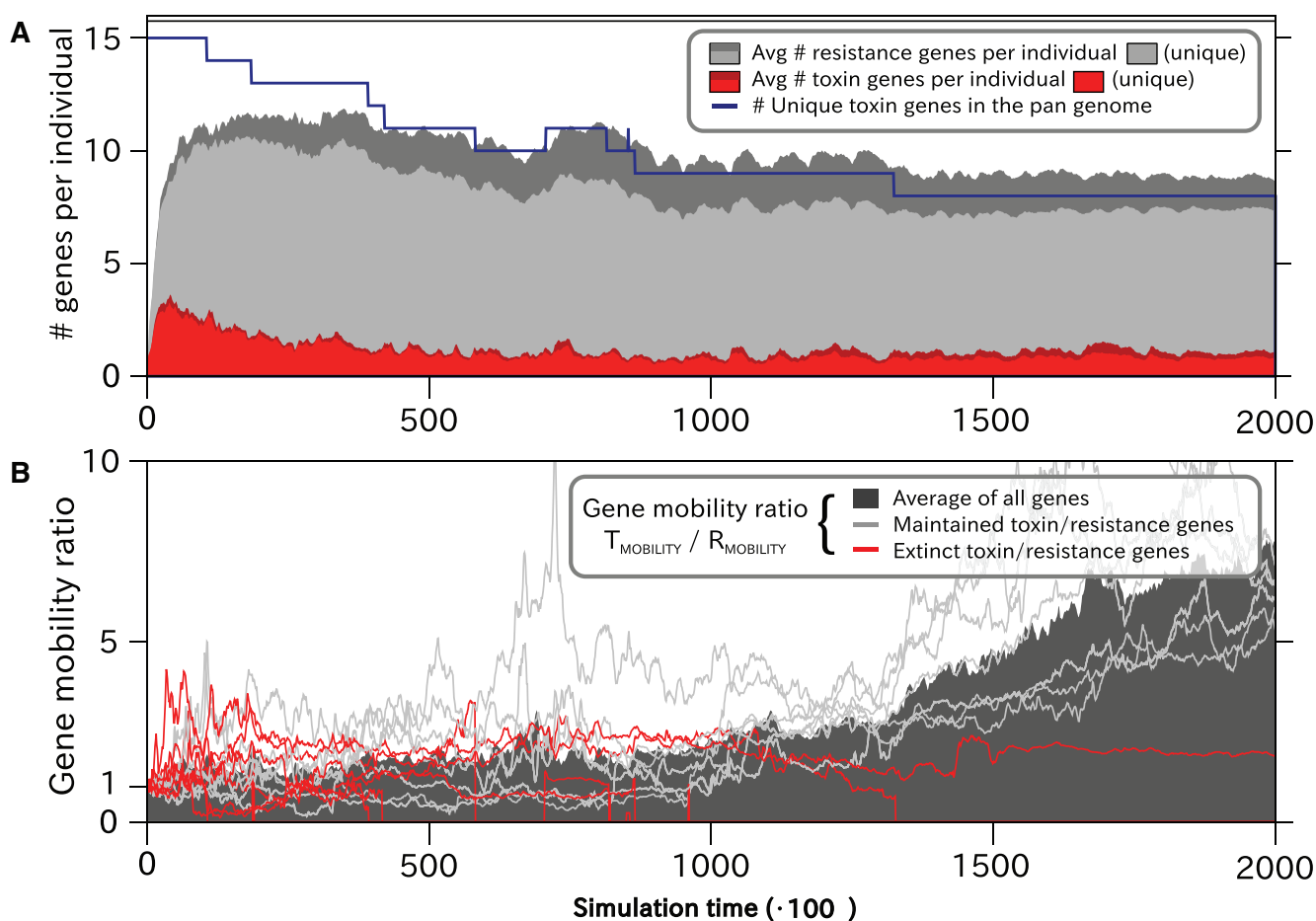


FIG. 3.—High levels of resistance promote the evolution of differential gene mobility. (A) Average number of different types of toxin genes (red) and resistance genes (gray) per individual. The darker shadow displays the total of nonunique genes of both types. The total number of toxins in the pan genome is depicted by the blue line. (B) The ratio between T_{mobility} and R_{mobility} for all corresponding toxin and resistance genes is displayed, where maintained couples are displayed in gray, and couples that go extinct before 1×10^6 simulation steps are displayed in red. The average gene mobility ratio is displayed by the dark gray area.

on fitness, and is frequently deleterious (fig. 5C), whereas losing toxin genes is almost always slightly beneficial (fig. 5D). Together these fitness effects lead to the observed population with high levels of resistance and much lower levels of toxicity (fig. 3A), and the evolution of differential gene mobility (fig. 2A). Furthermore, the fitness landscapes show that long-term evolution leads to significantly higher neutrality, reflecting the long-term evolution to higher robustness known for static fitness landscapes (van Nimwegen et al. 1999; Wagner 2005). In contrast to these classical studies, the evolution of the fitness landscape is not only due to the evolution of genome composition but also due to the evolved population structure and local pattern formation. To quantify the role of local pattern formation, [supplementary figure S2, Supplementary Material online](#), displays the fitness landscape for the evolved population structure and genome composition, excluding the influence of local pattern by choosing

random interactors instead of neighbors. Most strikingly, the beneficial effects of gaining toxin genes are not present ([supplementary fig. S2B, Supplementary Material online](#)), showing that gaining a toxin gene hardly ever yields a “global” advantage. Spatial pattern formation is indeed essential for the outcome of the model. If spatial pattern formation is prevented, diversity collapses in the very short term (data not shown). We conclude that the functional differences between toxin and resistance genes, and therewith the differences in fitness landscapes, play an important role in the evolution of differential gene mobility, and its consequences.

Finally, the fitness landscape also clarifies why too low per gene costs/loss leads to the loss of most genes. We can infer from figure 5B that the advantageous effects of gaining a toxin are very rare. When gene cost or gene loss is too low, multiresistant cells readily arise. This makes the advantageous effects of gaining toxin genes become even more rare.

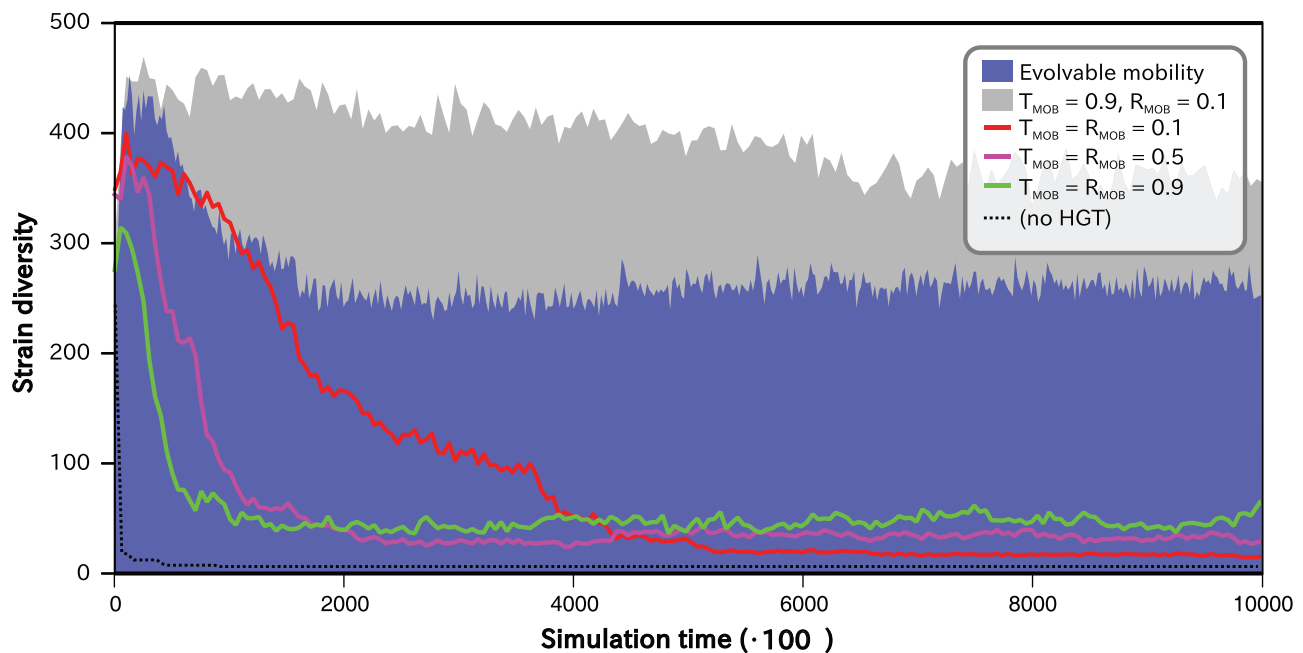


Fig. 4.—Differential gene mobility is required to maintain phenotypic diversity. Average strain diversity is depicted over evolutionary time, for different scenarios where T_{mobility} and R_{mobility} do not evolve (colored lines). For comparison, the diversity from the main simulations is also displayed (blue area). The gray area depicts simulations where toxins and resistance genes have differential gene mobility (0.9 and 0.1, respectively) that does not evolve.

Therefore, even very mobile toxin genes are lost from the population, which in the long run leads to the loss of the resistance genes, which are now useless. Indeed, [supplementary figure S3, Supplementary Material](#) online, shows that when gene cost or gene loss is low, the strain diversity is much lower. Even so, differential gene mobility still evolves. Increasing gene cost or gene loss generally yields clonal dynamics, and does not result in the evolution of differential gene mobility.

Core Resistance Repertoires and Accessory Toxins

We next focus on how toxin and resistance genes are distributed among the various strains, as T_{mobility} and R_{mobility} have diverged to nearly a 10-fold difference. For this, we have tagged both individuals and genes with a neutral bit-string marker (see Materials and Methods). Note that this marker is only used for observational purposes, and has no meaning otherwise. The markers for individuals can only be inherited vertically, which we use to generate a reference “species”-tree. Next, the phylogeny of genes is reconstructing using their own neutral markers. When the effective rate of HGT is high, closely related genes will appear in different clades in the reference tree. As shown in [figure 6A](#), the evolved differential gene mobility is reflected in how a typical toxin/resistance gene is mapped onto the phylogenetic tree. Closely related clusters of a typical resistance gene (depicted as different colors in the inner circle) appear together in the tree, whereas for the corresponding toxin, closely related genes appear in

many different clades (outer circle, also see [supplementary fig. S4, Supplementary Material](#) online, for a representation of all genes). To quantify this effect for all genes and in more replicate simulations, we use a Robinson–Foulds metric (Robinson and Foulds 1981), the fraction discrepant bipartitions, as a measure for tree-distance ([fig. 6B](#)). Clearly, resistance genes are a better representation of the species tree than toxin genes are. The evolved differential gene mobility leads to persistent and core-like resistance repertoires with more transient accessory-like toxin genes.

Strains Are Cohesive Units of Interaction

The sparing of close kin as described by Vetsigian et al. (2011) and Cordero et al. (2012) is a feature that requires a nonclonal, phylogenetic structure. The clonal populations in the model by Vetsigian et al. (2011) indeed fail to reproduce this observed pattern, as killing increases even for very small genetic distances. To study this feature for our simulations, [figure 7](#) displays how the cumulative probability of killing relates to the genetic distance with the target strain. Killing is not observed between highly related individuals, and higher frequencies of killing increase only for much greater genetic distances. As seen in the phylogenetic tree plotted along the same axis in [figure 7](#), the level of taxonomic structure below this threshold reflects the core-like resistance repertoires of strains that are used for [figure 6](#).

Similar to the all-against-all assay of *Vibrionaceae* sp. by Cordero et al. (2012), strains in our model are also cohesive

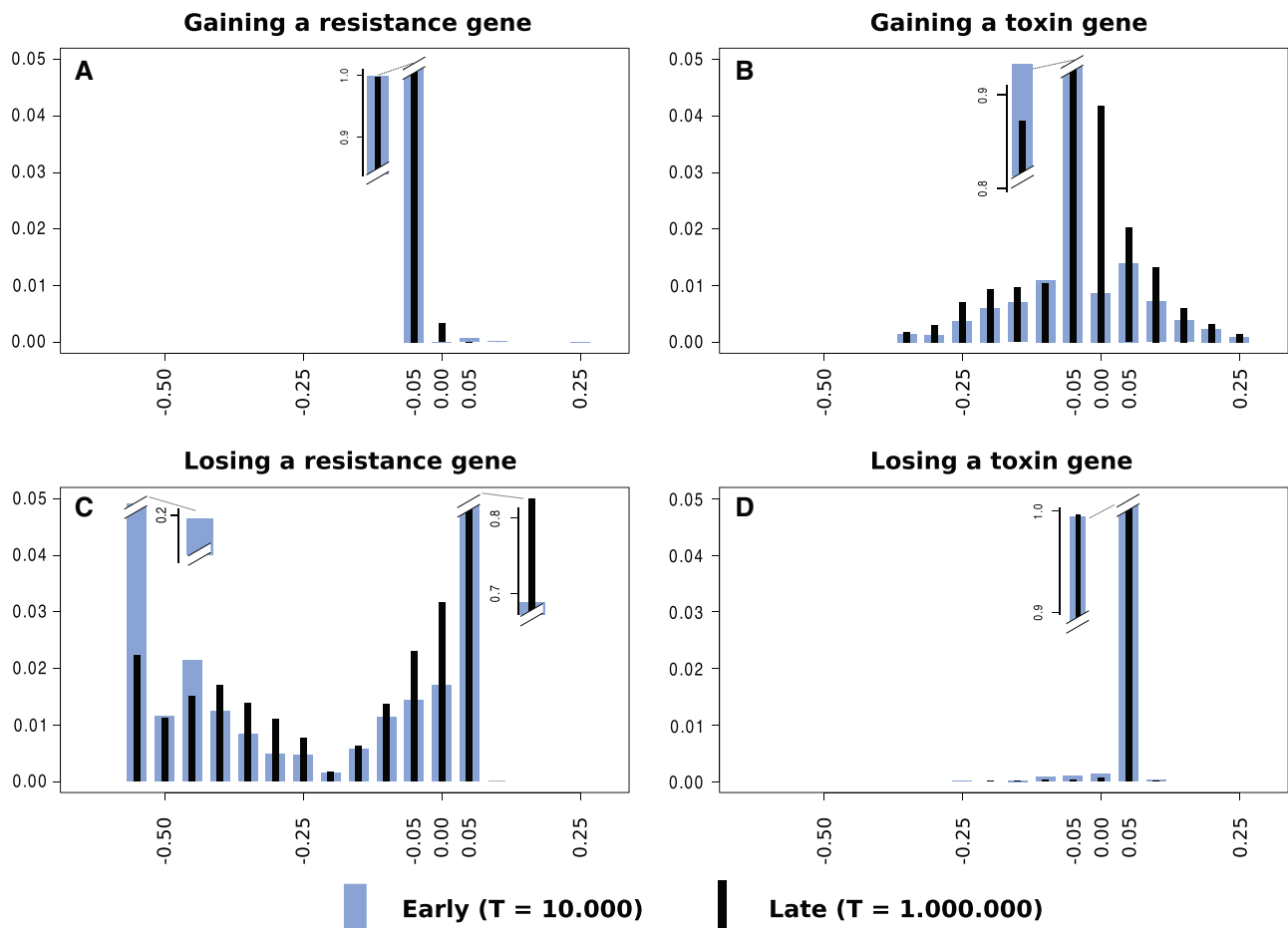


FIG. 5.—Evolution of the fitness landscape of gaining or losing toxin and resistance genes. The distribution of fitness ($B_j - D_j$) advantages with respect to local neighbors for gaining/losing genes. Blue bars indicate an early time point where T_{mobility} and R_{mobility} have not diverged yet. Black bars indicate the end of the simulation, where T_{mobility} and R_{mobility} have been diverged for a long time. Each type of mutation (A–D) was repeated 1,000 times for every individual in the population. All error bars were very small, and omitted for clarity.

populations, which act as emerging units of interaction. Killing happens more likely between clades than within, in accordance to the high phylogenetic structure and the core-like behavior of resistance repertoires. Spatial structure plays an important role in maintaining this pattern. Strains that lose resistance are less likely to survive if the corresponding toxin is produced by neighboring (i.e., more related) strains. The surprising interactions observed in all-against-all assays (e.g., superkillers) are not the interactions that determine the eco-evolutionary outcome of our model.

Diversity Is Robust to Selective Sweeps

Despite the small influx of genes, the rare discovery of new toxins happens in three of the ten replicated experiments. Surprisingly, diversity increases in all three cases, despite the fact that these events are followed by what looks like a selective sweep (supplementary movie S2 and fig. S5,

Supplementary Material online). The distance to the most recent common ancestor (MRCA) and most of the underlying population structure is maintained even after the sweep, as sufficient HGT happens between the sweeping strain and other pre-existing strains. We varied the diffusion of eDNA as a proxy for interstrain gene flow and subjected these populations to similar selective sweeps. When diffusion of eDNA is very low, HGT happens more frequently within strains than between them. When we subjected populations with low eDNA diffusion to a selective sweep, two out of four sweeps lead to a phylogenetic bottleneck, which never happens with higher rates of eDNA diffusion (see supplementary fig. S6, Supplementary Material online). This shows that when the interstrain gene flow is sufficiently faster than a selective sweep, the genes sweep the population on their own, allowing not only strain diversity but also phylogenetic diversity to persist.

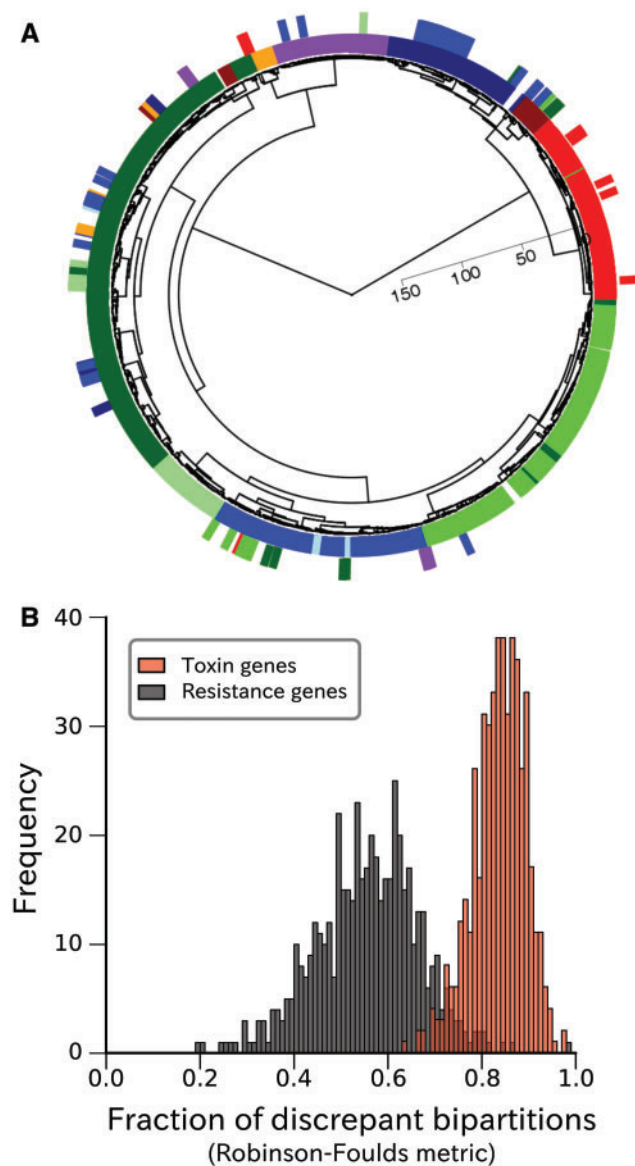


FIG. 6.—Phylogenetic reconstruction shows core-like resistance with a patchy distribution of toxin genes. (A) For a typical corresponding toxin and resistance gene, closely related genes (different colors) are mapped onto the phylogenetic tree of 500 randomly sampled individuals from a typical simulation. The inner circle depicts how the phylogeny of the resistance gene closely resembles the phylogenetic tree. The corresponding toxin in the outer circle displays that the toxin gene has a much more scrambled distribution. Although this effect exists on many levels (i.e., does not strictly depend on how many gene clusters are made), this example uses ten gene clusters. (B) Frequency histograms of the fraction of discrepant bipartitions, a Robinson–Foulds metric (Robinson and Foulds 1981), predicted by both gene-types for a total of ten simulations. The distribution for resistance genes and toxin genes has significantly different means ($P \ll 10e-14$, Welch’s t -test).

Differential Gene Mobility Evolves Only in a Closed Ecosystem

The results discussed so far are examples from a relatively closed ecosystem, that is, with low influx of genes. Next, we focus on gradually more open ecosystems by increasing the external influx of genes into the eDNA plane (fig. 8). For very open ecosystems, the difference between T_{mobility} and R_{mobility} is much smaller, as both T_{mobility} and R_{mobility} evolve high values. Although this increasing influx mostly has a positive effect on the strain diversity, the branch length distribution shows that this scenario describes an evolutionary arms race where novel clones continuously replace one another (see [supplementary movie S3, Supplementary Material](#) online). Indeed, the phylogenetic trees show that at high influx phylogenetic diversity is much lower than at lower influx. Clearly, the core-like resistance repertoire as discussed above no longer exists at this parameter-range. Differential gene mobility and the resulting core and accessory genomes are most strongly observed at low influx, that is, when most eDNA originates from the local population itself.

Discussion

We showed that differential gene mobility evolves in a simple model where HGT occurs by incorporation of locally present eDNA. Toxin genes evolve high mobility, whereas resistance genes are mostly vertically inherited. This differential gene mobility is furthermore required to maintain a diverse population, as diversity collapses when mobility is equal for toxin and resistance genes. Moreover, for parameter values which do not evolve differential gene mobility ([supplementary fig. S3, Supplementary Material](#) online), the high diversity as observed in vivo does not occur. Importantly, the differential gene mobility appears to generate the within-species population structure and interaction dynamics as observed in vivo (Vetsigian et al. 2011; Cordero et al. 2012; Cui et al. 2015). Our model shows that many observations on microbial ecosystems can be explained by local, internal HGT and the resulting interplay of gene- and cell-level evolution.

Closed Ecosystems with High Diversity

Interspecies HGT challenges our concept of species (Zhaxybayeva and Doolittle 2011). On the other hand, within-species HGT on a more local scale is now challenging our expectations on the eco-evolutionary dynamics of prokaryotes (Shapiro et al. 2012; Overballe-Petersen and Willerslev 2014; Takeuchi et al. 2014). We have shown in figure 8 that the dynamics of a closed ecosystem, that is, internal HGT of with a local pool of eDNA, is very different from open ecosystems. The closed ecosystem is very congruent with the observation of core-like resistomes in natural ecosystems (Forsberg et al. 2014). Furthermore, despite the closed and relatively small ecosystem, adaptations such as

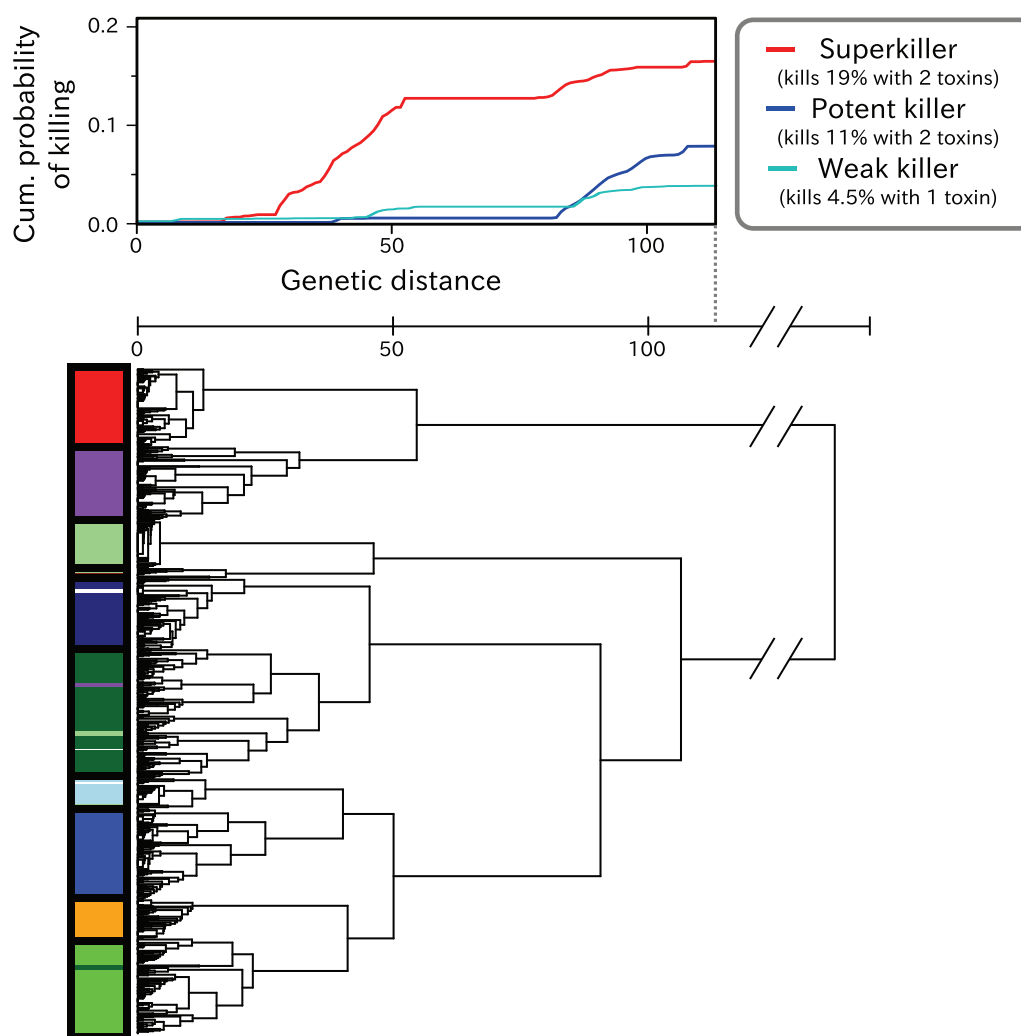


Fig. 7.—Strains spare close kin from killing. In an all-against-all killing assay, we compute the cumulative probability that the focal individual can kill a target individual given the genetic distance between them. Along the same axis the phylogenetic tree of 500 randomly sampled individuals is displayed, clustered into ten clades based on a core-like resistance gene (different colors). These ten clades correspond to the genetic distance where all profiles stop sparing close kin (~40).

toxin production can be present because of gene-level selection. This shows that diversity and the presence of superkillers (Vetsigian et al. 2011; Cordero et al. 2012) do not necessarily contradict one another. Moreover, the diversity and dynamics of the closed ecosystem do not depend on high rates of de novo gene discovery, as in other modelling approaches (Vetsigian et al. 2011; Cui et al. 2015), but on local HGT.

Although observations in our model are strikingly similar to in vivo observations, bacteria in agricultural soil (Heuer et al. 2011) or hospital wastewater (Naiemi et al. 2005) might have the dynamics of an open ecosystem (e.g., as a result of anthropogenic influences). We have shown that increasing the external influx of genes results in very rapid dynamics of clonal expansions driven by the continuous discovery of novel genes (fig. 8 and [supplementary movie S3, Supplementary Material](#) online). Although superficially this ongoing de novo gene discovery

results in higher strain diversity, figure 8 shows that repeated genome-wide selective sweeps actually negatively impact phylogenetic diversity. In other words, our results suggest that, counterintuitive, genetic diversity is highest when the majority of available adaptations originates from the population itself.

Earlier research has also challenged the idea diversity is reduced by genome-wide selective sweeps of beneficial adaptations. Shapiro et al. (2012) propose that HGT can actually result in gene-specific selective sweeps, and how the diversity of the population is not necessarily purged by this process. Indeed, Takeuchi et al. (2015) have studied a model of gene-specific sweeps of a generic beneficial adaptation through negative frequency-dependent selection. Here, we have also shown gene-specific sweeps, but now in a context in which the benefit of a gene depends on the local neighborhood and the genetic background of the host cell. As such, newly

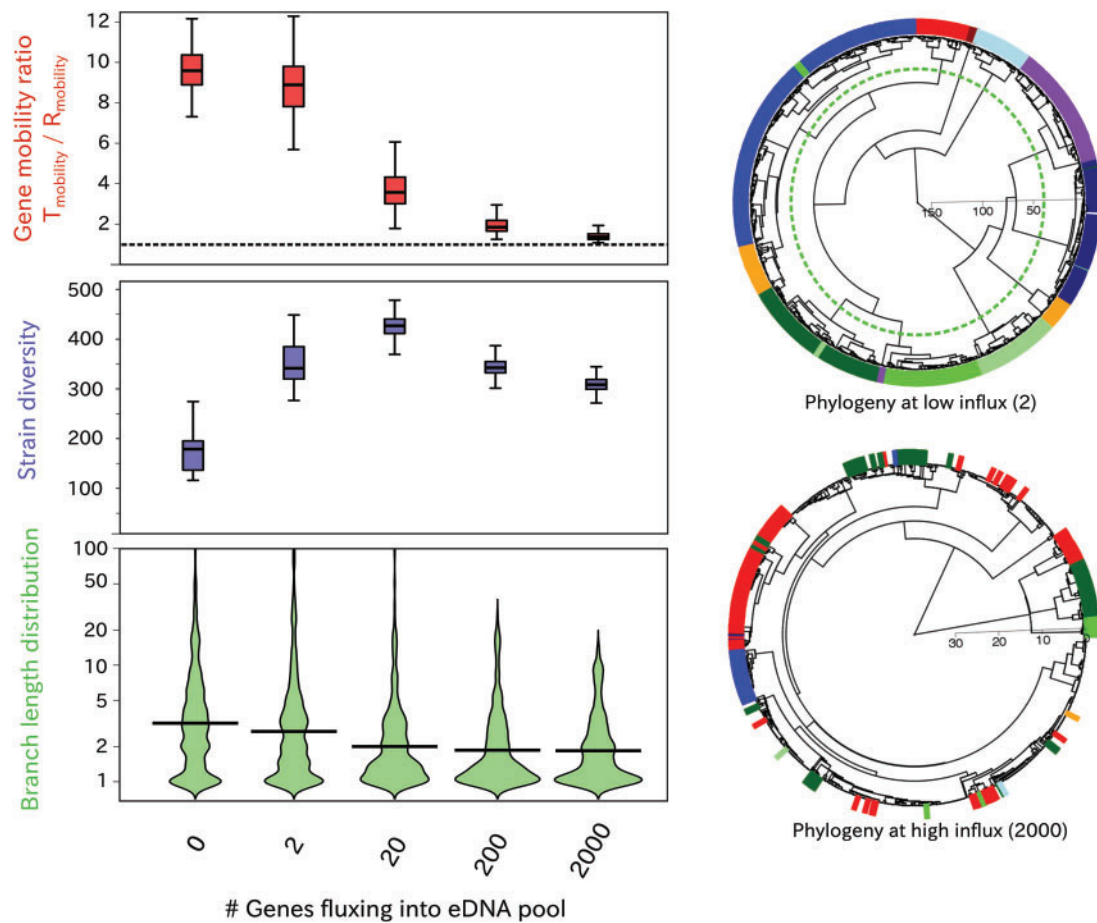


Fig. 8.—Differential gene mobility evolves only in a closed ecosystem. The effect of increasing influx on the gene mobility ratio (red), the strain diversity (blue), and the branch length distribution (green). Two phylogenetic trees display the difference in the population structure between a typical ecosystem with low influx and a typical ecosystem with high in flux. Clusters of an arbitrary resistance gene (different colors) are mapped to the trees to visualize the presence/absence of core genomes. The dashed green line displays how a single core-genome at low influx has a distance to MRCA comparable to the entire tree of the ecosystem with high influx.

discovered toxin and resistance genes spread through the population on their own, maintaining phylogenetic diversity (supplementary fig. S5, Supplementary Material online).

The evolutionary impact of HGT, and evolution of the other HGT-related processes, can be very different depending on the scale, the donor and recipient species, and the environment. Together with recent studies, our results show that microbial evolution is not driven by clonal dynamics, and that even small, closed niches can be expected to harbor a lot of genetic and phenotypic diversity.

Rarely Beneficial Genes Evolve High Mobility

The question of how and when HGT could be favorable is in most studies focused on the potential benefits of the cell (Martinez 2009; Heuer et al. 2011). However, in our model we take into account both gene- and the cell-level selection, and show that it is actually the least beneficial genes, toxin genes, that evolve high mobility. As shown in figure 3, as

highly resistant individuals evolve, only toxin genes with mobility higher than their corresponding resistance gene are able to survive. We interpreted the low mobility of resistance in closed ecosystems as reduction of fitness cost of redundant genes, as individuals in the vicinity are already resistant. We have indeed shown that taking up resistance genes is rarely beneficial (fig. 5). Congruent with these interpretations is the fact that that high diffusion of eDNA leads to increased gene mobility for both types of genes (supplementary fig. S1, Supplementary Material online), once again revealing how locality impacts the process of HGT. Despite the sparse presence of toxin producing individuals, resistance is sufficiently selected on the cell-level, whereas toxin genes need to evolve high gene mobility to survive. We tested the notion that less beneficial genes evolve higher gene mobility in a simple model, showing that there is indeed a significant anticorrelation between gene mobility and the frequency of selection (P value $\ll 2e-3$, $R^2 = 0.4611$, see supplementary material, Supplementary Material online, for implementation). As can

be seen from figure 6, the process described above results in core-like resistance repertoires with a patchy distribution of toxin genes. A similar pattern is observed in core- and accessory genomes in vivo as accessory genes are only occasionally useful (i.e., pathogenicity or secondary metabolism [Hacker and Carniel 2001; Norman et al. 2009]). This pattern could be explained by the dominant role of loss of genes (Wolf and Koonin 2013), but this does not explain the discrepancy between the phylogeny of genes and cells. We show that HGT plays an important role in shaping this observed genome mosaicism (Zhaxybayeva et al. 2004).

Who Wants to Take Up eDNA?

In our model, we assumed that the uptake of eDNA by cells is fixed. Although not all routes of HGT can be prevented by the cell, for example, transduction, many bacterial species are known to have evolved regulation of DNA-uptake (Seitz and Blokesch 2013). This suggests that uptake of eDNA is only occasionally useful. Indeed, when the uptake of eDNA is made evolvable in our model, cells avoid uptake under many conditions except high gene-loss (data not shown). What conditions make the uptake of eDNA favorable? Although DNA can serve as a nutrient (Finkel and Kolter 2001), which might lead to transformation as a side-effect (Macfadyen et al. 2001), more direct benefits of natural competence are poorly understood. Contrasting the regulated uptake of eDNA that is observed for most bacteria, the human pathogens of *Neisseria* sp. and *Helicobacter pylori* both constitutively take up eDNA (Seitz and Blokesch 2013). The strong environmental pressure by the immune system might directly select for diversity, potentially explaining the constitutive uptake of eDNA. Indeed, earlier studies have shown that the uptake of eDNA can prevent the loss of genomic information (Vogan and Higgs 2011; Johnston et al. 2013; Takeuchi et al. 2014). Indeed, by generating and maintaining diversity HGT could contribute to the evolvability of prokaryotes (Hindré et al. 2012), shaping the standing variation on which natural selection acts. In other words, an important entry point for future research would be the evolution of natural competence in relation to the environmental variability, and the signals used by the cells to induce this process.

Genes in our model are not only taken up with a fixed rate but also strictly independent of each other. In other words, two genes cannot get transferred together in a single event, as we assume no genetic linkage between genes. Although such decoupled behavior is observed in the *Vibrio*'s discussed by Cordero et al. (2012), other research commonly observes both genes together on a plasmid or MGE (Alonso et al. 2000; Van Melderen and Saavedra De Bast 2009). Linking genes can circumvent some of the risks of HGT, and could even feedback on the earlier questions regarding the evolution of natural transformation. What conditions favor independent toxin and resistance genes, and when does the alternative solution

of, for example, toxin–antitoxin systems evolve? To investigate this, the model can be extended to include the evolution of genome structure, where HGT of fragmented DNA or plasmids could include multiple genes in a single event. Answering the aforementioned questions will further increase our understanding on multilevel evolution and the selection pressures that acts on these different biological levels.

We conclude that HGT and the resulting interplay between gene-level and cell-level selection can structure genomes and populations of microbial communities as observed in vivo.

Supplementary Material

Supplementary movies S1–S3 and figures S1–S6 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

The authors thank the editor and referees for their critical view, resulting in improvement of both the content and style of the manuscript. This work was supported by the European Commission 7th Framework Programme (FPFP7-ICT-2013.9.6 FET Proactive: Evolving Living Technologies) EvoEvo project (ICT-610427).

Literature Cited

- Alonso G, Vilchez G, Rodríguez Lemoine V. 2000. How bacteria protect themselves against channel-forming colicins. *Int Microbiol.* 3:81–88.
- Cordero OX, et al. 2012. Ecological populations of bacteria act as socially cohesive units of antibiotic production and resistance. *Science* 337:1228–1231.
- Cui Y, et al. 2015. Epidemic clones, oceanic gene pools, and Eco-LD in the free living marine pathogen *Vibrio parahaemolyticus*. *Mol Biol Evol.* 32:1396–1410.
- de Vries J, Wackernagel W. 2002. Integration of foreign DNA during natural transformation of *Acinetobacter* sp. by homology-facilitated illegitimate recombination. *Proc Natl Acad Sci U S A.* 99:2094–2099.
- Finkel SE, Kolter R. 2001. DNA as a nutrient: novel role for bacterial competence gene homologs. *J Bacteriol.* 183:6288–6293.
- Forsberg KJ, et al. 2014. Bacterial phylogeny structures soil resistomes across habitats. *Nature* 509:612–616.
- Fraser C, Alm EJ, Polz MF, Spratt BG, Hanage WP. 2009. The bacterial species challenge: making sense of genetic and ecological diversity. *Science* 323:741–746.
- Hacker J, Carniel E. 2001. Ecological fitness, genomic islands and bacterial pathogenicity: a Darwinian view of the evolution of microbes. *EMBO Rep.* 2:376–381.
- Heuer H, Schmitt H, Smalla K. 2011. Antibiotic resistance gene spread due to manure application on agricultural fields. *Curr Opin Microbiol.* 14:236–243.
- Hindré T, Knibbe C, Beslon G, Schneider D. 2012. New insights into bacterial adaptation through in vivo and in silico experimental evolution. *Nat Rev Microbiol.* 10:352–365.
- Johnston C, Martin B, Granadel C, Polard P, Claverys JP. 2013. Programmed protection of foreign DNA from restriction allows pathogenicity island exchange during pneumococcal transformation. *PLoS Pathog.* 9:e1003178.
- Kolstø AB. 1997. Dynamic bacterial genome organization. *Mol Microbiol.* 24:241–248.

- Koonin EV, Wolf YI. 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* 36:6688–6719.
- Macfadyen LP, et al. 2001. Competence development by *Haemophilus influenzae* is regulated by the availability of nucleic acid precursors. *Mol Microbiol.* 40:700–707.
- Martinez JL. 2009. The role of natural environments in the evolution of resistance traits in pathogenic bacteria. *Proc R Soc Lond B Biol Sci.* 276:2521–2530.
- Mell JC, Redfield RJ. 2014. Natural competence and the evolution of DNA uptake specificity. *J Bacteriol.* 196:1471–1483.
- Naiemi NA, et al. 2005. Widespread transfer of resistance genes between bacterial species in an intensive care unit: implications for hospital epidemiology. *J Clin Microbiol.* 43:4862–4864.
- Norman A, Hansen LH, Sørensen SJ. 2009. Conjugative plasmids: vessels of the communal gene pool. *Philos Trans R Soc Lond B Biol Sci.* 364:2275–2289.
- Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* 405:299–304.
- Overballe-Petersen S, Willerslev E. 2014. Horizontal transfer of short and degraded DNA has evolutionary implications for microbes and eukaryotic sexual reproduction. *BioEssays* 36:1005–1010.
- Preheim SP, Timberlake S, Polz MF. 2011. Merging taxonomy with ecological population prediction in a case study of Vibrionaceae. *Appl Environ Microbiol.* 77:7195–7206.
- Puigbò P, Lobkovsky AE, Kristensen DM, Wolf YI, Koonin EV. 2014. Genomes in turmoil: quantification of genome dynamics in prokaryote supergenomes. *BMC Biol.* 12:66.
- Robinson DF, Foulds LR. 1981. Comparison of phylogenetic trees. *Math Biosci.* 53:131–147.
- Seitz P, Blokesch M. 2013. Cues and regulatory pathways involved in natural competence and transformation in pathogenic and environmental Gram-negative bacteria. *FEMS Microbiol Rev.* 37:336–363.
- Shapiro BJ, et al. 2012. Population genomics of early events in the ecological differentiation of bacteria. *Science* 336:48–51.
- Snel B, Bork P, Huynen MA. 2002. Genomes in flux: the evolution of archaeal and proteobacterial gene content. *Genome Res.* 12:17–25.
- Takeuchi N, Cordero OX, Koonin EV, Kaneko K. 2015. Gene-specific selective sweeps in bacteria and archaea caused by negative frequency-dependent selection. *BMC Biol.* 13:20.
- Takeuchi N, Kaneko K, Koonin EV. 2014. Horizontal gene transfer can rescue prokaryotes from Muller's ratchet: benefit of DNA from dead cells and population subdivision. *G3 (Bethesda, MD)* 4:325–339.
- Toffoli T, Margolus N. 1987. Cellular automata machines: a new environment for modeling (Scientific Computation). The MIT Press.
- Van Melderen L, Saavedra De Bast M. 2009. Bacterial toxin–antitoxin systems: more than selfish entities? *PLoS Genet* 5:e1000437.
- van Nimwegen E, Crutchfield JP, Huynen M. 1999. Neutral evolution of mutational robustness. *Proc Natl Acad Sci U S A.* 96:9716–9720.
- Vetsigian K, Jajoo R, Kishony R. 2011. Structure and evolution of *Streptomyces* interaction networks in soil and in silico. *PLoS Biol.* 9:e1001184.
- Vogan AA, Higgs PG. 2011. The advantages and disadvantages of horizontal gene transfer and the emergence of the first species. *Biol Direct.* 6:1.
- Wagner A. 2005. Robustness, evolvability, and neutrality. *FEBS Lett.* 579:1772–1778.
- Wolf YI, Koonin EV. 2013. Genome reduction as the dominant mode of evolution. *BioEssays* 35:829–837.
- Zhaxybayeva O, Doolittle WF. 2011. Lateral gene transfer. *Curr Biol.* 21:R242–R246.
- Zhaxybayeva O, Lapierre P, Gogarten JP. 2004. Genome mosaicism and organismal lineages. *Trends Genet.* 20:254–260. Erratum in: 21(1):36.

Associate editor: Howard Ochman