



Genic Selection Within Prokaryotic Pangenomes

Gavin M. Douglas  and B. Jesse Shapiro *

Department of Microbiology and Immunology, McGill University, Montréal, Québec, Canada

*Corresponding author: E-mail: jesse.shapiro@mcgill.ca.

Accepted: 14 October 2021

Abstract

Understanding the evolutionary forces shaping prokaryotic pangenome structure is a major goal of microbial evolution research. Recent work has highlighted that a substantial proportion of accessory genes appear to confer niche-specific adaptations. This work has primarily focused on selection acting at the level of individual cells. Herein, we discuss a lower level of selection that also contributes to pangenome variation: genic selection. This refers to cases where genetic elements, rather than individual cells, are the entities under selection. The clearest examples of this form of selection are selfish mobile genetic elements, which are those that have either a neutral or a deleterious effect on host fitness. We review the major classes of these and other mobile elements and discuss the characteristic features of such elements that could be under genic selection. We also discuss how genetic elements that are beneficial to hosts can also be under genic selection, a scenario that may be more prevalent but not widely appreciated, because disentangling the effects of selection at different levels (i.e., organisms vs. genes) is challenging. Nonetheless, an appreciation for the potential action and implications of genic selection is important to better understand the evolution of prokaryotic pangenomes.

Key words: genic selection, pangenome, horizontal gene transfer, mobile genetic elements, gene's eye view, selfish DNA.

Significance

Closely related bacteria often encode different sets of genes, resulting in a pangenome of genes found in some but not necessarily all genomes. Selection on individual organisms—offering adaptation to specific niches—is commonly offered as an explanation for this pangenome variation. Another explanation is that selection at the level of individual genes (genic selection) drives at least some of the variation. Entirely selfish elements are clear examples of genic selection. Here, we review the literature on genic selection in prokaryotes and argue that both gene-level and organism-level selection are important forces. Understanding their relative roles and interactions represents a key challenge, and also a way forward, in the study of pangenomes.

Background

Closely related prokaryotes often encode highly variable gene sets. For instance, strains of *Escherichia coli* have long been known to share only a minority of genes in their shared core genome (Welch et al. 2002). Similar results have been observed across many other prokaryotes: one analysis determined that a range of 19–64% of genes are core to all strains of common bacterial genera (Innamorati et al. 2020). Given these observations it has become common to categorize genes within a set of genomes as part of the “core” or

“accessory” genome. The diversity in accessory genome content is largely shaped by horizontal gene transfer (HGT) and gene loss, and the collection of all genes across a set of genomes is referred to as the “pangenome” (Tettelin et al. 2005).

Pangenome diversity is closely tied to several fundamental questions in microbial evolution. In particular, it is key to understanding rapid adaptation and niche specialization across prokaryotes (Azarian et al. 2020). Pangenome diversity also has major implications for which species concepts can be

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

sensibly applied, because HGT obfuscates the traditional tree-like structure of relationships between prokaryotes (Koonin and Wolf 2008).

Given the importance of pangenomes for understanding these and other concepts, there is interest in identifying to what degree natural selection and genetic drift each shape pangenome variation (Andreani et al. 2017; McInerney et al. 2017; Bobay and Ochman 2018). Although there can be reasonable disagreements regarding the relative contributions of these two evolutionary forces, there is no doubt that many horizontally transferred genes confer adaptive benefits to the strains that encode them, at least in certain contexts (Domingo-Sananes and McInerney 2021). For instance, antibiotic resistance genes on conjugative plasmids are generally taken to be examples of selection at the level of individual prokaryotic cells, because they help cells survive in the presence of antibiotics.

There are also contrasting examples of genes that have no (or even a negative) impact on an organism's fitness, but that nonetheless spread through populations (Werren 2011). Such genes are referred to as selfish mobile genetic elements (MGEs). At the surface level this appears paradoxical: if certain MGEs are detrimental, then why do they spread?

The reason is that the transmission patterns of these genes differ from most other genes in a genome, which enables them to spread horizontally even while having a neutral or negative effect on the host organism's fitness. Such genetic conflict between genes in the same genome is known to arise whenever transmission patterns differ (Werren 2011). Indeed, intragenomic conflict has been long appreciated in evolutionary biology, particularly in eukaryotes. For example, meiotic drive refers to conflict between alleles at a locus, where one allele manipulates the process of meiosis so that it has an increased chance of being transmitted to the next generation (Maynard Smith and Szathmary 1995) (fig. 1a). Due to this bias in transmission, alleles favored through meiotic drive can rapidly spread in a population despite having no effect, or even a deleterious effect, on organismal fitness. Similar pressures are acting upon selfish MGEs: selection for enhanced mobility through HGT can benefit a particular gene at the cost of the host's fitness (fig. 1b). Genic selection (selection at the level of individual genes) could drive the high mobility of such selfish elements and maintain them at high frequencies while decreasing the host's fitness (Werren 2011). Although selfish elements are the clearest example of this phenomenon, elements with beneficial impacts on host fitness could also be under genic selection. In particular, genes that have a higher fitness relative to other genes in a genome are also considered to be under some degree of genic selection (Okasha 2006). Such genes are transmitted through a combination of both horizontal and vertical transmissions (Novick and Doolittle 2020). In any case, determining that an element is under genic selection is distinct from the common heuristic of taking

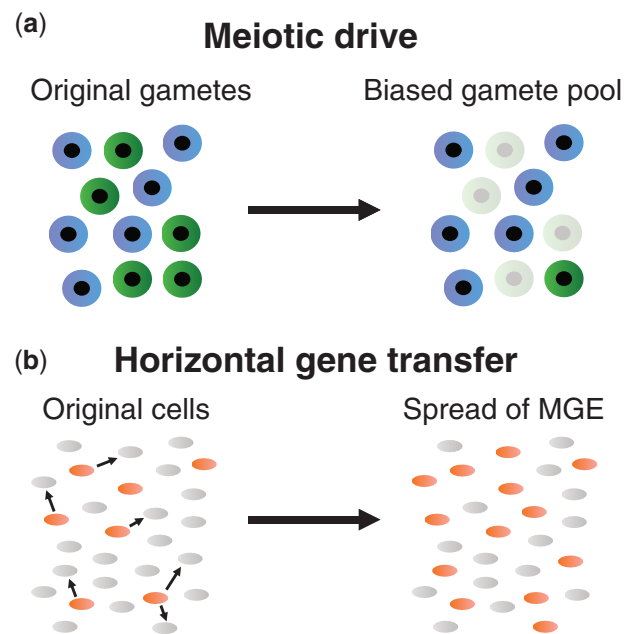


FIG. 1.—Examples of genetic transmission biases that can lead to genic selection. (a) Meiotic drive: a classical example of genic selection. This phenomenon generally occurs due to certain alleles that manipulate the meiosis process to ensure that they have higher rates of survival in eukaryotic male gametes (shown in blue). (b) Horizontal gene transfer also leads to biased transmission of certain genes, which have increased rates of transmission compared with genes that are solely vertically transmitted in a population of bacterial cells (ovals). In this example, red cells represent those that encode a MGE that can rapidly spread horizontally in the population (small arrows). Such biases in transmission can lead to genic selection for genes that are highly horizontally mobile. In extreme cases of both phenomena, genic selection can lead to genes that have deleterious effects on the host, but that nonetheless have high rates of transmission. In both (a) and (b), the large horizontal arrow denotes the passage of a short period of time (i.e., less than one generation).

the gene's eye view of evolution, which is largely agnostic to the actual targets of selection (box 1).

More generally, genic selection refers to some subset of genes being the units of selection. In this context, gene refers to an independent unit of heredity, a segment of the genome which can vary in length and contain multiple (or zero) elements that encode RNA or protein. Units of selection are entities which have the capability to undergo evolution by natural selection. A population of such entities has been referred to as a "Darwinian population" (Godfrey-Smith 2009). For such a collection of entities, such as individual genes within an organism, to be considered a Darwinian population, the three classical features for evolution through natural selection need to be present (Lewontin 1970). These features are: 1) that different entities have different phenotypes; 2) that these different phenotypes result in differential fitness between entities; and 3) that there is covariance between parent and offspring fitness (i.e., that fitness is

Box 1: The gene's eye view versus genic selection

There has been a long-standing debate regarding which units besides individual organisms could be targets of natural selection (Lewontin 1970; Hull 1980; Dawkins 1982). Myriad potential units of selection have been proposed besides individuals, such as species, populations, and genes. Some of the disagreements in this area are related to the usage of inconsistent definitions, such as varying explanations for what it means to be a “unit of selection.” This is particularly true when considering the validity of a gene as a unit of selection. For instance, many have criticized the view that genes are the ultimate targets of natural selection, because although gene frequencies may shift due to natural selection, this is in most cases due to selection on individual organisms (Gould 2002; Godfrey-Smith 2009).

Others have taken a pluralistic view where selection acting on genes is taken to be a complementary perspective of the same process acting on organisms (Dawkins 1982; Okasha 2006; Ågren 2016). In other words, they argue that it is useful to take the gene's eye view of evolution through natural selection even if genes themselves are not directly under selection. This is because considering the process from the gene's perspective can be a useful heuristic for gaining novel insights into the evolutionary process. For example, kin selection in animals has been formulated from both the perspectives of genes and the individual organisms with equivalent outcomes (Okasha 2006). The advantage of considering the gene-level perspective in this case is that it is typically more intuitive than considering the perspective of individual organisms. However, kin selection does not causally act at the gene-level and is instead a form of organism-level selection (Okasha 2006). Accordingly, although we accept that using the gene's eye view heuristic can be more intuitive, it can also lead to confusion regarding which entities are actually under selection.

Genic selection takes a stronger causal stance, and refers to selection acting directly on genes, which can lead to differential replication of genes within the same organism. Selfish genetic elements, such as transposable elements, are the classic examples of genic selection, which can spread while having negative impacts on fitness at the organism-level (Okasha 2006). Throughout the text we primarily focus on paradigmatic cases of genic selection, such as selfish mobile genetic elements, which evolve almost entirely through selection at the gene-level. However, in many cases mobile genetic elements in prokaryotic pangenomes can only marginally be considered independent units of selection. This is because they are not clearly transmitted independently of their hosts and thus it is unclear how to distinguish any genic phenotypes from organism-level phenotypes. For instance, such elements might confer adaptive benefits to their host and be consistently vertically transmitted, while being only infrequently horizontally transmitted. We argue that in such cases genic selection could still be necessary to explain the evolutionary history of such an element, although it would likely be insufficient to provide a full account without considering organism-level selection as well.

In summary, the gene's eye view is agnostic to whether selection acts at the gene-level, whereas genic selection refers specifically to genes as the direct targets of selection.

heritable). Based on this framework, genes that possess these characteristics should be considered units of selection. Phenotypes are typically attributed to organisms, not genes; however, genes can be considered to have phenotypes in some cases. For instance, mobile genes can encode their own transfer machinery, which can be thought of as a gene-level phenotype. Genes possessing this phenotype would have increased fitness relative to other genes in the same genome that are less frequently transferred, and this fitness benefit would be heritable through DNA replication. This example highlights that, at least in principle, genes can be the units of selection.

Although it is tempting to generalize this idea to all classes of genetic elements, in many cases an element's transmission (or the manifestation of its phenotype) is too tightly linked with that of its host for genic selection to be relevant. For other genetic elements, the minimal characteristics may be only marginally or transiently fulfilled (Godfrey-Smith 2009). In such cases, as discussed below, selection could be acting at

both the levels of individual genetic elements and individual organisms.

Herein, we discuss these and other concepts regarding the relevance of genic selection to prokaryotic pangenomes. We begin by reviewing the known literature on MGEs in prokaryotic genomes, which represent the clearest potential targets for genic selection. We discuss how genic selection for increased mobility and particular genomic insertion locations can enable the persistence of MGEs with neutral or even deleterious impacts on host organism fitness. We then discuss under what other conditions genic selection could occur within pangenomes, including what additional genetic characteristics could enable genes to have differential fitness. A major aspect of this discussion focuses on to what degree MGEs that are beneficial for host organisms can be under genic selection. Last, we discuss the implications of considering genic selection in the context of prokaryotic pangenomes and how this process could be better studied.

Major Classes of MGEs

Numerous MGEs, sometimes collectively called the mobilome, have been identified in prokaryotic genomes, which are major contributors to the size of pangenomes. These elements differ in their distributions across and within prokaryotic lineages (Zimmerly 2005). For instance, one investigation found that 41% of accessory genes in a subset of *E. coli* genomes were encoded within prophages, although the number of prophages per genome varied widely (Bobay et al. 2013). More generally, pathogenic bacteria often encode numerous prophages, whereas closely related benign strains encode only a few (Busby et al. 2013). MGEs are thus a major contributor to pangenome variation, perhaps more in free-living bacteria than symbionts with less access to a diverse gene pool and governed more by gene loss (Moran and Plague 2004; Kuo and Ochman 2009). Although the precise magnitude of the effect of genic selection on pangenome evolution is difficult to measure, it is plausibly an important force in many prokaryotic lineages.

Different classes of MGEs also differ in terms of the mechanisms used for within and between-genome transposition (Zimmerly 2005). Many MGEs do not encode machinery for self-sufficient HGT (i.e., between-genome transposition), and instead can only mobilize within a cell. However, these elements are often enriched in MGEs that are capable of autonomous HGT. For instance, plasmids are enriched for both DNA transposons (Redondo-Salvo et al. 2020) and group II introns (Tourasse and Kolstø 2008). This highlights that there is an overlap in mobilization strategies within and between genomes. Specifically, enhanced within-genome transposition can increase an element's chance of integrating into a horizontally transferrable MGE. In this section we describe several classes of MGEs that leverage diverse mechanisms for mobilization (table 1). We also highlight several ways that conflict can arise between prokaryotes and their constituent genes.

The most relevant characteristic that modulates the relative strengths of genic and organism-level selection is how often an MGE mobilizes to different cells. This has previously been outlined in terms of the types of symbioses that can occur between MGEs and their host (Jalasuuri and Koonin 2015), similar to those that can occur in classic host–parasite relationships. Highly mobile elements, such as prophages, are more likely to have solely neutral or deleterious effects on hosts. In contrast, elements with lower mobility, such as those that are only passively mobile, more commonly encode host-beneficial genes. These elements can be considered mutualists with the host (Jalasuuri and Koonin 2015) as their fitness is tied to host fitness. Importantly, these are not absolute categorizations of mobile elements: many prophage sequences are known to provide benefits to their host (Wang et al. 2010), whereas certain plasmids less mobile than temperate phages can be entirely selfish (Bahl et al. 2009). Nonetheless, this

relationship between mobility and degree of parasitism provides a valuable framework for understanding dynamics between MGEs and hosts.

This framework, and other evaluations of the relationship between mobile elements and host genomes (Rankin et al. 2011; Hall et al. 2020), have framed the question from the perspective of symbionts and host interactions. This is equivalent to taking the gene's eye view of accessory genes in prokaryotic pangenomes, which is closely linked to, but distinct from, genic selection (box 1). MGEs, especially those that are highly mobile, are the clearest possible targets of genic selection. Accordingly, pangenome variation in certain MGEs could be driven by factors besides organism-level selection or genetic drift. Many of these elements are also vectors for the horizontal transmission of otherwise immobile genes, which can further disrupt vertical transmission within a host genome. The potential for genic selection arises in prokaryotic genomes due to this disruption of vertical transmission.

Prophages

Temperate bacteriophages typically insert their genome into the chromosomal genome of their bacterial host, which results in a largely quiescent element referred to as a prophage. This inactive state is maintained by a virus-specific protein that represses the transcription of most phage genes (Fortier and Sekulovic 2013). Prophage induction most commonly proceeds through proteolytic cleavage of this repressor following a DNA-damage SOS response.

While in the prophage state the viral genome is replicated with the host genome. Inducible prophages can exist in this state in bacterial genomes for many generations. Consequently, the bacteriophage fitness during this stage is closely tied to the fitness of the bacterial host. Depending on one's perspective, bacteriophages can be considered independent organisms that face selective pressures like any other parasite. Accordingly, it is debatable whether prophage genetic elements involved solely in viral propagation should be categorized as genic or phage-level selection.

More important than this distinction, however, is the fact that when these prophages are induced, they can allow host genes to be horizontally transferred between cells. This is through a mechanism called transduction, which refers to prokaryotic genes being erroneously packaged into bacteriophage capsids during viral production (Yin and Stotzky 1997). In the case of temperate phages, this can occur when they are incorrectly excised from the host genome so that neighboring genes are transferred, which is referred to as specialized transduction. However, generalized transduction can occur through any bacteriophage, including virulent bacteriophages, as it simply involves the random packaging of prokaryotic DNA during viral assembly, including fragments from anywhere in the genome. Accordingly, all bacteriophages

Glossary

- **Accessory genome**—Genes variably present across a set of genomes. Often further divided into genes that are present at intermediate or low frequencies, which are referred to as the shell and cloud genomes, respectively.
- **Core genome**—Genes shared across all genomes in a set.
- **Gene**—We refer to two meanings of the term gene in this article. First, a gene can refer to a genetic element that encodes an RNA molecule (either coding for a protein or not). This is the intended sense in the context of typical pangenome analyses, where comparisons are limited to such genetic elements. Second, a gene can also refer to a genetic element that is an independent unit of heredity. This sense is used in the context of our discussions of HGT and genic selection. In other words, genetic elements of varying lengths, that need not encode an open-reading frame, can be transferred through HGT mechanisms and be the units of selection.
- **Gene's eye view**—A perspective of evolution where natural selection is interpreted from the viewpoint of individual genetic elements. Most commonly, this perspective is used to view selection acting on individuals from the perspective of their constituent genes. This can be a useful heuristic (e.g., for optimization models), but it is not equivalent to genic selection.
- **Genic selection**—Selection acting at the level of individual genetic elements. This is often in conflict with selection at the level of individual organisms. Genic selection occurs when there is (or potentially could be) differential transmission among genetic elements in a genome.
- **HGT** (also known as lateral gene transfer)—DNA transfer between organisms outside parent–offspring (i.e., vertical) transmission. The primary mechanisms of HGT in prokaryotes are conjugation (direct cell–cell contact, usually to transfer a plasmid), transduction (virus-mediated transfer), transformation (uptake of free DNA), and gene-transfer agents (GTAs; transfer by virus-like particles encoded by prokaryotes).
- **Pangenome**—The set of all genes (including both core and accessory genes) within a set of genomes.
- **Mobile genetic elements**—Units of DNA that encode proteins and/or RNA molecules that enable transposition within and across genomes. Elements which have no or a negative effect on a host are often referred to as selfish genetic elements. Some of these elements are also referred to as “addictive” because they impose pressures on a host to keep the element intact (e.g., a toxin–antitoxin system).
- **Units of selection**—The entities under selection. The level of selection refers to which hierarchical level is occupied by the entities under selection (Okasha 2006). For instance, individual prokaryotic cells are under selection at the organism level. Herein our focus is on differentiating selection acting at the gene level from the level of individual organisms.

represent potential vectors for mobilization of bacterial genetic elements between cells, whereas prophages represent the specific case where the bacteriophage genomes themselves are MGEs.

Prophage diversity has been profiled across numerous lineages, which has highlighted that they can represent varying proportions of host genome content. For instance, one investigation involved profiling prophages across the genomes of 48 *Escherichia* strains and found that they varied in abundance from 2 to 20 unique prophages per genome (Bobay et al. 2013). In the most extreme case, prophage sequences represented 13.5% of a strain's genome. Although such analyses are valuable, it can be difficult to determine whether prophages are inducible, or have become degenerated, without laborious experimental data. One approach for addressing this problem is to compare the ratios of phage with their host in whole-genome and/or metagenomics sequencing data sets. Potential inducible phages can be identified if they have relatively higher read depth compared with their bacterial host (Zünd et al. 2021).

Prophage elements are frequently co-opted by prokaryotic hosts, which is primarily why partially degenerated prophages are commonly found in prokaryotic genomes (Bobay et al. 2014). Certain prophage genes can also be specifically co-opted for the purposes of intercell DNA transfer. GTAs are the likely result of this process, which are phage-like particles that transfer DNA between closely related cells (Lang et al. 2017). GTAs are not MGEs as they typically transfer DNA that does not encode the GTA machinery. As they are often deeply conserved, they are presumed to be strongly host-adaptive, although the precise benefits that GTAs confer remain controversial.

Conjugative and Mobilizable Plasmids

Conjugation refers to transfer of genetic elements through direct cell–cell contact. Most commonly, this transfer is mediated by proteins in type IV secretion-like systems (Zaneveld et al. 2008). The exact mechanisms can differ, but the general steps are that TraG-like proteins within the conjugative system

Table 1
Highlighted Prokaryotic MGE Classes

Element Type	Brief Description
Plasmids	
Conjugative plasmids	Extrachromosomal DNA (typically double-stranded and circularized) that encodes genes necessary for autonomous HGT through conjugation.
Mobilizable plasmids	Plasmids that are dependent on conjugative plasmid machinery for mobilization.
Prophages	
	Bacteriophage genome inserted into a prokaryotic genome. This occurs during the lysogenic phase of a temperate bacteriophage life cycle. Excised and replicated following activation. Many of these elements eventually acquire mutations and undergo pseudogenization.
DNA transposons	
General DNA transposons	Diverse elements that encode a transposase to enable transposition. Primarily transposed within a single genome but can be inserted (or formed directly from integrative conjugative elements) into plasmids to enable HGT.
Insertion sequences	Elements that only encode machinery for transposition. They typically encode a transposase and a regulatory protein. The coding region is flanked by inverted repeats.
Integrative conjugative elements (ICEs)	Also known as conjugative transposons. Elements that encode all required machinery for transposition with a cell, creating circularized double-stranded DNA, and conjugation between cells.
Integrative and mobilizable elements (IMEs)	Also known as mobilizable transposons. Elements that encode their own excision and integration machinery and can insert into (often unrelated) conjugative elements to enable intercell transfer. There are many different mobilization strategies used by these elements across prokaryotes.
Intervening sequence elements	
Archaeal bulge–helix–bulge introns	Archaeal introns found primarily in ribosomal and transfer RNA genes. Contain conserved bulge–helix–bulge motif at intron–exon junction. Many contain open-reading frames that encode homing endonucleases.
Group I introns	Self-splicing introns diversely found across life, but absent in archaea and relatively rare in bacteria. Primarily found in transfer and ribosomal RNA genes in bacteria. They are highly variable at the primary sequence level. They also often encode homing endonucleases that enable them to spread to uninserted genes that share the target sequence.
Group II introns	Similar to group I introns, but almost all encode a reverse transcriptase gene, which enables retrotransposition. They home to specific 30 bp target sites through reverse-transcriptase-mediated insertion. Primarily in horizontally transferred elements (e.g., plasmids).
Inteins	Also known as internal proteins. These are similar to group I introns, except they are excised at the protein level after being translated with the host protein. They often encode homing endonucleases and have been found in a variety of conserved proteins in bacteria.

NOTE.—These classes of elements are not mutually exclusive (e.g., transposons can occur within plasmids, etc.).

initiate mating-pair formation, which is then followed by signaling that transfer can commence, and then the activation of other proteins directly involved in DNA transfer (Frost et al. 2005). Conjugative systems typically only transfer single-stranded DNA, which is produced by a relaxase protein nicking double-stranded DNA at a certain motif, referred to as the origin of transfer (Zaneveld et al. 2008).

Plasmids are the primary elements transferred through conjugation, which themselves represent an extremely diverse set of MGEs. The copy number of each element is plasmid-specific: certain plasmids replicate only during the host genome replication cycle, whereas others replicate at much higher rates. Conjugative plasmids are extrachromosomal elements that encode all necessary components for autonomous conjugation (Ramsay et al. 2016), which in conventional

conjugation systems includes all proteins involved in the general conjugation steps described above. In contrast, mobilizable plasmids encode some, but not all, of the necessary genes required for conjugation. Most often these plasmids are missing genes required for the mating pore formation step. These elements can also be transferred through conjugation by taking advantage of the mating pore formation machinery of conjugative elements in the same cell (Ramsay et al. 2016).

Through conjugation, plasmid inheritance can be partially unlinked from vertical transmission. As for prophages, this shift in transmission could in principle result in selection specifically for enhanced replication of the MGEs, potentially at the expense of host fitness. This is most clearly shown by postsegregational killing systems encoded by certain plasmids,

which enforce plasmid segregation in all viable daughter cells by various mechanisms that result in cell death if the plasmid is absent (Bahl et al. 2009).

Toxin–antitoxin systems are the best characterized postsegregational killing systems, which are often associated with plasmids. These involve production of both a toxin and an antitoxin, the latter of which typically both neutralizes the toxin and represses toxin transcription (Chan et al. 2016). However, the antitoxin degrades faster than the toxin, which leaves daughter cells lacking plasmids that encode an antitoxin to be vulnerable to the toxin. The toxins are typically proteins or RNA molecules that interfere with essential cellular processes, such as DNA replication. Such systems can greatly increase the inheritance stability of plasmids to daughter cells.

Restriction modification systems are involved in cleaving foreign DNA such as the genomes of invading bacteriophages. These systems can sometimes act analogously to toxin–antitoxin systems, particularly in the case of type II restriction–modification systems (Mruk and Kobayashi 2014), to contribute to plasmid maintenance. The analogous proteins to the toxin and antitoxin are an endonuclease and a methyltransferase, respectively. The methyltransferase methylates specific sites in the host genome, which protects it from cleavage by the endonuclease. However, within daughter cells lacking a methyltransferase this protection is soon lost and is once again vulnerable to the endonuclease activity (fig. 2).

Both these general postsegregational killing systems (sometimes referred to as “addictive” systems) highlight the potential for selfish plasmids (or other MGEs) carrying these systems to increase their fitness in a population of cells. They are widespread across prokaryotes and appear to be frequently horizontally transmitted (Kobayashi 2001; Makarova et al. 2009). In addition, there is also evidence that certain restriction–modification systems themselves could be independent MGEs (Furuta et al. 2010; Zheng et al. 2016). Regardless of these characteristics, it is also important to appreciate that these systems often confer many host-beneficial functions and can be integrated into the host chromosome. For instance, toxin–antitoxin systems are also involved in regulating key functions for the host, such as biofilm formation and phage defense (Otsuka 2016). Accordingly, identifying plasmids that encode such systems is not sufficient proof that they are solely selfish elements.

DNA Transposons

Many plasmids contain elements that are capable of self-transposition between regions of a host genome. These elements that have a DNA intermediate are known as DNA transposons, which are the predominant transposon class in prokaryotes. DNA transposons generally contain terminal inverted repeats and encode a transposase enzyme that carries out transposition. Transposases are diverse enzymes and have differing biochemical actions and target sites (Chandler

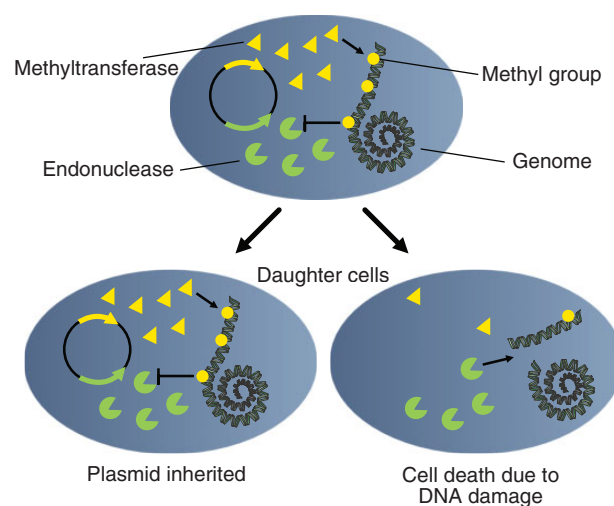


FIG. 2.—Example of a self-propagating system associated with certain MGEs: the general type II restriction modification system. This restriction modification system can enforce postsegregational killing, much like a toxin–antitoxin system. In each cell the blunted arrow indicates enzyme inhibition whereas the regular arrow indicates enzymatic action.

2017). For instance, certain transposases act only on the transposon where they are encoded and are not reused elsewhere. This is likely due to a need for these transposases to bind to the transposon DNA sequence while being translated, which could have evolved to prevent activation of other transposons in the same cell (Chandler 2017).

DNA transposons can also greatly vary in terms of their effect on host fitness. Insertion sequences (ISs) are the simplest DNA transposons, which only encode proteins for transposition. These usually include a transposase and regulatory protein to modulate the transposition activity (Siguier et al. 2014). ISs alone typically confer no benefit to the host and are known to greatly increase in copy number in bacterial species that have undergone recent population bottlenecks, such as intracellular symbionts (Parkhill et al. 2003; Plague et al. 2008). IS proliferation likely reflects the decreased efficacy of selection at the organism-level to counteract the genic selection acting on ISs (Moran and Plague 2004), which would result from the decrease in effective population size. However, it is worth noting that these elements are eventually expunged from the genomes of intracellular symbionts as the genomes become reduced over evolutionary time.

In addition, ISs can modulate the expression of host-beneficial genes by inserting into regulatory regions or alternatively by picking up other genes (Siguier et al. 2014). For example, DNA transposons commonly carry antibiotic resistance genes (Chandler 2017). This is common in composite transposons, which are larger elements flanked by two independent ISs. Clearly such composite MGEs that are host-beneficial could lead to higher replication for each constituent IS. However, composite transposons may not be a stable long-term strategy for individual ISs, due to the increased risk of

being lost through recombination when in this form (Wagner 2006).

Nonetheless, DNA transposons more complex than ISs are frequently observed in prokaryotic genomes. This is particularly true for transposons that are transferred through conjugation and that integrate into the host genome. Integrative conjugative elements (ICEs) and integrative and mobilizable elements (IMEs) are two key examples of such transposons. ICEs encode all machinery for transposition within a cell, as well as for creating circularized double-stranded DNA, and conjugation to different cells (Chandler 2017). In contrast, IMEs encode their own machinery for transposition within a cell, but do not encode conjugative machinery (Guédon et al. 2017). Instead, IMEs preferentially integrate into conjugative plasmids to be transmitted horizontally. These elements, in addition to other commonly genome-integrated MGEs, such as prophages, are the underlying causes of frequently observed “genomic islands” of differential nucleotide composition in regions of prokaryotic genomes (Juhas et al. 2009). In addition, both ICEs and IMEs often lack canonical transposases and instead encode site-specific recombinases, which makes their classification as DNA transposons somewhat debatable (Guédon et al. 2017).

Intervening Sequence Elements

Several subclasses of prokaryotic MGEs are intervening sequence elements, meaning that they are encoded in genes, but are not present in the final gene product. These elements are primarily introns, which are elements spliced from RNA molecules. Similar elements, called inteins, are removed post-translationally instead.

Archaeal bulge–helix–bulge introns are the simplest type of intron found in prokaryotes (Zimmerly 2005). They are characterized based on the presence of a conserved motif consisting of 3 bp bulges flanking a 4 bp helix. In many cases they have been identified as encoding homing endonucleases (Jay and Inskip 2015). These key enzymes cleave loci without the intron, which commences double-stranded DNA repair pathways that use the intron-containing locus as a template for repair. Through this mechanism, elements can spread within a genome. The target sites for homing endonucleases are typically conserved structures, which means that they are present in closely related genomes. These conserved sites are primarily found in transfer and ribosomal RNA genes, where these and certain other introns are mainly located (Zimmerly 2005).

Group I introns are similar elements that contrast by having self-splicing activity, although this is typically dependent on intron and/or host-encoded factors (Hausner et al. 2014). Besides maturase proteins, which are involved in the self-splicing action, these introns also often encode homing endonucleases. However, although group I introns have been observed across numerous prokaryotic genomes, such

as in *Thermotoga neapolitana* and *Bacillus cereus*, they are much rarer than other intron types.

Group II introns are much more common in prokaryotes. These elements also have self-splicing activity and encode maturase proteins, but they are mobilized through a reverse transcriptase mechanism (Zimmerly 2005). Group II introns can also encode homing endonucleases, most commonly the LAGLIDADG subtype (Zimmerly and Semper 2015). They are also primarily found in MGEs such as plasmids and prophages and are almost never found in highly conserved genes.

Inteins, or “internal proteins,” differ from introns in that they are translated with the entire protein sequence and then self-excise (Raghavan and Minnick 2009). They have been observed in a range of conserved proteins, such as DNA polymerases, helicases, and gyrases. These elements also commonly encode homing endonucleases, which enable their mobility within a genome. These homing endonucleases make a single-stranded DNA break at the site of the intein in the genome, which enables the element to be recombined into different locations during DNA repair.

The effects of these elements on host fitness are difficult to assess. At least in some cases intervening sequences likely provide benefits to prokaryotic hosts (Sandegren and Sjöberg 2007; Edgell et al. 2011). However, it is possible that many of these elements are neutral or even slightly deleterious from the perspective of the host. This is particularly true for group II introns, which are more commonly transferred horizontally as they are often present in horizontally transferred elements, and for those that encode homing endonucleases. Indeed, the idea that certain prokaryotic intervening sequences are primarily selfish has long been hypothesized (Morinaga et al. 2002).

General Strategies for MGEs

All the above MGEs have the potential to be under genic selection. However, due to the deletional bias in prokaryote genomes (box 2) when these elements do not provide a benefit to the host, they quickly accumulate deletions and are lost (Kuo and Ochman 2009). This genomic environment means that such elements must leverage more than genetic drift to be retained over long periods, particularly if they have deleterious impacts on host fitness.

It has long been hypothesized that mobile elements that have neutral or deleterious impacts on host fitness are maintained in populations through sufficiently high rates of HGT. This question was recently investigated through a modeling approach that focused on DNA transposons, conjugative plasmids, and toxin–antitoxin modules across prokaryotes (Iranzo and Koonin 2018). This investigation inferred that many such MGEs are likely deleterious over long timescales. In addition, they identified a clear positive link between the negative impact on host fitness of an MGE and the mobility needed to

Box 2: Deletional bias in prokaryotic genomes

Genome size is tightly linked to the number of genes in prokaryotes as typically there is only a small proportion of noncoding DNA (Mira et al. 2001). Overall, this is unlikely to be due to selection for genome streamlining to optimize replication rates, as genome sizes are independent from prokaryotic growth rates (Westoby et al. 2021). Exceptions include certain lineages where non-host-beneficial DNA is lost more quickly than expected by genetic drift alone (Kuo and Ochman 2009). However, in general, compact prokaryotic genomes can be explained by a mutational bias toward deletions. In particular, in most prokaryotes the rate of deletions to insertions is approximately 10:1 (Kuo and Ochman 2009). This bias means that, based on genetic drift alone, DNA that provides no benefit to host fitness is much more likely to be lost through deletions than expanded through insertions.

The rapid deletion of deteriorating, nonfunctional genes, or pseudogenes, in prokaryotes clearly demonstrates this phenomenon. Approximately 1–5% of genes in prokaryotic genomes are considered pseudogenes (Liu et al. 2004), which generally rapidly accumulate deletions (Kuo and Ochman 2009). There are some important exceptions, such as genomes of *Rickettsia prowazekii* and *Mycobacterium leprae*, which contain higher proportions of pseudogenes, and larger intergenic regions (Mira et al. 2001). These cases likely represent the early stages of genome degradation in endosymbionts. The relatively high number of pseudogenes in these cases is likely due to the decreased efficacy of negative selection in maintaining these genes. This decreased efficacy of selection is expected to follow from the decrease in effective population size associated with endosymbionts.

Small deletions have primarily been the focus of the work investigating deletional bias. These small deletions profiled from pseudogenes are disproportionately multiples of three, which could be due to errors in DNA repair associated with microhomologies between codons (Danneels et al. 2018). It remains to be seen whether this bias is restricted to former protein-coding genes or not. However, in addition to these short deletions, in experimental settings large-scale deletions have also been observed over short time-scales (Nilsson et al. 2005). In many cases these deletions can result in the loss of multiple genes with little detectable phenotypic effect (under the tested conditions). The relative frequency of such large-scale deletions relative to their smaller counterparts remains to be determined in natural communities of free-living prokaryotes. A better understanding of these and other related questions would provide improved insights into the selection pressures facing MGEs.

maintain it in the population. Overall, the transmission rate of these MGEs was inferred to be higher than the minimum level required for their maintenance. This observation is consistent with high transmission rates being highly adaptive for MGEs.

This mechanism of selfish MGE maintenance suggests that a decrease in HGT rates could enable many MGEs to be purged. Why then is HGT so rampant across prokaryotes? From the host perspective, one reason is that it enables host-beneficial genes to be rapidly acquired, which could be particularly needed in unstable environments. Another explanation for why HGT is commonplace is that it is necessary to counteract Muller's ratchet in asexual populations (Takeuchi et al. 2014). Asexual populations will periodically fix deleterious mutations which are unlikely to be purged without recombination. The accumulation of these deleterious mutations leads to mutational meltdown: the deterioration of a population due to the deleterious load of mutations exacerbating the problem by decreasing the population size. HGT introduces genetic diversity, including functional copies of genes that may have acquired deleterious mutations in the recipient genome, that can be recombined to purge deleterious mutations from the genome (Takeuchi et al. 2014). It has been estimated that even the low levels of HGT required to avoid Muller's ratchet are sufficient for selfish MGEs to persist

(Iranzo et al. 2016; Van Dijk et al. 2020). One example of potentially deleterious mutations that arise in lineages with ineffective selection and low rates of HGT is the genome-wide bloom of ISs in intracellular symbionts (Moran and Plague 2004). Accordingly, to some degree the cost of receiving selfish MGEs through HGT may be worth the introduction of genetic diversity to purge within-genome selfish MGEs.

Although most work investigating selfish MGEs has been based on modeling and genomic comparisons, corroborating results have been found in several experimental and natural systems (Brockhurst et al. 2019). One major advantage of such approaches is that they enable the spread of MGEs to be studied over short (e.g., days to months) rather than long (e.g., millions of years) evolutionary timescales. In particular, conjugative plasmids can rapidly spread through populations in the absence of selection at the host level (Lundquist and Levin 1986; Eberhard 1990). One recent example is a plasmid encoding the mercury resistance operon *mer* (Stevenson et al. 2017). This plasmid became fixed in populations of *Pseudomonas fluorescens* over eight days regardless of whether there was active selection for mercury resistance. In contrast, when the gene was chromosomally encoded it only reached fixation when mercury resistant cells were positively selected. This distinction highlights that this gene is

unlikely to have spread on the plasmid due to providing some other host-beneficial trait. Instead, it suggests that the HGT rate of the plasmid was sufficient to explain its rapid spread through the population. This example is particularly intriguing as this plasmid was expected to have a negative impact on host fitness in the absence of mercury (Stevenson et al. 2017).

One of us recently investigated related questions in the context of the gut microbiome of Fiji islanders (N'Guessan et al. 2021). By mapping metagenomic reads from stool samples to a data set of known mobile genes, our team was able to identify single-nucleotide variants in these genes. Many of these genes were enriched for rare mutations, which could result from several deviations from a standard neutral evolutionary model: a recent population expansion, a selective sweep, and/or purifying selection that eliminated variation. The observed excess of rare mutations in these mobile genes was recapitulated in simulations where the mobile genes were neutral to host fitness, and with only stabilizing selection on the overall genome size. Based on these simulations, the HGT rate of each gene was sufficient to account for the frequencies of mutations segregating per gene. This result suggests that, at least on short time-scales (i.e., on the order of a human lifespan), mobile gene sequence evolution can be explained without invoking any adaptive value to either human or bacterial hosts.

Taken together, these results highlight that at least in some circumstances high mobility can be a sufficient strategy for the maintenance and spread of MGEs. An intriguing complementary strategy for chromosomal MGEs is for them to insert in safe havens: locations in the genome where they are less likely to be lost and/or where they are unlikely to deleteriously affect host fitness (Werren 2011). There is substantial evidence for this latter trend, but it remains unclear to what degree genic selection drives insertions into safe havens that confer protection from inactivation or deletion.

Nonetheless, the distribution of certain MGEs could be compatible with a model of genic selection driving insertion into such protected regions. For example, as discussed in the above section, several classes of introns are enriched in RNA genes. These elements are primarily thought to be present in these genes because they contain highly conserved insertion targets, and also because the self-splicing process could interfere with the translation of protein-coding genes (Hausner et al. 2014). However, in cases where introns are in single-copy RNA genes (such as the 16S rRNA gene in many genomes), inactivation or deletion of the intron would be complicated due to the possibility of disrupting the essential function. Although this is less true for multicopy genes, introns and inteins in any gene could still be partially protected from deletions. This is because deletions that either disrupt the splicing/excising processes or that otherwise create missense products could deleteriously affect host fitness. Accordingly, to some degree the genomic distribution of introns and inteins could represent locations where they are partially protected from deletion.

Several other MGEs, such as temperate phages, are known to insert within or nearby conserved genes, particularly transfer RNA genes (Williams 2002). However, this is more likely to reflect integration into redundant genes which contain conserved target sites for insertion across diverse genomes. In other words, prophage locations across genomes likely reflect the second meaning of safe haven in this context: a location where they are unlikely to negatively affect host fitness (Bobay et al. 2013).

Although inserting into nonessential regions could be considered beneficial for individual MGEs, it is also beneficial for hosts to limit MGE insertion to such restricted genomic regions. Indeed, in many organisms there are biases in where MGEs insert (Touchon and Rocha 2016). For instance, in *E. coli*, MGE insertions are primarily limited to several hotspots throughout the genome, which are frequently located near the terminus of replication. In certain cases, accessory genes are relegated to secondary chromosomes as well (Cooper et al. 2010), which could help protect essential functions from MGEs. Although this is an enticing hypothesis to explain prokaryotic genome organization, it could also be possible that these genomic biases in gene content could reflect selection for robustness to large-scale deletions (Hosseini and Wagner 2018). Nonetheless, these examples highlight the potential for the localization of MGE insertion sites to be driven by both organism-level and genic selection. Disentangling these two forms of selection is a nontrivial problem, as discussed below.

Potential Genic Selection on Host-Beneficial Genetic Elements

Above we primarily discussed genic selection acting on MGEs that are deleterious or neutral to the host, which are commonly referred to as selfish DNA (Doolittle and Sapienza 1980; Orgel and Crick 1980), or genetic parasites (Iranzo et al. 2016). Although these represent the clearest targets of genic selection in prokaryotic genomes, many other elements are potential targets as well. In particular, MGEs that are adaptive for hosts, at least under certain conditions, could also be targets of genic selection. This is also true for adaptive genes that are commonly horizontally transferred but are not strictly classified as MGEs.

In many ways this is an obvious observation: clearly the rapid spread of an adaptive genetic element is beneficial from the perspective of both the host and the genetic elements. This is only controversial to the extent that the validity of the gene's eye view of evolution is disputed (box 1). However, our point here is that in many cases such host-beneficial genetic elements could be themselves, at least partially, the units of selection. We are not claiming that organism-level selection is not also active: undoubtedly selection acting on an organism's phenotype is organism-level selection, even if the phenotype is modulated by an MGE.

However, it would be incorrect to infer from this that such an MGE was only distributed in a population through organism-level selection. Instead, the genetic element in question might disproportionately benefit from the adaptive advantage that it provides to their host. Such differential fitness at the gene-level could arise through several different mechanisms. For instance, an element would have increased fitness relative to other genes if the increased frequency of the element due to selection at the organism-level also disproportionately increased the probability that it would be horizontally transmitted to future uncolonized populations. In other words, such elements should also be considered to be under genic selection if they benefit more from the increase in host fitness compared with most other genes encoded by the host (Okasha 2006).

Temperate phages are a noncontroversial example of such dynamics. Prophages frequently encode genes that confer benefits to their prokaryotic host, particularly for functions related to virulence (Fortier and Sekulovic 2013). Although these functions can be beneficial for hosts in the short-term, clearly the spread of inducible prophage sequences disproportionately benefits prophages in the long term, to the detriment of their bacterial hosts. In other words, there is a temporal offset in terms of which unit is under selection: the host during the lysogenic phase and the phage during the lytic phase. Accordingly, selection at both levels is responsible for the distribution of phages that have undergone both lysogenic and lytic phases in a community. Although this is an extreme example, similar dynamics could be at play for any other class of MGE which possesses the necessary characteristics to be considered an independent unit of selection.

The above example corresponds to a case where organism-level selection leads to vertical transmission of an element, which in turn leads to increased opportunities for the element to spread horizontally under genic selection. However, in different contexts the converse could also be true: horizontal transmission of an element could aid a slightly host-beneficial element to spread through a population more predictably. This has been clearly shown through past work that investigated to what degree the fixation probability of a newly introduced MGE in a population is dependent on the rate of horizontal transmission, compared with the adaptive advantage conferred to hosts where it is found. A theoretical investigation into this question found that the fixation probability of such an MGE is affected equally by these two factors (Tazzyman and Bonhoeffer 2013). In other words, high rates of horizontal transmission (at least when the elements are rare) can contribute to the fixation probability of an MGE just as much as the selective benefit to the host. Disentangling the relative contributions of these two factors is challenging in nature, but in theory this relationship is quite simple (fig. 3). These findings highlight that horizontal transmission rates can in principle be the main determinant of whether a newly introduced host-beneficial MGE would be

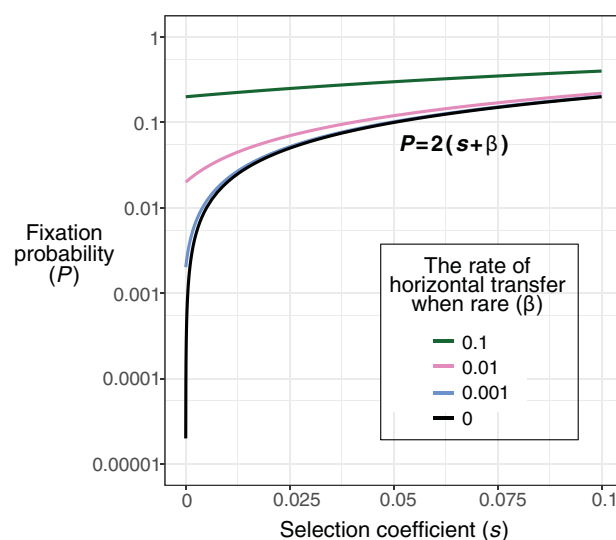


Fig. 3.—Illustrative examples of fixation probabilities for a rare MGE in a population with varying horizontal transmission rates and adaptive benefits to the host. These probabilities are based on the findings that selection (s) and horizontal transmission (β) can proportionally contribute to a mobile element's probability of fixation (P), which was formalized in the equation: $P = 2(s + \beta)$ (Tazzyman and Bonhoeffer 2013). We computed the fixation probability based on varying the values of the horizontal transfer rate and selection coefficient in this equation to generate these curves. These examples highlight that classifying an accessory gene as adaptive or selfish is overly simplistic: in many cases slightly host-beneficial genes might be spread faster than expected given the organism-level benefit conferred.

fixed in a population, particularly for those that are only weakly beneficial. Although this is noteworthy, it is important to appreciate that this result is for an idealized population, which was made up of a single species with a fixed census population size. It remains to be seen how important the rate of horizontal transmission and host-level selective benefit are to the spread of weakly beneficial MGEs in natural communities, with complicating factors such as fluctuating selective pressures, spatial structure, and interspecies interactions.

Despite its simplicity, this model could be relevant to the observed distribution of many accessory genes across prokaryotic pangenomes, as many are predicted to be only weakly beneficial to their hosts (Bobay and Ochman 2018). Although the connection to this model is most clearly seen for genes encoded on MGEs, it is possible that it could also be relevant to weakly beneficial genetic elements that are not self-mobilizable. For instance, even without mobilization machinery, certain elements might be more likely to be transmitted horizontally through transformation if, for example, they encode host-specific DNA uptake sequences (Zaneveld et al. 2008). Elements could also differ in how likely they are to be successfully recombined into the genome of a closely related host, which could depend on the genomic location and accessibility of the element. Such dynamics could influence which elements are expected to spread through populations

via gene-specific sweeps (Shapiro 2016): in addition to providing host-level benefits these elements might be localized in recombination hotspots and be more likely than other elements to spread between the genomes of closely related organisms.

The above discussion focused on interactions between genic and organism-level selection that could be important to understand prokaryotic pangenome structure within a species. However, similar dynamics could be occurring for MGEs that can be transmitted across multiple host species, which would provide protection against being lost from an environment. This potential strategy could be particularly relevant for weakly or transiently host-beneficial MGEs, which might be frequently lost and regained in specific populations. Such a process could be related to recent observations that phage host ranges are wider than traditionally thought. This conclusion was based on a large-scale assembly of phage genomes from human metagenomics data sets (Camarillo-Guerrero et al. 2021). The authors were able to characterize the host tropism of the profiled phages by defining phage clusters that shared at least 90% identity over a large region of alignment. Of the 21,012 clusters identified, 36% were predicted to infect more than one bacterial species.

Conjugative plasmids are a clearer example of where wider MGE host tropism could reflect gene-level adaptation to persist in a community. This is particularly true for plasmids that encode genes that are not widely connected to gene interaction networks, and that are more likely to be host-beneficial in diverse genetic backgrounds (Jain et al. 1999; Li et al. 2020; Novick and Doolittle 2020). For instance, a high proportion of genes with environment-specific carbohydrate metabolism functions encoded on plasmids have previously been identified in activated sludge samples (Sentchilo et al. 2013). The authors hypothesized that such functions would be adaptive for many members of the community and are likely linked to multiple bacterial species.

To explore this general question more directly, the link between plasmid host tropism and persistence in a community was investigated in an experimental system (Hall et al. 2016). The authors created mixed communities of *P. fluorescens* and *Pseudomonas putida* in addition to monocultures of each species. They investigated whether the persistence of a plasmid encoding the *mer* mercury resistance operon was heightened in the mixed communities in the presence of mercury. In the single-species communities the *mer* locus frequently integrated into the *P. putida* genome and the plasmid was lost, whereas the plasmid persisted in *P. fluorescens*. In contrast, in the mixed community the plasmid persisted in both species, which meant that accessory genes also encoded on the plasmid were retained in *P. putida* that otherwise would be lost when *mer* integrated into the genome. This example highlights that cross-species transmission can help maintain MGEs and genes that would otherwise be lost.

It is sometimes hypothesized that cases of host-beneficial MGEs identified across multiple species in a community reflect selection for fitness at the community-level. This could be mediated through enhanced cooperation between species that share beneficial MGEs in a common gene pool (Norman et al. 2009). Such cooperation could make the overall community more robust to environmental changes (Heuer and Smalla 2012). These possibilities were considered in the context of the above *Pseudomonas* experiments, which the authors considered unlikely (Hall et al. 2016). We also think such dynamics are unlikely to reflect community-level selection. Instead, we believe that a combination of genic and organism-level selection could likely account for these observations, without requiring selection to act on higher levels of complexity. Nonetheless, the relevance of community-level selection in maintaining MGEs across multiple species remains an exciting area for future exploration.

A final point regarding genic selection on host-beneficial elements is that in many cases adaptive genes may be in competition with other genes that fulfill the same functional niche (Francino 2012). Incompatibilities that arise between plasmids with similar replication mechanisms likely reflect such competition (Frost et al. 2005). A less clear example is that of two siderophores in *Salinispora*, which are highly mobile and functionally identical, but are never found together (Bruns et al. 2018; Hall et al. 2020). Although in this case the genes are thought to provide identical benefits, gene–gene competition could in principle occur with somewhat functionally distinct genes as well. If a functionally similar, yet less beneficial rare gene, had a higher rate of horizontal transmission compared with another rare gene, they could nonetheless have similar fixation probabilities (fig. 3) and remain in conflict.

Outlook

Herein we have highlighted the role of realized and potential genic selection occurring across prokaryotes. One major observation is that adaptive genes distributed across prokaryotic pangenomes could be spread by both selection at the level of the gene and the individual cell. The relative importance of each level of selection will depend primarily on the mobility of the genetic element in question (Jalasvuori and Koonin 2015) and to what degree it possesses the necessary features to be considered an independent unit of selection.

Highly mobile elements are more likely to be targets of genic selection given the higher potential for them to be differentially transmitted from the rest of the host genome. In many cases such elements likely confer an adaptive advantage to the host. However, this is no guarantee of long-term retention as the impact that an element has on host fitness is expected to be context-dependent (Domingo-Sananes and McInerney 2021). Consequently, in most cases it is hard to conclude that an element has solely one type of effect on host

fitness. For instance, it may be that in an unstable environment certain elements could sometimes be highly host-beneficial, but slightly deleterious or neutral otherwise. A high rate of horizontal transmission and broad host tropism could help such elements be retained in a community during periods when they are not beneficial.

Selfish elements are also widespread in prokaryotic pangenomes. For instance, selfish MGEs were the third most frequent class of accessory gene identified across 228 *E. coli* genomes, after those with unknown functions and metabolism-related genes (McInerney et al. 2017). As discussed above, selfish MGEs are inevitable with even low levels of HGT and so novel and newly introduced selfish MGEs are expected to continually arise. Accordingly, rare accessory genes of unknown function may be disproportionately enriched for selfish MGEs, although this remains to be determined.

Although we argue that such elements are important to consider, this point should not be mistaken to mean that genic selection alone can account for most variation across prokaryotic genomes. Indeed, the efficacy of selection, which is determined by the effective population size, is strongly positively correlated with the number of accessory genes in a species (Sela et al. 2016; McInerney et al. 2017). An important point regarding such analyses is that pangenome structure can be confounded by subpopulations or “ecotypes” within the overall species. Nonetheless, this is consistent with a model where most accessory genes are weakly host-beneficial and are only fixed in populations with sufficiently high efficacy of selection (Bobay and Ochman 2018). That is not to say that genic selection is irrelevant for such accessory genes, but only that genic selection alone is unlikely to account for the large-scale variation in pangenome sizes across prokaryotic species.

In addition, although this review has primarily focused on known MGEs, it is important to emphasize that genic selection could be relevant to the distribution of non-self-mobilizable genetic elements as well. Most prokaryotic genes have likely been encoded by an MGE at some point, simply due to the ubiquity of MGEs and ample opportunity for such interactions to arise over evolutionary time (Eberhard 1990). Accordingly, the extent to which genic selection would have a significant impact on the distribution of a nonmobile gene would be dependent on how long and how frequently the gene was transmitted by MGEs. It would also be dependent on the degree to which the gene was transmitted through transformation, which itself could be impacted by a range of factors. For instance, in certain lineages, the presence of specific DNA-uptake sequences can lead to higher rates of transformation between closely related organisms (Zaneveld et al. 2008). Although this is an intriguing process, transformation rates in general are likely more impacted by more practical matters, such as whether other potential recipients in the community are competent and share sufficient sequence homology for frequent recombination.

The timescales under consideration are also relevant when assessing the importance of genic selection. Genomic comparisons of taxa that have diverged over deep evolutionary time, compared with investigations of genomes within the same contemporary natural communities, may lead to different inferences about the relative importance of genic selection. The diverse interactions a particular genetic element can have with host fitness is why others have recommended that assessments of the fitness impacts of such elements be based on long timescales only (Iranzo and Koonin 2018). Although this is a sensible approach, it will not necessarily provide insight into the causal reasons for why a genetic element is present in a genome. Considering shorter eco-evolutionary time scales may help reveal the mechanisms that lead to long-term patterns. As discussed above, accessory genes could be at least partially retained through the action of genic selection, particularly during periods where they are not beneficial for the host, which might be obfuscated by looking at long timescales alone. More generally, genic selection is important to consider over short timescales as elements of the most practical concern, such as antibiotic resistance elements, may only be transiently adaptive for the host organism (i.e., in the presence of antibiotics).

Given that genic selection may leave only ambiguous traces in prokaryotic genomes, how can the relative contributions of genic and individual-level selection be determined? This is a colossal task but continued work in simple, experimental systems could help explore fundamental questions in this area. For instance, key work is being performed with conjugative plasmids that encode a selective marker that can be fixed by either individual or gene-level selection, depending on the conditions (Stevenson et al. 2017). One key observation of this work (as described under General Strategies for MGEs) is that genome-wide variation was maintained when the tested plasmid swept through the population horizontally (i.e., through genic selection), but was lost when there was positive selective pressure for individual cells to possess the plasmid. In this latter case, the signature reflected a standard genome-wide selective sweep.

These divergent signatures could be valuable for identifying the action of each level of selection, but there is a major complication in natural systems: diversity could also be maintained through high recombination in the face of a selective sweep. This is considered to be one mechanism underlying gene-specific sweeps in natural microbial populations (Shapiro 2016). Such examples could also be driven by other mechanisms, such as soft sweeps. Regardless, gene sweeps are typically considered to be examples of organism-level adaptive genes, including those conferring niche adaptation. It is noteworthy that this signature would be challenging to distinguish from that of genic selection alone.

Keeping this challenge in mind, continued investigation of natural microbial genomic diversity with shotgun metagenomics could be used to identify putative cases of genic

selection. This work could be conducted analogously to the experimental work described above (Stevenson et al. 2017) by focusing on genetic elements that are predicted to be host-beneficial under certain conditions. The prevalence of these elements could then be compared in natural communities with varying selection pressure for the host-level phenotype. For instance, the distribution of a plasmid encoding the *mer* operon could be profiled in representative samples from an environmental gradient of low to high mercury levels. Genic selection for a high rate of horizontal transmission could be responsible for any cases where the plasmid is widely prevalent in the absence of selection for mercury resistance. More generally, this approach could be used to identify widely horizontally transferred elements that are not associated with a genome-wide selective sweep. Such an approach could be especially valuable when used in concert with other promising technologies to produce high-quality metagenome-assembled genomes (Douglas and Langille 2019). Indeed, several recent projects have demonstrated the potential for studying prokaryotic pangenomes in natural environments using shotgun metagenomics data (Delmont and Eren 2018; Utter et al. 2020), which has been referred to as profiling the meta-pangenome (Ma et al. 2020).

Regardless of which approaches are used to quantify its importance, an appreciation of how genic selection may affect pangenome variation raises many questions. In particular, models based on more traditional forms of intragenomic conflict, such as meiotic drive (fig. 1a), could also be relevant to understanding prokaryotic pangenomes. For example, a long-standing question in the intragenomic conflict literature is why phenotypes are largely optimized at the level of individuals despite widespread selfish genetic elements. Such elements are expected to disrupt optimal phenotypes at the organism-level to enhance their own transmission. One potential explanation for this observation is that selection throughout the rest of the genome could suppress selfish elements and prevent them from perturbing phenotypes. This selection could arise if most genes were solely vertically transmitted, and thus tied to the individual. This model has been called the “parliament of genes” (Leigh 1971), and it has been shown to be a robust explanation under diverse conditions (Scott and West 2019). However, there is one exception: the suppression of selfish elements is expected to occur less reliably as the genomic proportion of these elements increases. With most forms of intragenomic conflict this is not an issue, as selfish elements tend to be restricted to a small number of loci. But with certain forms of HGT this is not the case. For example, the F plasmid is well known to transfer entire bacterial chromosomes (at least in the lab) (Virolle et al. 2020), whereas a recently discovered phage of *Staphylococcus aureus* has been shown to transfer large proportions of the host genome, through a mechanism called lateral transduction (Chen et al. 2018). Accordingly, in some cases genic selection contributing to pangenome variation may not be curbed as much as traditional forms of

intragenomic conflict. These exceptional cases could lead to a hung parliament of genes.

This is but one example of how an understanding of the potential for genic selection in the context of pangenome variation brings to light novel questions. Granted, many of these questions have no easy answers: disentangling the action of both genic and individual-level selection on the same elements is no simple task. Nonetheless, continued investigation of the potential for genic selection across prokaryotes is needed to develop a clearer understanding of pangenome variation.

Acknowledgments

G.M.D. is funded by a National Sciences and Research Council (NSERC) of Canada Post-Doctoral Fellowship and B.J.S. is supported by an NSERC Discovery Grant. We would like to thank the following individuals for helpful feedback and discussions regarding this manuscript: Joe Bielawski, Louis-Marie Bobay, Ford Doolittle, Jacob Nearing, and two anonymous reviewers. The image of DNA in figure 2 was taken from the Smart Servier Medical Art repository, which is under a Creative Commons Attribution 3.0 Unported License.

Data Availability

There are no new data associated with this article.

Literature Cited

- Ågren JA. 2016. Selfish genetic elements and the gene’s-eye view of evolution. *Curr Zool.* 62(6):659–665.
- Andreani NA, Hesse E, Vos M. 2017. Prokaryote genome fluidity is dependent on effective population size. *ISME J.* 11(7):1719–1721.
- Azarian T, Huang IT, Hanage WP. 2020. Structure and dynamics of bacterial populations: pangenome ecology. In: Tettelin H, Medini D, editors. *The pangenome: diversity, dynamics and evolution of genomes.* Cham, Switzerland: Springer. p. 115–128.
- Bahl MI, Hansen LH, Sørensen SJ. 2009. Persistence mechanisms of conjugative plasmids. In: Gogarten MB, Gogarten JP, Olendzenski L, editors. *Horizontal gene transfer: genomes in flux.* New York: Humana Press. p. 73–102.
- Bobay LM, Ochman H. 2018. Factors driving effective population size and pan-genome evolution in bacteria. *BMC Evol Biol.* 18(1):153.
- Bobay LM, Rocha EPC, Touchon M. 2013. The adaptation of temperate bacteriophages to their host genomes. *Mol Biol Evol.* 30(4):737–751.
- Bobay LM, Touchon M, Rocha EPC. 2014. Pervasive domestication of defective prophages by bacteria. *Proc Natl Acad Sci U S A.* 111(33):12127–12132.
- Brockhurst MA, et al. 2019. The ecology and evolution of pangenomes. *Curr Biol.* 29(20):R1094–R1103.
- Bruns H, et al. 2018. Function-related replacement of bacterial siderophore pathways. *ISME J.* 12(2):320–329.
- Busby B, Kristensen DM, Koonin EV. 2013. Contribution of phage-derived genomic islands to the virulence of facultative bacterial pathogens. *Environ Microbiol.* 15(2):307–312.
- Camarrillo-Guerrero LF, Almeida A, Rangel-Pineros G, Finn RD, Lawley TD. 2021. Massive expansion of human gut bacteriophage diversity. *Cell* 184(4):1098–1109.e9.

- Chan WT, Espinosa M, Yeo CC. 2016. Keeping the wolves at bay: anti-toxins of prokaryotic type II toxin-antitoxin systems. *Front Mol Biosci*. 3:9.
- Chandler M. 2017. *Prokaryotic DNA transposons: classes and mechanism*. Chichester: John Wiley & Sons, Ltd.
- Chen J, et al. 2018. Genome hypermobility by lateral transduction. *Science* 362(6411):207–212.
- Cooper VS, Vohr SH, Wrocklage SC, Hatcher PJ. 2010. Why genes evolve faster on secondary chromosomes in bacteria. *PLoS Comput Biol*. 6(4):e1000732.
- Danneels B, Pinto-Carbó M, Carlier A. 2018. Patterns of nucleotide deletion and insertion inferred from bacterial pseudogenes. *Genome Biol Evol*. 10(7):1792–1802.
- Dawkins R. 1982. *The extended phenotype: the gene as the unit of selection*. Oxford: Oxford University Press.
- Delmont TO, Eren EM. 2018. Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. *PeerJ* 6:e4320.
- Domingo-Sananes MR, McInerney JO. 2021. Mechanisms that shape microbial pangenomes. *Trends Microbiol*. 29(6):493–503.
- Doolittle WF, Sapienza C. 1980. Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284(5757):601–603.
- Douglas GM, Langille MGI. 2019. Current and promising approaches to identify horizontal gene transfer events in metagenomes. *Genome Biol Evol*. 11(10):2750–2766.
- Eberhard WG. 1990. Evolution in bacterial plasmids and levels of selection. *Q Rev Biol*. 65(1):3–22.
- Edgell DR, Chalamcharla VR, Belfort M. 2011. Learning to live together: mutualism between self-splicing introns and their hosts. *BMC Biol*. 9(1):22.
- Fortier L-C, Sekulovic O. 2013. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* 4(5):354–365.
- Francino MP. 2012. The ecology of bacterial genes and the survival of the new. *Int J Evol Biol*. 2012:394026.
- Frost LS, Leplae R, Summers AO, Toussaint A. 2005. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol*. 3(9):722–732.
- Furuta Y, Abe K, Kobayashi I. 2010. Genome comparison and context analysis reveals putative mobile forms of restriction-modification systems and related rearrangements. *Nucleic Acids Res*. 38(7):2428–2443.
- Godfrey-Smith P. 2009. *Darwinian populations and natural selection*. Oxford: Oxford University Press.
- Gould S. 2002. *The structure of evolutionary theory*. Cambridge: Harvard University Press.
- Guédon G, Libante V, Coluzzi C, Payot S, Leblond-Bourget N. 2017. The obscure world of integrative and mobilizable elements, highly widespread elements that pirate bacterial conjugative systems. *Genes (Basel)*. 8(11):337.
- Hall JPJ, Wood AJ, Harrison E, Brockhurst MA. 2016. Source-sink plasmid transfer dynamics maintain gene mobility in soil bacterial communities. *Proc Natl Acad Sci U S A*. 113(29):8260–8265.
- Hall RJ, Whelan FJ, McInerney JO, Ou Y, Domingo-Sananes MR. 2020. Horizontal gene transfer as a source of conflict and cooperation in prokaryotes. *Front Microbiol*. 11:1569.
- Hausner G, Hafez M, Edgell DR. 2014. Bacterial group I introns: mobile RNA catalysts. *Mob DNA*. 5(1):8.
- Heuer H, Smalla K. 2012. Plasmids foster diversification and adaptation of bacterial populations in soil. *FEMS Microbiol Rev*. 36(6):1083–1104.
- Hosseini SR, Wagner A. 2018. Genomic organization underlying deletional robustness in bacterial metabolic systems. *Proc Natl Acad Sci U S A*. 115(27):7075–7080.
- Hull DL. 1980. Individuality and selection. *Annu Rev Ecol Syst*. 11(1):311–332.
- Innamorati KA, Earl JP, Aggarwal SD, Ehrlich GD, Hiller NL. 2020. The bacterial guide to designing a diversified gene portfolio. In: Tettelin H, Medini D, editors. *The pangenome: diversity, dynamics and evolution of genomes*. Cham, Switzerland: Springer. p. 51–87.
- Iranzo J, Koonin EV. 2018. How genetic parasites persist despite the purge of natural selection. *EPL* 122(5):58001.
- Iranzo J, Puigbò P, Lobkovsky AE, Wolf YI, Koonin EV. 2016. Inevitability of genetic parasites. *Genome Biol Evol*. 8(9):2856–2869.
- Jain R, Rivera MC, Lake JA. 1999. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci U S A*. 96(7):3801–3806.
- Jalasvuori M, Koonin EV. 2015. Classification of prokaryotic genetic replicators: between selfishness and altruism. *Ann N Y Acad Sci*. 1341:96–105.
- Jay ZJ, Inskip WP. 2015. The distribution, diversity, and importance of 16S rRNA gene introns in the order Thermoproteales. *Biol Direct*. 10:35.
- Juhas M, et al. 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev*. 33(2):376–393.
- Kobayashi I. 2001. Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res*. 29(18):3742–3756.
- Koonin EV, Wolf YI. 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res*. 36(21):6688–6719.
- Kuo C-H, Ochman H. 2009. Deletional bias across the three domains of life. *Genome Biol Evol*. 1:145–152.
- Lang AS, Westbye AB, Beatty JT. 2017. The distribution, evolution, and roles of gene transfer agents in prokaryotic genetic exchange. *Annu Rev Virol*. 4(1):87–104.
- Leigh EG. 1971. *Adaptation and diversity: natural history and the mathematics of evolution*. San Francisco: Freeman.
- Lewontin RC. 1970. The units of selection. *Annu Rev Ecol Syst*. 1(1):1–18.
- Li L, et al. 2020. Plasmids persist in a microbial community by providing fitness benefit to multiple phylotypes. *ISME J*. 14(5):1170–1181.
- Liu Y, Harrison PM, Kunin V, Gerstein M. 2004. Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. *Genome Biol*. 5(9):R64.
- Lundquist PD, Levin BR. 1986. Transitory derepression and the maintenance of conjugative plasmids. *Genetics* 113(3):483–497.
- Ma B, France M, Ravel J. 2020. Meta-pangenome: at the crossroad of pangenomics and metagenomics. *Pangenome Divers Dyn Evol Genomes*. 205–218.
- Makarova KS, Wolf YI, Koonin EV. 2009. Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *Biol Direct*. 4:19.
- Maynard Smith J, Szathmáry E. 1995. *The major transitions in evolution*. Oxford: Oxford University Press.
- McInerney JO, McNally A, O'Connell MJ. 2017. Why prokaryotes have pangenomes. *Nat Microbiol*. 2:17040.
- Mira A, Ochman H, Moran NA. 2001. Deletional bias and the evolution of bacterial genomes. *Trends Genet*. 17(10):589–596.
- Moran NA, Plague GR. 2004. Genomic changes following host restriction in bacteria. *Curr Opin Genet Dev*. 14(6):627–633.
- Morinaga Y, Nomura N, Sako Y. 2002. Population dynamics of archaeal mobile introns in natural environments: a shrewd invasion strategy of the latent parasitic DNA. *Microb Environ*. 17(4):153–163.
- Mruk I, Kobayashi I. 2014. To be or not to be: regulation of restriction-modification systems and other toxin-antitoxin systems. *Nucleic Acids Res*. 42(1):70–86.
- N'Guessan A, Brito IL, Serohijos AWR, Shapiro BJ. 2021. Mobile gene sequence evolution within individual human gut microbiomes is better explained by gene-specific than host-specific selective pressures. *Genome Biol Evol*. 13(8):evab142. <https://doi.org/10.1093/gbe/evab142>

- Nilsson AI, et al. 2005. Bacterial genome size reduction by experimental evolution. *Proc Natl Acad Sci U S A*. 102(34):12112–12116.
- Norman A, Hansen LH, Sørensen SJ. 2009. Conjugative plasmids: vessels of the communal gene pool. *Philos Trans R Soc Lond B Biol Sci*. 364(1527):2275–2289.
- Novick A, Doolittle WF. 2020. Horizontal persistence and the complexity hypothesis. *Biol Philos*. 35:2.
- Okasha S. 2006. *Evolution and the levels of selection*. Oxford: Oxford University Press.
- Orgel LE, Crick FHC. 1980. Selfish DNA: the ultimate parasite. *Nature* 284(5757):604–607.
- Otsuka Y. 2016. Prokaryotic toxin–antitoxin systems: novel regulations of the toxins. *Curr Genet*. 62(2):379–382.
- Parkhill J, et al. 2003. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat Genet*. 35(1):32–40.
- Plague GR, Dunbar HE, Tran PL, Moran NA. 2008. Extensive proliferation of transposable elements in heritable bacterial symbionts. *J Bacteriol*. 190(2):777–779.
- Raghavan R, Minnick MF. 2009. Group I introns and inteins: disparate origins but convergent parasitic strategies. *J Bacteriol*. 191(20):6193–6202.
- Ramsay JP, et al. 2016. An updated view of plasmid conjugation and mobilization in *Staphylococcus*. *Mob Genet Elements*. 6(4):e1208317.
- Rankin DJ, Rocha EPC, Brown SP. 2011. What traits are carried on mobile genetic elements, and why? *Heredity* 106(1):1–10.
- Redondo-Salvo S, et al. 2020. Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids. *Nat Commun*. 11(1):3602.
- Sandegren L, Sjöberg BM. 2007. Self-splicing of the bacteriophage T4 group I introns requires efficient translation of the pre-mRNA in vivo and correlates with the growth state of the infected bacterium. *J Bacteriol*. 189(3):980–990.
- Scott TW, West SA. 2019. Adaptation is maintained by the parliament of genes. *Nat Commun*. 10(1):5163.
- Sela I, Wolf YI, Koonin EV. 2016. Theory of prokaryotic genome evolution. *Proc Natl Acad Sci U S A*. 113(41):11399–11407.
- Sentchilo V, et al. 2013. Community-wide plasmid gene mobilization and selection. *ISME J*. 7(6):1173–1186.
- Shapiro BJ. 2016. How clonal are bacteria over time? *Curr Opin Microbiol*. 31:116–123.
- Siguiet P, Gourbeyre E, Chandler M. 2014. Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol Rev*. 38(5):865–891.
- Stevenson C, Hall JJP, Harrison E, Wood AJ, Brockhurst MA. 2017. Gene mobility promotes the spread of resistance in bacterial populations. *ISME J*. 11(8):1930–1932.
- 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)* 4(2):325–339.
- Tazzyman SJ, Bonhoeffer S. 2013. Fixation probability of mobile genetic elements such as plasmids. *Theor Popul Biol*. 90:49–55.
- Tettelin H, et al. 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial 'pan-genome'. *Proc Natl Acad Sci U S A*. 102:3950–3955.
- Touchon M, Rocha EPC. 2016. Coevolution of the organization and structure of prokaryotic genomes. *Cold Spring Harb Perspect Biol*. 8(1):a018168.
- Tourasse NJ, Kolstø AB. 2008. Survey of group I and group II introns in 29 sequenced genomes of the *Bacillus cereus* group: insights into their spread and evolution. *Nucleic Acids Res*. 36(14):4529–4548.
- Utter D, Borisy G, Eren AM, Cavanaugh C, Welch J. 2020. Metapangenomics of the oral microbiome provides insights into habitat adaptation and cultivar diversity. *Genome Biol*. 21(1):293.
- Van Dijk B, Hogeweg P, Doekes HM, Takeuchi N. 2020. Slightly beneficial genes are retained by bacteria evolving DNA uptake despite selfish elements. *Elife* 9:e56801.
- Violle C, Goldlust K, Djermoun S, Bigot S, Lesterlin C. 2020. Plasmid transfer by conjugation in Gram-negative bacteria: from the cellular to the community level. *Genes (Basel)*. 11(11):1239.
- Wagner A. 2006. Cooperation is fleeting in the world of transposable elements. *PLoS Comput Biol*. 2(12):e162.
- Wang X, et al. 2010. Cryptic prophages help bacteria cope with adverse environments. *Nat Commun*. 1:147.
- Welch RA, et al. 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A*. 99(26):17020–17024.
- Werren JH. 2011. Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc Natl Acad Sci U S A*. 108(Suppl_2):10863–10870.
- Westoby M, et al. 2021. Cell size, genome size, and maximum growth rate are near-independent dimensions of ecological variation across bacteria and archaea. *Ecol Evol*. 11(9):3956–3976.
- Williams KP. 2002. Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. *Nucleic Acids Res*. 30(4):866–875.
- Yin X, Stotzky G. 1997. Gene transfer among bacteria in natural environments. *Adv Appl Microbiol*. 45:153–212.
- Zaneveld JR, Nemergut DR, Knight R. 2008. Are all horizontal gene transfers created equal? Prospects for mechanism-based studies of HGT patterns. *Microbiology* 154(Pt 1):1–15.
- Zheng H, Dietrich C, Hongoh Y, Brune A. 2016. Restriction-modification systems as mobile genetic elements in the evolution of an intracellular symbiont. *Mol Biol Evol*. 33(3):721–725.
- Zimmerly S. 2005. Mobile introns and retroelements in bacteria. In: Mullany P, editor. *The dynamic bacterial genome*. Cambridge: Cambridge University Press. p. 121–148.
- Zimmerly S, Semper C. 2015. Evolution of group II introns. *Mob DNA*. 6:7.
- Zünd M, et al. 2021. High throughput sequencing provides exact genomic locations of inducible prophages and accurate phage-to-host ratios in gut microbial strains. *Microbiome* 9(1):77.

Associate editor: Adam Eyre-Walker