

## Mobility of Plasmids†

Chris Smillie,<sup>1,2,‡</sup> M. Pilar Garcillán-Barcia,<sup>3,‡</sup> M. Victoria Francia,<sup>4</sup>  
Eduardo P. C. Rocha,<sup>1,2,\*</sup> and Fernando de la Cruz<sup>3,\*</sup>

*Institut Pasteur, Microbial Evolutionary Genomics, CNRS, URA2171, F-75015 Paris, France<sup>1</sup>; UPMC Univ. Paris 06, Atelier de BioInformatique, F-75005 Paris, France<sup>2</sup>; Departamento de Biología Molecular e Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC), Universidad de Cantabria-CSIC-IDICAN, C. Herrera Oria s/n, 39011 Santander, Spain<sup>3</sup>; and Servicio de Microbiología, Hospital Universitario Marqués de Valdecilla e Instituto de Formación e Investigación Marqués de Valdecilla (IFIMAV), Av. Valdecilla s/n, 39008 Santander, Spain<sup>4</sup>*

<b>INTRODUCTION</b> .....	<b>435</b>
<b>GENERAL PROPERTIES OF PLASMID MOBILITY</b> .....	<b>435</b>
Biological function and agents of plasmid mobility .....	435
Essential lexicon of plasmid conjugation .....	435
Mechanism of conjugation.....	436
Dimensions of the plasmid pool: methodological approach.....	436
<b>OVERVIEW OF PLASMID GENOMES</b> .....	<b>436</b>
Plasmid database with respect to prokaryotic phyla.....	436
Association between plasmid size, genome size, and coding density.....	437
<b>ASSESSMENT OF MOBILITY</b> .....	<b>437</b>
Relaxases .....	437
Plasmid relaxases belong to six protein families .....	437
Possibility of new relaxase families .....	439
Mating Pair Formation .....	439
Classification of T4SSs: the case of VirB4 .....	439
Cases of other protein families in the T4SS: specific and nonspecific proteins.....	439
Diversity of Conjugative Systems .....	440
The plasmid T4SS classification is virtually complete for the proteobacteria .....	440
Abundance of each T4SS in proteobacteria.....	440
Mobility classes in proteobacterial plasmids.....	440
Hybrid and Multiple Conjugation Systems.....	441
<b>MOBILITY OF GenBank PLASMIDS</b> .....	<b>441</b>
Association between plasmid size and mobility.....	441
Size of conjugative plasmids .....	441
Relationship between MOB family, MPF family, and plasmid size.....	442
When Inventories Do Not Fit.....	442
What is the frequency with which one finds the key elements of plasmid transmissibility among different clades? .....	442
<b>SYSTEMATICS AND MOBILITY</b> .....	<b>443</b>
<b>COEVOLUTION OF MOBILIZATION AND CONJUGATION</b> .....	<b>444</b>
Phylogeny of T4CPs and VirB4 in the light of relaxases.....	444
T4CP and VirB4 subgroups follow prokaryotic phylogeny .....	444
VirB4 phylogeny matches T4SS archetypes .....	444
<b>HOW MOBILE ARE NONMOBILIZABLE PLASMIDS?</b> .....	<b>446</b>
Are Large Plasmids Becoming Secondary Chromosomes? .....	446
<b>TOWARDS COMPARATIVE GENOMICS OF PLASMIDS</b> .....	<b>447</b>
<b>CONCLUSION</b> .....	<b>448</b>
<b>ACKNOWLEDGMENTS</b> .....	<b>449</b>
<b>REFERENCES</b> .....	<b>449</b>

\* Corresponding author. Mailing address for Eduardo P. C. Rocha: Institut Pasteur, Microbial Evolutionary Genomics, CNRS, URA2171, F-75015 Paris, France. Phone: 33-140-613353. Fax: 33-144-387251. E-mail: erocha@pasteur.fr. Mailing address for Fernando de la Cruz: Departamento de Biología Molecular e Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC), Universidad de Cantabria-CSIC-IDICAN, C. Herrera Oria s/n, 39011 Santander, Spain. Phone: 34-942-201945. Fax: 34-942-201945. E-mail: delacruz@unican.es.

† Supplemental material for this article may be found at <http://mmb.asm.org/>.

‡ C.S. and M.P.G.-B. contributed equally.

## INTRODUCTION

Plasmids are, along with phages and integrative conjugative elements, the key vectors of horizontal gene transfer and essential tools in genetic engineering. They often code for genes involved in detoxication, virulence, ecological interactions, and antibiotic resistance. Hence, an understanding of plasmid mobility is essential to an understanding of the evolution of these important bacterial traits, often involved in human health or well-being. However, while the last decades have seen a remarkable enlightenment of plasmid mobility mechanisms in model systems (reviewed in references 52 and 82) (5, 36, 50, 54, 55, 123), the identification and characterization of plasmid mobility on a global scale remain unexplored. There is a need to devise such a tool for several reasons. First, if we want to give a systems biology answer to the problems of plasmid dissemination, e.g., the dissemination of multiple-drug-resistant bacteria in hospital or farm environments or the evolution of catabolic plasmids in contaminated environments, we must understand plasmid mobility. A global understanding of plasmid mobility can help us control the dissemination of these fastidious genes and thus help curb the recent increase in human mortality produced by infectious diseases in advanced countries. By the same token, it can help the efficiency of bioremediation efforts or the biological fight against other types of damaging bacteria. Second, microbial ecology is undergoing a revolution caused by the availability of metagenomic approaches that allow the sampling of the genetic diversity of microbial communities. Knowing how information flows in such communities is essential to interpreting how communities respond to changes. Third, the ability of plasmids to move between different hosts is of technical and industrial importance. The identification of the diversity of mobility mechanisms might allow the development of new, more-efficient, and better-adapted vectors for genetic engineering and the release of genetically modified microorganisms for bioremediation and pest control, etc. Most prokaryotes remain genetically intractable, and an understanding of the natural mechanisms of gene mobility is bound to allow the creation of new tools. Finally, conjugation is a secretion system that must adapt to cell physiology. An understanding of its diversity might enlighten how existing variants of this secretion mechanism are adapted to peculiar cellular envelopes or environments.

In this review we make use of the abundant available genomic data to extract a few general concepts that, we hope, will help our understanding of plasmid mobility. To carry out such an analysis, we first established a computational protocol to identify conjugation and mobilization genetic modules in 1,730 plasmids and used these data to establish a plasmid classification system. This allowed an accurate classification of proteobacterial conjugative or mobilizable systems in a combination of four mating pair formation (MPF) and six mobilization families (the term family is used here as it refers to protein families, that is, a set of proteins that are related in sequence and share a biological function). Few genes, e.g., those coding for conjugative coupling proteins, relaxases, and VirB4 proteins, are at the core of plasmid conjugation. Together with several auxiliary genes, they have evolved into systems with specific adaptations to the cell physiology and to

ecological strategies. Second, we used this inventory of plasmid mobility and family-specific genes to characterize systems in other prokaryotic clades. We found that, globally, one-fourth of the plasmids are conjugative, and as many are mobilizable. Half of all plasmids are classed as being nonmobilizable. Third, evolutionary analysis allowed us to trace the evolution of conjugation elements and showed an ancient divergence between mobility systems, with relaxases and type IV coupling proteins (T4CPs) often following separate paths from type IV secretion systems (T4SSs). Phylogenetic patterns of mobility proteins are consistent with the phylogeny of the host prokaryotes, suggesting that plasmid mobility is in general circumscribed within large clades. Surprisingly, most very large plasmids are nonmobilizable. We have made no attempt to change the naming of GenBank replicons. Therefore, we used all replicons named "plasmid" and none named "chromosome." There is no consensual distinction between plasmids and chromosomes. We therefore review the claims that many very large plasmids are becoming secondary chromosomes. Indeed, we find that these plasmids tend to be nonmobilizable and contain appreciable amounts of essential genes. The last section of this article discusses outstanding issues in comparative genomics of plasmids in relation to the understating of their mobility.

## GENERAL PROPERTIES OF PLASMID MOBILITY

**Biological function and agents of plasmid mobility.** Genetic information flows at high rates among prokaryotes, leading bacteria and archaea to have a small core genome that is conserved within a species and a large pangenome that is highly variable (137). Widespread horizontal gene transfer has profound evolutionary implications (35, 42, 90, 109). First, it allows homologous recombination between closely related strains or species in a process resembling eukaryotic sex (55, 142). Second, it leads to the integration of new genetic information, creating large functional leaps that allow fast adaptation to new environments or to stressful conditions (62). Third, gene mobility has been proposed to drive microbial cooperative processes (105). From the three classical mechanisms of horizontal transfer, transformation, transduction, and conjugation, the latter is thought to be quantitatively more important (66). This is because phages have restricted host ranges and small cargo regions, whereas some plasmids can conjugate between remotely related organisms, including conjugation from bacteria to eukaryotes (10, 69). Plasmids, along with integrative conjugative elements (ICEs), are the major players in conjugation processes. Many of the genes allowing bacteria to metabolize toxic organic compounds such as antibiotics are carried by plasmids (9, 106, 138). Plasmids also often code for information essential for the interaction of bacteria with multicellular eukaryotes, including nitrogen fixation by rhizobia (95), plant cell manipulation by *Agrobacterium* species (59), and virulence by *Shigella* species (21), among many other human pathogens. Plasmids are also molecular biology workhorses whose mobility opened up the possibility of genetic manipulation (28).

**Essential lexicon of plasmid conjugation.** Mobility is an essential part of plasmid fitness. It is also a key element to an understanding of the epidemiology of plasmid-carried traits such as virulence and antibiotic resistance. As such, two func-

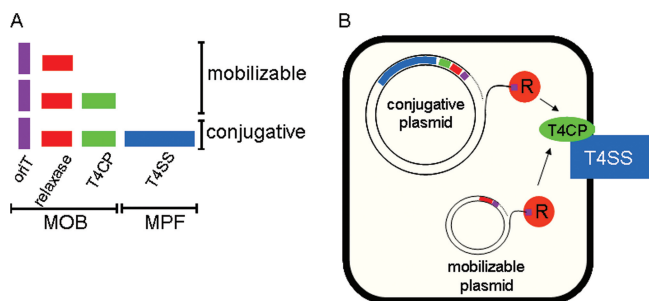


FIG. 1. (A) Schematic view of the genetic constitution of transmissible plasmids. Self-transmissible or conjugative plasmids code for the four components of a conjugative apparatus: an origin of transfer (*oriT*) (violet), a relaxase (R) (red), a type IV coupling protein (T4CP) (green), and a type IV secretion system (T4SS) (blue). The T4SS is, in fact, a complex of 12 to 30 proteins, depending on the system (see text). Mobilizable plasmids contain just a MOB module (with or without the T4CP) and need the MPF of a coresident conjugative plasmid to become transmissible by conjugation. (B) Scheme of some essential interactions in the process of conjugation. The relaxase cleaves a specific site within *oriT*, and this step starts conjugation. The DNA strand that contains the relaxase protein covalently bound to its 5' end is displaced by an ongoing conjugative DNA replication process. The relaxase interacts with the T4CP and then with other components of the T4SS. As a result, it is transported to the recipient cell, with the DNA threaded to it. Subsequently, the DNA is pumped into the recipient by the ATPase activity of the T4CP (93).

tions are deemed essential for plasmid survival: DNA replication and horizontal spread. The latter may occur by conjugation if a plasmid carries two sets of genes. The set of mobility (MOB) genes is essential and allows conjugative DNA processing (the MOB genes were also called Dtr genes, for DNA-transfer replication). Besides, a membrane-associated mating pair formation (MPF) complex, which is a form of a type 4 secretion system (T4SS), provides the mating channel. A plasmid that codes for its own set of MPF genes is called self-transmissible or conjugative. If it uses an MPF of another genetic element present in the cell, it is called mobilizable. Some plasmids are called nonmobilizable because they are neither conjugative nor mobilizable. They spread by natural transformation or by transduction. Hence, plasmids can be classified into three categories according to mobility: conjugative, mobilizable, and nonmobilizable.

**Mechanism of conjugation.** The only protein component of the conjugative machinery that is common to all transmissible, i.e., conjugative or mobilizable, plasmids is the relaxase (Fig. 1). The relaxase is a key protein in conjugation, since it recognizes the origin of transfer (*oriT*), a short DNA sequence which is the only sequence required in *cis* for a plasmid to be conjugally transmissible. The relaxase catalyzes the initial and final stages in conjugation, that is, the initial cleavage of *oriT* in the donor, to ultimately produce the DNA strand that will be transferred, and the final ligation of the transported DNA in the recipient cell that reconstitutes the conjugated plasmid (see reference 36 for a recent review). Conjugative relaxases are structurally related to rolling-circle replication initiator proteins, and they catalyze similar biochemical reactions. However, they are easily distinguished because relaxase amino acid sequences are linear permutations of the replication initiator sequences, as discussed previously (57, 63). Mobilizable plas-

mids carry only the relaxosomal components *oriT*, a relaxase gene, and one or more nicking auxiliary proteins. On the other hand, conjugative plasmids carry all the machinery needed for self-transfer. This includes, besides the above-mentioned relaxosome components, the type IV coupling protein (T4CP) and the components of the mating channel that assemble a T4SS. The T4CP is involved in the connection between the relaxosome and the transport channel (11, 41, 94, 101). It is also thought to energize the process of DNA transport (135, 136). The conjugative mating channel is basically a protein secretion channel, which transports the relaxase protein bound to the DNA to be transferred (43, 58). According to the nomenclature of protein secretion mechanisms, it is a T4SS (54). The phylogenetic relationship among relaxases has been traced, leading to a classification of conjugative systems into six MOB families: MOB<sub>F</sub>, MOB<sub>H</sub>, MOB<sub>O</sub>, MOB<sub>C</sub>, MOB<sub>P</sub>, and MOB<sub>V</sub> (52, 57). This classification extends to the entire mobility region, which includes the nicking auxiliary proteins (145) and the T4CPs (57). There have been reports on the classification and phylogenetic relationships among NTPases (53, 112), yet little is known about the classification of conjugative T4SSs in prokaryotes, except for some specific clades such as *Rickettsia* (147) or for VirB4 of T4SSs involved in pathogenicity (53). While the manuscript was being finished, a review giving an extensive description of known T4SSs in prokaryotes, including many conjugative systems, was published (5). That analysis showed some common themes among T4SSs as well as an important diversity, and here we aim to quantify them.

#### Dimensions of the plasmid pool: methodological approach.

No estimation exists on the proportion of plasmids that are conjugative, mobilizable, or nontransmissible. Even if there are over 1,700 complete plasmid genomes in GenBank, there is experimental evidence of transmissibility for just a small fraction of them (57). Since several proteins involved in conjugation have been extensively studied, we figured that it should be possible to draw an educated hypothesis of plasmid mobility just by analyzing plasmid genomes. Thus, to carry out this review, we identified and classified key proteins in the two major modules, i.e., sets of functionally related genetic elements involved in plasmid mobility: relaxases and T4SSs. By identifying relaxases, we obtained a list of putatively transmissible and nontransmissible plasmids. Along the following discussion, we will assume that plasmids containing relaxases are either mobilizable or conjugative. We then defined the major characteristics of T4SSs and identified them on plasmids, thereby separating conjugative from mobilizable plasmids. We did much of this survey in two steps. Initially, an intensive expert study allowed the identification and classification of plasmids. Using these results as guidelines, we created a set of methodologies to automate the analysis. Our results (see Table S1 in the supplemental material for the complete list of plasmids analyzed) suggest that the majority of plasmids are not transmissible and that many conjugative systems remain yet to be found in most bacterial clades.

## OVERVIEW OF PLASMID GENOMES

**Plasmid database with respect to prokaryotic phyla.** By early 2009 there were 1,730 complete plasmid sequences in

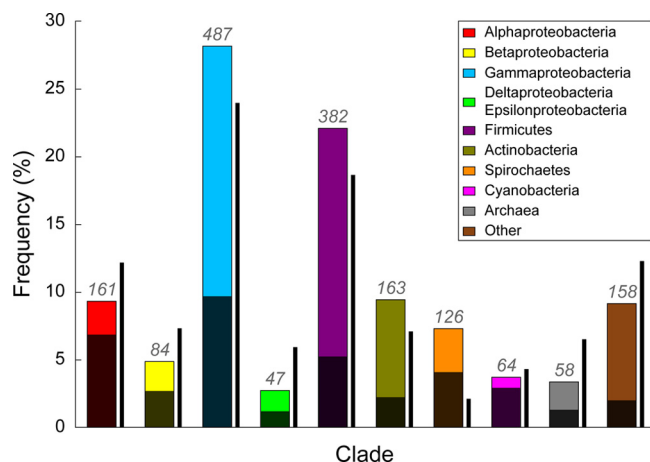


FIG. 2. Distribution of the 1,730 GenBank plasmids by taxon. The large bars represent, for a given clade, the proportion of plasmids sequenced along with prokaryotic genomes (dark color) and sequenced independently of the host (light color). Bars are topped with the numbers of plasmids for the clades. The thin bars represent the proportion of plasmids of each clade among the plasmids sequenced with the prokaryotic host.

GenBank. According to their annotation files, these plasmids code for a total of 106,288 proteins. Some plasmids are very small, the smallest being *Thermotoga petrophila* RKU1 plasmid pRKU1, with 846 bp and carrying only the *rep* gene. Nine small plasmids carry no annotated gene. On the other hand, some other plasmids are very large, for example, 22 plasmids carrying more than 500 genes. The largest replicon marked as a plasmid, *Ralstonia solanacearum* GMI1000 plasmid pGMI1000MP, carrying 1,674 genes, is larger than 20% of the published bacterial chromosomes. It follows that the coding potential of plasmids is both large and diverse. The set of plasmids that we analyzed comes from two different sources, plasmid-sequencing projects (~62%), which are motivated by scientific interest in the plasmid itself, and microbial genome projects (~38%), which typically show a marginal interest in plasmids. The latter sample is thus less biased by the interests of the plasmid biology community. Both samples overrepresent culturable prokaryotes and often include strains that were extensively cultured before sequencing. It is unclear if this fact changes the structure of plasmids significantly. An analysis of *Escherichia coli* K-12 strains MG1655 and W3110, which diverged in the laboratory over 50 years ago, showed very few differences between these strains, suggesting that genomes are relatively stable in the laboratory in the absence of extensive transposition (68). One likely effect of extensive subculturing of bacteria is plasmid loss. This leads to an undersampling of the natural pool of plasmids.

Most plasmids have a description of the original source, even if this is sometimes not retrievable in a simple manner. According to this information, a few bacterial phyla contain most of the sequenced bacterial genomes. As shown in Fig. 2, almost half of the plasmids come from proteobacteria (46%) and especially from gammaproteobacteria (29%). Firmicutes, spirochetes, and actinobacteria make most of the remaining data set. Overall, these four clades correspond to nearly 80% of all sequenced plasmids. Unsurprisingly, qualitatively similar,

albeit fewer, biases are found among the completely sequenced prokaryotic chromosomes (Fig. 2). Firmicutes and gammaproteobacteria are extensively sampled because they include the plasmids that are best studied and also the ones most associated with the dissemination of antibiotic resistance among human-pathogenic bacteria. Noticeably, clades containing many large plasmids, such as alphaproteobacteria and cyanobacteria, have been sampled mostly thanks to microbial genome projects.

**Association between plasmid size, genome size, and coding density.** Larger plasmids are hosted by prokaryotes with larger chromosomes (Spearman's  $\rho$  correlation between chromosome and plasmid lengths of 0.34;  $P < 0.001$ ). Within genome sequences, chromosomes have a higher coding density than plasmids (86% versus 75%;  $P < 0.001$  by a Wilcoxon test). Surprisingly, larger plasmids have higher coding densities than do smaller ones (Spearman's  $\rho$  between gene density and plasmid size of 0.45;  $P < 0.001$ ) (see Fig. S1 in the supplemental material). No such trend is found for prokaryotic chromosomes ( $\rho = -0.03$ ;  $P = 0.37$ ). To test if these results were due to poor annotations, we made automatic reannotations using Glimmer v3.02 (37). In one assay we used the plasmid sequence for the training set, whereas in the other we used the host chromosome. In both cases the extents of the correlation between plasmid size and gene density were similar or increased (data not shown). Hence, while larger plasmids have coding densities approaching those of chromosomes, small plasmids are less coding dense. This surprising result might be due to the inability of current annotation protocols to identify small protein-encoding genes and stable RNA genes. If so, small plasmids might have a higher proportion of these small genes. Alternatively, one might imagine that, below a certain size, the gene coding density decreases because some incompressible part of the plasmid is occupied by sequences governing the DNA structural scaffold or signals regulating replication or transmission. Since the automatic reannotation efforts did not significantly increase the number of coding sequences, we used the gene positions of the GenBank files in this assessment of plasmid mobility. The key proteins discussed in this work, relaxase, T4CP, and VirB4, are large, and therefore, their genes are unlikely to have passed unnoticed by annotation software. We have not attempted at this stage to extensively identify pseudogenes associated with plasmid mobility. However, we did make extensive analyses of pseudogenization events using tBLASTn for intergenic regions of plasmids carrying all genes expected from a conjugative plasmid except a key one, e.g., *virD4* or *virB4*.

## ASSESSMENT OF MOBILITY

### Relaxases

**Plasmid relaxases belong to six protein families.** An expert classification of MOB families was carried out previously for a large set of plasmids (52, 57). In this review, we followed this classification methodology as a first step and extended it manually to an analysis of 1,000 plasmids, which were classified into six families and 31 subfamilies. We then used these results to automatically cluster relaxases into six families (MOB<sub>P</sub>, MOB<sub>F</sub>, MOB<sub>V</sub>, MOB<sub>O</sub>, MOB<sub>H</sub>, and MOB<sub>C</sub>), as shown in Fig.

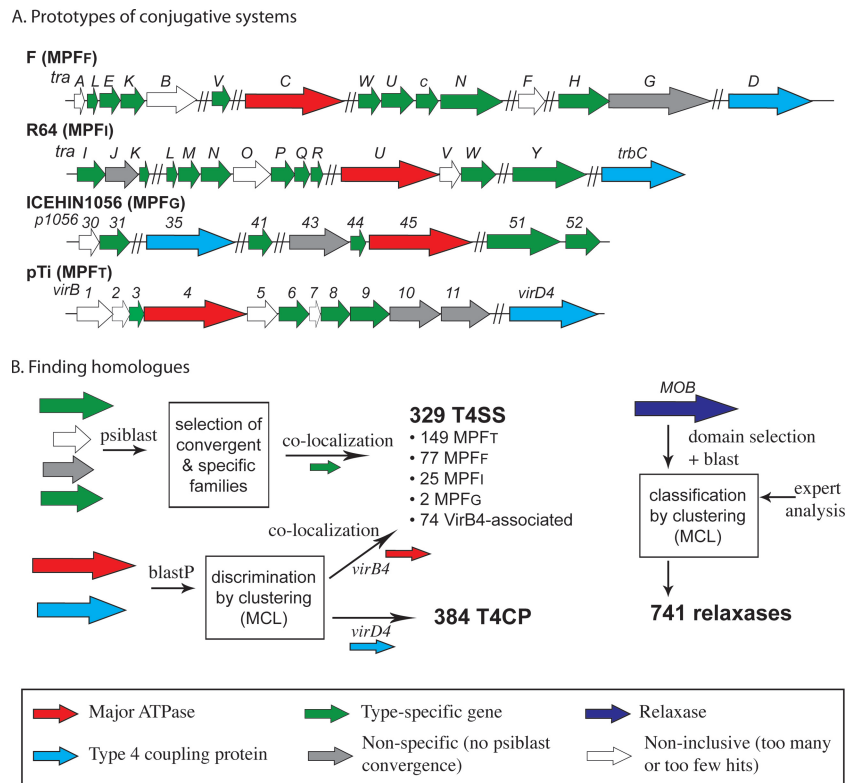


FIG. 3. Overall diagram of the analysis to identify and classify relaxases, T4CPs, and T4SSs. To classify T4SSs, we used selected genes from four well-known systems (A), which were then used as templates to search for homology (B). To cluster homologous proteins, we used all-against-all BLASTP (4) (E value of  $<1e-4$ ) and the Markov cluster algorithm (MCL) (47) (after extensive searches,  $I = 1.12$ , using a log transformation of the BLASTP E values as edge weights). We identified VirB4s and T4CPs by BLASTP. These families being homologous, a disambiguation step was necessary. To discriminate between VirB4s and T4CPs, we used an expert annotated data set of these proteins, elaborated by the authors, which was used to query the MCL clusters resulting in a resolution between the two families ( $I = 1.16$ ). Phylogenetic analyses were then used to remove spurious hits. To classify T4SSs into archetypes, we analyzed the indicated genes of plasmid F (MPF<sub>F</sub>) (GenBank accession number NC\_002483), plasmid Ti (MPF<sub>T</sub>) (accession number NC\_002377), plasmid R64 (MPF<sub>R</sub>) (accession number NC\_005014), and the genomic loci of ICEHin1056 (MPF<sub>C</sub>) (accession number AJ627386). We made PSI-BLAST searches against the plasmid database (maximum of 30 iterations). Genes with convergent PSI-BLAST results were grouped into colocalized systems (defined as sets of genes  $<50$  genes apart), and the distribution of each gene across all systems was calculated. For each system, we defined a set of marker genes that were nearly universally present in plasmids containing T4SSs and absent from plasmids lacking T4SSs (they were called “type-specific” genes and are shown in green to distinguish them from the remaining “nonspecific” genes, shown in gray). The final list of T4SSs resulted from the joint analysis of type-specific T4SS- and VirB4-containing plasmids. The automatic discovery of relaxases first used iterative BLASTP searches using the 1,730 plasmids and the 300 N-terminal amino acids for each of the six prototypical relaxases previously described (57). The E values used were those for TrwC\_R388 (MOB<sub>F</sub>;  $1e-8$ ), TraI\_R27 (MOB<sub>H</sub>;  $1e-4$ ), TraI\_RP4 (MOB<sub>P</sub>;  $1e-4$ ), MobA\_RS1010 (MOB<sub>C</sub>;  $1e-4$ ), MobM\_pMV158 (MOB<sub>V</sub>;  $1e-5$ ), and MobC\_CloDF13 (MOB<sub>C</sub>;  $1e-4$ ). After completing a search, the hits with the lowest E values were used as queries to retrieve distantly related sequences. We used this set to query the MCL clusters, with each relaxase family being unambiguously assigned to a cluster, which were then manually validated.

3 and explained in the legend of Fig. 3. All classes could be distinguished by using the automatic clustering procedure, with eight clusters accounting for 98.5% of all relaxases (see Fig. S2 in the supplemental material). However, some relaxases were in small clusters (e.g., two to three items). Changing the clustering parameters (i.e., the inflation parameter in the Markov cluster algorithm [MCL]) did show a lower-level structure, suggesting some heterogeneity in the classes and justifying their further division into subclasses (as attempted in reference 57). In this review we have not extended our automatic classification procedure to subclasses because many classes still contain few elements, precluding the establishment of reliable automatic classification. Among 1,730 plasmids, we found a total of 741 relaxases in 673 plasmids (Fig. 4). All families were found in the database, with MOB<sub>H</sub> (28 plasmids) and MOB<sub>C</sub> (24 plasmids) being the least represented, while MOB<sub>P</sub> was the

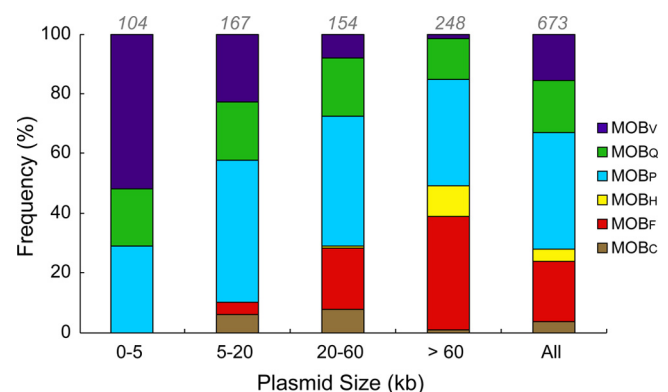


FIG. 4. Distribution of relaxase families according to plasmid size. Each bar shows the abundance of each relaxase family for the given plasmid size range. The rightmost bar represents the overall distribution of relaxases in all plasmids.

most represented (273 plasmids). As shown previously (57), within each family there are subfamilies corresponding to phylogenetic clades that contain more members, e.g., MOB<sub>F12</sub>, which harbors relaxases of IncF plasmids; MOB<sub>P11</sub>, which groups relaxases of IncP plasmids; MOB<sub>P5</sub>, which contains mostly MOB<sub>HEN</sub> (ColE1-like) relaxases; MOB<sub>V1</sub>, which groups relaxases of plasmids belonging to several Inc groups from firmicutes; and MOB<sub>C2</sub>, producing aggregation substances.

MOB<sub>P</sub> and MOB<sub>O</sub> are uniformly distributed in plasmids of all size ranges, but the remaining families are associated with plasmids of different sizes ( $P < 0.001$  by a Wilcoxon test on similar sizes of plasmid families) (Fig. 4). MOB<sub>F</sub> and MOB<sub>H</sub> are typical, or exclusive, of large plasmids. MOB<sub>C</sub> is present in mid-sized plasmids. MOB<sub>V</sub> is almost absent from large plasmids and present in over 50% of all plasmids less than 5 kb long. This finding suggests that particular MOB modules (and, thus, specific DNA-processing mechanisms) have adapted to plasmids of different sizes, possibly as a result of the interaction between conjugation and replication systems. For example, MOB<sub>V</sub> plasmids of firmicutes, which replicate by a rolling circle, are known to become unstable with increasing size (19, 20, 71, 86).

**Possibility of new relaxase families.** We were surprised by the large fraction of large plasmids lacking relaxases. In an attempt to find out if there might be relaxases that were not detected by our automated analysis, all plasmids within the 30- to 60-kb size range that did not give a hit against relaxases (70 out of a total of 140 plasmids in January 2008) were manually inspected for the presence of relaxase-like proteins. This can be done because most relaxase families belong to a single protein superfamily, the so-called 3H family, because of a signature of a cluster of three histidines in the catalytic site (57). After a thorough examination, only one plasmid that contained a possible new relaxase was found. *Bacteroides thetaiotaomicron* VPI-5482 plasmid p5482 (33,038 bp) contains some genes coding for T4SS components, a putative T4CP, MobC (25% identity to that of plasmid pBFY46), and also a putative relaxase, MobB, which show 27% and 28% identities to similar proteins in plasmids pBIF10 and pBUN24, respectively. These three putative relaxases were not retrieved by PSI-BLAST starting from any of the six known relaxase families. Since the three motifs typical of the 3H class of relaxases could be identified, they might represent the first members of a seventh relaxase family. There is no experimental information concerning the transfer abilities of these plasmids, and currently, the number of known elements in the family is small. Hence, we left it out of the graphical representations. If the same approach that we followed here with plasmids in the size range of 30 to 60 kb was applied to larger plasmids not showing a relaxase gene (similar to those used as queries up to now), additional relaxase families might be uncovered. It should be emphasized, nevertheless, that for each of the six relaxase classes analyzed in this work, there is at least one prototype plasmid for which detailed molecular studies warrant the inclusion of the relevant genes within the realm of conjugative relaxases. Since relaxases show ancestral homology to other DNA-processing proteins (RC replication proteins, IS91-like transposases, and so on [57]), extreme care has to be exerted before new gene families are proposed as relaxases. As will

become evident at the end of this work, we are convinced that new relaxase families remain to be discovered, particularly in phyla other than the proteobacteria. However, before claiming such a discovery, it is imperative that some experimental data within the proposed family rigorously identify the protein as a relaxase.

### Mating Pair Formation

**Classification of T4SSs: the case of VirB4.** After automating the classification of relaxases, we set up to identify putative T4SSs associated with the conjugative pilus. For this, we classified known T4SSs according to protein homology among a dozen plasmids from different MOB and Inc types. This allowed the clustering of all known proteobacterial T4SSs into four groups (Fig. 3), as suggested previously (72). For each group we used the nomenclature of one model T4SS representative of the group, notably the *vir* system for MPF<sub>T</sub> (25, 123), F for MPF<sub>F</sub> (82), R64 for MPF<sub>I</sub> (77), and ICEHIN1056 for *Haemophilus influenzae* genome island-like MPF<sub>G</sub> (73). For each group we selected a dozen genes for which the function was known and/or for which inactivation was shown experimentally to strongly reduce conjugation rates (see the legend of Fig. 3). We then classified genes into three categories: nearly ubiquitous, group specific, and others. Finally, we built phylogenetic trees to find potential false assignments.

In accordance with data reported in the literature (5), we found that the only gene present in practically all T4SSs codes for the VirB4 family (TraC in the F plasmid and TraU in R64). However, even though TraU is generally believed to be a homologue of VirB4, it could not be retrieved in our analysis due to the lack of significant sequence similarity. Instead, the search for distant homologues of VirB4 with PSI-BLAST resulted in the retrieval of other homologous but functionally different proteins such as T4CPs. The low level of similarity between TraU and VirB4, and the highest level of similarity of the latter with VirD4, is puzzling. One might suppose that TraU evolved extremely fast, but we found no very long branches in the TraU tree, and one would still expect TraU to diverge among proteobacterial VirB4 proteins, since its distribution seems to be confined to this clade. It is therefore more likely that either TraU arose before the split of VirB4 and the T4CP or it arose independently from another ATPase. Current data do not allow one to unambiguously choose between these hypotheses.

We disentangled VirB4 and T4CP proteins with a clustering procedure and fine-tuned the clusters using phylogenetic analyses (see the legend of Fig. 3). From the 255 T4CPs having been found by expert annotation in 1,000 plasmids, we found 100% of them in just two clusters. We also found all proteins annotated as VirB4 in a few separate clusters. We retrieved a total of 327 VirB4-encoding genes (including homologues of *virB4* and *traU*). Importantly, these genes were always inside the region coding for the T4SS, whereas the gene coding for the T4CP often matched other regions of the plasmid.

**Cases of other protein families in the T4SS: specific and nonspecific proteins.** Other nearly ubiquitous protein families, most notably VirB11, were not reliable markers of the presence of a T4SS. They were also poorly specific of T4SS groups, and thus, we called them “nonspecific” proteins. For example,

VirB11 is homologous to ATPases of type II secretion systems (24), while it is absent from most T4SSs of firmicutes and archaea and from model plasmids such as F (5). These “non-specific” proteins were used in the expert analysis to confirm the presence or absence of T4SSs but were ignored in the automatic analysis, without consequences for its accuracy. The remaining genes were separated into families according to their discriminative powers. Some families matched functionally unrelated proteins, were too rare, or did not converge in PSI-BLAST searches (Fig. 3). These genes were discarded. The remaining genes were used as T4SS group-specific markers, as they were present in the vast majority of members of each T4SS group, while they were absent from other groups of T4SSs. They are likely to have important roles in conjugation. Some of the genes that are poor markers may also have important roles in some specific systems. However, when we analyzed genes whose inactivation had little effect on the frequency of conjugation, they turned out to be bad markers of specific T4SSs (e.g., p1056.30 for chromosomally encoded MPF<sub>G</sub> [72] or TraO for MPF<sub>I</sub> T4SSs [77]). These sporadic genes may provide accessory functions, evolve quickly, or be prone to analogous gene replacement. For example, many known T4SSs contain a lipoprotein, but this protein shows wide variations in size and sequence, and we found it to be irretrievable by sequence similarity searches, often even among closely related plasmids.

In summary, six families of MOB modules and four families of MPF modules represent most of the diversity of known conjugative transfer systems. Extensive work in microbial genetics and ecology needs to be carried out to explain why some of these modules associate preferentially among themselves (and with specific replication modules) and how they have adapted to particular ecosystems. Naturally, the ensuing question is how complete this set of known systems is and how it is represented among prokaryotes.

### Diversity of Conjugative Systems

**The plasmid T4SS classification is virtually complete for the proteobacteria.** We found 253 regions with several hits to T4SS type-specific genes in prokaryotic plasmids. There were 66 plasmids where VirB4 did not colocalize with a prototypical proteobacterial T4SS, of which 40 were from firmicutes and 13 were from archaea (see Table S1 in the supplemental material). Four proteobacterial plasmids had VirB4 but no clearly identifiable T4SS: the *Magnetospirillum gryphiswaldense* MSR-1 plasmid, *Gluconacetobacter diazotrophicus* PAI 5 plasmid pGDIPal5I, the *Pseudomonas syringae* pv. phaseolicola 1448A large plasmid, and *Achromobacter denitrificans* plasmid pEST4011. Inspection of these plasmids showed evidence of a degraded T4SS. One system (*Vibrio* species plasmid p23023) coded for homologues of several T4SS proteins, e.g., VirD4, VirB10, VirB1, and VirB11, but not of genes among those that we used to specify the MPF type (Fig. 3), i.e., no clearly identifiable type-specific genes. For all cases where we found a putative T4SS but no VirB4, a detailed inspection revealed either a *virB4* pseudogene or an incomplete T4SS. In short, among the 798 plasmids of proteobacteria, we found 236 plasmids coding for VirB4 or TraU and 236 plasmids coding for type-specific T4SS proteins. Two hundred thirty-two of these,

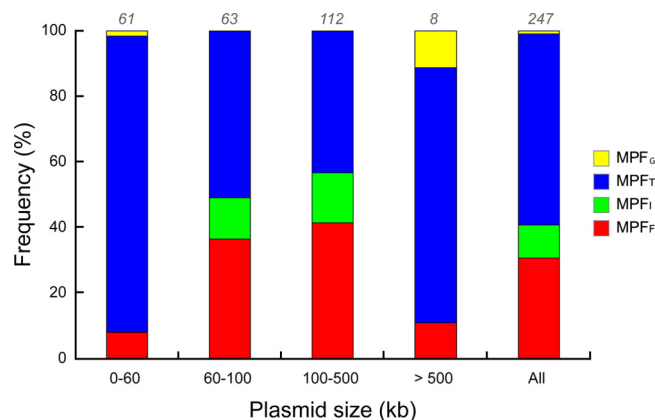


FIG. 5. Distribution of T4SS types in proteobacterial plasmids according to plasmid size. Each bar shows the abundance of each T4SS family for the given plasmid size range. The rightmost bar represents the overall distribution of T4SSs in all plasmids.

i.e., an outstanding 98%, coincided exactly. Since the search for T4SS classes and VirB4 was carried out independently and with different methods, the large overlap of the results strongly suggests that we can classify practically all VirB4-containing conjugative T4SSs in proteobacteria. Concomitantly, these four classes of T4SSs represent practically all the diversity of conjugative plasmids of sequenced proteobacteria.

**Abundance of each T4SS in proteobacteria.** Among classified conjugative plasmids of proteobacteria (Fig. 5), there was a clear overrepresentation of MPF<sub>T</sub> plasmids (142 plasmids [58%]) over MPF<sub>F</sub> plasmids (77 plasmids [31%]). MPF<sub>I</sub> plasmids were even rarer (25 plasmids [10%]). The same analysis restricted to plasmids sequenced with the microbial chromosomes showed similar frequencies (see Fig. S3 in the supplemental material). Thus, the biased distribution of these plasmids was not caused by the specific biases of plasmid models and might be representative of their true frequency in nature. We found two occurrences of MPF<sub>G</sub> in plasmids: *Marinobacter aquaeolei* VT8 plasmid pMAQU02 and *Haemophilus influenzae* plasmid ICEhin1056. Thus, contrary to previous suggestions (73), this type of T4SS is not associated exclusively with ICEs, although it is indeed rare among sequenced plasmids. The distribution of the three most abundant T4SSs showed a significant association with genome size ( $P < 0.001$  by a Wilcoxon test) (Fig. 5). For instance, MPF<sub>T</sub> was found more frequently in very large and in small plasmids, whereas MPF<sub>I</sub> was absent from small plasmids and MPF<sub>F</sub> was more frequent in intermediate-sized plasmids. The overrepresentation of MPF<sub>T</sub> in small plasmids may result from its lower complexity. Thus, model plasmids with MPF<sub>T</sub> systems, such as RP4 or R388, are composed of about 12 genes and allow mating only on solid surfaces. On the other hand, model plasmids with MPF<sub>F</sub> and MPF<sub>I</sub> systems, such as F and R64, respectively, are composed of around 30 genes and also allow mating at high frequencies in liquid culture (3, 15). It is not surprising that an additional set of genes is required for a functional liquid mating system. Results regarding the largest plasmids are harder to interpret given the small sample size (eight plasmids).

**Mobility classes in proteobacterial plasmids.** We could then analyze the joint distribution of MOB and MPF modules in the

most populated set in our database, that of proteobacteria. Only 28% of all proteobacterial plasmids contained both a relaxase and a T4SS and were therefore classed as conjugative (MOB plus MPF). These plasmids correspond to about half (54%) of the transmissible plasmids. The high frequency of cooccurrence of VirB4 with both T4CP- and T4SS-specific genes in proteobacteria supports the idea that VirB4 can be used as a marker of T4SSs.

### Hybrid and Multiple Conjugation Systems

Most plasmids contain just one MPF module and one associated MOB module, with surprisingly few exceptions. We found only one case of a cooccurrence of T4SS-specific genes of different MPF types in the same plasmid. In the transferrable plasmid pP99-018 of *Photobacterium damsela* subsp. *pisicida* (74), we found two MPF<sub>T</sub> genes (*virB8* and *virB9*) and seven MPF<sub>F</sub> genes (*traLEKVVUN*). A more detailed inspection of the DNA sequence of this plasmid showed a complete MPF<sub>F</sub> module and an incomplete MPF<sub>T</sub>, which is part of a recently acquired transposon-like element. Besides, these MPF<sub>T</sub> genes are absent in a closely related plasmid, pP99-018, from the same host. Therefore, our survey excludes the existence of hybrid T4SSs. While we found other plasmids carrying homologues of different systems, these invariably showed evidence of pseudogenization. Only 50 plasmids harbored two relaxases, and nine plasmids contained three relaxases. We found 10 plasmids that contained two T4SSs, in all cases two MPF<sub>T</sub>, usually associated with one MOB<sub>O</sub> and one MOB<sub>P</sub> relaxase and two T4CPs (see Table S2 in the supplemental material). It would be tempting to suggest that one T4SS is responsible for conjugation and that the other is responsible for the secretion of unrelated proteins. However, the finding of multiple relaxases in these plasmids argues against this view. Interestingly, 9 of the 10 plasmids with two MPFs were in pathogens or mutualists of plants. Several of these include the *Agrobacterium tumefaciens* Ti plasmid, which indeed codes for two T4SSs, one dedicated to Ti plasmid conjugation and the other dedicated to the transfer of T-DNA to plants (59). We are therefore inclined to think that all these plasmids carry two conjugation systems. These may either arise from plasmid cointegration or be used for different purposes/target cells. The existence of multiple relaxases and T4SSs may result from the cointegration of plasmids. For example, plasmids pDOJH10L and pK214 contain more than one relaxase and also contain more than one replication initiation protein (85, 111). Since there is a clear overrepresentation of the same MOB family and T4SS type in the same plasmids (63% and 100% of the total), it is possible that cointegration is less damaging for long-term plasmid stability if the relaxases and T4SSs are from the same type. On the other hand, this might simply reflect the preference of a given T4SS for that host or environment or a higher frequency of cointegration among plasmids of the same type. Indeed, very similar plasmids are expected to recombine at higher rates since they carry very similar genes, such as those coding for MOB and MPF, due to recent common ancestry.

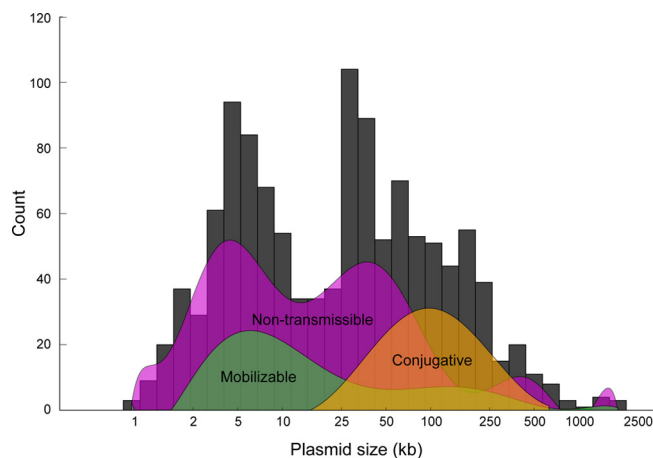


FIG. 6. Distribution of conjugative, mobilizable, and nonconjugative plasmids according to plasmid size (curves were created from a polynomial interpolation of the histograms of each class).

### MOBILITY OF GenBank PLASMIDS

**Association between plasmid size and mobility.** The plasmid database shows a bimodal distribution of plasmid sizes, with a clear local minimum at ~20 kb (Fig. 6). It has often been suggested that mobilizable plasmids tend to be less than 30 kb, while conjugative plasmids are usually larger. Indeed, plasmids coding for relaxases tend to be larger than the others (median sizes of 35 kb and 11 kb;  $P < 0.001$  by a Wilcoxon test), and conjugative plasmids are even larger (median, 181 kb;  $P < 0.001$ ). The classification of plasmids in terms of mobility shows that conjugative plasmids distribute around an average of 100 kb, while mobilizable plasmids have a mean peak at 5 kb and a broad, flat, secondary peak at around 150 kb (Fig. 6). Thus, speaking very broadly, this dichotomy has true value. Besides, plasmids smaller than 30 kb that code for a relaxase rarely code for a T4CP (6%), whereas larger plasmids generally do (86%).

Nevertheless, the distribution of nonmobilizable plasmids is also multimodal (Fig. 6) and suggests that things may not be that simple. Nonmobilizable plasmids show three distinctive peaks for small (around 5 kb), average (around 35 kb), and large (more than 300 kb) plasmids. The percentages of nonmobilizable plasmids larger than 20 kb are 42% in proteobacteria and 41% in the other clades, showing that the identification of large nonmobilizable plasmids is not an artifact caused by poorly known plasmid mobility in some clades. Interestingly, the two largest plasmids in our data set are classified as being nonmobilizable: *Burkholderia phymatum* STM815 plasmid pBPHY01 and *Ralstonia solanacearum* GMI1000 plasmid pGMI1000MP. Both are more than 1.9 Mb in size. Thus, a lack of relaxases, while more frequent in the very small plasmids, is also found among average- and large-sized plasmids. This result was surprising and perhaps breaks some misconceptions in plasmid biology. It will be discussed more extensively below (see How Mobile Are Nonmobilizable Plasmids?).

**Size of conjugative plasmids.** From 673 plasmids harboring relaxases, conjugative transfer was experimentally verified for only 123 of them (57), comprising all size ranges considered here except those larger than 1 Mb. In proteobacteria, the



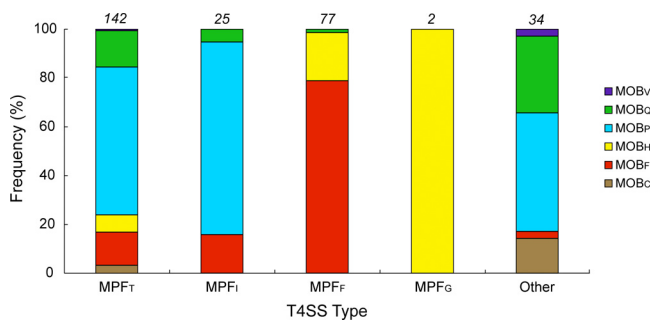


FIG. 7. Distribution of relaxase families within each T4SS type.

smallest plasmid that we identified as being putatively conjugative is 21.8 kb long (pCRY from *Yersinia pestis*) and is highly homologous to plasmids shown to be conjugative (128). In other clades, the smallest VirB4-containing plasmid was pSci1, with only 13 kb, found in *Spiroplasma citri* GII3. This strain harbors several plasmids for which transmission has been demonstrated, but it is unclear if the plasmid codes for a self-sufficient T4SS (17). Mollicutes contain small conjugative elements measuring much less than 30 kb (23, 98, 107). Since they lack most T4SS genes, one could imagine that they are “degraded” nonconjugative derivatives of ancestral plasmids. However, most mollicutes are not undergoing rapid genome degradation, have few pseudogenes, and have highly coding-dense genomes (125). Since there is evidence that horizontal gene transfer is frequent in the clade (126), it is more parsimonious to think that they code for a simpler conjugative machinery. The largest plasmid encoding a T4SS is pSymA from *Sinorhizobium meliloti* 1021, which has 1.35 Mb. The largest mobilizable plasmid is pSymB, also from *Sinorhizobium meliloti* 1021, with 1.68 Mb. Thus, most of the very large plasmids are not conjugative (Fig. 6 and see Table S1 in the supplemental material).

**Relationship between MOB family, MPF family, and plasmid size.** MOB families were differentially represented among mobilizable and conjugative plasmids. MOB<sub>v</sub> was found almost exclusively among mobilizable plasmids, while MOB<sub>f</sub> and MOB<sub>h</sub> were present almost exclusively in conjugative plasmids. Although we know of no molecular reason that can explain this difference, we speculate that large plasmids might need additional regulation. Accordingly, MOB<sub>f</sub> and MOB<sub>h</sub> relaxosomes are structurally more complex and contain comparably large *oriTs* (>300 bp) (2, 51, 81, 92, 118). On the other hand, MOB<sub>o</sub> and certain MOB<sub>v</sub> relaxosomes are simple, contain short *oriTs* (<100 bp), and are promiscuous in their use of different helper T4SSs as mating channels (49, 52, 100). MOB<sub>c</sub> plasmids are a case apart, since they are mobilizable but code for their own T4CP (22). If we assume that conjugation requires two recognition steps, relaxosome-T4CP and T4CP-T4SS, the genetic composition of MOB<sub>c</sub> plasmids explains why they can use the amplest set of helper T4SSs for mobilization (22), although this fact has to be confirmed in natural settings. The association between classes of relaxases and T4SSs in proteobacteria is far from random (Fig. 7). While MOB<sub>p</sub> is the largest relaxase family, it is absent from MPF<sub>f</sub> plasmids, which show a clear overrepresentation of MOB<sub>f</sub> relaxases and, to a lesser extent, MOB<sub>h</sub>. MOB<sub>c</sub> relaxases are

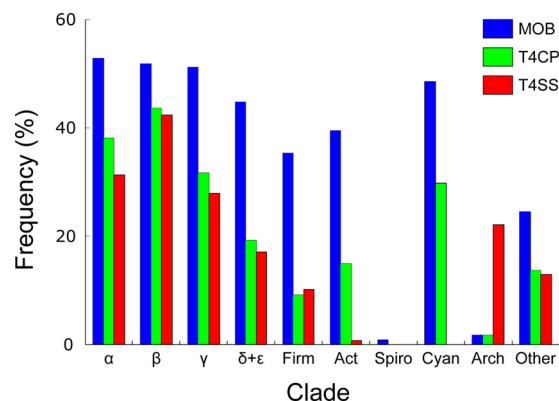


FIG. 8. Frequency with which key elements of plasmid mobility are present in plasmid genomes of different clades. The figure indicates the abundance of relaxases (MOB) in all clades and MPF types in proteobacteria. It also indicates the overall classification of plasmids in terms of mobility.

particularly abundant in MPF-unclassified plasmids, i.e., in plasmids from clades other than proteobacteria. Once again, this suggests large differences between the archetypical plasmids of proteobacteria and the ones of other clades.

#### When Inventories Do Not Fit

**What is the frequency with which one finds the key elements of plasmid transmissibility among different clades?** As expected by the accumulated knowledge of plasmid conjugation, in plasmids there are more relaxases than T4CPs, and there are more of these than T4SSs (Fig. 8). Plasmid transfer has been demonstrated only for four plasmids that do not contain (known) relaxases: the circular plasmid SCP2 of *Streptomyces coelicolor* and the linear plasmid SLP2 of *Streptomyces lividans*, which transfer between mycelia through an FtsK-like protein (18, 84); plasmid pC194 of *Staphylococcus aureus*, which is transferred by the conjugative transposon Tn916 (104); and pKPN2 of *Klebsiella pneumoniae*, which contains two *oriTs* related to plasmids F and R64 and was mobilized by plasmid F (150). Thus, in all four cases, their conjugative behavior can be explained by existing knowledge.

Surprisingly, we found 34 plasmids lacking a T4CP while having VirB4. Most of these were from firmicutes (13 cases), archaea (13 cases), or proteobacteria (7 cases). This might indicate the presence of T4SSs specialized in the transport of proteins rather than in conjugation. Among proteobacteria, all seven plasmids had a recognizable T4SS with conserved VirB4 or TraU homologues, and three of them contained relaxases. The association of T4CP-lacking T4SSs with relaxases suggests that their T4CPs have been recently lost and that these T4SSs were involved in conjugation. Surprisingly, 22% of the archaeal plasmids have VirB4 homologues, but only one has T4CP and relaxase (*Haloarcula marismortui* plasmid pNG500). There are a few plasmids from archaea (e.g., pING and pNOB) that are known to be conjugative (60), and indeed, we found *virB4* homologues in their genomes. However, they do not contain genes coding for a recognizable relaxase or T4CP. Considering that T4CP is the most conserved protein of those analyzed in this work and that it can be found in all other clades containing

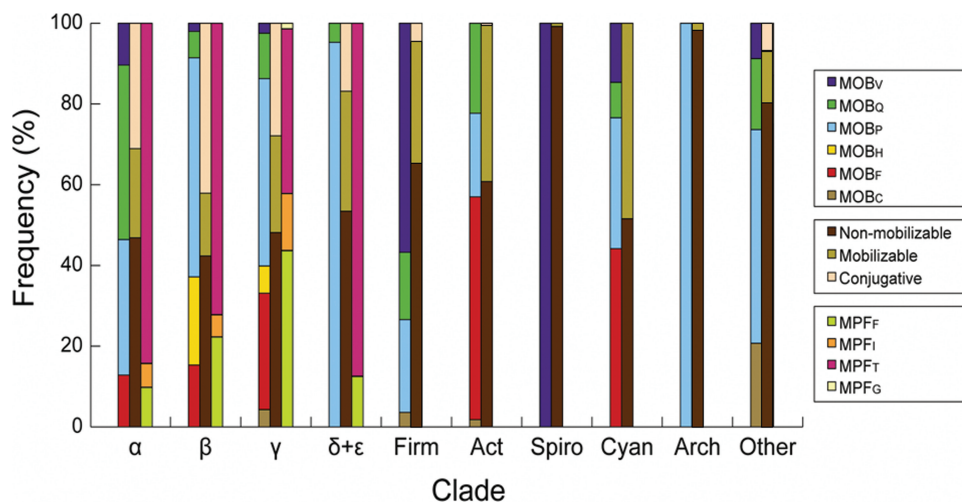


FIG. 9. Clade distribution of plasmid relaxases, T4SSs, and mobility class.

relaxases, this result is puzzling. Given the very large family of NTPases to which T4CPs belong (112), it may be that a more distant homologue than the ones considered in this work performs the T4CP role in archaea (60, 114, 124, 130). These results suggest that T4CPs and relaxases are (radically) different in archaea but still interact with VirB4-containing T4SSs.

Some 90 plasmids encode a T4CP but not VirB4. Of these, some belong to MOB<sub>CI</sub>, the only group of mobilizable plasmids shown to code for their own T4CP (52, 57). Most of the remaining ones correspond to plasmids with uncharacterized T4SSs, such as the cyanobacterial plasmids (see below). In fact, contrary to archaea, in cyanobacteria and actinobacteria, we found many relaxases and T4CPs but few T4SSs. The above-mentioned plasmids of *Streptomyces* contain FtsK-like proteins involved in double-stranded DNA (dsDNA) mobility in mycelia (61). The existence of this alternative mobilization mechanism in *Streptomyces* could be invoked to explain the low number of conjugative plasmids found in actinobacteria. However, our data set contains 147 plasmids of actinobacteria, of which only 25 are from *Streptomyces*. Since many of the other actinobacteria do not produce mycelia, it is unclear why so many plasmids code for relaxases (39%) or T4CPs (15%) but no known T4SS or VirB4. The results concerning cyanobacteria are even more surprising: we found 31 mobilizable plasmids, many of which are larger than 100 kb, but no T4SS. Naturally, if no plasmid is conjugative, the mobilizable plasmids are effectively not mobile. This strongly suggests that cyanobacteria use an as-yet-uncharacterized system to conjugate. This system lacks a clear homologue of *virB4* or any T4SS-specific genes of proteobacteria. To test if more-sensitive sequence searches could detect distant homologues of *virB4*, we built profiles for this protein using hmmer (43a). This allowed the retrieval of very distant homologues of VirB4 in both cyanobacteria and actinobacteria. Experimental work to test the validity of these results is ongoing. If they were correct, then the conjugation systems of cyanobacteria and actinobacteria are very divergent from but phylogenetically related to the VirB4-associated T4SSs of proteobacteria, firmicutes, and archaea.

## SYSTEMATICS AND MOBILITY

We investigated how variants of the key elements of transmissibility, relaxases and T4CPs, are distributed among clades (Fig. 9). Firmicutes lack MOB<sub>F</sub> and MOB<sub>H</sub> but contain an overrepresentation of MOB<sub>V</sub>, which is rare in proteobacteria. Actinobacteria seem to be dominated by the MOB<sub>F</sub> family and in particular by MOB<sub>F2</sub> relaxases; they lack MOB<sub>H</sub> and MOB<sub>V</sub> and have few MOB<sub>C</sub> relaxases. Within proteobacteria, the gamma subdivision contains representatives of the six MOB groups, whereas the alphaproteobacteria lack MOB<sub>H</sub> and MOB<sub>C</sub> and show many MOB<sub>O</sub> relaxases. This is due mainly to MOB<sub>O2</sub> relaxases of tumorigenic and symbiotic plasmids of *Agrobacterium* and rhizobiales. Surprisingly, the association of a T4CP with the relaxases of different MOB families is far from random. While 62% of proteobacterial relaxases have an associated T4CP, this drops to 36% in actinobacteria and 19% in firmicutes. Relaxases of some subfamilies of MOB<sub>F2</sub>, MOB<sub>O1</sub>, MOB<sub>P5</sub> (previously MOB<sub>HEN</sub>), and MOB<sub>V</sub> are never associated with a cognate T4CP.

All four classes of T4SSs were found exclusively in plasmids of proteobacteria, with the exception of *Prosthecochloris aestuarii* plasmid pPAES01 (a member of the *Chlorobi*) and a few plasmids from noncultivated bacteria for which the phylogenetic classification of the host is unknown. While MPF<sub>T</sub> clearly predominates in most clades, MPF<sub>F</sub> is equally abundant in gammaproteobacteria (Fig. 9). MPF<sub>I</sub> is less frequent, being more abundant in gammaproteobacteria and absent from delta- and epsilonproteobacteria, possibly due to a small sample size. The finding that prototypical conjugative T4SSs are exclusive of proteobacteria is puzzling because some of these plasmids can conjugate into cells of very distant clades (99, 139), such as cyanobacteria (75), where we found no such types of plasmids. Plasmids of firmicutes have also been shown to transfer to proteobacteria and actinobacteria (79, 143). Hence, our results suggest that although some proteobacterial plasmids have a broad host range, they tend to reside in one bacterial clade.

Overall, these analyses suggest that our understanding of

plasmid mobility is excellent for proteobacteria and poor for most other clades, including clades as environmentally important as archaea, cyanobacteria, and actinobacteria. They also show a complex evolutionary history, with different clades sharing or lacking different components of the conjugation machinery. We therefore set up an evolutionary analysis of the coevolution between mobilization and conjugation.

### COEVOLUTION OF MOBILIZATION AND CONJUGATION

**Phylogeny of T4CPs and VirB4 in the light of relaxases.** The T4CP is the most informative protein of the conjugation machinery because it is a large, highly conserved protein with homologues in almost all conjugative plasmids. Furthermore, the phylogeny of T4CPs matched that of relaxases closely (57). We selected 184 T4CPs sharing less than 95% similarity to reduce sampling redundancy. The phylogenetic tree of the T4CPs confirmed on a larger scale that relaxases and T4CPs evolve in parallel, since large clusters in the T4CP tree corresponded almost invariably to a similar relaxase family or to an absence of known relaxases (Fig. 10). The VirB4 tree showed a more scattered distribution of relaxase families, especially among MPF<sub>T</sub> plasmids. From a functional point of view, although T4CPs interact with components of both the relaxosome and the T4SS, the T4CP-relaxosome interactions seem to be more system specific (94, 96, 97, 101). Besides, T4CP-encoding genes are usually adjacent to the genes coding for the relaxosome components (50, 57), with gene order conservation being another sign of coevolution and functional relatedness. In the light of these results, it is becoming evident that T4CPs are part of the conjugative DNA-processing module (what we call in this review the MOB module). There are, however, some exceptions to this rule, mostly concerning the MOB<sub>P</sub> family, showing a less tight linkage between MOB and T4CP. The ample diversity within the MOB<sub>P</sub> family might be the cause of this weaker association, which is consistent with the clustering of these relaxases into separate groups based on global sequence similarity (see Fig. S2 in the supplemental material). Interestingly, plasmids without known relaxases are often clustered together in both phylogenetic trees. It is tempting to view this as a further indication of the existence of unknown novel relaxases in some clades.

**T4CP and VirB4 subgroups follow prokaryotic phylogeny.** The basal positions in the T4CP tree show a few plasmids grouped into three clusters corresponding to MPF<sub>I</sub>-containing plasmids, a variety of MPF<sub>T</sub> plasmids from different phylogenetic origins containing all MOB<sub>C</sub> relaxases, and a group of MOB<sub>H</sub>-containing MPF<sub>T</sub> and MPF<sub>F</sub> plasmids (Fig. 10). The inner groups are constituted essentially by two clusters of MOB<sub>F</sub>-containing plasmids, separating MOB<sub>F</sub> plasmids from cyanobacterial plasmids lacking identifiable T4SSs, and a somewhat intermingled group of MOB<sub>P5</sub>- and MOB<sub>O</sub> (or both)-associated plasmids. Figure 10 (middle) shows the phylogenetic origin of the plasmid hosts containing T4CP or VirB4. It is apparent from Fig. 10 that the large clusters of T4CP sequences tend to correspond to coherent taxonomic clusters. While this might seem surprising given the mobile nature of plasmids and, in particular, of broad-host-range plasmids, it is perfectly compatible with our analysis of gene rep-

ertoires of T4SSs, showing a strict association of archetypical systems with proteobacteria. In fact, the only groups containing large phylogenetic diversity, e.g., MOB<sub>C</sub>-associated T4CPs, correspond to branches rooting very deeply in the tree. In our analysis, closely related genes usually correspond to plasmids in closely related clades. For example, archaeal VirB4s, for which we could find no matching T4CP, cluster together, and so do cyanobacterial and actinobacterial T4CPs, for which we could not find T4SSs. Hence, these results further suggest an important evolutionary inertia in the mechanisms of plasmid mobility, where mobility between distant clades is sporadic and short-lived.

**VirB4 phylogeny matches T4SS archetypes.** The VirB4 tree is highly concordant with our classification of the archetypical T4SSs (Fig. 10). As mentioned above, TraU and VirB4 trees are independent because the similarities between the two families are too weak. The early divergence of MPF<sub>I</sub> is consistent with the basal position of the T4CP associated with MPF<sub>I</sub> in the T4CP tree. MPF<sub>I</sub> plasmids are known to produce a functionally idiosyncratic type of pilus. They can mate in liquid medium, so in this respect, they bear resemblance to MPF<sub>F</sub>. However, they contact other cells by a specific fimbrial system related to type 2 secretion (76). This originality might be the hallmark of an early divergence or independent origin from the remaining conjugative systems.

Among the remaining plasmids, there is a neat division of the VirB4 proteins into two groups. One group includes MPF<sub>T</sub> and MPF<sub>F</sub>, suggesting that the divergence between these two classes of T4SSs was comparatively recent, in agreement with the higher level of similarity of their T4SS loci (25, 123). Interestingly, there are some MPF<sub>T</sub>/MOB<sub>P11</sub> plasmids on the side of the MPF<sub>F</sub> group at the basal position (Fig. 10). This finding suggests that MPF<sub>F</sub> and MPF<sub>G</sub> derived from the same lineage of MPF<sub>T</sub>. More plasmid sequences are required to ascertain this branching that, if correct, suggests that the ancestral proteobacterial VirB4-based conjugation system was MPF<sub>T</sub>. It is tempting to speculate that the ancestral MPF<sub>T</sub>-containing plasmids, exclusively surface maters, led to the more-complex MPF<sub>F</sub> systems by the acquisition of additional functions that allow plasmids to mate in liquid. Indeed, the other related T4SSs found in firmicutes also seem to carry fewer genes than MPF<sub>F</sub> systems. On the other hand, MPF<sub>T</sub> systems include plasmids with a broad host range (140) and that are capable of retrotransfer (134). Thus, along with the capacity to mate in liquid medium, MPF<sub>F</sub> evolved toward a narrower host range. This may represent an inevitable tradeoff between the complexity of the conjugation machinery and plasmid host range, since the known complex, independently derived, liquid-mating MPF<sub>I</sub> systems also have a narrow host range (33).

The presence of a given T4SS cannot be trivially correlated with the life-style of the host. For instance, the genus *Vibrio* is composed of planktonic organisms that thrive in brackish waters. Out of the six *Vibrio* plasmids listed in Table S1 in the supplemental material that contain mobility genes, *V. harveyi* plasmid pVIBHAR and *V. fischeri* plasmid pES100 contain MPF<sub>T</sub>, while *V. vulnificus* plasmids pC4602-1 and pYJ016 and *V. tapetis* plasmid pVT1 contain MPF<sub>F</sub>. Finally, *V. fischeri* plasmid pMJ100 lacks relaxases but contains a T4CP and MPF<sub>I</sub>. Thus, there are examples of all three MPF systems in *Vibrio*

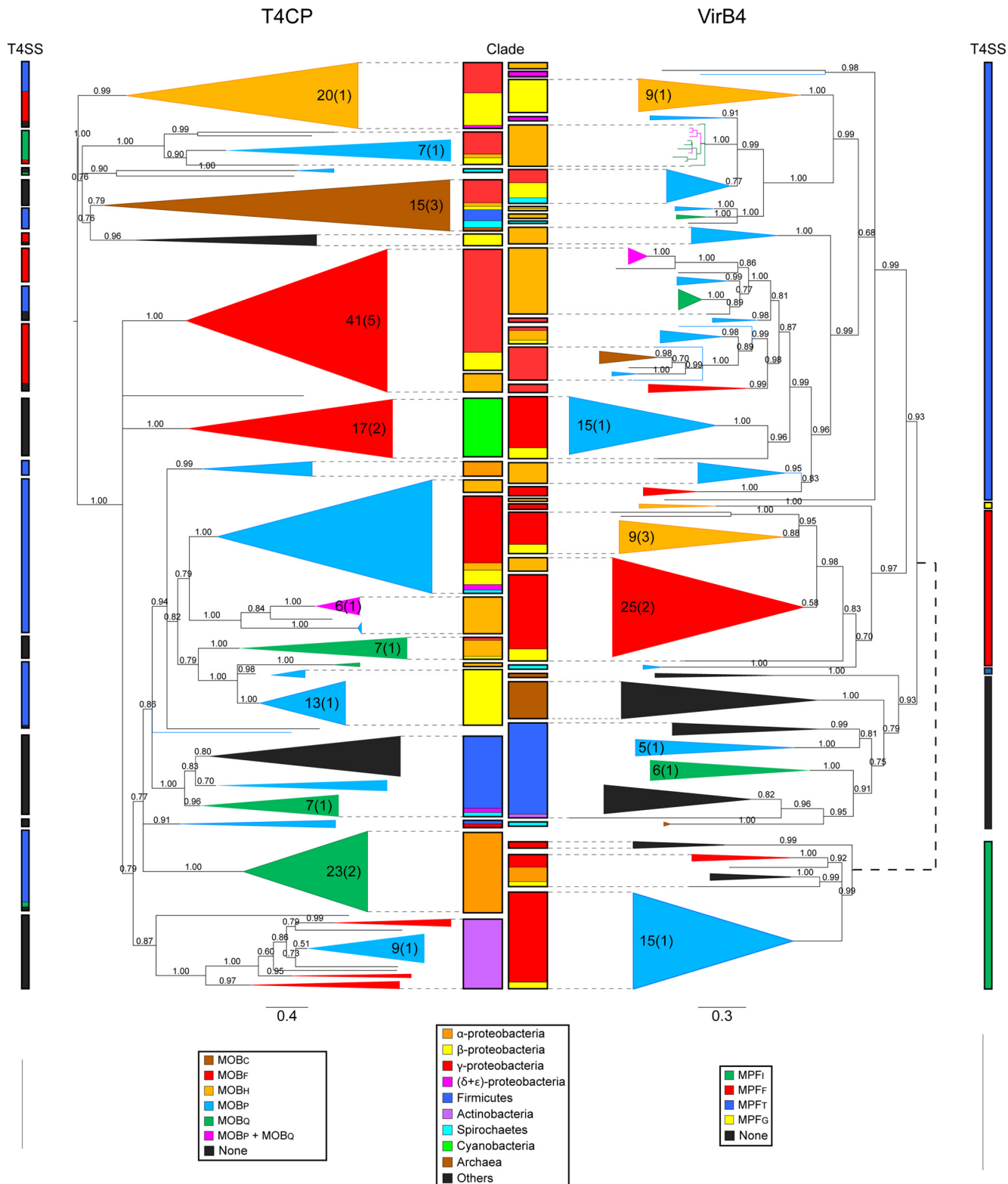


FIG. 10. Phylogenetic analysis of the T4CP family (left) compared to the T4SS main ATPase (VirB4 and TraU) (right). The middle panels indicate the phylogenetic origin of the plasmids harboring the T4CPs and the T4SSs. The phylogenetic trees were cartooned; i.e., tips were transformed into triangles whose height is proportional to the number of proteins and its depth is proportional to the most-deep-branching element, for clarity, when the groups were consistent and when support for the branches was high. When some large triangles included a few elements that do not have the same relaxase family, they are indicated by numbers [e.g., 41(5) indicates that out of 41 proteins, 5 have a different classification]. Colors in the tree cartoons correspond to families of relaxases. Numbers in nodes correspond to the values of the aLRT statistical test of branch support (6). The lateral panels represent the distribution of the prototypical T4SS in the trees. To reduce computational time, sequences with  $\geq 95\%$  identity were collapsed into a single sequence corresponding to the longest in the set. MUSCLE v3.6 (44) was used to make multiple alignments (default settings), and poorly aligned columns of the alignment were removed by visual inspection using Seaview (56). Phym1 v3.0 was used to compute the phylogenies using maximum likelihood [model WAG+ $\Gamma(6)$ +I] (64).

plasmids. As a case in point, the best-studied bacterium, *E. coli*, has plasmids with all three MPF types. There is a lack of studies on the possible relative advantages of each mating type in a given environment.

Finally, there is a large congruent group of VirB4 proteins in plasmids from archaea and firmicutes (plus other Gram-positive bacteria), with the two groups being well-separated. This strongly suggests that these systems are more closely related among themselves than with the T4SSs of proteobacteria. Since firmicutes are more distantly related to archaea than to other bacteria, these results suggest an origin of VirB4-based MPFs dating from after the divergence of archaea and bacteria. These MPFs would have then spread among prokaryotes by horizontal gene transfer, possibly by conjugation. The close relationship between VirB4 homologues of firmicutes and archaea might then result from the adaptation of a lineage of ancestral T4SSs that was particularly plastic and capable of adapting to different cellular envelope structures.

### HOW MOBILE ARE NONMOBILIZABLE PLASMIDS?

Most evolutionary models assume that plasmids must be mobile to persist because plasmid carriage incurs a fitness cost (39, 78, 89, 91). Plasmids are regarded as selfish genetic elements because they can spread and survive without necessarily increasing their host fitness (7, 113). In this light, how can one interpret the existence of so many nontransmissible plasmids? Some of the very small plasmids may be remnants of larger replicons that lost genetic information and thus remain transiently carried in spite of being nonmobile. These plasmids would be expected to be on their way to extinction. However, since our survey uncovered that more than half of the plasmids are in fact nontransmissible, this explanation seems unsatisfactory.

Plasmids might move between cells by processes other than direct conjugation, e.g., by transduction, natural transformation, or cointegration in mobile plasmids. However, these mechanisms are logically expected to affect the transfer of plasmids at much lower rates than conjugation. They are also expected to favor small plasmids. Transformation requires the presence of full plasmids in the environment in spite of environmental nucleases, their entry into the cell, and eventually their reconstitution there, when transformation includes a preliminary step of DNA restriction. Transduction requires plasmids to be packaged and fit into a phage capsid. Therefore, transduction is restricted to plasmids with genomes the size of, or smaller than, the phage. Transfer by cointegration in other replicons requires the presence of other mobilizable elements and integration mechanisms. Since half of the plasmids are nonmobilizable, including the largest ones, these results seriously question the purely parasitic nature of all plasmids, which require high rates of transfer (13, 91).

It follows that the persistence of nonmobile plasmids has to be explained by the potentially useful genes (for the host) that they carry. Plasmids of *Borrelia* take up a significant part of their cellular genomes (26). Plasmids of these intracellular obligate parasites might have rare opportunities for plasmid transfer. The 39 different virulence plasmids of *Borrelia garinii* recombine frequently (32), and selection for gene conversion by antigenic variation might offset the cost of the loss of the

capacity for frequent horizontal transfer by conjugation. There is also evidence of some exchange of genetic material between *Borrelia* species, since both plasmid and chromosomal sequences show marks of homologous recombination between different lineages (115). Some plasmids of *Borrelia*, the cp32 family, transfer by transduction (27, 151). These plasmids might then have a dual nature, like the bacteriophage P1 plasmid of *E. coli*; that is, they are maintained in cells like plasmids and transfer like phages (121). Plasmids in *Buchnera* species have most likely been kept without horizontal transfer but with exchanges with the chromosome for millions of years, possibly to increase the gene dosage of genes involved in mutualism with aphids (80). Traits encoded in mobile elements are both functionally and ecologically peculiar, and they might be durably and adaptively associated with plasmids (116). There is thus some evidence suggesting that plasmids with little or no mobility may survive for long periods of time by natural selection.

### Are Large Plasmids Becoming Secondary Chromosomes?

Surprisingly, most plasmids larger than 300 kb are nonmobilizable (Fig. 6). Such large, coding-dense plasmids are expected to have an important impact on cell fitness, and given their size, they are unlikely to be transferred by transformation or transduction. Their loss can be precluded only if their presence is selected for in the long term. One is then tempted to suggest that many of such very large plasmids are in the way of being domesticated into secondary chromosomes (recently called chromids [67]). In multichromosomal bacteria, occurring among clades such as *Rhizobium*, *Burkholderia*, and *Vibrio* (108), secondary chromosomes are smaller, contain few essential genes, and code for niche-specific functions (45, 67, 127). Some of these chromosomes contain a plasmid-like origin of replication, e.g., chromosome 2 in *Vibrio cholerae* resembles the *oriVs* of P1 and F plasmids (46) and replicates at a different moment in the cell cycle (117).

While there is no consensual definition of the difference between a chromosome and a plasmid, the former is often seen as any replicon containing essential housekeeping genes (108). Many large plasmids carry essential RNA genes, such as tRNA and rRNA (Fig. 11). In *Deinococcus geothermalis* and *Ralstonia solanacearum*, some chromosomal copies of the 16S rRNA genes are more similar to plasmid copies than to other chromosomal ones. In general, intact rRNA gene copies present in plasmids are as identical to the copies in chromosomes as they are among themselves, showing in all cases more than 99% identity. Since the threshold identity commonly used to define species in prokaryotes from 16S copies is 97% (129), this strongly suggests that these plasmids carry rRNA genes acquired from the species that hosted them at the moment of natural isolation. Our failure to observe rRNA genes significantly different from the chromosomal copies is a further indication that these plasmids have little, if any, mobility.

Genes coding for essential proteins are nearly always absent from small plasmids (Fig. 11) but are present in ~60% of large plasmids (100 to 400 kb) and 90% of very large plasmids (>400 kb). Highly persistent genes, i.e., genes present in most strains, many of which are not strictly essential (48), are also frequently found in a small number in large replicons (67). Contrary to

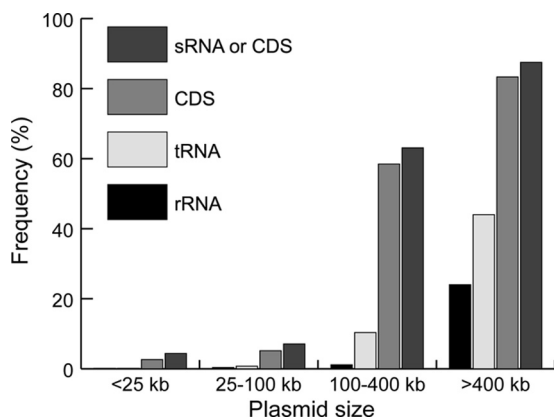


FIG. 11. Presence of tRNA, rRNA, or protein-encoding genes best homologous to *E. coli* or *B. subtilis* essential genes in plasmids classified according to genome size. Small plasmids (<25 kb) rarely contain such genes, whereas very large plasmids (>400 kb) often contain them. We searched for homologues of essential *B. subtilis* genes in plasmids of firmicutes and for those of *E. coli* in the remaining plasmids. If the resident chromosome had a homologue more similar to the essential gene than the plasmid homologue, the plasmid hit was excluded from the analysis.

tRNA and rRNA, genes encoding essential or persistent proteins rarely have multiple copies in a genome, and therefore, there is no possibility of a chromosomal compensation for the plasmid loss. Naturally, the existence of essential genes in a plasmid does not preclude plasmid mobility to other cells; it just prevents its segregation from the host. The data presented in this review suggest that most very large plasmids cannot be segregated without large negative fitness effects. In fact, even very large plasmids that are conjugative tend to show little mobility. For example, the large symbiotic plasmid of *Rhizobium leguminosarum* (>1 Mb) is conjugative but shows little mobility in natural populations (148). As a case in point, the largest plasmid of three strains of *Rhizobium leguminosarum* bv. trifolii, which is nonmobilizable in our analysis, could not be cured (8).

Plasmids tend to be A+T rich relative to the host (119) and have different oligonucleotide frequencies (144). However, large plasmids and secondary chromosomes tend to have compositions closer to that of the primary chromosome (67, 131). This is interpreted as the result of the long-standing presence of the plasmid in the host lineage, leading to the plasmid genome amelioration to the chromosomal composition (83), since the replication machineries are shared between replicons. It was recently proposed that secondary chromosomes and most very large plasmids should be called “chromids,” defined as replicons with plasmid-like maintenance and replication systems, nucleotide composition close to that of the primary chromosome, and carriage of some essential or persistent genes (67). It has been proposed that chromids allow bacteria to have a larger genome without incurring a chromosome replication time penalty leading to slower growth (149). However, bacterial replication is uncoupled from cell duplication, and the existence of multiple simultaneous rounds of chromosomal replication renders the genome partition into multiple replicons unnecessary. Furthermore, there is no association between the maximal growth rate and genome size

(146). Accordingly, few of the fastest-growing and a number of the slow-growing bacteria have chromids (29).

It might be thought that large domesticated plasmids would be doomed to disappear because genes coding for adaptive traits could migrate to the chromosome, rendering the costly and nonmobile plasmid useless to the cell. Nevertheless, it has been proposed that the above-mentioned highly dynamic *Borrelia* plasmids do not integrate into the chromosome to stabilize chromosome organization (120). In itself, this might lead to a longstanding presence of plasmids, but plasmids also co-evolve to impose low or no cost at all to hosts (14, 40). A plasmid conferring an adaptive advantage to the host does not require the maintenance of costly functions such as surface exclusion, poison-antidote systems, or even mobility. For example, both the fusion of *Sinorhizobium meliloti* replicons (65) and division of the *Bacillus subtilis* genome (70) showed little effect on growth in rich media. Overall, these results are consistent with the idea that very large nonmobilizable plasmids are in the way of becoming secondary chromosomes.

## TOWARDS COMPARATIVE GENOMICS OF PLASMIDS

One of the major contributions that we expect from this review is the demonstration that many conjugative transfer systems can be found and classified just by searching for similarities to known relaxases, T4CPs and T4SS main ATPase sequences (VirB4). The use of cluster techniques allows matching the expert work usually involved in the classification of relaxases and T4SSs with great accuracy. The simplicity of the methodology will allow anyone to reuse it to characterize newly sequenced plasmids. In this respect, we are now creating ready-to-use hidden Markov model-based protein domains specific to each of the proteins analyzed in this study to make them available on a dedicated website. However, for plasmid comparative genomics to take off as an important part of plasmid biology, there are still several hindrances. These hindrances relate to resources for data storage and analysis and how these might allow using the extensive information on individual plasmids to understand the biology of other plasmids. Some effort has been done by comprehensive databases such as ACLAME (87). However, such databases are often short-lived or updated at a much slower pace than GenBank, as is the case of the Plasmid Genome Database (102), the Database of Plasmid Replicons, or the Genome Database of Naturally Occurring Plasmids. The ideal plasmid database should also establish links between the plasmid and its host chromosomes (or range thereof) and provide a rationale for functional classification. The ACLAME (87) database provides these links and a growing gene ontology that facilitates the expert analysis of plasmid sequence data from a systems biology perspective.

Gene nomenclature can also defeat the neophyte enthusiasm for plasmid biology. It is confusing that sets of orthologous genes can have different names in different plasmids. For example, VirB4 is called over 50 different names in the database, among which are VirB4, TrbE, TraE, TraC, MpfC, and TraB, etc., not to mention the distant homologue TraU family. Inversely, different plasmids have similarly named genes that have no relation of homology. This is largely a historical problem. While some genes were named by their order in the

mobility locus, other genes were named by homology, and yet others, because homology was not obvious, were named yet differently. It is also not intuitive to name one given T4SS—MPF<sub>I</sub>—as a type 2 conjugative pilus or to name T2SSs as “type IV pili.” Rationalization of the nomenclature in the field would facilitate comparative approaches but naturally requires a community-level effort in establishing new names and assessing synonyms. While care must be taken in gene renaming, as sequence similarity may not be an indication of an exact function analogy, the current naming of genes carried by plasmids makes genome comparison using annotation files impossible. Our plasmid mobility classification scheme can be automatically generated by sequence similarity searches followed by clustering. This allows analyses such as the ones that we presented here but also allows analyses of environmental plasmid sets, such as those found in metagenomic data sets (12, 110, 122, 132, 133).

We now have an automated method for the systematic classification of relaxases, T4CPs, and T4SSs that allows the monitoring of the evolutionary patterns of plasmids in phylogenetic trees. The evolutionary histories of relaxases, T4CPs, and T4SSs are not fully congruent but suggest coevolution. The relaxases can thus be considered markers for plasmid classification in a way resembling the 16S RNA sequence in genomes, with the exception that around half of the plasmids lack relaxases and that different relaxase families exist. Still, MOB is a broader phylogenetic marker than origins of replication used to define Inc types (30), which evolve too quickly to obtain deep phylogenetic trees. The T4SSs are also good phylogenetic markers but are more limited in that T4SSs are absent from most plasmids. The T4CP is an excellent marker because it so far consists of one single protein family, allowing the establishment of the deepest evolutionary relationships. An understanding of the evolution of the conjugation machinery in clades other than proteobacteria and firmicutes will certainly be gained by the study of T4CP evolution and its interaction with the relaxosome and the T4SS.

## CONCLUSION

Within proteobacteria, the 98% overlap between VirB4-based and specific T4SS-based searches suggests that we have attained a good knowledge of the conjugative pilus that can be used to automatically classify plasmid mobility. Our analysis suggests that the classical division of T4SSs into two groups, T4SSa and T4SSb (31), should be revised into a classification into four groups, one of which (MPF<sub>G</sub>) is rare and another of which (MPF<sub>T</sub>) is by far the most frequent. However, one must emphasize that such a classification scheme is applicable only to plasmids of proteobacteria. Outside this clade, our analyses suggest that new relaxases and T4SSs remain to be found. Some clades, notably in archaea, have known mobilizable plasmids containing a putative T4SS but no relaxase or T4CP. Some plasmids contain a small but coherent set of mobility genes of low homology with known relaxases, such as the above-mentioned new family represented by *Bacteroides thetaiotaomicron* VPI-5482 plasmid p5482. Other clades, notably cyanobacteria, contain many plasmids with relaxases and T4CPs but no close VirB4 homologues. Given that cyanobacterial cells are among the most abundant on earth and that

cyanobacteria genomes have many plasmids, uncovering how plasmids spread in this clade should become a priority. Indeed, the 408-kb *Anabaena* MOB<sub>V</sub> plasmid pCC7120 $\alpha$  was reported to be transmissible (103), although the exact mechanism of transmission was not investigated. As mentioned above, the VirB4 analogue of MPF<sub>I</sub> is less similar to VirB4 than to VirD4. It is thus likely that other unknown types of proteins energize unknown types of T4SSs. Mining for relaxases and T4SSs in large genomes is therefore bound to produce novel families. This should be done in connection with the identification of ICEs in genomes, which remain largely unexplored by comparative genomics.

It is tempting to relate the current caveats in relaxases and T4SSs with our results showing that known broad-host-range proteobacterial plasmids are in fact found only in proteobacteria. This reinforces previous observations that the conjugation range of plasmids is larger than the range of hosts that they typically occupy (34) and that plasmids tend to have nucleotide compositions close to that of the host where they are often found (131). While some plasmids can conjugate between very different clades, they are not naturally found there (at least not significantly often), and our phylogenetic analysis suggests that the diversification of mobility-associated protein families takes place in narrower clades. Plasmids do not shuffle modules freely; they tend to cluster within given clades, and this preference will somehow be related to specific features of a given plasmid design and with the host physiology (50). As a case in point, the T4SSs of firmicutes seem smaller than those of proteobacteria, possibly because of the lack of an outer membrane in these cells. Elements of the T4SS with homologues between proteobacteria and firmicutes are found to interact in equivalent ways (1), but the remaining machinery may work very diversely; e.g., T4SSs of firmicutes are not known to form a mating pilus (1). The T4SSs of mollicutes, lacking an outer membrane and a cell wall, seem even simpler. While conjugation systems have certainly adapted to the peculiarities of bacterial membranes, it makes little sense in opposing Gram-positive and Gram-negative plasmids. As a case in point, T4SSs of plasmids of proteobacteria seem to have more in common with the ones of firmicutes than with the ones of cyanobacteria, which also have two membranes. Thus, the emerging picture that seems to arise is that conjugation systems are adapted to taxonomic clades, and more research on differences between plasmids of proteobacteria, firmicutes, actinobacteria, cyanobacteria, and archaea, etc., should be carried out. As a result of these adaptive processes, few elements of the conjugation machinery are common between firmicutes and proteobacteria, and even fewer are common between cyanobacteria and archaea.

Plasmid design or host adaptation is also likely to account for phylogenetic specificities in MOB and MPF modules. This is analogous to the evolution of prokaryotic chromosomes. Although rampant recombination could have taken place and eroded phylogenetic lineages, this has not happened, because some core genes are rarely successfully exchanged (88, 141). What makes prokaryotic classification useful and meaningful appears to do the same job in plasmid classification, respecting mobility systems most likely because of the adaptive coevolution of the different elements of the mobility machinery with the host.

According to the interpretation of the sequence analysis carried out in this work, we found 15% conjugative, 24% mobilizable, and 61% nontransmissible plasmids in prokaryotes. In proteobacteria, for which our predictions are more accurate, the percentages are not so different, 28%, 23%, and 49%, respectively. Thus, about half of the plasmids are nontransmissible, and the remaining ones are divided more or less evenly between conjugative and mobilizable plasmids. This finding suggests that many evolutionary models of plasmid evolution requiring high rates of horizontal spread for plasmid survival might have to be revised to account for the fact that at least half of the plasmids probably have low transfer rates. Our phylogenetic analysis shows rare transfer between distant phyla and describes the evolutionary history of T4SSs. MPF<sub>T</sub> is by far the most abundant T4SS, and MPF<sub>G</sub> and MPF<sub>F</sub> might derive from an ancestral MPF<sub>T</sub>. However, MPF<sub>I</sub> and T4SSs from other clades have not derived from MPF<sub>T</sub>, which seems to be a proteobacterial invention. The high frequency of MPF<sub>T</sub> occurrence might reflect a particularly successful design, even though such plasmids are notoriously poorly functioning for mating in liquid media (16, 38). It might also result from some bias toward sequencing host-associated proteobacteria. The incoming metagenomic data should be able to provide more-accurate and less-biased estimates of the diversity and frequency of the different types of MPF and MOB. In any case, a full account of the evolutionary history of conjugation will require a parallel study of ICEs and the uncovering of conjugation mechanisms in prokaryotes lacking identifiable T4SSs and/or relaxases, that is, the vast majority of prokaryotes.

#### ACKNOWLEDGMENTS

Work in the laboratory of F.D.L.C. was supported by grants BFU2008-00995/BMC (Spanish Ministry of Education), REIPI RD06/0008/1012 (RETICS Research Network, Instituto de Salud Carlos III, Spanish Ministry of Health), and LSHM-CT-2005\_019023 (European VI Framework Program). Research in the laboratory of E.P.C.R. is supported by the CNRS and the Institut Pasteur. C.S. was supported by a BRAVO fellowship from the University of Arizona (HHMI grant 52005889). M.V.F.'s research was supported by grants FIS PI07/0664 (Spanish Fondo de Investigación Sanitaria, Instituto de Salud Carlos III) and CES08/008 (Spanish Fondo de Investigación Sanitaria, Instituto de Salud Carlos III, and Fundación Marqués de Valdecilla-IFIMAV), REIPI RD06/0008/0031 (Ministerio de Sanidad y Consumo, Instituto de Salud Carlos III), and LSHE-CT-2007-037410 (European VI Framework Program).

We thank Pascal Sirand-Pugnet for comments on the manuscript.

#### REFERENCES

- Abajy, M. Y., J. Kopec, K. Schiwon, M. Burzynski, M. Doring, C. Bohn, and E. Grohmann. 2007. A type IV-secretion-like system is required for conjugative DNA transport of broad-host-range plasmid pIP501 in gram-positive bacteria. *J. Bacteriol.* **189**:2487–2496.
- Aho, T., and E. Ohtsubo. 1995. Characterization of the functional sites in the oriT region involved in DNA transfer promoted by sex factor plasmid R100. *J. Bacteriol.* **177**:4350–4355.
- Achtman, M. 1975. Mating aggregates in *Escherichia coli* conjugation. *J. Bacteriol.* **123**:505–515.
- Altschul, S. F., T. L. Madden, A. A. Schäfer, J. Zhang, Z. Zhang, W. Miller, and D. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.
- Alvarez-Martinez, C. E., and P. J. Christie. 2009. Biological diversity of prokaryotic type IV secretion systems. *Microbiol. Mol. Biol. Rev.* **73**:775–808.
- Anisimova, M., and O. Gascuel. 2006. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.* **55**:539–552.
- Bahl, M. I., L. H. Hansen, and S. J. Sorensen. 2009. Persistence mechanisms of conjugative plasmids. *Methods Mol. Biol.* **532**:73–102.
- Baldani, J. I., R. W. Weaver, M. F. Hynes, and B. D. Eardly. 1992. Utilization of carbon substrates, electrophoretic enzyme patterns, and symbiotic performance of plasmid-cured clover rhizobia. *Appl. Environ. Microbiol.* **58**:2308–2314.
- Barlow, M. 2009. What antimicrobial resistance has taught us about horizontal gene transfer. *Methods Mol. Biol.* **532**:397–411.
- Bates, S., A. M. Cashmore, and B. M. Wilkins. 1998. IncP plasmids are unusually effective in mediating conjugation of *Escherichia coli* and *Saccharomyces cerevisiae*: involvement of the tra2 mating system. *J. Bacteriol.* **180**:6538–6543.
- Beranek, A., M. Zettl, K. Lorenzoni, A. Schauer, M. Manhart, and G. Koraimann. 2004. Thirty-eight C-terminal amino acids of the coupling protein TraD of the F-like conjugative resistance plasmid R1 are required and sufficient to confer binding to the substrate selector protein TraM. *J. Bacteriol.* **186**:6999–7006.
- Beres, S. B., and J. M. Musser. 2007. Contribution of exogenous genetic elements to the group A *Streptococcus* metagenome. *PLoS One* **2**:e800.
- Bergstrom, C. T., M. Lipsitch, and B. R. Levin. 2000. Natural selection, infectious transfer and the existence conditions for bacterial plasmids. *Genetics* **155**:1505–1519.
- Bouma, J. E., and R. E. Lenski. 1988. Evolution of a bacteria/plasmid association. *Nature* **335**:351–352.
- Bradley, D. E. 1984. Characteristics and function of thick and thin conjugative pili determined by transfer-derepressed plasmids of incompatibility groups I1, I2, I5, B, K and Z. *J. Gen. Microbiol.* **130**:1489–1502.
- Bradley, D. E., D. E. Taylor, and D. R. Cohen. 1980. Specification of surface mating systems among conjugative drug resistance plasmids in *Escherichia coli* K-12. *J. Bacteriol.* **143**:1466–1470.
- Breton, M., S. Duret, N. Arricau-Bouvery, L. Beven, and J. Renaudin. 2008. Characterizing the replication and stability regions of *Spiroplasma citri* plasmids identifies a novel replication protein and expands the genetic toolbox for plant-pathogenic spiroplasmas. *Microbiology* **154**:3232–3244.
- Brolle, D. F., H. Pape, D. A. Hopwood, and T. Kieser. 1993. Analysis of the transfer region of the *Streptomyces* plasmid SCP2. *Mol. Microbiol.* **10**:157–170.
- Bron, S., and E. Luxen. 1985. Segregational instability of pUB110-derived recombinant plasmids in *Bacillus subtilis*. *Plasmid* **14**:235–244.
- Bron, S., E. Luxen, and P. Swart. 1988. Instability of recombinant pUB110 plasmids in *Bacillus subtilis*: plasmid-encoded stability function and effects of DNA inserts. *Plasmid* **19**:231–241.
- Buchrieser, C., P. Glaser, C. Rusniok, H. Nedjari, H. D'Hauteville, F. Kunst, P. Sansonetti, and C. Parsot. 2000. The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of *Shigella flexneri*. *Mol. Microbiol.* **38**:760–771.
- Cabezon, E., J. I. Sastre, and F. de la Cruz. 1997. Genetic evidence of a coupling role for the TraG protein family in bacterial conjugation. *Mol. Gen. Genet.* **254**:400–406.
- Calcutt, M. J., M. S. Lewis, and K. S. Wise. 2002. Molecular genetic analysis of ICEF, an integrative conjugal element that is present as a repetitive sequence in the chromosome of *Mycoplasma fermentans* PG18. *J. Bacteriol.* **184**:6929–6941.
- Cao, T. B., and M. H. Saier, Jr. 2001. Conjugal type IV macromolecular transfer systems of Gram-negative bacteria: organismal distribution, structural constraints and evolutionary conclusions. *Microbiology* **147**:3201–3214.
- Cascales, E., and P. J. Christie. 2003. The versatile bacterial type IV secretion systems. *Nat. Rev. Microbiol.* **1**:137–149.
- Casjens, S., N. Palmer, R. van Vugt, W. M. Huang, B. Stevenson, P. Rosa, R. Lathigra, G. Sutton, J. Peterson, R. J. Dodson, D. Haft, E. Hickey, M. Gwinn, O. White, and C. M. Fraser. 2000. A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*. *Mol. Microbiol.* **35**:490–516.
- Casjens, S., R. van Vugt, K. Tilly, P. A. Rosa, and B. Stevenson. 1997. Homology throughout the multiple 32-kilobase circular plasmids present in Lyme disease spirochetes. *J. Bacteriol.* **179**:217–227.
- Cohen, S. N., A. C. Chang, H. W. Boyer, and R. B. Helling. 1973. Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **70**:3240–3244.
- Couturier, E., and E. P. C. Rocha. 2006. Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Mol. Microbiol.* **59**:1506–1518.
- Couturier, M., F. Bex, P. L. Bergquist, and W. K. Maas. 1988. Identification and classification of bacterial plasmids. *Microbiol. Rev.* **52**:375–395.
- Christie, P. J., K. Atmakuri, V. Krishnamoorthy, S. Jakubowski, and E. Cascales. 2005. Biogenesis, architecture, and function of bacterial type IV secretion systems. *Annu. Rev. Microbiol.* **59**:451–485.
- Dai, Q., B. I. Restrepo, S. F. Porcella, S. J. Raffel, T. G. Schwan, and A. G. Barbour. 2006. Antigenic variation by *Borrelia hermsii* occurs through recombination between extragenic repetitive elements on linear plasmids. *Mol. Microbiol.* **60**:1329–1343.
- Datta, N., and R. W. Hedges. 1972. Host ranges of R factors. *J. Gen. Microbiol.* **70**:453–460.



34. De Gelder, L., J. M. Ponciano, P. Joyce, and E. M. Top. 2007. Stability of a promiscuous plasmid in different hosts: no guarantee for a long-term relationship. *Microbiology* **153**:452–463.
35. de la Cruz, F., and J. Davies. 2000. Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends Microbiol.* **8**:128–133.
36. de la Cruz, F., L. S. Frost, R. J. Meyer, and E. Zechner. 2010. Conjugative DNA metabolism in Gram-negative bacteria. *FEMS Microbiol. Rev.* **34**:18–40.
37. Delcher, A. L., K. A. Bratke, E. C. Powers, and S. L. Salzberg. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**:673–679.
38. Dennison, S., and S. Baumberg. 1975. Conjugational behaviour of N plasmids in *Escherichia coli* K12. *Mol. Gen. Genet.* **138**:323–331.
39. Dionisio, F. 2005. Plasmids survive despite their cost and male-specific phages due to heterogeneity of bacterial populations. *Evol. Ecol. Res.* **7**:1089–1107.
40. Dionisio, F., I. C. Conceicao, A. C. Marques, L. Fernandes, and I. Gordo. 2005. The evolution of a conjugative plasmid and its ability to increase bacterial fitness. *Biol. Lett.* **1**:250–252.
41. Disque-Kochem, C., and B. Dreiseikelmann. 1997. The cytoplasmic DNA-binding protein TraM binds to the inner membrane protein TraD in vitro. *J. Bacteriol.* **179**:6133–6137.
42. Doolittle, R. F. 1998. Microbial genomes opened up. *Nature* **392**:339–342.
43. Draper, O., C. E. Cesar, C. Machon, F. de la Cruz, and M. Llosa. 2005. Site-specific recombinase and integrase activities of a conjugative relaxase in recipient cells. *Proc. Natl. Acad. Sci. U. S. A.* **102**:16385–16390.
- 43a. Eddy, S. R. 1998. Profile hidden Markov models. *Bioinformatics* **14**:755–763.
44. Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**:1792–1797.
45. Egan, E. S., M. A. Fogel, and M. K. Waldor. 2005. Divided genomes: negotiating the cell cycle in prokaryotes with multiple chromosomes. *Mol. Microbiol.* **56**:1129–1138.
46. Egan, E. S., and M. K. Waldor. 2003. Distinct replication requirements for the two *Vibrio cholerae* chromosomes. *Cell* **114**:521–530.
47. Enright, A. J., S. Van Dongen, and C. A. Ouzounis. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**:1575–1584.
48. Fang, G., E. P. C. Rocha, and A. Danchin. 2005. How essential are nonessential genes? *Mol. Biol. Evol.* **22**:2147–2156.
49. Farias, M. E., and M. Espinosa. 2000. Conjugal transfer of plasmid pMV158: uncoupling of the pMV158 origin of transfer from the mobilization gene mobM, and modulation of pMV158 transfer in *Escherichia coli* mediated by IncP plasmids. *Microbiology* **146** (Pt. 9):2259–2265.
50. Fernandez-Lopez, R., M. P. Garcillan-Barcia, C. Revilla, M. Lazaro, L. Vielva, and F. de la Cruz. 2006. Dynamics of the IncW genetic backbone imply general trends in conjugative plasmid evolution. *FEMS Microbiol. Rev.* **30**:942–966.
51. Finlay, B. B., L. S. Frost, and W. Paranchych. 1986. Origin of transfer of IncF plasmids and nucleotide sequences of the type II oriT, traM, and traY alleles from ColB4-K98 and the type IV traY allele from R100-1. *J. Bacteriol.* **168**:132–139.
52. Francia, M. V., A. Varsaki, M. P. Garcillan-Barcia, A. Latorre, C. Drinas, and F. de la Cruz. 2004. A classification scheme for mobilization regions of bacterial plasmids. *FEMS Microbiol. Rev.* **28**:79–100.
53. Frank, A. C., C. M. Alsmark, M. Tholleson, and S. G. Andersson. 2005. Functional divergence and horizontal transfer of type IV secretion systems. *Mol. Biol. Evol.* **22**:1325–1336.
54. Fronzes, R., E. Schafer, L. Wang, H. R. Saibil, E. V. Orlova, and G. Waksman. 2009. Structure of a type IV secretion system core complex. *Science* **323**:266–268.
55. Frost, L. S., R. Leplae, A. O. Summers, and A. Toussaint. 2005. Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* **3**:722–732.
56. Galtier, N., M. Gouy, and C. Gautier. 1996. SEAVIEW and PHYLO\_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput. Appl. Biosci.* **12**:543–548.
57. Garcillan-Barcia, M. P., M. V. Francia, and F. de la Cruz. 2009. The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol. Rev.* **33**:657–687.
58. Garcillan-Barcia, M. P., P. Jurado, B. Gonzalez-Perez, G. Moncalian, L. A. Fernandez, and F. de la Cruz. 2007. Conjugative transfer can be inhibited by blocking relaxase activity within recipient cells with intrabodies. *Mol. Microbiol.* **63**:404–416.
59. Gelvin, S. B. 2003. Agrobacterium-mediated plant transformation: the biology behind the “gene-jockeying” tool. *Microbiol. Mol. Biol. Rev.* **67**:16–37.
60. Greve, B., S. Jensen, K. Brugger, W. Zillig, and R. A. Garrett. 2004. Genomic comparison of archaeal conjugative plasmids from *Sulfolobus*. *Archaea* **1**:231–239.
61. Grohmann, E., G. Muth, and M. Espinosa. 2003. Conjugative plasmid transfer in gram-positive bacteria. *Microbiol. Mol. Biol. Rev.* **67**:277–301.
62. Groisman, E. A., and H. Ochman. 1996. Pathogenicity islands: bacterial evolution in quantum leaps. *Cell* **87**:791–794.
63. Guasch, A., M. Lucas, G. Moncalian, M. Cabezas, R. Perez-Luque, F. X. Gomis-Ruth, F. de la Cruz, and M. Coll. 2003. Recognition and processing of the origin of transfer DNA by conjugative relaxase TrwC. *Nat. Struct. Biol.* **10**:1002–1010.
64. Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**:696–704.
65. Guo, X., M. Flores, P. Mavingui, S. I. Fuentes, G. Hernandez, G. Davila, and R. Palacios. 2003. Natural genomic design in *Sinorhizobium meliloti*: novel genomic architectures. *Genome Res.* **13**:1810–1817.
66. Halary, S., J. W. Leigh, B. Cheaib, P. Lopez, and E. Bapteste. 2010. Network analyses structure genetic diversity in independent genetic worlds. *Proc. Natl. Acad. Sci. U. S. A.* **107**:127–132.
67. Harrison, P. W., R. P. Lower, N. K. Kim, and J. P. Young. 2010. Introducing the bacterial ‘chromid’: not a chromosome, not a plasmid. *Trends Microbiol.* **18**:141–148.
68. Hayashi, K., N. Morooka, Y. Yamamoto, K. Fujita, K. Isono, S. Choi, E. Ohtsubo, T. Baba, B. L. Wanner, H. Mori, and T. Horiuchi. 2006. Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Mol. Syst. Biol.* **2**:2006.0007.
69. Heinemann, J. A. 1991. Genetics of gene transfer between species. *Trends Genet.* **7**:181–185.
70. Itaya, M., and T. Tanaka. 1997. Experimental surgery to create subgenomes of *Bacillus subtilis* 168. *Proc. Natl. Acad. Sci. U. S. A.* **94**:5378–5382.
71. Jannière, L., A. Gruss, and S. D. Ehrlich. 1993. Plasmids, p. 625–644. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other Gram-positive bacteria: biochemistry, physiology, and molecular genetics. American Society for Microbiology, Washington, DC.
72. Juhas, M., D. W. Crook, I. D. Dimopoulou, G. Lunter, R. M. Harding, D. J. Ferguson, and D. W. Hood. 2007. Novel type IV secretion system involved in propagation of genomic islands. *J. Bacteriol.* **189**:761–771.
73. Juhas, M., P. M. Power, R. M. Harding, D. J. Ferguson, I. D. Dimopoulou, A. R. Elamin, Z. Mohd-Zain, D. W. Hood, R. Adegbola, A. Erwin, A. Smith, R. S. Munson, A. Harrison, L. Mansfield, S. Bentley, and D. W. Crook. 2007. Sequence and functional analyses of *Haemophilus* spp. genomic islands. *Genome Biol.* **8**:R237.
74. Kim, M. J., I. Hirono, K. Kurokawa, T. Maki, J. Hawke, H. Kondo, M. D. Santos, and T. Aoki. 2008. Complete DNA sequence and analysis of the transferable multiple-drug resistance plasmids (R plasmids) from *Photobacterium damsela* subsp. *piscicida* isolates collected in Japan and the United States. *Antimicrob. Agents Chemother.* **52**:606–611.
75. Koksharova, O. A., and C. P. Wolk. 2002. Genetic tools for cyanobacteria. *Appl. Microbiol. Biotechnol.* **58**:123–137.
76. Komano, T., S. R. Kim, T. Yoshida, and T. Nisioka. 1994. DNA rearrangement of the shufflon determines recipient specificity in liquid mating of Inc11 plasmid R64. *J. Mol. Biol.* **243**:6–9.
77. Komano, T., T. Yoshida, K. Narahara, and N. Furuya. 2000. The transfer region of Inc11 plasmid R64: similarities between R64 tra and *Legionella icm*/dot genes. *Mol. Microbiol.* **35**:1348–1359.
78. Krone, S. M., R. Lu, R. Fox, H. Suzuki, and E. M. Top. 2007. Modelling the spatial dynamics of plasmid transfer and persistence. *Microbiology* **153**:2803–2816.
79. Kurenbach, B., C. Bohn, J. Prabhu, M. Abudukerim, U. Szewzyk, and E. Grohmann. 2003. Intergeneric transfer of the *Enterococcus faecalis* plasmid pIP501 to *Escherichia coli* and *Streptomyces lividans* and sequence analysis of its tra region. *Plasmid* **50**:86–93.
80. Latorre, A., R. Gil, F. J. Silva, and A. Moya. 2005. Chromosomal stasis versus plasmid plasticity in aphid endosymbiont *Buchnera aphidicola*. *Hereditas* **95**:339–347.
81. Lawley, T. D., M. W. Gilmour, J. E. Gunton, L. J. Standeven, and D. E. Taylor. 2002. Functional and mutational analysis of conjugative transfer region 1 (Tra1) from the IncH11 plasmid R27. *J. Bacteriol.* **184**:2173–2180.
82. Lawley, T. D., W. A. Klimke, M. J. Gubbins, and L. S. Frost. 2003. F factor conjugation is a true type IV secretion system. *FEMS Microbiol. Lett.* **224**:1–15.
83. Lawrence, J. G., and H. Ochman. 1997. Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.* **44**:383–397.
84. LeBlanc, D. J., Y. Y. Chen, and L. N. Lee. 1993. Identification and characterization of a mobilization gene in the streptococcal plasmid, pVA380-1. *Plasmid* **30**:296–302.
85. Lee, J. H., and D. J. O’Sullivan. 2006. Sequence analysis of two cryptic plasmids from *Bifidobacterium longum* DJO10A and construction of a shuttle cloning vector. *Appl. Environ. Microbiol.* **72**:527–535.
86. Leer, R. J., N. van Luijk, M. Posno, and P. H. Pouwels. 1992. Structural and functional analysis of two cryptic plasmids from *Lactobacillus pentosus* MD353 and *Lactobacillus plantarum* ATCC 8014. *Mol. Gen. Genet.* **234**:265–274.
87. Leplae, R., A. Hebrant, S. J. Wodak, and A. Toussaint. 2004. ACLAME: a classification of mobile genetic elements. *Nucleic Acids Res.* **32**:D45–D49.
88. Lerat, E., V. Daubin, and N. A. Moran. 2003. From gene trees to organismal

- phylogeny in prokaryotes: the case of the gamma-proteobacteria. *PLoS Biol.* **1**:E19.
89. **Levin, B. R.** 1993. The accessory genetic elements of bacteria: existence conditions and (co)evolution. *Curr. Opin. Genet. Dev.* **3**:849–854.
  90. **Levin, B. R., and C. T. Bergstrom.** 2000. Bacteria are different: observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in prokaryotes. *Proc. Natl. Acad. Sci. U. S. A.* **97**:6981–6985.
  91. **Lili, L. N., N. F. Britton, and E. J. Feil.** 2007. The persistence of parasitic plasmids. *Genetics* **177**:399–405.
  92. **Llosa, M., S. Bolland, and F. de la Cruz.** 1991. Structural and functional analysis of the origin of conjugal transfer of the broad-host-range IncW plasmid R388 and comparison with the related IncN plasmid R46. *Mol. Gen. Genet.* **226**:473–483.
  93. **Llosa, M., F. X. Gomis-Ruth, M. Coll, and F. de la Cruz.** 2002. Bacterial conjugation: a two-step mechanism for DNA transport. *Mol. Microbiol.* **45**:1–8.
  94. **Llosa, M., S. Zunzunegui, and F. de la Cruz.** 2003. Conjugal coupling proteins interact with cognate and heterologous VirB10-like proteins while exhibiting specificity for cognate relaxosomes. *Proc. Natl. Acad. Sci. U. S. A.* **100**:10465–10470.
  95. **Long, S. R.** 1989. Rhizobium-legume nodulation—life together in the underground. *Cell* **56**:203–214.
  96. **Lu, J., and L. S. Frost.** 2005. Mutations in the C-terminal region of TraM provide evidence for in vivo TraM-TraD interactions during F-plasmid conjugation. *J. Bacteriol.* **187**:4767–4773.
  97. **Lu, J., J. J. Wong, R. A. Edwards, J. Manchak, L. S. Frost, and J. N. Glover.** 2008. Structural basis of specific TraD-TraM recognition during F plasmid-mediated bacterial conjugation. *Mol. Microbiol.* **70**:89–99.
  98. **Marenda, M., V. Barbe, G. Gourgues, S. Mangenot, E. Sagne, and C. Citti.** 2006. A new integrative conjugative element occurs in *Mycoplasma agalactiae* as chromosomal and free circular forms. *J. Bacteriol.* **188**:4137–4141.
  99. **Mazodier, P., R. Petter, and C. Thompson.** 1989. Intergeneric conjugation between *Escherichia coli* and *Streptomyces* species. *J. Bacteriol.* **171**:3583–3585.
  100. **Meyer, R.** 2009. Replication and conjugative mobilization of broad host-range IncQ plasmids. *Plasmid* **62**:57–70.
  101. **Mihajlovic, S., S. Lang, M. V. Sut, H. Strohmaier, C. J. Gruber, G. Koraimann, E. Cabezón, G. Moncalian, F. de la Cruz, and E. L. Zechner.** 2009. Plasmid r1 conjugative DNA processing is regulated at the coupling protein interface. *J. Bacteriol.* **191**:6877–6887.
  102. **Molbak, L., A. Tett, D. W. Ussery, K. Wall, S. Turner, M. Bailey, and D. Field.** 2003. The plasmid genome database. *Microbiology* **149**:3043–3045.
  103. **Muro-Pastor, A. M., T. Kuritz, E. Flores, A. Herrero, and C. P. Wolk.** 1994. Transfer of a genetic marker from a megaplasmid of *Anabaena* sp. strain PCC 7120 to a megaplasmid of a different *Anabaena* strain. *J. Bacteriol.* **176**:1093–1098.
  104. **Naglich, J. G., and R. E. Andrews, Jr.** 1988. Tn916-dependent conjugal transfer of PC194 and PUB110 from *Bacillus subtilis* into *Bacillus thuringiensis* subsp. *israelensis*. *Plasmid* **20**:113–126.
  105. **Nogueira, T., D. J. Rankin, M. Touchon, F. Taddei, S. P. Brown, and E. P. Rocha.** 2009. Horizontal gene transfer of the secretome drives the evolution of bacterial cooperation and virulence. *Curr. Biol.* **19**:1683–1691.
  106. **Nojiri, H., M. Shintani, and T. Omori.** 2004. Divergence of mobile genetic elements involved in the distribution of xenobiotic-catabolic capacity. *Appl. Microbiol. Biotechnol.* **64**:154–174.
  107. **Nouvel, L. X., P. Sirand-Pugnet, M. S. Marenda, E. Sagne, V. Barbe, S. Mangenot, C. Schenowitz, D. Jacob, A. Barre, S. Claverol, A. Blanchard, and C. Citti.** 2010. Comparative genomic and proteomic analyses of two *Mycoplasma agalactiae* strains: clues to the macro- and micro-events that are shaping *Mycoplasma* diversity. *BMC Genomics* **11**:86.
  108. **Ochman, H.** 2002. Bacterial evolution: chromosome arithmetic and geometry. *Curr. Biol.* **12**:R427–R428.
  109. **Ochman, H., J. G. Lawrence, and E. A. Groisman.** 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**:299–304.
  110. **Palenik, B., Q. Ren, V. Tai, and I. T. Paulsen.** 2009. Coastal *Synechococcus* metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity. *Environ. Microbiol.* **11**:349–359.
  111. **Perreten, V., F. Schwarz, L. Cresta, M. Boeglin, G. Dasen, and M. Teuber.** 1997. Antibiotic resistance spread in food. *Nature* **389**:801–802.
  112. **Planet, P. J., S. C. Kachlany, R. DeSalle, and D. H. Figurski.** 2001. Phylogeny of genes for secretion NTPases: identification of the widespread tadA subfamily and development of a diagnostic key for gene classification. *Proc. Natl. Acad. Sci. U. S. A.* **98**:2503–2508.
  113. **Ponciano, J. M., L. De Gelder, E. M. Top, and P. Joyce.** 2007. The population biology of bacterial plasmids: a hidden Markov model approach. *Genetics* **176**:957–968.
  114. **Prangishvili, D., S. V. Albers, I. Holz, H. P. Arnold, K. Stedman, T. Klein, H. Singh, J. Hiort, A. Schweier, J. K. Kristjansson, and W. Zillig.** 1998. Conjugation in archaea: frequent occurrence of conjugative plasmids in *Sulfolobus*. *Plasmid* **40**:190–202.
  115. **Qiu, W. G., S. E. Schutzer, J. F. Bruno, O. Attie, Y. Xu, J. J. Dunn, C. M. Fraser, S. R. Casjens, and B. J. Luft.** 2004. Genetic exchange and plasmid transfers in *Borrelia burgdorferi* sensu stricto revealed by three-way genome comparisons and multilocus sequence typing. *Proc. Natl. Acad. Sci. U. S. A.* **101**:14150–14155.
  116. **Rankin, D. J., E. P. Rocha, and S. P. Brown.** 24 March 2010. What traits are carried on mobile genetic elements, and why? *Heredity*. doi:10.1038/hdy.2010.24.
  117. **Rasmussen, T., R. B. Jensen, and O. Skovgaard.** 2007. The two chromosomes of *Vibrio cholerae* are initiated at different time points in the cell cycle. *EMBO J.* **26**:3124–3131.
  118. **Revilla, C., M. P. Garcillan-Barcia, R. Fernandez-Lopez, N. R. Thomson, M. Sanders, M. Cheung, C. M. Thomas, and F. de la Cruz.** 2008. Different pathways to acquiring resistance genes illustrated by the recent evolution of IncW plasmids. *Antimicrob. Agents Chemother.* **52**:1472–1480.
  119. **Rocha, E. P. C., and A. Danchin.** 2002. Competition for scarce resources might bias bacterial genome composition. *Trends Genet.* **18**:291–294.
  120. **Rocha, E. P. C., A. Danchin, and A. Viari.** 1999. Functional and evolutionary roles of long repeats in prokaryotes. *Res. Microbiol.* **150**:725–733.
  121. **Rosner, J. L.** 1972. Formation, induction, and curing of bacteriophage P1 lysogens. *Virology* **48**:679–689.
  122. **Schluter, A., L. Krause, R. Szczepanowski, A. Goesmann, and A. Puhler.** 2008. Genetic diversity and composition of a plasmid metagenome from a wastewater treatment plant. *J. Biotechnol.* **136**:65–76.
  123. **Schroder, G., and E. Lanka.** 2005. The mating pair formation system of conjugative plasmids—a versatile secretion machinery for transfer of proteins and DNA. *Plasmid* **54**:1–25.
  124. **She, Q., H. Phan, R. A. Garrett, S. V. Albers, K. M. Stedman, and W. Zillig.** 1998. Genetic profile of pNOB8 from *Sulfolobus*: the first conjugative plasmid from an archaeon. *Extremophiles* **2**:417–425.
  125. **Sirand-Pugnet, P., C. Citti, A. Barre, and A. Blanchard.** 2007. Evolution of molluscites: down a bumpy road with twists and turns. *Res. Microbiol.* **158**:754–766.
  126. **Sirand-Pugnet, P., C. Lartigue, M. Marenda, D. Jacob, A. Barre, V. Barbe, C. Schenowitz, S. Mangenot, A. Couloux, B. Segurens, A. de Daruvar, A. Blanchard, and C. Citti.** 2007. Being pathogenic, plastic, and sexual while living with a nearly minimal bacterial genome. *PLoS Genet.* **3**:e75.
  127. **Slater, S. C., B. S. Goldman, B. Goodner, J. C. Setubal, S. K. Farrand, E. W. Nester, T. J. Burr, L. Banta, A. W. Dickerman, I. Paulsen, L. Otten, G. Suen, R. Welch, N. F. Almeida, F. Arnold, O. T. Burton, Z. Du, A. Ewing, E. Godsy, S. Heisel, K. L. Houmiel, J. Jhaveri, J. Lu, N. M. Miller, S. Norton, Q. Chen, W. Phoolcharoen, V. Ohlin, D. Ondrusek, N. Pride, S. L. Stricklin, J. Sun, C. Wheeler, L. Wilson, H. Zhu, and D. W. Wood.** 2009. Genome sequences of three *Agrobacterium* biovars help elucidate the evolution of multichromosome genomes in bacteria. *J. Bacteriol.* **191**:2501–2511.
  128. **Soler Bistue, A. J., D. Birshan, A. P. Tomaras, M. Dandekar, T. Tran, J. Newmark, D. Bui, N. Gupta, K. Hernandez, R. Sarno, A. Zorreguieta, L. A. Actis, and M. E. Tolmashy.** 2008. Klebsiella pneumoniae multiresistance plasmid pMET1: similarity with the *Yersinia pestis* plasmid pCRY and integrative conjugative elements. *PLoS One* **3**:e1800.
  129. **Stackebrandt, E., and B. M. Goebel.** 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Bacteriol.* **44**:846–849.
  130. **Stedman, K. M., Q. She, H. Phan, I. Holz, H. Singh, D. Prangishvili, R. Garrett, and W. Zillig.** 2000. pING family of conjugative plasmids from the extremely thermophilic archaeon *Sulfolobus islandicus*: insights into recombination and conjugation in Crenarchaeota. *J. Bacteriol.* **182**:7014–7020.
  131. **Suzuki, H., M. Sota, C. J. Brown, and E. M. Top.** 2008. Using Mahalanobis distance to compare genomic signatures between bacterial plasmids and chromosomes. *Nucleic Acids Res.* **36**:e147.
  132. **Szczepanowski, R., T. Bekel, A. Goesmann, L. Krause, H. Kromeke, O. Kaiser, W. Eichler, A. Puhler, and A. Schluter.** 2008. Insight into the plasmid metagenome of wastewater treatment plant bacteria showing reduced susceptibility to antimicrobial drugs analysed by the 454-pyrosequencing technology. *J. Biotechnol.* **136**:54–64.
  133. **Szczepanowski, R., B. Linke, I. Krahn, K. H. Gartemann, T. Gutzkow, W. Eichler, A. Puhler, and A. Schluter.** 2009. Detection of 140 clinically relevant antibiotic-resistance genes in the plasmid metagenome of wastewater treatment plant bacteria showing reduced susceptibility to selected antibiotics. *Microbiology* **155**:2306–2319.
  134. **Szipirer, C., E. Top, M. Couturier, and M. Mergeay.** 1999. Retrotransfer or gene capture: a feature of conjugative plasmids, with ecological and evolutionary significance. *Microbiology* **145**(Pt. 12):3321–3329.
  135. **Tato, I., I. Matilla, I. Arechaga, S. Zunzunegui, F. de la Cruz, and E. Cabezón.** 2007. The ATPase activity of the DNA transporter TrvB is modulated by protein TrwA: implications for a common assembly mechanism of DNA translocating motors. *J. Biol. Chem.* **282**:25569–25576.
  136. **Tato, I., S. Zunzunegui, F. de la Cruz, and E. Cabezón.** 2005. TrvB, the coupling protein involved in DNA transport during bacterial conjugation, is a DNA-dependent ATPase. *Proc. Natl. Acad. Sci. U. S. A.* **102**:8156–8161.
  137. **Tettelin, H., D. Riley, C. Cattuto, and D. Medini.** 2008. Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* **11**:472–477.

138. **Thomas, C. M.** 2000. Horizontal gene pool: bacterial plasmids and gene spread. CRC, Amsterdam, Netherlands.
139. **Tolonen, A. C., G. B. Liszt, and W. R. Hess.** 2006. Genetic manipulation of *Prochlorococcus* strain MIT9313: green fluorescent protein expression from an RSF1010 plasmid and Tn5 transposition. *Appl. Environ. Microbiol.* **72**:7607–7613.
140. **Top, E., I. De Smet, W. Verstraete, R. Dijkmans, and M. Mergeay.** 1994. Exogenous isolation of mobilizing plasmids from polluted soils and sludges. *Appl. Environ. Microbiol.* **60**:831–839.
141. **Touchon, M., C. Hoede, O. Tenaillon, V. Barbe, S. Baeriswyl, P. Bidet, E. Bingen, S. Bonacorsi, C. Bouchier, O. Bouvet, A. Calteau, H. Chiapello, O. Clermont, S. Cruveiller, A. Danchin, M. Diard, C. Dossat, M. El Karoui, E. Frapy, L. Garry, J. Ghigo, A. Gilles, J. Johnson, C. Le Bouguenec, M. Lescat, S. Mangenot, V. Martinez-Jehanne, I. Matic, X. Nassif, S. Oztas, M. Petit, C. Pichon, Z. Rouy, C. Saint Ruf, D. Schneider, J. Tourret, B. Vacherie, D. Vallenet, C. Medigue, E. Rocha, and E. Denamur.** 2009. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet.* **5**:e1000344.
142. **Treangen, T. J., O. H. Ambur, T. Tonjum, and E. P. Rocha.** 2008. The impact of the neisserial DNA uptake sequences on genome evolution and stability. *Genome Biol.* **9**:R60.
143. **Trieu-Cuot, P., M. Arthur, and P. Courvalin.** 1987. Origin, evolution and dissemination of antibiotic resistance genes. *Microbiol. Sci.* **4**:263–266.
144. **van Passel, M. W., E. E. Kuramae, A. C. Luyf, A. Bart, and T. Boekhout.** 2006. The reach of the genome signature in prokaryotes. *BMC Evol. Biol.* **6**:84.
145. **Varsaki, A., G. Moncalian, M. del Pilar Garcillan-Barcia, C. Drinas, and F. de la Cruz.** 2009. Analysis of ColE1 MbeC unveils an extended ribbon-helix-helix family of nicking accessory proteins. *J. Bacteriol.* **191**:1446–1455.
146. **Vieira-Silva, S., M. Touchon, and E. P. Rocha.** 2010. No evidence for elemental-based streamlining of prokaryotic genomes. *Trends Ecol. Evol.* **25**:319–320.
147. **Weinert, L. A., J. J. Welch, and F. M. Jiggins.** 2009. Conjugation genes are common throughout the genus *Rickettsia* and are transmitted horizontally. *Proc. Biol. Sci.* **276**:3619–3627.
148. **Werngreen, J. J., E. E. Harding, and M. A. Riley.** 1997. *Rhizobium* gene native: unexpected plasmid stability of indigenous *Rhizobium leguminosarum*. *Proc. Natl. Acad. Sci. U. S. A.* **94**:5483–5488.
149. **Yamaichi, Y., T. Iida, K. S. Park, K. Yamamoto, and T. Honda.** 1999. Physical and genetic map of the genome of *Vibrio parahaemolyticus*: presence of two chromosomes in *Vibrio* species. *Mol. Microbiol.* **31**:1513–1521.
150. **Zakharova, M. V., I. V. Beletskaya, M. M. Denjmukhametov, T. V. Yurkova, L. M. Semenova, M. G. Shlyapnikov, and A. S. Solonin.** 2002. Characterization of pECL18 and pKPN2: a proposed pathway for the evolution of two plasmids that carry identical genes for a type II restriction-modification system. *Mol. Genet. Genomics* **267**:171–178.
151. **Zhang, H., and R. T. Marconi.** 2005. Demonstration of cotranscription and 1-methyl-3-nitroso-nitroguanidine induction of a 30-gene operon of *Borrelia burgdorferi*: evidence that the 32-kilobase circular plasmids are prophages. *J. Bacteriol.* **187**:7985–7995.

**Chris Smillie** received his B.S. in Molecular and Cellular Biology and Mathematics from the University of Arizona, where he worked with Dr. Howard Ochman on bacterial genome dynamics and next-generation sequencing. The following year, he was funded by a BRAVO research fellowship to study plasmid evolution and ecology in Dr. Eduardo Rocha's laboratory at the University of Paris. He is currently enrolled as a graduate student in the Computational and Systems Biology Ph.D. program at the Massachusetts Institute of Technology, where he is a student in Dr. Eric Alm's laboratory. His research interests include the development of computational methods for understanding microbial evolution and ecology, particularly from metagenomics sequencing projects.



**M. Pilar Garcillán-Barcia**, Ph.D. in Molecular Biology and M.Sc. in Biochemistry, is currently a postdoctoral researcher at the Institute of Biomedicine and Biotechnology of Cantabria. Her work has been focused on the molecular mechanisms and diversity of transposable elements and plasmid conjugation. Her main interests are the study of plasmid evolution and its applicability to synthetic biology.



**Dr. M. Victoria Francia** is a molecular biologist interested in the regulation of gene transfer in bacteria. Her research focuses on mobile genetic elements, mainly plasmids and integrons, in the context of the emergence and dissemination of multiple-antibiotic resistance. Her research is also concerned with studying the emergence and spread of important nosocomial pathogens such as MRSA (methicillin-resistant *Staphylococcus aureus*) and VRE (vancomycin-resistant enterococci).



**Eduardo P. C. Rocha** received M.Sc.'s in Biochemical Engineering and in Applied Mathematics, a Ph.D. in Bioinformatics, and an Habilitation à Diriger des Recherches in Biology. He is a CNRS senior researcher at the Institut Pasteur in Paris, where he leads the Microbial Evolutionary Genomics group. His main interests are the study of genome organization and dynamics and especially the ensuing tradeoffs from an evolutionary perspective.



**Fernando de la Cruz** is Professor of Genetics at the University of Cantabria, Santander, Spain. He has worked for more than 30 years on various aspects of plasmid biology, with a focus on the mechanism of bacterial conjugation. His main interests at present revolve around the diversity and evolution of plasmids and their application in Systems and Synthetic Biology.

