Published online 29 November 2022
Nucleic Acids Research, 2023, Vol. 51, No. 7 **3001–3016**
https://doi.org/10.1093/nar/gkac1079

NAR Breakthrough Article

Origins of transfer establish networks of functional dependencies for plasmid transfer by conjugation

Manuel Ares-Arroyo*, Charles Coluzzi and Eduardo P.C. Rocha*

Institut Pasteur, Université de Paris Cité, CNRS UMR3525, Microbial Evolutionary Genomics, Paris, France

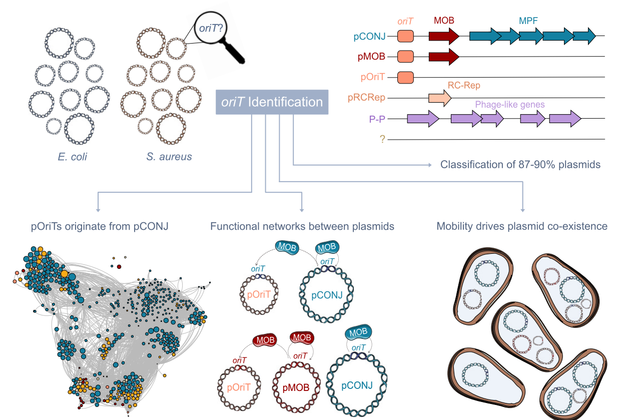
Received August 30, 2022; Revised October 17, 2022; Editorial Decision October 20, 2022; Accepted October 27, 2022

Downloaded from https://academic.oup.com/nar/article/51/7/3001/6849495 by guest on 30 September 2023

ABSTRACT

Plasmids can be transferred between cells by conjugation, thereby driving bacterial evolution by horizontal gene transfer. Yet, we ignore the molecular mechanisms of transfer for many plasmids because they lack all protein-coding genes required for conjugation. We solved this conundrum by identifying hundreds of plasmids and chromosomes with conjugative origins of transfer in *Escherichia coli* and *Staphylococcus aureus*. These plasmids (pOriT) hijack the relaxases of conjugative or mobilizable elements, but not both. The functional dependencies between pOriT and other plasmids explain their co-occurrence: pOriT are abundant in cells with many plasmids, whereas conjugative plasmids are the most common in the others. We systematically characterized plasmid mobility in relation to conjugation and alternative mechanisms of transfer and can now propose a putative mechanism of transfer for ∼90% of them. In most cases, plasmid mobility seems to involve conjugation. Interestingly, the mechanisms of mobility are important determinants of plasmid-encoded accessory traits, since pOriTs have the highest densities of antimicrobial resistance genes, whereas plasmids lacking putative mechanisms of transfer have the lowest. We illuminate the evolutionary relationships between plasmids and suggest that many pOriT may have arisen by gene deletions in other types of plasmids. These results suggest that most plasmids can be transferred by conjugation.

GRAPHICAL ABSTRACT



INTRODUCTION

Plasmids are extra-chromosomal DNA molecules that have an important role in horizontal gene transfer (1), being key contributors to the spread of antimicrobial resistance genes, virulence factors, and metabolic traits (2). The transfer of a plasmid between cells can take place by several processes (3). Some plasmids can be transferred passively, i.e. without dedicated genetic determinants encoded in the plasmid, by natural transformation (4), in vesicles (5), or by transducing bacteriophages (phages) (6). Some plasmids are also phages, phage-plasmids (P-P) and transfer by producing viral particles where they package their own DNA (7). Yet, one commonly considers that conjugation is the major mechanism of plasmid transfer (8).

Conjugation involves the recognition by the relaxase (MOB) of a small DNA sequence in the plasmid called the origin of transfer (*oriT*) (9). The relaxase cleaves the *oriT* at the *nic* site and binds covalently to the single-stranded

*To whom correspondence should be addressed. Tel: +33 1 86 46 74 77; Email: manuel.ares-arroyo@pasteur.fr
Correspondence may also be addressed to Eduardo P.C. Rocha. Tel: +33 1 40 61 33 53; Email: eduardo.rocha@pasteur.fr

© The Author(s) 2022. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://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

DNA. This nucleoprotein complex, named relaxosome, interacts with a type 4 coupling protein that connects it to the mating pair formation (MPF), including a Type 4 Secretion System (T4SS) that transfers the nucleoprotein complex to another cell (10). Once the relaxosome has been transferred, the relaxase catalyzes the DNA ligation of the plasmid in the recipient cell to produce a circular single stranded molecule that is replicated by the replication machinery of the recipient cell (9). At the end of conjugation there is one copy of the plasmid in each cell. Some conjugative elements remain in cells as plasmids whereas others integrate into the chromosome as integrative conjugative elements (ICEs) (11). The conjugation machineries of ICEs and plasmids are very similar and have intermingled evolutionary histories (12).

Plasmids or integrative elements encoding the three functional elements—*oriT*, relaxase and MPF—may conjugate autonomously between bacteria. They are called *conjugative* (8). However, plasmids encoding the MPF represent only ∼1/4 of all plasmids. Those lacking an MPF but encoding a relaxase and *oriT* are called *mobilizable*. In this case, the relaxase interacts with the plasmid *oriT*, and the resulting nucleoprotein complex is transported by the MPF of a conjugative element co-occurring in the donor cell. Plasmids encoding a relaxase but lacking a complete MPF are as numerous as the conjugative plasmids (8). This means that half of all plasmids lack a relaxase and an MPF. We will refer to them as pMOBless hereinafter. Even though pMOBless lack all proteins required for conjugation, there is epidemiological evidence that some of them transfer between cells (13–15). The mobility of pMOBless may occur by several mechanisms: (i) they may have an *oriT* and be mobilized by a relaxase and an MPF encoded *in-trans* by a conjugative plasmid (16); (ii) they may interact with a relaxase of a mobilizable plasmid, and the nucleoprotein complex further interacts with an MPF of a third plasmid (17); (iii) or they may transfer using other mechanisms, e.g. conjugation through a rolling circle replication protein (18), co-integration with a conjugative plasmid (19) or the alternative transfer mechanisms mentioned above. Similar mechanisms could be used by integrative elements lacking a complete MPF, commonly named integrative mobilizable elements (IMEs) (20).

The observation over a decade ago that slightly more than half of all plasmids lack genes for relaxases was paradoxical, because genetic mobility is thought to be necessary for plasmid maintenance in populations (21,22). Of note, some pMOBless with an *oriT* (pOriT hereinafter) were shown to be mobilized by a conjugative plasmid decades ago (17). Yet, the few available sequences of *oriT* have precluded systematic identification of these plasmids. Recently, pioneering studies on *Staphylococcus aureus*, a species that has unusually few conjugative plasmids and few types of *oriT*, showed that 50% of the pMOBless can be mobilized since they carry *oriT*s similar to those of pWBG749 (23) or pSK41 (24). Subsequent studies with three additional *oriT*s, suggested that *oriT*-based mobilization is common in this species (25,26). If this is true for other species, including those with numerous conjugative plasmids, is not known. Unfortunately, most *oriT*s remain unknown, precluding their systematic study across bacteria. Here, we focused on *S. aureus*, for which plasmid diversity is low and well-characterized and *Escherichia coli*, the best described species of bacteria and one with numerous well-known plasmid families (27). These two species are of particular importance because they are responsible for the greatest number of deaths associated to antimicrobial resistance in the world (28), a trait that is spread by plasmids (29). We first complement previous studies and test if ICEs could be involved in the mobilization of pOriTs in *S. aureus*. We also test if the same approach can be extended to *E. coli*. The confirmation that we can identify homologs of experimentally verified *oriT*s in the plasmids of these species paved the way to answer some outstanding questions. We don't know how these plasmids contribute to the spread of functions across bacteria. We don't know the functional dependencies associated with pOriTs, i.e. if they tend to be associated with one single conjugative plasmid or if they often require a third plasmid encoding a relaxase. We don't know how these plasmids arose in natural history. We also ignore how the existence of pOriTs affects the patterns of co-occurrence of plasmids in cells. Finally, we would like to know how many plasmids remain without a hypothetical mechanism of transfer once pOriT plasmids and phage-plasmids are accounted for. By tackling these questions, this study contributes to unravel the mechanisms of plasmid mobility.

MATERIALS AND METHODS

Genome data

We retrieved all the *E. coli* and *S. aureus* complete genomes available in the NCBI non-redundant RefSeq database in March 2021. This resulted in a set of 1585 genomes of *E. coli* and 582 genomes of *S. aureus*, including 3409 and 462 plasmids, respectively. The information on the plasmids (including accession numbers) is available in the Supplementary Table S1. The information on the chromosomes is available in the Supplementary Table S2.

Collection of the *oriT* database and its identification in the complete genomes

We built a collection of experimentally validated origins of transfer. First, we retrieved the 52 *oriT*s with a status '*experimental*' from the already published *oriT* database by Li *et al.* (30). We expanded this collection by consulting the literature, using as a query '*oriT*' in the PubMed database (available in September 2021). Among the 708 entries, we screened for experimentally validated *oriT*s not included in the aforementioned database. This resulted in the retrieval of 47 additional *oriT*s. However, one *oriT* from the published database and seven *oriT*s from the literature were discarded from the collection as only the *nic*-site sequence was available. This resulted in a final dataset of 91 origins of transfer. Information on this collection is available in Supplementary Table S3.

We used the BLAST suite of programs, version 2.9.0+, to identify *oriT*s (31). The complete genomes of *E. coli* and *S. aureus* were indexed with makeblastdb (default parameters). Then, we used blastn to search for occurrences of each of the 91 *oriT*s (query) against the database of complete genomes. Due to the short length of the origins of transfer, blastn was used with the option -*task blastn-short* and an

Downloaded from https://academic.oup.com/nar/article/51/7/3001/6849495 by guest on 30 September 2023

E-value threshold of 0.01 following the developer's instructions. In cases in which two different *oriT*s were identified in the same region of a plasmid (overlapping), only the *oriT* hit with the best *E*-value was retrieved.

We identified during this screening an exceptional case of a ∼50 kb plasmid with 23 identical *oriT*s. This plasmid (NZ_CP019265.1) was discarded from further analysis as we considered it to be a sequencing artifact.

Characterization of conjugative systems and relaxases and plasmid classification on the mobility

We used the module CONJscan of MacSyFinder, version 2.0 (32) to identify all the complete MPF systems. The individual hidden Markov model (HMM) hits that were not associated with MPFs deemed complete were used to identify incomplete MPF systems.

Relaxases were identified using HMMER version 3.3.2 (33), and the HMM profiles employed by the software MOBscan (34). We used the tool hmmsearch (default options) to screen for relaxases in all the proteins annotated in the dataset and kept the 2195 significant hits with >50% coverage on the profile. A careful analysis of the results revealed that this version of the RefSeq annotations sometimes missed genes encoding relaxases, especially when these genes overlapped others (Supplementary Figure S1). To correct this problem, we introduced a preliminary step of re-annotation. This ensured a coherent annotation of the genes throughout all the genomes, which was then used to identify the MPF and the relaxases. For the annotation, we used the software Prodigal, version 2.6.3 (35), with the recommended mode for plasmids and viruses to identify all open reading frames. Hits were then identified as mentioned above. When two different profiles matched the same protein, we kept the one with the lowest *E*-value.

Plasmids were classified in different mobility categories depending on their composition in terms of *oriT*, relaxase, and MPF genes. Plasmids encoding a putatively complete MPF system (including a relaxase) were considered to be conjugative (pCONJ). Plasmids encoding relaxases and lacking a complete MPF system were classified as mobilizable (pMOB). The remaining plasmids were classified as pMOBless, and were split into different categories: pOriTs when they had an *oriT,* P-Ps when they were phage-related elements (see below), pRC-Rep plasmids if they were not included into a prior category but encode for a RC-Rep protein (see below), and presumably non-transmissible plasmids (pNTs) otherwise. In addition, some plasmids were classified as decayed conjugative plasmids (pdCONJ). These plasmids encode two or more MPF genes, but not enough to form a complete MPF system. Therefore, pdCONJ show a close evolutionary relationship with conjugative plasmids (36), but are functionally equivalent to pMOB, pOriT or pNT in terms of mobility (Supplementary Figure S2). Similarly, the loci encoding presumably complete MPF systems in chromosomes were classed as ICE (Integrative and Conjugative Element), even if often we ignore the precise limits of the element. Chromosomal genes encoding relaxases that were distant from genes encoding MPFs (>60 genes) were classed as IME (Integrative and Mobilizable Element).

Identification of rolling circle replication proteins

We identified Rolling Circle Replication (RC-Rep) proteins involved in plasmid conjugation by retrieving the RC-Rep protein sequence of the *S. aureus* plasmid pC194 (NC_002013.1), a pMOBless plasmid known to be mobilized through *in trans* conjugation (37). We used its Pfam profile (38), Rep_1 (PF01446), to look for related RC-Rep proteins in all the plasmids of *E. coli* and *S. aureus* using the HMMER tool hmmsearch (default options, *E*-value < 0.001), version 3.3.2 (23).

Identification of phage-plasmids

We identified P-Ps using the data on *E. coli* and *S. aureus* that was recently published (39). The database used in the cited work corresponds to the same RefSeq database (retrieved on March 2021).

Analysis of the pangenome of *E. coli* and *S. aureus* plasmids

The pangenome of the plasmid-encoded genes of *E. coli* and *S. aureus* was identified using the module pangenome of the software PanACoTa, version 1.3.1 (40). Briefly, gene families were built with MMseqs2, version 13.45111, with an identity threshold of 80%. This is the typical threshold for the determination of the *E. coli* pangenome (41). This way, the 227428 plasmid-encoded proteins in *E. coli* were grouped into 11530 gene families. In *S. aureus*, the 7902 proteins were grouped into 1010 gene families. Some plasmids were not used in the analysis because their annotations lacked protein coding genes: 32 of the 3409 plasmids in *E. coli* (0.94%) and 20 of the 482 in *S. aureus* (4,15%). Rarefaction curves were performed with the R package vegan, version 2.5–6 (https://CRAN.R-project.org/package=vegan). The later package was additionally employed to infer the plasmid pangenome of *S. aureus* until matching the same sample size as *E. coli* following an Arrhenius model. Additionally, the Gleason model and Gitay model were used to extrapolate the rarefaction curves of the pangenome for *S. aureus* (Supplementary Figure S3). Rarefaction curves were plotted with sample sizes increasing by a step of 100 plasmids.

Determination of sequence similarity between plasmids

We assessed sequence similarity for all pairs of the 3869 plasmids using two different approaches.

At the micro-evolutionary scale, to analyze very closely related plasmids, we classified them based on their average nucleotide identity (ANI) into the existing catalogue of Plasmid Taxonomic Units (PTUs) (27). The clustering was performed using COPLA (42), version 1.0 (default parameters).

At the macro-evolutionary scale, to analyze more distantly related plasmids, we assessed the gene relatedness within and between PTUs, using the weighted Gene Repertoire Relatedness (wGRR) (43). For this, we searched for sequence similarity between all the proteins identified in the plasmids using MMseqs2 (version 9-d36de) (44), retrieving the hits with *E*-value $<10^{-4}$ and coverage >50%. Best bi-directional hits (BBH) between pairs of plasmids were used

Downloaded from https://academic.oup.com/nar/article/51/7/3001/6849495 by guest on 30 September 2023

to calculate the wGRR as previously described (43):

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

where A_i and B_i are the ith 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 in the smallest plasmid of the pair. This way, the wGRR value varies between 0 (no BBH between the plasmids) and 1 (all genes of the smallest plasmid have an identical homolog in the larger one). The wGRR values were used to identify related plasmids between and within PTUs, setting the threshold in wGRR > 0.75 as previously described (36). With this purpose, only plasmid pairs with wGRR >0.75 were retrieved for visualizations, i.e. at least the 75% of genes encoded in the smallest plasmid are shared between the pair.

Clustering *oriT*s by sequence similarity

We clustered the *oriT*s by searching for sequence similarity between all pairs of *oriT*s in the reference dataset using blastn (31) (Supplementary Figure S4). BLAST was used with the option *-task blastn-short* and an *E*-value threshold of 0.01. Only matches with >80% identity and >70% coverage of the smallest *oriT* were kept for the clustering analysis. The clustering was performed with the hierarchical method available in the R package pheatmap, version 1.0.12 (default options) (https://CRAN.R-project.org/package=pheatmap). The clusters were named after well-known *oriT*s contained in the cluster: F-like, R6K-like, R64-like, ColE1-like, RP4-like and R46-like. The association of each *oriT* to their *oriT* family is available in the Supplementary Table S3 and Supplementary Figure S4.

Identification of antimicrobial resistance genes

We identified antimicrobial resistance genes in the plasmid dataset using AMRFinderPlus (45), version 3.10, with the default options. This tool combines BLASTP and HMMER to identify the 6189 resistance determinants available in the NCBI Pathogen Detection Reference Gene Catalog (April 2022). The latter is the result of the curated merging of various widespread-used databases, including CARD (46), and ResFinder (47) databases, among others (45).

Statistical analysis

Except where explicitly stated, all statistical analyses were done with R, version 3.5.2. Additionally, all visualizations were performed with the R package ggplot2 (https://CRAN.R-project.org/package=ggplot2), version 3.3.5, occasionally supported by the R packages ggsignif (https://CRAN.R-project.org/package=ggsignif), version 0.6.0 and ggridges (https://CRAN.R-project.org/package=ggridges), version 0.5.3. For the construction and visualization of the networks, we used the R package igraph (https://CRAN.R-project.org/package=igraph), version 1.2.4.1 and the software Gephi 0.9.2 (48), respectively.

RESULTS

Characterization of *E. coli and S. aureus* plasmid repertoires

We analyzed the complete genomes available in RefSeq of *E. coli* (n = 1585) and *S. aureus* (n = 581) to characterize the size and diversity of their plasmids. *E. coli* isolates carry almost three times more plasmids per genome than *S. aureus* isolates ($t_{(2068.9)} =$ 20.65; P < 2.2e–16) (Figure 1A). Moreover, *E. coli* plasmids tend to be larger (Kolmogorov–Smirnov test, D = 0.586, P < 2.2e–16) (Figure 1B) and have higher GC% than *S. aureus* plasmids ($t_{(1074.7)}$ = 191.23, P < 2.2e–16) (Supplementary Figure S5). They are also more diverse in terms of gene repertoires. *E. coli* plasmids encode on average four times more gene families than those of *S. aureus* ($t_{(2817.9)}$ = 43.129, P < 2.2e–16) (Supplementary Figure S5). The plasmid pangenome of *E. coli* (11 530 gene families) is much larger than that of *S. aureus* (ca. 1000), even when comparing samples with similar numbers of plasmids (Figure 1C). Overall, plasmids contribute many genes to the species pangenomes. This is particularly striking in *E. coli*, where the plasmid pangenome is more than double the average size of a strain genome (41).

We characterized the plasmids in terms of the protein-coding genes involved in conjugation: pCONJ encode an MPF and a relaxase, pMOB encode a relaxase, and pMOBless lack a relaxase. In *E. coli* ∼35% of the plasmids are pCONJ, ∼25% pMOB, and ∼40% pMOBless (Figure 1D). These values are close to previously published ones across Bacteria (8). In contrast, only 4% of the *S. aureus* plasmids were classed as pCONJ, 18% as pMOB and 77% as pMOBless. Hence, *S. aureus* seems a more atypical bacteria, where conjugative plasmids are rare. We then tested the hypothesis that ICEs could compensate for the paucity of conjugative plasmids in the species. We searched the chromosomes for loci associated with ICEs (encoding MPF and relaxase) and IMEs (encoding a relaxase), and found that 46% of the chromosomes of *S. aureus* encode MPF systems (Figure 1E). In contrast, conjugative systems were identified in only ∼7% of *E. coli* chromosomes. Interestingly, many genomes in both species have either conjugative plasmids or ICEs, but rarely both. The integration of these analyses provides a more nuanced view of the differences between the species in terms of the fraction of genomes containing a conjugative element: ∼52% of *E. coli* and ∼47% of *S. aureus* (Figure 1F). The precise identification of the limits of ICEs and IMEs in the chromosome is difficult and precludes systematic comparisons between elements in terms of gene content. Still, these results suggest that the existence of ICEs could explain the mobility of some pMOBless, especially in *S. aureus*. In summary, the two species show different patterns in terms of the mobility of plasmids and integrative elements, but both contain many plasmids lacking relaxases.

oriTs are frequent in plasmids of *E. coli* and *S. aureus*

To unveil the mechanisms of mobilization of the many plasmids lacking a relaxase, we searched the literature for plasmids with just an *oriT* that could be mobilized by conjugative or mobilizable plasmids (Supplementary Tables S4 and S5). All these plasmids are artificial constructions, so they cannot be used for validation of our method. Yet, they

Downloaded from https://academic.oup.com/nar/article/51/7/3001/6849495 by guest on 30 September 2023

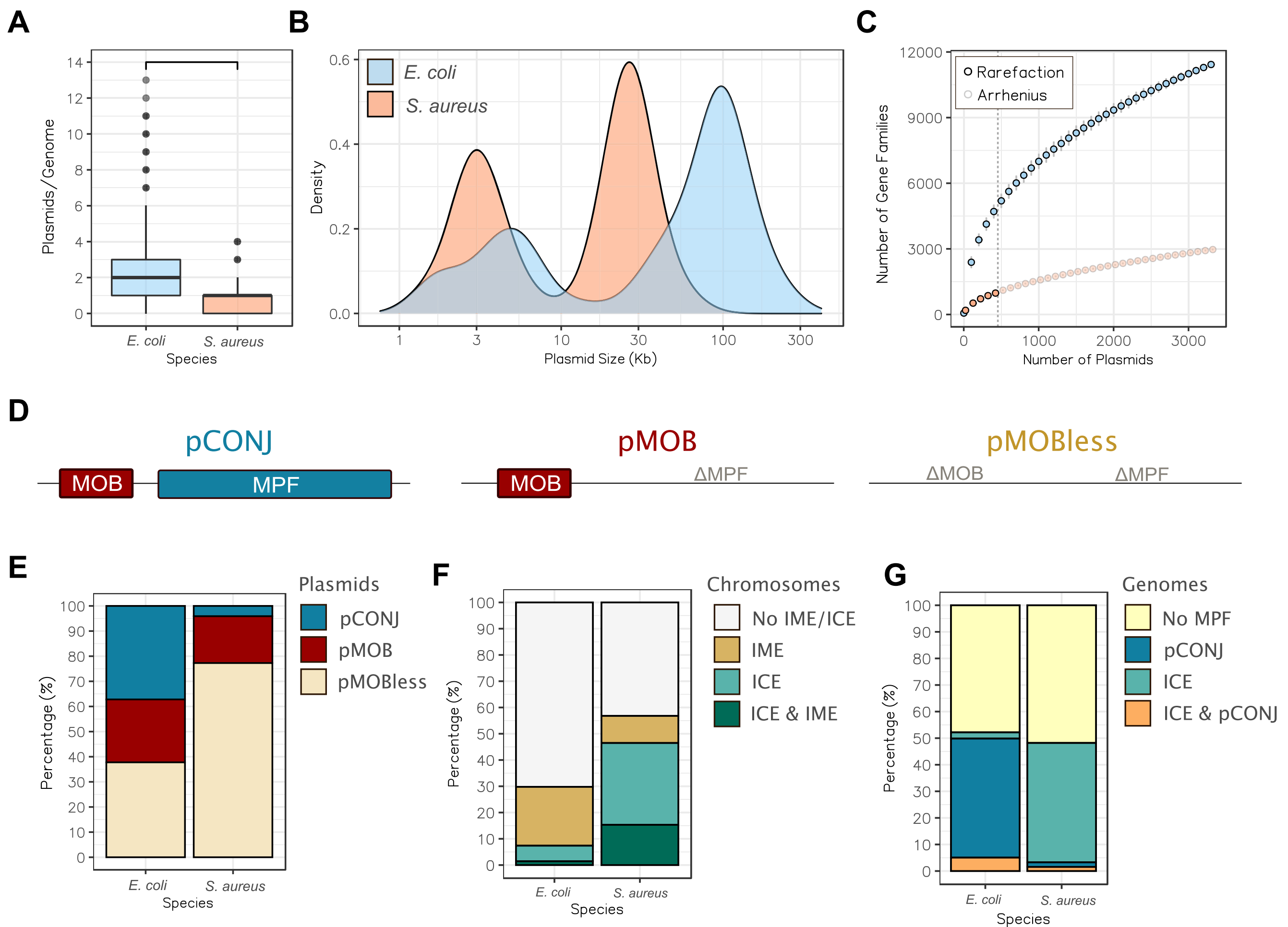


Figure 1. *E. coli* and *S. aureus* plasmids and integrative elements. (**A**) Number of plasmids per genome. The horizontal lines represent the median value, while the lower and upper hinges correspond to the first and third quartiles. The whiskers extend from the hinge to 1.5 times the range between the first and third quartile. Data beyond these values are shown as dots. The horizontal bar over the plot denotes statistically significant difference ($t_{(2068.9)}$ = 20.65; $P < 0.0001$). (**B**) Plasmid size distribution. The curves were drawn using a kernel density estimate. (**C**) Plasmid pangenome of *E. coli* and *S. aureus* attending to the number of plasmids sampled. The vertical dashed grey line at $x = 455$ represents the number of plasmids from which *S. aureus* pangenome is inferred following an Arrhenius model. (**D**) Classification of plasmids according to their protein-coding genes involved in conjugation: conjugative (pCONJ, MOB + MPF); mobilizable (pMOB, only MOB); and pMOBless (neither MPF, nor MOB). (**E**) Percentage of each mobility type among the plasmid repertoire of both species. (**F**) Percentage of the chromosomes with at least one ICE (complete MPF and relaxase), IME (relaxase without a complete MPF), both ICE and IME, or none of them. (**G**) Percentage of genomes with a complete MPF in plasmids (pCONJ), in the chromosome (ICE), or in both (ICE and pCONJ).

confirm that these *oriT*-carrying pMOBless can be transferred by conjugation using genes from other plasmids. To screen for *oriT*s in our plasmid collection, we collected 51 *oriT* from the '*oriT* database' (30) and added 40 new ones from the literature (Supplementary Table S3). Most of these 91 experimentally validated *oriT*s (mean size ∼131 bp) were originally identified and verified in plasmids of γ-Proteobacteria ($n = 44$) and Bacilli ($n = 22$) (Supplementary Figure S6). We then used the collection of *oriT* sequences to search for origins of transfer in the 1585 *E. coli* and 581 *S. aureus* genomes by sequence similarity (see Methods). We identified 2831 putative *oriT*s in 2626 plasmids, almost the totality of which locate in intergenic regions (Supplementary Figure S7). Even if *E. coli* has more diverse plasmids and more types of *oriT*s ($n = 37$) than *S. aureus* ($n = 7$), *oriT*s were found at similar frequencies in the plasmids of the two species (ca. 70%) (Figure 2A). We also identified 336 *oriT*s in 282 chromosomes. These chromosomal *oriT* were much more abundant in *S. aureus* (25% of the genomes) than in *E. coli* (9%), in line with the higher frequency of ICEs in the former (Figure 2A). Although many *oriT*s were identified in both types of replicons, a given family tends to be present either in plasmids or in chromosomes (Figure 2B). Importantly, none of the *oriT*s was identified in both species.

Most *oriT*-encoding plasmids have just one *oriT* (∼88% *E. coli*, ∼85% *S. aureus*), although a few can have up to 5 (Supplementary Figure S7). Expectedly, plasmids showing multiple *oriT*s tend to encode multiple relaxases ($r_{(3868)}$ = 0.32, $P < 2.2e{-}16$) (Supplementary Figure S7). To study the co-occurrence of *oriT*s and relaxases, we retrieved the families of *oriT*s identified in >10 plasmids. The *oriT*s of a given family are usually associated with plasmids of a specific size range, i.e. they tend to be associated to either small or large plasmids (Figure 2C). Yet, in a few cases, the *oriT* families associated with large plasmids are also found on

Downloaded from https://academic.oup.com/nar/article/51/7/3001/6849495 by guest on 30 September 2023

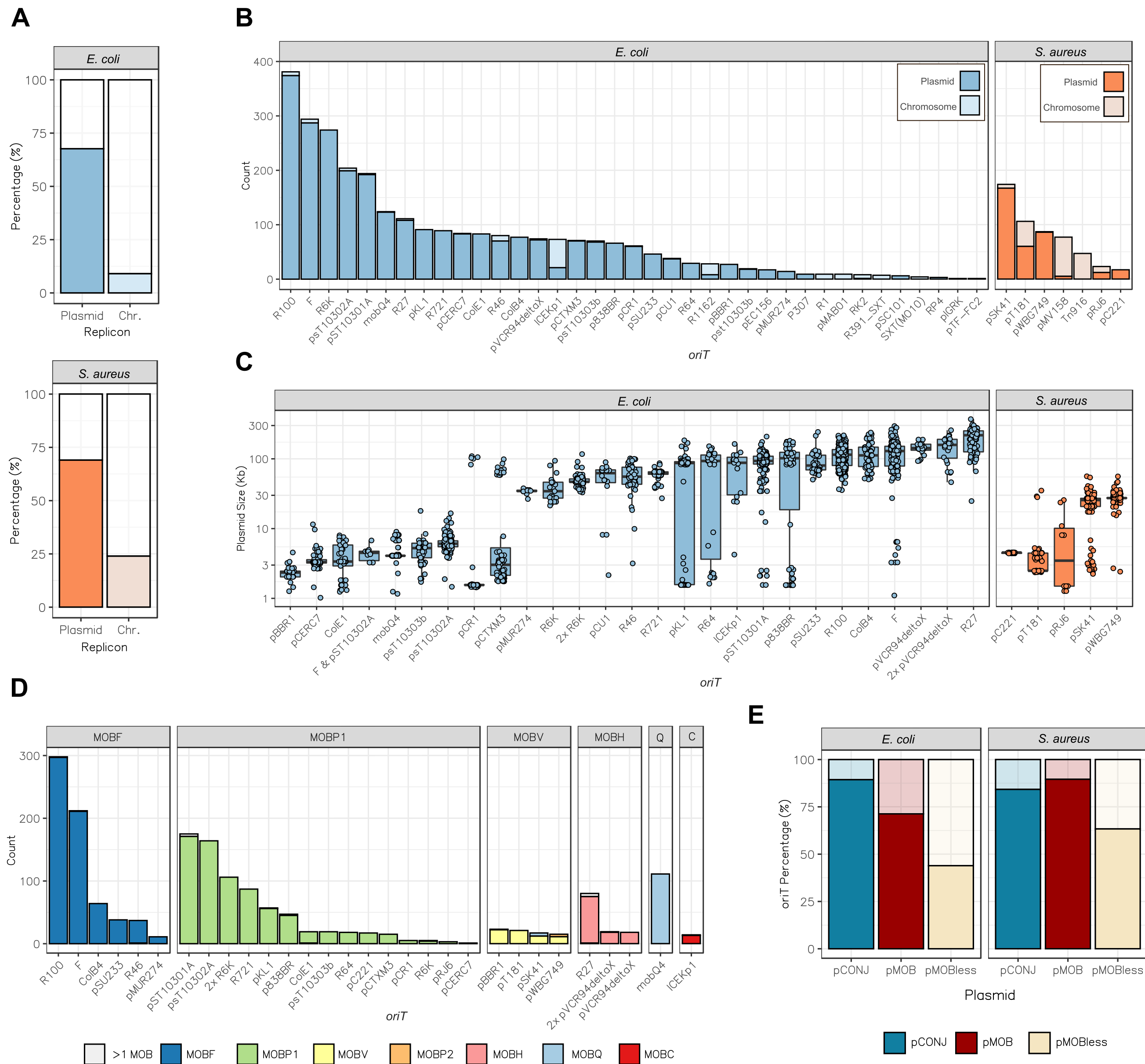


Figure 2. Identification of *oriT*s. (**A**) Percentage of plasmids and chromosomes with at least one *oriT* in *E. coli* (top) and *S. aureus* (down). (**B**) Counts of *oriT*s in the genomes of *E. coli* (left) and *S. aureus* (right). (**C**) Size of plasmids containing at least one *oriT* for families of *oriTs* present in at least 10 plasmids. (**D**) MOB families associated to the *oriT*s in (C). (**E**) Percentage of plasmids in which at least one *oriT* was identified, classed by mobility type.

smaller ones. Finally, the *oriT*s of a given family tend to be in plasmids with the same class of relaxases (Figure 2D). All things considered, although experimental validation would be needed to confirm their functionality, the identification of *oriT*s in most plasmids, usually in a single copy, the strict association between the *oriT* and the MOB, and their identification in plasmids of homogeneous size, suggest that most *oriT*s we identified are true positives.

oriT-MOBless plasmids are abundant and usual carriers of antimicrobial resistance genes

We identified at least one *oriT* in more than 80% of pCONJ and pMOB (Figure 2E). Hence, the *oriT*s in our collection are represented in a very large fraction of the *oriT*s used by the conjugative plasmids of these species. Importantly, we found an *oriT* in 790 pMOBless. Hereinafter, we will refer to these *oriT*-carrying pMOBless as pOriT. pOriTs constitute 65% of *S. aureus* plasmids lacking relaxases and more than 40% of those of *E. coli*. These results are subject to caution. We cannot ascertain the functionality of all these *oriT*, even if they are homologous to experimentally verified sequences. Also, we may have missed some *oriTs*, since a few pCONJ and pMOB lack identifiable *oriTs*. Despite these limitations, most plasmids have only one identifiable *oriT*, suggesting that we have identified most of them. If so, around half of the plasmids lacking relaxases are mobilizable by conjugation.

Downloaded from https://academic.oup.com/nar/article/51/7/3001/6849495 by guest on 30 September 2023

Due to the importance of *E. coli* and *S. aureus* as multidrug resistant pathogens (28), we inquired on the role of their different plasmids in the spread of antimicrobial resistance genes (ARG). Previous studies showed that conjugative plasmids tend to carry more ARGs than the other plasmids (29). This is the case of pCONJ in *E. coli* (∼64% of the genes) but not in *S. aureus*, where pOriTs carry most of these genes (∼76%) (Figure 3A). Furthermore, the number of ARGs per kilobase is highest in pOriT in both species (Figure 3B). The differences between plasmid types seem to be more important in terms of the number and density of ARGs than in the class of antibiotic resistance provided by the gene. Indeed, we found no obvious differences in the relative distribution of classes of ARGs between plasmids of different types (Supplementary Figure S8). Of note, the pOriT carry many ARG of clinical relevance, with high frequency of those conferring resistance for aminoglycosides and β-lactams, as is the case of the other plasmids. Interestingly, the plasmids with fewer ARGs, and lowest density, are those lacking both a relaxase and an *oriT* (presumably non-transmissible, pNT). These results show that plasmids lacking relaxases can be split in two categories, where those with an *oriT* have an important role in the spread of antibiotic resistance.

pOriTs exploit either conjugative or mobilizable plasmids

Many plasmids have *oriT*s from the same family. This information allows to study the functional dependencies between plasmids because we can link the pOriTs with sets of pMOBs or pCONJs. We have previously proposed that relaxases of pMOB evolve to interact with multiple types of MPF encoded in pCONJ, whereas those of pCONJ coevolve with the MPF to optimize their mutual interaction (49,50). As a consequence, one might expect that *oriT*s of pMOB and pCONJ would often be very different, in order to allow the respective relaxases to identify the plasmid with which to interact (26). In our dataset, many families of *oriT*s are present in either pCONJ or pMOB, but few are present in both (Figure 4A). The exceptions tend to correspond to 'pCONJ-like *oriT*s' (*oriTs* typical of pCONJ) that were found in large pMOB plasmids. We hypothesized that these might be decayed conjugative plasmids (pdCONJ) (36). These elements have some MPF genes, but not enough to be functional. Their analysis revealed that key genes, such as *virB4*, are often missing in pdCONJ (Supplementary Table S6), and seem to have derived from pCONJ by gene deletion (36). Hence, we split the pMOB into those encoding at least two MPF genes (pdCONJ) and the others. The pdCONJ are indeed 80% of the mobilizable plasmids with pCONJ-like *oriT*s. In contrast, pdCONJ do not have 'pMOB-like *oriT*s' (*oriTs* typical of pMOB) (Figure 4A). After this analysis, only three *oriT*s remained in a significant fraction of both pCONJ and pMOB (excluding pdCONJ): $oriT_{\mathrm{pKL1}}$, $oriT_{\mathrm{pWBG749}}$ and $oriT_{\mathrm{pSK41}}$. We then inquired on the possibility that ICEs or IMEs show similar trends, i.e. have specific *oriT*s. We found many *oriTs* in their chromosomes, but the precise *in silico* delimitation of ICEs and IMEs is technically challenging. Hence, we only analyzed if certain *oriT*s are present in chromosomes encoding an ICE, an IME or both. Our results showed that indeed, *oriT*s tend to be associated with either ICEs or IMEs (Supplementary Figure S9). We conclude that conjugative and mobilizable elements tend to use different *oriT*s.

The pOriT with a pCONJ-like *oriT* can recruit relaxases of a conjugative plasmid, in which case the relaxase will interact with the cognate MPF. In contrast, the pOriT with a pMOB-like *oriT* recruit relaxases of mobilizable plasmids, and these must then recruit a MPF from another plasmid. We know very little about the fitness costs of pOriT or pMOB on conjugative plasmids. But some mobilizable elements are known to strongly antagonize the cognate conjugative plasmids (51), which suggests competition for the conjugation pilus. Hence, mobilization of other elements may decrease the fitness of conjugative plasmids. If this is the case, then the pOriT with a pCONJ-like *oriT* is a parasite of the conjugative plasmid and the pOriT with a pMOB-like *oriT* is a parasite of pMOB and an hyper-parasite (a parasite of a parasite) of conjugative plasmids. One could expect that the most efficient strategy for a pOriT would be to take advantage of a unique conjugative plasmid rather than requiring on two other plasmids for transfer. However, since pMOB are often able to interact with multiple pCONJ, a pMOB-like *oriT* might allow a pOriT to have a higher chance of transfer under certain circumstances. Since the *oriT*s of pOriTs are homologous to those of conjugative or mobilizable elements (Figure 4B), we could infer the relations of dependence between pOriT and the other plasmids. We used *E. coli* data for this analysis because it includes much more diversity of *oriT*s for both pMOB and pCONJ than that of *S. aureus*. Interestingly, the frequency of pOriTs in *E. coli* with a pCONJ-like *oriT* (∼56%) or a pMOB-like one (35%) is very close to the relative frequency of each of these types of plasmids in the species (Figure 4C). Hence, the relative frequency of each type of pOriT matches the relative frequency of the hijacked plasmids.

pOriT may originate from both conjugative and mobilizable plasmids

Given the large number of pOriTs, we inquired on their evolutionary origin. It was recently suggested that some pMOBless derived from conjugative or mobilizable plasmids by gene deletion (36). Since pOriTs have either a pCONJ-like or a pMOB-like *oriT*, we thought they might have emerged by gene deletion in plasmids that lost the protein-coding genes associated with conjugation but kept an *oriT*. To evaluate this hypothesis, we grouped the 3,869 plasmids into Plasmid Taxonomic Units (PTUs) (27) and analyzed their mobility and *oriTs*. Most plasmids in a PTU have the same type of mobility, reflecting the short evolutionary distances between plasmids in the same PTU. Interestingly, the rare plasmids that have different types of mobility still tend to have *oriT*s of the same family (Supplementary Figure S10). This suggests that the *oriT* family is more conserved than the mobility type.

To test the possibility that some pOriTs originated from conjugative plasmids, we selected two PTUs and explored the relation between the pOriTs and pCONJ within a PTU. We analyzed the PTU-F_e (Figure 5) and the PTU-C (Supplementary Figure S11). Most of the plasmids in these PTUs are pCONJ with a pCONJ-like *oriT* (Figure 5C, D,

Downloaded from https://academic.oup.com/nar/article/51/7/3001/6849495 by guest on 30 September 2023

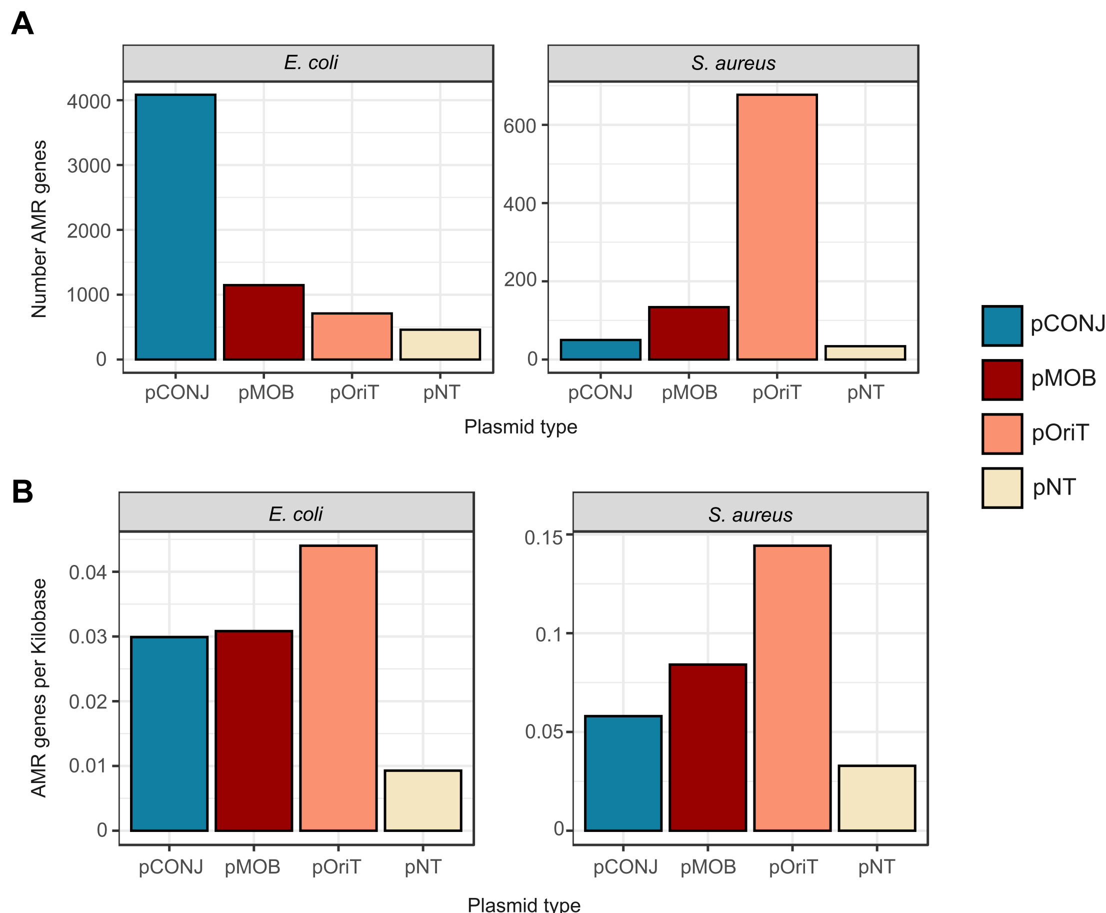


Figure 3. Plasmid types and antimicrobial resistance (AMR). (**A**) Number of AMR genes in plasmids of each type. (**B**) Density of AMR genes (genes per kilobase) in plasmids of each type.

Supplementary Figure S11B, C). Yet, both include a few plasmids with other mobility type that have *oriTs* of the same family but are significantly smaller (Figure 5B, Supplementary Figure S11A). This supports the idea that these replicons derived from conjugative plasmids by gene deletion. To further test this idea, we sought and selected pairs of pCONJ/pOriT within the PTUs having similar gene repertoires (wGRR > 0.75, see Materials and Methods). This analysis suggests that these pOriTs were generated by staggered degradation of the MPF system in pCONJ (see a representative example in Figure 5E, Supplementary Figure S11D). Crucially, the derived replicons are likely to be mobilized through *in-trans* conjugation because of the maintenance of their ancestral *oriT*.

We then selected two PTUs with a majority of pMOB (E1, E22) and analyzed them as above (Figure 6, Supplementary Figure S12). Both include ColE1-like plasmids, with a MOB_P (Figure 6C, Supplementary Figure S12B) and the pMOB-like family $oriT_{\text{ColE1-like}}$ (Figure 6D, Supplementary Figure S12C). As before, these PTUs include other types of plasmids, notably pOriTs and pNTs, which tend to have similar *oriTs* and smaller sizes (Figure 6B, Supplementary Figure S12A). The alignment of the closely related pMOB/pOriT pairs further suggest that small pOriTs arise by the loss of the relaxase in an ancestral pMOB (see a representative example in Figure 6E, Supplementary Figure S12D). Interestingly, we identified a subgroup of plasmids of the PTU-E1 that has another family of *oriTs* ($oriT_{\text{pCERC7}}$). This origin of transfer is related to the $oriT_{\text{R64}}$ of conjugative plasmids (52) (Figure 6D, Supplementary Figure S4). This finding suggests that recombination events may result in the exchange of the *oriT* of the plasmid. Overall, these results show at the micro-evolutionary scale how pOriTs can derive by gene deletion from other types of plasmids.

Most plasmids may be mobilized by known mechanisms of transfer

Our results suggest that ∼80% of *E. coli* and >70% *S. aureus* plasmids use an *oriT* to transfer by conjugation. To this, one may add other genetic elements that spur plasmid transfer (Figure 7A). Notably, some rolling-circle replication proteins (RC-Rep) act as replicative relaxases (37). They interact with the MPF system of a conjugative element and trigger plasmid conjugation in an *oriT*-independent manner (53). We searched for these proteins to test if this alternative pathway could be involved in the mobilization of plasmids

Downloaded from https://academic.oup.com/nar/article/51/7/3001/6849495 by guest on 30 September 2023

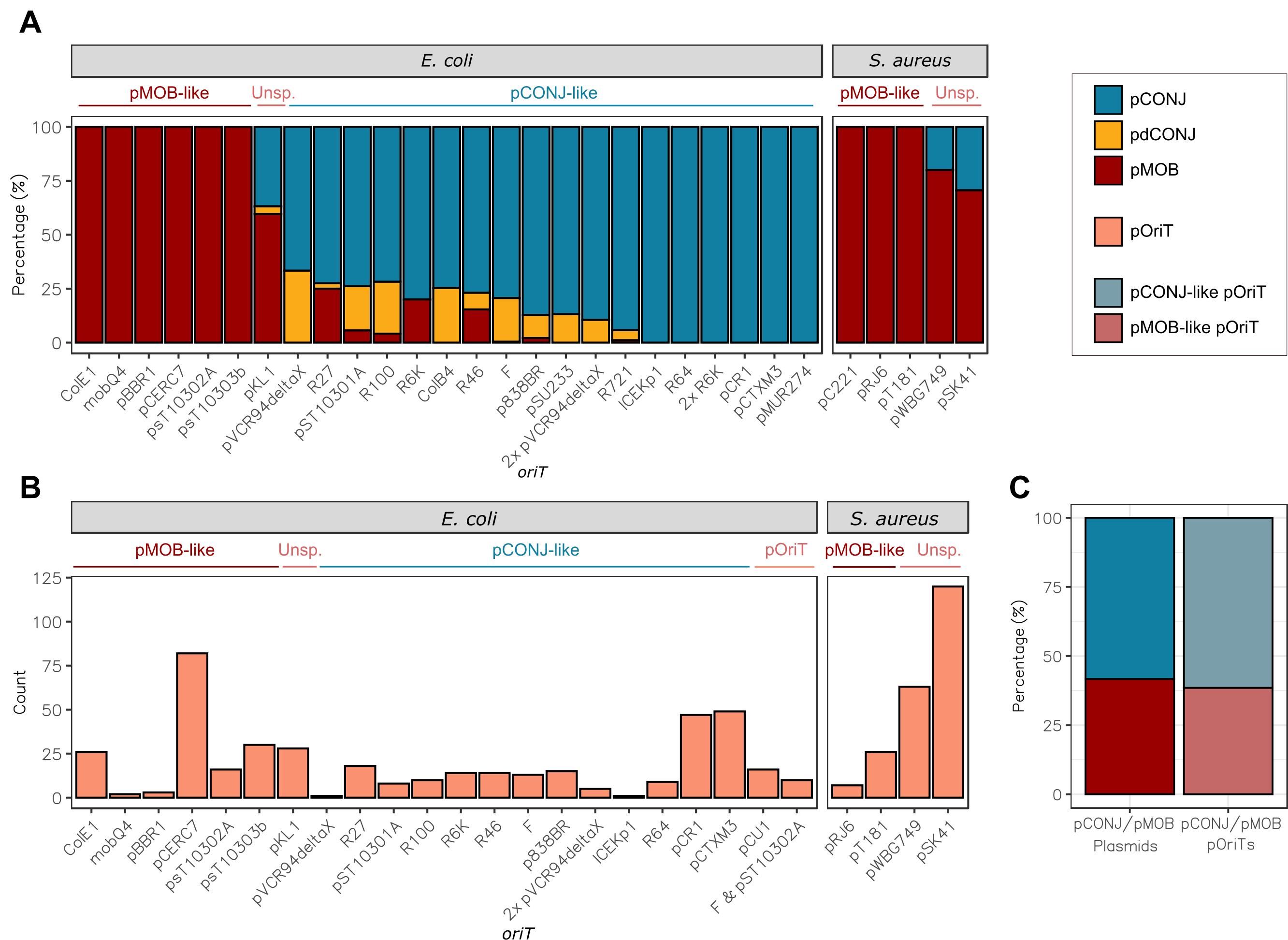


Figure 4. (**A**) Percentage of plasmid types with a given *oriT* (for *oriTs* occurring in >10 plasmids). (**B**) Number of pOriTs (*oriT*-encoding MOBless plasmids) found for each *oriT*. pCONJ-like: *oriT*s identified mostly (>75%) in conjugative plasmids; pMOB-like: *oriTs* identified mostly (>75%) in mobilizable plasmids; Unsp.: *oriTs* identified in many conjugative and mobilizable plasmids; pOriT: *oriT*s identified only in pOriTs. (**C**) Ratio of pCONJ/pMOB plasmids (left) and ratio of pOriTs with pCONJ-like/pMOB-like *oriT*s (right) in *E. coli*.

lacking *oriT* and classical relaxases. We identified 225 homologs of RC-Rep proteins in 208 plasmids. These plasmids are frequent in *S. aureus* (∼30%), but rare in *E. coli* (1.9%). As expected, there is an overrepresentation of RC-Rep in non-*oriT* pMOBless ($\chi^2_{(4)} = 103.12$, $P < 2.2e{-}16$) (Supplementary Figure S13). The unexpected abundance of RC-Rep in plasmids lacking an *oriT* suggests that such proteins could mediate the mobility of many plasmids in *S. aureus*.

Some plasmids can be transferred within viral particles. The propensity of a plasmid to be transduced cannot be predicted from its sequence. But ca. 6% of the plasmids are also phages (P-Ps) (7), and encode viral particles, virion assembly packaging, and cell lysis (54). Phage-plasmids can be identified from the plasmid sequences by searching for these genes. We identified 222 P-Ps in *E. coli* and 1 in *S. aureus,* which is consistent with the reported uneven distribution of P-Ps across bacteria (7). P-Ps correspond to a third of the pMOBless without *oriT* in *E. coli* (n = 216/702). In agreement with the idea that P-Ps provide an alternative mechanism of plasmid transfer, only six P-Ps encode conjugation-related elements (Supplementary Figure S14). The latter are much larger (∼175 kb) than the remaining P-Ps (∼90 kb), and might be the result of co-integration events or assembly artifacts (Supplementary Figure S14).

At the end of these analyses, we could assign a putative mechanism of mobility for most plasmids in each species. In *E. coli*, 80% of the plasmids were classed as conjugative or mobilizable by conjugation, and ∼7% as P-Ps. In *S. aureus*, 90% were classed as conjugative or mobilizable by some type of conjugation and only 1 is a P-P. Hence, when one accounts for MPF, relaxases, RC-Rep, *oriT*, and P-Ps, few plasmids lack a hypothetical mechanism of transfer, i.e. few remain putatively non-transmissible (pNT) (Figure 7A): 13.7% in *E. coli* and 10.4% is *S. aureus*. We inquired on the possible mechanisms of mobility of the remaining plasmids. Around 50% of the *E. coli* pNTs are related to the large plasmid pO157 (PTU-E5) (Supplementary Figure S15). These are well-known non-transmissible plasmids that have disseminated in *E. coli* O157:H7 (55). The mechanisms of mobility of the few remaining plasmids (if any) remains unknown.

The distribution of the size of plasmids is bi-modal and associated with their type of mobility (Figure 7B). The mode associated with the largest plasmids is characteris-

Downloaded from https://academic.oup.com/nar/article/51/7/3001/6849495 by guest on 30 September 2023

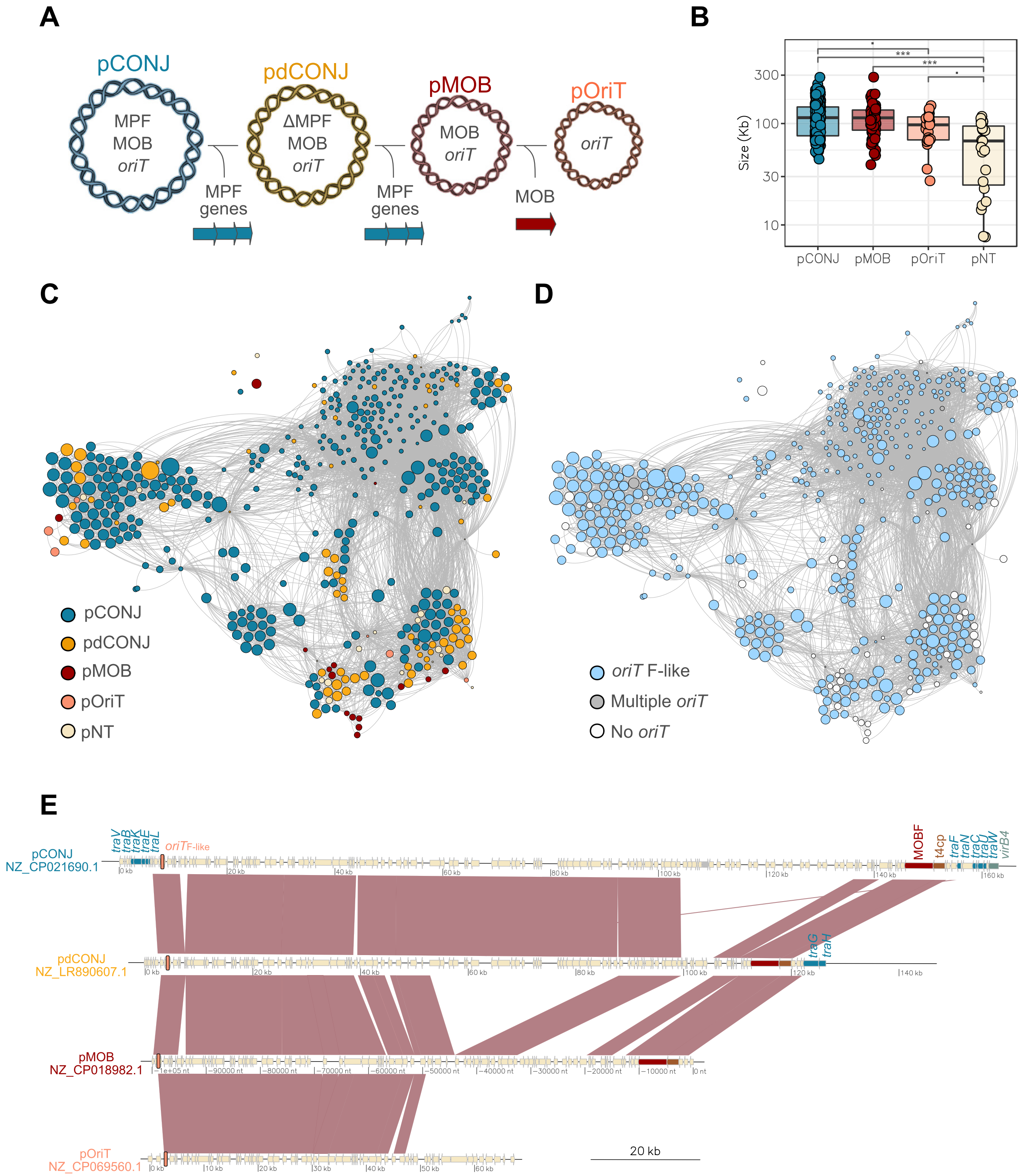


Figure 5. Evolution of pCONJ-like pOriTs. (**A**) Proposed evolutionary hypothesis for the origin of pCONJ-like pOriTs. (**B**) Size of plasmids of the PTU-Fe (IncF/MOB_F/MPF_F) according to their mobility type. The horizontal bars over the plot denote statistically significant difference (pairwise *t*-tests): $***P < 0.001$, $**P < 0.01$, $*P < 0.05$, $\cdot P < 0.1$. (**C** and **D**) Graphs showing the plasmids of the PTU-Fe. Nodes (circles) represent the plasmids and edges (grey lines) connect highly similar plasmids (wGRR > 0.75). The colors of the nodes represent the plasmid mobility type (**C**) and the plasmid *oriT* family (**D**). (**E**) Multiple alignment of a pCONJ, pdCONJ, pMOB and pOriT from the PTU-Fe. Brown shadings between sequences denote >80% identity between the sequences. Conjugative genes are represented by blue arrows, the relaxase is red, the coupling protein in brown, *virB4* is in green and the *oriT* is represented as an orange rectangle.

Downloaded from https://academic.oup.com/nar/article/51/7/3001/6849495 by guest on 30 September 2023

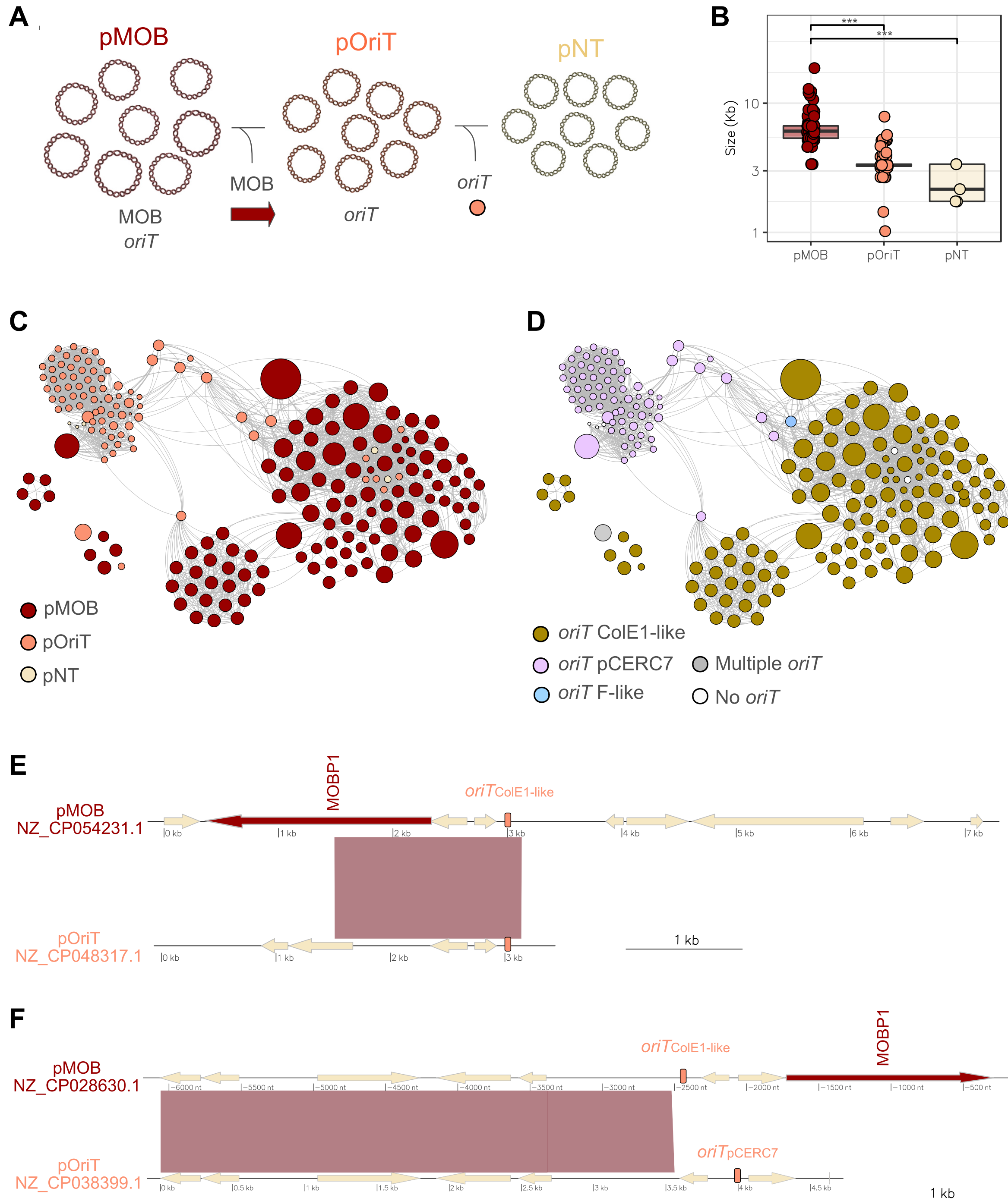


Figure 6. Evolution of pMOB-like pOriTs. (**A**) Proposed evolutionary hypothesis for the origin of pMOB-like pOriTs. (**B**) Size of plasmids of the PTU-E1 (ColRNAI/Col440I/MOB_P) according to their mobility type. The horizontal bars over the plot denote statistically significant difference (pairwise *t*-tests): $***P < 0.001$, $**P < 0.01$, $*P < 0.05$, $\cdot P < 0.1$. (**C** and **D**) Graphs showing the plasmids of the PTU-E1. Nodes (circles) represent the plasmids and edges (grey lines) connect highly similar plasmids (wGRR > 0.75). The colors of the nodes represent the plasmid mobility type (**C**) and the plasmid *oriT* family (**D**). (**E** and **F**) Alignments of pMOB and pOriT from the PTU-E1. Brown shadings between sequences denote > 80% identity between the sequences. The relaxase is indicated as a red arrow and the *oriT* as an orange rectangle.

Downloaded from https://academic.oup.com/nar/article/51/7/3001/6849495 by guest on 30 September 2023

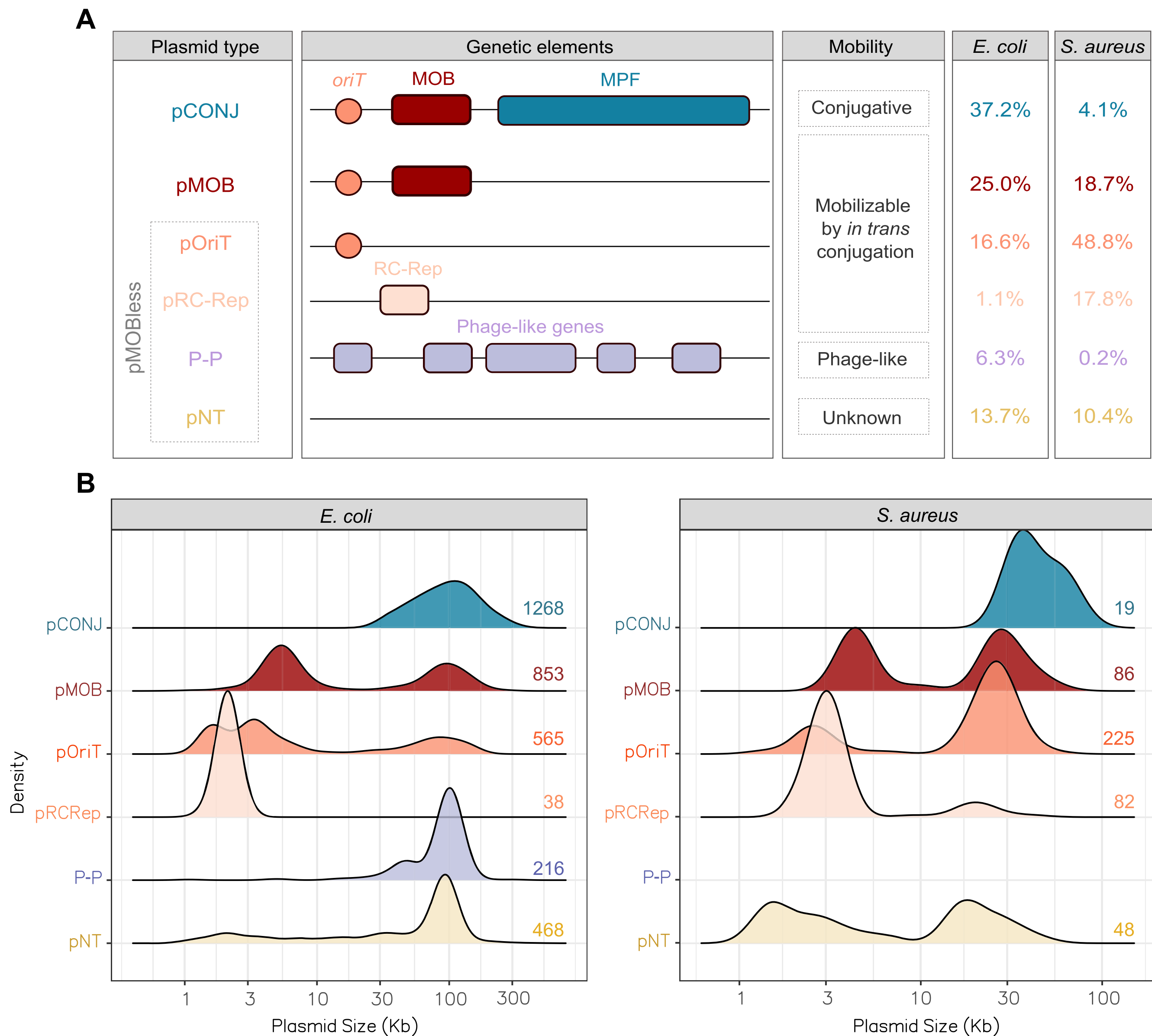


Figure 7. Classification of plasmid mobility. (**A**) Representation of plasmids in function of their category, genetic composition, and mechanism of mobility. The frequency (%) of each plasmid type in *E. coli* and *S. aureus*, respectively, is shown at the right columns of the figure. (**B**) Distribution of plasmid sizes attending to the mobility type. The curves were drawn using a scaled kernel density to simplify the representation (sample sizes at the right of each row). The size distribution of P-Ps is shown in the Supplementary Figure S9.

tic of pCONJ, but also found among certain pMOB and pOriT in both species. For the latter, we observed a shift of the peak to lower values of plasmid size. Similarly, the mode of the smaller plasmids is characteristically associated with pMOB, but is also found among pOriT, with a shift of the peak to lower values of plasmid size. These small downwards shifts observed among pOriT are consistent with our hypothesis that they often originate from pCONJ or pMOB by gene deletion (Supplementary Figure S16). The patterns for pNT are less clear. In *E. coli* they are shaped by the many large pO157-like plasmids, whereas in *S. aureus* they seem to follow the trends of pOriT, suggesting that maybe some *oriT* remain to be uncovered in the species.

Mobilization explains patterns of plasmid co-existence

The dependence of certain plasmids, e.g. pOriT, on others, notably pCONJ, for conjugative transfer means that the type of mobility of plasmids may affect the patterns of their co-occurrence in cells. We can now test this hypothesis by analyzing which plasmids tend to co-occur. The number of plasmids per genome is much more variable (and on average higher) in *E. coli* than in *S. aureus*. Hence, we concentrated on the *E. coli* data for this analysis. We identified the most common patterns of occurrence among the 1207 plasmid-bearing *E. coli* genomes, focusing on pCONJ, pMOB, pOriT and pNT (Figure 8A). The most common pattern is the presence of only conjugative plas-

Downloaded from https://academic.oup.com/nar/article/51/7/3001/6849495 by guest on 30 September 2023

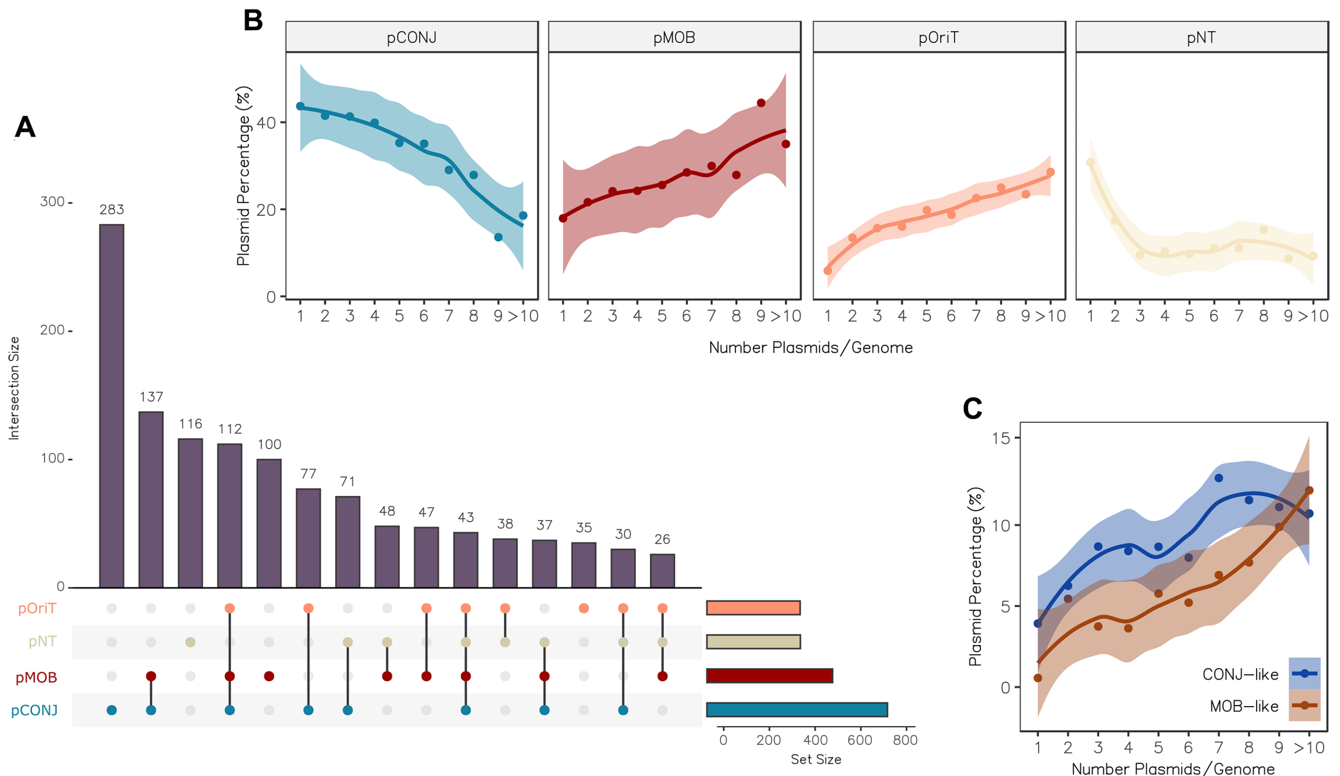


Figure 8. (**A**) Upset plot showing the frequency of co-occurrences of pCONJ, pMOB, pOriT and pNT. (**B**) Percentage of plasmid types in genomes attending to the number of different plasmids present in the genome. (**C**) Percentage of pCONJ-like and pMOB-like pOriTs attending to the number of plasmids in their hosts' genomes.

mids in the cell. The second and fourth most frequent patterns are a pair of pCONJ-pMOB and the triplet pCONJ-pMOB-pOriT. Interestingly, the third most frequent pattern is the single presence of MOBless pNTs, in contrast to the much rarer event of having single pOriTs in the cell. This further reinforces the idea that while MOBless pNTs are non-transmissible and vertically transmitted with their host cells, pOriTs co-transfer with other elements in the cell.

If the pMOB and pOriT require a pCONJ to transfer between cells, one would expect that the frequency of each type of plasmids would vary with the number of plasmids per genome. Notably, genomes with few plasmids would tend to have more pCONJ and those with many plasmids would have progressively a larger fraction of other types of plasmids. Indeed, the frequency of pCONJ in *E. coli* is highest in genomes with a single plasmid and constantly decreases with the increase in the number of plasmids (Figure 8B). As expected, pMOB and pOriT show the inverse trend. These plasmids are rarely found alone in the genome and become increasingly frequent when cells contain more and more plasmids. The frequency of these plasmids is very high (almost 70%) in genomes with more than 10 different plasmids. Hence, the relative frequency of plasmids of a given mobility type varies in a predictable way with the number of plasmids in the cell.

We showed above that some pOriTs may only require a pCONJ (since they have a pCONJ-like *oriT*), whereas others may require a pCONJ and a pMOB to transfer (pMOB-like *oriT*). The latter might be found preferentially in genomes with more plasmids, since they require a combination of two compatible plasmids to transfer. Indeed, while pCONJ-like pOriTs reach a frequency plateau in genomes with ≥7 plasmids, pMOB-like pOriTs increase steeply in frequency up to 10 plasmids/genome (Figure 8C). All these findings suggest that functional dependencies between plasmids shape the co-occurrence of plasmids in cells.

DISCUSSION

To understand how plasmids lacking relaxases could be transferred between bacteria, we searched for homologs of experimentally verified *oriT*, the only genetic element a plasmid needs *in-cis* for conjugation. The search for homologs of *oriT*s could result in misidentifications, but our observations suggest that most of the identified *oriT*s are correct. (i) While most plasmids have an *oriT*, most chromosomes lack them, in spite of their much longer sequences. (ii) At least one *oriT* has been identified in most plasmids that were expected to have it (pMOB or pCONJ). (iii) There are no cross matches between *E. coli* and *S. aureus oriT*s. (iv) There are almost no cross matches between pCONJs/pdCONJs and pMOBs, allowing to identify pCONJ-like and pMOB-like *oriT*s. (v) Most plasmids have one single *oriT*, and the others often have multiple relaxases, seem to be plasmid co-integrates, or have been already described (56). (vi) Almost all *oriT*s identified are located in non-coding regions. (vii) There is a strict association between the *oriT*s and their associated relaxase family. (viii) The *oriT*s were not

Downloaded from https://academic.oup.com/nar/article/51/7/3001/6849495 by guest on 30 September 2023

found where they were not expected, e.g. in phage-plasmids that rely on alternative mechanisms rather than conjugation (45), or in pO157-like plasmids, which are known to be non-conjugative (55). Finally, previous work in *S. aureus* validated the identification of *oriT*s in plasmids (25). The results (i), (ii), (v) and (vi) suggest that we have identified *oriT*s with high sensitivity, i.e. we found most plasmids with a given type of *oriT*. The results (i), (iii), (iv), (vi), (vii) and (viii) suggest that we have identified them with high specificity, i.e. we have few false positives. Hence, our *oriT* screening seems accurate.

Despite the previous results, we believe that some *oriT*s remain to be identified in these two species because we found that some pCONJ and pMOB lack known *oriT*s (Figure 2E, Supplementary Figure S17). This could result from occasional *oriT* deletion. Yet, we found a small number of PTUs of conjugative (e.g. PTU X4) and mobilizable plasmids (e.g. PTU-E7) lacking identifiable *oriTs* in most elements (Supplementary Figure S17). This strongly suggests the existence of still uncharacterized origins of transfer in the plasmids of these widely studied species. The number of unknown *oriTs* seems much larger in elements integrated in the chromosome. For example, we identified relaxases in ∼30% of *E. coli* chromosomes, but *oriT*s in only 9%. Further work on integrative elements will require identification of novel *oriTs* and development of methods to accurately delimit conjugative and mobilizable elements in chromosomes. We are actively working on both of these aspects.

The observation that pOriTs usually have *oriT*s from either pCONJ or pMOB, suggests that these elements have evolved to either hijack the relaxase of a conjugative or that of a mobilizable plasmid. The latter results in a more complex succession of ecological dependencies since transfer of the plasmid needs the presence in the cell of an additional pCONJ (see below). In natural environments, these two types of pOriT could have arisen by gene deletion of the conjugation protein coding genes in pCONJ (first) or pMOB (second), while the ancestral *oriT* was kept. This is consistent with the observation that novel pOriTs may emerge within PTUs (which are sets of very closely related plasmids). More complex scenarios are also possible, e.g. the translocation of an *oriT* to a plasmid lacking one. The hypothesis of frequent pOriT genesis by gene deletion from pMOB or pCONJ is further supported by the analysis of the distribution of pOriT size (Figure 7B, Supplementary Figure S16). This distribution has two peaks, each at values of plasmid size that are slightly smaller than the corresponding peaks for pMOB and pCONJ. We have proposed that a fraction of pMOB derived recently from pCONJ (36). Our present results further suggest that a part of pOriT originated from either pCONJ or pMOB.

Why would some plasmids evolve towards less autonomous mobilization, i.e. to depend on other plasmids for mobility? One might expect that such evolutionary processes would be counter-selected because they make plasmids dependent on the presence of other plasmids for horizontal transmission. We speculate that pOriT may be advantageous in terms of carriage cost. The *oriT* is a small non-coding sequence that may have little impact on bacterial fitness. In contrast, MPF systems and relaxases are costly and may hamper the successful vertical transmission of the plasmid (57–59). This is why the genetic components of conjugative plasmids are usually repressed (60) and occasionally lost (61). Hence, the loss of protein-coding genes for conjugation may decrease horizontal transfer but increase the success of vertical transmission. In contrast, the loss of *oriT*s precludes horizontal transmission by conjugation without providing significant advantages for vertical transmission. Hence, the conditions that favor loss of conjugation-related protein coding genes may not favor the loss of *oriT*.

The transfer of pOriTs depends on their co-occurrence in cells with plasmids whose relaxases and MPF are compatible with the *oriT* of the plasmid. We observed that the frequency of pOriT with pCONJ-like and pMOB-like *oriT*s was in direct proportion of the frequency of the ‘helper’ plasmids. The dependence of pOriT on the presence of other plasmids in the cell might suggest that pOriTs should evolve to have a pCONJ-like *oriT* and dispense the requirement for a pMOB. This is not what we observed. Instead, the two types of pOriT seem frequent. It was previously known that pMOBs are frequent and can in some cases be mobilized by many different pCONJ (49,50). We speculate that pOriT with pMOB-like *oriT*s have an advantage in certain cases over those with pCONJ-like *oriT*s because pMOBs may hijack many different pCONJ. Genomes with many plasmids might thus often have the right combinations of pMOB/pCONJ allowing the transfer of the pOriT. Furthermore, if the mobilization of a pOriT and/or pMOB entails the co-transfer of the helper pCONJ as it has been suggested (62), the pOriT will find in this novel host cell all the plasmids that are required for its subsequent mobility.

Possibly because of processes of co-mobilization, plasmid mobility type seems to shape plasmid distribution within cells. Large and small plasmids were previously found to co-occur more often than expected in bacteria (63). Since large plasmids are often pCONJ and smaller ones are typically pMOB or pOriT, this fits our observations of co-occurrence of the different types of plasmids. Interestingly, pMOBs and pOriTs are particularly abundant in genomes bearing many plasmids, where the chances to find helper pCONJ are high. In contrast, pCONJ, which conjugate autonomously, are the most common plasmids in cells having one or a few elements. The simplest mechanism to explain these results is that these plasmids often arrive at the cell together, i.e. using the same mating event. But additional interactions may also contribute to further stabilize the presence of these plasmids in cells. For example, the cost of carrying small plasmids was smaller in a *Pseudomonas* strain already carrying a large plasmid (63).

Our results suggest that most plasmids can conjugate either autonomously or by recruiting the required functions from other plasmids. Notably, around 90% of *S. aureus* plasmids have the genetic elements needed for horizontal transfer by conjugation. Notwithstanding, alternative mechanisms of plasmid mobility have been recently described. Among *E. coli* plasmids, 7% are phage-plasmids that can transfer within viral particles. In *S. aureus*, phage-plasmids are rare, but plasmids can be transduced by phages and their satellites (64). This creates constraints in the size of plasmids because phages and satellites transduce pieces of DNA with the length of their own genomes. Interestingly, the size of the

Downloaded from https://academic.oup.com/nar/article/51/7/3001/6849495 by guest on 30 September 2023

genomes of temperate phages matches one of the two peaks in the distribution of sizes of pMOBless and the size of the satellite genomes matches the other peak. It was proposed that plasmids were selected to have sizes compatible with transduction by phages and satellites, which explains the bi-modal distribution of plasmid sizes (Figure 7B). If correct, transduction by phages and their satellites would explain the enigmatic bi-modality of plasmid sizes, while gene deletions causing the transitions between pCONJ or pMOB to pOriT would explain why the latter tend to follow the size distribution of the former.

In summary, 9 out of 10 plasmids have identifiable genetic elements that facilitate their horizontal transfer, most of them involving conjugation. There are only ∼10% plasmids lacking known genetic elements associated with horizontal transfer. Such plasmids may still occasionally be transferred through alternative mechanisms leaving little trace in the plasmid sequence, such as transformation or transduction. With this work, we provide strong evidence suggesting that there is no conundrum of plasmid mobility. Most plasmids are mobile. They just differ in the mechanism and in their degree of autonomy for transfer.

DATA AVAILABILITY

The databases used in this work are publicly available (RefSeq database) or indicated within the text.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Eugen Pfeifer for providing the wGRR and PTUs data. Fernando de la Cruz and Maria Pilar Garcillán Barcia for discussion along the years on plasmid mobility. Microbial Evolutionary Genomics Unit for scientific discussions.

FUNDING

INCEPTION project [PIA/ANR-16-CONV-0005]; Fédération pour la Recherche Médicale [Equipe FRM/EQU201903007835]; Labex IBEID [ANR-10-LABX-62-IBEID]; HORIZON-MSCA-2021-PF-01–01 EvoPlas-101062386 (to M.A.-A.). Funding for open access charge: Institut Pasteur.

Conflict of interest statement. None declared.

REFERENCES

1. Treangen,T.J. and Rocha,E.P.C. (2011) Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet.*, **7**, e1001284.
2. Frost,L.S., Leplae,R., Summers,A.O. and Toussaint,A. (2005) Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.*, **3**, 722–732.
3. Wein,T. and Dagan,T. (2020) Plasmid evolution. *Curr. Biol.*, **30**, R1158–R1163.
4. Zhang,X., Cai,X., Qi,Y., Liu,Y., Cao,Q., Wang,X., Chen,H. and Xu,X. (2018) Improvement in the efficiency of natural transformation of *Haemophilus parasuis* by shuttle-plasmid methylation. *Plasmid*, **98**, 8–14.
5. Erdmann,S., Tschitschko,B., Zhong,L., Raftery,M.J. and Cavicchioli,R. (2017) A plasmid from an antarctic haloarchaeon uses specialized membrane vesicles to disseminate and infect plasmid-free cells. *Nat. Microbiol.*, **2**, 1446–1455.
6. Canosi,U., Lüder,G. and Trautner,T.A. (1982) SPP1-mediated plasmid transduction. *J. Virol.*, **44**, 431–436.
7. Pfeifer,E., Moura de Sousa,J.A., Touchon,M. and Rocha,E.P.C. (2021) Bacteria have numerous distinctive groups of phage–plasmids with conserved phage and variable plasmid gene repertoires. *Nucleic Acids Res.*, **49**, 2655–2673.
8. Smillie,C., Garcillán-Barcia,M.P., Francia,M.V., Rocha,E.P.C. and de la Cruz,F. (2010) Mobility of plasmids. *Microbiol. Mol. Biol. Rev.*, **74**, 434–452.
9. De La Cruz,F., Frost,L.S., Meyer,R.J. and Zechner,E.L. (2010) Conjugative DNA metabolism in Gram-negative bacteria. *FEMS Microbiol. Rev.*, **34**, 18–40.
10. Guglielmini,J., de la Cruz,F. and Rocha,E.P.C. (2013) Evolution of conjugation and type IV secretion systems. *Mol. Biol. Evol.*, **30**, 315–331.
11. Johnson,C.M. and Grossman,A.D. (2015) Integrative and conjugative elements (ICEs): what they do and how they work. *Annu. Rev. Genet.*, **49**, 577–601.
12. Cury,J., Oliveira,P.H., de la Cruz,F. and Rocha,E.P.C. (2018) Host range and genetic plasticity explain the coexistence of integrative and extrachromosomal mobile genetic elements. *Mol. Biol. Evol.*, **35**, 2230–2239.
13. Branger,C., Ledda,A., Billard-Pomares,T., Doublet,B., Barbe,V., Roche,D., Médigue,C., Arlet,G. and Denamur,E. (2019) Specialization of small non-conjugative plasmids in *Escherichia coli* according to their family types. *Microbial Genomics*, **5**, e000281.
14. Gu,D., Dong,N., Zheng,Z., Lin,D., Huang,M., Wang,L., Chan,E.W., Shu,L., Yu,J., Zhang,R. *et al.* (2018) A fatal outbreak of ST11 carbapenem-resistant hypervirulent *Klebsiella pneumoniae* in a chinese hospital: a molecular epidemiological study. *Lancet Infect. Dis.*, **18**, 37–46.
15. Maeda,S., Ito,M., Ando,T., Ishimoto,Y., Fujisawa,Y., Takahashi,H., Matsuda,A., Sawamura,A. and Kato,S. (2006) Horizontal transfer of nonconjugative plasmids in a colony biofilm of *Escherichia coli*. *FEMS Microbiol. Lett.*, **255**, 115–120.
16. Lambert,C.M., Hyde,H. and Strike,P. (1987) Conjugal mobility of the multicopy plasmids NTP1 and NTP16. *Plasmid*, **18**, 99–110.
17. Chang,A.C. and Cohen,S.N. (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol.*, **134**, 1141–1156.
18. Lee,C.A., Thomas,J. and Grossman,A.D. (2012) The *Bacillus subtilis* conjugative transposon ICE *Bs1* mobilizes plasmids lacking dedicated mobilization functions. *J. Bacteriol.*, **194**, 3165–3172.
19. Xie,M., Chen,K., Ye,L., Yang,X., Xu,Q., Yang,C., Dong,N., Chan,E.W., Sun,Q., Shu,L. *et al.* (2020) Conjugation of virulence plasmid in clinical *Klebsiella pneumoniae* strains through formation of a fusion plasmid. *Adv. Biosyst.*, **4**, e1900239.
20. Daccord,A., Ceccarelli,D. and Burrus,V. (2010) Integrating conjugative elements of the SXT/R391 family trigger the excision and drive the mobilization of a new class of vibrio genomic islands: ICE-mediated GI mobilization. *Mol. Microbiol.*, **78**, 576–588.
21. Brockhurst,M.A. and Harrison,E. (2022) Ecological and evolutionary solutions to the plasmid paradox. *Trends Microbiol.*, **30**, 534–543.
22. Hall,J.P.J., Wood,A.J., Harrison,E. and Brockhurst,M.A. (2016) Source–sink plasmid transfer dynamics maintain gene mobility in soil bacterial communities. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 8260–8265.
23. O'Brien,F.G., Yui Eto,K., Murphy,R.J., Fairhurst,H.M., Coombs,G.W., Grubb,W.B. and Ramsay,J.P. (2015) Origin-of-transfer sequences facilitate mobilisation of non-conjugative antimicrobial-resistance plasmids in *Staphylococcusaureus*. *Nucleic Acids Res.*, **43**, 7971–7983.
24. Pollet,R.M., Ingle,J.D., Hymes,J.P., Eakes,T.C., Eto,K.Y., Kwong,S.M., Ramsay,J.P., Firth,N. and Redinbo,M.R. (2016). Processing of nonconjugative resistance plasmids by conjugation nicking enzyme of staphylococci. *J. Bacteriol.*, **198**, 888–897.
25. Ramsay,J.P., Kwong,S.M., Murphy,R.J., Yui Eto,K., Price,K.J., Nguyen,Q.T., O'Brien,F.G., Grubb,W.B., Coombs,G.W. and Firth,N. (2016) An updated view of plasmid conjugation and mobilization in *Staphylococcus*. *Mob. Genet. Elements*, **6**, e1208317.

Downloaded from https://academic.oup.com/nar/article/51/7/3001/6849495 by guest on 30 September 2023

26. Ramsay,J.P. and Firth,N. (2017) Diverse mobilization strategies facilitate transfer of non-conjugative mobile genetic elements. *Curr. Opin. Microbiol.*, **38**, 1–9.
27. Redondo-Salvo,S., Fernández-López,R., Ruiz,R., Vielva,L., de Toro,M., Rocha,E.P.C., Garcillán-Barcia,M.P. and de la Cruz,F. (2020) Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids. *Nat. Commun.*, **11**, 3602.
28. Antimicrobial Resistance Collaborators (2022) Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet*, **12**, 629–655.
29. Partridge,S.R., Kwong,S.M., Firth,N. and Jensen,S.O. (2018) Mobile genetic elements associated with antimicrobial resistance. *Clin. Microbiol. Rev.*, **31**, e00088-17.
30. Li,X., Xie,Y., Liu,M., Tai,C., Sun,J., Deng,Z. and Ou,H.Y. (2018) oriTfinder: a web-based tool for the identification of origin of transfers in DNA sequences of bacterial mobile genetic elements. *Nucleic Acids Res.*, **46**, W229–W234.
31. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.
32. Cury,J., Abby,S.S., Doppelt-Azeroual,O., Néron,B. and Rocha,E.P.C. (2020) Identifying Conjugative Plasmids and Integrative Conjugative Elements with CONJscan. *Methods Mol. Biol.*, **2075**, 265–283.
33. Wheeler,T.J. and Eddy,S.R. (2013) nhmmer: DNA homology search with profile HMMs. *Bioinformatics*, **29**, 2487–2489.
34. Garcillán-Barcia,M.P., Redondo-Salvo,S., Vielva,L. and Cruz,F. (2020) MOBscan: automated annotation of MOB relaxases. *Methods Mol. Biol.*, **2075**, 295–308.
35. Hyatt,D., Chen,G.L., Locascio,P.F., Land,M.L., Larimer,F.W. and Hauser,L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, **11**, 119.
36. Coluzzi,C., Garcillán-Barcia,M.P., de la Cruz,F. and Rocha,E.P.C. (2022) Evolution of plasmid mobility: origin and fate of conjugative and nonconjugative plasmids. *Mol. Biol. Evolut.*, **39**, msac115.
37. Lee,C.A., Thomas,J. and Grossman,A.D. (2012) The *Bacillus subtilis* conjugative transposon ICEBs1 mobilizes plasmids lacking dedicated mobilization functions. *J. Bacteriol.*, **194**, 3165–3172.
38. Mistry,J., Chuguransky,S., Williams,L., Qureshi,M., Salazar,G.A., Sonnhammer,E.L.L., Tosatto,S.C.E., Paladin,L., Raj,S., Richardson,L.J. *et al.* (2021) Pfam: the protein families database in 2021. *Nucleic Acids Res.*, **49**, D412–D419.
39. Pfeifer,E., Bonnin,R. and Rocha,E.P.C. (2022) Phage-Plasmids spread antibiotic resistance genes through infection and lysogenic conversion. *Mbio*, **26**, e0185122.
40. Perrin,A. and Rocha,E.P.C. (2021) PanACoTA: a modular tool for massive microbial comparative genomics. *NAR Genom. Bioinform.*, **3**, lqaa106.
41. Touchon,M., Perrin,A., de Sousa,J.A.M., Vangchhia,B., Burn,S., O'Brien,C.L., Denamur,E., Gordon,D. and Rocha,E.P.C. (2020) Phylogenetic background and habitat drive the genetic diversification of *Escherichia coli*. *PLoS Genet.*, **16**, e1008866.
42. Redondo-Salvo,S., Bartomeus-Peñalver,R., Vielva,L., Tagg,K.A., Webb,H.E., Fernández-López,R. and de la Cruz,F. (2021) COPLA, a taxonomic classifier of plasmids. *BMC Bioinformatics*, **22**, 390.
43. Cury,J., Touchon,M. and Rocha,E.P.C. (2017) Integrative and conjugative elements and their hosts: composition, distribution and organization. *Nucleic Acids Res.*, **45**, 8943–8956.
44. Steinegger,M. and Söding,J. (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.*, **35**, 1026–1028.
45. Feldgarden,M., Brover,V., Gonzalez-Escalona,N., Frye,J.G., Haendiges,J., Haft,D.H., Hoffmann,M., Pettengill,J.B., Prasad,A.B., Tillman,G.E. *et al.* (2021) AMRFinderPlus and the reference gene catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci. Rep.*, **11**, 12728.
46. Alcock,B.P., Raphenya,A.R., Lau,T.T.Y., Tsang,K.K., Bouchard,M., Edalatmand,A., Huynh,W., Nguyen,A.V., Cheng,A.A., Liu,S. *et al.* (2020) CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.*, **48**, D517–D525.
47. Bortolaia,V., Kaas,R.S., Ruppe,E., Roberts,M.C., Schwarz,S., Cattoir,V., Philippon,A., Allesoe,R.L., Rebelo,A.R., Florensa,A.F. *et al.* (2020) ResFinder 4.0 for predictions of phenotypes from genotypes. *J. Antimicrob. Chemother.*, **75**, 3491–3500.
48. Bastian,M., Heymann,S. and Jacomy,M. (2009) Gephi: an open source software for exploring and manipulating networks. *ICWSM*, **3**, 361–362.
49. Cabezón,E., Lanka,E. and 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.
50. Sastre,J.I., Cabezón,E. and 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.
51. Durand,R., Huguet,K.T., Rivard,N., Carraro,N., Rodrigue,S. and Burrus,V. (2021) Crucial role of *salmonella* genomic island 1 master activator in the parasitism of IncC plasmids. *Nucleic Acids Res.*, **49**, 7807–7824.
52. Moran,R.A. and Hall,R.M. (2017) Analysis of pCERC7, a small antibiotic resistance plasmid from a commensal ST131 *Escherichia coli*, defines a diverse group of plasmids that include various segments adjacent to a multimer resolution site and encode the same NikA relaxase accessory protein enabling mobilisation. *Plasmid*, **89**, 42–48.
53. Garcillán-Barcia,M.P., Pluta,R., Lorenzo-Díaz,F., Bravo,A. and Espinosa,M. (2022) The facts and family secrets of plasmids that replicate via the rolling-circle mechanism. *Microbiol. Mol. Biol. Rev.*, **86**, e0022220.
54. Łobocka,M.B., Rose,D.J., Plunkett,G., Rusin,M., Samojedny,A., Lehnherr,H., Yarmolinsky,M.B. and Blattner,F.R. (2004) Genome of bacteriophage p1. *J. Bacteriol.*, **186**, 7032–7068.
55. Lim,J.Y., Yoon,J. and Hovde,C.J. (2010) A brief overview of *Escherichia coli* O157:H7 and its plasmid O157. *J. Microbiol. Biotechnol.*, **20**, 5–14.
56. Avila,P., Núñez,B. and de la Cruz,F. (1996) Plasmid R6K contains two functional oriTs which can assemble simultaneously in relaxosomes in vivo. *J. Mol. Biol.*, **261**, 135–143.
57. San Millan,A. and MacLean,R.C. (2017) Fitness costs of plasmids: a limit to plasmid transmission. *Microbiol. Spectr.*, **5**, https://doi.org/10.1128/microbiolspec.MTBP-0016-2017.
58. Turner,P.E., Cooper,V.S. and Lenski,R.E. (1998) Tradeoff between horizontal and vertical modes of transmission in bacterial plasmids. *Evolution*, **52**, 315–329.
59. Moleres,J., Santos-López,A., Lázaro,I., Labairu,J., Prat,C., Ardanuy,C., González-Zorn,B., Aragon,V. and Garmendia,J. (2015) Novel $bla_{\mathrm{ROB-1}}$-bearing plasmid conferring resistance to β-lactams in *Haemophilus parasuis* isolates from healthy weaning pigs. *Appl. Environ. Microbiol.*, **81**, 3255–3267.
60. Koraimann,G. and Wagner,M.A. (2014) Social behavior and decision making in bacterial conjugation. *Front. Cell Infect. Microbiol.*, **4**, 54.
61. Hooton,S.P.T., Pritchard,A.C.W., Asiani,K., Gray-Hammerton,C.J., Stekel,D.J., Crossman,L.C., Millard,A.D. and Hobman,J.L. (2021) Laboratory stock variants of the archetype silver resistance plasmid pMG101 demonstrate plasmid fusion, loss of transmissibility, and transposition of Tn7/pco/sil into the host chromosome. *Front. Microbiol.*, **12**, 723322.
62. Dionisio,F., Zilhão,R. and Gama,J.A. (2019) Interactions between plasmids and other mobile genetic elements affect their transmission and persistence. *Plasmid*, **102**, 29–36.
63. San Millan,A., Heilbron,K. and MacLean,R.C. (2014) Positive epistasis between co-infecting plasmids promotes plasmid survival in bacterial populations. *ISME J.*, **8**, 601–612.
64. Humphrey,S., San Millán,Á., Toll-Riera,M., Connolly,J., Flor-Duro,A., Chen,J., Ubeda,C., MacLean,R.C. and Penadés,J.R. (2021) Staphylococcal phages and pathogenicity islands drive plasmid evolution. *Nat. Commun.*, **12**, 5845.

Downloaded from https://academic.oup.com/nar/article/51/7/3001/6849495 by guest on 30 September 2023