Evolution of Conjugation and Type IV Secretion Systems

Julien Guglielmini,*[1,2] Fernando de la Cruz,[3] and Eduardo P.C. Rocha[1,2]

[1]Département Génomes et Génétique, Microbial Evolutionary Genomics, Institut Pasteur, Paris, France
[2]CNRS, UMR3525, Paris, France
[3]Departamento de Biología Molecular e Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC), Universidad de Cantabria-CSIC-SODERCAN, Santander, Spain
***Corresponding author:** E-mail: jgugliel@pasteur.fr.
Associate editor: Howard Ochman

Abstract

Genetic exchange by conjugation is responsible for the spread of resistance, virulence, and social traits among prokaryotes. Recent works unraveled the functioning of the underlying type IV secretion systems (T4SS) and its distribution and recruitment for other biological processes (exaptation), notably pathogenesis. We analyzed the phylogeny of key conjugation proteins to infer the evolutionary history of conjugation and T4SS. We show that single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA) conjugation, while both based on a key AAA$^+$ ATPase, diverged before the last common ancestor of bacteria. The two key ATPases of ssDNA conjugation are monophyletic, having diverged at an early stage from dsDNA translocases. Our data suggest that ssDNA conjugation arose first in diderm bacteria, possibly Proteobacteria, and then spread to other bacterial phyla, including bacterial monoderms and Archaea. Identifiable T4SS fall within the eight monophyletic groups, determined by both taxonomy and structure of the cell envelope. Transfer to monoderms might have occurred only once, but followed diverse adaptive paths. Remarkably, some Firmicutes developed a new conjugation system based on an atypical relaxase and an ATPase derived from a dsDNA translocase. The observed evolutionary rates and patterns of presence/absence of specific T4SS proteins show that conjugation systems are often and independently exapted for other functions. This work brings a natural basis for the classification of all kinds of conjugative systems, thus tackling a problem that is growing as fast as genomic databases. Our analysis provides the first global picture of the evolution of conjugation and shows how a self-transferrable complex multiprotein system has adapted to different taxa and often been recruited by the host. As conjugation systems became specific to certain clades and cell envelopes, they may have biased the rate and direction of gene transfer by conjugation within prokaryotes.

Key words: bacterial conjugation, horizontal gene transfer, type IV protein secretion, exaptation, plasmid evolution.

Introduction

Prokaryotic genomes adapt quickly to new environmental conditions largely because they can acquire pre-evolved traits by horizontal gene transfer (HGT) (de la Cruz and Davies 2000; Gogarten et al. 2002; Ochman et al. 2005). Conjugation is a mechanism of genetic transfer that allows single-event transfer of large DNA fragments, up to entire chromosomes. Conjugation can transfer nonhomologous genes to the recipient genome and has a broader host range than transduction or transformation (Amabile-Cuevas and Chicurel 1992; Llosa et al. 2002; Chen et al. 2005). Accordingly, recent work suggests that conjugation is the most frequent mechanism of HGT (Halary et al. 2010). Indeed, conjugative systems are major players in the spread of antibiotic resistance, metabolic pathways, symbiotic traits, and other mobile genetic elements (de la Cruz and Davies 2000; Thomas 2000; van der Meer and Sentchilo 2003; Frost et al. 2005; Ding and Hynes 2009; Allen et al. 2010). Conjugation is also involved in the establishment of social processes, promoting biofilm formation (Ghigo 2001) and spreading of cooperative traits (Nogueira et al. 2009; Rankin et al. 2011). There are two known modes of conjugation that differ both in the type of translocated DNA, single-stranded DNA (ssDNA) versus double-stranded DNA (dsDNA), and in the complexity of the transport system (de la Cruz et al. 2010; Vogelmann et al. 2011). Both types of conjugative systems are either encoded by autonomously replicating plasmids or inserted in chromosomes as integrative conjugative elements (ICEs) (Smillie et al. 2010; Wozniak and Waldor 2010). We recently made a large-scale identification of ssDNA conjugation systems, both in plasmids and ICEs, and found them to be essentially short-term variants of otherwise identical backbone elements (Guglielmini et al. 2011).

In the following, we note proteins from a given genetic element by GI$_{MGE}$, where GI refers to the gene identification and mobile genetic element (MGE) to the name of the element (e.g., TraC$_F$ corresponds to the TraC protein of the F plasmid). Conjugative systems involved in ssDNA conjugation include two major protein complexes: relaxosomes and type IV secretion systems (T4SS) (reviewed in Fronzes et al. 2009; de la Cruz et al. 2010). MGE delivery through the membranes of the donor and recipient cells is done by the T4SS (fig. 1). In Proteobacteria, the T4SS are a large protein complex, including a ubiquitous ATPase (VirB4$_{Ti}$ or the distant homolog TraU$_{R64}$), mating-pair formation (MPF) proteins that form the transport channel, and a pilus that attaches to the recipient cell (Alvarez-Martinez and Christie 2009;

Open Access

Article
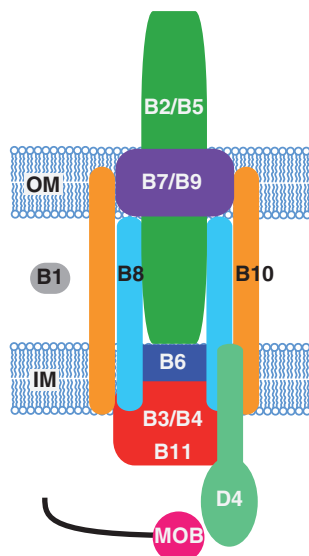
Fig. 1. Scheme of the most-studied T4SS, the *vir* system of *A. tumefaciens* Ti plasmid. The VirBX proteins are depicted as BX (e.g., B5 refers to the VirB5 protein). The coupling protein VirD4 (D4) and the mobilization complex, which includes the relaxase (MOB)-DNA complex are also represented. OM: outer membrane; IM: inner membrane.

Fronzes et al. 2009). The large (>70 kDa) VirB4 ATPase is highly conserved in sequence and the only protein with clear-sequence homologs in all known T4SS. It is therefore the marker of the presence of a T4SS (Alvarez-Martinez and Christie 2009). VirB4 is thought to energize the assembly or activity of the secretion channel and is essential for pilus biogenesis and substrate transfer (Berger and Christie 1993; Fullner et al. 1996; Wallden et al. 2012). Four MPF families have been described in Proteobacteria: MPF$_T$ (based on the T-DNA conjugation system of *A. tumefaciens* plasmid Ti), MPF$_F$ (based on plasmid F), MPF$_I$ (based on the IncI plasmid R64), and MPF$_G$ (based on ICEHIN1056) (Smillie et al. 2010). These four models describe all functionally studied and nearly all T4SS identified by bioinformatic methods among Proteobacteria, both in plasmids and chromosomes (Guglielmini et al. 2011). The best-studied system is the *vir* operon (MPF$_T$) from *A. tumefaciens* Ti plasmid. This small operon encodes 11 VirB proteins (Thompson et al. 1988; Ward et al. 1988), and we use these names as a template for naming the protein families of the MPF$_T$ system. T4SS from Cyanobacteria, Bacteroides, Firmicutes, Actinobacteria, and Archaea have homologs to VirB4 (Guglielmini et al. 2011). ssDNA-conjugative systems are very diverse, but very few studies have been done on the structure, function, and evolution of T4SS outside Proteobacteria and Firmicutes.

The two other essential components of the ssDNA conjugation machinery are the relaxosome and the type IV coupling protein (T4CP). The relaxosome is composed of the relaxase (MOB) and often includes auxiliary proteins. It nicks the dsDNA and binds the resulting ssDNA at the origin of transfer. The diversity and evolution of the different families of relaxases has been extensively studied (Garcillan-Barcia et al. 2009). The highly conserved T4CP binds the DNA-relaxase substrate and couples it to the T4SS, possibly

using ATP to translocate the complex across the inner membrane (Gomis-Ruth et al. 2004; Tato et al. 2005). The majority of T4CPs belong to the VirD4$_{Ti}$ family, but some T4SS were recently found to lack VirD4 and instead use a distantly related ATPase as T4CP (TcpA$_{pCW3}$) (Parsons et al. 2007; Steen et al. 2009). Protein secretion systems based on T4SS do not require relaxosomes. They usually require T4CP, albeit exceptions have been found in *Bordetella pertussis* and *Brucella* spp. (Alvarez-Martinez and Christie 2009). In these systems, proteins are translocated across the inner membrane by other means.

Conjugation of dsDNA takes place in mycelia-producing Actinobacteria (Grohmann et al. 2003; Ghinet et al. 2011). It relies on a single protein: TraB$_{pSG5}$ that translocates dsDNA between neighboring cells in mycelia (Possoz et al. 2001). This protein resembles, in sequence and function, the essential protein FtsK that segregates sister chromosomes in the last stages of chromosomal replication (Bigot et al. 2007; Vogelmann et al. 2011). They are both members of the AAA$^+$ motor ATPase family, which also includes both types of T4CP (VirD4 and TcpA) and both types of ATPases essential for the function of T4SS (VirB4 and TraU). Hence, all key proteins of the dsDNA and ssDNA conjugation systems are evolutionarily related. This association has not yet been clarified from a phylogenetic point of view.

T4SS are often recruited by bacterial pathogens to deliver effectors to eukaryotic cells (Weiss et al. 1993; Vogel et al. 1998; Seubert et al. 2003; Nystedt et al. 2008). These MOBless T4SS, called so because they do not contain a relaxase gene, are closely related to the T4SS of conjugative systems. Indeed, several T4SS can perform both conjugation between bacteria and protein delivery (Vogel et al. 1998; Llosa et al. 2003; Schroder et al. 2011). Protein delivery by T4SS is essential for the virulence of many plant and animal pathogens, including *Legionella pneumophila*, *Helicobacter pylori*, *Bartonella* spp., *Coxiella burnetii*, and *A. tumefaciens* (reviewed in Seubert et al. 2003; Juhas et al. 2008; Alvarez-Martinez and Christie 2009). Only T4SS among MPF$_T$ and MPF$_I$ have been experimentally shown to be used for protein delivery. The extreme flexibility of T4SS has allowed at least two other types of exaptations, i.e., evolutionary events in which part of the pre-existing machinery of conjugation was recruited for other functions (Gould and Vrba 1982). *H. pylori* genomes encode a MOBless T4SS that is used for natural transformation. It is necessary to import environmental DNA (Hofreuter et al. 2001). In *Neisseria gonorrhoeae*, one T4SS is responsible for DNA export to the extracellular space, an intermediate step in the process of natural transformation among these bacteria (Hamilton et al. 2005). Interestingly, in the case of *Neisseria*, the locus encodes a T4SS and a MOB$_H$-type relaxase that is necessary for DNA export (Salgado-Pabon et al. 2007). A previous analysis of MPF$_T$ systems suggests that exaptation of conjugative systems occurred several times in evolution (Frank et al. 2005). Because we recently found that MOBless T4SS are significantly more abundant than previously thought (Guglielmini et al. 2011), this point needs to be reassessed for MPF$_T$ and developed for other MPF types.

Although studies on conjugation are as old as molecular biology itself (Lederberg and Tatum 1946), several recent works have significantly changed our understanding of this process. These include the discovery of new conjugation systems (Juhas, Crook, et al. 2007), of new key elements in known conjugation systems, e.g., TcpA (Parsons et al. 2007) and of the important role of ICEs (Burrus et al. 2002; Wozniak and Waldor 2010). Recent functional studies explored the diversity of T4SS (Alvarez-Martinez and Christie 2009), and bioinformatics work unraveled the presence of T4SS in several new clades (Guglielmini et al. 2011). Finally, other works highlighted the close structural and functional relationship between T4SS used for protein secretion and conjugation (Fernandez-Gonzalez et al. 2011). This succession of works opens the opportunity to infer a global scenario for the evolution of conjugative systems and T4SS, which is the goal of the present work. To assess the uncertainty in the phylogenetic reconstruction, we used classical methods such as bootstrap analyses. Yet, because these large and deep phylogenetic reconstructions can be sensitive to alignment algorithms and to methods to extract informative positions (Philippe et al. 2011), we also tested the robustness of our results by comparing them with two automatic analyses that we did in parallel. To guide the comparisons between the three sets of analyses, we made an assessment of the quality of the multiple alignments using T-Coffee (Notredame et al. 2000). By default, we only mention the results of our expert analysis (typically, the one with highest alignment quality), but highlight differences between methods when they are relevant. The overall structure of the article is the following. First, we analyze the deep branching of the key proteins that have homologs among (nearly) all conjugative systems of a given kind. This allows uncovering the initial split of the proteins that became key to conjugative processes. Then, we focus on the early events of the diversification of ssDNA conjugation, by far the most frequent process among prokaryotes. Finally, we detail the diversification of the best-known conjugation families within ssDNA-based systems with a focus on the evolution of gene repertoires and MOBless T4SS. This analysis provides information that naturally leads to a revision of T4SS classification based on evolutionary biology.

Materials and Methods

Data

Data on complete chromosomes and plasmids of prokaryotes were taken from Genbank Refseq (ftp://ftp.ncbi.nih.gov/genomes/Bacteria/, last accessed November 2011). This included 1,207 chromosomes, 891 plasmids that were sequenced along with these chromosomes, and 1,391 plasmids that were sequenced independently. We used the annotations of the Genbank files, having removed all pseudogenes and proteins with inner stop codons. The information on T4SS was taken from Guglielmini et al. (2011).

Construction of Protein Profiles and Genome Searches

Unless mentioned explicitly, the protein profiles used are those described in Guglielmini et al. (2011). To study the presence/absence of the different components of the *vir* system, we made additional protein profiles, namely for VirB1, VirB2, VirB5, VirB7, VirB10, and VirB11. We first used PSI-Basic Local Alignment Search Tool (BLAST) (e value < 0.1) to search for distant homologs, using as query each of these genes from the VirB locus of the *A. tumefaciens* plasmid pTi SAKURA (Refseq entry NC_002147) and the aforementioned databank of completely sequenced replicons. Given the problems of convergence of PSI-BLAST when using complete genomes, and the extensive similarity of plasmid and chromosomal conjugative systems (Guglielmini et al. 2011), we restricted homology searches to plasmid sequences when building protein profiles. We retrieved the proteins with hits for each protein family and built multiple alignments using MUSCLE (Edgar 2004). We removed the few proteins with sizes very different from the average. We then rebuilt the multiple alignments with MUSCLE and trimmed them to remove the sites at the edges that were poorly aligned. We used HMMER 3.0 (Eddy 2011) to produce hidden Markov model (HMM) profiles and to perform searches within genomes. In the analysis of the evolution of the MPF_T system, we only considered the hits that colocalized with previously detected *vir* proteins (VirB3, VirB4, VirB6, VirB8, VirB9). FtsK proteins were retrieved directly by using the PFAM PF01580 profile. TraB proteins, being closely related to FtsK, were retrieved by BLASTP searches of TraB from *Streptomyces* plasmid pCQ3 (YP_003280879) on the Actinomycetales proteins from the Refseq database. We sampled the top results and then built a protein profile for this protein and searched for its occurrences as for the other profiles. We built a web server to allow running the protein profiles. This is available at http://mobyle.pasteur.fr/cgi-bin/portal.py#forms::CONJscan-T4SSscan.

Phylogenetic Analysis

Unless explicitly stated, all phylogenetic analyses were performed with the following procedure. First, sequences were aligned using MUSCLE with default parameters as implemented in SeaView (Gouy et al. 2010). Second, all columns in the multiple alignment matrix with more than 80% of gaps were removed. Third, 100 replicate trees were built with RAxML 7.2.7 (Stamatakis 2006) using the model GTRGAMMA. We kept the one with the best likelihood. We calculated bootstraps with the standard implementation and used the autoMR stop criterion to obtain confidence values for each node. There were two exceptions to this method. We aligned the ATPases using MAFFT (Katoh and Toh 2010) with the G-INSI algorithm and removed the sites containing more than 60% of gaps. We performed the phylogenetic inference as mentioned earlier and additionally with PhyML 3.0 (Gascuel et al. 2010) under the LG model and with the bioNJ starting tree to get aLRT support values. The alignment of the set of VirB4 and VirD4 was built with MAFFT with the E-INSI algorithm, since these two proteins show different domain organization, and then manually edited. MAFFT was used instead of MUSCLE because it provided better alignments in these cases. The computation of

100 replicates plus hundreds of bootstrap trees was excessively time consuming, given the size of the data set in the VirB4/VirD4 analysis. Thus, we used PhyML 3.0 to build the phylogenetic tree, under the LG model and with the bioNJ starting tree. aLRT support values were also calculated for each node.

The support tests we conducted revealed in this last tree some weak support that conflict with the aLRT values. To further investigate this, we used a reduced data set composed of VirB4 proteins, excluding the distant homolog TraU. Using this data set, we performed the tests described later. All multiple alignments and phylogenetic reconstructions are freely available on DRYAD (http://data dryad.org/).

Tests to the Phylogenetic Analysis

To test the robustness of our conclusions based on phylogenetic analysis, we made a number of tests. These analyses aimed at testing the robustness of the conclusions to the multiple alignments, to the identification of informative sites in multiple alignments, and to the use of a protein model matrix. We therefore produced two automatic methods where we make the alignment of the protein using MAFFT and MUSCLE. Informative sites were extracted from the alignments using BMGE (Criscuolo and Gribaldo 2010). We fine-tuned BMGE parameters for each alignment to obtain a good compromise between the quality and the number of informative sites. The best model to analyze the data was chosen with ProtTest (Darriba et al. 2011). Note that ProtTest does not analyze the GTR model for proteins, so we cannot assess whether the model chosen by ProtTest is better than ours. Trees were built as before using RAxML, and we generated 100 bootstrap trees for each analysis. To compare the different analyses, we computed the quality of multiple alignment score using the Core component of T-Coffee (Notredame et al. 2000) for the three methods (our expert analysis, the MAFFT and MUSCLE-based analyses). This score, ranging from 0 to 100, is computed by comparing the consistency of the alignment with a list of precomputed pairwise alignments called library. We used the default "Mproba_pair" library. The key results, e.g., monophyly or basal position of certain clades, were tested for the three methods and are displayed in table 1 and supplementary table S1, Supplementary Material online. Each of these tests has an identification number in the tables. This number is displayed in the respective node in the phylogenetic trees. For example, in figure 2, the node with ID no. 3 refers to the monophyly of TraB and is indicated in table 1 as having 99% bootstrap support in our expert analysis, 100% in the automatic analysis using MAFFT, and 96% in the automatic analysis using MUSCLE. In supplementary table S1, Supplementary Material online, it is indicated that for this analysis the best alignment, as given by T-Coffee, is the one of the expert alignment (score 88), followed by MAFFT (76) and then MUSCLE (67). The node no. 3 in figure 2 is thus indicated in a black circle (high bootstrap support).

Relative Decrease in Protein Similarity with Divergence

For each pair of T4SS loci, we made pairwise alignments of each of the orthologous pairs of genes. Alignments were done using an end-gap free version of the Needleman–Wunsch algorithm (Mount 2004), with a BLOSUM60 matrix, open penalty of 1.2, and extension penalty of 0.8. We then plotted the percentage of similarity between VirB4 homologs and each of the other pairs of homologs. The points for each scatter plot were then fitted with a spline ($\lambda = 1,500$), and the curves were superimposed.

Results and Discussion

Early Evolutionary Split of the Key Conjugation ATPases

The two families of T4CPs (with prototypes given by the VirD4$_{pTi}$ and TcpA$_{pCW3}$), the two families of ATPases (based on VirB4$_{Ti}$ and TraU$_{R64}$), the dsDNA conjugation protein TraB$_{pSG5}$, and FtsK are all part of the superfamily of AAA$^+$ motor ATPases. Hence, we investigated the events at the onset of the natural history of conjugation from the analysis of the phylogeny-linking homologs for all these protein profiles among 3,489 replicons (see Materials and Methods). The tree was rooted using the distantly related protein family derived from VirB11$_{Ti}$ (Planet et al. 2001). The monophyly of VirB11 is robust in both expert and the automatic analyses (table 1). This phylogenetic reconstruction separates a monophyletic VirD4/VirB4 clade (67% boostrap) from the others. This fits previous genomic and structural analysis showing the similarity between the dsDNA translocators FtsK and TraB on the one hand and between the ssDNA translocators VirD4 and VirB4 on the other (Iyer et al. 2004; Cabezon et al. 2011).

The previous analysis allows rooting the tree and highlights the early split between ssDNA and dsDNA translocases. But the inclusion of the distantly related VirB11 produces a multiple alignment with few sufficiently conserved positions, increasing uncertainty in the process of phylogenetic inference (supplementary table S1, Supplementary Material online). This reduced the power of this data set to robustly resolve the more recent splits. Thus, we excluded VirB11 from the analysis and made a new phylogenetic reconstruction of the remaining five families. This tree shows the same dichotomy at the base (fig. 2), with strong support for all five monophyletic groups with our expert analysis and in the best automatic method (table 1). These results fit our observation that our VirB4 protein profiles often match VirD4 proteins and vice versa, albeit with weak scores, and that none of these match significantly proteins from the families TraB/FtsK. T4CPs and VirB4s show clear structural similarities, underscoring a common functional mechanism (Cabezon et al. 2011). The most conspicuous structural difference between T4CPs and VirB4s is the existence of three alpha helices that are conserved in the C terminus of VirB4 proteins but are absent in T4CPs. Deletion of these helical structures in the VirB4 homolog TrwK$_{R388}$ resulted in a large increase in its ATPase activity, suggesting that the C-terminal end of VirB4 proteins functions as an autoregulatory element (Pena et al. 2011). Overall, these analyses fit structural work, suggesting that the common ancestor of the VirB4/VirD4

Table 1. Analysis of the Robustness of Key Phylogenetic Results.

Data Set[a]	Hypotheses[b]	Number[c]	Bootstrap Values[d]			Other Informative Traits
			Original Alignment	MAFFT+ BMGE	MUSCLE+ BMGE	
ATPases + VirB11	VirB11 monophyly	1	100	96	87	
	TraB, TcpA, FtsK monophyly		67	30	1	
ATPases without VirB11	VirB4, VirD4 monophyly	2	78	82	19	
	TraB monophyly	3	99	100	96	
	TcpA monophyly	4	96	99	95	
	FtsK monophyly	5	62	38	69	Essential ubiquitous gene in bacteria
VirB4, VirD4	MPF_I monophyly	6	100	100	100	Ten specific genes for this MPF
	MPF_{FA} monophyly	7	98	95	94	
	MPF_T monophyly	9	99	62	89	Four specific genes
	MPF_C monophyly	10	100	100	99	
	MPF_G monophyly	11	100	100	96	Five specific genes
	MPF_F monophyly	12	100	53	85	Nine specific genes for this MPF
	MPF_B monophyly	13	98	87	94	
	Basal MPF_I	14	26	45	23	Specificity of primary sequence of TraU
	Basal MPF_C		33	37	18	
VirB4 (without MPF_I)	MPF_{FA} monophyly	7		100	100	
	MPF_{FATA} monophyly	8		97	74	
	Monoderms monophyly	15		67	55	
	MPF_F, MPF_B monophyly	16		33	33	
	MPF_F, MPF_B, Monoderms monophyly	17		36	18	
VirB4 (MPF_T)	Same basal clade in all analyses	18	100	100	100	
	Two big clades corresponding to the gene order conservation	19	98	90	57	Gene order analysis shows the same split
VirB4 (MPF_F)	Same basal clade in all analyses	20	100	100	77	
VirB4 (MPF_{FATA})	Archaea monophyly		21	29	17	
	Euryarchaeota monophyly	21	96	91	95	
	Crenarchaeota monophyly	22	100	100	100	
	Tenericutes monophyly	23	96	99	97	
	Actinobacteria monophyly		0	0	0	
	Firmicutes monophyly		0	0	0	
VirB4 (MPF_{FA})	Firmicutes I monophyly	24	87	99	52	Use of TcpA and MOB_T
	Firmicutes II monophyly	25	79	98	24	
	Basal Firmicutes III	26	100	90	52	
	Actinobacteria monophyly		79	2	4	
	Actinobacteria I monophyly	27	85	11	91	
	Actinobacteria II monophyly	28	100	100	97	

[a]Proteins included in the data set.

[b]The different hypotheses for which we present the bootstrap supports.

[c]When the hypothesis correspond to what we observe in the reference phylogeny, and if the support value is greater than 50, it is displayed here and in the corresponding figure with a number.

[d]Bootstrap values for each hypothesis and for each alignment technique.
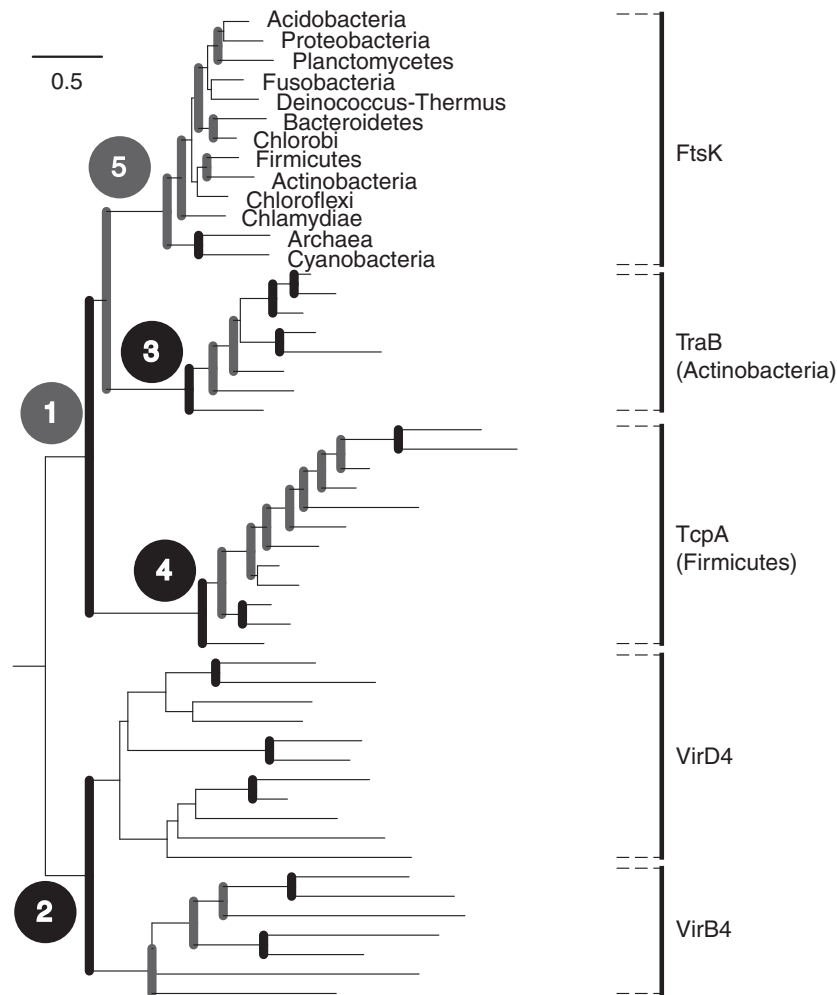
Fig. 2. Phylogenetic analysis of the AAA+ ATPases associated with conjugation. The position of the root was determined using the AAA+ ATPase VirB11 in a separate analysis. Names along the FtsK tips correspond to the taxonomic origins of each protein, reflecting the width of sampling. Bold vertical black lines represent nodes with a high support value (bootstrap >70% and aLRT >0.7). Bold gray lines represent nodes with high aLRT score (>0.7) but a weaker bootstrap (<70%). The homologs of TcpA are found only in Firmicutes. The homologs of TraB are found only in Actinobacteria. Numbers in circles refer to the analysis of robustness in table 1 (identified in the third column of table 1); black background stands for a high support (≥70% bootstrap in the best-scoring alignment) and gray background for a moderate support (≥50% bootstrap in the best-scoring alignment).

families consisted of a soluble protein engaged in polypeptide transport (as it's still the case in most studied VirB4 proteins). VirB4 later became membrane bound by association with the VirB3 component of T4SS (as in VirB4$_{R388}$). This association can be covalent (as in VirB4$_{R6K}$) (Pena et al. 2011). The protein that specialized in ssDNA transport (T4CP) also acquired an integral-membrane protein domain in its N-terminus. This component is involved in its interaction with another T4SS component, in this case VirB10 (Llosa et al. 2003; de Paz et al. 2010).

The other basal branch in the phylogeny includes TraB, TcpA, and FtsK, all with strong to moderate evidence of monophyly (99%, 96%, and 62% bootstraps, respectively) (fig. 2). The relative order of the split between the three clades is different from a previously published one, but its bootstrap support is weak in our tree (and not documented in Parsons et al. 2007). SpoIIIE, a protein involved in segregation of chromosomes during *Bacillus subtilis* sporulation (Wu and Errington 1998), branches within the FtsK clade (data not shown). The elements of the TraB family are found only in

Actinobacteria and are related with FtsK, but they do not emerge from within the FtsK. Instead, they derive independently from the ancestor of this protein. FtsK is an essential protein that, contrary to some previous suggestions (Iyer et al. 2004), includes at least one member among Archaea (YP_503307.1). The latter is annotated as FtsK-like protein, and it is not closely related with HerA proteins, which branch closer to the VirD4/B4 branches, and its study falls outside the scope of this article. FtsK phylogeny follows approximately the one of bacteria (Gupta 2004) and thus provides a guideline to the timing of the diversification of these protein families. The tree in figure 2 shows that proteins have widely diverse tip-to-root branch lengths, i.e., the proteins do not evolve according to a strict molecular clock. Therefore, we cannot assume a molecular clock that would allow dating the split of these families and thus presumably that of conjugation processes. Yet, this data does place the origin of ssDNA conjugation extremely early in the history of life. While TraB and TcpA seem to diversify after FtsK, in agreement with their presence only in Firmicutes and

Actinobacteria, the diversification of the pair VirB4/VirD4 could be contemporaneous or shortly subsequent to that of FtsK. These results suggest that the two conjugation mechanisms, ssDNA and dsDNA conjugation, are based on ATPases that diverged before the last common ancestor of bacteria.

T4SS Phylogeny

We aligned the proteins matching the VirB4 and TraU profiles to infer the evolutionary history of all VirB4 homologs. We then used VirD4 to root this tree. Despite relatively weak support in the bootstrap tests (48% in the best automatic alignment and 69% in our expert analysis), this rooting shows a good aLRT support value (0.82), consistent with the literature in terms of phylogeny and biochemical function (Iyer et al. 2004; de la Cruz et al. 2010; Smillie et al. 2010) and with the previous analysis of the five ATPases (78% bootstrap). The tree shows that all VirB4 and TraU-related proteins can be classified into eight groups, which are represented by eight well-supported clades (fig. 3). The two basal groups in the VirB4 phylogenetic reconstruction are MPF_I followed by a group specific to Cyanobacteria (MPF_C). This is in agreement with the low similarity between $TraU_{R64}$ (MPF_I) and $VirB4_{Ti}$ (MPF_T) that had prevented previous phylogenetic reconstructions of all VirB4 homologs (Smillie et al. 2010). With the availability of more sequences of these proteins, notably cyanobacteria, and the inclusion of the T4CP, we could now reconstruct a reliable phylogeny. However, the position of MPF_I at the basis of the tree must be taken with care. Our expert method and the two controls produce MPF_I at the basis of the phylogeny but with relatively low support (45% bootstrap in the best automatic alignment) (table 1). The MPF_C clade often arises at the basis in the bootstrap trees or as a sister clade of MPF_I. In any case, this analysis places one of these two clades at the root of the tree in more than 85% of the boostrap analyses.

Some mobile elements encoding an MPF_I, e.g., the R64 plasmid from the MOB_{P12} family, besides encoding a thick rigid pilus, with homology to MPF_T, also encode a thin pilus that is only required for conjugation in liquid and that is homologous to type IV pili (Kim and Komano 1997). This led to the classification of MPF_I as T4SSb in opposition to MPF_F and MPF_T, both classed as T4SSa (Christie and Vogel 2000). However, other MPF_I elements, e.g., plasmid CTX-M3, lack a thin pilus and are still able to mate in liquid at high frequency (Golebiewski et al. 2007). Thus, the thin pilus of MOB_{P12} plasmids is just an additional feature of some MPF_I systems, acting probably just as a facilitator of liquid mating and a selector of recipients (Kim and Komano 1997), while the core MPF_I machinery forms the basis of this conjugation system. In any case, the highly divergent nature of $TraU_{R64}$ is a signature for this whole family of liquid maters. Nothing is known experimentally about MPF_C. Because cyanobacteria diverged early on from Proteobacteria, MPF_C might also contain peculiarities relevant to the genetic or physical environment of these organisms. MPF_G is the next most basal group in the tree. This system was recently discovered, was identified

only in Proteobacteria, and its features are largely unknown (Juhas, Crook, et al. 2007; Juhas, Power, et al. 2007). Interestingly, an MPF_G encoding element, the PAPI-1 pathogenicity island of *Pseudomonas aeruginosa*, has several genes homologous to the thin pilus of R64 (Carter et al. 2010). Hence, the association between MPF and thin pili might be an ancestral trait.

Four groups correspond to the different T4SS families of Proteobacteria (MPF_F, MPF_G, MPF_I, MPF_T) (Juhas, Crook, et al. 2007; Smillie et al. 2010). These four groups are clearly separated because they all have strong bootstraps in the analysis of monophyly (table 1), and each contains a set of four to nine genes that are specific, i.e., their protein profiles match loci of a given MPF but not those of the other MPF types (Smillie et al. 2010). Interestingly, 307 out of 327 (94%) of the T4SS of Proteobacteria are classed in one of these four clades. We investigated the loci of the 20 remaining VirB4 proteins. One of them does not colocalize with any of the other conjugation protein profiles, including relaxases and T4CP. The other 19 VirB4 are encoded near genes specific of one, and only one, MPF type. They were not classed as a given MPF just because the number of these specific genes is below the quorum we set up as a minimum for a putative complete T4SS (Guglielmini et al. 2011). Many of these 20 unclassed elements are thus probably inactive, enduring a genetic degradation that results in incomplete loci. Alternatively, they may correspond to highly modified versions of T4SS; the *H. pylori* Cag-pathogenicity island is notably found within these elements.

A few genomes of species not classed among Proteobacteria encode T4SS classed within MPF_F and MPF_T. All these bacteria are diderms, i.e., they have both an inner and an outer membrane. This list includes MOBless T4SS in one Aquificae (MPF_F) and one Protochlamydia (MPF_F), and conjugative T4SS in one Chlorobi (MPF_T), one Deferribacteres (MPF_F), one Acidobacteria (MPF_T), and two Fusobacteria (MPF_T). These elements are scattered in the trees of MPF_T and MPF_F (figs. 4 and 5), suggesting different events of horizontal transfer from Proteobacteria. Indeed, they do not cluster together in the phylogenetic trees (0% in bootstrap trees). The elements of each given bacterial clade are always monophyletic, suggesting one single transfer event, but the very small number of such elements does not allow any robust conclusions for the moment. Only one nonproteobacterial clade, Acidobacteria, is basal in the tree of MPF_T (100% bootstraps in the expert analysis and the controls). Acidobacteria are often regarded as a sister clade of Proteobacteria (Ciccarelli et al. 2006), and therefore, we cannot discard the possibility of a diversification of MPF_T before the split between Acidobacteria and Proteobacteria. However, since MPF_G and MPF_I are more basal in the tree of VirB4 (fig. 3), and both only found in Proteobacteria, the scenario of a transfer from Proteobacteria to Acidobacteria remains more parsimonious. Interestingly, all T4SS predicted in these six nonproteobacterial clades were classed among MPF_F and MPF_T. Nothing is known about conjugation in these clades, but this data suggest they might use mechanisms closely related to, and originating from, those of Proteobacteria.
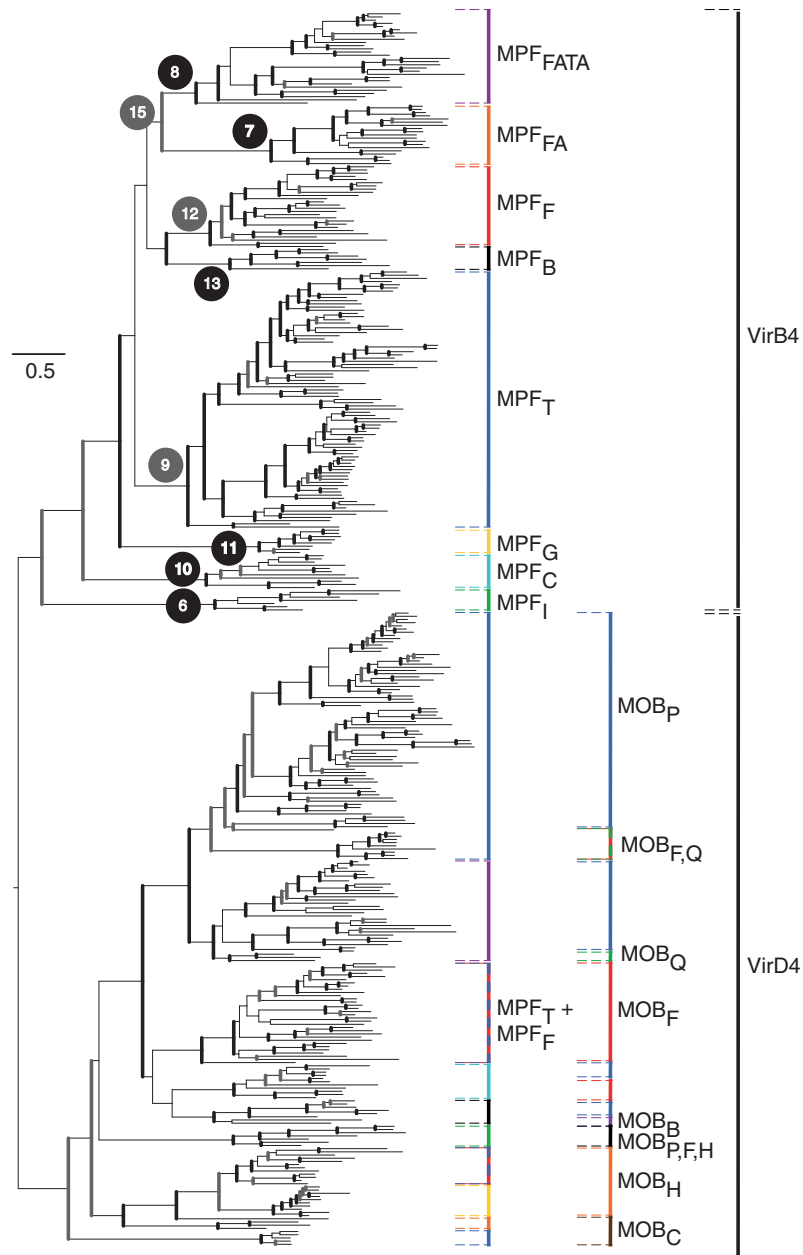
Fig. 3. Joint phylogenetic reconstruction of the VirD4 and VirB4/TraU families of proteins from conjugative systems. Bold vertical black lines represent nodes with a high support value (aLRT ≥ 0.9), and black vertical gray lines represent nodes with a support value between 0.7 and 0.9. Black square brackets indicate the VirB4 and VirD4 clades; colored square brackets on the left delimit the different MPF clades (purple: MPF_{FATA}, orange: MPF_{FA}, red: MPF_F, black: MPF_B, blue: MPF_T, yellow: MPF_G, cyan: MPF_C, green: MPF_I); colored square brackets on the right delimit the relaxase clades within the VirD4 part of the tree (blue: MOB_P, green: MOB_Q, red: MOB_F, purple: MOB_B, orange: MOB_H, brown MOB_C, red/green dashed brackets: clades with a mix of MOB_F and MOB_Q; black: mix of MOB_P, MOB_F and MOB_H). Numbers in circles refer to the analysis of robustness in table 1 (identified in the third column of table 1); black background stands for a high support (≥70% bootstrap in the best-scoring alignment) and gray background for a moderate support (≥50% bootstrap in the best-scoring alignment).

Phylogeny of the T4CP at the Light of VirB4 Phylogeny

The trees of VirD4 and VirB4 are not congruent (ELW confidence value: 0, and SH P value < 0.01). Yet, they share many features (fig. 3). The proteins encoded by the *virD4* genes colocalizing in replicons with *virB4* tend to form similar clades. Notably, the VirD4 associated with each of six of the eight VirB4 clades also clustered in nearly monophyletic clades of T4CP (MPF_{FA}, MPF_{FATA}, MPF_B, MPF_G, MPF_I, and

MPF_C). VirD4 of the two remaining clades (MPF_T and MPF_F) are scattered in a small number of clades. Most of the MPF_{FA} use TcpA instead of a VirD4-like T4CP (see later). The few VirD4 proteins found in MPF_{FA} are also monophyletic (orange in the bottom of fig. 3). It was previously shown that plasmid T4CP are sometimes scattered in different groups corresponding to given relaxases (Smillie et al. 2010). This result is still valid with the present much larger data set. For example, the T4CP clade with a mixture of MPF_T
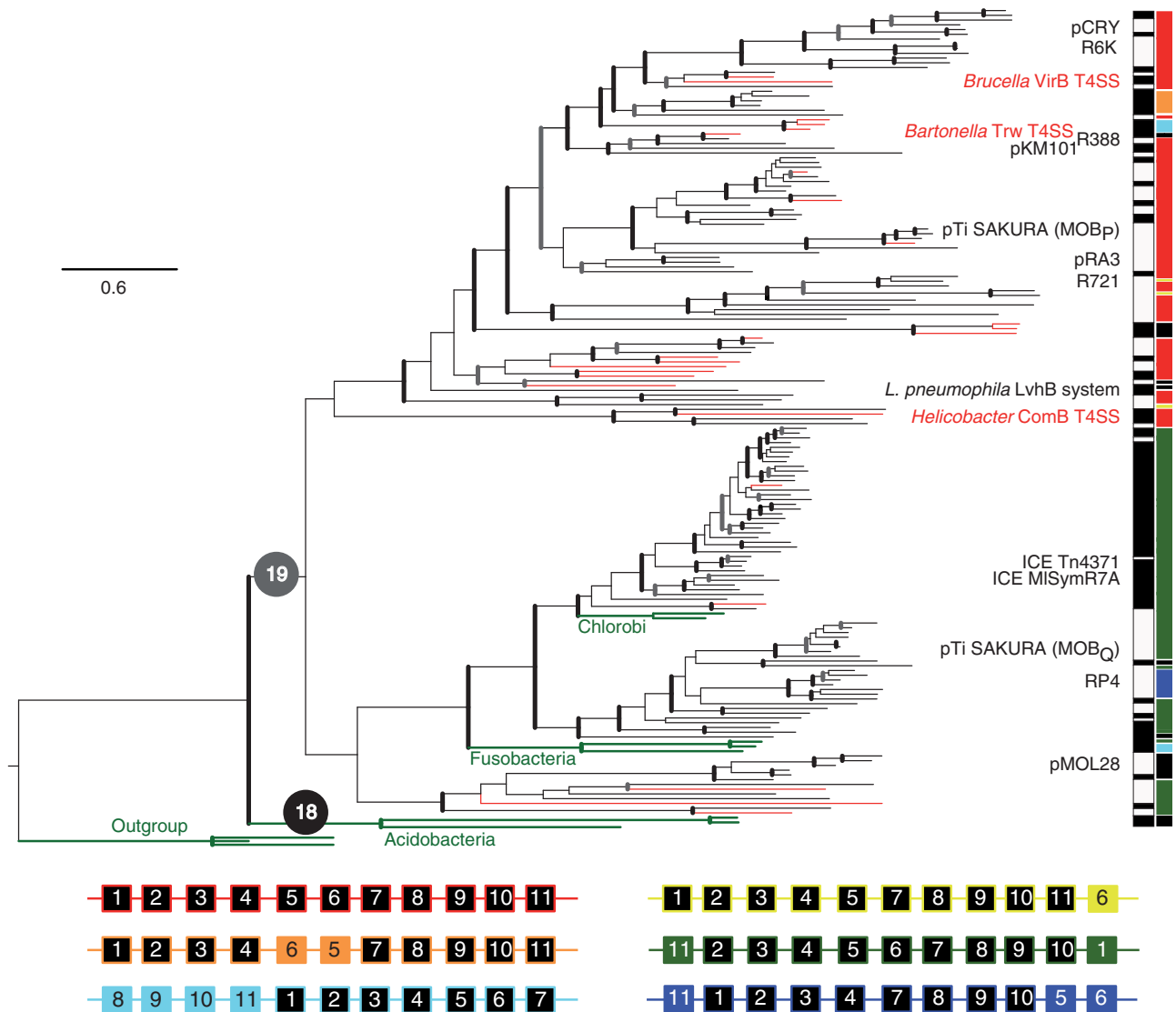
Fig. 4. Phylogenetic analysis of MPF$_T$ VirB4 proteins. Bold vertical black lines represent nodes with a high support value (bootstrap > 90%), and bold vertical gray lines represent nodes with a support value between 70% and 90%. Green branches correspond to taxa that are not within Proteobacteria (or the outgroup). Red branches represent VirB4 not associated to a relaxase (MOBless T4SS). The leftmost vertical bar on the right stands for chromosomal (black) or plasmidic (white) proteins. The colored bar represents the different gene order patterns found; the patterns and their corresponding color are depicted at the bottom (the numbers represent the corresponding *virB* gene); a pattern is attributed to a system if, considering the possibly missing *vir* genes, the gene order is preserved. For example, a system composed of the genes *virB1*, *virB4*, *virB6*, *virB5*, *virB8*, *virB9*, and *virB10* in this order will be assigned to the orange pattern. Unique or atypical patterns are depicted in black. Known representative systems are labeled. Numbers in circles refer to the analysis of robustness in table 1 (identified in the third column of table 1); black background stands for a high support ($\geq$70% bootstrap in the best-scoring alignment) and gray background for a moderate support ($\geq$50% bootstrap in the best-scoring alignment).

and MPF$_F$ has one type of relaxase in common (MOB$_F$). On the other hand, some relaxase types are scattered among different VirD4 clades that follow MPF types, e.g., the VirD4 associated with MPF$_C$ is monophyletic and includes three different relaxases, which are also found in other MPF types. Hence, evolution of conjugation is driven by two main constraints, one acting mainly on the T4SS, represented by VirB4, and other on the relaxosome, represented by the relaxases. T4CP tends to coevolve with both components.

Cell Envelope Adaptation in Monoderms

The most basal clades in both VirB4 and VirD4 phylogenies correspond to bacteria with both inner and outer

membranes, i.e., diderms (98–100% of the bootstrap trees in all three analyses). This strongly suggests that ssDNA conjugation was invented among diderms. In this scenario, ssDNA conjugation would have been acquired by monoderm prokaryotes, i.e., organisms devoid of an outer membrane, by HGT. This also fits the observation that all monoderm conjugation systems are in two sister clades: MPF$_{FA}$ and MPF$_{FATA}$ (monophyletic in 67–55% of the bootstrap trees).

MPF$_{FATA}$ includes six distinct groups of Firmicutes (monophyly of all Firmicutes supported by 0% of the bootstrap trees, table 1), two of Actinobacteria (monophyly of all Actinobacteria supported by 0% bootstrap trees), one of Tenericutes (monophyly of the clade supported by 96–99% bootstrap
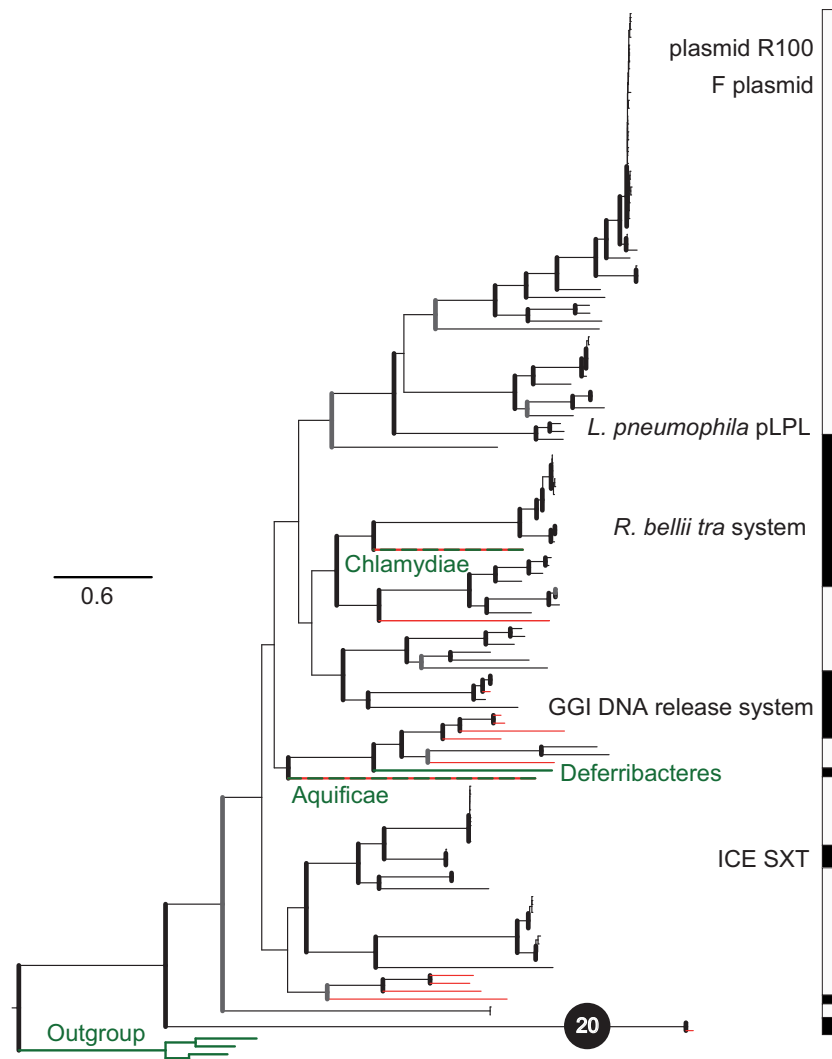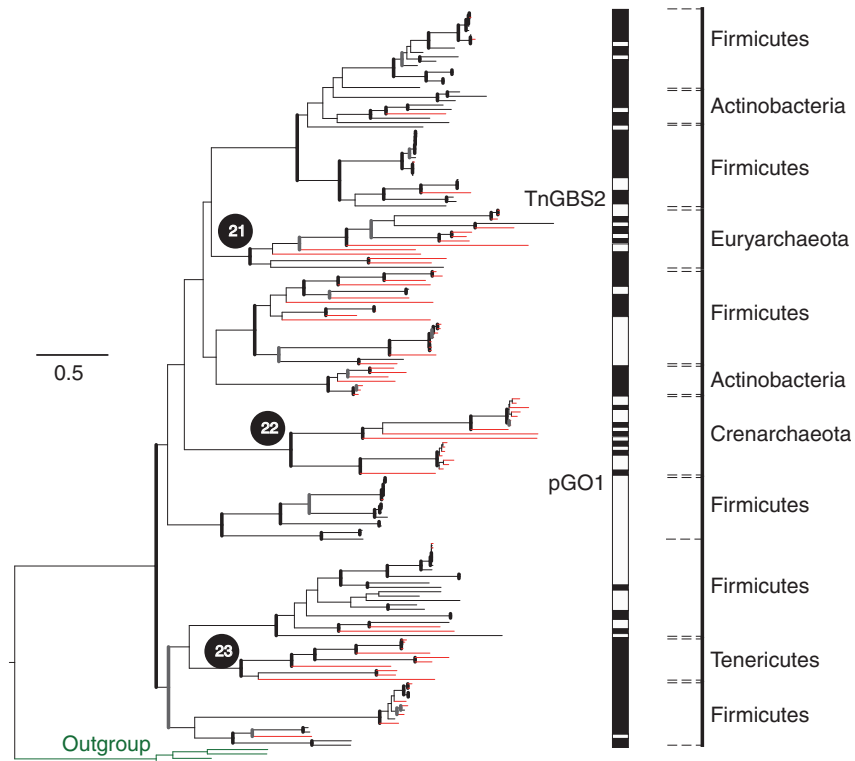
Fig. 5. Phylogenetic analysis of MPF$_F$ VirB4 proteins. Bold vertical black lines represent nodes with a high support value (bootstrap >90%), and bold vertical gray lines represent nodes with a support value between 70% and 90%. Green branches correspond to taxa that are not from Proteobacteria (plus the outgroup). Red branches represent the VirB4 not associated to a relaxase (MOBless T4SS). Green and red dotted branches represent MOBless T4SS that are not from Proteobacteria. The bar on the right stands for the chromosomal (black) or plasmidic (white) proteins. Known representative systems are labeled. The GGI DNA release system corresponds to the *N. gonorrhoeae* gonococcal genetic island (Hamilton et al. 2005). Number in circles refers to the analysis of robustness in table 1 (identified in the third column of table 1); black background stands for a high support (≥70% bootstrap in the best-scoring alignment) and gray background for a moderate support (≥50% bootstrap in the best-scoring alignment).
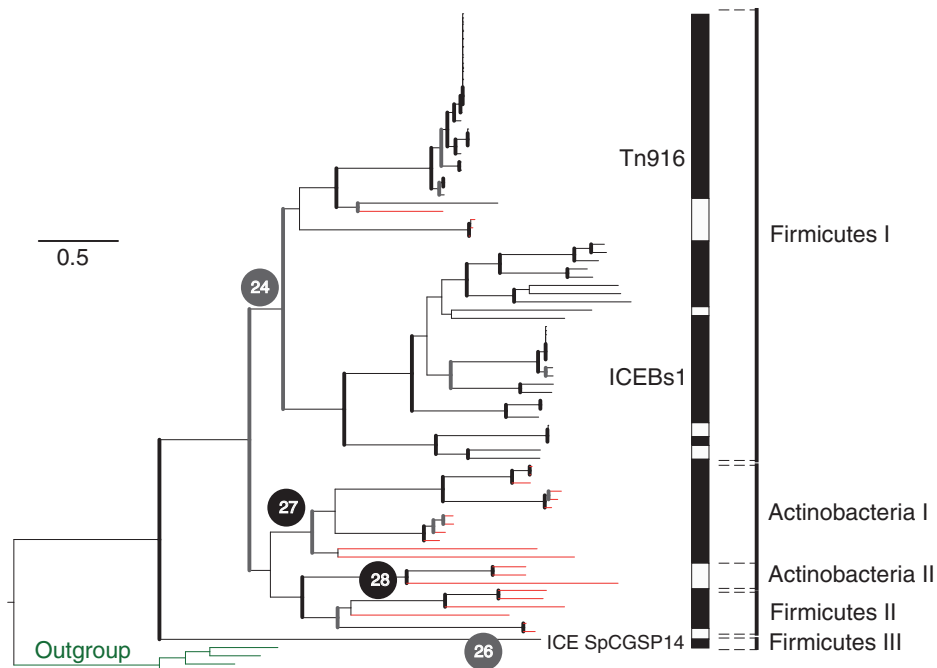
trees), and a group of Archaea unlikely to be monophyletic (bootstrap of only 17–29%) with a clear separation between Euryarchaeota and Crenarchaeota (91–96%, respectively, and 100% bootstrap support for each clade) (fig. 6). The deeper relations between these clades are difficult to disentangle, given the low bootstrap supports of the basal nodes. Within the Firmicutes clades, we find the main divisions, i.e., Bacillales, Lactobacillales, and Clostridia, scattered in the tree. This suggests that, once a conjugative system arose in this phylum, it spreads early among the main divisions, and transfers between divergent clades were maintained through a certain moment in evolution. The monophyly of monoderms in the VirB4 tree suggests that monoderms acquired conjugative systems by transfer from diderms. This early acquisition was followed by the adaptation of the T4SS to monoderms. Finally, frequent conjugation between diderms

contributed to the scattered distribution of taxa in the phylogenetic tree of MPF$_{FATA}$ and MPF$_{FA}$.

The MPF$_{FA}$ clade includes two groups of Actinobacteria intermingled with three groups of Firmicutes (<5% bootstrap support for a net separation of the two clades) (fig. 7). The most basal group (Firmicutes III in fig. 7) is constituted by a few elements from Firmicutes (bootstrap support for this basal position of 52–100%, table 1). This suggests that the ancestral conjugative system might have arisen within Firmicutes from which it was transferred to Actinobacteria. This is consistent with the observation of a basal group, including only Firmicutes and Tenericutes in the sister MPF$_{FATA}$ tree (fig. 6). The subsequent split in the MPF$_{FA}$ group separates a clade with Actinobacteria and Firmicutes II from Firmicutes I (fig. 7). The latter encodes TcpA as a putative T4CP, which further supports the monophyly of

FIG. 6. Phylogenetic analysis of MPF$_{FATA}$ VirB4 proteins. Bold vertical black lines represent nodes with a high support value (bootstrap >90%), and bold vertical gray lines represent nodes with a support value between 70% and 90%. Squared brackets delimit the different taxonomic clades (plus the outgroup). Red branches represent the VirB4 not associated to a relaxase (MOBless T4SS). The bar on the right stands for the chromosomal (black) or plasmidic (white) proteins. Numbers in circles refer to the analysis of robustness in table 1 (identified in the third column of table 1); black background stands for a high support (≥70% bootstrap in the best-scoring alignment) and gray background for a moderate support (≥50% bootstrap in the best-scoring alignment).

FIG. 7. Phylogenetic analysis of MPF$_{FA}$ VirB4 proteins. Bold vertical black lines represent nodes with a high support value (bootstrap >90%), and bold vertical gray lines represent nodes with a support value between 70% and 90%. Squared brackets delimit the different taxonomic clades (plus the outgroup). Red branches represent the VirB4 not associated to a relaxase (MOBless T4SS). The bar on the right stands for the chromosomal (black) or plasmidic (white) proteins. Numbers in circles refer to the analysis of robustness in table 1 (identified in the third column of table 1); black background stands for a high support (≥70% bootstrap in the best-scoring alignment) and gray background for a moderate support (≥50% bootstrap in the best-scoring alignment).

Firmicutes I based on VirB4 sequences (52–99% of bootstrap support). Homologs of TcpA were found in the plasmid pCW3 of *Clostridium perfringens*, in ICE*Bs1* of *B. subtilis*, and in Tn916 of *Enterococcus faecalis* (Teng et al. 2008). We found that 63% of the TcpA$_{pCW3}$ hits were colocalized with VirB4 in MPF$_{FA}$ systems of Firmicutes, and all 47 of these regions lacked a VirD4-like protein. This gives further credit to the hypothesis that TcpA is an alternative T4CP (Parsons et al. 2007; Steen et al. 2009). TcpA-associated systems are, with one single exception, also associated with MOB$_T$. The MOB$_T$ relaxase of Tn916 (Orf20), when assisted by the accessory protein Int, produces strand- and sequence-specific cleavage generating a 3′-OH (Rocco, Churchward 2006). Thus, although phylogenetically different, TcpA and VirD4 T4CPs seem to be both alternatives for ssDNA conjugation, suggesting the recruitment of a new dsDNA translocase to make ssDNA conjugation in this subclade of MPF$_{FA}$. This process was concomitant with the acquisition of a very atypical relaxase, which has no similarity with other relaxases, and instead resembles replication initiator factors of phages and plasmids (Garcillan-Barcia et al. 2009). Interestingly, ICE*Bs1* transfers extremely fast within chains of bacteria (Babic et al. 2011). It is currently unknown if this behavior reminiscent of TraB, which as we showed earlier is a closer homolog of TcpA than VirD4, has associated mechanistic analogies, e.g., if TcpA might have maintained a dsDNA translocase activity.

Evolution of MPF$_T$

Except for VirB4, which has homologs in every T4SS, most of our protein profiles for a given MPF type allow identifying homologs only within the respective MPF system. Several of these are nearly ubiquitous within a given MPF type, and we have previously used them to class MPF types in plasmids and chromosomes (Smillie et al. 2010; Guglielmini et al. 2011). To analyze in detail the patterns of presence and absence of MPF specific genes, we analyzed the MPF$_T$ system, the best studied and most frequently found in sequenced genomes. Its prototype is the *vir* system of the *A. tumefaciens* plasmid Ti, which encodes 11 genes: *virB1* to *virB11*. We built HMM profiles for each protein and used them to scan plasmids for homologs. We excluded chromosomes from this particular analysis because these are more likely to contain inactivated T4SS ongoing genetic degradation, and this would lead to the introduction of false positives in the analysis. Most systems include between 8 and 11 out of the 11 genes, but not always the same genes are missing (supplementary fig. S1, Supplementary Material online). The only gene nonessential for conjugation in this system, the lytic transglycosylase *virB1* (Berger and Christie 1994), is often missing or not identified (absent in 48% of the MPF$_T$). The small VirB7 lipoprotein interacts with VirB9 and performs some sort of stabilizing function (Spudich et al. 1996) and is also often missed in the search (67%). The most basal branches within the MPF$_T$ tree show an increasing number of proteins that we fail to detect, most notably the minor component of the pilus VirB5 (missing in 25%). VirB5 and VirB7 are the most exposed proteins at the cell outer membrane (Christie and Vogel 2000; Fronzes et al. 2009) and

are cell receptors for phages and the immune system (Haase et al. 1995; Harris and Silverman 2002; Alvarez-Martinez and Christie 2009). They are therefore likely to evolve rapidly because of these two types of selection pressure. Accordingly, both VirB5 and VirB7 show evidence of positive selection in the T4SS$_T$ of *Bartonella* (Engel et al. 2011). Hence, the patterns of gene absence are probably caused by both gene absence and rapid evolution of some T4SS components.

The names of the different *vir* genes correspond to their order within the prototype VirB$_{Ti}$ system. This prototype gene order pattern (from 1 to 11 in ascending order) is conserved in a large fraction of the MPF$_T$ (fig. 4). For almost all MPF$_T$ loci, the order is strictly conserved for a core composed of *virB2*, *virB3*, *virB4*, *virB8*, *virB9*, and *virB10*. As mentioned earlier, *virB7* is often missed by our scan. The gene *virB11* can be found before *virB2*, and *virB1* after *virB10*; this defines the gene order depicted in green in figure 4. Importantly, the node separating the two large clades of MPF$_T$ relative to gene order is also highly supported by the analysis of the VirB4 phylogeny (98% bootstrap). The genes *virB5* and *virB6* are sometimes placed after *virB10* (fig. 4, in dark blue), which seems a derivation from the previous pattern. These three patterns of gene order represent more than 80% of all the MPF$_T$. Interestingly, the prototype pattern is less often found on chromosomes, the "green" pattern being more represented. It is difficult to say for the moment if this difference is a simple consequence of the higher frequency of chromosomal T4SS in this part of the tree or if this gene order is adaptive in chromosomal loci. Importantly, the clusters of gene order in the tree accurately reflect the phylogeny of VirB4. This is further evidence that recombination of distant VirB4 variants rarely occurs, even within MPF types.

Considering the number of possible permutations and the relatively low number of different patterns, these data suggest that the gene order within *vir* systems is highly constrained in most genes, with four genes often being found in different positions (*virB1*, *virB5*, *virB6*, and *virB11*). The gene succession is also preserved; indeed, the vast majority of *virB* genes are directly adjacent, suggesting strong counterselection for insertions in the loci (data not shown). Highly conserved gene order at a locus is a sign of selection for a given organization of transcription (Rocha 2006). In the case of large protein complexes, such organization can give rise to an ordered assembly of the complex, as it has been shown for the flagellum (Kutsukake et al. 1990). Gene order conservation thus suggests conservation of a developmental plan. The variants we see, outlined in figure 4, could reflect innovations in this plan.

T4SS Exaptation

We recently uncovered that a large fraction of T4SS lack neighboring relaxases (Guglielmini et al. 2011). A few observations suggest that most of these are not genetic elements ongoing degradation. First, these MOBless T4SS are more often chromosomal than plasmidic. Second, many of these chromosomal elements lack neighboring integrases. Third, the T4SS known to deliver proteins were classed as

MOBless T4SS. These observations suggest that many MOBless T4SS are not undergoing degradation but that, instead, they result from recruitment of conjugation systems for other functions (exaptation). The VirB4 phylogeny confirms that, within MPF$_T$, the loss of the relaxase occurred many times and that this pattern is also found among the other MPF types (figs. 4–7). Just like conjugative systems of ICEs and plasmids are interspersed in the phylogenetic trees (Gugliel-mini et al. 2011), MOBless T4SS are interspersed with conjugative systems. This shows that MOBless T4SS arose frequently and in independent instances. The only exception concerns the Archaea and the Actinobacteria, for which the lack of known relaxases has been pointed out before (Garcillan-Barcia et al. 2009). In these clades, it is likely that the abundance of MOBless T4SS predominantly reflects the presence of unknown relaxases. Importantly, the T4SS that are experimentally known to have nonconjugation-related functions are interspersed in the trees of MPFT and MPFF (figs. 4 and 5). This suggests that conjugative T4SS have been frequently recruited for other functions.

An Evolution-Based Classification System for MPF

The lack of an all-encompassing classification scheme for conjugative systems and the extreme diverse gene nomenclature for homologous conjugation genes greatly and unnecessarily complicates the analysis of the literature of the domain. We suggest that the phylogeny of VirB4, the only ubiquitously recognizable protein of T4SS, could be used to class ssDNA conjugative systems and other T4SS. This could be the foundation for the much-needed gene name standardization in the literature and databases. The model systems of the *vir* operon of *A. tumefaciens* Ti plasmid (MPF$_T$), F plasmid (MPF$_F$), R64 plasmid (MPF$_I$), and ICEHin1056 (MPF$_G$) could be used for all Proteobacteria and possibly for other diderm clades such as Acidobacteria. Four other MPF types for now cover the diversity of all the other systems in so far as the VirB4 phylogeny is concerned. These would include a type that for the moment only includes Bacteroides (MPF$_B$) and another that includes only Cyanobacteria (MPF$_C$). The classification would also include the two types that are specific to monoderms, the MPF$_{FA}$ and MPF$_{FATA}$. The MPF$_{FA}$ type, given its heterogeneity in the use of T4CP, might be split into two groups when more is known about the differences in the biochemistry of conjugation in the group. The advantage of this classification is that it is based on evolutionary biology, tends to reflect similarity between elements, and can be done even when one knows yet relatively little of the biochemistry of the elements being classed.

We believe there is little risk of an excessive inflation in the number classes of MPF with the uncovering of new uncultivated bacterial clades. First, all monoderms seem to cluster in only two sister clades. Second, MPF of a string of poorly sampled clades of diderms are classed along with the four common MPF types of Proteobacteria. Some previous classifications of conjugation systems have been based on the type of replicon or on the secretion substrate. The former, separating conjugative plasmids from ICEs, are pertinent to class

mobile elements but are inadequate to separate conjugative systems because MPF cannot be discriminated based on the type of the host replicon (Guglielmini et al. 2011). Classifications regarding the secretion substrate, i.e., proteins or DNA–protein complexes, pertain to the role of the T4SS and its impact on genetic mobility. They are extremely important to understand the adaptive role of T4SS in a bacterium. However, as shown in this work, they carry little information allowing classification of the T4SS.

T4SS were divided on structural grounds in two classes: T4SSa-including elements from MPF$_T$ and MPF$_F$ and T4SSb including elements from MPF$_I$ (Christie and Vogel 2000). These two classes can easily be mapped into the VirB4 phylogeny in these three different MPF types. Although this classification reflects important differences in terms of conjugative pili among Proteobacteria, it no longer represents the diversity of T4SS. It is unclear how MPF$_G$ or any MPF type not present in Proteobacteria should be classed in this scheme (fig. 3). Our analysis provides a natural classification scheme for T4SS and may also help highlight the commonalities and differences between systems. Together with the classification of relaxases (Garcillan-Barcia et al. 2009; Guglielmini et al. 2011), it can be easily extended to class ssDNA conjugative systems. Furthermore, this classification system can be applied to partial data, e.g., from metagenomics, because it requires the identification of a single gene.

Conclusion

Our work provides a scenario for the evolution of conjugation and T4SS from their origin to recent exaptations (fig. 8). These results suggest that conjugation is a very ancient process that arose in two independent ways for ssDNA and dsDNA mechanisms, starting from ancestrally related AAA+ ATPases involved in DNA translocation. Conjugation of ssDNA is by far the best studied and also the mechanism most frequently found in prokaryotes. It probably appeared very early among bacteria with two cell envelopes, possibly ancient Proteobacteria, and from there it spread to all clades of prokaryotes. The T4SS of monoderms seem less complex, in that they involve fewer genes (Grohmann et al. 2003), and could initially evolve by gene deletion from the larger T4SS of diderms. Our evolutionary scenario links together all known ssDNA conjugative systems, and their T4SS, by the common ancestry of VirB4. Several observations show the validity of the use of this protein for the classification of T4SS. First, it is the only ubiquitous protein in T4SS. Second, its phylogeny closely matches those of other conserved proteins, notably the VirD4. Third, patterns of the presence/absence of MPF specific genes match the VirB4 phylogeny. Fourth, the order of MPF-specific genes, at least in MPF$_T$, also matches the VirB4 phylogeny.

The structure of the VirB4 tree, with its robust separation in eight large clades, reflects in part an effect of the cell envelope. Indeed, once systems arose within a clade with a peculiar membrane structure, they tended to adapt to this cell structure and were not further passed on to other clades. This resulted in large clades of VirB4, including monoderms—such as Archaea or Firmicutes—or diderms with peculiar
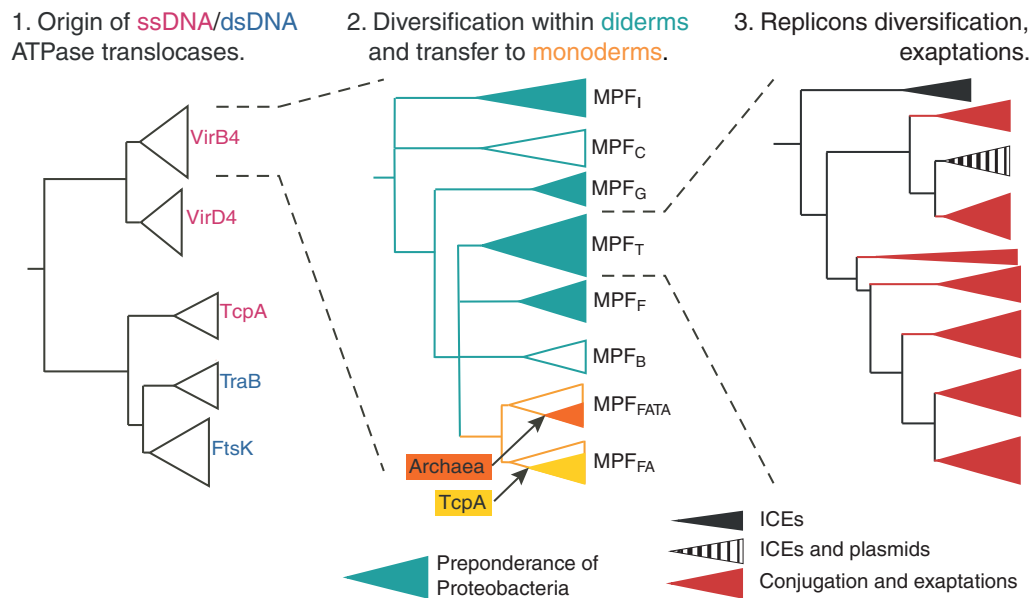
Fig. 8. Model for the evolution of conjugation. First, DNA translocases diversify into a number of families that are involved in conjugation (ssDNA for VirB4, VirD4, and TcpA, and dsDNA for TraB). Second, ssDNA conjugation diversified in a series of clades that are the basis of MPF classes. Several of these show a preponderance of Proteobacteria. Transfer of a conjugative system to monoderms led to the diversification and further spread within Firmicutes, Actinobacteria, Archaea, and Tenericutes. Among MPF$_{FA}$, some elements engaged in a dramatically different system, including TcpA and the relaxase MOB$_T$. Finally, at much shorter evolutionary distances, we observe diversification of conjugative systems among integrative (ICEs) and extrachromosomal (plasmids) elements. Exaptation of the conjugative systems for protein delivery, DNA uptake and other, also arise relatively late in the evolutionary scale.

membrane compositions such as Cyanobacteria (Wada and Murata 1998) or Bacteroides (An et al. 2011). Adaptation of the T4SS to such cell envelopes is likely to increase the efficiency of conjugation within taxa but at the cost of reducing its efficiency between taxa, effectively leading to T4SS specialization. This process has the potential to bias the rate and direction of genetic transfer between prokaryotes and thus shape the networks of gene sharing (Halary et al. 2010; Dagan 2011). Notably, it might contribute to the observed coherence between high bacterial taxonomic ranks (Philippot et al. 2010).

Surprisingly, one group of ssDNA T4SS has radically changed into a system with a new T4CP (TcpA) and relaxase (MOB$_T$). Although the cognate VirB4 protein fits clearly in our T4SS classification and is presumably representative of the evolutionary history of the remaining proteins of the MPF$_{FA}$ T4SS, the replacement of the T4CP suggests that the evolution of the coupling protein can in certain cases differ radically from the one of the T4SS. In several cases (fig. 3), this seems to reflect the double evolutionary constraint of T4CP in adapting to both the T4SS and to the relaxase.

Our work also shows that exaptations of T4SS can occur frequently in the evolutionary history. Conjugation consists in the secretion of a nucleoprotein complex. Passing from this function to a protein secretion system can probably occur in few evolutionary steps. Accordingly, several systems are known to transfer both proteins and relaxosomes (Vogel et al. 1998; Fernandez-Gonzalez et al. 2011; Schroder et al. 2011). Furthermore, conjugation systems and MOBless T4SS can interchange components without loss of function (de Paz et al. 2005). The exaptation of *H. pylori comB*

system is more surprising because this system has evolved into a DNA import mechanism (Hofreuter et al. 2001). Several other protein secretion systems are thought to be exaptations, e.g., nonflagellum T3SS are related with the bacterium flagellum and T6SS show structural homologies with phages (Ginocchio et al. 1994; Pell et al. 2009). Yet, T4SS present an uncommon case in that exaptations occurred multiple times in the evolutionary history. Given the present results, it is not unlikely that novel exaptations, e.g., protein transfer among bacteria, are present among the poorly studied MOBless T4SS of free-living bacteria.

Supplementary Material

Supplementary table S1 and figure S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

References

Allen HK, Donato J, Wang HH, Cloud-Hansen KA, Davies J, Handelsman J. 2010. Call of the wild: antibiotic resistance genes in natural environments. *Nat Rev Microbiol.* 8:251–259.

Alvarez-Martinez CE, Christie PJ. 2009. Biological diversity of prokaryotic type IV secretion systems. *Microbiol Mol Biol Rev.* 73:775–808.

Amabile-Cuevas CF, Chicurel ME. 1992. Bacterial plasmids and gene flux. *Cell* 70:189–199.

An D, Na C, Bielawski J, Hannun YA, Kasper DL. 2011. Membrane sphingolipids as essential molecular signals for Bacteroides survival in the intestine. *Proc Natl Acad Sci U S A.* 108(Suppl 1):4666–4671.

Babic A, Berkmen MB, Lee CA, Grossman AD. 2011. Efficient gene transfer in bacterial cell chains. *MBio* 2:e00027–00011.

Berger BR, Christie PJ. 1993. The *Agrobacterium tumefaciens* virB4 gene product is an essential virulence protein requiring an intact nucleoside triphosphate-binding domain. *J Bacteriol.* 175:1723–1734.

Berger BR, Christie PJ. 1994. Genetic complementation analysis of the *Agrobacterium tumefaciens* virB operon: *virB2* through *virB11* are essential virulence genes. *J Bacteriol.* 176:3646–3660.

Bigot S, Sivanathan V, Possoz C, Barre FX, Cornet F. 2007. FtsK, a literate chromosome segregation machine. *Mol Microbiol.* 64:1434–1441.

Burrus V, Pavlovic G, Decaris B, Guedon G. 2002. Conjugative transposons: the tip of the iceberg. *Mol Microbiol.* 46:601–610.

Cabezon E, Lanza VF, Arechaga I. 2012. Membrane-associated nanomotors for macromolecular transport. *Curr Opin Biotechnol.* 23:537–44.

Carter MQ, Chen J, Lory S. 2010. The *Pseudomonas aeruginosa* pathogenicity island PAPI-1 is transferred via a novel type IV pilus. *J Bacteriol.* 192:3249–3258.

Chen I, Christie PJ, Dubnau D. 2005. The ins and outs of DNA transfer in bacteria. *Science* 310:1456–1460.

Christie PJ, Vogel JP. 2000. Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends Microbiol.* 8:354–360.

Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P. 2006. Toward automatic reconstruction of a highly resolved tree of life. *Science* 311:1283–1287.

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

Dagan T. 2011. Phylogenomic networks. *Trends Microbiol.* 19:483–491.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27:1164–1165.

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

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

de Paz HD, Larrea D, Zunzunegui S, Dehio C, de la Cruz F, Llosa M. 2010. Functional dissection of the conjugative coupling protein TrwB. *J Bacteriol.* 192:2655–2669.

de Paz HD, Sangari FJ, Bolland S, Garcia-Lobo JM, Dehio C, de la Cruz F, Llosa M. 2005. Functional interactions between type IV secretion systems involved in DNA transfer and virulence. *Microbiology* 151:3505–3516.

Ding H, Hynes MF. 2009. Plasmid transfer systems in the rhizobia. *Can J Microbiol.* 55:917–927.

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

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.

Engel P, Salzburger W, Liesch M, et al. (11 co-authors). 2011. Parallel evolution of a type IV secretion system in radiating lineages of the host-restricted bacterial pathogen *Bartonella*. *PLoS Genet.* 7:e1001296.

Fernandez-Gonzalez E, de Paz HD, Alperi A, Agundez L, Faustmann M, Sangari FJ, Dehio C, Llosa M. 2011. Transfer of R388 derivatives by a pathogenesis-associated type IV secretion system into both bacteria and human cells. *J Bacteriol.* 193:6257–6265.

Frank AC, Alsmark CM, Thollesson M, Andersson SG. 2005. Functional divergence and horizontal transfer of type IV secretion systems. *Mol Biol Evol.* 22:1325–1336.

Fronzes R, Christie PJ, Waksman G. 2009. The structural biology of type IV secretion systems. *Nat Rev Microbiol.* 7:703–714.

Frost LS, Leplae R, Summers AO, Toussaint A. 2005. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol.* 3:722–732.

Fullner KJ, Lara JC, Nester EW. 1996. Pilus assembly by Agrobacterium T-DNA transfer genes. *Science* 273:1107–1109.

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

Gascuel O, Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59:307–321.

Ghigo JM. 2001. Natural conjugative plasmids induce bacterial biofilm development. *Nature* 412:442–445.

Ghinet MG, Bordeleau E, Beaudin J, Brzezinski R, Roy S, Burrus V. 2011. Uncovering the prevalence and diversity of integrating conjugative elements in Actinobacteria. *PLoS One* 6:e27846.

Ginocchio CC, Olmsted SB, Wells CL, Galan JE. 1994. Contact with epithelial cells induces the formation of surface appendages on *Salmonella typhimurium*. *Cell* 76:717–724.

Gogarten JP, Doolittle WF, Lawrence JG. 2002. Prokaryotic evolution in light of gene transfer. *Mol Biol Evol.* 19:2226–2238.

Golebiewski M, Kern-Zdanowicz I, Zienkiewicz M, Adamczyk M, Zylinska J, Baraniak A, Gniadkowski M, Bardowski J, Ceglowski P. 2007. Complete nucleotide sequence of the pCTX-M3 plasmid and its involvement in spread of the extended-spectrum beta-lactamase gene blaCTX-M-3. *Antimicrob Agents Chemother.* 51:3789–3795.

Gomis-Ruth FX, Sola M, de la Cruz F, Coll M. 2004. Coupling factors in macromolecular type-IV secretion machineries. *Curr Pharm Des.* 10:1551–1565.

Gould SJ, Vrba ES. 1982. Exaptation—a missing term in the science of form. *Paleobiology* 8:4–15.

Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol.* 27:221–224.

Grohmann E, Muth G, Espinosa M. 2003. Conjugative plasmid transfer in gram-positive bacteria. *Microbiol Mol Biol Rev.* 67:277–301.

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

Gupta RS. 2004. The phylogeny and signature sequences characteristics of Fibrobacteres, Chlorobi, and Bacteroidetes. *Crit Rev Microbiol.* 30:123–143.

Haase J, Lurz R, Grahn AM, Bamford DH, Lanka E. 1995. Bacterial conjugation mediated by plasmid RP4: RSF1010 mobilization, donor-specific phage propagation, and pilus production require the same Tra2 core components of a proposed DNA transport complex. *J Bacteriol.* 177:4779–4791.

Halary S, Leigh JW, Cheaib B, Lopez P, Bapteste E. 2010. Network analyses structure genetic diversity in independent genetic worlds. *Proc Natl Acad Sci U S A.* 107:127–132.

Hamilton HL, Dominguez NM, Schwartz KJ, Hackett KT, Dillard JP. 2005. *Neisseria gonorrhoeae* secretes chromosomal DNA via a novel type IV secretion system. *Mol Microbiol.* 55:1704–1721.

Harris RL, Silverman PM. 2002. Roles of internal cysteines in the function, localization, and reactivity of the TraV outer membrane lipoprotein encoded by the F plasmid. *J Bacteriol.* 184:3126–3129.

Hofreuter D, Odenbreit S, Haas R. 2001. Natural transformation competence in *Helicobacter pylori* is mediated by the basic components of a type IV secretion system. *Mol Microbiol.* 41:379–391.

Iyer LM, Makarova KS, Koonin EV, Aravind L. 2004. Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. *Nucleic Acids Res.* 32:5260–5279.

Juhas M, Crook DW, Dimopoulou ID, Lunter G, Harding RM, Ferguson DJ, Hood DW. 2007. Novel type IV secretion system involved in propagation of genomic islands. *J Bacteriol.* 189:761–771.

Juhas M, Crook DW, Hood DW. 2008. Type IV secretion systems: tools of bacterial horizontal gene transfer and virulence. *Cell Microbiol.* 10: 2377–2386.

Juhas M, Power PM, Harding RM, et al. (16 co-authors). 2007. Sequence and functional analyses of *Haemophilus* spp. genomic islands. *Genome Biol.* 8:R237.

Katoh K, Toh H. 2010. Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* 26:1899–1900.

Kim SR, Komano T. 1997. The plasmid R64 thin pilus identified as a type IV pilus. *J Bacteriol.* 179:3594–3603.

Kutsukake K, Ohya Y, Iino T. 1990. Transcriptional analysis of the flagellar regulon of *Salmonella typhimurium. J Bacteriol.* 172:741–747.

Lederberg J, Tatum EL. 1946. Gene recombination in *E. coli. Nature* 158:558.

Llosa M, Gomis-Ruth FX, Coll M, de la Cruz Fd F. 2002. Bacterial conjugation: a two-step mechanism for DNA transport. *Mol Microbiol.* 45:1–8.

Llosa M, Zunzunegui S, de la Cruz F. 2003. Conjugative 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.

Mount DW. 2004. Bioinformatics: sequence and genome analysis. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Nogueira T, Rankin DJ, Touchon M, Taddei F, Brown SP, Rocha EP. 2009. Horizontal gene transfer of the secretome drives the evolution of bacterial cooperation and virulence. *Curr Biol.* 19:1683–1691.

Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol.* 302: 205–217.

Nystedt B, Frank AC, Thollesson M, Andersson SG. 2008. Diversifying selection and concerted evolution of a type IV secretion system in *Bartonella. Mol Biol Evol.* 25:287–300.

Ochman H, Lerat E, Daubin V. 2005. Examining bacterial species under the specter of gene transfer and exchange. *Proc Natl Acad Sci U S A.* 102(Suppl. 1):6595–6599.

Parsons JA, Bannam TL, Devenish RJ, Rood JI. 2007. TcpA, an FtsK/SpoIIIE homolog, is essential for transfer of the conjugative plasmid pCW3 in *Clostridium perfringens. J Bacteriol.* 189:7782–7790.

Pell LG, Kanelis V, Donaldson LW, Howell PL, Davidson AR. 2009. The phage lambda major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. *Proc Natl Acad Sci U S A.* 106:4160–4165.

Pena A, Ripoll-Rozada J, Zunzunegui S, Cabezon E, de la Cruz F, Arechaga I. 2011. Autoinhibitory regulation of TrwK, an essential VirB4 ATPase in type IV secretion systems. *J Biol Chem.* 286:17376–17382.

Philippe H, Brinkmann H, Lavrov DV, Littlewood DT, Manuel M, Worheide G, Baurain D. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol.* 9: e1000602.

Philippot L, Andersson SG, Battin TJ, Prosser JI, Schimel JP, Whitman WB, Hallin S. 2010. The ecological coherence of high bacterial taxonomic ranks. *Nat Rev Microbiol.* 8:523–529.

Planet PJ, Kachlany SC, DeSalle R, Figurski DH. 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.

Possoz C, Ribard C, Gagnat J, Pernodet JL, Guerineau M. 2001. The integrative element pSAM2 from *Streptomyces*: kinetics and mode of conjugal transfer. *Mol Microbiol.* 42:159–166.

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

Rocco JM, Churchward G. 2006. The integrase of the conjugative transposon Tn916 directs strand- and sequence-specific cleavage of the origin of conjugal transfer, oriT, by the endonuclease Orf20. *J Bacteriol.* 188:2207–2213.

Rocha EPC. 2006. Inference and analysis of the relative stability of bacterial chromosomes. *Mol Biol Evol.* 23:513–522.

Salgado-Pabon W, Jain S, Turner N, van der Does C, Dillard JP. 2007. A novel relaxase homologue is involved in chromosomal DNA processing for type IV secretion in *Neisseria gonorrhoeae. Mol Microbiol.* 66:930–947.

Schroder G, Schuelein R, Quebatte M, Dehio C. 2011. Conjugative DNA transfer into human cells by the VirB/VirD4 type IV secretion system of the bacterial pathogen *Bartonella henselae. Proc Natl Acad Sci U S A.* 108:14643–14648.

Seubert A, Hiestand R, de la Cruz F, Dehio C. 2003. A bacterial conjugation machinery recruited for pathogenesis. *Mol Microbiol.* 49: 1253–1266.

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

Spudich GM, Fernandez D, Zhou XR, Christie PJ. 1996. Intermolecular disulfide bonds stabilize VirB7 homodimers and VirB7/VirB9 heterodimers during biogenesis of the *Agrobacterium tumefaciens* T-complex transport apparatus. *Proc Natl Acad Sci U S A.* 93: 7512–7517.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.

Steen JA, Bannam TL, Teng WL, Devenish RJ, Rood JI. 2009. The putative coupling protein TcpA interacts with other pCW3-encoded proteins to form an essential part of the conjugation complex. *J Bacteriol.* 191:2926–2933.

Tato I, Zunzunegui S, de la Cruz F, Cabezon E. 2005. TrwB, 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.

Teng WL, Bannam TL, Parsons JA, Rood JI. 2008. Functional character-ization and localization of the TcpH conjugation protein from *Clostridium perfringens*. *J Bacteriol.* 190:5075–5086.

Thomas CM. 2000. Horizontal gene pool: bacterial plasmids and gene spread. Amsterdam: CRC.

Thompson DV, Melchers LS, Idler KB, Schilperoort RA, Hooykaas PJ. 1988. Analysis of the complete nucleotide sequence of the *Agrobacterium tumefaciens* virB operon. *Nucleic Acids Res.* 16:4621–4636.

van der Meer JR, Sentchilo V. 2003. Genomic islands and the evolution of catabolic pathways in bacteria. *Curr Opin Biotechnol.* 14:248–254.

Vogel J, Andrews H, Wong S, Isberg R. 1998. Conjugative transfer by the virulence system of *Legionella pneumophila*. *Science* 279:873–876.

Vogelmann J, Ammelburg M, Finger C, Guezguez J, Linke D, Flotenmeyer M, Stierhof YD, Wohlleben W, Muth G. 2011. Conjugal plasmid transfer in *Streptomyces* resembles bacterial chromosome segrega-tion by FtsK/SpoIIIE. *EMBO J.* 30:2246–2254.

Wada H, Murata N. 1998. Membrane lipids in cyanobacteria. In: Siegenthaler P-A, Murata N, editors. Lipids in photosynthesis: structure, function and genetics. Dordrecht (The Netherlands): Springer. p. 65–81.

Wallden K, Williams R, Yan J, Lian PW, Wang L, Thalassinos K, Orlova EV, Waksman G. 2012. Structure of the VirB4 ATPase, alone and bound to the core complex of a type IV secretion system. *Proc Natl Acad Sci U S A.* 109:11348–11353.

Ward JE, Akiyoshi DE, Regier D, Datta A, Gordon MP, Nester EW. 1988. Characterization of the virB operon from an *Agrobacterium tumefaciens* Ti plasmid. *J Biol Chem.* 263:5804–5814.

Weiss AA, Johnson FD, Burns DL. 1993. Molecular characterization of an operon required for pertussis toxin secretion. *Proc Natl Acad Sci U S A.* 90:2970–2974.

Wozniak RA, Waldor MK. 2010. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Microbiol.* 8:552–563.

Wu LJ, Errington J. 1998. Use of asymmetric cell division and spoIIIE mutants to probe chromosome orientation and organization in *Bacillus subtilis*. *Mol Microbiol.* 27:777–786.