

Evolution of Plasmid Mobility: Origin and Fate of Conjugative and Nonconjugative Plasmids

Charles Coluzzi, Maria Pilar Garcillán-Barcia, Fernando de la Cruz, Eduardo P.C. Rocha

▶ **To cite this version:**

Charles Coluzzi, Maria Pilar Garcillán-Barcia, Fernando de la Cruz, Eduardo P.C. Rocha. Evolution of Plasmid Mobility: Origin and Fate of Conjugative and Nonconjugative Plasmids. Molecular Biology and Evolution, 2022, 39 (6), pp.msac115. 10.1093/molbev/msac115 . pasteur-03818036

HAL Id: pasteur-03818036

https://pasteur.hal.science/pasteur-03818036

Submitted on 17 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial| 4.0 International License

Evolution of Plasmid Mobility: Origin and Fate of Conjugative and Nonconjugative Plasmids

Charles Coluzzi [1],*,[1] Maria Pilar Garcillán-Barcia,[2] Fernando de la Cruz,[2] and Eduardo P.C. Rocha [1]

[1]Institut Pasteur, Université de Paris Cité, CNRS, UMR3525, Microbial Evolutionary Genomics, Paris 75015, France

[2]Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC), Universidad de Cantabria-CSIC, C/Albert Einstein 22, 39011 Santander, Spain

***Corresponding authors:** E-mails: charles.coluzzi@pasteur.fr; erocha@pasteur.fr.

Associate Editor: Irina Arkhipova

Abstract

Conjugation drives the horizontal transfer of adaptive traits across prokaryotes. One-fourth of the plasmids encode the functions necessary to conjugate autonomously, the others being eventually mobilizable by conjugation. To understand the evolution of plasmid mobility, we studied plasmid size, gene repertoires, and conjugation-related genes. Plasmid gene repertoires were found to vary rapidly in relation to the evolutionary rate of relaxases, for example, most pairs of plasmids with 95% identical relaxases have fewer than 50% of homologs. Among 249 recent transitions of mobility type, we observed a clear excess of plasmids losing the capacity to conjugate. These transitions are associated with even greater changes in gene repertoires, possibly mediated by transposable elements, including pseudogenization of the conjugation locus, exchange of replicases reducing the problem of incompatibility, and extensive loss of other genes. At the microevolutionary scale of plasmid taxonomy, transitions of mobility type sometimes result in the creation of novel taxonomic units. Interestingly, most transitions from conjugative to mobilizable plasmids seem to be lost in the long term. This suggests a source-sink dynamic, where conjugative plasmids generate nonconjugative plasmids that tend to be poorly adapted and are frequently lost. Still, in some cases, these relaxases seem to have evolved to become efficient at plasmid mobilization in trans, possibly by hijacking multiple conjugative systems. This resulted in specialized relaxases of mobilizable plasmids. In conclusion, the evolution of plasmid mobility is frequent, shapes the patterns of gene flow in bacteria, the dynamics of gene repertoires, and the ecology of plasmids.

Key words: **plasmids, conjugation, horizontal gene transfer, bacteria, genomes.**

Introduction

Bacteria acquire DNA from other cells, eventually of different species, by multiple mechanisms of horizontal gene transfer (HGT). The expression of these novel genes can provide adaptive phenotypic shifts (de la Cruz and Davies 2000; Soucy et al. 2015). In many species, HGT is in large part, or entirely, driven by the transfer of mobile genetic elements (MGEs) carrying accessory genes of adaptive value (Rankin et al. 2011; Hall et al. 2017; Partridge et al. 2018). Some of the most abundant MGEs, conjugative plasmids and integrative conjugative elements (ICEs), encode a mating pair formation (MPF) machinery with a conserved set of genes responsible for pilus biogenesis and mating junctions. The MPF includes a type 4 secretion system (T4SS) which transfers one DNA strand from the donor to the recipient cell (de la Cruz et al. 2010) (fig. 1). In plasmids, conjugation starts by the action of a relaxase, a multi-domain protein with transesterification activity that nicks the DNA molecule at the origin of transfer (*oriT*) and links covalently to a single strand (Gonzalez-Perez et al. 2007). The nicking of *oriT* also initiates a rolling-circle replication of the plasmid in the donor cell. The nucleoprotein filament is then presented to the T4SS by a type 4 coupling protein and transferred to the other cell where it is circularized. Finally, the replication machinery of the recipient cell restores the double strand of the plasmid. The amounts of contiguous DNA that can be transferred by conjugation are without equal among the mechanisms of HGT, since entire chromosomes can be transferred in one single event (Adelberg and Pittard 1965). Conjugation can also transfer genes across very distantly related species (Jain and Srivastava 2013).

Both ICEs and conjugative plasmids facilitate the conjugative transfer of MGEs that are not capable of autonomous conjugation, so-called mobilizable plasmids and integrative mobilizable elements (reviewed in Ramsay and Firth 2017). Mobilization in trans occurs when the element's horizontal transfer uses a conjugative system encoded by another one (fig. 1A). The best studied cases concern plasmids lacking an MPF machinery but encoding an *oriT* and a relaxase. In such plasmids, the relaxase interacts with the *oriT* and then the nucleoprotein filament is transferred by the T4SS encoded by another element. In this study, plasmids are called conjugative (pCONJ) when they encode a presumably complete machinery for

© The Author(s) 2022. Published by Oxford University Press on behalf of 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

Open Access

Downloaded from https://academic.oup.com/mbe/article/39/6/msac115/6593704 by CaRIS user on 18 October 2022

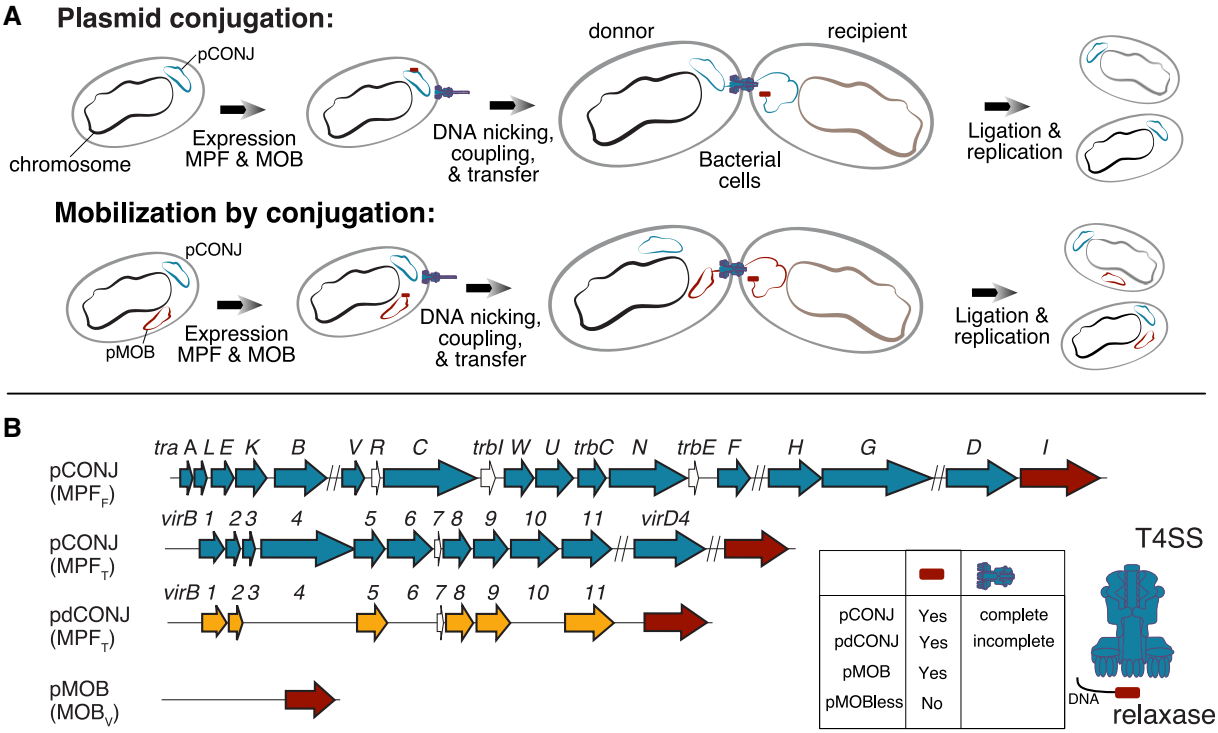


FIG. 1. Conjugative transfer and archetypal sets of conjugative transfer genes of conjugative and mobilizable mobile plasmids. (*A*) Different steps of the transfer by conjugation of conjugative and mobilizable plasmids. pCONJ, conjugative plasmid; pMOB, mobilizable plasmid. The red rectangle represents the relaxase. (*B*) Archetypal sets of transfer genes for conjugative plasmids (F and pTi), decayed conjugative plasmids (pdCONJ) and mobilizable plasmids (pMV158). Only conjugative genes are represented. The blue and orange arrows represent MPF genes and the red arrows represent relaxases. T4SS, type 4 secretion system.

conjugation (relaxase and MPF), decayed conjugative (pdCONJ) when they encode a relaxase but an incomplete MPF machinery, and mobilizable when they encode a relaxase gene but no or very few MPF genes (pMOB) (fig. 1*B*). It has often been observed that rates of conjugation of mobilizable plasmids are lower than those of the co-occurring conjugative plasmids (Perez-Mendoza et al. 2006; Blanca-Ordóñez et al. 2010; Klümper et al. 2014). Other MGEs, plasmids or elements integrated into the chromosome, only have an *oriT* (O'Brien et al. 2015; Yui Eto et al. 2021). In this case, the *oriT* interacts with a relaxase encoded in trans that presents the nucleoprotein filament to a T4SS. Since such plasmids only encode *oriT* for their mobility, and these sequences are more difficult to identify using bioinformatic tools, we will refer to all plasmids lacking relaxases as pMOBless. We usually ignore if they can be mobilized by conjugation. Work done a decade ago has shown that many completely sequenced genomes have plasmids, among which there is an approximately similar number of conjugative and mobilizable plasmids (Smillie et al. 2010). Mobilizable plasmids tend to be smaller than conjugative plasmids, presumably because they do not need to encode the large locus of the MPF machinery that may include dozens of genes (Smillie et al. 2010). Interestingly, associations between small and large plasmids are more common than expected by chance alone (San Millan et al. 2014), which suggests that the presence of large conjugative elements favors the presence of smaller mobilizable ones.

Phylogenetic and functional studies have revealed that conjugative plasmids and ICEs are very similar from the point of view of their conjugation machinery and interconversions between the two can occur (Guglielmini et al. 2011; Johnson and Grossman 2015; Cury et al. 2018). There are eight large groups of MPF, two are specific to monoderms (cells lacking an outer membrane), including Archaea, and the remaining are found in diderms (bacteria with an outer membrane) (Guglielmini et al. 2013). There are nine well-known classes of relaxases that vary in their sequence, number and type of protein domains, and sometimes in how they catalyze chemical reactions (Garcillan-Barcia et al. 2020). Novel putative families of relaxases are still being unraveled by computational studies (Coluzzi et al. 2017). The relaxase (that defines the MOB class) is one of the few key taxonomic traits of plasmids (Garcillan-Barcia et al. 2009). Recently, the analysis of the average nucleotide identity (ANI) between plasmids revealed that closely related plasmids cluster in so-called plasmid taxonomic units (PTUs) (Redondo-Salvo et al. 2020). PTUs group plasmids with many highly similar homologs in common. They tend to have one single type of relaxase and characteristic host ranges that go beyond the species barrier.

Downloaded from https://academic.oup.com/mbe/article/39/6/msac115/6593704 by Institut Pasteur - CeRIS user on 18 October 2022

Conjugative and mobilizable elements transfer many kinds of traits, including virulence factors and genes involved in symbiosis (Nuti et al. 1979; Johnson and Nolan 2009; Carattoli 2013). Their ability to transfer large amounts of DNA per event allows the spread of complex traits encoded in long genetic loci (Kobayashi 2018; Geng et al. 2021). The host range of conjugation can be very broad, for example, conjugation can take place between Proteobacteria and Firmicutes (Trieu-Cuot et al. 1987) or from Proteobacteria to plants (Lacroix and Citovsky 2018). This trait may explain why conjugative plasmids are key vectors of the transfer of antibiotic resistance genes from distantly related bacteria to nosocomial pathogens (Pedersen et al. 2018; Che et al. 2021). But conjugation is also costly because it requires production, assembly, and functioning of a large protein complex that crosses multiple membranes and renders bacteria sensitive to certain phages (San Millan and Craig MacLean 2019). In this context, the added cost of having a mobilizable plasmid might be low because the costly conjugative apparatus is encoded by the conjugative element. Hence, mobilizable elements depend on an MPF encoded in trans by a conjugative element but may be less costly to the cell than the latter. The dependency of these plasmids on other conjugative elements can have some advantages. Some pMOB can be mobilized by different conjugative systems in trans (Garcillan-Barcia et al. 2019), allowing mobilizable plasmids to increase their frequency of transfer. As a result, traits encoded in conjugative or mobilizable elements (such as virulence factors, metabolic pathways, heavy metal, or antibiotic resistance genes) may use different transmission paths. Such traits may also transit from one type of plasmid to another, since there are interconversions between types of plasmids and there is gene flow between conjugative and mobilizable elements (Liu et al. 2013). To understand the evolution of conjugative elements and their impact on bacterial genomes, one thus needs to understand the evolutionary interplay between conjugative and mobilizable elements.

Although the macroevolution of the machinery for conjugation, the MPF, has been described in detail (Guglielmini et al. 2013), we know relatively little of the general patterns of evolution of plasmids relative to their categorization as mobilizable or conjugative. To shed light on this issue, we analyzed a large set of plasmids in terms of their mobility by conjugation. Plasmids are better suited for such studies than ICEs because they are better known and easier to delimit accurately in complete genome sequences. To understand the evolution of conjugative and mobilizable plasmids, we assessed the relative frequency of each relaxase in mobilizable and conjugative plasmids and used the relaxases as phylogenetic markers to trace past events of change from one type of mobility to another, that is, pCONJ to pMOB or vice-versa. In this study, we assume that if a relaxase is present in a plasmid it renders it mobilizable because we cannot assess its level or conditions of expression at this scale. We also cannot know if a gene has inactivating point mutations (but we can and have removed pseudogenes). Such limitations are probably not very impactful at the statistical level since one expects most complete systems to be functional and expressed at least under certain conditions. Transitions between mobility types can occur by several mechanisms. Deletions may change a pCONJ into a pMOB (but not the other way around). Cointegration of a pMOB and a pCONJ (or translocation of the relaxase gene into a pCONJ) results in a relaxase of a pMOB suddenly becoming a relaxase of a pCONJ. Here, we analyzed the co-occurrence of relaxases in plasmids and the frequencies with which relaxases that were associated with a given type of plasmid become associated with another. These analyses revealed complex events, but were often consistent with a source-sink dynamics where conjugative plasmids frequently evolve to generate pMOBs and pMOBless. Transitions in terms of mobility are associated with major disruptions of gene repertoires and sometimes result in novel plasmid types. They also suggest that relaxases endure different selective pressures in conjugative and mobilizable elements.

Materials and Methods

Plasmid Dataset

The dataset used in this study consists of 13,525 complete genomes from 2,421 species retrieved from NCBI RefSeq database of high-quality complete nonredundant prokaryotic genomes (ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/, last accessed in May 2019). These genomes contain 11,805 plasmids. To avoid the misidentification of ICEs as conjugative plasmids in chromids or secondary chromosomes, we excluded from further study the 419 plasmids larger than 500 kb (Harrison et al. 2010; Smillie et al. 2010). The complete list of 11,386 plasmids can be found in supplementary table S1, Supplementary Material online.

Detection of Conjugative Systems

The detection of conjugative systems was performed using MacSyFinder v.1.0.5 (Abby et al. 2014). Briefly, MacSyFinder uses HMM protein profiles and a set of rules (which are defined in MPF model files) about their presence and genetic organization to identify occurrences of given MPFs and relaxases in a genome. MacSyFinder was used with default parameters (HMMer *e*-value < 0.001, HMM (Hidden Markov Model) profile alignment coverage $>50\%$) and was run independently for each MPF type. To detect conjugative systems in bacterial chromosomes, we used protein HMM profiles and the MPF system definitions as described before (Cury et al. 2017).

To detect conjugative systems in plasmids, we changed the default procedure to allow for relaxase genes that are distant from the genes encoding the MPF. Hence, relaxases and T4SS were searched independently. We searched for relaxases in plasmids and in chromosomes using the same protein profiles. We used HMMer v.3.2.1 (*e*-value < 0.001, HMM profile alignment coverage $>50\%$)

Downloaded from https://academic.oup.com/mbe/article/39/6/msac115/6593704 by Institut Pasteur - CeRIS user on 18 October 2022

(Eddy 2011). When a relaxase matched two different HMM profiles, the hit having the smallest *e*-value was selected. The T4SS were searched using MacSyFinder with modified MPF system definitions (command line: macsyfinder <model> –db-type ordered_replicon -d < definitions_path> -p < profiles_path> –profile-suffix.hmm –sequence-db < Proteins_path.prt> -o < out_path>): the modified definitions allow the relaxase to be absent and the MPF proteins to be at any distance from each other's in the replicon (contrary to the 60 kb of space for the chromosome models). Replicons encoding both a relaxase and a complete MPF were considered conjugative. The macsyfinder models adapted specifically for this study can be found in supplementary materials, Supplementary Material online.

Recently, a previously described class of relaxase, MOB_{MG} (Tomita and Ike 2005) was named MOB_L after it was identified in Firmicutes (Ramachandran et al. 2017). We made a protein profile for this group of proteins with HMMER v.3.2.1 using the 817 protein sequences described as MOB_L homologs in (Ramachandran et al. 2017). The search for elements of this family of relaxases revealed an important overlap with relaxases detected with the MOBP2 HMM profile, in agreement with previous works (Garcillan-Barcia et al. 2020), and we decided not to use the MOBL profile in this study.

Assessment of Plasmid Mobility

A plasmid was considered conjugative (pCONJ) when it encoded one or several relaxases, a VirB4, a T4CP, and a minimum number of additional MPF proteins. The latter threshold varies according to the MPF type: two proteins for MPF_{FA} and MPF_{FATA} and three for the other MPF types (B, C, F, G, I, and T). These are values lower than previously used in Cury et al. (2018) to make the assessment of a transition from pCONJ to pMOB conservative. A plasmid was considered mobilizable (pMOB) when it encoded a relaxase but lacked the conditions to be conjugative (misses T4CP, VirB4, or a sufficient number of the other components). Plasmids encoding a relaxase and <5 MPF proteins were classed pMOBs (supplementary fig. S1, Supplementary Material online). To pinpoint cases where plasmids encode several genes typical of conjugative elements, but not enough to allow classing them as pCONJ, we classed mobilizable plasmids containing more than five proteins involved in the conjugative transfer (apart from the first relaxase, either MPF proteins, T4CP, or additional relaxases) as pdCONJ. Script used to parse mobility data can be found in supplementary materials, Supplementary Material online.

Relaxase Clustering, Alignments, and Phylogeny

We clustered the 5,666 relaxase protein sequences using MMSeqs2 *cluster* (v. 9-d36de) with 99% identity and alignment coverage (on both query and target) thresholds (parameters –min-seq-id 0.99 –cov-mode 0 -c 0.99) (Hauser et al. 2016). The protein sequences of the relaxases of each type were then aligned using MAFFT v.7.245 with the E-INS-I algorithm (parameters –genafpair –maxiterate 1000) (Katoh and Standley 2014). The multiple alignments were trimmed using ClipKIT with the "kpic-gappy" algorithm which keeps only parsimony informative and constant sites and removed all sites above 90% gappyness (command line: clipkit <MSA> -m kpic-gappy) (Steenwyk et al. 2020). The phylogenetic analyses were performed on the resulting multiple alignment using IQ-TREE (v.1.5.5, Nguyen et al. 2015) with the ultrafast bootstrap option (1000 bootstraps) and with the best fitting model estimated *ModelFinder Plus* (-MPF) for each class of relaxases according to the BIC criterion (command line: Command: iqtree –s <MSA.clipkit> -m MPF -bb 1000). Trees were rooted using the midpoint function from the phangorn packages (v.2.5.5) for R (command line: midpoint[tree]).

HMM–HMM Alignments

The protein alignments of the relaxases were retrieved from Guglielmini et al. (2011) and then used to produce HMM profiles with hhmake (command line: hhmake -I <msa> -o HHM -cons) from HH-suite3.0 program v.3.3.0 (Steinegger et al. 2019). The alignment of HMM profiles was performed using *hhalign* from the same suite of programs with default parameters (command line: hhalign -t HMM1 -i HMM2 -M 50 -add_cons). We took into consideration only bidirectional alignments with a probability superior to 70% and an *e*-value inferior to 10^{-4}.

Typing of Plasmid Replicons

To identify plasmid replicons, we used PlasmidFinder 2.0.1 with the 2020-07-13 database and default parameters (Carattoli et al. 2014). We identified 200 replicon groups in 4,827 of the 11,386 plasmids.

wGRR Estimation

We searched for significant similarity (*e*-value $< 10^{-4}$, identity $\geq 35\%$, coverage $\geq 50\%$) among all pairs of plasmid proteins using MMseqs2 (v. 9-d36de). The best bidirectional hits (BBH) between pairs of elements were then used to calculate the weighted gene repertoire relatedness (wGRR) (Cury et al. 2018):

$$\mathrm{wGRR}(A, B) = \frac{\sum_i^P id(A_i, B_i)}{\min(\#A, \#B)}$$

where *Ai* and *Bi* are the *i*th BBH pair of *P* total pairs, $id(A_i, B_i)$ is the identity between the BBH pair, and $\min(\#A, \#B)$ is the number of genes encoded by the plasmid encoding fewer genes. The wGRR varies between 0 and 1. A wGRR close to 1 means that both plasmids are very similar (all genes in an element have a very similar BBH in the other) and a wGRR of 0 means that plasmids lack homologs.

Downloaded from https://academic.oup.com/mbe/article/39/6/msac115/6593704 by Institut Pasteur - CeRIS user on 18 October 2022

PTU Predictions

The taxonomic classifier of plasmids, COPLA, was used to assign plasmids to taxonomic units with default parameters (Redondo-Salvo et al. 2021).

ANI Calculation

ANI scores were calculated using the ani.rb Ruby script from Enveomics Collection (https://github.com/lmrodriguezr/enveomics/tree/master/Scripts), as detailed in Redondo-Salvo et al. (2020). Briefly, pairwise ANI scores were obtained by splitting every genome into fragments of 1.0 kb using a sliding window algorithm with a step of 200 bp. The resulting sets of fragments were compared by all against all BLASTn, with a 70% identity threshold over 70% of the fragment length. BBH were selected and, if the number of selected fragments covered at least 50% of the smallest plasmid, an ANI score was assigned. The network layout was obtained using Gephi (Bastian et al. 2009).

Plasmid Proteome Network Analysis

We used AcCNET v 1.2 to build the plasmid protein sequence similarity network (command line: perl accnet.pl –threshold 1.1 –kp "-s 5.03 -c 0.8 -e 1e-4 -M 35000MB" –fast = yes –clustering = yes –clean = no –in *.faa) (Lanza et al. 2017). Homologous protein clusters were generated using kClust (v. 1.0) (Hauser et al. 2013), with >95% protein identity, >80% alignment coverage, and clustering *e*-value $< 10^{-4}$. All edges were assigned equal weights. The network layout was obtained using Gephi (Bastian et al. 2009).

Inference of Ancestral States

We inferred the ancestral state of each plasmid mobility type (pCONJ, pMOB, and pdCONJ) with PastML (v1.9.33) (Ishikawa et al. 2019). We used the maximum-likelihood algorithm marginal posterior probabilities approximation (MPPA) and the F81 model, as recommended by the authors. The MPPA algorithm chooses for every node a subset of ancestral states that minimizes the prediction error measured by the Brier score. Hence, it may keep multiple state predictions per node but only when they have similar and high probabilities (command line: pastml -t MOB_tree.nwk -d Mobility_states.csv -s ',' –prediction_method MPPA -o pastml.out). To avoid overestimation of the transitions between states, we counted transitions only when the ancestral node had a unique state that was different from the derived state. This may result in a conservative estimate of the total number of transitions. Out of the 287 observed transitions for all trees combined, 38 were removed because of this stringent criterion, leaving 249 transitions to be analyzed.

Cumulative Probability of Transitions

For each tree, we computed the patristic distance matrix using the cophenetic.phylo function from the package ape (v5.3) for R (command line: PatristicDist Matrix<-cophenetic[tree]) (Paradis and Schliep 2019). Based on the patristic distance matrices, we measured the patristic distance separating each relaxase in our dataset from the closest relaxase in the phylogenetic tree that was associated with a different mobility type. For relaxases not present in the tree, that is, those that are in a cluster but were not included in tree as the cluster representative, we used the patristic distance separating the representative of the cluster and the closest relaxase in the tree having a different mobility type. When a relaxase had a different mobility than its cluster representative we used a patristic distance of 0. This provides for each plasmid a patristic distance to the closest plasmid of a different type. The empirical cumulative distribution function (CDF) of these patristic distances was computed for each MOB class and each mobility type, using the ECDF built-in function for python (v3.6.10). Empirical cumulative distribution plots were computed using the ecdfplot function from the seaborn package (v0.11.0) for python (command line: sns.ecdfplot()) (Waskom 2021).

Mobile Plasmid-pMOBless Associations

We paired each pMOB, pCONJ, and pdCONJ with the pMOBless plasmid with the highest wGRR (when wGRR > 0.75). To avoid the overrepresentation of the same transition event, if a pMOBless plasmid was paired that way with multiple mobile plasmids, we only kept the mobile/pMOBless pair with the highest wGRR value.

Partition Proteins and Toxin–Antitoxins Proteins Detection

To detect partition proteins carried by plasmids, we used HMMER v.3.2.1 (*e*-value < 0.001 and coverage of 50%) with HMM protein profiles for partition proteins described in Cury et al. (2017). To detect toxin and antitoxin proteins in plasmids, we used HMMER v.3.2.1 (*e*-value < 0.001 and coverage of 50%) with toxin and antitoxin HMM profiles from the TASmania database (Akarsu et al. 2019).

Homologies Between Plasmids and Chromosomal Sequences

To assess if plasmids that recently changed in terms of type of mobility had acquired chromosomal sequences from their hosts, we searched for homologies between the plasmids and chromosomal sequences. For each plasmid, we collected all chromosomal sequences from the whole host species. We then searched for homologies longer than 5 kb (2 kb in a complementary analysis) having more than 90% nucleotide identity using Blastn with default parameters. The percentage of coverage was calculated by adding the nonoverlapping length of the alignments and dividing it by the length of the plasmid.

Statistics

Unless mentioned otherwise all statistics were performed within R (v3.6.2). Statistics between two variables were

Downloaded from https://academic.oup.com/mbe/article/39/6/msac115/6593704 by Institut Pasteur - CeRIS user on 18 October 2022

done using standard nonparametric tests (Wilcoxon test). χ^2 tests were performed using the R (v3.6.2) χ^2 test built-in function.

Results

Cooccurrence and Wide Distribution of Conjugative and Mobilizable Plasmids

We identified relaxases and conjugative systems among 11,386 plasmids smaller than 500 kb from the completely assembled bacterial genomes of RefSeq (supplementary table S1, Supplementary Material online). The upper size threshold was imposed to avoid the inclusion in the analysis of mis-assigned secondary chromosomes that might contain ICEs or IMEs. Plasmids were classed in four classes (fig. 2A): 53% of the plasmids lacked relaxases (pMOBless), 23% were classed as conjugative (pCONJ) because they encode a relaxase and the required MPF genes (see Materials and Methods), 3% were classed as pdCONJ because they encoded more than five MPF genes but lacked the quorum of genes required to be conjugative, and 21% were classed as pMOB because they encoded a relaxase and five or fewer MPF genes. Hence, mobilizable and conjugative plasmids are present at similar frequencies and make up about half of all plasmids, consistent with previous results with smaller datasets (Smillie et al. 2010). The average size of plasmids is correlated with the type of mobility, with pCONJs (median size of 111.7 kb) being slightly, but significantly, larger than pdCONJs (96 kb, $P = 0.0015$) and the latter being much larger than pMOBs (37 kb, $P < 0.001$) or pMOBless (39 kb, $P < 0.001$, Wilcoxon tests, fig. 2B). Although the average size of plasmids is different depending on the class of relaxase, the trends in terms of plasmid size in relation to the mobility type are similar across MOB classes (fig. 2C). Many plasmids encode insertion sequences (IS) that are known to promote recombination, cointegration, and transposition of genes between and within DNA molecules. We identified ISs in all plasmids using ISEscan v1.7.2.2 default options and found the highest densities in pdCONJs and the lowest in pMOBless (fig. 2D).

The dataset used in this work includes 5,666 relaxases of all the different MOB classes (fig. 2E). The classes of relaxases are usually described as MOB_X where X is the class identifier (Garcillan-Barcia et al. 2009). However, some of the subclasses of relaxases are very divergent. Hence, here we will use the nomenclature MOBX to refer to the proteins retrieved by the HMM protein profile for the relaxases class or subclass X, and MOB_X when we refer generically to the class. The most prevalent classes are MOB_F and MOB_P (particularly those retrieved by the MOBP1 HMM profile), which account for more than 62% of all relaxases and are present in both pCONJs and pMOBs. It is important to note that the distribution of relaxase families differs between conjugative and mobilizable plasmids ($\chi^2 = 1680.6$, $df = 18$, $P < 0.001$). For example, 95% of the MOB_V relaxases and 99% of those identified by the MOBP2 HMM profile are in mobilizable plasmids (pMOBs or pdCONJs) (fig. 2E). In contrast, 94% of MOB_H relaxases are in pCONJs. The pdCONJs have relaxases typical of pCONJs, sizes that are almost as large as those of pCONJ, and the largest density of transposable elements. This suggests that they were recently derived from pCONJ by gene loss.

We assessed the distribution of plasmid mobility across bacterial taxa (fig. 3). Except for the Chlorobi and the Chlamydiae, which have respectively one and two plasmids, all the taxa with conjugative plasmids also have mobilizable plasmids. Extensively sampled clades rarely lack conjugative elements. Their absence in Deinococcus-Thermus (67 plasmids) might be related with the existence of a novel transfer mechanism, transjugation, in these bacteria (Blesa et al. 2017). Actinobacteria and Bacilli have many more mobilizable than conjugative plasmids, possibly because conjugative systems in some of these clades are less described. Alternatively, it has been suggested that nonconjugative plasmids in *Staphylococcus* spp. could be mobilized in trans by conjugative plasmids (Ramsay et al. 2016) or by transduction (Humphrey et al. 2021). Also, it was previously observed that ICEs are much more frequent than conjugative plasmids in these clades (Guglielmini et al. 2011), and these pMOBs might be mobilized by ICEs. Finally, given the frequency of mobilizable and conjugative plasmids, we assessed how often mobilizable plasmids and conjugative systems cooccur in the same host cell. We found that 46% of pMOBs are in genomes that encode at least one conjugative system in a plasmid or in the chromosome. Hence, many pMOBs are in cells with a conjugative element and the others may eventually meet one by conjugation because most bacterial classes have both types of plasmids.

Classes of Relaxases Rarely Cooccur, but can Interact with Different MPFs

Although relaxases from conjugative systems are expected to have coevolved tightly with their cognate T4SS, those of pMOBs may have evolved the ability to interact with different systems to maximize their chances of transfer. To study this hypothesis, we analyzed the distribution of relaxases in plasmids. Most relaxase classes are present across bacterial phyla (fig. 3). For example, the MOB_P class was found in Firmicutes, Proteobacteria, and Bacteroidetes. In contrast, MPF types tend to occur in specific phyla, in agreement with Guglielmini et al. (2013). As a result, the most frequent classes of relaxases cooccur with many MPF types. For example, the hits of the profile MOBP1 cooccur with six different MPF types and the one of MOBF with four (supplementary fig. S2, Supplementary Material online). Mobilizable plasmids could encode multiple relaxases to increase the range of MPFs that a plasmid can use. Yet, this rarely occurs, since only 258 of the 11,386 plasmids encode two relaxases and only 33 encode more than two. Furthermore, the frequencies of these

Downloaded from https://academic.oup.com/mbe/article/39/6/msac115/6593704 by Institut Pasteur - CeRIS user on 18 October 2022

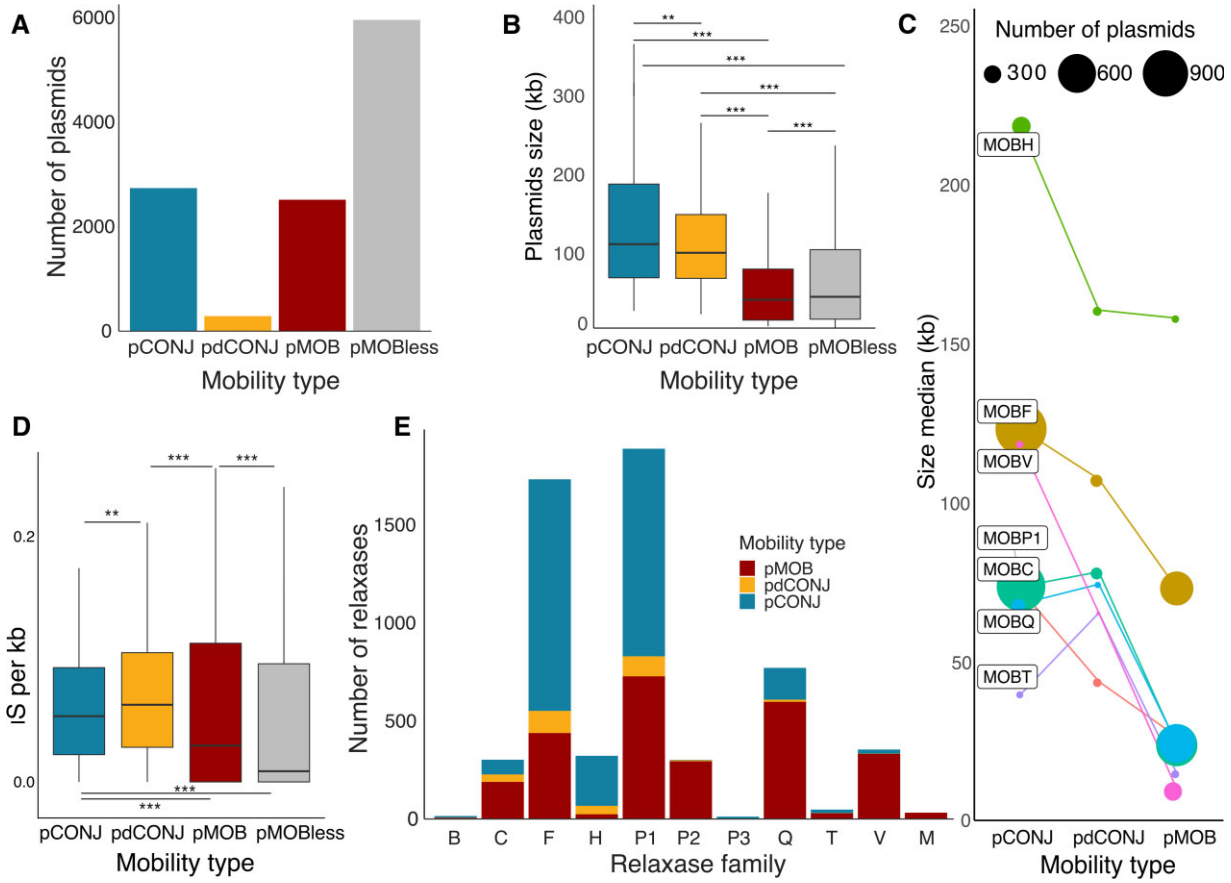


FIG. 2. Plasmid abundance (*A* and *E*), size (*B* and *C*), and IS density (*D*) according to mobility type and relaxase class. (*C*) The median plasmid size for relaxase classes with a significant number of elements for each mobility type. The width of the circles indicates the frequency of the plasmids (see inset legend). Tests of differences (Wilcoxon paired-tests): $^{***}P < 0.001$, $^{**}P < 0.01$.

cooccurrences are similar among pCONJs (143/291) and pMOBs (130/291), which is not in agreement with the expectation that pMOBs would use multiple relaxases to exploit different conjugative plasmids. Actually, more than 55% of the plasmids with two relaxases have proteins of the same MOB class (fig. 4A), and there is no evidence for more cooccurrence of different MOB classes than expected by chance ($\chi^2 = 3.2609$, $df = 1$, $P = 0.071$). These results suggest that relaxases of a given class can interact with very different MPFs. They also show that co-occurrence of multiple relaxases in a plasmid is rare, suggesting little selection for a pMOB to encode different relaxases to use a broader set of MPF.

Since these results were not suggestive of maximization of the diversity of relaxases cooccurring in plasmids, we tested if they reflected the deep evolutionary relations between relaxase classes using profile–profile alignments (fig. 4B). The relaxase classes MOB_F, MOB_H, MOB_T, MOB_M, and MOB_C showed no significant sequence homology with the others, whereas the remaining were all related. Although fitting previous studies (Garcillan-Barcia et al. 2009), this does not show clear parallels between sequence similarity and cooccurrence in the same plasmid. Importantly, relaxase classes lacking homology with MOB_P (MOB_H, MOB_F, MOB_C) are not specific to mobilizable plasmids (fig. 2). The apparent exceptions could be MOB_T and MOB_M, which are rare in conjugative plasmids. However, MOB_T is very frequent in ICEs (Soler et al. 2019) and MOB_M was first described in the conjugative plasmid pCW3 (Wisniewski et al. 2016). Therefore, they are also associated with conjugative elements.

The few cooccurrences of relaxases in the same plasmid could result from structural changes that affect plasmid mobility. Notably they could result from cointegration of a pMOB into a pCONJ (or another pMOB), in which case they could provide interesting information on these events. The most frequent relaxases of pMOBs (e.g., MOB_V, MOB_Q, or the MOB_P identified with the MOBP2 HMM profile) cooccur more frequently in pMOBs than in pCONJ. In contrast, relaxase classes found in both pCONJs and pMOBs (those identified with the profiles MOBF and MOBP1), or those that are mostly found in conjugative plasmids (MOB_H) cooccur more in pCONJ. These results indicate that either there is no frequent cointegration of pMOB and pCONJ or this cointegration rapidly leads to the loss of the relaxases of the plasmid that was originally a pMOB.

Early Specialization of Relaxases and Frequent Recent Changes in Mobility

To study the evolution of plasmid mobility we focused initially on the relaxases identified by MOBP1 (from the

Downloaded from https://academic.oup.com/mbe/article/39/6/msac115/6593704 by Institut Pasteur - CeRIS user on 18 October 2022

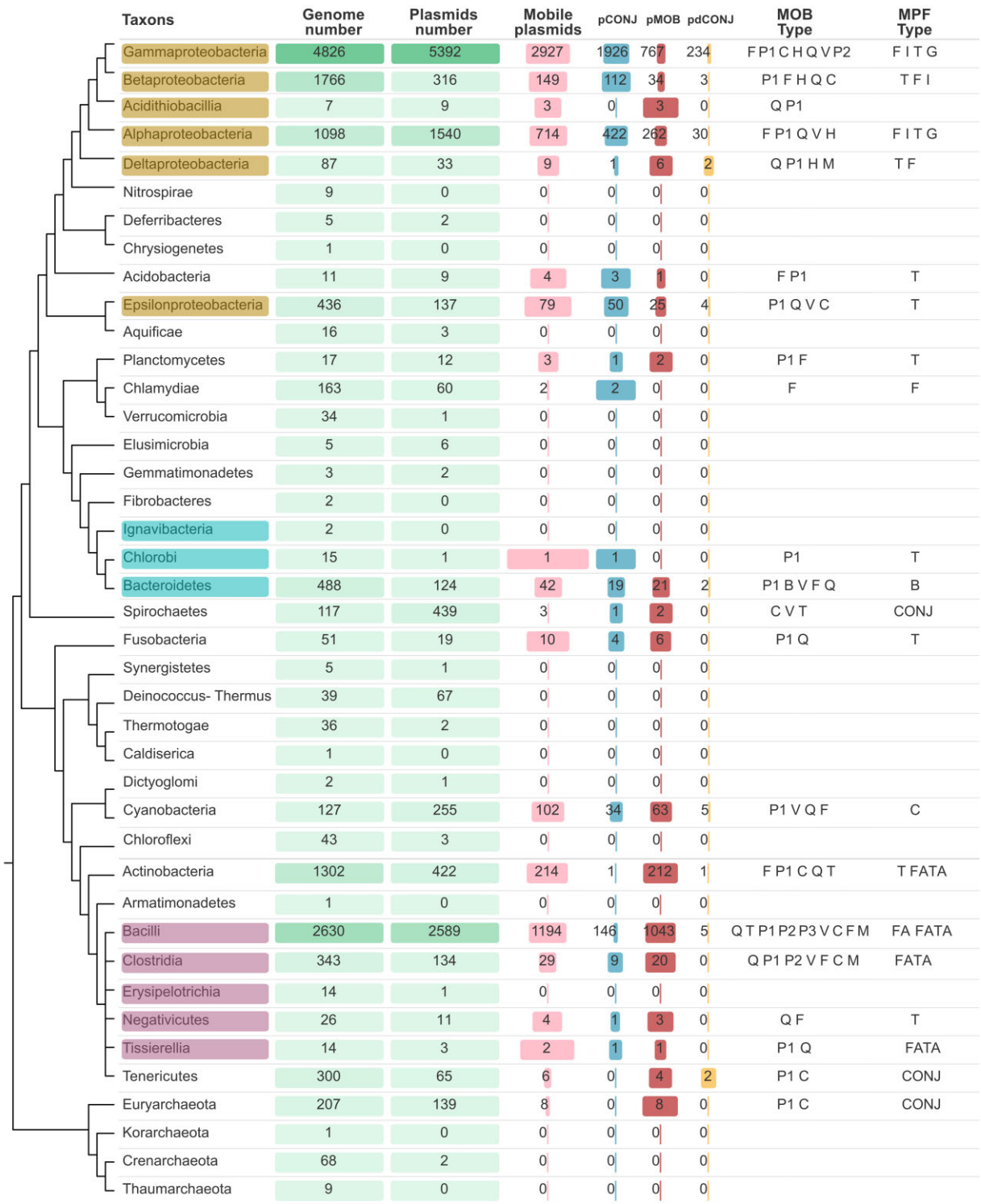


FIG. 3. Overview of plasmids and their mobility across the tree of life. The plasmids were grouped according to their NCBI taxonomic classification of the host bacteria. This sketch tree was drawn from the compilation of different published phylogenetic analyses (Denise et al. 2019). The Proteobacteria, Bacteroidetes, and Firmicutes phyla are highlight in gold, turquoise, and purple, respectively.

MOB_P family) and MOBF profiles. These are the most abundant (fig. 1*E*), they rarely cooccur in the same plasmid (fig. 4*A*), they are sufficiently conserved to be good phylogenetic markers, and they are frequent in both mobilizable and in conjugative plasmids (fig. 1*E*). Hence, we used the protein sequences of these relaxases for phylogenetic inference. Some of the relaxases were nearly identical in protein sequence. We removed this redundancy to accelerate phylogenetic reconstructions and to avoid uncertainty in the inferred ancestral states (which is inevitable when proteins are identical). To keep most of the genetic diversity of the dataset, we clustered the proteins that

Downloaded from https://academic.oup.com/mbe/article/39/6/msac115/6593704 by Institut Pasteur - CeRIS user on 18 October 2022

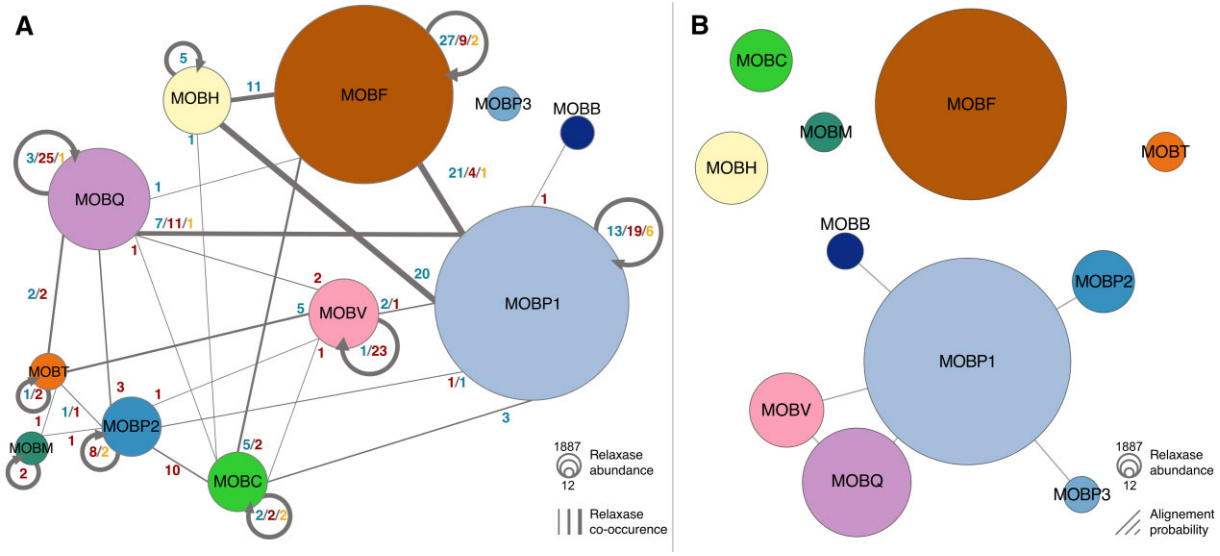


Fig. 4. Associations between relaxases. (*A*) Cooccurrence (edges) of relaxases within the same plasmid. The numbers in blue, red, and orange represent the number of cooccurrences found in pCONJ, pMOB, and pdCONJ, respectively. The thickness of the edges is proportional to the number of cooccurrences of the two relaxases in the same plasmids. (*B*) Representation of the results of profile–profile alignments between classes of relaxases. Edges represent bidirectional alignments between HMM profiles, and their length represents the probability of the alignment (short links represent a higher probability [up to 96%] and long links represent lower probability [down to 76%]). Are only represented alignments having a probability superior to 70% and an *e*-value inferior to 10^{-4}.

aligned over at least 99% of the length with at least 99% identity. This resulted in 850 MOB_P and 735 MOB_F homogeneous clusters of relaxases: 94.9% of MOB_P and 92.5% of MOB_F are present on only one type of plasmids (pCONJ, pMOB, or pdCONJ). We picked one representative protein per group and used them to make phylogenetic reconstructions by maximum likelihood (see details in Materials and Methods). The two trees are well supported, that is, they have high bootstrap values, in most branches (see fig. 5). In both cases, there is a clear tendency for closely related relaxases to be in plasmids with similar MPF types. This leads to large clades overrepresenting one MPF type (fig. 5). In the MOB_P tree, there is an overrepresentation of MPF_T which suggests it was the original MPF associated with this relaxase. The MOB_F tree is too variable at the basal clades to make any kind of conclusion regarding the ancestral associations. Differences in plasmid size are very pronounced across the tree, since many plasmids that are close in the tree exhibit wide variations in size. Hence, plasmid size varies fast relatively to the protein sequence of the relaxases.

The distribution of plasmid mobility in the two trees is suggestive of two types of processes. At a large evolutionary scale, we observe specialization of the relaxases in terms of plasmid mobility leading to a few large clades where almost all relaxases are in pMOB and other clades where most relaxases are in pCONJ. There are no large clades containing similar numbers of both types of plasmids intermingled in the phylogeny. This suggests a process of specialization of these relaxases that took place very early in natural history. In contrast, we observed at the microevolutionary scale a few pCONJ relaxases in the pMOB clades and some pMOB and pdCONJ in the pCONJ clades. These correspond to relatively recent changes in the classification of these relaxases: they used to be associated with one type of plasmid and recently became associated with another. These changes may have resulted from modifications in the plasmid, such as the loss of MPF genes, or from the translocation of the relaxases to a plasmid with a different type of mobility. Of note, there are more small groups of mobilizable plasmids within the pCONJ clades than vice-versa (see section below for a precise quantification). These groups of recent pMOBs (or pdCONJs) tend to be small and are present at or close to the terminal branches of the phylogenetic trees, which suggests that they rarely last for a long period of time. This can be due to either reversion of the process or to loss of the novel mobilizable plasmid by natural selection. To quantify these observations, the next sections focus on the quantification of the variation of plasmids' mobility and gene repertoires.

Plasmid Size and Gene Repertoire Evolve Fast Relative to the Relaxase Sequence

Gene repertoires of plasmids are known to vary very fast. To quantify their rates of change, we analyzed the differences of gene repertoires between pairs of plasmids in the light of the divergence of their relaxases (see schema in fig. 6). These analyses used all relaxases, that is, we did not remove those in clusters of very similar proteins, to maximize the statistical power of the analysis. We assessed the variations in gene repertoires between pairs of the 3,501 plasmids of our dataset encoding relaxases identified by the protein profiles MOBP1 and MOBF using the wGRR.

Downloaded from https://academic.oup.com/mbe/article/39/6/msac115/6593704 by Institut Pasteur - CeRIS user on 18 October 2022

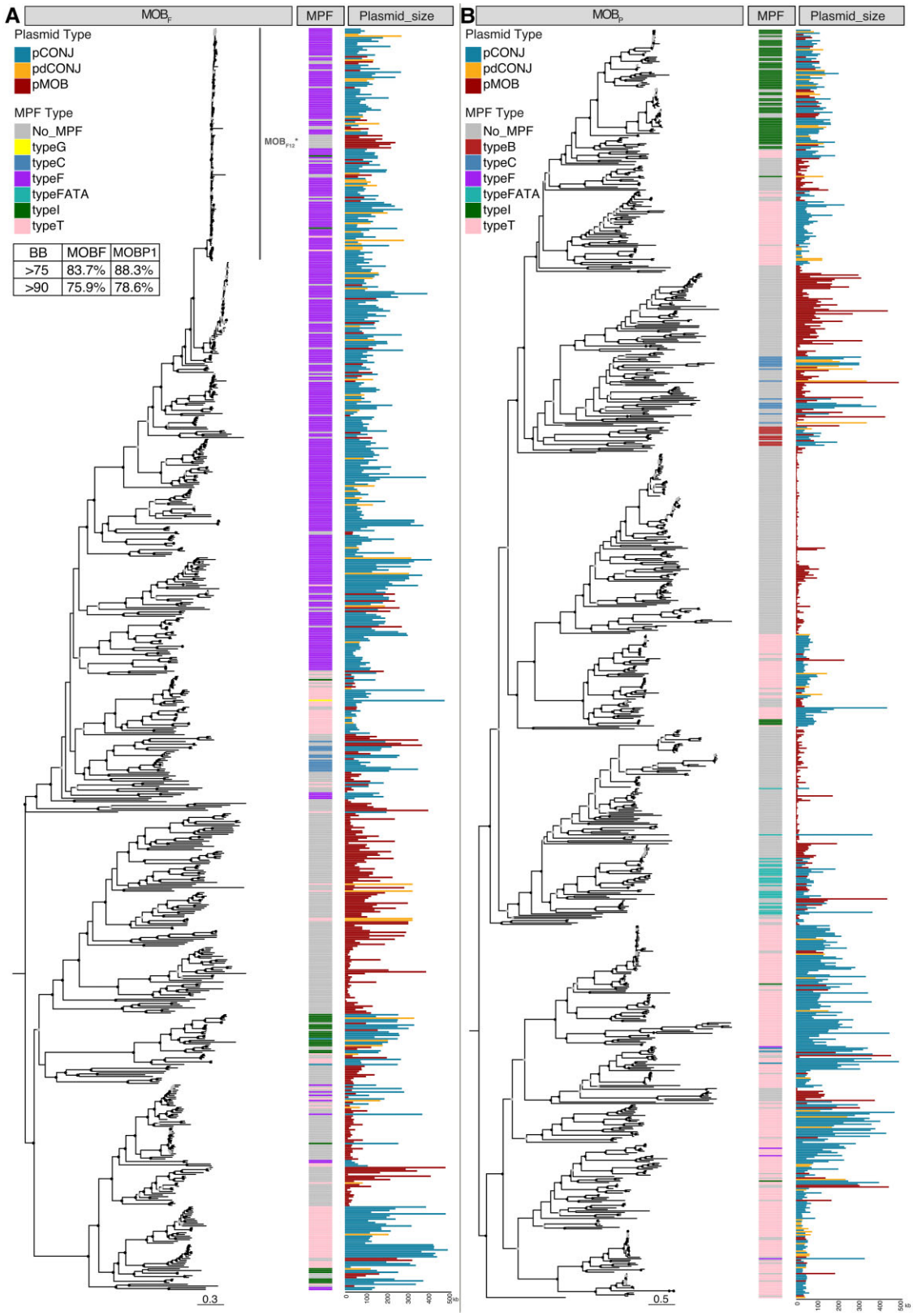


FIG. 5. Phylogenetic trees of MOB_F relaxases and MOB_{P1} relaxases. Ultrafast bootstrap values superior to 75 are shown with a light gray circle and values superior to 90 with a black circle. Percentage of Ultrafast bootstrap values superior to 75 and 90 for each the tree are reported in the table. The trees were rooted using the midpoint root. (*A*) The phylogenetic tree was built using 735 MOB_F proteins with maximum likelihood with IQTree (model WAG + F + R10 and 1000 ultrafast bootstraps, Nguyen et al. 2015). *: clade containing MOB_{F12} plasmids used in fig. 7. (*B*) The phylogenetic tree was built using 850 MOB_P proteins (all retrieved with the MOBP1 HMM profile) with maximum likelihood with IQTree (model VT + F + R10 and 1000 ultrafast bootstraps).

Downloaded from https://academic.oup.com/mbe/article/39/6/msac115/6593704 by Institut Pasteur - CeRIS user on 18 October 2022

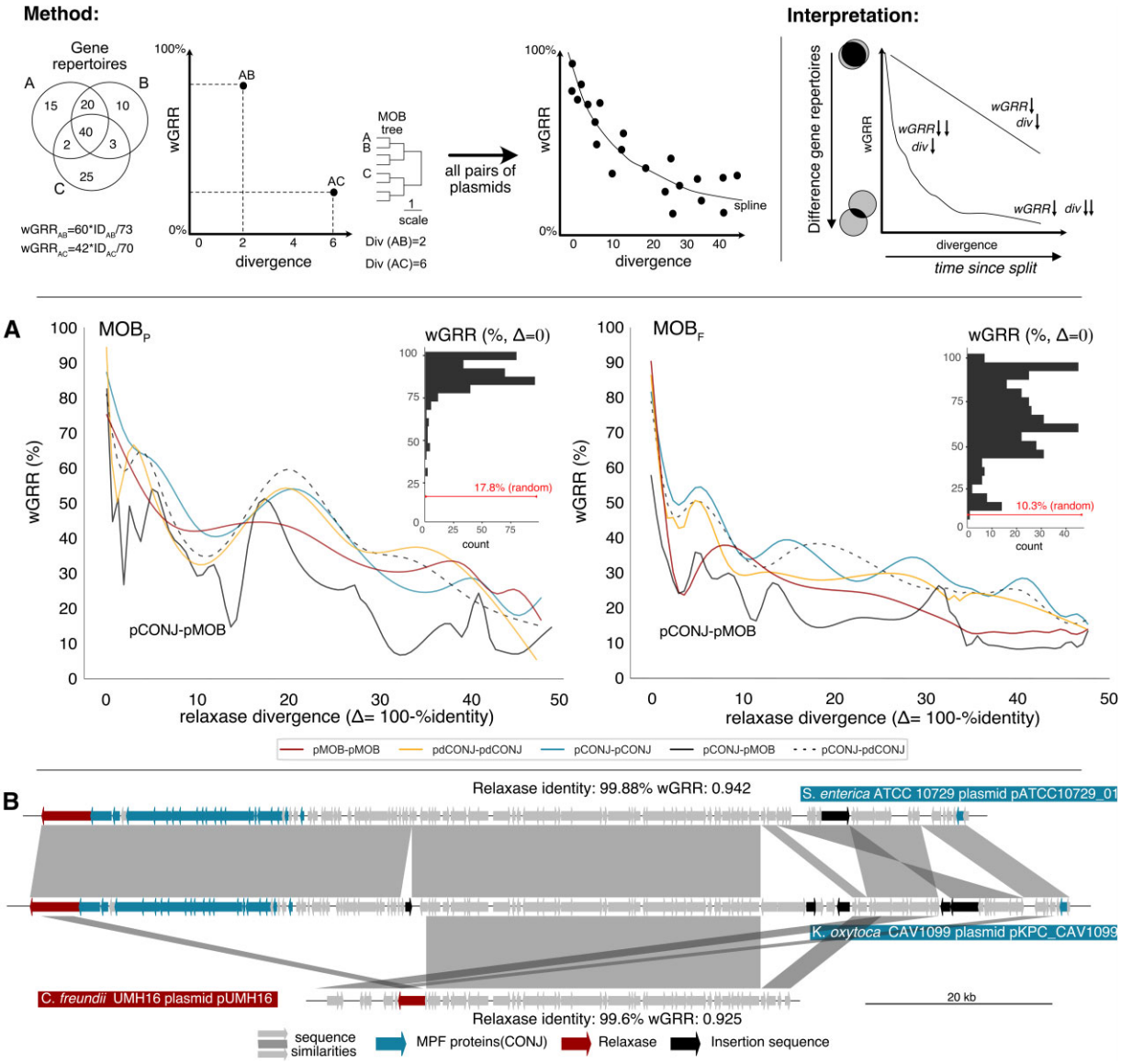


FIG. 6. (*A*) Relationship between relaxase protein sequence divergence and wGRR in pairs of plasmids classed in terms of mobility for MOB_P (left) and MOB_F (right). Each curve represents the smoothing spline for comparisons between: pdCONJ (yellow), conjugative (blue), mobilizable (red), conjugative and pdCONJ (dotted black), and comparisons between mobilizable and conjugative plasmids (black). (*B*) Comparison between selected conjugative and mobilizable plasmids. Plotted using GenoPlotR v.0.8.10 (Guy et al. 2010), based on Blastn analysis of the plasmid nucleotide sequences (*e*-value $< 10^{-4}$). Comparison between two pCONJs (blue) and a pMOB (red). For clarity, we only represent regions of homology larger than 1 kb with more than 50% identity. Genes involved in the plasmid mobility (T4SS and relaxase[s]) and insertion sequences are highlighted.

The wGRR is the number of bi-directional best hits between two plasmids weighted by their sequence similarity and divided by the number of genes of the smaller element (see Materials and Methods). A wGRR close to 1 means that plasmids are very similar (or one is a subset of the other) and a wGRR of 0 indicates the absence of homologous genes. Only around 20,278 out of the 2,616,636 pairs of plasmids show a wGRR higher than 0.9 (<0.77%), and 79% of the pairs of plasmids have a wGRR lower than 0.2 (supplementary fig. S3, Supplementary Material online). We searched for pairs of plasmids with high wGRR (>0.9) and with different classes of relaxases. We found only 22 cases, of which none implicated a pair of MOB_P–MOB_F plasmids. Hence, plasmids with similar gene repertoires tend to have relaxases of the same type.

We then analyzed in detail the 27,225 pairs of plasmids with relaxases 100% identical in nucleotide sequence. This corresponds to the most recent transitions in terms of mobility and most of these pairs of plasmids have similar gene repertoires (median wGRR = 0.928). More precisely, pairs of plasmids of the same type of mobility show median values of wGRR of 0.925 (pCONJ), 0.972 (pdCONJ), and 0.987 (pMOB). Lower values of wGRR are found in pairs of plasmids pCONJ–pdCONJ (0.875) and pCONJ–pMOB (0.822). However, even the latter values are very high, considering that the average wGRR values between random plasmids

Downloaded from https://academic.oup.com/mbe/article/39/6/msac115/6593704 by Institut Pasteur - CeRIS user on 18 October 2022

containing the relaxases of the same class are below 0.2 (fig. 6). Plasmids with very similar relaxases also tend to be part of the same PTUs (only 18.2% of the 307 clusters of nearly identical relaxases have more than one PTU, supplementary fig. S4, Supplementary Material online). In conclusion, plasmids with identical relaxases tend to be very similar, even when they have different type of mobility. This suggests that most changes in mobility take place by structural changes in the plasmid, not by translocation of the relaxases to an unrelated plasmid.

Comparisons between plasmids with very similar relaxases, but not identical (between 1% and 5% protein sequence divergence), reveal rapidly decreasing wGRR with increasing divergence of the relaxases (fig. 6). The average trends are similar for pMOBs and pCONJs, suggesting comparable rates of change in gene repertoires. In contrast, there is a steepest loss of wGRR for pairs pCONJ–pMOB, which rapidly fall to values of wGRR around 0.3 suggesting that the change in mobility is followed by accelerated differentiation of gene repertoires. Interestingly, pairs pCONJ–pdCONJ show a lower decrease in wGRR with relaxase divergence than pairs pCONJ–pMOB. This suggests that transitions pMOB–pCONJ are associated with larger changes in gene repertoires.

The analysis of pairs of plasmids with divergent relaxases (<95% protein sequence identity) reveals much lower values of wGRRs, often lower than 50% (MOBP1) or 40% (MOBF). This means that such plasmids are poorly related or entirely unrelated, even when they share the same type of mobility. At 50% identity between the relaxases, the average wGRR is <30% for every type of mobility (even though the plasmids have necessarily the relaxase as homolog). Such distant comparisons show little variation in wGRR in relation to the divergence of the relaxases, suggesting that at such long evolutionary distances there is little correlation between the similarity in genome repertoires and the similarity in the sequences of the homologous relaxases. Hence, to understand changes in types of mobility, one must focus at short evolutionary distances.

Variation in Plasmid Taxonomy with Changes in Mobility

To understand how the difference in mobility impacts the taxonomic classification of the plasmids, we focused on the plasmids from the MOB_{F12} family from *Escherichia coli* and *Shigella spp*. (fig. 5A). We chose this family because it includes many closely related well-characterized plasmids (Fernandez-Lopez et al. 2016). We analyzed these plasmids using pairwise ANI comparisons. This resulted in a graph that we colored with the information on plasmid mobility or their PTUs (fig. 7A). The plasmids that have a different mobility type tend to be less connected and are thus placed at the outer edge of the graph component with the main PTU (PTU-F_E). Alternatively, some are classified as different PTUs. Hence, plasmids that change in terms of gene mobility tend to bud out of the original PTU graph and sometimes lead to novel PTUs (such as E5, F_{Sh}, E41, etc.). In this specific example, the emerging novel PTUs seem to be associated with specific strains (e.g., PTU-F_{Sh} to *Shigella* spp., and PTU-E5 to *E. coli* O157:H7), which may be an indication that novel PTUs emerge by change of mobility and adaptation to strains with specific ecology.

The largest component of the MOB_{F12} family ANI graph corresponds to the PTU-F_E plasmids. In this PTU most plasmids are conjugative (68%) and large (average 112 kb), implicating that the most parsimonious hypothesis is that changes in mobility correspond to loss of the ability to conjugate. Accordingly, the pMOBs and pdCONJs in this PTU are much larger than usual in mobilizable plasmids (see fig. 2 for the general trends): 117.4 kb (pCONJ), 118 kb (pdCONJ), 112 kb (pMOB), and 95.3 kb (pMOBless). We used AcCNET to analyze the gene repertoires shared by the members of this PTU (fig. 7B). The pMOBs and pdCONJs that presumably transited from pCONJs in this PTU, are scattered across the network. These results are consistent with those at larger evolutionary scales, where recent transitions to pMOB and pdCONJ were dispersed in the phylogenetic trees of the relaxases (fig. 5). In both cases, a group with a preponderance of pCONJs includes a few pMOBs, pdCONJs, and pMOBless that are scattered far apart, suggesting frequent independent transitions. Nevertheless, these plasmids that have recently changed in terms of type of mobility have many properties of the original pCONJ (e.g., the PTU and the plasmid size).

Conjugative Plasmids are a Frequent Source of Other Mobile Plasmids

To quantify the rates of transition between mobility types, we inferred the ancestral type of mobility of plasmids using PastML (see schema in fig. 8). We then focused on changes of plasmid mobility type identified in the terminal branches of the tree of relaxases (fig. 5, supplementary figs. S5–S11, Supplementary Material online). The focus on terminal branches is due to both a technical issue, the inference of ancestral states is more accurate in these locations, and a biological reason: the previous results suggest that it is better to focus on recent events of change before phylogenetic information becomes a poor predictor of the similarity in gene repertoires (fig. 6). We identified 249 terminal branches associated with a change in the type of mobility. We then computed the rates of transition of type A to type B as the observed number of transitions A to B divided by the number of plasmids of type A (Rocha et al. 2006). The most frequent transitions (116) are from pCONJ to pMOB, corresponding to a rate of 4.2% and pCONJ to pdCONJ (77, rate = 2.8%) (fig. 8A). The transition from pdCONJ to pMOB (1.4%) is more

Downloaded from https://academic.oup.com/mbe/article/39/6/msac115/6593704 by Institut Pasteur - CeRIS user on 18 October 2022

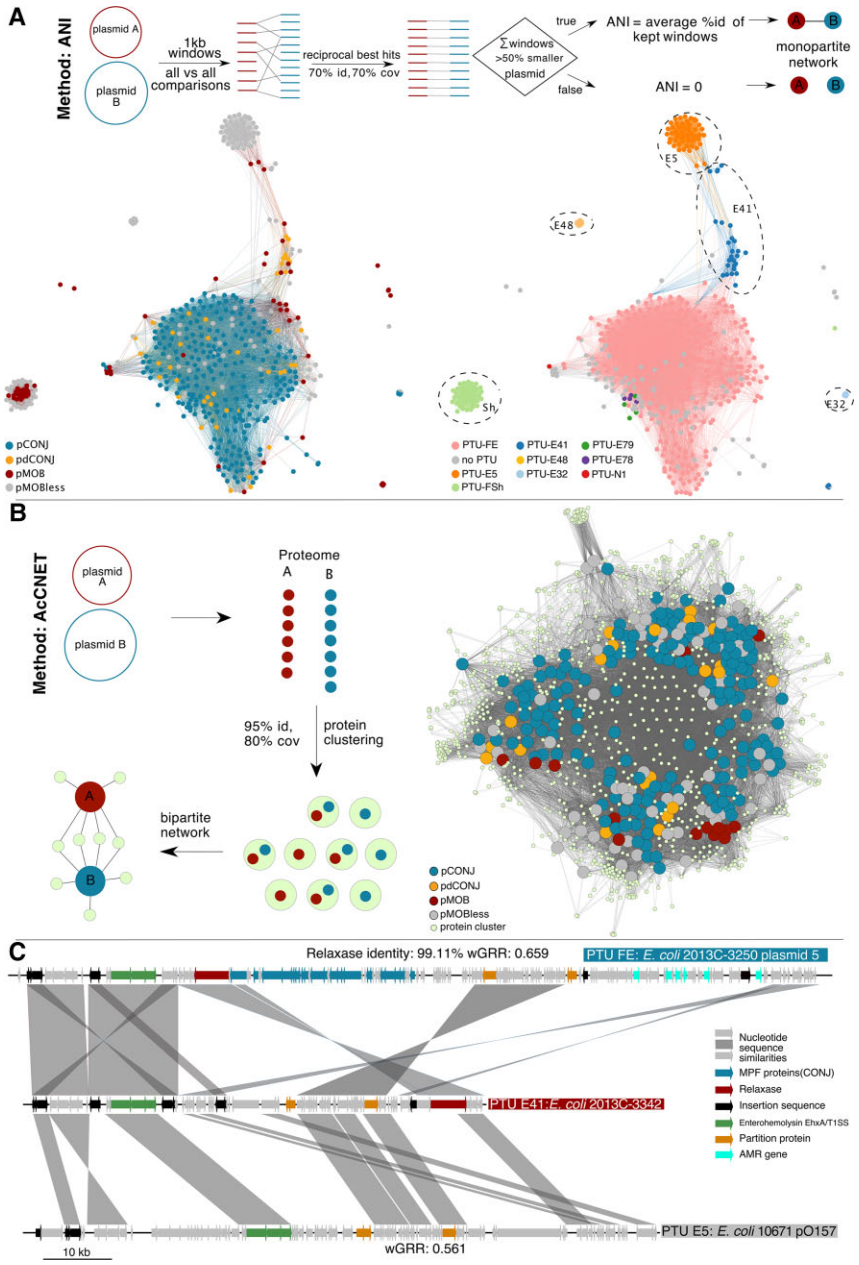


FIG. 7. (*A*) Graphical representation of the algorithm used for the ANI calculation: each plasmid nucleotide sequence was divided into overlapped 1 kb windows and for each plasmid pair, all stretches were compared. ANI scores were obtained by averaging the percentage of identity of all considered windows with identity and coverage >70%, whenever the sum of windows covered at least the 50% of the smallest plasmid in the pair. ANI scores different from 0 were represented as edges connecting the members of the plasmid pair (nodes) in the ANI network. ANI similarity network of the MOB_{F12} plasmid family from *E. coli* and *Shigella* spp: The genomic relatedness of 353 transmissible MOB_{F12} plasmids hosted in *E. coli* and *Shigella* spp., and 140 MOBless plasmids belonging to the same PTUs was estimated by pairwise ANI calculations. At the left, nodes are colored by their mobility type, and at the right, by their PTU. "no PTU" means nonassigned PTU. In the right panel, PTUs that are separated from the main group (PTU-F_E) are surrounded by a circle. (*B*) Graphical representation of the method used for the plasmid proteome network analysis: the protein set of all plasmids to be compared is clustered at 95% amino acid identity and 80% alignment coverage. The homologous protein clusters and their corresponding plasmids are the two kinds of nodes represented in the network and edges connect both types whenever a plasmid contains a member in a given protein cluster. Proteome network of the PTU-F_E plasmids. The proteins of 250 PTU-F_E plasmids were clustered at 95% identity and 80% coverage. Whenever a plasmid has a member in a protein cluster, an edge is linking them. Homologous protein clusters present in a single plasmid were removed from the figure. Plasmids are colored by their mobility type. (*C*) Comparison between selected PTU-F_E, PTU-E_{41}, and PTU-E_5 plasmids. Plotted using GenoPlotR v.0.8.10 (Guy et al. 2010), based on Blastn analysis of the plasmid nucleotide sequences (*e*-value $< 10^{-4}$ and alignment length > 1000 bp). Comparison between a pCONJ (blue), a pMOB (red), and a pMOBless plasmid (gray). For clarity, we only represent best bidirectional regions of homology larger than 1 kb with more than 50% identity. Genes involved in the plasmid mobility (T4SS and relaxase[s]), antimicrobial resistance (AMR), plasmid partitioning, pathogenicity, and insertion sequences are highlighted.

Downloaded from https://academic.oup.com/mbe/article/39/6/msac115/6593704 by Institut Pasteur - CeRIS user on 18 October 2022

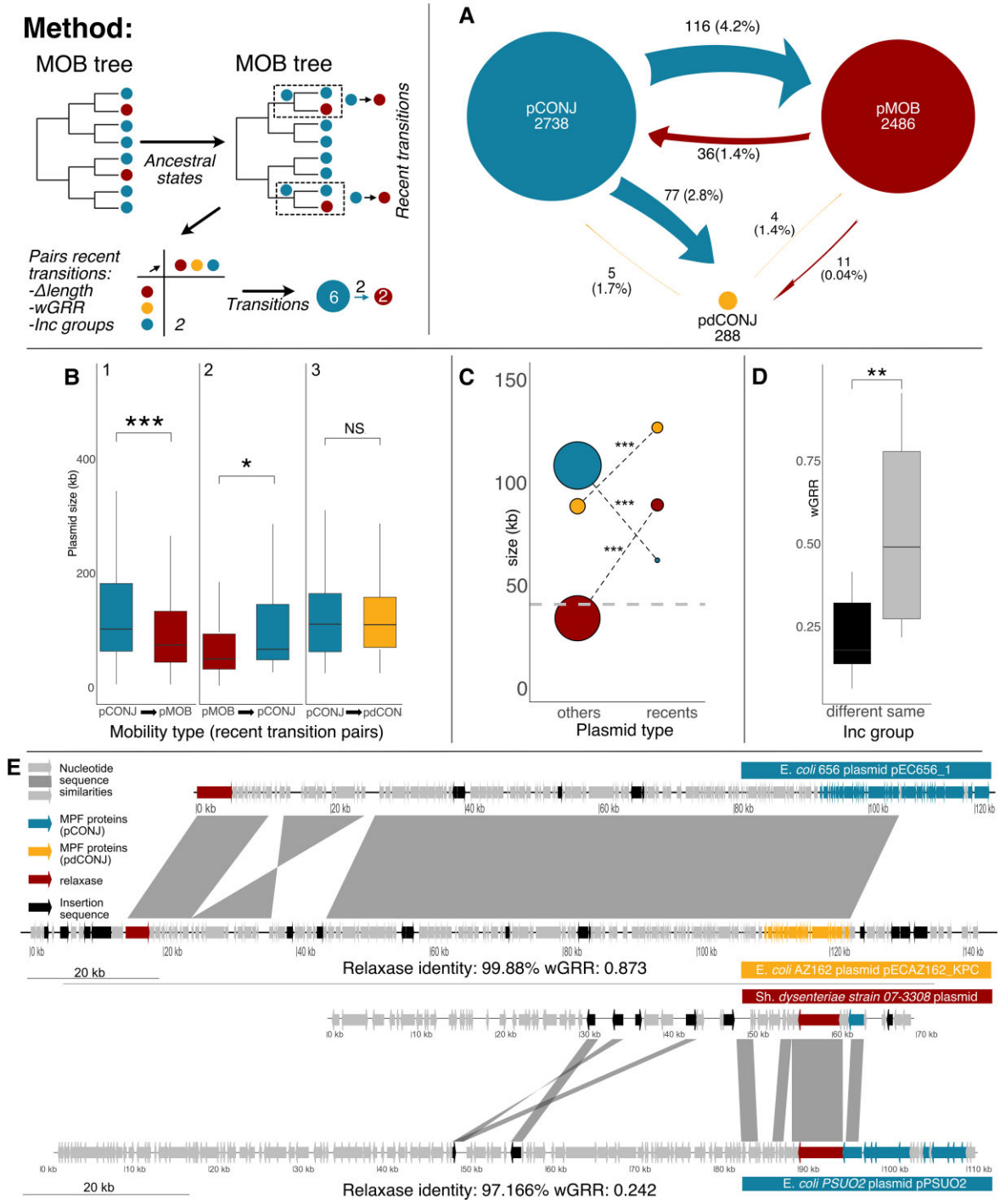


FIG. 8. Characterization of mobility transitions. Graphical representation of the method: starting from the phylogenetic tree of the relaxases and the reconstruction of the ancestral states, we inferred the direction of the changes in terms of mobility and paired recently transited plasmids with the closest plasmid of another mobility. (*A*) Transitions inferred in terminal branches of the relaxase phylogenetic trees. Arrows and numbers represent the direction and number of transitions, respectively. The size of circle indicates the abundance of each plasmids type. (*B*) Boxplots representing the size of plasmids that recently changed in terms of mobility (in comparison to the sister-taxa plasmid in the relaxase tree having the ancestral state). Boxplots 1, 2, and 3 represent the size distribution associated with transitions from pCONJ to pMOB, from pMOB to pCONJ, and from pCONJ to pdCONJ respectively. (*C*) Bubble plot of the median size between plasmid that recently transited mobility and others. The red bubbles represent pMOB, the blue bubbles pCONJ, and the yellow bubbles pdCONJ. The gray dotted line represents the median size of pMOBless. (*D*) Boxplot of the wGRR between pairs of plasmids where one recently changed from pCONJ to pMOB. The gray (black) boxplot represents the wGRR for plasmids of the same (different) replicon type. (*E*) Comparison between selected plasmids, plotted using GenoPlotR v.0.8.10 (Guy et al. 2010), based on Blastn analysis of the plasmid nucleotide sequences (*e*-value $< 10^{-4}$ and alignment length > 1000 bp). Top: Comparison between a pdCONJ (yellow) and a conjugative plasmid (blue). Bottom: Comparison between a pMOB (red) and a pCONJ (blue). For clarity, we only represent regions of homology larger than 1 kb with more than 50% identity. Genes involved in the plasmid mobility (T4SS and relaxase[s]) are highlighted. Tests of differences (Wilcoxon paired-tests): $^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$, NS (nonsignificant).

Downloaded from https://academic.oup.com/mbe/article/39/6/msac115/6593704 by Institut Pasteur - CeRIS user on 18 October 2022

frequent than the inverse (0.04%), suggesting that pdCONJ are likely to be further degraded and result in a plasmid with few or no MPF gene. In contrast, we find much lower rates of transitions from pMOB to pCONJ (1.4%) and from pdCONJ to pCONJ (1.7%), confirming that pCONJs are more often the source of other plasmids than vice-versa.

The analysis of the wGRR values for the comparisons between pCONJs and pMOBs suggested that transitions in type of mobility are associated with large changes in plasmid gene repertoires (fig. 6). However, our previous analysis could not distinguish the direction of the changes in plasmid mobility. So, we took advantage of the identification of the 249 transitions by inference of ancestral states to study the characteristics of pairs of closely related plasmids with different mobility types. We observe that transitions pCONJ to pMOB are associated with a decrease in plasmid size (fig. 8*B.1*). The second most frequent transitions are those from pCONJ to pdCONJ and they are associated with much smaller changes in plasmid size since in this case most of the MPF genes are still present in the plasmid (fig. 8*B.3*). As expected, the few transitions pMOB to pCONJ involve an increase in plasmid size (fig. 8*B.2*). We then compared the plasmids that recently changed in terms of type of mobility with the other plasmids of the same mobility type. Notably, we compared the average size of the recent pMOBs with those of the remaining pMOBs. This analysis clearly shows that the plasmids that recently became pMOBs or pdCONJs have much larger sizes than the average plasmids of the same type (fig. 8*C*). For pMOBs, even though those that recently transited from pCONJs are 23% smaller than the most closely related pCONJs, they are still 165% larger than the other pMOBs (fig. 8*C*). If these recent pMOBs derive by deletion of ancestral pCONJs, then one might expect to find in their sequences a few MPF genes. Indeed, the recent pMOBs have 44% more MPF genes than the remaining ones ($P < 0.01$, Wilcoxon test, supplementary fig. S12, Supplementary Material online). This shows that transitions from pCONJ to pMOB are associated with gene deletions. Yet, during the initial stages of this process of genome reduction, the pMOB still resemble pCONJ in many respects.

A pCONJ that lost the ability to conjugate becomes dependent on the presence in the cell of a pCONJ with a compatible MPF. The most closely related pCONJ are likely to be the most compatible with the relaxase of this pMOB. However, these very closely related pCONJs are also likely to encode similar replication initiation proteins and thus be incompatible (same Inc type), that is, they cannot be stabilized in the same cell (Novick 1987). If the transfer of the novel pMOB depends on the presence of conjugative plasmids whose replication initiators are incompatible with its own, then its viability is at risk. To understand how incompatibility evolves upon change in plasmid mobility, we typed the replicons and analyzed if those 249 that recently changed in terms of mobility also changed in terms of incompatibility group. Interestingly, this is often the case among the few such plasmids that could be typed: 17 out of 25 of the novel pMOBs have an Inc type different from the one of the closest related pCONJ. Expectedly, pMOBs that have different replication initiation proteins have a much smaller wGRR relative to the closest pCONJ than those that have similar ones (fig. 8*D*). Other processes in plasmid biology may affect the stability of plasmids in cells, including plasmid partition and postsegregation killing (e.g., toxin–antitoxin systems) (Bouet et al. 2007; Diaz-Orejas et al. 2017). Unfortunately, there are no established methods to identify classes of incompatibility between such systems, albeit works suggest that they have a weaker effect on incompatibility than replication (Nordstrom et al. 1980). We identified a partition system in 242 out of the 249 plasmids and found that 107 (42.9%) encoded a partition system different from the one of the closest related plasmid. Most of the 249 plasmids (182) encoded at least one toxin of a toxin–antitoxin system, which often (143) lacked a homolog in the closest related plasmid of a different mobility type. Hence, modifications in the mechanisms associated with plasmid stability may be part of the broader changes resulting in rapid decrease of wGRR values after transition from pCONJ to pMOB. Such changes may facilitate the coexistence of the novel mobilizable plasmid with the closest related pCONJ, which is the one that encodes the most compatible MPF system for the pMOB.

In this study, we did not analyze ICEs or IMEs because there is no sufficiently accurate method to delimit them from chromosomes in a large scale. This should not affect the phylogenetic analyses of relaxases nor the pairwise comparisons between plasmids mentioned above. However, if some of these integrative elements were very similar to plasmids, their analysis could shed light on recent changes of plasmid mobility. To assess this possibility, we searched for large chromosomal regions (>5 kb) with high sequence similarity (>90% identity) to the 249 plasmids that recently changed in terms of mobility. Most (63.6%) of the plasmids lacked such chromosomal homologs. For the others, these regions covered <10% of the plasmids in 83% of the cases and only one plasmid matched the chromosome along more than 70% of its length (supplementary fig. S13, Supplementary Material online). Similar qualitative results were found when analyzing smaller regions of homology (>2 kb, supplementary fig. S14, Supplementary Material online). Hence, disregarding the chromosome sequences does not seem to remove a lot of information when the goal is to analyze recent events of change in plasmid mobility.

Persistence of Mobility Type

The previous results raise the question of the persistence of the different types of plasmids, that is, of how much time the novel plasmids persist in populations, eventually become fixed, and then diversify. The precise quantification of the persistence of a given of mobility is difficult because plasmids evolve very fast and the inference of ancestral states deep in the tree is unreliable (this is why

Downloaded from https://academic.oup.com/mbe/article/39/6/msac115/6593704 by Institut Pasteur - CeRIS user on 18 October 2022

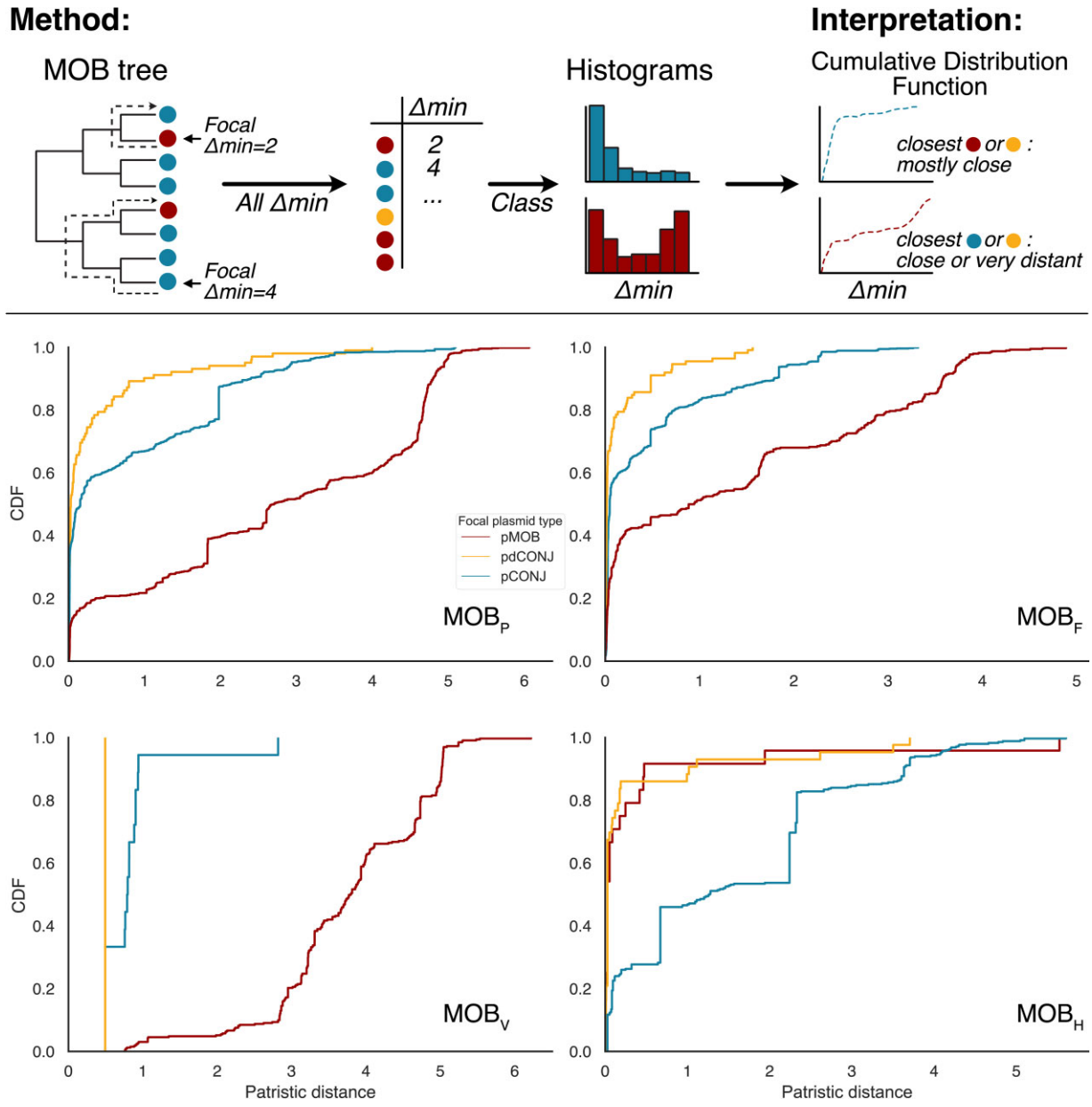


Fig. 9. CDF of the minimal patristic distances in the relaxase phylogenetic trees from a relaxase to another relaxase of a different mobility type. The upper panel depicts the method used to compute the CDF. The patristic distance between each relaxase and its closest closest homolog associated with a different mobility type was retrieved for all trees. When the CDF approaches 1 for low patristic distances, this means that all relaxases are close to a relaxase of a plasmid with a different mobility type in the tree. Here, are indicated the CDF of relaxases identified with protein profiles for MOBP1, MOBF, MOBV, and MOBH.

above we focused on terminal branches). Instead, we computed the shortest patristic distances between each plasmid and the closest plasmid of a different kind of mobility as a proxy of persistence. For the very similar relaxases clustered in figure 5, we used the patristic distance separating the representative of the cluster and the closest relaxase in the tree having a different mobility type (if this protein is in the same cluster, then the patristic distance is 0). This gives an indication of how far one must go back in time to find a plasmid that has a different mobility type (see schema in fig. 9). We then computed the CDF of these patristic distances for each type of mobility (fig. 9). A rapid initial increase in the CDF with increasing patristic distance indicates that the state is not very persistent, that is, there are plasmids with different types of mobility at small patristic distances. In contrast, a slow initial increase in the CDF means that the closest plasmid with a different mobility type is usually distant in the tree. The results for MOB_P and MOB_F are qualitatively similar. Conjugative plasmids have intermediate CDFs, which is consistent with the observation that they are frequently the source of pdCONJs and pMOBs. The pdCONJ have a steeper CDF indicative that they are always close to other types of plasmids in the tree of relaxases. This is consistent with frequent

Downloaded from https://academic.oup.com/mbe/article/39/6/msac115/6593704 by Institut Pasteur - CeRIS user on 18 October 2022

transitions pCONJ to pdCONJ and with the little persistence of the latter (either because they change to pMOBs or because they are lost from populations). Hence, these results confirm, and quantify, the trends observed in figure 5.

The CDF varies with patristic distances in a more complex way for pMOBs. We observe an initial steep increase for low patristic distances, like for pCONJs and pdCONJs, followed by a region of slow increase, and a final steep increase (especially visible among plasmids encoding a MOB_P relaxase). This is consistent with the observations made on the phylogenetic trees of the relaxases. We had observed there that many pMOBs recently transited from pCONJs. These correspond to the initial increase in the CDF, that is, they are caused by the existence of many pMOBs with closely related pCONJs or pdCONJs. The remaining pMOBs are placed in clades of the relaxase phylogenetic tree that have almost only pMOBs. For such pMOBs, the closest related pCONJ or pdCONJ may be at very high patristic distances.

MOB_P and MOB_F are present in many conjugative and mobilizable plasmids. To understand if the persistence a type of mobility differs from the above in plasmids with classes of relaxases associated with either one or the other, but not both, types of plasmid mobility, we analyzed the CDF for MOB_V (mostly pMOBs) and MOB_H (mostly pCONJs). The results for MOB_V mirror those of MOB_P and MOB_F, with the exception that the CDF increases faster for conjugative plasmids and slower for mobilizable (fig. 9). The low persistence of the few pCONJ in the MOB_V analysis is caused by the fact they are all recent, which suggests that such plasmids rapidly disappear from populations. In contrast, the CDF increases much slower for conjugative elements with MOB_H than for those with MOB_P and MOB_F, suggesting that the former rarely result in persistent pMOBs. In contrast, the high persistence of a subset of pMOB relaxases in MOB_P and MOB_F and across MOB_V is consistent with a specialization of certain relaxases in mobilizable plasmids.

Transitions to and from pMOBless are Frequent

In the precedent sections, we focused on conjugative and mobilizable plasmids because they can be analyzed in terms of the presence and evolution of the relaxase. Yet, around half of the plasmids lack a recognizable relaxase (pMOBless, fig. 2A). If plasmids can change between conjugative and mobilizable types, they may certainly also gain or lose the relaxase. The study of the transitions to and from pMOBless is difficult because they lack conserved phylogenetic markers. The only alternative to the relaxase, the replication genes, are very diverse, often unrecognizable, and they can be protein or RNA genes, making deep evolutionary studies in a phylogenetic framework impossible. To study the transition to and from pMOBless, we identified nonredundant pairs of pMOBless-mobile plasmids (pCONJ, pMOB, pdCONJ) that presumably diverged recently because their wGRR is high (superior to 0.75, see details in Materials and Methods and schema in fig. 10). We detected 345 mobile plasmids that had one closely related pMOBless (fig. 10A): 154 pCONJs, 172 pMOBs, and 19 pdCONJs. One might have expected an overrepresentation of pairs pMOBless/pMOBs in this data, since both types of plasmids tend to be small (fig. 2), and it only takes the gain/loss of a relaxase to transition from one into the other. Yet, this is not the case and there is no significative difference between the number of pCONJ/pMOBless pairs (44.5%) and pMOB/pMOBless pairs (50%) ($\chi^2 = 2.8209$, $df = 1$, $P = 0.09$). Hence, transitions to or from pMOBless may be common for all other types of plasmids.

To assess the consequences of these transitions, we compared the sizes of the plasmids of each pair. When the analysis was stratified by plasmid mobility type, it revealed that pMOBless tend to be significantly smaller than the other element of the pair independently of the type of mobility of the latter (fig. 10). Unfortunately, we lack a phylogenetic marker that allows to identify the directions of change, but these results suggest that transitions to pMOBless are associated with gene losses and transitions from pMOBless are associated with gene acquisition (translocations) or cointegrations. The relative rates of each type of event will require substantial further work.

Discussion

In this study we aimed at understanding the evolution of plasmid mobility. We focused on plasmids encoding relaxases because these are expected to be able to transfer horizontally. Furthermore, the relaxase is a well-conserved protein that allows the study of plasmid evolution using a rigorous phylogenetic framework, including the inference of rates of change (and their direction) in the type of mobility. Conjugative systems, and thus relaxases, appeared early in the history of life, as discussed in Guglielmini et al. (2013). Since plasmids with a relaxase can only be mobilizable if there is a conjugative system available, one may presume that conjugative plasmids arrived first and mobilizable plasmids arrived later. Nevertheless, relaxases may have predated the emergence of conjugative systems, for example, functioning as plasmid replication initiators, and been coopted later for existing conjugative systems. This would contribute to explain why relaxases are so diverse, when the key components of the MPF are homologous (Guglielmini et al. 2014). Since a gene deletion is sufficient for a conjugative plasmid to become mobilizable, pMOBs may have arisen very quickly after pCONJ. This would explain the existence of large ancient clades of relaxases specific to mobilizable plasmids in the MOB_P and MOB_F trees (fig. 5). Importantly, relaxase classes lacking homology with MOB_P are all associated with conjugative elements. The specialization of relaxases could also explain the large separate clades dominated by either pMOB or pCONJ in the phylogenetic trees of MOB_P and MOB_F relaxases (fig.

Downloaded from https://academic.oup.com/mbe/article/39/6/msac115/6593704 by Institut Pasteur - CeRIS user on 18 October 2022

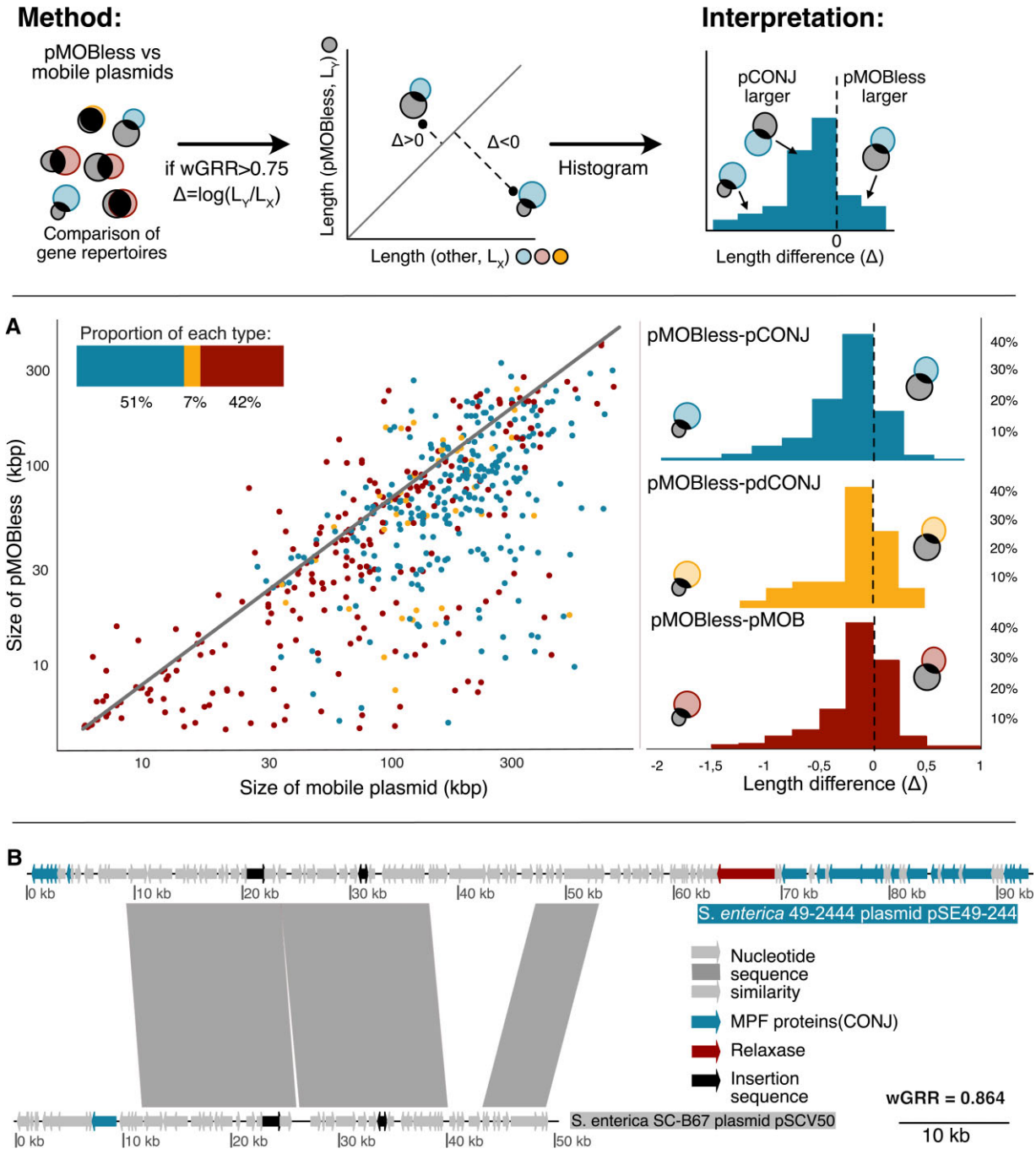


FIG. 10. Relationship between the size of mobilizable or conjugative plasmid and the closest related pMOBless. Graphical representation of the method: mobile plasmids were paired with the pMOBless with the highest wGRR. Then the difference in size of both plasmids was calculated and compared with linear model. (*A*) Each data point in the scatter plot represents the size of pairs of plasmids that are very similar (wGRR $>$ 0.75) and one is pMOBless whereas the other is not. Blue, red, and yellow dots represent pairs between MOBless and pCONJ, pMOB, and pdCONJ, respectively. The histograms on the right represent the distribution of the length difference compared with a linear model (distance to the identity line in gray). All three distributions have an average significantly lower than zero (all Wilcoxon tests, $P < 0.0001$). (*B*) Comparison between selected plasmids, plotted using GenoPlotR v.0.8.10, based on Blastn analysis of the plasmid nucleotide sequences (*e*-value $< 10^{-4}$ and alignment length $>$ 1000 bp). Comparison between a pCONJ (blue) and a MOBless plasmid (gray). For clarity, we only represent regions of homology larger than 1 kb with more than 50% identity. Genes involved in the plasmid mobility (T4SS and relaxase[s]) and insertion sequences are highlighted. The pMOBless pSCV50 also encode a relaxase pseudogene not highlighted.

4). This suggests that some relaxases evolved specific traits to become specialized in mobilizable plasmids. The alternative hypothesis, that relaxases were separately coopted to be relaxases in conjugative and mobilizable plasmids and were never transferred into another type of plasmid seems extremely unlikely given the fast rate of change of plasmid gene repertoires, the transitions between mobility types in some clades of the trees, and the high gene flow between plasmids (Eberhard 1990; Revilla et al. 2008; Redondo-Salvo et al. 2020).

Downloaded from https://academic.oup.com/mbe/article/39/6/msac115/6593704 by Institut Pasteur - CeRIS user on 18 October 2022

Why would there be a specialization of relaxases? Tight coevolution between the T4SS and the cognate relaxase of conjugative plasmids is expected because the two components must interact efficiently and they are genetically linked. This may result in very specific interactions. For example, the pCONJ R751 and RP4 have very similar origins of transfer, but are unable to mobilize each other (Fürste et al. 1989). In contrast, mobilizable plasmids require a conjugative element to transfer and may have evolved to interact with very different conjugative systems. Accordingly, MOB_{Q4} family plasmids can be transferred by at least the MPF_I and MPF_T systems (Garcillan-Barcia et al. 2019), RSF1010 is mobilized by MPF_I, MPF_F, and other uncharacterized systems (Meyer 2009), MOB_{P5} mobilizable plasmids (ColE1) are mobilized by MPF_F and MPF_I plasmids (Cabezon et al. 1997), MOB_{V1} mobilizable plasmids, such as pMV158, are mobilized by MPF_T and MPF_{FATA} (Lorenzo-Diaz et al. 2014), and SGI1 genomic islands from *Salmonella enterica* can be mobilized by IncA and IncC plasmids (Szabó et al. 2021). A recent study showed that IncQ plasmids, which are pMOBs, have exceptionally large host range (Stalder et al. 2019). Finally, whereas the pCONJ R388 and RP4 cannot mobilize each other, both can mobilize the unrelated pMOBs RSF1010 and ColE1 (Cabezón et al. 1994). The ability of pMOB relaxases to interact with multiple T4SS may implicate less efficient interactions with each specific T4SS (Sastre et al. 1998). Hence, relaxases of mobilizable plasmids would evolve to interact with many different T4SS at the cost of interacting less efficiently with any single one.

Our analyses of the rates of recent transitions in terms of mobility show a clear pattern of more frequent transitions from pCONJ to pdCONJ or to pMOB than the opposite. Furthermore, the rates of transition of pdCONJ to pMOB are higher than the inverse, even if in this case the much larger number of pMOBs evens out the exchanges in terms of the number of events. If this dynamic is constant in time and there is no moderation of these events by natural selection, then one would expect that pMOBs would have become much more abundant than pCONJ. The fact that there are so many pCONJ suggests that many of the transitions from pCONJ are eventually lost or compensated by relative higher propagation of conjugative plasmids. This fits the phylogenetic studies, since clades dominated by conjugative plasmids have pMOBs systematically at or close to terminal branches, that is, they are recent. Such dynamics is usually classed under source-sink models (Sokurenko et al. 2006), where one population, here pCONJ, constantly provides novel individuals to another, here the pMOB. If there is evolution in the sink, and such conditions have been proposed to increase genetic plasticity (Chevin and Lande 2011), then the sink population may occasionally be salvaged. Here, the source-sink model does not concern migration, but genetic modifications resulting in the transition of mobility types. In the case of plasmids, if the novel pMOB is persistent enough, then it may adapt to the novel conditions and the "sink" population becomes a viable, differentiated group of mobilizable plasmids. This fits the data showing the existence of some large clades of pMOB relaxases. Hence, whereas many transitions pCONJ to pMOB seem to result in plasmid lineages that go extinct, we propose that sometimes such transitions led to successful creation of novel lineages of relaxases. We presume that one possible advantage of these plasmids is their ability to interact with multiple MPF, as discussed above. Adaptation resulting in the salvage of the pMOB may also occur if the latter is less costly. This might explain why the pMOBs that emerged a long time ago tend to be much smaller than the most recent ones. Similar processes may explain how plasmids with relaxases become pMOBless.

The quantification of similarities of gene repertoires showed that they change very fast in plasmids, relative to the evolution of relaxases. This observation fits numerous previous reports of rapid evolution in epidemic plasmids (Porse et al. 2016; Peter et al. 2020; Reid et al. 2022). The pace of change is further accelerated when there is a change in the mobility type, which suggests a link between the genetic mechanisms driving changes in plasmid mobility and those changing plasmid gene repertoires. Transitions between types of mobility can be due to accretion/deletion of genetic material or translocation/cointegration events (fig. 11). It should be noted that given the rapid pace of gene repertoire evolution, it is difficult to distinguish large translocations from plasmid cointegrations. Still, several arguments suggest that most transitions of pCONJ to other types of plasmids tend to begin with deletions of genetic material. (1) The relaxase gene is not mobile in itself, thus requiring other genetic elements, like transposable elements, to transfer between plasmids. Importantly, the homology between closely related plasmids of different mobility type usually extends beyond the relaxase. (2) The analyses of pairs of plasmids with different mobility types but identical relaxases reveal wGRR values that are almost one order of magnitude higher than those of random comparisons between plasmids carrying homologous relaxases (fig. 6). This is consistent with genetic changes within the plasmid driving the transitions in terms of mobility and is inconsistent with the hypothesis of relaxase translocation to another plasmid. (3) Given the wGRR values and the similarities in size, the pCONJ to pdCONJ transitions seem to be the result of small deletions (fig. 8*B*. 3 *and E*). It is possible that many transitions between pCONJ and pMOB have also rapidly passed by a pdCONJ intermediate state that we are no longer able to observe. (4) The pdCONJs and pMOBs that recently transited from pCONJs have much larger sizes than the typical pMOBs (fig. 8C), suggesting they derived from pCONJs by gene deletions and not by translocation of the relaxase to a pMOBless (which tend to be small). Both groups of families also have more MPF genes than the other pMOBs. (5) The analysis at the microevolutionary level of PTUs from the MOB_{F12} family exemplifies how pMOBs, pdCONJs and pMOBless emerge multiple times within a group of pCONJs (fig. 7). Finally, gene loss is frequent in bacteria (Mira et al.

Downloaded from https://academic.oup.com/mbe/article/39/6/msac115/6593704 by Institut Pasteur - CeRIS user on 18 October 2022

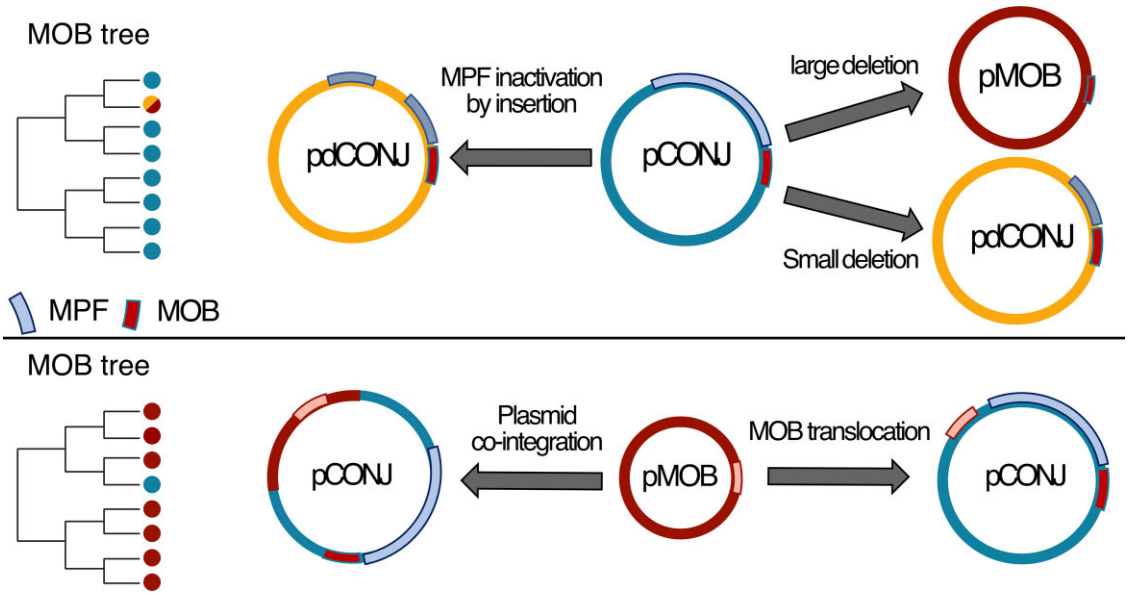


Fig. 11. Snapshots of the aftermath of the mobility transition events. Possible scenarios for plasmid transitions in terms of mobility type. The similarities in genes repertoires and size are given for the moment right after the transition occurred. MPF represents mating pair formation genes. MOB represents the relaxase.

2001) and provides a very simple mechanism for pCONJ to pMOB transitions. Deletion rates are increased in the presence of transposable elements (Cerveau et al. 2011), and we found these elements to be more abundant in pCONJs than in pMOBs and especially abundant in pdCONJs (fig. 2*D*). Naturally, the existence of numerous deletions resulting in pCONJ to pMOB transitions does not exclude the possibility that some transitions are caused by more complex mechanisms.

Our results suggest that some pCONJ to pMOB transitions have given rise to specialized relaxases that may be advantageous in certain circumstances, as described above. Since most bacterial clades have both conjugative and mobilizable plasmids and that half of the genomes with a mobilizable plasmid also contains a conjugative plasmid, there are many opportunities for the novel mobilizable plasmid to meet a conjugative element and transfer to other cells. In such a context, why should many novel pMOBs or pdCONJs be purged by natural selection as suggested by our data? If the relaxase is initially specialized in a T4SS (the one previously encoded in *cis*), then the novel mobilizable plasmid may have low transfer rates because it requires a very similar T4SS to transfer. This difficulty is amplified by plasmid incompatibility. If the novel mobilizable plasmid keeps its replication initiator protein, then it is incompatible with the most closely related pCONJ. Incompatibility means the two plasmids cannot coexist in a stable manner and this may decrease the rate of transmission of the newly formed mobilizable plasmid to a point where it will become extinct (if not adaptive to the host). This may explain why transitions between pCONJ and pMOB always involve significant changes in wGRR (fig. 6*A*): plasmids changing the replication machinery have a higher likelihood of surviving the initial stages after the transition. This fits our observation that many novel pMOBs have evolved to be part of a different incompatibility family, have divergent plasmid partition genes, encode different toxin–antitoxins, and endured a global change in gene repertoires (fig. 8*D*). Other reasons may contribute to the counterselection of novel mobilizable plasmids, including lack of coordination in expression of the relaxase and the compatible T4SS encoded in trans, or the existence of remnants of degraded T4SS that can be costly and even toxic to the cell.

Transitions from pMOB to pCONJ are rarer and the few observed cases are associated with increased plasmid size (fig. 8*B*,*E*). Such transitions may occur by translocation of a MPF in a pMOB or cointegration of a pMOB and a pCONJ (fig. 11). In both cases, the pMOB relaxase would now appear in the phylogenetic tree of the relaxases marked as a pCONJ, whereas remaining very close to relaxases marked as pMOB, that is, inside a pMOB clade. It should be noted that cointegration or large translocations are difficult to disentangle because they create hybrids that can then rapidly evolve by loss of genetic material. Of note, we observed that cooccurrence of relaxases in one plasmid is rare and the analysis of the observed cases shows that cooccurrence of typical pMOB and pCONJ relaxases is even rarer (fig. 4). This suggests that cointegrations of pCONJs (relaxase + MPF) and pMOBs are often followed by the loss of one of the two relaxases. Since the cooccurrence of pMOB and pCONJ relaxases is very rare, the relaxase lost after cointegration may often be the one of the pMOB. Alternatively, translocations of MPF (without relaxase) into a pMOB could result in a pMOB to pCONJ transition. In general, transitions between pMOB and pCONJ seem to require more complicated genetic mechanisms. They may also be less likely to result in well-adapted plasmids. pMOBs are generally small and often present in multiple copies in the cell.

Downloaded from https://academic.oup.com/mbe/article/39/6/msac115/6593704 by Institut Pasteur - CeRIS user on 18 October 2022

If the large hybrids resulting from the events described above replicate like small multi-copy plasmids, then they may be very costly to the host cells.

Transitions in terms of mobility are associated with changes in the classification of replication types of plasmids and the changes in gene repertoires can also affect their classification in terms of PTUs. We show that PTUs frequently contain several mobility types (fig. 7A). This is consistent with the observation that changes in mobility type are frequent in plasmid evolution. Many of these recent changes are not drastic enough to expel a plasmid from a PTU (fig. 7B). However, plasmids that endure substantial changes in their genomes tend to diverge from the original PTU. If these transitions are successful, that is, these plasmids persist and propagate in populations, this may generate novel PTUs, as observed for E5, F_{Sh}, and E41, which originated from F_E. These new PTUs may represent specific adaptations of the plasmid genetic structure to the constraints of a new condition, in this case a novel type of mobility. Hence, substantial diversity of genomic content can be found within a PTU, but very substantial diversification of plasmids following changes in mobility may result in novel PTUs.

Supplementary Material

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

Acknowledgements

Eugen Pfeifer for providing the wGRR data and Marie Touchon for providing the ISEScan analysis. Manuel Ares-Arroyo for comments and suggestions. Jorge Moura de Sousa, Matthieu Haudiquet, and Olaya Rendueles-Garcia for scientific discussions. INCEPTION project (PIA/ANR-16-CONV-0005), Equipe FRM (EQU201903007835), Laboratoire d'Excellence IBEID (ANR-10-LABX-62-IBEID) to E.P.C.R. and C.C. PID2020-117923GB-I00 project funded by Spanish Ministry of Science and Innovation to FdlC and MPG-B.

Data Availability

All major data, or identifiers to public repositories with the data, are incorporated into the article and its online supplementary material. Other data, e.g., intermediary results files, underlying this article will be shared on reasonable request to the corresponding author.

References

Abby SS, Neron B, Menager H, Touchon M, Rocha EP. 2014. MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-cas systems. *PLoS One*. **9**:e110726.

Adelberg EA, Pittard J. 1965. Chromosome transfer in bacterial conjugation. *Bacteriol Rev*. **29**:161–172.

Akarsu H, Bordes P, Mansour M, Bigot D-J, Genevaux P, Falquet L. 2019. TASmania: a bacterial toxin–antitoxin systems database. *PLoS Comput Biol*. **15**:e1006946.

Bastian M, Heymann S, Jacomy M. 2009. Gephi : an open source software for exploring and manipulating networks. *In: Third International ICWSM Conference*. p. 361–362.

Blanca-Ordóñez H, Oliva-García JJ, Pérez-Mendoza D, Soto MJ, Olivares J, Sanjuán J, Nogales J. 2010. pSymA-dependent mobilization of the Sinorhizobium meliloti pSymB megaplasmid. *J Bacteriol*. **192**:6309–6312.

Blesa A, Baquedano I, Quintáns NG, Mata CP, Castón JR, Berenguer J. 2017. The transjugation machinery of Thermus thermophilus: identification of TdtA, an ATPase involved in DNA donation. *PLoS Genet*. **13**:e1006669.

Bouet JY, Nordstrom K, Lane D. 2007. Plasmid partition and incompatibility—the focus shifts. *Mol Microbiol*. **65**:1405–1414.

Cabezón E, Lanka E, De La Cruz F. 1994. Requirements for mobilization of plasmids RSF1010 and ColE1 by the IncW plasmid R388: trwB and RP4 traG are interchangeable. *J Bacteriol*. **176**:4455–4458.

Cabezon E, Sastre JI, De La Cruz F. 1997. Genetic evidence of a coupling role for the TraG protein family in bacterial conjugation. *Mol Gen Genet*. **254**:400–406.

Carattoli A. 2013. Plasmids and the spread of resistance. *Int J Med Microbiol*. **303**:298–304.

Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, Møller Aarestrup F, Hasman H. 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother*. **58**:3895–3903.

Cerveau N, Leclercq S, Bouchon D, Cordaux R. 2011. Evolutionary dynamics and genomic impact of prokaryote transposable elements. *Evolutionary biology—concepts, biodiversity, macroevolution and genome evolution*: Springer. p. 291–312.

Che Y, Yang Y, Xu X, Brinda K, Polz MF, Hanage WP, Zhang T. 2021. Conjugative plasmids interact with insertion sequences to shape the horizontal transfer of antimicrobial resistance genes. *Proc Natl Acad Sci U S A*. **118**:e2008731118.

Chevin LM, Lande R. 2011. Adaptation to marginal habitats by evolution of increased phenotypic plasticity. *J Evol Biol*. **24**:1462–1476.

Coluzzi C, Guédon G, Devignes M-D, Ambroset C, Loux V, Lacroix T, Payot S, Leblond-Bourget N. 2017. A glimpse into the world of integrative and mobilizable elements in streptococci reveals an unexpected diversity and novel families of mobilization proteins. *Front Microbiol*. **8**:443.

Cury J, Oliveira PH, De La Cruz F, Rocha EPC. 2018. Host range and genetic plasticity explain the coexistence of integrative and extrachromosomal mobile genetic elements. *Mol Biol Evol*. **35**:2230–2239.

Cury J, Touchon M, Rocha EP. 2017. Integrative and conjugative elements and their hosts: composition, distribution and organization. *Nucl Acid Res*. **45**:8943–8956.

De La Cruz F, Davies J. 2000. Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends Microbiol*. **8**:128–133.

De La Cruz F, Frost LS, Meyer RJ, Zechner E. 2010. Conjugative DNA metabolism in gram-negative bacteria. *FEMS Microbiol Rev*. **34**:18–40.

Denise R, Abby SS, Rocha EP. 2019. Diversification of the type IV filament superfamily into machines for adhesion, protein secretion, DNA uptake, and motility. *PLoS Biol*. **17**:e3000390.

Diaz-Orejas R, Espinosa M, Yeo CC. 2017. The importance of the expendable: toxin–antitoxin genes in plasmids and chromosomes. *Front Microbiol*. **8**:1479.

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

Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol*. **7**:e1002195.

Downloaded from https://academic.oup.com/mbe/article/39/6/msac115/6593704 by Institut Pasteur - CeRIS user on 18 October 2022

Fernandez-Lopez R, De Toro M, Moncalian G, Garcillan-Barcia MP, De La Cruz F. 2016. Comparative genomics of the conjugation region of F-like plasmids: five shades of F. *Front Mol Biosci*. **3**:71.

Fürste JP, Pansegrau W, Ziegelin G, Kröger M, Lanka E. 1989. Conjugative transfer of promiscuous IncP plasmids: interaction of plasmid-encoded products with the transfer origin. *Proc Natl Acad Sci U S A*. **86**:1771–1775.

Garcillan-Barcia MP, Cuartas-Lanza R, Cuevas A, De La Cruz F. 2019. Cis-acting relaxases guarantee independent mobilization of MOBQ4 plasmids. *Front Microbiol*. **10**:2557.

Garcillan-Barcia MP, Francia MV, De La Cruz F. 2009. The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol Rev*. **33**:657–687.

Garcillan-Barcia MP, Redondo-Salvo S, Vielva L, De La Cruz F. 2020. MOBscan: automated annotation of MOB relaxases. *Methods Mol Biol*. **2075**:295–308.

Geng P, Cheng J, Yuan Z, Xiong H, Wang H, Hu X. 2021. Horizontal transfer of large plasmid with type IV secretion system and mosquitocidal genomic island with excision and integration capabilities in Lysinibacillus sphaericus. *Environ Microbiol*. **23**: 5131–5146.

Gonzalez-Perez B, Lucas M, Cooke LA, Vyle JS, De La Cruz F, Moncalián G. 2007. Analysis of DNA processing reactions in bacterial conjugation by using suicide oligonucleotides. *EMBO J*. **26**: 3847–3857.

Guglielmini J, De La Cruz F, Rocha EPC. 2013. Evolution of conjugation and type IV secretion systems. *Mol Biol Evol*. **30**:315–331.

Guglielmini J, Neron B, Abby SS, Garcillan-Barcia MP, La Cruz FD, Rocha EP. 2014. Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. *Nucleic Acids Res*. **42**:5715–5727.

Guglielmini J, Quintais L, Pilar Garcillan-Barcia M, De La Cruz F, Rocha EPC. 2011. The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. *PLoS Genet*. **7**:e1002222.

Guy L, Kultima JR, Andersson SG. 2010. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* **26**:2334–2335.

Hall JPJ, Brockhurst MA, Harrison E. 2017. Sampling the mobile gene pool: innovation via horizontal gene transfer in bacteria. *Philos Trans R Soc Lond B Biol Sci*. **372:**20160424.

Harrison PW, Lower RP, Kim NK, Young JP. 2010. Introducing the bacterial 'chromid': not a chromosome, not a plasmid. *Trends Microbiol*. **18**:141–148.

Hauser M, Mayer CE, Söding J. 2013. kClust: fast and sensitive clustering of large protein sequence databases. *BMC Bioinform*. **14**:248.

Hauser M, Steinegger M, Soding J. 2016. MMseqs software suite for fast and deep clustering and searching of large protein sequence sets. *Bioinformatics* **32**:1323–1330.

Humphrey S, San Millán Á, Toll-Riera M, Connolly J, Flor-Duro A, Chen J, Ubeda C, Maclean RC, Penadés JR. 2021. Staphylococcal phages and pathogenicity islands drive plasmid evolution. *Nat Commun*. **12**:5845.

Ishikawa SA, Zhukova A, Iwasaki W, Gascuel O. 2019. A fast likelihood method to reconstruct and visualize ancestral scenarios. *Mol Biol Evol*. **36**:2069–2085.

Jain A, Srivastava P. 2013. Broad host range plasmids. *FEMS Microbiol Lett*. **348**:87–96.

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

Johnson TJ, Nolan LK. 2009. Pathogenomics of the virulence plasmids of *Escherichia coli*. *Microbiol Mol Biol Rev*. **73**:750–774.

Katoh K, Standley DM. 2014. MAFFT: iterative refinement and additional methods. *Methods Mol Biol*. **1079**:131–146.

Klümper U, Droumpali A, Dechesne A, Smets BF. 2014. Novel assay to measure the plasmid mobilizing potential of mixed microbial communities. *Front Microbiol*. **5**:730.

Kobayashi H. 2018. Regeneration of *Escherichia coli* from minicells through lateral gene transfer. *J Bacteriol*. **200**:e00630-17.

Lacroix B, Citovsky V. 2018. Beyond Agrobacterium-mediated transformation: horizontal gene transfer from bacteria to eukaryotes. *Agrobacterium Biol*. 443–462.

Lanza VF, Baquero F, De La Cruz F, Coque TM. 2017. AcCNET (Accessory Genome Constellation Network): comparative genomics software for accessory genome analysis using bipartite networks. *Bioinformatics* **33**:283–285.

Liu MA, Kwong SM, Jensen SO, Brzoska AJ, Firth N. 2013. Biology of the staphylococcal conjugative multiresistance plasmid pSK41. *Plasmid* **70**:42–51.

Lorenzo-Diaz F, Fernandez-Lopez C, Garcillan-Barcia MP, Espinosa M. 2014. Bringing them together: plasmid pMV158 rolling circle replication and conjugation under an evolutionary perspective. *Plasmid* **74**:15–31.

Meyer R. 2009. Replication and conjugative mobilization of broad host-range IncQ plasmids. *Plasmid* **62**:57–70.

Mira A, Ochman H, Moran NA. 2001. Deletional bias and the evolution of bacterial genomes. *Trends Genet*. **17**:589–596.

Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. **32**:268–274.

Nordstrom K, Molin S, Aagaard-Hansen H. 1980. Partitioning of plasmid R1 in Escherichia coli. II. Incompatibility properties of the partitioning system. *Plasmid* **4**:332–339.

Novick RP. 1987. Plasmid incompatibility. *Microbiol Rev*. **51**:381–395.

Nuti M, Lepidi A, Prakash R, Schilperoort R, Cannon F. 1979. Evidence for nitrogen fixation (nif) genes on indigenous Rhizobium plasmids. *Nature* **282**:533–535.

O'Brien FG, Yui Eto K, Murphy Riley JT, Fairhurst Heather M, Coombs GW, Grubb WB, Ramsay JP. 2015. Origin-of-transfer sequences facilitate mobilisation of non-conjugative antimicrobial-resistance plasmids in Staphylococcus aureus. *Nucleic Acids Res*. **43**:7971–7983.

Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**: 526–528.

Partridge SR, Kwong SM, Firth N, Jensen SO. 2018. Mobile genetic elements associated with antimicrobial resistance. *Clin Microbiol Rev*. **31**:e00088-00017.

Pedersen T, Sekyere JO, Govinden U, Moodley K, Sivertsen A, Samuelsen Ø, Essack SY, Sundsfjord A. 2018. Spread of plasmid-encoded NDM-1 and GES-5 carbapenemases among extensively drug-resistant and pandrug-resistant clinical enterobacteriaceae in durban, South Africa. *Antimicrob Agents Chemother*. **62**:e02178-02117.

Perez-Mendoza D, Lucas M, Munoz S, Herrera-Cervera JA, Olivares J, De La Cruz F, Sanjuan J. 2006. The relaxase of the Rhizobium etli symbiotic plasmid shows nic site cis-acting preference. *J Bacteriol*. **188**:7488–7499.

Peter S, Bosio M, Gross C, Bezdan D, Gutierrez J, Oberhettinger P, Liese J, Vogel W, Dorfel D, Berger L, *et al*. 2020. Tracking of antibiotic resistance transfer and rapid plasmid evolution in a hospital setting by nanopore sequencing. *mSphere* **5**:e00525-20.

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

Ramachandran G, Miguel-Arribas A, Abia D, Singh PK, Crespo I, Gago-Cordoba C, Hao JA, Luque-Ortega JR, Alfonso C, Wu LJ. 2017. Discovery of a new family of relaxases in Firmicutes bacteria. *Plos Genet*. **13**:e1006586.

Ramsay JP, Firth N. 2017. Diverse mobilization strategies facilitate transfer of non-conjugative mobile genetic elements. *Curr Opin Microbiol*. **38**:1–9.

Ramsay JP, Kwong SM, Murphy RJ, Yui Eto K, Price KJ, Nguyen QT, O'brien FG, Grubb WB, Coombs GW, Firth N. 2016. An updated view of plasmid conjugation and mobilization in Staphylococcus. *Mob Genet Elem*. **6**:e1208317.

Rankin DJ, Rocha EPC, Brown SP. 2011. What traits are carried on mobile genetic elements, and why? *Heredity* **104**:1–10.

Downloaded from https://academic.oup.com/mbe/article/39/6/msac115/6593704 by Institut Pasteur - CeRIS user on 18 October 2022

Redondo-Salvo S, Bartomeus-Peñalver R, Vielva L, Tagg KA, Webb HE, Fernández-López R, De La Cruz F. 2021. COPLA, a taxonomic classifier of plasmids. *BMC Bioinform*. **22**:390.

Redondo-Salvo S, Fernández-López R, Ruiz R, Vielva L, De Toro M, Rocha EP, Garcillán-Barcia MP, De La Cruz F. 2020. Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids. *Nat Commun*. **11**:1–13.

Reid CJ, Cummins ML, Börjesson S, Brouwer MSM, Hasman H, Hammerum AM, Roer L, Hess S, Berendonk T, Nešporová K, *et al.* 2022. A role for ColV plasmids in the evolution of pathogenic *Escherichia coli* ST58. *Nat Commun*. **13**:683.

Revilla C, Garcillán-Barcia MP, Fernández-López R, Thomson NR, Sanders M, Cheung M, Thomas CM, De La Cruz F. 2008. Different pathways to acquiring resistance genes illustrated by the recent evolution of IncW plasmids. *Antimicrob Agents Chemother*. **52**:1472–1480.

Rocha EPC, Touchon M, Feil EJ. 2006. Similar compositional biases are caused by very different mutational effects. *Genome Res*. **16**:1537–1547.

San Millan A, Craig Maclean R. 2019. Fitness costs of plasmids: a limit to plasmid transmission. *Microb Transm*. **5**:65–79.

San Millan A, Heilbron K, Maclean RC. 2014. Positive epistasis between co-infecting plasmids promotes plasmid survival in bacterial populations. *ISME J*. **8**:601–612.

Sastre JI, Cabezón E, De La Cruz F. 1998. The carboxyl terminus of protein TraD adds specificity and efficiency to F-plasmid conjugative transfer. *J Bacteriol*. **180**:6039–6042.

Smillie C, Pilar Garcillan-Barcia M, Victoria Francia M, Rocha EPC, De La Cruz F. 2010. Mobility of plasmids. *Microbiol Mol Biol Rev*. **74**:434–452.

Sokurenko EV, Gomulkiewicz R, Dykhuizen DE. 2006. Source–sink dynamics of virulence evolution. *Nat Rev Microbiol*. **4**:548–555.

Soler N, Robert E, de Beauchêne IC, Monteiro P, Libante V, Maigret B, Staub J, Ritchie DW, Guédon G, Payot S, *et al.* 2019. Characterization of a relaxase belonging to the MOB(T) family, a widespread family in Firmicutes mediating the transfer of ICEs. *Mob DNA* **10**:18.

Soucy SM, Huang J, Gogarten JP. 2015. Horizontal gene transfer: building the web of life. *Nat Rev Genet*. **16**:472.

Stalder T, Press MO, Sullivan S, Liachko I, Top EM. 2019. Linking the resistome and plasmidome to the microbiome. *ISME J*. **13**:2437–2446.

Steenwyk JL, Buida III TJ, Li Y, Shen X-X, Rokas A. 2020. ClipKIT: a multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biol*. **18**:e3001007.

Steinegger M, Meier M, Mirdita M, Vohringer H, Haunsberger SJ, Soding J. 2019. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinform*. **20**:473.

Szabó M, Murányi G, Kiss J. 2021. IncC helper dependent plasmid-like replication of Salmonella Genomic Island 1. *Nucleic Acids Res*. **49**:832–846.

Tomita H, Ike Y. 2005. Genetic analysis of transfer-related regions of the vancomycin resistance enterococcus conjugative plasmid pHTβ: identification of oriT and a putative relaxase gene. *J Bacteriol*. **187**:7727–7737.

Trieu-Cuot P, Carlier C, Martin P, Courvalin P. 1987. Plasmid transfer by conjugation from *Escherichia coli* to Gram-positive bacteria. *FEMS Microbiol Lett*. **48**:289–294.

Waskom ML. 2021. Seaborn: statistical data visualization. *J Open Source Softw*. **6**:3021.

Wisniewski JA, Traore DA, Bannam TL, Lyras D, Whisstock JC, Rood JI. 2016. TcpM: a novel relaxase that mediates transfer of large conjugative plasmids from Clostridium perfringens. *Mol Microbiol*. **99**:884–896.

Yui Eto K, Kwong SM, LaBreck PT, Crow JE, Traore DAK, Parahitiyawa N, Fairhurst HM, Merrell DS, Firth N, Bond CS, *et al.* 2021. Evolving origin-of-transfer sequences on staphylococcal conjugative and mobilizable plasmids-who's mimicking whom? *Nucleic Acids Res*. **49**:5177–5188

Downloaded from https://academic.oup.com/mbe/article/39/6/msac115/6593704 by Institut Pasteur - CeRIS user on 18 October 2022