Mol Genet Genomics (2008) 280:397–408
DOI 10.1007/s00438-008-0373-y

ORIGINAL PAPER

Distant horizontal gene transfer is rare for multiple families of prokaryotic insertion sequences

Andreas Wagner · Nicole de la Chaux

Received: 20 June 2008 / Accepted: 11 August 2008 / Published online: 28 August 2008

Abstract Horizontal gene transfer in prokaryotes is rampant on short and intermediate evolutionary time scales. It poses a fundamental problem to our ability to reconstruct the evolutionary tree of life. Is it also frequent over long evolutionary distances? To address this question, we analyzed the evolution of 2,091 insertion sequences from all 20 major families in 438 completely sequenced prokaryotic genomes. Specifically, we mapped insertion sequence occurrence on a 16S rDNA tree of the genomes we analyzed, and we also constructed phylogenetic trees of the insertion sequence transposase coding sequences. We found only 30 cases of likely horizontal transfer among distantly related prokaryotic clades. Most of these horizontal transfer events are ancient. Only seven events are recent. Almost all of these transfer events occur between pairs of human pathogens or commensals. If true also for other, non-mobile DNA, the rarity of distant horizontal transfer increases the odds of reliable phylogenetic inference from sequence data.

Keywords Insertion sequences · Lateral gene transfer · Molecular evolution

Communicated by D. Ussery.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-008-0373-y) contains supplementary material, which is available to authorized users.

A. Wagner (✉) · N. de la Chaux
Department of Biochemistry, University of Zurich,
Bldg. Y27, Winterthurerstrasse 190, 8057 Zurich, Switzerland
e-mail: aw@bioc.uzh.ch

A. Wagner
The Santa Fe Institute, Santa Fe, NM, USA

A. Wagner · N. de la Chaux
The Swiss Institute of Bioinformatics, Basel, Switzerland

A. Wagner
The University of New Mexico, Albuquerque, NM, USA

Introduction

In this paper, we provide evidence that successful horizontal transfer over large phylogenetic distances may be rare among prokaryotic insertion sequences, an important class of transposable elements. Transposable elements are important components of many bacterial genomes (Mahillon and Chandler 1998; Craig et al. 2002; Siguier et al. 2006a). To understand their evolutionary dynamics is important for two unrelated reasons. The first comes from the observation that transposable elements may have a net deleterious effect on their host, despite their ability to occasionally cause beneficial mutations (Doolittle and Sapienza 1980; Orgel and Crick 1980; Hartl et al. 1983; Lawrence et al. 1992; Blot 1994; Charlesworth et al. 1994; Zeyl et al. 1996; Treves et al. 1998; Capy et al. 2000; Cooper et al. 2001; Edwards and Brookfield 2003; Schneider and Lenski 2004; Wagner 2006). Their continued sustenance in prokaryotic populations and metapopulations may thus depend on horizontal gene transfer (Lawrence et al. 1992; Ochman et al. 2000; Wagner 2006), which is analogous to infection in the epidemiology of infectious diseases: A human disease agent may cause mortality of individuals, but may persist in a population through horizontal transfer from infected hosts. The incidence of such horizontal transfer determines the evolutionary fate of disease agents (Anderson and May 1991), and the same may hold for families of transposable elements. However, we know little about this incidence, especially among distantly related species, even though a rich literature exists on the evolution of transposable elements (Sawyer and Hartl 1986; Ajioka and Hartl 1989; Charlesworth and Langley 1989; Vonsternberg et al. 1992; Wilke and Adams 1992; Blot 1994; Maside et al. 2000; Bartolome et al. 2002; Vieira et al. 2002; Edwards and Brookfield 2003; Fingerman et al. 2003; Petrov et al. 2003; Witherspoon and Robertson 2003; Pasyukova et al. 2004; Vieira and Biemont 2004; Arkhipova 2005; Garfinkel 2005; Maside et al. 2005; Sanchez-Gracia et al. 2005; Wagner 2006; Touchon and Rocha 2007).

An unrelated reason to study the evolutionary dynamics of transposable elements is that it may shed light on the evolution of prokaryotes themselves. Pervasive horizontal gene transfer is the major challenge in reconstructing prokaryotic phylogenies from gene trees (Doolittle 1999; Lake et al. 1999; Snel et al. 1999; Gogarten et al. 2002; Lawrence and Ochman 2002; Brown 2003; Daubin et al. 2003; Philippe and Douady 2003; Daubin and Ochman 2004; Delsuc et al. 2005; Kurland 2005; Lerat et al. 2005; Ochman et al. 2005). The magnitude of this problem varies with the extent to which horizontal gene transfer occurs among distantly related species. If horizontal gene transfer were largely restricted to closely related species, then species trees would be ill-resolved on small time scales, but well-resolved on large time scales. Broad-scale prokaryotic phylogenies would not be in danger. If horizontal gene transfer, however, were also abundant among distantly related species, then prokaryotic phylogenetic relationships might be ill-resolved on all time scales.

While only a few studies focus on the incidence of horizontal gene transfer for transposable elements, considerable effort has been devoted to the genome-wide incidence of horizontal transfer (involving mobile and other kinds of DNA) (Lawrence et al. 1992; Nelson et al. 1999; Ochman et al. 2000; Nakamura et al. 2004; Choi and Kim 2007). Taken together, existing work suggests that horizontal gene transfer is frequent on short and intermediate evolutionary time scales (Lawrence et al. 1992; Ochman et al. 2000; Nakamura et al. 2004), but that transfer may be rarer among more distantly related species (Brugger et al. 2002; Ge et al. 2005; Choi and Kim 2007).

Most existing studies on horizontal transfer focus on genes whose products play important roles in an organism's life cycle. Because transposable elements are often not essential per se, and because they can easily migrate between different DNA molecules, such as chromosomes, plasmids, and viral genomes, they are more easily transferred than other, non-mobile DNA. Insertion sequences are among the simplest kinds of mobile DNA. If their incidence of transfer is representative of that of other kinds of mobile DNA, then a systematic survey of this incidence may provide a "worst-case-scenario" of the overall extent of horizontal transfer.

Many analyses of single insertion sequences in individual genomes exist, but large surveys of many IS families in multiple completely sequenced genomes are scarce. In a previous paper (Wagner et al. 2007), we introduced the computational tool IScan that can scan multiple genomes for insertion sequences and other transposable elements. IScan can identify not only the coding regions of these insertion sequences, but also associated DNA such as direct or indirect repeats. In this earlier work, we used a large data set produced by IScan to demonstrate that the within-genome sequence divergence of insertion sequences in a given family is generally low, which provides evidence that insertion sequences may generally not reside long in the genomes that they have infected (Wagner 2006). An unrelated analysis, based on an independently generated large-scale data set (Touchon and Rocha 2007), focused on the question what determines IS abundance in a genome. It concluded that genome size is the only significant predictor of insertion sequence abundance.

The data set generated by IScan for our earlier analysis (Wagner et al. 2007) is large and comprises 2,091 insertion sequences (ISs) from all major 20 IS families, and their abundance in more than 400 completely sequenced bacterial genomes. We here use this data in a phylogenetic analysis of insertion sequence evolution. The results show that horizontal transfer of insertion sequences among distantly related prokaryotic species is rare. Most distant transfer events are very old, underscoring their rarity.

Methods

The departure point of our analysis was a data set produced by our previously published tool IScan (Wagner et al. 2007). Briefly, this data resulted from a search for ISs (Wagner et al. 2007) that represent the 20 major IS families listed in Table 1 (Mahillon and Chandler 1998; Siguier et al. 2006a; Toleman et al. 2006). We had carried out this search in 438 curated prokaryotic genomes (consisting of 790 sequenced DNA molecules) available from GenBank (ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/). The curated query ISs are listed in column 2 of Table 1, and had been obtained from the IS repository IS Finder (http://www-is.biotoul.fr; Siguier et al. 2006b). We retained BLAST hits to IS ORFs with an E value of ≤ 1 and at least 35% amino acid identity to the query sequence. Our approach identified a total of 2,091 insertion sequences (Table 1). Establishing the completeness of IScan's results is difficult, because no gold standard of a set of genomes with a bona fide set of sequences that constitute insertion sequences is known, and because most recently sequenced genomes are automatically annotated, and can thus not be used as a reference. However, a comparison of IScan's results with the annota-

Table 1 Reference IS and numbers of insertion sequences for each IS family studied here

Family	Reference IS	Number of ISs	Distinct bacterial clades
IS1	**IS1A**	**863**	**8**
IS481	**IS481**	**259**	**3**
IS3	**IS2**	**242**	**2**
IS5	**IS5**	**239**	**11**
IS4	**IS4**	**171**	**3**
IS110	**IS110**	**88**	**4**
IS982	**IS982**	**57**	**1**
IS630	**IS630**	**55**	**4**
IS256	**IS256**	**29**	**2**
IS21	**IS21**	**25**	**1**
IS91	**IS91**	**19**	**1**
Tn3	**IS1071**	**18**	**2**
IS30	**IS30**	**13**	**1**
ISL3	**ISL3**	**7**	**1**
IS66	ISRm14	3	ND
ISCR	ISCR1	2	ND
IS6	IS15	1	ND
ISAs1	ISAs1	0	ND
IS1380	IS380A	0	ND
IS605	IS605	0	ND
Total		2,091	44

The data in the first three columns of this table are reproduced from Table 1 in Wagner et al. (2007), and are shown here only for clarity. Column 1 shows the IS families we studied, and column 2 shows the particular IS within a family that was used as a query sequence for our tool IScan. The query sequences can be found at the ISfinder database (http://www-is.biotoul.fr; Siguier et al. 2006b). IS families shown in black have sufficient members for a meaningful phylogenetic analysis

tion of the perhaps best-annotated genome, that of *E. coli*, suggests that IScan's result match known genomic IS content well. For instance, the curated genome sequence of *E. coli* K-12 (file NC_000913.gbk, available from ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/) contains 7, 4, and 1 non-truncated copies of the insertion sequences IS1, IS30, and IS4, respectively. IScan detects all these copies, in the right position, and it detects no additional copies. We note that the original genome sequencing paper (Blattner et al. 1997) for *E. coli* K-12 reported only 3, 3, and 0 copies for the insertion sequences IS1, IS30, and IS4. This suggests that even the presumably high-quality manual annotation of the earliest sequenced genomes is subject to error, suggesting that it will be difficult to establish an annotation gold standard for transposable elements.

A few bacterial species have multiple chromosomes, not all of which contain 16S rDNA. Because one of our main goals was to study the distribution of ISs on the bacterial 16S rDNA tree, we excluded ISs on molecules that did not contain 16S rDNA from further analysis.

For generation of the 16S rDNA phylogenetic tree, prokaryotic 16S rDNA sequences were extracted from genbank files (ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/) of bacterial genomes and aligned with the GreenGenes NAST alignment program at http://greengenes.lbl.gov. A prokaryotic 16S rDNA maximum-likelihood tree was constructed using the package phyml (Guindon and Gascuel 2003b), with the Hasegawa–Kishino–Yano (Hasegawa et al. 1985) substitution model, where the transition–transversion ratio and the proportion of variable sites were estimated from the data. To accommodate variable substitution rates among sites, we allowed for four different substitution rates and estimated the parameter of the gamma distribution determining the rate variation from the data. A tree generated by neighbor joining (Higgs and Attwood 2005) was used as the starting tree to be refined by the maximum likelihood algorithm. The major features of the resulting tree are concordant with other recently published trees using different approaches, such as that by (Ciccarelli et al. 2006).

For those ISs where more than three copies existed in the hundreds of genomes we studied, we also generated phylogenetic trees of individual IS families, both within a given genome, as well as for all family members, regardless of genome provenance. Because some of the IS families that we studied had more than one open reading frame (ORF), we first merged these ORFs for reasons of computational tractability, as described below. In each set of insertion sequences for which a phylogenetic tree was to be constructed, we then identified subsets of ISs whose coding region was identical within a genome, and used only one representative of each such subset for further analysis. We then aligned the coding sequence of the ISs using clustalw (Thompson et al. 1994), and constructed a maximum-likelihood phylogenetic tree from the resulting alignment using phyml with the same parameters as listed above.

For some of our analyses, it was necessary to estimate synonymous divergence K_s among IS coding regions. We prefer to use K_s rather than raw DNA sequence divergence, because synonymous changes are under weak selection, accumulate rapidly, and are thus more sensitive to detect recent horizontal transfer. We note that for low divergence (e.g., $K_s < 0.2$), K_s estimates sequence divergence well. For example, a value of $K_s = 0.1$ implies that two sequences differ approximately at 10% of their synonymous sites. To estimate K_s, we first merged ORFs for ISs whose coding region contains more than two ORFs. Specifically, we calculated the number of nucleotides that overlap in the two ORFs, and eliminated from a sequence containing both ORFs the segment containing the overlap, and any additional nucleotides upstream or downstream of the overlapping segment required to retain the reading frames of the two ORFs. On average, IS ORFs were shortened by only four nucleotides through this procedure. We then used our previously published tool GenomeHistory (Conant and Wagner 2002) to estimate K_s.

Results

Only few distant transfers are required to explain the global distribution of ISs

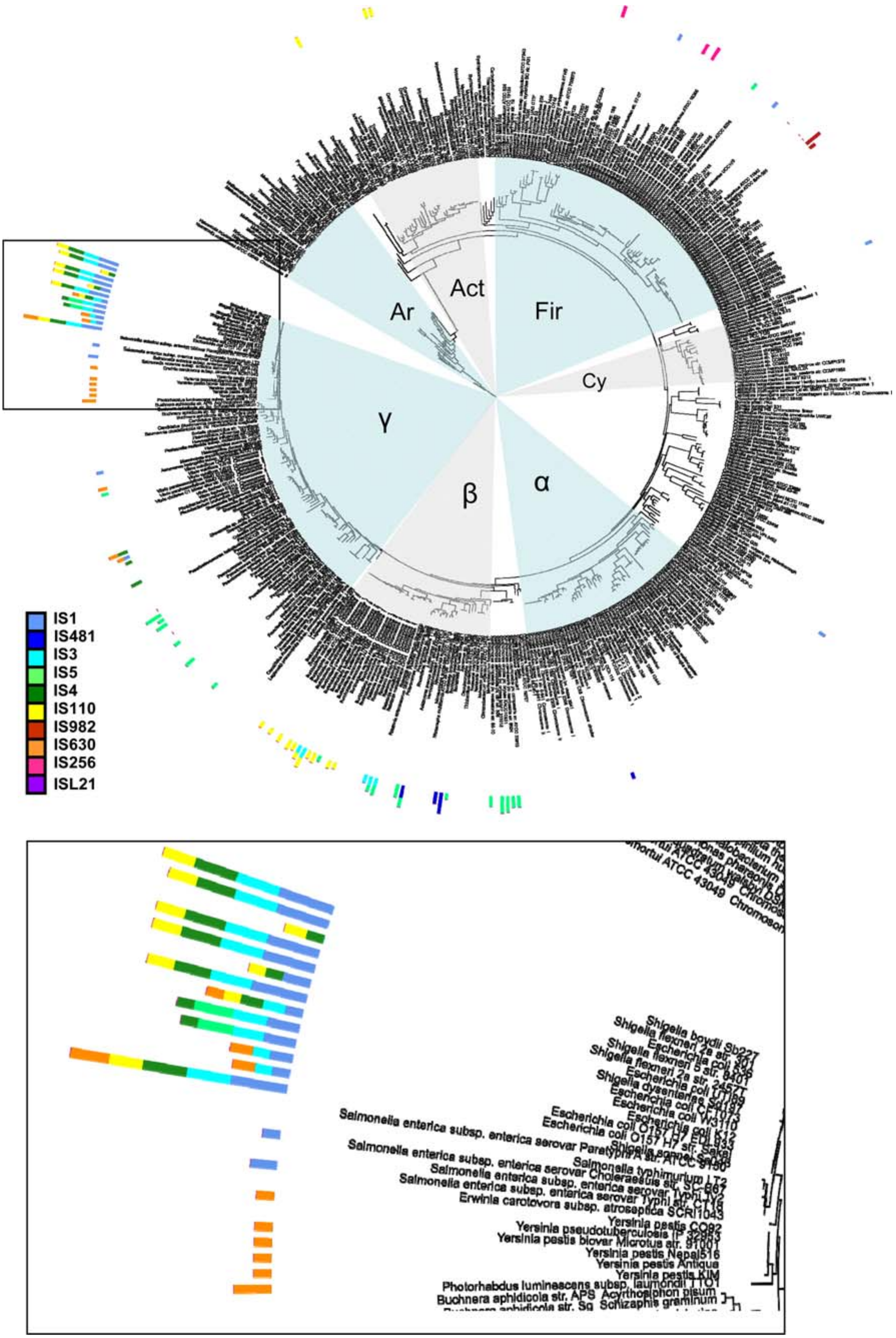
To study the phylogenetic distribution of insertion sequences, we first identified 438 curated, completely sequenced prokaryotic genomes, and constructed a maximum-likelihood phylogenetic tree of all 16S-rDNA-containing molecules in this data set ("Methods").

This tree serves as a scaffold to place IS relationships. The use of 16S rDNA strikes a compromise between computational feasibility and phylogenetic accuracy: The tree's major features are in good agreement with phylogenies based on more sophisticated multi-locus sequence analysis (Gevers et al. 2005), such as that by (Ciccarelli et al. 2006), which are computationally very costly. We then identified members of 20 different insertion sequence families (Table 1) in the curated genomes, and mapped them onto this phylogenetic tree. Figure 1 indicates the structure of the 16S tree, as well as the distribution and abundance (length of bars) for the ten most abundant IS families in the data set (Table 1). Close examination shows that each IS family has a patchy and sporadic distribution on the tree, with modest concentrations of ISs found in only a small number of species, such as the extremely closely related *Escherichia coli*/*Shigella* clade (box in Fig. 1). Available completely sequenced genomes are not an unbiased sample from the prokaryotic world, because many sequencing projects have focused on human-associated species. This bias in the data may partly account for the concentrations of ISs in closely related species. Together, the ten most abundant families shown in Fig. 1 encompass almost 97% of the 2,091 IS copies we identified. Six of the 20 families that we had examined had fewer than three representatives (Table 1). No meaningful phylogenetic analysis is possible for such small numbers of ISs, and we thus did not study these families further. We analyzed each of the remaining 14 families separately, and also constructed maximum-likelihood phylogenetic trees (Guindon and Gascuel 2003a) for family members within a given genome.

Past horizontal gene transfer can reveal itself through several possible signatures (de la Cruz and Davies 2000; Koonin et al. 2001; Ragan 2001), including phylogenetic signatures and DNA composition signatures. None of these is without limitations. For our analysis, phylogenetic signatures are better suited, because over time, the DNA composition of horizontally transferred genes approaches that of the host genome, which limits the time horizon for the detection of transfer. There are two major phylogenetic signatures (Figure S1 in Electronic supplementary material). The first involves incongruences between gene trees (e.g., between 16S rDNA and IS trees). Unfortunately, biased gene deletions from either tree, rapid expansion of gene families, or taxonomic sampling artefacts can cause erroneous results with this signature. Preferable in our case is the second phylogenetic signature: a patchy distribution of genes in disjoint clades of a large phylogeny (Figure S1b). In such a distribution, only a small fraction of genomes contain a gene or IS of interest. These genomes occur in small clades (patches) on a tree that are separated by deep branches, and by many taxa that do not contain the IS. In principle, such a patchy distribution could also be explained by independent loss of an IS from all taxa that do not contain it. However, with a phylogenetic tree of more than 400 taxa spanning vast phylogenetic distances, and relatively few taxa containing ISs, this explanation is exceedingly unlikely. For example, the most abundant IS we study (IS1) occurs in fewer than 5% (20/438) of completely sequenced bacterial genomes. In addition, no known IS has a broad phylogenetic distribution that would be required as ancestral under the independent-loss scenario. We thus attribute IS occurrence in clearly disjoint clades to horizontal gene transfer.

Figure 2 shows examples of such anomalous distributions for three different ISs. For IS110 (Fig. 2a), there are four IS-containing clades, *Escherichia coli*/*Shigella* (40 copies), *Burkholderia* spp. (24), *Corynebacterium* spp (5), and *Streptomyces coelicolor* (5), requiring three horizontal transfer events between these clades. IS5 (Fig. 2b) occurs in eleven disjoint bacterial clades or species (*Escherichia coli*, *Vibrio vulnificus*, *Pseudomonas syringae*, *Marinobacter aquaeolei*, *Methylococcus capsulatus*, *Burkholderia cepacia*, *Ralstonia solanacearum*, *Acidovorax spp.*, *Azoarcus sp. EbN1*, *Xanthomonas spp.*, *Staphylococcus aureus*), requiring ten horizontal transfer events. We observe IS1, by far the most prolific IS element (Table 1), in eight disjoint clades (Fig. 2c), which are (counterclockwise beginning at 9 o'clock) the *Shigella* spp/*Escherichia coli* clade (>700

Fig. 1 The *upper panel* shows a maximum-likelihood phylogenetic tree of 16S rDNA in more than 400 completely sequenced prokaryotic genomes, where the following major clades are indicated: Archaea (*Ar*), actinobacteria (*Act*), cyanobacteria (*Cy*), firmicutes (*Fir*), α-, β-, and γ-proteobacteria (α-, β-, and γ-, respectively). Lengths of *colored bars* are proportional to IS numbers within a genome in the ten most abundant IS families, as indicated by the *color legend*. Note the patchy distribution of individual families. The *rectangular box* and the *lower (boxed) panel* highlight the *Escherichia coli*/*Shigella* clade, which contain the greatest numbers of ISs. Trees were displayed with ITOL (Letunic and Bork 2007) ▶


Act
Ar
Fir
Cy
γ
β
α
IS1
IS481
IS3
IS5
IS4
IS110
IS982
IS630
IS256
ISL21
Shigella boydii Sb227
Shigella flexneri 2a str. 301
Escherichia coli 536
Shigella flexneri 5 str. 8401
Escherichia coli UTI89
Shigella flexneri 2a str. 2457T
Escherichia coli CFT073
Shigella dysenteriae Sd197
Escherichia coli W3110
Escherichia coli K12
Escherichia coli O157 H7 EDL933
Escherichia coli O157 H7 str. Sakai
Shigella sonnei Ss046
Salmonella enterica subsp. enterica serovar Paratyphi A str. ATCC 9150
Salmonella typhimurium LT2
Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67
Salmonella enterica subsp. enterica serovar Typhi str. CT18
Salmonella enterica subsp. enterica serovar Typhi Ty2
Erwinia carotovora subsp. atroseptica SCRI1043
Yersinia pestis CO92
Yersinia pseudotuberculosis IP 32953
Yersinia pestis biovar Microtus str. 91001
Yersinia pestis Nepal516
Yersinia pestis Antiqua
Yersinia pestis KIM
Photorhabdus luminescens subsp. laumondii TTO1
Buchnera aphidicola str. APS Acyrthosiphon pisum
Buchnera aphidicola str. Sg Schizaphis graminum
ATCC 43049 Chromosome

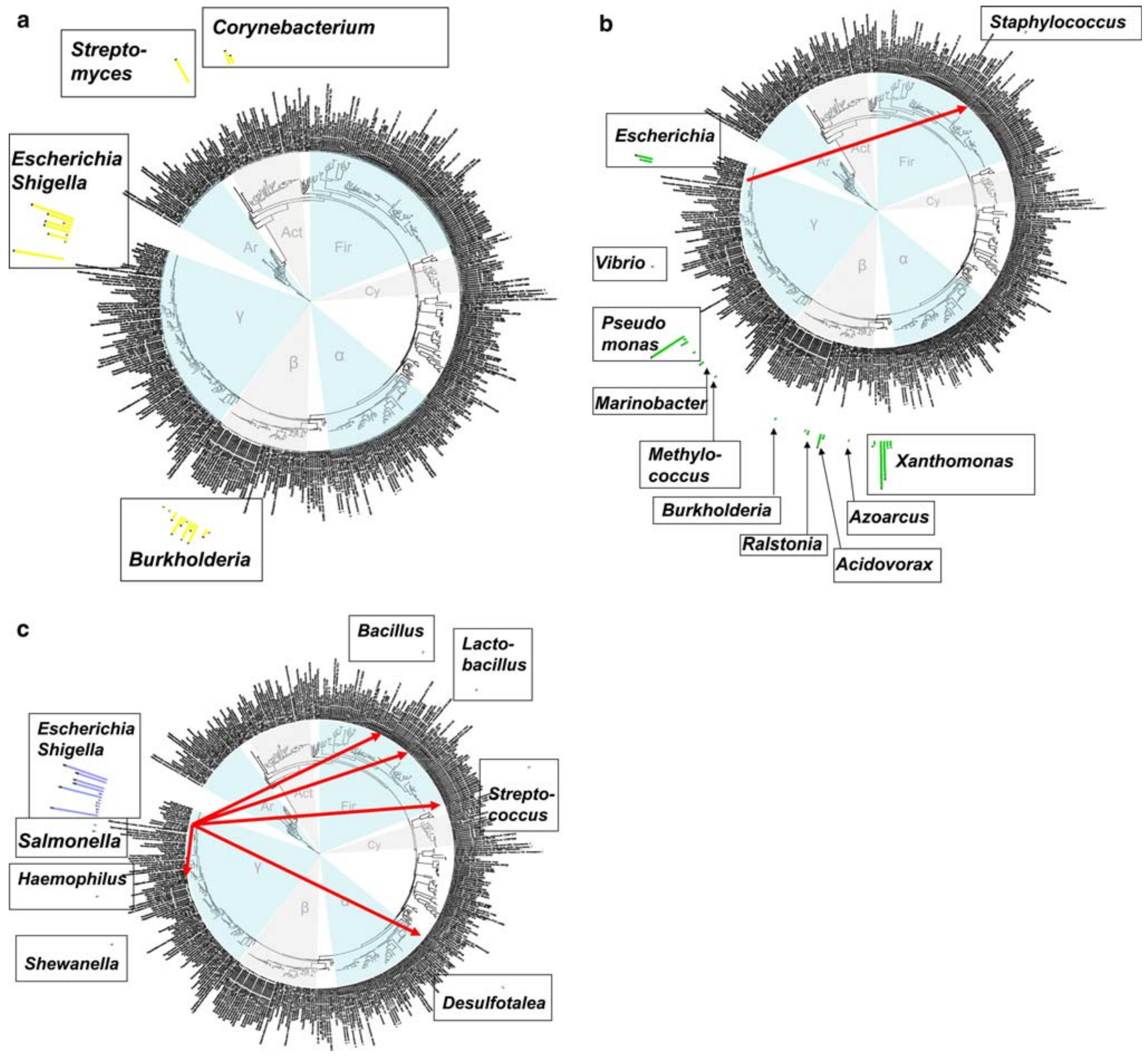


Fig. 2 Incidence of **a** IS110, **b** IS5, and **c** IS1 on the maximum-likelihood phylogenetic tree of prokaryotic 16S rDNA from Fig. 1. *Boxes* and inscribed names indicate genera in which ISs occur. Where space permitted, the *bars* indicating IS numbers were included in the *box*, and indicated by *arrows* otherwise. *Red arrows* indicate likely directions of recent horizontal gene transfer. Tree layout, *symbols* for major clades, and *color coding* of ISs are as in Fig. 1. Trees are displayed with ITOL (Letunic and Bork 2007)

copies), *Salmonella enterica* (4 copies), *Haemophilus ducreyi* (1), *Shewanella sp. W3-18-1* (1), *Desulfotalea psychrophila* (1), *Streptococcus pyogenes* (1), *Lactobacillus sakei* (1), and *Bacillus cereus* (1). A minimum of seven horizontal transfer events would be required to explain this phylogenetic distribution. In general, the majority of clades among which distant transfers occur are associated with humans. For example, among the eight IS1 clades, six are associated with humans either as pathogens or commensals (Albritton 1989; Kotiranta et al. 2000; Ryan and Ray 2004; Chaillou et al. 2005). The remaining two (*D. psychrophila* and *Shewanella sp. W3-18-1*) are psychrophilic (cold-loving) marine bacteria. To preserve space, we do not show 16S trees for the remaining IS elements, but we list (Table 1) the numbers of distinct clades containing these elements. Among the 14 IS families with sufficient copy numbers for a phylogenetic analysis, 9 families occurred in more than one clade. To explain the phylogenetic distribution of ISs among these clades, merely 30 horizontal transfer events would be necessary.

Most of the few distant transfers are old

We next asked whether any of these likely horizontal transfer events might have occurred recently. To this end, we compared the sequence similarity of the 16S rDNA sequences considered here with the divergence of insertion sequences among genomes. Figure S2a shows the distribution of pairwise nucleotide divergence among the 16S rDNA molecules considered here. Figure S2b shows the distribution of synonymous divergence K_s, the fraction of synonymous substitutions at synonymous sites (Li 1997), for all pairs of ISs of the same family that occur in different genomes. A signature of a recent horizontal transfer would involve distantly related species (high 16S rDNA divergence) with closely related ISs (low synonymous divergence K_s). Such pairings are very rare. The typical pattern of association observed for all IS families that we have studied is exemplified by IS110 in Fig. 3a. The figure shows that IS elements in highly diverged bacterial species are also highly diverged. This means that no horizontal transfers of IS110 involving distantly related species occured recently. A similar pattern holds for 11 of the 14 families of ISs that we have studied.

The three exceptions are IS256 (phylogeny not shown), IS5 (Figs. 2b, 3b), and IS1 (Figs. 2c, 3c). The case of IS256 is simple: only two distantly related clades (*Enterococcus faecalis* V583 and *Staphylococcus epidermidis*; 16S divergence ≈0.1) harbor copies of IS256. All these copies are

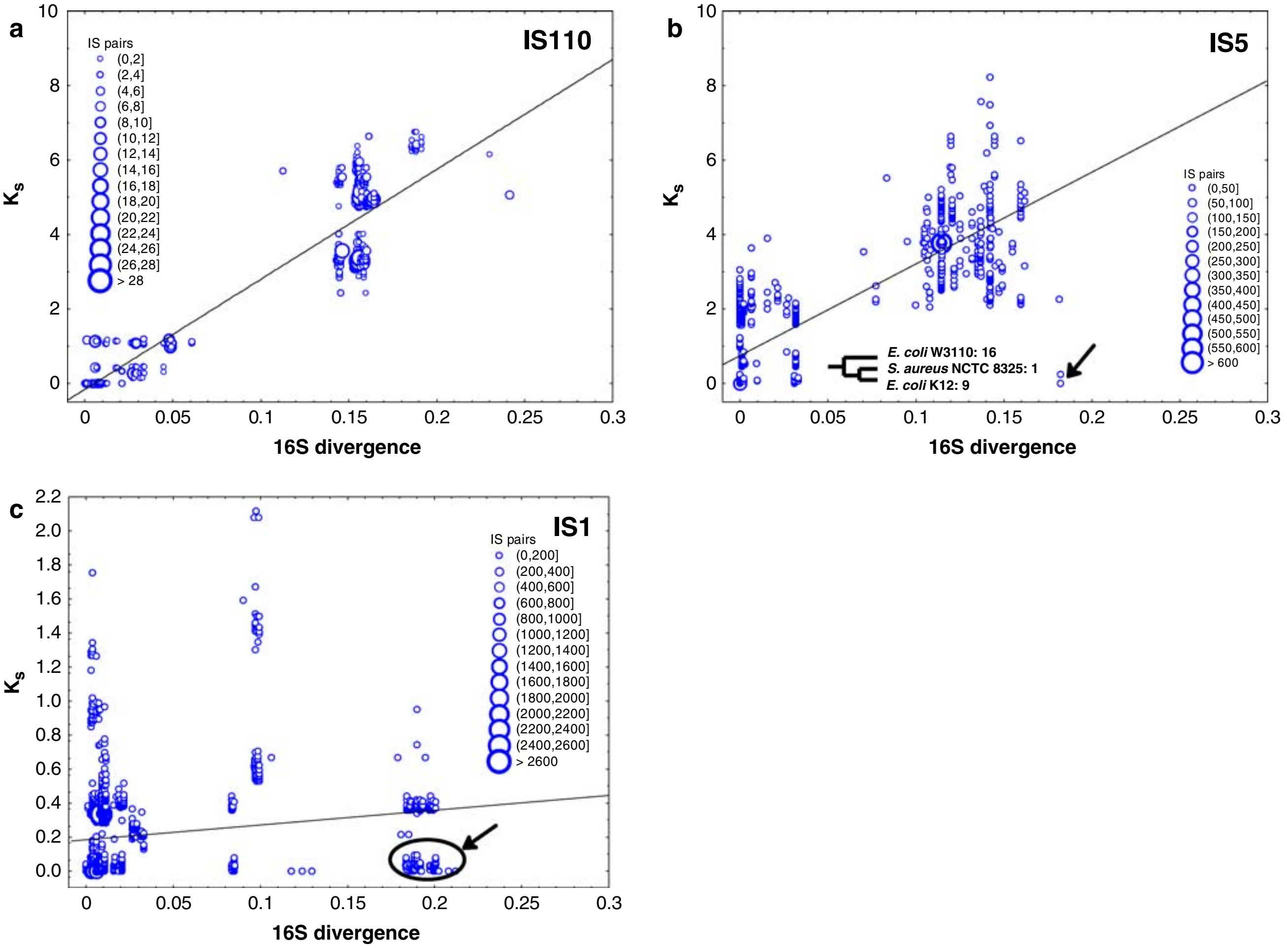


Fig. 3 Association between 16S rDNA divergence (*horizontal axis*) and synonymous divergence (K_s) of IS pairs among genomes (*vertical axis*) for IS families **a** IS110, **b** IS5, and **c** IS1. The size of the *circles* correspond to the number of IS pairs, as indicated in the *inset*, that occur in the genomes of the 16S divergence shown on the *horizontal axis*, and that has the divergence shown on the *vertical axis*. In **b**, two data points indicating very similar IS5 elements in highly diverged genomes are indicated by an *arrow*. A phylogenetic tree based on IS5 coding region divergence is shown for the three genomes in which these IS elements occur. See text for details. In **c**, the most highly diverged group of taxa containing very similar IS1 elements is *circled*. *Solid lines* are linear regression lines. Note the different range on the vertical axis in **c**, due to the lower overall divergence of IS1 elements. Highly divergent ISs in different genomes may reflect either the long time that has elapsed between the most recent common ancestors of the two genomes, or it may be caused by multiple ancient and highly diverged copies of the IS in one genome, some of which may also show high divergence to ISs in the other genome

identical to one another, suggesting a recent transfer. In IS5 there are two species-IS pairs with very divergent 16S rDNA, yet highly similar IS5 sequences (indicated by an arrow in Fig. 3b). These involve two strains of *E. coli* (W3110 with 18 IS copies, and K12 with 11 IS copies) on one hand, and *Staphylococcus aureus* (1 copy) on the other hand. Only three numerically different values of K_s are observed between the IS5 elements in these *E. coli*–*S. aureus* species pairs, $K_s = 0$, $K_s = 0.007$, and $K_s = 0.23$. The value $K_s = 0$ indicates a very recent transfer. Can we infer the direction of this transfer? If the *E. coli* copies were derived from a very recent transfer ($K_s = 0$) from *S. aureus* to *E. coli*, and if they thus had also expanded recently, then we would expect all the IS5 copies in *E. coli* to have the same divergence ($K_s = 0$) to the single IS copy in *S. aureus*. However, there are IS copies with greater values of $K_s = 0.007$ and $K_s = 0.23$ in *E. coli*, which cannot be explained by this scenario. In contrast, a transfer from one of the *E. coli* species to *S. aureus* is consistent with the data. The phylogenetic tree shown in Fig. 3b shows the phylogenetic relationship between the *E. coli* IS5 copies and the single *S. aureus* copy. We can infer that this copy is derived from one of the 16 identical IS5 copies in *E. coli*. The red arrow in Fig. 2b reflects the direction of this transfer.

The only other examples of recent horizontal transfers among distantly related species are observed for IS1 (Touchon and Rocha 2007; Wagner et al. 2007). Here, we see a large cluster of species with highly similar ISs and divergent 16S rDNAs (circled in Fig. 3c). Note that the large number of data points in this cluster does not necessarily imply multiple distant transfer events. It is caused by large numbers of highly similar ISs in one clade of closely related species. Specifically, all of the species pairs involve a member of the *E. coli/Shigella* group (high IS copy numbers) on one hand, and the following species (low IS copy numbers) on the other hand (counterclockwise in Fig. 2c from the *E. coli/Shigella* clade): *Haemophilus ducreyi* (1 copy), *Desulfotalea psychrophila* (1), *Streptococcus pyogenes* (1), *Lactobacillus sakei* (1), *Bacillus cereus* (1). The single IS1 element of each of these species has at least one identical ($K_s = 0$) counterpart in the *E.coli/Shigella* clade (and many other IS1 elements with greater divergence). The five species are highly diverged (16S divergence >0.18) from the *E.coli/Shigella* clade, and contain only a single IS1 element. Because the *E.coli/Shigella* clade contains multiple IS1 pairs with $K_s > 0$, we can infer, with the same reasoning as above for IS5, that the transfer occurred from the *E.coli/Shigella* clade to these other species, and not vice versa. This pattern is plausible if one considers the large number (>700) IS1 copies in this clade. We can, however, not completely exclude IS1 transfers between those species that have only one IS1 element. In sum, we observe only *seven* recent horizontal gene transfer events (one for each of IS256 and IS5, as well as five for IS1) among distantly related prokaryotic species. All but one (*Desulfotalea psychrophila*) of these involve transfers between species with well-known human associations.

Analysis of IS trees

Thus far, we have focused on the analysis of 16S rDNA-based trees of prokaryotic species to identify the likely cases of horizontal gene transfer. A second possibility is to analyze the phylogenetic relationships of IS elements themselves. Two reasons, however, render the interpretation of such trees difficult. The first is the often large number of ISs within any given genome (see also below). The second is the fact that most transfer events are ancient, meaning that the clade from which they originated may not be unambiguously identifiable. However, a few IS trees are informative. The information they provide is consistent with analyses of the 16S bacterial trees. A case in point is the IS5 tree. Here, the recent transfer of IS5 from the *E.coli/Shigella* clade to *Staphylococcus aureus* discussed earlier is clearly identifiable (Fig. 4a; red arrow), and two other likely ancient transfer events can also be identified. They include a possible transfer event between *Methylococcus capsulatus* and *Azoarcus sp. EbN1*, as well as another event involving *Acidovorax sp*. JS42 and the *Xanthomonas* clade (Fig. 4a; blue arrows). Another example, involving very clear separation of ancient clades, involves IS110. As discussed earlier (Fig. 2a), IS110 occurs only in four well separated very distantly related clades and shows no evidence of recent transfer. The IS110 tree itself (Fig. 4b) shows that ISs in the clades *Burkholderia* and *E.coli/Shigella* group together, suggesting that the more recent of two ancient transfer events occurred between these clades.

Within clades of closely related taxa, the 16S rDNA based phylogenetic approach fails for our data set. Part of the reason is that 16S rDNA evolves slowly, and thus does not resolve the phylogenies of closely related taxa well; another part is the especially frequent horizontal transfer among closely related genomes. Figure S3 illustrates two examples of these problems, manifested in the low-bootstrap support of phylogenetic trees for IS1 in the *E.coli/Shigella* clade and for IS110 in a clade of closely related *Burkholderia* species. This limitation means that our analysis cannot answer some important questions about the evolutionary dynamics of transposable elements. Examples include whether IS copy numbers generally increase in a clade over the time, or whether individual ISs get frequently lost from genomes.

A final aspect of horizontal gene transfer regards the question whether ISs in a family are often transferred *into* a genome more than once. Although recent evidence (Tou-

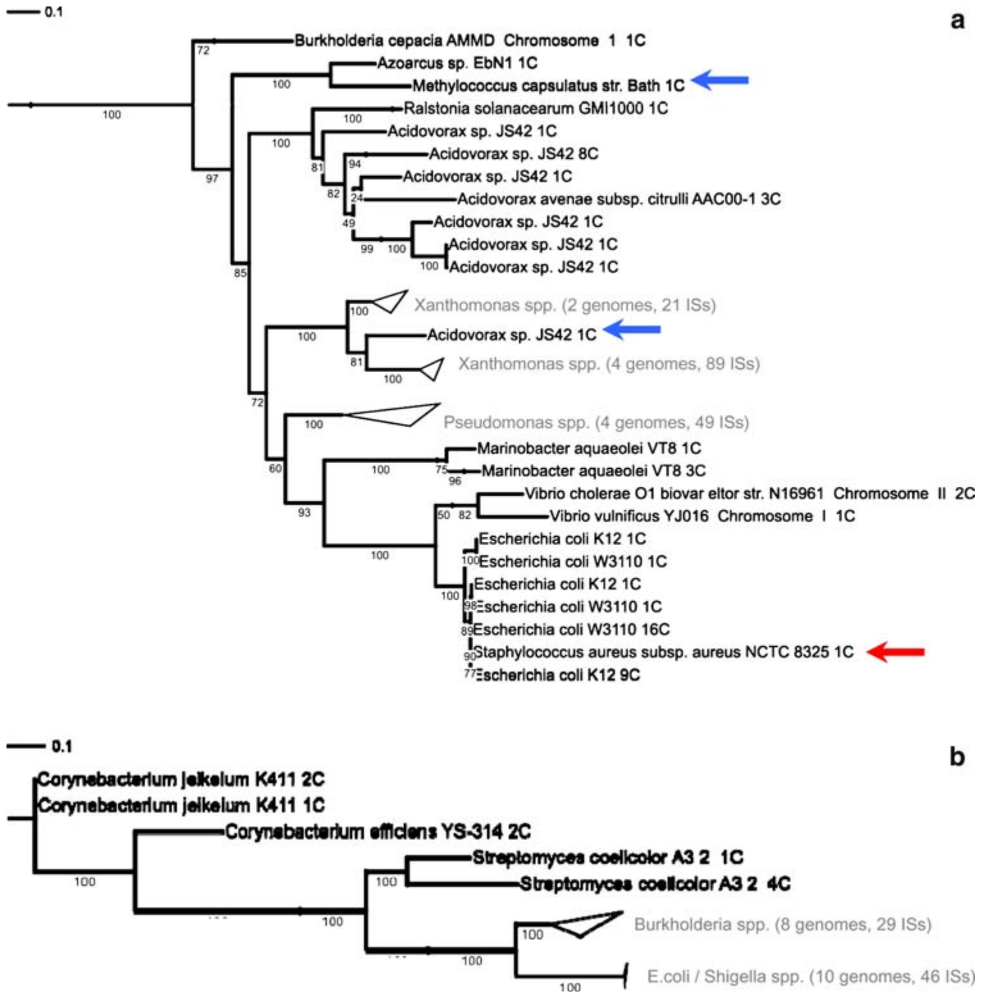

0.1
a
Burkholderia cepacia AMMD Chromosome 1 1C
Azoarcus sp. EbN1 1C
Methylococcus capsulatus str. Bath 1C
Ralstonia solanacearum GMI1000 1C
Acidovorax sp. JS42 1C
Acidovorax sp. JS42 8C
Acidovorax sp. JS42 1C
Acidovorax avenae subsp. citrulli AAC00-1 3C
Acidovorax sp. JS42 1C
Acidovorax sp. JS42 1C
Acidovorax sp. JS42 1C
Xanthomonas spp. (2 genomes, 21 ISs)
Acidovorax sp. JS42 1C
Xanthomonas spp. (4 genomes, 89 ISs)
Pseudomonas spp. (4 genomes, 49 ISs)
Marinobacter aquaeolei VT8 1C
Marinobacter aquaeolei VT8 3C
Vibrio cholerae O1 biovar eltor str. N16961 Chromosome II 2C
Vibrio vulnificus YJ016 Chromosome I 1C
Escherichia coli K12 1C
Escherichia coli W3110 1C
Escherichia coli K12 1C
Escherichia coli W3110 1C
Escherichia coli W3110 16C
Staphylococcus aureus subsp. aureus NCTC 8325 1C
Escherichia coli K12 9C
100
72
100
97
100
100
81
94
82
24
49
99
100
100
85
100
100
81
100
72
100
60
100
75
96
93
50
82
100
100
100
98
89
90
77
0.1
b
Corynebacterium jeikeium K411 2C
Corynebacterium jeikeium K411 1C
Corynebacterium efficiens YS-314 2C
Streptomyces coelicolor A3 2 1C
Streptomyces coelicolor A3 2 4C
Burkholderia spp. (8 genomes, 29 ISs)
E.coli / Shigella spp. (10 genomes, 46 ISs)
100
100
100
100
100
100

Fig. 4 A maximum likelihood phylogenetic tree of **a** IS5 and **b** IS110 elements in all prokaryotic species. *Red* (*blue*) *arrows* indicate likely recent (ancient) horizontal gene transfer events. Each genome may contain multiple IS elements, but IS elements that are identical to each other within a genome are phylogenetically not informative. Each leaf on the tree corresponds to one IS element within the named genome, or to a group of identical ISs within the genome. The species name associated with each leaf is followed by the suffix *i*C, where *i* is the number of identical IS copies this leaf corresponds to. (For example, 1C means that there is only one IS of a given sequence within the named genome.) Several clades of numerous very closely related ISs from different genomes were collapsed, as indicated by the *triangles* and the genus names in *gray*. *Numbers* on branches correspond to boostrap values based on 100 generated bootstrap samples

chon and Rocha 2007) suggests that such frequent transfer is not likely to account for differences in IS numbers among genomes, we should not exclude this possibility a priori, especially because the phenomenon of transposition immunity is not widespread among insertion sequences, with the possible exception of Tn3 and IS21 (Mahillon and Chandler 1998). Figure S4a shows, as an example, a hypothetical IS tree consistent with two independent transfer events.

In analyzing the data, we need to distinguish three possible scenarios. The first of them involves recent transfers among closely related prokaryotic species. We here face an often poorly-resolved phylogeny of ISs within genomes. A case in point is the phylogeny of IS982, which occurs in only two species. The phylogeny (Figure S4b) shows very low-boostrap support along many branches. Fundamentally, the reason for this problem is that ISs within a genome are usually highly similar to one another, indicating their recent acquisition by the genome (Mahillon and Chandler 1998). Figure S4c illustrates that this pattern holds more generally. The figure shows the distribution of $K_{s,max}$, the maximal within-genome synonymous divergence K_s of IS copies (pooled for all families). $K_{s,max}$ is the synonymous divergence of the most highly diverged IS pair within a given host genome and IS family. If synonymous divergence of ISs accumulates at a clock-like rate, then this maximal K_s can be used as an estimate of the time of most recent common ancestry of ISs within a genome. The median $K_{s,\ max}$ of 0.0087 is very low for most genomes. This indicates that most of the ISs entered their host genome too recently to resolve multiple transfer events with molecular evolution data. It also suggests that it may be difficult to resolve recent horizontal transfer events among closely related prokaryotes, even though such transfer events may be abundant.

The second scenario involves recent transfers among distantly related clades. As discussed above, there are only seven such events and they involve either ISs that are all identical (IS256), or transfer events into species that have only one IS. Such transfer events are thus useless to identify multiple transfers into a genome. The third scenario involves ancient IS transfers among distantly related clades. Here, as discussed above, it is usually not only difficult to trace individual transfer events, but the direction of transfer events is unclear. Thus, even the great abundance of existing sequence data is insufficient to provide convincing examples of multiple IS transfer events into a genome.

Discussion

In sum, in our survey of 2,091 insertion sequences, with representatives from all 20 major families in 438 completely sequenced prokaryotic genomes, we found only 30 cases of likely horizontal transfer among distantly related prokaryotic clades. The vast majority (23 of 30) of these horizontal transfer events are ancient. Only seven events are recent. Almost all of these transfer events occur between pairs of human pathogens or commensals. This bias towards human-associated species is at least partly explicable by a bias in the data set of available completely sequenced genomes: Genome sequencing projects preferably focus on human-associated species, because of their medical relevance. Our small numbers of distant horizontal transfer events may even be overestimates, because ISs from different families may sometimes be transferred at the same time on the same vector, thus further reducing the actual number of transfer events.

Previous studies that focus on genomic DNA in general, and not just on transposable elements, indicate that horizontal gene transfer is frequent on short and intermediate evolutionary time scales (Lawrence et al. 1992; Ochman et al. 2000; Nakamura et al. 2004). However, transfer, especially recent transfer, may be rarer among more distantly related species (Brugger et al. 2002; Ge et al. 2005; Choi and Kim 2007). With possible exceptions (Nelson et al. 1999), our observations are thus consistent with previous work.

A variety of barriers for distant horizontal transfer of genes are known (Thomas and Nielsen 2005; Sorek et al. 2007). Among them is high-gene expression. It is not a priori a likely candidate for the sequences we study, because ISs are generally lowly expressed and tightly regulated (Nagy and Chandler 2004). However, this tight regulation may depend on host factors. It is tempting to speculate that distant transfer increases the likelihood of host death by uncontrolled expression and proliferation of transposase genes. Other possible barriers include incompatible restriction–modification systems, or conjugative plasmids with limited host range (Thomas and Nielsen 2005).

Even the enormous amounts of available sequence data do not allow us to answer several questions about the evolutionary dynamics of ISs. These include how often horizontal transfer occurs between closely related species, whether the number of IS copies in a clade shows a net decrease or increase over time, whether ISs often get lost from genomes, and whether genomes usually get "infected" by an IS multiple times.

The reasons are threefold: first, phylogenetic trees of closely related prokaryotes are often ill-resolved. This problem might be remedied for some clades by more sophisticated multi-locus approaches (Godoy et al. 2003), but only at a computational cost too large for large-scale surveys like this one. Second, IS phylogenies within given species are often poorly resolved. Fundamentally, the reason is that many ISs within a genome are highly similar (Lawrence et al. 1992; Wagner 2006). If the median synonymous divergence K_s for the two most diverged ISs within a genome is less than 0.01 (Figure S4c), then each IS will contain very few phylogenetically informative sites. Third, and relatedly, different closely related genomes often contain identical ISs (e.g., Fig. 3c). Their origin in a genome through vertical or horizontal transfer is thus often unclear. Some of these problems (for example ambiguous IS phylogenies) may not be solvable through a simple accumulation of more data, but may represent fundamental limitations of molecular evolution approaches. Other problems would disappear if we were able to analyze horizontal gene transfer among many distantly related genomes. However, because distant transfer is so rare, we are not able to do that.

We now turn to some limitations of our analysis. First, computational constraints prevent us from analyzing truncated and very short sequences, or sequences with very low (and often dubious) sequence similarity to the reference insertion sequence we used. Because most truncated ISs would be inactive, and because passive proliferation of inactive ISs through active copies is probably less prevalent than for eukaryotic DNA transposons (Mahillon and Chandler 1998), such elements may be less likely to transpose between DNA molecules. If so, then their propensity to become successfully transferred horizontally may be lower, and their distant transfer even rarer, but our data do not allow us to determine by how much. Second, because some IS families consist of multiple extremely diverse sub-families (Mahillon and Chandler 1998), and because we use only one query sequence per family, our approach does not yield an exhaustive enumeration of ISs in the genomes we analyzed. Rather, it represents a statistical survey, sufficient for our purpose, which ensures that the major families are represented. Third, it would be highly instructive to study the phylogenetic relationships of ISs on plasmids. Some of the completely sequenced genomes have associated plasmids. However, only 5% of the ISs we identified occurred on plasmids, and because these are distributed over multiple families, their numbers in our data set are too small for a meaningful phylogenetic analysis. However, combined with data from dedicated plasmid sequencing efforts, such an analysis may become possible. We leave it to a future contribution. Finally, we have no knowledge of the environmental conditions from which the prokaryotes whose genomes we analyzed have been sampled. These conditions may affect transposition rates. For example, long-term stab cultures of *E.coli* show increased transposition rates (Naas et al. 1994).

The pervasiveness of horizontal gene transfer has led some researchers to question our ability to resolve the broad phylogeny of prokaryotes, with much ensuing debate (Doolittle 1999; Lake et al. 1999; Snel et al. 1999; Gogarten et al. 2002; Brown 2003; Philippe and Douady 2003; Delsuc et al. 2005; Kurland 2005). In this regard, the rarity of distant horizontal transfer we observe is reassuring, especially since it comes from a highly mobile class of sequences. The ISs we study can be much more easily transferred than many other, non-mobile genetic elements, because they can autonomously change location from chromosomes to transferable plasmids, and vice versa. Embedded in composite transposons mediating antibiotic resistance, or in pathogenicity islands allowing conversion from a free-living to a pathogenic lifestyle, natural selection can further facilitate their spreading. The rarity of distant transfer for the many IS families and many genomes we study suggests that distant transfer among most other genes might be even rarer. Observations like these give reason to hope that the broad evolutionary history of prokaryotes can be reliably inferred from sequence data.

Acknowledgments Support through SNF grant 315200-116814 is gratefully acknowledged. We thank C. Lewis for technical assistance.

References

Ajioka J, Hartl D (1989) Population dynamics of transposable elements. In: Berg D, Howe M (eds) Mobile DNA. American Society for Microbiology Press, Washington, DC, pp 185–210

Albritton W (1989) Biology of *Haemophilus ducreyi*. Microbiol Mol Biol Rev 53:377–389

Anderson R, May R (1991) Infectious diseases of humans. Dynamics and control. Oxford University Press, Oxford, UK

Arkhipova IR (2005) Mobile genetic elements and sexual reproduction. Cytogenet Genome Res 110:372–382

Bartolome C, Maside X, Charlesworth B (2002) On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. Mol Biol Evol 19:926–937

Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Colladovides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of *Escherichia coli* K-12. Science 277:1453–1474

Blot M (1994) Transposable elements and adaptation of host bacteria. Genetica 93:5–12

Brown JR (2003) Ancient horizontal gene transfer. Nat Rev Genet 4:121–132

Brugger K, Redder P, She QX, Confalonieri F, Zivanovic Y, Garrett RA (2002) Mobile elements in archaeal genomes. FEMS Microbiol Lett 206:131–141

Capy P, Gasperi G, Biemont C, Bazin C (2000) Stress and transposable elements: co-evolution or useful parasites? Heredity 85:101–106

Chaillou S, Champomier-Verges MC, Cornet M, Crutz-Le Coq AM, Dudez AM, Martin V, Beaufils S, Darbon-Rongere E, Bossy R, Loux V, Zagorec M (2005) The complete genome sequence of the meat-borne lactic acid bacterium *Lactobacillus sakei* 23K. Nat Biotechnol 23:1527–1533

Charlesworth B, Langley CH (1989) The population genetics of *Drosophila* transposable elements. Annu Rev Genet 23:251–287

Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371:215–220

Choi IG, Kim SH (2007) Global extent of horizontal gene transfer. Proc Natl Acad Sci USA 104:4489–4494

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

Conant GC, Wagner A (2002) GenomeHistory: a software tool and its applications to fully sequenced genomes. Nucleic Acids Res 30:1–10

Cooper VS, Schneider M, Blot M, Lenski RE (2001) Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli*. J Bacteriol 183:2834–2841

Craig N, Craigie R, Gellert M, Lambowitz AL (eds) (2002) Mobile DNA II. ASM Press, Washington, DC

Daubin V, Ochman H (2004) Quartet mapping and the extent of lateral transfer in bacterial genomes. Mol Biol Evol 21:86–89

Daubin V, Moran N, Ochman H (2003) Phylogenetics and the cohesion of bacterial genomes. Science 301:829–832

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

Delsuc F, Brinkmann H, Philippe H (2005) Phylogenomics and the reconstruction of the tree of life. Nat Rev Genet 6:361–375

Doolittle WF (1999) Phylogenetic classification and the universal tree. Science 284:2124–2128

Doolittle WF, Sapienza C (1980) Selfish genes, the phenotype paradigm, and genome evolution. Nature 284:601–607

Edwards RJ, Brookfield JFY (2003) Transiently beneficial insertions could maintain mobile DNA sequences in variable environments. Mol Biol Evol 20:30–37

Fingerman EG, Dombrowski PG, Francis CA, Sniegowski PD (2003) Distribution and sequence analysis of a novel Ty3-like element in natural *Saccharomyces paradoxus* isolates. Yeast 20:761–770

Garfinkel DJ (2005) Genome evolution mediated by Ty elements in Saccharomyces. Cytogenet Genome Res 110:63–69

Ge F, Wang LS, Kim J (2005) The cobweb of life revealed by genome-scale estimates of horizontal gene transfer. PLoS Biol 3:1709–1718

Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, Stackebrandt E, Van de Peer Y, Vandamme P, Thompson FL, Swings J (2005) Re-evaluating prokaryotic species. Nat Rev Microbiol 3:733–739

Godoy D, Randle G, Simpson AJ, Aanensen DM, Pitt TL, Kinoshita R, Spratt BG (2003) Multilocus sequence typing and evolutionary relationships among the causative agents of melioidosis and glanders, *Burkholderia pseudomallei* and *Burkholdefia mallei*. J Clin Microbiol 41:4913 (vol 41, pg 2068, 2003)

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

Guindon S, Gascuel O (2003a) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704

Guindon S, Gascuel O (2003b) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704

Hartl DL, Dykhuizen DE, Miller RD, Green J, de Framond J (1983) Transposable element IS50 improves growth rate of *E. coli* cells without transposition. Cell 35:503–510

Hasegawa M, Kishino H, Yano TA (1985) Dating of the human ape splitting by a molecular clock of mitochondria. J Mol Evol 22:160–174

Higgs P, Attwood T (2005) Bioinformatics and molecular evolution. Blackwell, Oxford, UK

Koonin EV, Makarova KS, Aravind L (2001) Horizontal gene transfer in prokaryotes: Quantification and classification. Annu Rev Microbiol 55:709–742

Kotiranta A, Lounatmaa K, Haapasalo M (2000) Epidemiology and pathogenesis of *Bacillus cereus* infections. Microbes Infect 2:189–198

Kurland CG (2005) What tangled web: barriers to rampant horizontal gene transfer. Bioessays 27:741–747

Lake JA, Jain R, Rivera MC (1999) Genomics—mix and match in the tree of life. Science 283:2027–2028

Lawrence JG, Ochman H (2002) Reconciling the many faces of lateral gene transfer. Trends Microbiol 10:1–4

Lawrence JG, Ochman H, Hartl DL (1992) The evolution of insertion sequences within enteric bacteria. Genetics 131:9–20

Lerat E, Daubin V, Ochman H, Moran NA (2005) Evolutionary origins of genomic repertoires in bacteria. PLoS Biol 3:e130

Letunic I, Bork P (2007) Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23:127–128

Li W-H (1997) Molecular evolution. Sinauer, MA, USA

Mahillon J, Chandler M (1998) Insertion sequences. Microbiol Mol Biol Rev 62:725–774

Maside X, Assimacopoulos S, Charlesworth B (2000) Rates of movement of transposable elements on the second chromosome of *Drosophila melanogaster*. Genet Res 75:275–284

Maside X, Assimacopoulos S, Charlesworth B (2005) Fixation of transposable elements in the *Drosophila melanogaster* genome. Genet Res 85:195–203

Naas T, Blot M, Fitch WM, Arber W (1994) Insertion sequence-related genetic variation in resting *Escherichia coli* K-12. Genetics 136:721–730

Nagy Z, Chandler M (2004) Regulation of transposition in bacteria. Res Microbiol 155:387–398

Nakamura Y, Itoh T, Matsuda H, Gojobori T (2004) Biased biological functions of horizontally transferred genes in prokaryotic genomes. Nat Genet 36:760–766

Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson LD, Nelson WC, Ketchum KA, McDonald L, Utterback TR, Malek JA, Linher KD, Garrett MM, Stewart AM, Cotton MD, Pratt MS, Phillips CA, Richardson D, Heidelberg J, Sutton GG, Fleischmann RD, Eisen JA, White O, Salzberg SL, Smith HO, Venter JC, Fraser CM (1999) Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. Nature 399:323–329

Ochman H, Lawrence J, Groisman E (2000) Lateral gene transfer and the nature of bacterial innovation. Nature 405:299–304

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

Orgel LE, Crick FHC (1980) Selfish DNA: the ultimate parasite. Nature 284:604–607

Pasyukova EG, Nuzhdin SV, Morozova TV, Mackay TFC (2004) Accumulation of transposable elements in the genome of *Drosophila melanogaster* is associated with a decrease in fitness. J Hered 95:284–290

Petrov DA, Aminetzach YT, Davis JC, Bensasson D, Hirsh AE (2003) Size matters: non-LTR retrotransposable elements and ectopic recombination in *Drosophila*. Mol Biol Evol 20:880–892

Philippe H, Douady CJ (2003) Horizontal gene transfer and phylogenetics. Curr Opin Microbiol 6:498–505

Ragan MA (2001) Detection of lateral gene transfer among microbial genomes. Curr Opin Genet Dev 11:620–626

Ryan K, Ray C (eds) (2004) Sherris medical microbiology. McGraw Hill, New York

Sanchez-Gracia A, Maside X, Charlesworth B (2005) High rate of horizontal transfer of transposable elements in *Drosophila*. Trends Genet 21:200–203

Sawyer S, Hartl DL (1986) Distribution of transposable elements in prokaryotes. Theor Popul Biol 30:1–16

Schneider D, Lenski RE (2004) Dynamics of insertion sequences elements during experimental evolution of bacteria. Res Microbiol 155:319–327

Siguier P, Filee J, Chandler M (2006a) Insertion sequences in prokaryotic genomes. Curr Opin Microbiol 9:526–531

Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M (2006b) ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res (Database issue) 34:D34–D36

Snel B, Bork P, Huynen MA (1999) Genome phylogeny based on gene content. Nat Genet 21:108–110

Sorek R, Zhu YX, Creevey C, Francino M, Bork P, Rubin E (2007) Genome-wide experimental determination of barriers to horizontal gene transfer. Science 318(5855):1449–1452

Thomas CM, Nielsen KM (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat Rev Microbiol 3:711–721

Thompson JD, Higgins DG, Gibson TJ (1994) Clustal-W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting; position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Toleman MA, Bennett PM, Walsh TR (2006) ISCR elements: novel gene-capturing systems of the 21st century? Microbiol Mol Biol Rev 70(2):296–316

Touchon M, Rocha EPC (2007) Causes of insertion sequences abundance in prokaryotic genomes. Mol Biol Evol 24:969–981

Treves DS, Manning S, Adams J (1998) Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli*. Mol Biol Evol 15:789–797

Vieira C, Biemont C (2004) Transposable element dynamics in two sibling species: *Drosophila melanogaster* and *Drosophila simulans*. Genetica 120:115–123

Vieira C, Nardon C, Arpin C, Lepetit D, Biemont C (2002) Evolution of genome size in *Drosophila*. Is the invader's genome being invaded by transposable elements? Mol Biol Evol 19:1154–1161

Vonsternberg RM, Novick GE, Gao GP, Herrera RJ (1992) Genome canalization: the coevolution of transposable and interspersed repetitive elements with single copy DNA. Genetica 86:215–246

Wagner A (2006) Periodic extinctions of transposable elements in bacterial lineages: evidence from intragenomic variation in multiple genomes. Mol Biol Evol 23:723–733

Wagner A, Lewis C, Bichsel M (2007) A survey of bacterial insertion sequences using IScan. Nucleic Acids Res 35:5284–5293

Wilke CM, Adams J (1992) Fitness effects of Ty transposition in *Saccharomyces cerevisiae*. Genetics 131:31–42

Witherspoon DJ, Robertson HM (2003) Neutral evolution of ten types of mariner transposons in the genomes of *Caenorhabditis elegans* and *Caenorhabditis briggsae*. J Mol Evol 56:751–769

Zeyl C, Bell G, Green DM (1996) Sex and the spread of retrotransposon Ty3 in experimental populations of *Saccharomyces cerevisiae*. Genetics 143:1567–1577