Functional Divergence and Horizontal Transfer of Type IV Secretion Systems

A. Carolin Frank, Cecilia M. Alsmark,1 Mikael Thollesson, and Siv G. E. Andersson
Department of Molecular Evolution, Evolutionary Biology Center, Uppsala University, Uppsala, 752 36, Sweden

The type IV secretion system (TFSSs) is a multifunctional family of translocation pathways that mediate the transfer of DNA among bacteria and deliver DNA and proteins to eukaryotic cells during bacterial infections. Horizontal transmission has dominated the evolution of the TFSS, as demonstrated here by a lack of congruence between the tree topology inferred from components of the TFSS and the presumed bacterial species divergence pattern. A parsimony analysis suggests that conjugation represents the ancestral state and that the divergence from conjugation to secretion of effector molecules has occurred independently at multiple sites in the tree. The result shows that the nodes at which functional shifts have occurred coincide with those of horizontal gene transfers among distantly related bacteria. We suggest that it is the transfer between species that paved the way for the divergence of the TFSSs and discuss the general role of horizontal gene transfers for the evolution of novel gene functions.

Introduction

Gene duplication is one of the major mechanisms by which novel gene functions arise. The two copies derived from a duplication event initially have the same function, but the accumulation of mutations in both copies may result in the loss of one copy or a functional divergence between the two copies (Ohno 1970; Ohta 1988a; Ohta 1988b; Taylor and Raes 2004). Given that the rate of detrimental mutations is normally much higher than the rate of beneficial mutations, selection is required for the persistence of a novel or divergent gene function. This is the basis for the evolution of gene families, small and large, in which the individual members have related, but not identical protein functions.

Likewise, it can be anticipated that if a horizontally transferred gene initially confers a weak or nonessential function to the recipient lineage, it can be a source of raw material for molecular innovation. If the accumulation of mutations results in the emergence of a new protein function that is beneficial to the cell, the gene will be retained in the population. The role of horizontal transfer in innovation and adaptation is well studied, and it is widely accepted that ready-to-use horizontally transferred genes (e.g., antibiotic resistance genes) can confer a fitness advantage to a recipient organism. However, the possibility that novel gene functions arise as a result of horizontal transfer has to our knowledge not been much discussed.

To investigate if horizontal transfers provide an opportunity for novel gene functions to evolve, we have here studied plasmid-encoded conjugation systems, a family with both large potential for horizontal transfers and documented diverse functions. The bacterial conjugation machinery belongs to a multifunctional family of translocation systems, the so-called type IV secretion systems (TFSSs) (Ding, Atmakuri, and Christie 2003). These transport systems are functionally diverse both in terms of the transported substrate, which can be both DNA and protein, and in terms of the recipient cell, which can be a cell of the same or different bacterial species or a eukaryotic cell.

The TFSS family members have broadly been grouped into three categories; conjugation, effector translocation, and DNA uptake and transformation (Ding, Atmakuri, and Christie 2003). Plasmid-encoded conjugation systems mediate the transfer of plasmids and chromosomal DNA in bacterial populations. Effector translocators inject protein or DNA-protein complexes into eukaryotic host cells or secrete toxic molecules into the extracellular environment during the infectious process. DNA-uptake and release systems have so far only been found in a few species, like Helicobacter pylori (Hofreuter, Odenbreit, and Haas 2001), Campylobacter jejuni (Bacon et al. 2002), and Neisseria gonorrhoeae (Dillard and Seifert 2001).

The virB operon of the plant pathogen Agrobacterium tumefaciens was the first TFSS system to be identified (Matthysse and Stump 1976; Depicker, Van Montagu, and Schell 1978; Zupan et al. 2000) and is currently the best characterized. This system, which is located on the Ti-plasmid, mediates the transfer of the so-called transfer-DNA (T-DNA), in the form of a DNA-protein complex. The T-DNA encodes virulence genes, which after injection and integration into the plant genome, are expressed to transform the plant into a nutrient factory for the bacterium. Other members of the TFSSs serve as virulence factors in human pathogens. For example, the TFSS of H. pylori directs the transfer of the CagA protein to mammalian cells, where it causes host cytoskeleton rearrangements and other physiological changes (Segal et al. 1999). In Bordetella pertussis, the agent of whooping cough, the TFSS mediates the secretion of pertussis toxin that interferes with receptor-mediated signaling pathways (Antoine, Raze, and Locht 2000). Additionally, TFSSs have been shown to contribute to invasion and virulence in several γ-proteobacterial species of genera such as Erhlichia and Anaplasma (Ohashi et al. 2002), Bartonella (Schulein and Dehio 2002; Seubert et al. 2003; Schmid et al. 2004; Schulein et al. 2004), and Brucella (Boschirolli et al. 2002).

The diverse functions of the TFSSs and their importance in bacterial pathogenesis call for an examination of their evolutionary history. At issue are questions of whether the diversification of the TFSSs has happened once or multiple times in evolution and which of the functions represents the ancestral state. For example, the idea that TFSSs have evolved from preexisting sources, first into...
conjugation systems and later into virulence systems has been proposed (Cao and Saier 2001). Furthermore, it has been suggested that conjugation systems may be a subgroup of protein translocation systems that have evolved the ability to also translocate DNA in the form of DNA-protein complexes (Christie 2001). The argument for this hypothesis is that all TFSSs are capable of protein transport, whereas only some support the transfer of DNA (Christie 2001). According to this scenario, protein transfer represents the ancestral state whereas the transfer of DNA-protein complexes is a derived function.

To examine the role of horizontal transfer for the emergence of novel gene functions in general and to learn more about the basis for the functional diversity of the TFSSs in particular, we have here examined the evolutionary relationships of functionally divergent TFSSs. We present evidence suggesting that plasmid-encoded conjugation represents the ancestral function of the TFSSs and that the conversion of these into effector translocation systems was associated with horizontal gene transfer across the species boundaries.

Materials and Methods

We have used a total of four different data sets to compare the phylogeny of the TFSSs to the corresponding bacterial phylogeny. The divergence pattern of the TFSSs was inferred from a data set consisting of homologs to the A. tumefaciens VirB4 proteins from a large number of taxa (called the virB4 data set) and from a data set consisting of the concatenated gene sequences of the virB operon from a reduced number of taxa (called the virB operon data set). In addition, phylogenetic analyses were performed for each of the individual virB genes selected for concatenation. The bacterial phylogeny was inferred from a concatenated protein data set and a 16S rRNA data set (these are referred to as the species data sets). The assembly and methods used for the phylogenetic inference of each data set is described below. GI- or accession numbers for all genes used are available in Supplementary Materials.

Assembly of the Species Data Set

The species phylogeny was inferred for a representative subset of proteobacterial species containing complete virB operons that encode the full set of TFSS components. We selected the cyanobacterium Synechocystis as outgroup for the analysis. The 16S rRNA data set was obtained from the Ribosomal Database Project (Maidak et al. 2001). Because there was no rRNA entry for Anaplasma phagocytophilum, which contains a complete virB operon, we substituted it for Anaplasma marginale in the rRNA data set. The topology of the individual protein trees within the Rhizobiales (Bartonella spp., Brucella spp., A. tumefaciens, Sinorhizobium meliloti, and Mesorhizobium loti) differs from the 16S rRNA gene tree and from each other (data not shown). To increase the chances of obtaining true orthologs within this group, a set of 83 genes with colinear organization in all the Rhizobiales in this study was chosen. Homologs in the other taxa were inferred by Blast (E ≤ 10⁻¹⁰). Genes with representatives in more than 11 out of the 13 x-proteobacterial species were retained after this step. Finally, an additional nine genes were discarded after manual inspection of the alignments, leaving us with a data set of 17 protein-coding genes. Protein sequences were extracted to limit the effects of biased base composition on phylogenetic inference.

Assembly of the virB4 Data Set

Homologs of the A. tumefaciens VirB4 protein were obtained from two iterations of PSI-Blast (Altschul et al. 1997) with a cutoff E value of 0.001. Supplementary table 1 shows the species, accession number, gene name, and genome location of the taxa included in the analysis of the VirB4 proteins. A maximum likelihood and a Neighbor-Joining analysis were carried out to obtain the tree topology (see Phylogeny of the VirB4 Protein Homologs). The topology of the VirB4 protein tree separated the virB homologs (the subclade containing the A. tumefaciens virB operon) from the trb homologs (the subclade containing the A. tumefaciens trb operon). We selected 31 taxa in the virB clade and as an out-group one taxon in the trb clade (Azotobacter rhizogenes trb) for the analysis (see below).

Assembly of the virB Operon Data Set

To improve the phylogenetic signal we concatenated the genes in the virB operon, including in the analysis only the selected taxa of the virB subgroup. The selected taxa were first examined, using BlastP (Altschul et al. 1990) for the identification of homologs of the other 10 genes in the operon (virB1–11). The genes were clustered in groups linked by a BlastP cutoff E value below 0.0001. Species with homologs to 7 or more genes out of the 11 virB genes in A. tumefaciens and/or Bartonella henselae were selected only if the homologs formed an operon (e.g., were found at the same chromosomal location). To obtain a set that represented all major lineages of the x-proteobacteria, we also included Ehrlichia chaffeensis and A. phagocytophilum, although the virB genes in these species are found in two locations in the genome.

Some genes in the operon are less conserved among taxa (e.g., virB5 and virB7) and are in some cases annotated solely on the basis of their location in the operon. These formed several clusters. In order to avoid the alignment of nonhomologous sequences, only genes below the threshold value for sequence similarity (as inferred by BlastP) across the whole group were considered. These criteria left us with the following seven genes for analysis, in 32 taxa: virB3, virB4, virB6, virB8–virB11.

Sequence Alignments

All alignments were made using ClustalW (Thompson, Higgins, and Gibson 1994) with default parameters. The 17 proteins selected for the species phylogeny were aligned individually and concatenated into one alignment after manual correction.

Because distantly related homologs of VirB4 were retrieved using PSI-Blast and are highly divergent in sequence, we were concerned that ambiguously aligned regions would influence the result of the phylogenetic
analysis. To evaluate this, we used the program SOAP (Loytynoja and Milinkovitch 2001), which compares alignments produced with different parameters and cleans a reference alignment from blocks that are sensitive to parameter change.

Alignments were produced for each of the possible combination of parameters with the gap-opening penalty ranging from 8 to 12 (step = 1) and the gap-extension penalty ranging from 0.1 to 0.6 (step = 0.2). Positions supported by less than 90% of the alignments produced for the possible combination of parameters obtained were excluded (positions 138–146, 194–255, 291–521, 556–586, 617–631, 727–728, 740–756, 788–801, 828–964, 971–985, 998–1003, 1023–1028, 1042–1045, 1104–1109, 1111–1129, 1142–1229). This left 517 sites, i.e., 59% of the complete alignment was considered as ambiguously aligned. The tree topology produced from the cleaned alignment differed very slightly from the topology inferred from the complete alignment and only involving branches lacking bootstrap support. Although there was no resolution in the order of divergences for distantly related sequences in either tree, all clades with bootstrap support were robust to including regions of ambiguous alignment.

DNA sequences for the seven virB genes were aligned separately according to the protein sequence alignments and concatenated in the order virB3–virB11. Because this analysis only includes a subgroup of the many taxa in the virB phylogeny, sequences are less diverged and alignment ambiguity less of a problem. The virB homologs form an operon (as inferred by homology) and therefore probably function together to form protein secretion or conjugation machineries. Thus, it is safe to assume that they have evolved as a single unit and share the same evolutionary history. Moreover, comparisons of groups with high bootstrap support and posterior probability (>80% and 0.97, respectively) across the seven individual maximum likelihood and Bayesian virB gene trees failed to identify any differences in the topology (data not shown), suggesting that each gene in the operon has a similar evolutionary history. This is in line with earlier results (Cao and Saier 2001) and validates the approach.

The alignments are available in supplementary information and on our Web site, http://artedi.ebc.uu.se/molev/.

Model Selection

Adequate models of DNA evolution for the 16S rRNA gene and codon positions 1 and 2 of the virB genes were determined using a hierarchical likelihood ratio test approach. We used the same test hierarchy (and thus model selection) as implemented in the program Modeltest (Posada and Crandall 1998) at $P < 0.01$. For the 16S rRNA and the virB6 genes the optimal model was GTR + I + Γ, i.e., general time reversible with invariant sites and a gamma rate distribution. For the virB3, virB4, and virB10 genes, the model TVM + I + Γ, i.e., a model with variable transversion frequencies and equal transition frequencies with invariant sites and a gamma rate distribution was shown to be most adequate, whereas the best model for virB8 was HKY + I + Γ (Hasegawa, Kishino, and Yano 1985). Finally, the selected model for virB11 was the Jukes and Cantor model (Jukes and Cantor 1969).

Inference of the Species Phylogeny

The species phylogeny was inferred from the rRNA alignment and the concatenated protein alignment using a Bayesian analysis. In addition, a maximum likelihood analysis was performed for the rRNA alignment, and a Neighbor-Joining analysis using maximum likelihood distances on the protein alignment was performed.

Bayesian inference of phylogeny for the 16S rRNA data set was done using MrBayes 3.0 (Ronquist and Huelsenbeck 2003). Four chains were run for 1,000,000 generations, sampling a tree at every 100 generations. Convergence of parameters was reached before the first 100,000 generations for all trees. A consensus of the last 9,000 trees was calculated using PAUP* 4.0b8 (Swofford 2000). The Markov chain was initialized from a random tree. Likelihood analysis of the 16S rRNA data set was done using PAUP*. A heuristic search algorithm was used along with simple stepwise addition and tree bisecting and reconnection branch swapping.

The concatenated protein alignment was analyzed using the Neighbor-Joining method on likelihood distances computed with Tree-Puzzle using the Jones, Taylor, and Thornton (1992) (JTT) rate matrix for protein evolution and four rate categories (the likelihood of the hypothesis was higher with JTT than with the other models available in Tree-Puzzle). A Neighbor-Joining tree was computed from the likelihood distances using the program neighbor from the PHYLP package (Felsenstein 2004). Moreover, a Bayesian analysis was performed on the same data set, using the JTT model for protein evolution, gamma distributed rates, and the proportion of invariable sites estimated from the data.

Phylogeny of the VirB4 Protein Homologs

We used maximum likelihood and Neighbor-Joining methods to infer the phylogeny of the VirB4 protein homologs. PHYML (Guindon and Gascuel 2003) was used to infer a maximum likelihood tree from the VirB4 protein alignment under the JTT model of protein evolution (which had a higher likelihood with the data than the other models available in PHYML) with four substitution rate categories. The Neighbor-Joining analysis was performed as described in Inference Of The Species Phylogeny.

Phylogeny of the Concatenated virB Gene Alignments

To infer the phylogeny of the virB operons, we used Bayesian analysis on nucleotide sequences. The primary reason for this was that the adequate models for nucleotide substitution varied between the individual genes in the operon alignment, and MrBayes allows the user to specify different models for different partitions. For the Bayesian analysis, only codon positions 1 and 2 were included. The different selected models were specified for the different partitions (genes) of the data, and the gamma shape parameter was unlinked across all partitions. Two runs of four and five chains, respectively, each with
sequences (codon positions 1 and 2), with substitution models in the Bayesian analysis of the virB4 operon alignment.

Because we were concerned that the analysis might be influenced by biased nucleotide composition, we also inferred a tree from a protein alignment. We used PHYML to do the maximum likelihood analysis on protein sequences, as described for the analysis of the virB4 homologs.

To test the reliability of the inferred trees, we analyzed 100 bootstrap replicates generated with seqboot from the PHYLIP package (Felsenstein 2004) for all maximum likelihood and Neighbor-Joining analyses.

Character State Reconstructions

The distribution of the three functions (DNA uptake, conjugation, and effector translocation; Supplementary table 1) was mapped onto the VirB4 maximum likelihood tree. The root was placed between the virB-comB clade and the trb clade using MacClade (W. P. Maddison and D. R. Maddison 1989). The characters were unordered and equally weighted. The distribution of characters (main chromosome vs. auxiliary; Supplementary table 1) was mapped onto the VirB4 maximum likelihood tree in the same way.

Results

Inference of the Proteobacterial Species Tree

To obtain a species tree for an in-depth phylogenetic analysis of the TFSSs in the proteobacteria, we first inferred the phylogenetic relationship of a selected set of proteobacterial species using a concatenated alignment of 17 protein-coding genes (Supplementary table 2) as well as a 16S rRNA alignment. We included in the analysis only those species with virB operons that encode the complete set of TFSS components.

The tree topology obtained from the 16S rRNA and the concatenated alignments (fig. 1) is fully compatible with the prevailing view of proteobacterial species divergence patterns (Boussau et al. 2004). Thus, the α-proteobacterial species clustered together with strong support, as did also the βγ-proteobacterial species. The obligate intracellular bacteria Anaplasma and Ehrlichia represent the earliest diverging cluster in the α-proteobacteria, with the facultative intracellular genera Bartonella and Brucella clustering with plant-associated members of the Rhizobiales. Bordetella and Xylella are separated from the α-proteobacterial cluster, in which Legionella pneumophila is the earliest diverging lineage.

Phylogenetic inferences based on concatenated protein alignments are valid only if all individual genes share a common history. For the 17 proteins concatenated for the species phylogeny, we believe this is the most likely scenario because these show colinear organization in the Rhizobiales, consistent with vertical transmission at least in these species. Topologies of maximum likelihood phylogenies of the individual 17 genes are largely consistent (data not shown), the exception being the placement of M. loti. The position of M. loti in the α-proteobacterial tree has earlier been reported to be sensitive to methods and sampled genes, probably due to short branches separating it from its neighboring clades (Boussau et al. 2004).

Evolutionary Relationships of the TFSSs

The species phylogeny was compared with tree topologies inferred from components of the TFSSs. We first inferred the phylogeny of the TFSS based on the VirB4 protein, an ATPase that is encoded by the longest and most highly conserved gene in the operon (fig. 2). The phylogeny separates a cluster of genes annotated as trb from the remaining homologs. Some members of the trb clade encode documented parts of conjugation systems, such as the TrbE protein encoded by the transfer region of the A. tumefaciens Ti-plasmid. Another subclade, called the virB clade, contains the prototypical VirB4 protein encoded by the A. tumefaciens Ti-plasmid as well as other known effector translocation systems. The DNA-uptake system (virB4-comB) from H. pylori along with the virB4 genes in C. jejuni and Wolinella succinogenes represents a third, albeit small subgroup in the tree (the comB clade).

A closer inspection of the tree topology reveals a complex distribution of the trbE-virB4 genes, with homologs from closely related bacterial species being highly dispersed in the tree (fig. 2). For example, members of the α-proteobacteria (fig. 1) are present in at least seven different clusters in the VirB4 protein tree (fig. 2), and individual species often contain multiple genes that belong to more than one cluster. Five clusters contain exclusively genes from α-proteobacterial species (fig. 2, clusters A, D, E, F, and G), whereas in another two (fig. 2, clusters B and C) α-proteobacterial genes are interspersed with genes from nonrelated bacterial phyla or broad host-range plasmids.

Additionally, these clusters represent a mix of functions, with the prototypical virB4 gene on the pTi plasmid in A. tumefaciens clustering with homologs located on the symbiosis island of M. loti and on the pRi plasmid in A. rhizogenes (fig. 2, cluster A). A sister clade contains homologs from three species, including the avhB4 gene involved in conjugation in A. tumefaciens, the virB4 gene on the pSymA replicon in S. meliloti, and the chromosomally encoded virB4 gene in Bartonella involved in host-cell interactions (fig. 2, cluster A). Two copies of the virB4 genes are present in Rickettsia and Wolbachia, one of which evolves about four- to fivefold faster than the other (fig. 2, cluster E). The trb subclade has so far only been associated with conjugation, but the function of many systems in this cluster remains to be discovered. Taken together, the TFSS operon represents a functionally diverse set of genes that includes all of the currently identified effector translocation systems along with many conjugation systems.
Horizontal Transfer of the TFSS across Distantly Related Species

To contrast the divergence patterns of the proteins in the functionally divergent virB clade with the species phylogeny, we needed a better resolution of the deeper branching pattern in the TFSS phylogeny. To improve the signal, we concatenated the individual virB3–6 and virB8–11 gene alignments from a representative set of taxa (fig. 1) and compared it to the species tree. As expected, very little congruence was observed between the species tree and the TFSS operon tree (compare figs. 1 and 3). In contrast to the species tree, which clearly separates the different proteobacterial subdivisions, as many as two γ and three α subgroups were discerned in the TFSS operon tree. Even within subgroups, similarities in divergence patterns were rare.

The most notable inconsistency concerns the placement of Bartonella and Brucella. Although these genera
FIG. 2.—The VirB4 protein phylogeny. The topology and the branch lengths of the tree are according to maximum likelihood reconstructions. Bootstrap support values are from inferences with the maximum likelihood and Neighbor-Joining methods. Clusters A–G show clades of particular interest; these may contain only α-proteobacterial genes or a mixture of α-proteobacterial genes and genes from other bacterial phyla or broad host-range plasmids.
are sister clades in the species tree (fig. 1), the trw operon in *Bartonella* forms a clade with genes located on the broad host-range R388 plasmid from *Escherichia coli* (fig. 3, cluster B) rather than with *Brucella*, which instead clusters with genes located on two broad host-range plasmids isolated from the wheat and alfalfa rhizospheres (Schneiker et al. 2001; Tauch et al. 2002) and with *B. pertussis* (fig. 3, cluster C). Although it is not excluded that both systems were present in the common ancestor of *Bartonella* and *Brucella* (fig. 2), a more likely scenario is that two independent transfer events occurred subsequent to the divergence of *Bartonella* and *Brucella*.

Given the accessory nature of the TFSS operons, their wide distribution on plasmids and auxiliary replicons (fig. 2), and the presence of multiple copies in several species (e.g., *A. tumefaciens*, which encodes three homologs on two plasmids), we suggest that the incongruence between the species tree (fig. 1) and the TFSS operon tree (fig. 3) is best explained by multiple horizontal gene transfer events.

Divergence of Conjugation Systems into Effector Translocation Systems

To infer the ancestral state and the directionality of functional changes, parsimony analysis was used to trace the different functions of the TFSSs along the branches of the tree. Figure 4A shows the three functional groups, conjugation, effector translocation, and DNA uptake (Supplementary table 1) superimposed onto the VirB4 tree topology (fig. 1). The parsimony analysis predicts that the general direction of change has been from conjugation to secretion. Because effector translocation is polyphyletic and found at the tips of the tree, it is predicted to have arisen from conjugation fairly recently. This conclusion is reached irrespective of the taxa and nodes with unknown or equivocal function and, importantly, regardless of the rooting of the tree. Still, this result has to be treated with some caution because it may not be robust to the uncertainty of the inferred phylogeny, which contains several unsupported nodes (fig. 2). However, an identical analysis using the

---

**FIG. 3.**—The TFSS operon tree topology inferred from concatenated alignments of seven genes in the TFSS operon encoding virB3–5 and virB8–11. The topology and branch lengths are according to a maximum likelihood analysis of the seven protein sequences. Maximum likelihood bootstrap values for the phylogenetic inference of the concatenated protein alignment and posterior probabilities for a Bayesian analysis based on first and second nucleotide positions in the concatenated gene alignments are shown.
operon tree, which is well supported but has fewer taxa, lead to the same conclusions (data not shown).

There are several examples of clades in which the conjugation and effector translocation functions coexist. For example, the conjugation system encoded by the \textit{avhB} operon and the effector translocation system encoded by the \textit{virB} operon in \textit{A. tumefaciens} form a cluster together with putative effector translocation systems from other species (fig. 3, cluster A). Other examples are the clade containing the \textit{Bordetella} toxin secretion system, the putative effector translocation system in \textit{Brucella} as well as conjugation systems encoded by the plasmids pSB102 and pPO2T (fig. 3, cluster C), and the \textit{Bartonella} \textit{trw} operon along with the \textit{trw} operon from plasmid R388 (fig. 3, cluster B).

To establish the role of plasmids and auxiliary replicons for the evolution of TFSSs, we superimposed the location of genes on auxiliary replicons and chromosomes onto
the species tree topology (fig. 4B). As auxiliary, we counted all operons located in a part of the genome that is or has recently been mobile, including operons located on plasmids, megaplasmids, auxiliary chromosomes, pathogenicity islands, transposons, or integrated plasmids (Supplementary table 1). Interestingly, these represent as many as 57% of the 69 cases for which the genome location is known. This is probably an underestimate because some of the chromosomally encoded TFSSs may be the result of no longer detectable plasmid integration events. The relatively few cases of chromosomally encoded TFSSs are found in intracellular bacteria with small genomes and no plasmids, such as in *Rickettsia, Wobbachia, Bartonella*, and *Bordetella*. Taken together, the analysis suggests that the ancestor of the TFSS was a plasmid-located conjugation system, and that the direction of change has been from plasmid to chromosome.

**Discussion**

The evolutionary fate of a horizontally transferred gene resembles that of a duplicated gene copy in that the acquired gene may not necessarily be essential for the cell. In both cases, the most likely outcome is gene loss. For duplicated genes, it is predicted that mutations start accumulating and that these may occasionally generate new protein functions (Ohno 1970). Preservation of duplicated copies was originally suggested to happen through nonfunctionalization, whereby one copy acquires a new beneficial function and the other retains the original function (Ohno 1970; Ohta 1998). Later, other explanations for preservation have been put forward, such as subfunctionalization, where both gene duplicates experience degenerative mutations leading to partitioning of the original function between the copies (Force et al. 1999, Lynch and Force 2000).

Neofunctionalization should also be possible for those horizontally acquired genes initially not under strong selection in the recipient genome, and thus free to accumulate mutations and adopt new functions. Once established, selective constrains on the new function will slow down the fixation rate for new mutations, and this is to be expected irrespective of whether the gene was initially acquired by duplication or horizontal transfer. Our analysis of the TFSSs supports the hypothesis that horizontal transfer can lead to functional divergence of genes or operons. It is generally hard to infer the direction of a horizontal transfer. On the branches we believe are recipients for horizontal transfer, e.g., clusters A in figures 3 and 4, clusters B in figures 3 and 4, and clusters C in figures 3 and 4, parsimony analysis indicates a switch in function from conjugation to effector translocation. We estimated the probability that these switches (counting only unambiguous events) are not more common than expected by chance on the branches that are recipients for horizontal transfers. Given the inferred species tree, the operon tree, and the assignment of characters effector translocation-conjugation on the operon tree (data not shown), this probability was found to be low ($P \leq 0.024$, concentrated-changes test; Maddison 1990), and we thus conclude that there is a significant correlation.

Furthermore, according to our parsimony analysis, the ancestors of all chromosomally encoded TFSSs are plasmid borne, suggesting that transfers from plasmid to chromosome are more likely than vice versa (fig. 4B). For example, the *virB* operon of *Brucella*, located on the second replicon, is most closely related to two broad host-range plasmids isolated from the wheat and alfalfa rhizospheres (Schneiker et al. 2001; Tauch et al. 2002). Because *Brucella* survives in the soil between animal infections, the acquisition of a plasmid-borne conjugation system from a soil-growing organism is a likely scenario. Likewise, our analysis supports the hypothesis that one of the operons coding for TFSS in *Bartonella*, the *trw* operon, was acquired from a plasmid conjugation system (Seubert et al. 2003). In both cases, horizontal transfer overlaps with a switch in function from conjugation to involvement in host-cell interaction, although the exact mechanism by which the *trw* genes contribute to pathogenesis remains to be deciphered. It has been suggested that the proteins encoded by the *trw* genes have adapted primarily to mediate binding to host-cell structures (Seubert et al. 2003), which seems plausible considering that plasmid conjugation systems have been shown to play a role in biofilm formation (Ghigo 2001).

Several genomes encode more than one operon for TFSSs, possibly due to several independent transfer events. For example, the *B. henselae* genome encodes two TFSSs. In addition to the *trw* operon already mentioned, this genome contains a second *virB* operon that is most closely related to the *avhB* conjugation system located on plasmid pATC58 in *A. tumefaciens* and the *virB* operon on the auxiliary replicon pSymA in *S. melliloti*. The latter is essential for symbiosis but does not seem to play a role in nodulation and nitrogen fixation (Barnett et al. 2001) (fig. 3, cluster C). Thus, this particular *virB* operon may be circulating among α-proteobacterial plasmids and auxiliary replicons. In *B. henselae*, the *virB* operon is located within a region that is the putative remnant of an integrated megaplasmid (Alsmark et al. 2004) and was probably first acquired by a large plasmid or auxiliary replicon, analogous to the ones found in other members of the Rhizobiales. Despite being most closely related to a conjugation machinery (the *A. tumefaciens avh* operon), the *Bartonella virB* operon encodes an effector translocation machinery involved in pathogenesis (Schulein and Dehio 2002; Schulein et al. 2004) and might therefore represent yet another case of a horizontal transfer followed by a functional switch.

There are examples of incongruences between the TFSSs and the species phylogeny also outside the α-proteobacteria, some of which probably reflect plasmid transfers (i.e., the *tra* operon on *Haemophilus influenzae* plasmid pF3028 is most closely related to *tra* operon on the *E. coli* plasmid R721). However, because TFSSs in these species mostly remain uncharacterized, we do not know if they provide examples of functional divergence. Our PSI-Blast searches failed to identify a few systems previously described as members of the TFSS, such as the Gram-positive conjugation systems (Grohmann, Muth, and Espinosa 2003), the F-plasmid *tra* region (Lawley et al. 2003), the *N. gonorrhoeae* DNA-uptake system (Dillard and Seifert 2001), or the *L. pneumophila dot/com* system. Although it is possible that the criteria of our searches...
were too stringent, a more likely explanation is that all translocator systems which show sequence similarity to any plasmid conjugation system were called TFSSs in the previous classifications (Ding, Atmakuri, and Christie 2003). This means that some systems may be classified as TFSSs even if they have evolved from unrelated or very distantly related conjugation systems.

Bacteria have a range of mechanisms for the transport of proteins and DNA across the cell membranes, including the type I, type II, and type III secretion systems (TTSS) (Saier 2004). Phylogenetic analyses of TTSS suggest that these share an ancestor with the flagellar subunit export systems. A notable difference from the evolution of TFSS is that the flagellar transport system seems to have evolved by vertical descent and forms a monophyletic group to the exclusion of the TTSS, although it is not clear which of the two systems arose first (Nguyen et al. 2000; Gophna, Ron, and Graur 2003; Saier 2004). In contrast, our analysis strongly suggests that effector translocation is a function that has emerged multiple times during evolution. The TFSS of pathogens such as Brucella and Bartonella are the results of recent innovations made possible through the import of plasmid conjugation systems. This is remarkable, considering that both species are completely dependent on their TFSSs to establish infection. With the exception of very closely related species, such as B. henselae and Bartonella quintana, we find no evidence for vertical inheritance of effector translocation systems. Instead, most modern effector translocation systems appear to have been acquired recently from conjugation systems. This is consistent with our parsimony analysis, which suggests that the ancestor of the modern TFSSs was a plasmid-encoded conjugation system (fig. 4).

During switches from conjugation to effector translocation, there is a risk that the original function of the gene may be lost in the new host. Although this may be beneficial for the recipient cell, the overall fitness of the plasmid in the bacterial community may decrease because horizontal transfer is the survival mechanism of conjugative plasmids. Thus, for long-term plasmid maintenance it may be essential that a certain proportion of the TFSS systems circulating in the bacterial community retain the conjugation function.

As yet, we do not know whether the different functions of the virB operons correlate with defined amino acid substitutions or whether they merely result from differences in expression profiles. The Rickettsia prowazekii and Rickettsia conorii genomes contain two copies of the virB genes (RP103/RP784 and RC0141/RC1217, respectively) that are distantly related to each other and evolve by different functional constraints, as shown by differences in nonsynonymous substitution frequencies (Ka = 0.0135 for RP103–RC0141; Ka = 0.0454 for RP784–RC1217). This shows that fixation rates for nonsynonymous substitutions may differ by a factor of two or more even within individual genomes, indicative of different functions and counterselective constraints. However, it is possible that also TFSSs optimized for different functions are capable of complementing each other. Indeed, the trv operons located in the Bartonella tribocorum genome and on the broad host-range plasmid R388 in E. coli can functionally substitute for each other, as evidenced by in vitro gene replacement experiments (Seubert et al. 2003). Furthermore, studies of duplicated genes in Saccharomyces cerevisiae suggest that compensation effects might also exist for relatively divergent genes (Gu et al. 2003).

Taken together, our analysis supports the hypothesis that the acquisition of initially superfluous DNA in the recipient lineage enables novel gene functions to emerge. It also suggests that horizontal transfer has dominated the evolution of the TFSSs and that effector translocation systems used by bacterial pathogens have evolved from conjugation systems. The high prevalence, the ease of transmission, and the functional flexibility of bacterial conjugation systems provide a firm basis for the continuous emergence of novel types of secretion systems. Given the variety of different DNA and protein molecules among the identified effector molecules, it seems likely that new adaptive features will emerge and that many novel toxins and virulence factors remain to be discovered.

Supplementary Materials

Supplementary tables 1 and 2 and the alignments are available at Molecular Biology and Evolution online (www.mbe.oupjournals.org).

Acknowledgments

We thank two anonymous referees and the associate editor for suggestions and comments on the manuscript. This work was supported by the Swedish Research Council (VR), the Swedish Foundation for Strategic Research (SSF), the Knut and Alice Wallenberg Foundation (KAW) and the European Union (EU).

Literature Cited


Birot, A. M., and F. Casse-Delbart. 1988. Map location on Agrobacterium root-inducing plasmids of homologies with the


Herve Philippe, Associate Editor

Accepted February 21, 2005