Host Range and Genetic Plasticity Explain the Coexistence of Integrative and Extrachromosomal Mobile Genetic Elements

Jean Cury,*[,1,2] Pedro H. Oliveira,[†,1,2] Fernando de la Cruz,[3] and Eduardo P.C. Rocha[1,2]

[1]Microbial Evolutionary Genomics, Institut Pasteur, Paris, France

[2]CNRS, UMR3525, Paris, France

[3]Departamento de Biologia Molecular e Instituto de Biomedicina y Biotecnologia de Cantabria (IBBTEC), Universidad de Cantabria-CSIC, Santander, Spain

[†]Present address: Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY

***Corresponding author:** E-mail: jean.cury@normalesup.org.

Associate editor: Nicole T. Perna

Abstract

Self-transmissible mobile genetic elements drive horizontal gene transfer between prokaryotes. Some of these elements integrate in the chromosome, whereas others replicate autonomously as plasmids. Recent works showed the existence of few differences, and occasional interconversion, between the two types of elements. Here, we enquired on why evolutionary processes have maintained the two types of mobile genetic elements by comparing integrative and conjugative elements (ICE) with extrachromosomal ones (conjugative plasmids) of the highly abundant MPF_T conjugative type. We observed that plasmids encode more replicases, partition systems, and antibiotic resistance genes, whereas ICEs encode more integrases and metabolism-associated genes. ICEs and plasmids have similar average sizes, but plasmids are much more variable, have more DNA repeats, and exchange genes more frequently. On the other hand, we found that ICEs are more frequently transferred between distant taxa. We propose a model where the different genetic plasticity and amplitude of host range between elements explain the co-occurrence of integrative and extrachromosomal elements in microbial populations. In particular, the conversion from ICE to plasmid allows ICE to be more plastic, while the conversion from plasmid to ICE allows the expansion of the element's host range.

Key words: **mobile genetic elements, horizontal gene transfer, molecular evolution, microbial genomics, conjugation.**

Article

Introduction

The genomes of Prokaryotes have mobile genetic elements (MGEs) integrated in the chromosome or replicating as extrachromosomal elements. These MGEs usually encode non-essential but ecologically important traits (Ochman et al. 2000; Frost et al. 2005). Extrachromosomal elements, such as conjugative plasmids (CPs) and lytic phages, replicate autonomously in the cell using specialized replicases to recruit the bacterial DNA replication machinery (or to use their own). Plasmids and extrachromosomal prophages can also increase their stability in cellular lineages using partition systems, for proper segregation during bacterial replication (Ebersbach and Gerdes 2005), resolution systems, to prevent accumulation of multimers (Summers 1991), and restriction–modification or toxin–antitoxin systems, for postsegregation killing of their hosts (Kobayashi 2001). Alternatively, many MGEs integrate into the chromosome. This is the case of the vast majority of known prophages, of most conjugative elements (ICEs), and of many elements with poorly characterized mechanisms of genetic mobility (e.g., many pathogenicity islands) (Canchaya et al. 2003; Dobrindt et al. 2004; Guglielmini et al. 2011). The integrated elements are replicated along with the host chromosome and require an additional step of excision before being transferred between cells. The existence of both integrative and extrachromosomal elements was a fruitful source of controversy in the dawn of molecular biology, eventually leading to the discovery of the molecular mechanisms allowing both states (Jacob et al. 1960; Lederberg 1998). Yet, a complementary question does not seem to have been addressed in the literature: Why are there both types of elements? What are the relative benefits and disadvantages of the integrated and extrachromosomal MGE?

To address these questions, we analyzed the differences and similarities between ICEs and CPs. We focused on these elements because both forms are frequently found in bacteria, they can be easily detected in genomes, and the mechanism of conjugation is well known. Conjugative elements have a crucial role in spreading antibiotic resistance and virulence genes among bacterial pathogens (Bellanger et al. 2014; Carraro and Burrus 2014; Johnson and Grossman 2015; Delavat et al. 2017). Recently, several works suggested that the line separating integrative ICEs and CPs could be thinner than anticipated (Carraro and Burrus 2015), because some ICEs encode plasmid-associated functions like replication (Lee et al. 2010) or partition (Carraro et al. 2015), some plasmids

encode integrases (Nunes-Düby et al. 1998), and ICEs and CPs are intermingled in the phylogenetic tree of conjugative systems (Guglielmini et al. 2013). Finally, both forms—ICEs and CPs—are found throughout the bacterial kingdom, but their relative frequency depends on the taxa and on the mechanisms of conjugation (Guglielmini et al. 2011). These differences suggest that conjugative elements endure diverse selective pressures for being integrative or extrachromosomal depending on unknown environmental, genetic, or physiological variables.

We thought that key differences in the biology of integrative and extrachromosomal elements might provide them with different types of advantages. ICEs require an additional step of integration/excision during transfer, which may take time and requires genetic regulation. Their integration in the chromosome may affect the latter's organization and structure, and these collateral effects might depend on the size of the element. On the other hand, ICEs replicate as part of chromosomes and could thus be lost from the cell at lower rates than plasmids. Furthermore, plasmids must recruit the host replication machinery, which may render elements incompatible and is known to constrain their host range: many plasmids are able to conjugate into distantly related hosts, but are unable to replicate there (Guiney 1982; Zhong et al. 2005; Klümper et al. 2015). We thus hypothesized that ICEs might be favored when transfers occur between distant hosts, whereas plasmids might provide more genetic plasticity because their size is not constrained by chromosomal organization.

Here, we study conjugative elements of the type MPF_T. This is the most frequent and best-studied type of conjugative systems (Guglielmini et al. 2013), and the only one for which we can identify hundreds of elements of each of the forms (ICEs and CPs). We restricted our analysis to genera containing both CPs and ICEs, to avoid, as much as possible, taxonomical biases. We first describe the content of both types of elements and highlight their differences and similarities. Next, we quantify their genetic similarity and the extent of their gene exchanges. Finally, we show that chromosomal integration facilitates the colonization of novel taxa by a conjugative element.

Results

Functional and Genetic Differences between ICEs and CPs

We analyzed a set of 151 ICEs and 136 CPs of the same genera and of type MPF_T, most of which were from Proteobacteria (96.9%). Both ICEs and CPs were found to be AT-richer than their host chromosomes, which is a common feature in MGEs and horizontally transferred genes (Rocha and Danchin 2002). However, the difference was three times smaller for ICEs (fig. 1A), presumably because they replicate with the chromosome or remain a longer time in the same host. The average size of CPs is slightly larger (75 kb vs. 59 kb), and the median slightly smaller (46 kb vs. 52 kb) than that of ICEs. In contrast, CPs have more diverse sizes than ICEs (fig. 1B), showing a coefficient of variation twice as large (1.05 vs. 0.49). The

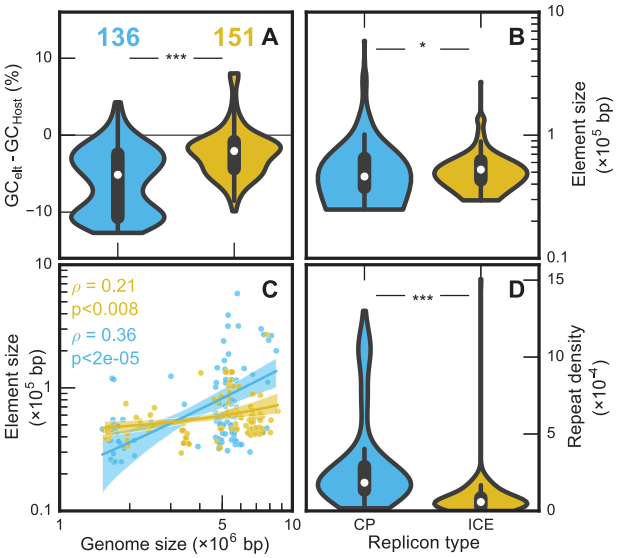

Fig. 1. Comparison between 136 CPs (blue) and 151 ICEs (yellow) in terms of their composition and size. (*A*) Violin plot showing the distribution of the differences between the elements' GC content and that of their hosts'. CPs are AT richer than ICEs relative to their hosts (Wilcoxon rank sum test, *P* value $< 10^{-3}$). (*B*) Violin plots showing the distribution of size of ICEs and CPs. ICEs and CPs have different distributions of size (same test, *P* value < 0.05). Median sizes: ICEs (52.5 kb) $>$ CPs (46.1 kb). Averages: ICEs (59 kb) $<$ CPs (74.6 kb) (*C*) Plot of the size of the element as a function of their hosts' genome size (decreased by the size of the mobile element itself). Shaded regions indicate the 95% confidence interval. (*D*) Violin plots showing the distribution of the density of repeats in the two types of elements. The density of repeats is higher in CPs than in ICEs (0.30 vs. 0.078 repeats per kb, same test, *P* value $<10^{-10}$).

size of the conjugative elements depends on the size of the bacterial genome (after discounting the size of their conjugative elements), this effect being much stronger for CPs (fig. 1C). CPs also have four times higher density of large DNA repeats than ICEs (fig. 1D). Repeats are targeted by homologous recombination leading to increased rates of genome rearrangements, amplifications, and deletions (Treangen et al. 2009), which could explain the higher coefficient of variation of the size of plasmids. These results suggest that CPs diversify faster than ICEs.

HGT concentrates in a few hotspots in bacterial chromosomes, presumably to minimize disruption in their organization (Oliveira et al. 2017). We used HTg50, a measure of the concentration of HGT in chromosomes that corresponds to the minimal number of spots required to account for 50% of horizontally transferred genes (Oliveira et al. 2017), to test if chromosomes with fewer integration hotspots had more plasmids. Indeed, there is a negative association between the number of plasmids, (expressed as the ratio of the total size of plasmid to the total the size of their genomes in a given species), and the chromosomes' HTg50 (Spearman $\rho = -0.35$, *P* value $= 0.0016$, supplementary fig. S1, Supplementary Material online).

We then quantified the differences between ICEs and CPs in terms of functions associated with their biology, with a focus on stabilization functions. Relaxases are part of the

rolling circle replication initiator proteins and some have been shown to act as replicases (Lee et al. 2010; Carraro et al. 2016) or site-specific recombinases (Francia and Clewell 2002; César et al. 2006). Since all conjugative elements encode a relaxase, by definition, they may also have relaxase-associated replicase functions. In the following, we focused on typical plasmid replication initiator proteins (>95% of them are involved in theta-replication, and none is matched by the protein profiles of relaxases), and serine or tyrosine recombinases as integrases. Expectedly, ICEs showed higher frequency of integrases, while CPs had more frequently identifiable partition and replication systems. Some ICEs encode partition systems (11%) and many encode a replicase (40%), while 37% of CPs encode at least one tyrosine or serine recombinase (fig. 2A). These results further illustrate a continuum between the two types of elements: about half of the elements (40% and 48%, ICEs and CPs, respectively) have functions usually associated with the other type and may (rarely) lack functions typically associated with its own type (fig. 2B). Some of the functions lacking—integrase in ICEs and replicase in CPs—are surprising. Their absence may have been caused by recent gene losses or annotation issues resulting from gaps in our knowledge. They may also pinpoint further examples of known exceptions: some conjugative elements have been shown to integrate using DDE recombinases (Brochet et al. 2009; Guérillot et al. 2013) and plasmid replication can be done by a rep protein in *trans* or by relaxases. Interestingly, we observe that ICEs containing replication or partition systems contain more repeats per nucleotide than the others (Wilcoxon rank-sum test, P value = 0.02). We identified plasmid incompatibility systems of diverse types, whereas ICE could not be typed in the current scheme (supplementary fig. S2, Supplementary Material online).

We then made similar analyses for functions usually regarded as accessory or unrelated to the biology of MGEs (fig. 2B). ICEs were more likely to carry restriction–modification systems (x2.8) than CPs (but not orphan methylases), suggesting that ICEs endure stronger selective pressure for stabilization in the genome. In contrast, they were significantly less likely to carry antibiotic resistance genes or integrons. They also had fewer identifiable entry-exclusion systems, which may reflect the ability of ICEs to tolerate the presence of multiple similar elements in the cell (Garcillán-Barcia and de la Cruz 2008). The classification of genes in the four major functional categories of the EggNOG database, showed that ICEs had relatively more genes encoding metabolic and cellular processes. We have previously shown that genes of unknown or poorly characterized function were overrepresented in ICEs relative to their host chromosome (Cury et al. 2017). The frequency of these genes is even higher in plasmids (61% vs. 46%). Hence, both types of elements have many functions in common, but their relative frequency often differs significantly.

Genetic Similarities between ICEs and CPs

The results of the previous section, together with previously published studies (see Introduction), suggest that ICEs and CPs either share a common history or often exchange genes

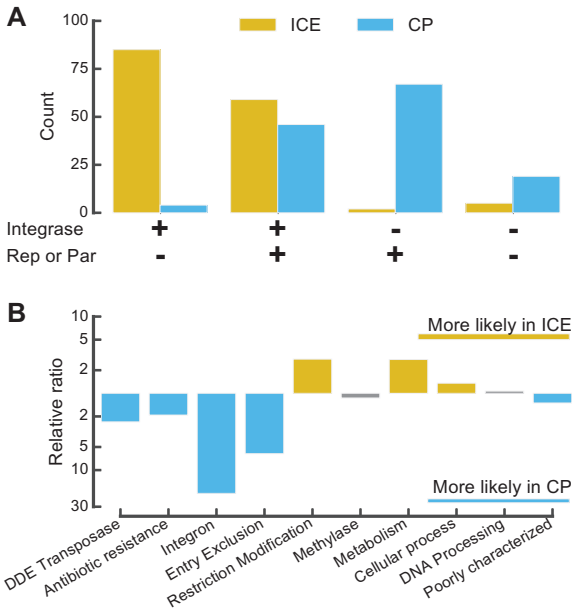

FIG. 2. Comparisons of the functions carried by ICEs and CPs. (*A*) Counts of the number of elements carrying Integrases on one hand and replicases or partition systems ("Rep or Par") on the other. The presence of the function is noted by (+) and its absence by (−). (*B*) Relative ratio of the different functions between the two types of elements. Accessory functions significantly overrepresented in ICEs (relative to CPs) are in yellow, whereas those underrepresented are in blue. Statistical test: P value <0.05 Fisher's exact test with Bonferroni–Holm correction for multiple tests. Gray bars indicate nonsignificant differences between ICEs and CPs.

(or both). We detailed the relationships of homology between ICEs and CPs using the weighted Gene Repertoire Relatedness (wGRR), which measures the frequency of bidirectional best hits between two elements weighted by their sequence similarity (see Materials and Methods). We clustered the matrix of wGRR using the Louvain algorithm (Blondel et al. 2008), and found six well-distinguished groups (fig. 3). Two groups (1 and 6) are only constituted of CPs, two are composed of >90% of ICEs (3 and 5) and two have a mix of both types of elements (2 and 4) (fig. 3, top bar). Bacterial species are scattered between groups, showing that they are not the key determinant of the clustering. Some groups are only from γ-proteobacteria, but others include bacteria from different classes. Groups where elements are from the same taxonomic classes tend to have either CPs or ICEs, whereas the others have mixtures of both elements. Group 4, includes many ICEs and CPs, where all ICEs have integrases while more than half of the CPs lack both replication and partition systems (fig. 4). This group includes almost only elements from ε-proteobacteria that may have specificities that we were not able to take into account. In contrast, almost all ICEs of groups 2 and 3 encode an integrase and all CPs have partition or replication systems.

We controlled for the effect of the MPF genes in the previous clustering analysis by redoing it without these genes (supplementary fig. S3, Supplementary Material online). This produced the same number of groups—N1 to N6—and 90% of the elements of the former groups were classed

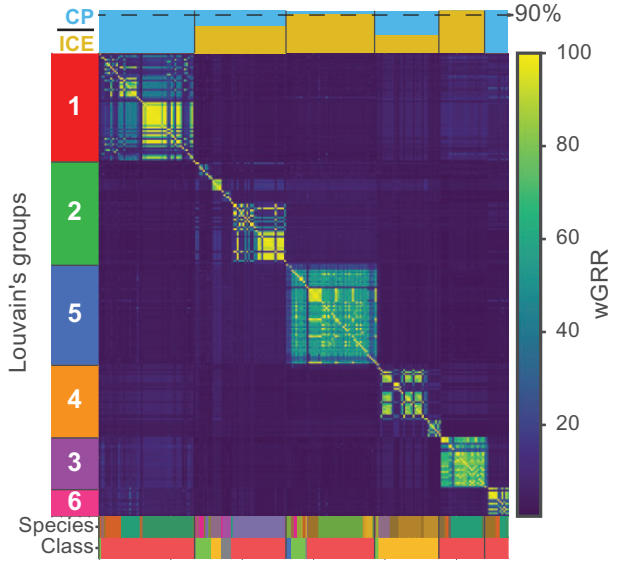

Fig. 3. Heatmap of the wGRR scores between pairs of elements (ICE or CP), ordered after the size of Louvain's group (depicted on the left bar). The top bar represents the proportion of ICEs (yellow) and CPs (blue) for each group. The bottom bar assigns a color corresponding to the host's species (top) or class (bottom). Species: colors indicate species identified in each group. Class: γ-proteobacteria are in red, β-proteobacteria in green, α-proteobacteria in blue, ϵ-proteobacteria in orange, Fusobacteria and Acidobacteria in gray). The vertical ribbon on the right indicates the scale of wGRR values.

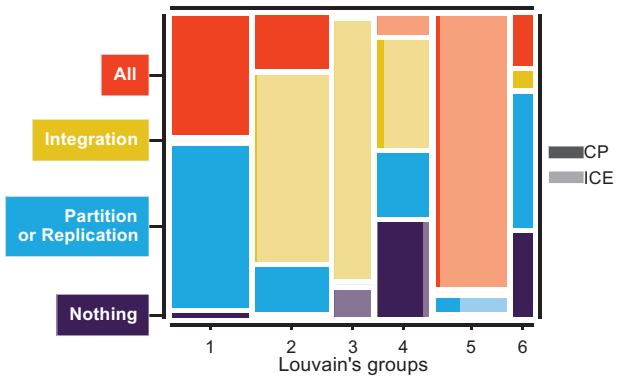

Fig. 4. Mosaic plot representing the frequency of key functions of conjugative elements in terms of the Louvain's groups (fig. 3). The width of the bar is proportional to the number of elements in a given Louvain's group (67, 64, 32, 45, 62, and 17 elements, respectively) and the heights reflect the proportion of the elements with the function in a given group. The colors represent the type of function, and their tint represents the proportion of ICEs (lighter) and CPs (darker).

in the same novel groups (supplementary fig. S4, Supplementary Material online). The only qualitatively significant difference between the two analyses concerned the group 2 for which 36% of the elements were now classed in groups N4 or N6. Overall, these controls confirm that ICEs and CPs can be grouped together, and apart from other elements of the same type. The grouping is not caused by sequence similarity between conjugative systems. Instead, it probably reflects either within group genetic exchanges

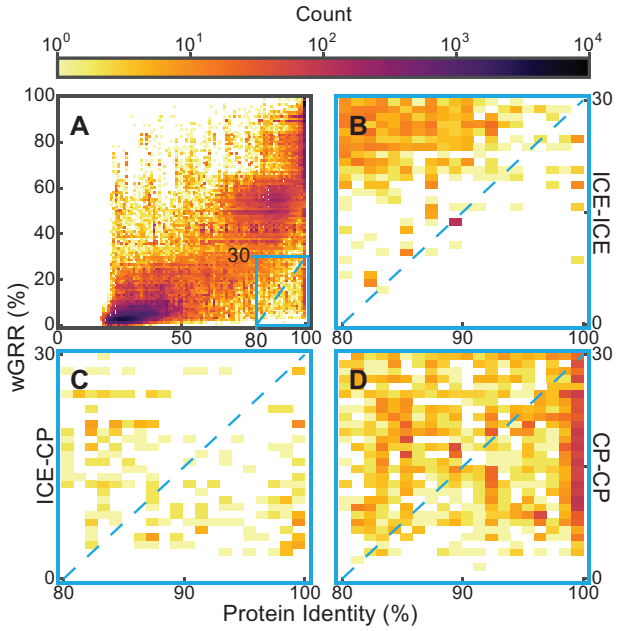

Fig. 5. 2D histogram of the wGRR score of pairs of conjugative elements as a function of the protein identity of their homologues. Each square indicates the number of proteins under comparison (e.g., comparisons between proteins of ICEs in *B*) for given bins of protein identity and wGRR. For a given pair of elements, there is a single value of wGRR (on the *y* axis) and as many values of protein identity (on the *x* axis) as there are pairs of homologs between these elements. (*A*) Distribution for the entire data set. wGRR values are correlated with protein identity of the elements' homologs ($\rho_{ICE-ICE} = 0.90$, $\rho_{CP-CP} = 0.83$). The blue rectangle zooms on a region where the pairs of elements are very different (GRR < 30%), yet they encode at least one very similar protein (identity >80%). The dashed line separates the elements where protein identity is higher than wGRR x 2/3 + 80. (*B*) Zoom for ICE–ICE comparisons. (*C*) ICE–CP comparisons. (*D*) CP–CP comparisons.

between ICEs and CPs, or interconversions of the two types of elements.

Genetic Exchanges: CPs Become ICEs for Broader Host Range

The clustering of ICEs and CPs could be explained by genetic transfer between them. To address this question, we represented the wGRR of pairs of conjugative elements as a function of the percentage of identity of the homologous proteins (fig. 5). Pairs of elements with low wGRR (<30%) but with highly similar homologs (>80% sequence identity) can be best explained by recent gene transfer between initially very distinct elements (fig. 5*A*). In agreement with our observations of higher genetic plasticity in CPs, most of these transfers took place between these types of elements (fig. 5*B–D*).

We hypothesized that ICEs could hold an advantage over CPs to colonize novel hosts, because replication restricts plasmid host range. To test this hypothesis, we analyzed the wGRR between pairs of ICEs and pairs of CPs in function of the phylogenetic distance between their bacterial hosts. This showed similar patterns for the two types of elements, with the notable exception that there are no pairs of highly similar

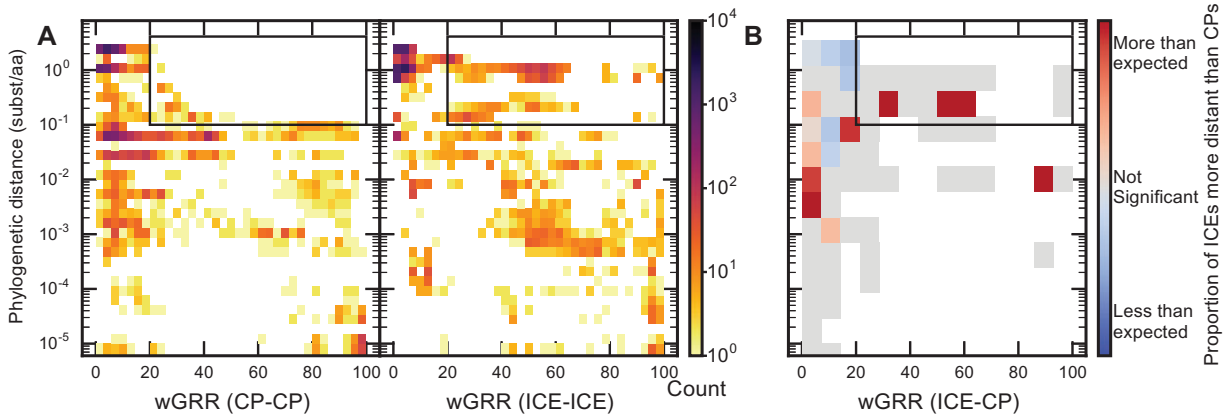

Fig. 6. wGRR scores between two elements, as function of the host phylogenetic distance (*A*) 2D histogram of the distribution of the wGRR score as a function of the phylogenetic distance for pairs of CPs (left) and pairs of ICEs (right). Each square represents a bin in the 2D histogram with the number of comparisons (e.g., between CPs in the left plot) for a given bin of phylogenetic distance between the hosts and wGRR value between the elements. The bottom row corresponds to all pairs with distance $<10^{-5}$ (including those in the same host). The black rectangle represents comparisons between very similar elements (high wGRR), present in very distant hosts (high phylogenetic distance). An alternative representation of the elements in the black rectangle can be found on a phylogenetic tree in supplementary figure S5, Supplementary Material online. (*B*) Proportion of ICEs more distinct in terms of trinucleotides from their host than CPs. Bins are larger than in (*A*) to increase the power of the statistical analysis. In each bin, the color represents the proportion of ICEs that are more distant to their host chromosome in terms of trinucleotide composition than CPs (relative to their hosts). This proportion is in red (blue) if there are significantly more (less) ICE than CP, and in gray if the test not significant (binomial test, *P* value $>10^{-2}$). White: no elements in the bin.

plasmids (wGRR > 50%) in distant hosts (>0.1 substitutions/position, e.g., the average distance in the tree between *Escherichia coli* and *Pseudomonas aeruginosa*). In contrast, a third of all ICEs (n = 50) are in these conditions (fig. 6*A* and supplementary fig. S5, Supplementary Material online). The same analysis after removing the MPF genes shows wGRR values shifted to lower wGRR values for all elements, but qualitatively similar trends (supplementary fig. S6, Supplementary Material online). This suggests a major difference in the ability of ICEs and CPs to be stably maintained after their transfer into a distant host.

We then analyzed the pairs ICE-CP. We found few pairs of highly similar ICEs and CPs in closely related hosts (bottom right corner of fig. 6, n = 8 for wGRR > 50% and d $< 10^{-2}$), suggesting that interconversion between these elements remains rare within a clade. A larger number of ICE-CP pairs were very similar but present in distant hosts (n = 38, fig. 6). The most parsimonious explanation for these observations, is the recent transfer of one of the elements (ICE or CP) to a distant bacterial host. We identified the latter element based on the differences in terms of trinucleotide composition between the elements and the host chromosomes (defined as pvalue in Suzuki et al. 2010). We then computed for each ICE–CP pair the difference between the pvalue of the pair ICE–host and that of the pair CP–host (see Materials and Methods). In agreement to our observation that ICEs have broader host ranges, these differences indicate that ICEs are relatively more distant from the host chromosome for pairs with high wGRR in distant hosts than for the rest of the pairs (Wilcoxon rank sum test, *P* value $< 10^{-20}$, fig. 6*B*, and supplementary fig. S7, Supplementary Material online). The rarity of ICE–CP pairs in closely related hosts, their abundance in distant hosts, and the identification that ICEs are the most compositionally atypical relative to the host in the latter, suggest that successful transfer of CPs to distant hosts is favored when they integrate the chromosome and become ICEs.

Discussion

In this study, we compared the genetic organization of ICEs and CPs to evaluate the hypothesis that they are essentially equivalent MGEs (Lee et al. 2010; Carraro and Burrus 2015; Cury et al. 2017). We found that numerous CPs have integrases (although these may serve for dimer resolution and not integration; Carnoy and Roten 2009), and numerous ICEs encode replication and partition functions. Relaxases—present in both ICEs and CPs—have been shown to act as integrases or replication initiators in certain elements (Wawrzyniak et al. 2017), which provides further functional overlapping between the elements. Hence, ICEs and CPs share many functions beyond those related to conjugation. These similarities explain why they can cluster together in terms of gene repertoires, even when excluding conjugation functions.

There are also some clear differences between CPs and ICEs explaining why they sometimes group separately. First, genes encoding replicases and partition systems are more frequent in plasmids, while tyrosine and serine recombinases are more frequent in ICEs. Interestingly, we could not attribute incompatibility groups to ICEs, suggesting that the replication module either is rarely exchanged between ICEs and CPs or evolves too rapidly. Second, the frequency of certain accessory traits is different: plasmids are more likely to encode antibiotic resistance genes whereas ICEs encode more metabolism-related

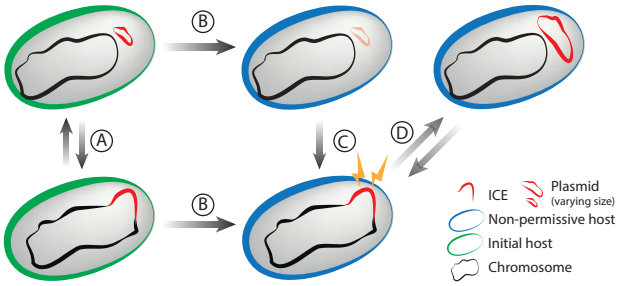

Fig. 7. Interconversion between ICEs and CPs allows access to the higher genetic plasticity of CPs and the broader host range of ICEs. (*A*) Many ICEs and CPs encode the functions of the other element, facilitating CP <–> ICE interconversion. (B) Conjugative transfer to a taxonomically distant host often precludes stabilization of CPs because they are nonpermissive for plasmid replication. ICEs seem at an advantage when conjugating to distant hosts, presumably because they integrate the chromosome. (*C*) CPs can be salvaged after transfer in a nonpermissive host if they integrate the chromosome (as ICE). This effect corresponds to an host-range expansion, as described in broad host range plasmids (De Gelder et al. 2008). (*D*) Mutations and/or gene acquisitions can result in an ICE with the ability to replicate as a plasmid. CPs have wider size distributions and exchange more genetic information. This allows them to acquire novel genetic information at higher rates, which may eventually be adaptive for their hosts.

genes. We controlled for taxonomical bias by analyzing ICEs and CPs from the same genera, yet the precise functions we observed are dependent on the data set which, here, is biased toward nosocomial pathogens. A data set of bacteria from other environments might present other functional differences in terms of the traits carried preferentially by one or the other type of elements. Finally, the difference in %GC content relative to the host, the number of repeats, the patterns of gene variation and exchange, and the host range are quantitatively different in the two types of elements. After integrating all this information, we propose that, in spite of their similarities, each type has traits that favored in specific situations. In particular, our results suggest that interconversion between ICEs and CPs gives them access to the higher genetic plasticity of CPs and the broader host range of ICEs (fig. 7).

Even if there are some known families of large (>200 kb) ICEs (Sullivan and Ronson 1998; Kers et al. 2005), these elements have a remarkably narrower variation in size than CPs in our data set. This suggests that CPs are more flexible than ICEs in terms of the amount of genetic information they can carry and in their ability to accommodate novel information. We also show that CPs exchange genes more frequently. Mechanistically, this higher rate of gene exchange between CPs may result from recombination between DNA repeats, gene acquisition by integrons, and genome rearrangements caused by transposable elements. We showed that CPs encode all of these elements in higher number. Plasmid copy number, when high, may also contribute to increase recombination rates in plasmids relative to ICEs. Interestingly, recombination mediated by transposable elements has been shown to drive the evolution of certain plasmids (He et al. 2016; Hall et al. 2017) and to accelerate the reduction of plasmid cost, thus stabilizing the element after horizontal transfer (Porse et al. 2016). The restrictions in size variation of ICEs are probably not due to the mechanism of integration or excision, because such reactions can occur between very distant recombination sites (Wu and Errington 2002). Instead, very large ICEs may disrupt chromosome organization by affecting the distribution of motifs, changing chromosome folding domains, or unbalancing the sizes of replichores (Touchon and Rocha 2016). Repeat-mediated recombination leads to replicon rearrangements, and may lead to stronger counter-selection of DNA repeats in ICEs than in CPs. This further restricts the variation of size in ICEs and their rate of gene exchange. Interestingly, the size of plasmids varies much more steeply with genome size than the size of ICEs, suggesting that CPs may play a particularly important role in the evolution of large bacterial genomes of Proteobacteria, which have higher rates of genetic exchanges (Oliveira et al. 2016), and often contain mega-plasmids (Smillie et al. 2010).

Some plasmids are known to be broad-host range and adapt to novel hosts, especially if they carry adaptive traits that compensate for the initially poor intrinsic persistence of the element (De Gelder et al. 2008; Loftie-Eaton et al. 2017). However, within the large phylogenetic span considered in this work, MPF_T ICEs have broader host ranges than CPs. Actually, the first known ICE, Tn916 (not MPF_T, thus not included in this study), became notorious due to its ability to spread antibiotic resistance between distant phyla (Clewell et al. 1995). We propose that the conversion of plasmids into ICEs can elicit conjugative transmission to otherwise nonpermissive hosts. This results in an effective expansion of the host range of the element (now an ICE). Once installed in the recipient chromosome, the ICE can incorporate a new replication system, or freely mutate its own. If it obtains a functional replication system in the new host the reverse interconversion may occur (fig. 7). This results in a plasmid with a new host range and the higher evolvability of an autonomous plasmid.

The similarities between certain groups of ICEs and CPs in terms of gene repertoires, the integration of CPs as ICEs upon long-range transfer, and the exchange of genetic information between them are novel evidence for interconversion between the two types of elements. This had previously been proposed based on the phylogeny of the conjugative system (Guglielmini et al. 2011). Interconversion of elements and occasional transfers between CPs and ICEs allow them to access the other elements' gene pool. These events may result in conjugative elements sharing many traits of the other type of element, as we observed in more than a third of all conjugative elements, and should produce many genetic similarities between ICEs and CPs. The latter probably facilitate further interconversions between the elements.

Other traits may provide advantages specifically to either ICEs or CPs. The ability of plasmids to modify their copy number may accelerate adaptive evolutionary processes, such as the acquisition of antibiotic resistance (San Millan et al. 2016). On the other hand, ICEs might be more stably maintained in lineages because they replicate within the chromosome. Finally, the carriage of ICEs and CPs may have

different costs. The cost of plasmids has been extensively studied and is strongly dependent on the traits they encode (San Millan and MacLean 2017). Much less is known about the costs of ICEs; several reports suggest that they have low fitness costs when conjugation is not expressed, but this may change dramatically during transfer (Delavat et al. 2017). Direct comparisons of the cost of carriage of ICEs and CPs carrying similar traits are unavailable. Further experimental work will be needed to test these hypotheses.

Many mobile elements are mobilizable but not able to conjugate independently (Smillie et al. 2010; Guédon et al. 2017). These elements often encode a relaxase that recognizes the element's origin of transfer and is able to interact with a T4SS from an autonomously conjugative element to transfer to other cells. Some other elements only contain an origin of transfer that is recognized by a relaxase of another element. Many of the disadvantages of CPs and ICEs are similar to those of mobilizable plasmids and integrative mobilizable elements, whether they encode a relaxase or not. Notably, the former must be replicated in the extrachromosomal state, and the latter integrate the genome where they must not disrupt genome organization. Patterns observed in conjugative elements are thus likely to be applicable to mobilizable ones.

These results may also be relevant to understand lysogeny by temperate phages. The vast majority of known prophages are integrated in the chromosome, but some replicate like plasmids (Łobocka et al. 2004; Ravin 2011). Considering that prophages share some of the constraints of conjugative elements, they are likely to be under similar trade-offs. However, phages are under additional constraints. Notably, their genome size is much less variable than that of conjugative elements, because it must be packaged into the virion (Touchon et al. 2017), and this may render the extrachromosomal prophages less advantageous in terms of accumulating novel genes. This could explain why most prophages are integrative whereas conjugative systems are more evenly split between integrative and extrachromosomal elements.

In summary, our results show that there are specific fitness benefits associated with the divergent lifestyles identified for pairs of highly similar ICEs and CPs. We should emphasize that our model proposes that plasmid to ICE transition results in broadening the host range of the element, with a concomitant fitness benefit associated with higher rates of its horizontal transfer. The ICE to plasmid transition, on the other hand, results in added versatility of the cargo content (increased size range), which enhances the evolvability of the mobile element. These factors may promote or even drive the interconversion between plasmids and ICEs in an ever-changing environment. The concepts presented in this work will provide a better understanding of the evolution of bacterial genomes and their mobile genetic elements.

Materials and Methods

Data

Conjugative systems of type T (MPF_T) were searched in the set of complete bacterial genomes from NCBI RefSeq (http://ftp.ncbi.nih.gov/genomes/refseq/bacteria/, last accessed in November 2016). We analyzed 5,562 complete genomes from 2,268 species, including 4,345 plasmids and 6001 chromosomes. The classification of the replicon in plasmid or chromosome was taken from the information available in the GenBank file. Our method to delimit ICEs is based on comparative genomics of closely related strains. Hence, we restricted our search for conjugative systems to the species for which we had at least five genomes completely sequenced (164 species, 2,990 genomes).

Detection of Conjugative Systems and Delimitation of ICEs

Conjugative systems were detected using the CONJscan module of MacSyFinder (Abby et al. 2014), with protein profiles and definitions of the MPF type T, published previously (Guglielmini et al. 2014). ICEs were delimited with the same methodology, as developed in a previous work (Cury et al. 2017). Briefly, we identified the core genomes of the species. The region between two consecutive genes of the core genome defined an interval in each chromosome. We then defined spots as the sets of intervals in the chromosome flanked by genes of the same two families of the core genome (Oliveira et al. 2017). We then identified the intervals and the spots with conjugative systems. The information on the sets of gene families of the spots with ICEs (i.e., the spot pan-genome) was used to delimit the element boundaries (script available at https://gitlab.pasteur.fr/gem/spot_ICE; last accessed April 2017). This methodology was shown to be accurate at the gene level (precise nucleotide boundaries are not identifiable by this method, see Cury et al. 2017).

Functional Analyses

Partition systems, replication systems, entry-exclusion systems and restriction modification systems were annotated with HMM profiles, as described in our previous work (Oliveira et al. 2014; Cury et al. 2017). Integrases were annotated with the PFAM profile PF00589 for the Tyrosine recombinases and the combination of PFAM profiles PF00239 and PF07508 for Serine recombinases. DDE Transposases were detected with Macsyfinder (Abby et al. 2014) with models used previously (Touchon et al. 2014). Antibiotic resistance genes were detected with ResFams profiles (core version v1.1) (Gibson et al. 2015) using the –cut_ga option. We determined the functional categories of genes using their annotation as provided by their best hits to the protein profiles of the EggNOG database for bacteria (version 4.5, bactNOG) (Huerta-Cepas et al. 2016). Genes not annotated by the EggNOG profiles were classed as "Unknown" and included in the "Poorly characterized" group. The HMM profiles were used to search the genomes with HMMER 3.1b2 (Eddy 2011), and we retrieved the hits with an e-value $<10^{-3}$ and with alignments covering at least 50% of the profile. Integrons were detected using IntegronFinder version 1.5.1 with the –local_max option for higher accuracy (Cury et al. 2016). Repeats (direct and inverted) were detected with Repseek (version 6.6) (Achaz et al. 2007) using the option -p 0.001 which set the *P* value for determining the minimum seed length.

Statistics

We tested the overrepresentation of a given function or group of functions using Fisher's exact tests on contingency tables. For partition, replication and integration, the contingency table was made by splitting replicons into those encoding or and those not encoding the function and between ICEs and CPs. The use of presence/absence data instead of the absolute counts was made because the presence of at least one occurrence of a system is sufficient to have the function and because the counts were always low. For the other functions, the contingency table was made by splitting the proteins of the element in those annotated for a given function and the remaining ones. This allowed to take into account the differences in the number of genes between elements. The Fisher's exact tests were considered as significant after sequential Holm–Bonferroni correction, with a family-wise error rate of 5% (the probability of making at least one false rejection in the multiple tests, the type I error). From the contingency table, we computed the relative ratio (or relative risk) of having a given function more often in ICEs than in CPs. The relative ratio is computed as follow: $RR = \frac{ICEwF/N_{ICE}}{CPwF/N_{CP}}$ where *ICEwF* is the number of ICE (or proteins in ICEs) with the given function, and N_{ICE}, the total number of ICE (or proteins in ICEs), and likewise for CP. The term $ICEwF/N_{ICE}$ is an estimation of the probability of an ICE (or a protein in an ICE) to carry a given a function.

Phylogenetic Distances

Phylogenetic distances were extract from the Proteobacterial tree of the Core-genome. To build the tree, we identified the genes present in at least 90% of the 2,897 genomes of Proteobacteria >1 Mb that were available in GenBank RefSeq in November 2016. A list of orthologs was identified as reciprocal best hits using end-gap free global alignment. Hits with <37% similarity in amino acid sequence and >20% difference in protein length were discarded. We then identified the protein families with relations of orthology in at least 90% of the genomes. They represent 341 protein families. We made multiple alignments of each protein family with MAFFT v.7.205 (with default options) (Katoh and Standley 2013) and removed poorly aligned regions with BMGE (with default options) (Criscuolo and Gribaldo 2010). Genes missing in a genome were replaced by stretches of "-" in each multiple alignment, which has been shown to have little impact in phylogeny reconstruction (Filipski et al. 2014). The tree of the concatenate alignment was computed with FastTree version 2.1 under the LG model (Price et al. 2009). We chose the LG model because it was the one that minimized the AIC.

Distance to the Host

We used the differences in trinucleotide composition to compute the genetic distance between the mobile element and its host chromosome, as previously proposed (Suzuki et al. 2008). The analysis of ICEs was done by comparing the element with the chromosome after the removal of its sequence from the latter. Briefly, we computed the trinucleotide relative abundance ($x_{ijk} \forall i,j,k \in \{A,T,C,G\}$) for the chromosomes (in windows of 5 kb) and for the conjugative elements (entire replicon), which is given by: $x_{ijk} = f_{ijk}/f_i f_j f_k$, with *f* the frequency of a given k-mer in the sequence (Suzuki et al. 2010). We first computed the Mahalanobis distance between each window and the host chromosome as follow:

$$D = \sqrt{(w-h)^T H^{-1}(w-h)}$$

with *w*, the vector of trinucleotide abundances (x_{ijk}) in a given window, and *h*, the mean of the vector of x_{ijk} (i.e., the average trinucleotide abundance in the chromosome). *H* is the covariance matrix of the trinucleotide relative abundances. The inverse of the covariance matrix (H^{-1}) downweights frequent trinucleotides, like the trinucleotides corresponding to start codons, which are common to conjugative elements and chromosome and could bias the distance. We computed the Mahalanobis distance between conjugative elements and their hosts' chromosomes (same formula as above, but *w* is now for a conjugative element instead of a chromosome window). We then computed the probability (*P* value) that the measured distance between a conjugative element and the host's chromosome is the same as any fragment of the host's chromosome. The authors of this method proposed different variants in terms of k-mer size (with k = 2 or k = 3) and distance definition (Suzuki et al. 2008, 2010). The ones used here were the ones regarded as providing the best results in these studies.

We compared ICEs and CPs in relation to their compositional distance to the host. For this, we made the null hypothesis that the proportion of ICEs having a *P* value lower than CPs follows a binomial distribution whose expected proportion is that of the entire data set (the proportion of ICEs having a *P* value lower than CP), precisely: $H_0 = N(P\ \text{value}_{ICE} < P\ \text{value}_{CP})/N_{Comparisons}$, where $N_{Comparisons}$ is the total number of ICE–CP pairs, that is, $151 \times 136 = 20,536$.

Clustering of the Elements with the Weighted Gene Repertoire Relatedness Score

The relationship between two elements was quantified with the weighted Gene Repertoire Relatedness score (wGRR). This score represents the number of homologous proteins between two elements, weighted by their sequence identity, as described in (Cury et al. 2017). The formula is:

$$wGRR_{A,B} = \sum_{i=1}^{N} \frac{id(A_i,B_i)}{min(A,B)} \overset{\leftrightarrow}{iff}\ evalue(A_i,B_i) < 10^{-5}$$

Where (A_i, B_i) is the *i*th pair among N pairs of homologous proteins between element A and element B, $id(A_i, B_i)$ is the sequence identity of their alignment, $min(A, B)$ is the number of proteins of the element with fewest proteins (A or B). The sequence identity was computed with blastp v.2.2.15 (default parameters) (Altschul et al. 1997) and kept all bidirectional best hits with an e-value $<10^{-5}$.

The groups were inferred from the wGRR matrix using the Louvain algorithm (Blondel et al. 2008). We controlled for the consistency of the heuristic used, by assessing that the group

founds are robust. We performed 100 clustering, which led to the same classification in 95% of the time.

Incompatibility Typing

We determined the incompatibility group of replicons using the method of PlasmidFinder (Carattoli et al. 2014). We used BLASTN (Altschul et al. 1997) to search the replicons for sequences matching the set of 116 probes used by PlasmidFinder. We kept the hits with a coverage >60% and sequence identity >80%, as recommended by the authors. Around 3% of the elements had multiple incompatibility types attributed.

Data Availability

The data produced by these analyses are available at: https://gitlab.pasteur.fr/gem/ICE_CP, Accessed June 2018.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This work was supported by an European Research Council grant (EVOMOBILOME no. 281605), and a grant from the Agence National de la Recherche (MAGISBAC, ANR-14-CE10-0007). Work in FdlC lab was supported by grants BFU2014-55534-C2-1-P and BFU2014-62190-EXP from the Spanish Ministry of Economy and Competitiveness. We thank Alan Grossman and Marie Touchon for comments and suggestions, and Aude Bernheim for providing the phylogenetic tree of Proteobacteria. J.C. is a member of the "Ecole Doctorale Frontière du Vivant (FdV)—Programme Bettencourt."

References

Abby SS, Néron B, Ménager H, Touchon M, Rocha EPC. 2014. MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems. *PLoS One* 9(10):e110726.

Achaz G, Boyer F, Rocha EPC, Viari A, Coissac E. 2007. Repseek, a tool to retrieve approximate repeats from large DNA sequences. *Bioinformatics* 23(1):119–121.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25(17):3389–3402.

Bellanger X, Payot S, Leblond-bourget N, Guédon G. 2014. Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. *FEMS Microbiol Rev*. 38(4):720–760.

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. 2008. Fast unfolding of communities in large networks. *J Stat Mech.* 2008(10):P10008.

Brochet M, Da Cunha V, Couvé E, Rusniok C, Trieu-Cuot P, Glaser P. 2009. Atypical association of DDE transposition with conjugation specifies a new family of mobile elements. *Mol Microbiol*. 71(4):948–959.

Canchaya C, Proux C, Fournous G, Bruttin A, Brussow H. 2003. Prophage genomics. *Microbiol Mol Biol Rev*. 67(2):238–276.

Carattoli A, Zankari E, Garcìa-Fernàndez A, Larsen MV, Lund O, Villa L, Aarestrup FM, Hasman H. 2014. In silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother*. 58(7):3895–3903.

Carnoy C, Roten C-A. 2009. The dif/Xer recombination systems in proteobacteria. *PLoS One* 4(9):e6531.

Carraro N, Burrus V. 2014. Biology of three ICE families: sXT/R391, ICEBs1, and ICESt1/ICESt3. *Microbiol Spectr*. 2:1–20.

Carraro N, Burrus V. 2015. The dualistic nature of integrative and conjugative elements. *Mob Genet Elem.* 5(6):98–102.

Carraro N, Libante V, Morel C, Charron-Bourgoin F, Leblond P, Guédon G. 2016. Plasmid-like replication of a minimal streptococcal integrative and conjugative element. *Microbiology* 162(4):622–632.

Carraro N, Poulin D, Burrus V. 2015. Replication and active partition of integrative and conjugative elements (ICEs) of the SXT/R391 family: the line between ICEs and conjugative plasmids is getting thinner. *PLoS Genet*. 11(6):e1005298.

César CE, Machón C, De La Cruz F, Llosa M. 2006. A new domain of conjugative relaxase TrwC responsible for efficient oriT-specific recombination on minimal target sequences. *Mol Microbiol*. 62(4):984–996.

Clewell DB, Flannagan SE, Jaworski DD. 1995. Unconstrained bacterial promiscuity: the Tn916–Tn1545 family of conjugative transposons. *Trends Microbiol*. 3(6):229–236.

Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol*. 10:210.

Cury J, Jové T, Touchon M, Néron B, Rocha EP. 2016. Identification and analysis of integrons and cassette arrays in bacterial genomes. *Nucleic Acids Res*. 44(10):4539–4550.

Cury J, Touchon M, Rocha EPC. 2017. Integrative and conjugative elements and their hosts: composition, distribution and organization. *Nucleic Acids Res*. 45:1–14.

De Gelder L, Williams JJ, Ponciano JM, Sota M, Top EM. 2008. Adaptive plasmid evolution results in host-range expansion of a broad-host-range plasmid. *Genetics* 178(4):2179–2190.

Delavat F, Miyazaki R, Carraro N, Pradervand N, van der Meer JR. 2017. The hidden life of integrative and conjugative elements. *FEMS Microbiol Rev*. 41(4):512–537.

Dobrindt U, Hochhut B, Hentschel U, Hacker J. 2004. Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol*. 2(5):414–424.

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

Eddy SR. 2011. Accelerated Profile HMM Searches. *PLoS Comput Biol*. 7(10):e1002195.

Filipski A, Murillo O, Freydenzon A, Tamura K, Kumar S. 2014. Prospects for building large timetrees using molecular data with incomplete gene coverage among species. *Mol Biol Evol*. 31(9):2542–2550.

Francia MV, Clewell DB. 2002. Transfer origins in the conjugative Enterococcus faecalis plasmids pAD1 and pAM373: identification of the pAD1 nic site, a specific relaxase and a possible TraG-like protein. *Mol Microbiol*. 45(2):375–395.

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.

Garcillán-Barcia MP, de la Cruz F. 2008. Why is entry exclusion an essential feature of conjugative plasmids? *Plasmid* 60(1):1–18.

Gibson MK, Forsberg KJ, Dantas G. 2015. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J*. 9(1):207–216.

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.

Guérillot R, Da Cunha V, Sauvage E, Bouchier C, Glaser P. 2013. Modular evolution of TnGBSs, a new family of integrative and conjugative elements associating insertion sequence transposition, plasmid replication, and conjugation for their spreading. *J Bacteriol*. 195(9):1979–1990.

Guglielmini J, de la Cruz F, Rocha EPC. 2013. Evolution of conjugation and type IV secretion systems. *Mol Biol Evol*. 30(2):315–331.

Guglielmini J, Néron B, Abby SS, Garcillán-Barcia MP, la Cruz FD, Rocha EPC. 2014. Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. *Nucleic Acids Res*. 42(9):5715–5727.

Guglielmini J, Quintais L, Garcillán-Barcia MP, de la Cruz F, Rocha EPC. 2011. The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. *PLoS Genet*. 7(8): e1002222.

Guiney DG. 1982. Host range of conjugation and replication functions of the *Escherichia coli* sex plasmid Flac. Comparison with the broad host-range plasmid RK2. *J Mol Biol*. 162(3):699–703.

Hall JPJ, Williams D, Paterson S, Harrison E, Brockhurst MA. 2017. Positive selection inhibits gene mobilization and transfer in soil bacterial communities. *Nat Ecol Evol*. 1(9):1348–1353.

He S, Chandler M, Varani AM, Hickman AB, Dekker JP, Dyda F. 2016. Mechanisms of evolution in high-consequence drug resistance. *MBio* 7(6):e01987-16.

Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, et al. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res*. 44(D1):D286–D293.

Jacob F, Schaeffer P, Wollman EL. 1960. Episomic element in bacteria. Symp Soc Gen Microbiol. 10:67–91.

Johnson CM, Grossman AD. 2015. Integrative and conjugative elements (ICEs): what they do and how they work. *Annu Rev Genet*. 49:577–601.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.

Kers JA, Cameron KD, Joshi MV, Bukhalid RA, Morello JE, Wach MJ, Gibson DM, Loria R. 2005. A large, mobile pathogenicity island confers plant pathogenicity on Streptomyces species. *Mol Microbiol*. 55(4):1025–1033.

Klümper U, Riber L, Dechesne A, Sannazzarro A, Hansen LH, Sørensen SJ, Smets BF. 2015. Broad host range plasmids can invade an unexpectedly diverse fraction of a soil bacterial community. *ISME J*. 9(4):934–945.

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.

Lederberg J. 1998. Plasmid (1952–1997). *Plasmid* 39(1):1–9.

Lee CA, Babic A, Grossman AD. 2010. Autonomous plasmid-like replication of a conjugative transposon. *Mol Microbiol*. 75(2): 268–279.

Łobocka MB, Rose DJ, Plunkett G, Rusin M, Samojedny A, Lehnherr H, Yarmolinsky MB, Blattner FR. 2004. Genome of bacteriophage P1. *J Bacteriol*. 186(21):7032–7068.

Loftie-Eaton W, Bashford K, Quinn H, Dong K, Millstein J, Hunter S, Thomason MK, Merrikh H, Ponciano JM, Top EM. 2017. Compensatory mutations improve general permissiveness to antibiotic resistance plasmids. *Nat Ecol Evol*. 1(9):1354–1363.

Nunes-Düby SE, Kwon HJ, Tirumalai RS, Ellenberger T, Landy a. 1998. Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res*. 26(2):391–406.

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

Oliveira PH, Touchon M, Cury J, Rocha EPC. 2017. The chromosomal organization of horizontal gene transfer in Bacteria. *Nat Commun*. 8(1):1–10.

Oliveira PH, Touchon M, Rocha EPC. 2014. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res*. 42(16):10618–10614.

Oliveira PH, Touchon M, Rocha EPC. 2016. Regulation of genetic flux between bacteria by restriction-modification systems. *Proc Natl Acad Sci U S A*. 113(20):5658–5663.

Porse A, Schønning K, Munck C, Sommer MOA. 2016. Survival and evolution of a large multidrug resistance plasmid in new clinical bacterial hosts. *Mol Biol Evol*. 33(11):2860–2873.

Price MN, Dehal PS, Arkin AP. 2009. Fasttree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol*. 26(7):1641–1650.

Ravin NV. 2011. N15: the linear phage-plasmid. *Plasmid* 65(2):102–109.

Rocha EPC, Danchin A. 2002. Base composition bias might result from competition for metabolic resources. *Trends Genet*. 18(6):291–294.

San Millan A, Escudero JA, Gifford DR, Mazel D, Maclean RC. 2016. Multicopy plasmids potentiate the evolution of antibiotic resistance in bacteria. *Nat Ecol Evol*. 1(1):0010–0018.

San Millan A, MacLean RC. 2017. Fitness costs of plasmids: a limit to plasmid transmission. *Microbiol Spectr*. 5(5):1–12.

Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EPC, de la Cruz F. 2010. Mobility of plasmids. *Microbiol Mol Biol Rev*. 74(3):434–452.

Sullivan JT, Ronson CW. 1998. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Proc Natl Acad Sci U S A*. 95(9):5145–5149.

Summers DK. 1991. The kinetics of plasmid loss. *Trends Biotechnol*. 9(8):273–278.

Suzuki H, Sota M, Brown CJ, Top EM. 2008. Using Mahalanobis distance to compare genomic signatures between bacterial plasmids and chromosomes. *Nucleic Acids Res*. 36(22):e147.

Suzuki H, Yano H, Brown CJ, Top EM. 2010. Predicting plasmid promiscuity based on genomic signature. *J Bacteriol*. 192(22):6045–6055.

Touchon M, Cury J, Yoon E-J, Krizova L, Cerqueira GC, Murphy C, Feldgarden M, Wortman J, Clermont D, Lambert T, et al. 2014. The genomic diversification of the whole acinetobacter genus: origins, mechanisms, and consequences. *Genome Biol Evol*. 6(10):2866–2882.

Touchon M, Moura de Sousa JA, Rocha EP. 2017. Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer. *Curr Opin Microbiol*. 38:66–73.

Touchon M, Rocha EPC. 2016. Coevolution of the organization and structure of prokaryotic genomes. *Cold Spring Harb Perspect Biol*. 8(1):a018168–a018118.

Treangen TJ, Abraham AL, Touchon M, Rocha EPC. 2009. Genesis, effects and fates of repeats in prokaryotic genomes. *FEMS Microbiol Rev*. 33(3):539–571.

Wawrzyniak P, Płucienniczak G, Bartosik D. 2017. The different faces of rolling-circle replication and its multifunctional initiator proteins. *Front Microbiol*. 8:1–13.

Wu LJ, Errington J. 2002. A large dispersed chromosomal region required for chromosome segregation in sporulating cells of *Bacillus subtilis*. *EMBO J*. 21(15):4001–4011.

Zhong Z, Helinski D, Toukdarian A. 2005. Plasmid host-range: restrictions to F replication in Pseudomonas. *Plasmid* 54(1):48–56.