Short- and long-term evolutionary dynamics of bacterial insertion sequences: insights from Wolbachia endosymbionts

Nicolas Cerveau, Sébastien Leclercq, Didier Bouchon, Richard Cordaux

Short- and Long-term Evolutionary Dynamics of Bacterial Insertion Sequences: Insights from *Wolbachia* Endosymbionts

Nicolas Cerveau, Sébastien Leclercq, Elodie Leroy, Didier Bouchon, and Richard Cordaux*

UMR CNRS 6556, Ecologie, Evolution, Symbiose, Université de Poitiers, France

*Corresponding author: E-mail: richard.cordaux@univ-poitiers.fr.

Accepted: 13 September 2011

Abstract

Transposable elements (TE) are one of the major driving forces of genome evolution, raising the question of the long-term dynamics underlying their evolutionary success. Long-term TE evolution can readily be reconstructed in eukaryotes, thanks to many degraded copies constituting genomic fossil records of past TE proliferations. By contrast, bacterial genomes usually experience high sequence turnover and short TE retention times, thereby obscuring ancient TE evolutionary patterns. We found that *Wolbachia* bacterial genomes contain 52–171 insertion sequence (IS) TEs. IS account for 11% of *Wolbachia w*Ri, which is one of the highest IS genomic coverage reported in prokaryotes to date. We show that many IS groups are currently expanding in various *Wolbachia* genomes and that IS horizontal transfers are frequent among strains, which can explain the apparent synchronicity of these IS proliferations. Remarkably, >70% of *Wolbachia* IS are nonfunctional. They constitute an unusual bacterial IS genomic fossil record providing direct empirical evidence for a long-term IS evolutionary dynamics following successive periods of intense transpositional activity. Our results show that comprehensive IS annotations have the potential to provide new insights into prokaryote TE evolution and, more generally, prokaryote genome evolution. Indeed, the identification of an important IS genomic fossil record in *Wolbachia* demonstrates that IS elements are not always of recent origin, contrary to the conventional view of TE evolution in prokaryote genomes. Our results also raise the question whether the abundance of IS fossils is specific to *Wolbachia* or it may be a general, albeit overlooked, feature of prokaryote genomes.

Key words: insertion sequence, transposable element, evolutionary dynamics, prokaryote, *Wolbachia*, molecular palaeontology.

Introduction

Transposable elements (TE) are discrete pieces of DNA that can move within (and sometimes, between) genomes. They are widely distributed in eukaryotes and prokaryotes, and they sometimes represent substantial fractions of genomes. For example, TEs encompass about half of the human genome (Lander et al. 2001) and nearly 85% of the maize genome (Schnable et al. 2009). Because of their mobility and accumulation, TEs are major drivers of genome evolution, with effects ranging from generating insertion mutations and genomic instability to altering gene expression and contributing to genetic innovation (Feschotte and Pritham 2007; Cordaux and Batzer 2009; Cerveau et al. 2011). Given their tremendous genomic impact, abundance, and widespread taxonomic distribution, the question arises as to what long-term dynamics have made TEs so prolific and evolutionary successful during the evolution of life.

In eukaryotes, long-term TE dynamics can readily be investigated because genomes often carry highly mutated and degraded TE relics constituting a genomic fossil record of past TE proliferations and evolution at various time depths (Lander et al. 2001; Kapitonov and Jurka 2003). For example, analyses of the human genomic fossil record have revealed that DNA transposons became extinct ~40 million years ago in the primate lineage, after having experienced intense activity during the mammalian radiation and early primate evolution, 60–150 million years ago (Lander et al. 2001; Pace and Feschotte 2007). By contrast, Alu and L1 retrotransposons have proliferated throughout primate evolution, although their activity has declined within the past ~20 million years (Lander et al. 2001; Xing et al. 2004; Khan et al. 2006). The relevance of genomic fossil records is also well illustrated with the emerging field of paleovirology, consisting in the study of ancient extinct viruses unearthed

from genome sequences, which are witnesses of ancient viral infections, and the effects these agents have had on their host evolution (Emerman and Malik 2010; Gilbert and Feschotte 2010).

In sharp contrast with eukaryotes, the gene repertoires of prokaryotes change quickly by lateral (or horizontal) gene transfer and gene deletion (Rocha 2008). This high turnover is well illustrated by pseudogenes and TEs in which retention times appear to be particularly short (Wagner 2006; Touchon and Rocha 2007; Wagner et al. 2007; Kuo and Ochman 2010; Cerveau et al. 2011). As a consequence, recent TE insertions are overrepresented in bacterial genomes, and our ability to infer ancient evolutionary patterns vanishes with the erosion of the past TE fossil record. For example, insertion sequences (IS), which are simple transposase-encoding TEs frequently found in prokaryotic genomes (Chandler and Mahillon 2002; Siguier, Filee, et al. 2006) are generally considered to be of recent origin. This is reflected in the very low nucleotide divergence generally observed between IS sequences within genomes. This result has been reported in early IS studies, as exemplified by three IS families of *Escherichia coli* in which copies are >99.7% similar to their family consensus sequences (Lawrence et al. 1992). Broader-scale studies on several hundreds of bacterial genomes and up to 20 IS families confirmed this trend as more than two thirds of transposase genes are identical within genomes (Wagner 2006; Wagner et al. 2007). IS recent origin is further supported by their usually patchy distribution among closely related strains (Sawyer et al. 1987; Parkhill et al. 2003; Yang et al. 2005; Cordaux et al. 2008; Qiu et al. 2010).

The analysis of recently integrated IS elements in bacterial genomes has suggested that IS may undergo extinction–reinfection cycles on the long term (Wagner 2006). Under this scenario, periodic IS reintroductions in genomes mediated by lateral transfers are crucial for their long-term survival (Wagner 2006; Bichsel et al. 2010). However, IS copy number is not directly correlated to the rate of lateral gene transfer in bacteria, suggesting that horizontal transfer may not be a major determinant of IS abundance in genomes (Touchon and Rocha 2007). Overall, our current understanding of long-term TE dynamics in prokaryotes lags far behind that of eukaryotic TE evolutionary dynamics, in part because no IS genomic fossil record has been reported and analyzed in prokaryotes.

In this study, we report an analysis of IS elements in the genomes of *Wolbachia* bacterial endosymbionts. These ancient obligate intracellular microorganisms have been associated with arthropod and nematode hosts for >100 million years, and they are considered one of the most abundant endosymbionts on Earth (Werren et al. 1995; Cordaux, Michel-Salzat, et al. 2004; Bouchon et al. 2008; Saridaki and Bourtzis 2010; Cordaux et al. 2011). Despite their reduced sizes, *Wolbachia* genomes show an unusually high proportion of repetitive and mobile DNA, including IS elements (Moran and Plague 2004; Wu et al. 2004; Bordenstein and Reznikoff 2005; Foster et al. 2005; Klasson et al. 2008, 2009; Cordaux 2009; Leclercq et al. 2011). Strikingly, we found that the vast majority of *Wolbachia* IS copies are more or less severely degraded as a result of the accumulation of nucleotide substitutions and deletions across time. Thus, they constitute an uncommon genomic fossil record for these bacterial TEs. This rich genomic archive gave us an opportunity to directly investigate the long-term dynamics of IS elements (and provide the first empirical test of Wagner's hypothesis that IS elements experience extinction–reinfection cycles on the long term) and the microevolutionary processes governing IS expansions in bacterial genomes.

Materials and Methods

Identification and Classification of IS Elements

The complete genome sequences and annotations of *Wolbachia* strains *w*Mel (Wu et al. 2004), *w*Bm (Foster et al. 2005), *w*Pel (Klasson et al. 2008), and *w*Ri (Klasson et al. 2009) were consulted and downloaded from the National Center for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html).

We used three different strategies to identify IS elements. We first queried the original genome annotations available in GenBank with the keywords "transposase" and "transposon." Next, we performed similarity searches against the IS reference database ISFinder (Siguier, Perochon, et al. 2006) using ISsaga (Varani et al. 2011). Finally, we performed a de novo repeat detection for each genome with Repeatscout software, using l-mers size of 15 bp (Price et al. 2005). These approaches were complementary because each method alone has its own advantages and drawbacks. For example, IS identification with Repeatscout requires at least three copies in the genome, but it does not require any a priori knowledge of IS sequences. By contrast, ISsaga can detect single copy IS elements, but it requires a library of IS sequences for querying genomes.

All originally annotated IS elements were recovered by ISsaga. Repeatscout results were used as queries for BlastN searches against GenBank to identify non-IS repeats. All repeats with significant identity to known non-IS repeats (e.g., phages, group II introns, duplicated genes, etc.) were discarded. The remaining repeats were subjected to TBlastX searches against ISFinder to identify known IS elements. For the few Repeatscout repeats with no homology to known transposases remaining at this stage, we manually aligned copies and searched for IS hallmarks, such as terminal inverted repeats and target site duplications (Chandler and Mahillon 2002). None of these repeats exhibited hallmarks of IS elements, and they were therefore discarded. In sum, all IS elements identified with Repeatscout were also recovered by ISsaga.

To refine our IS annotations (i.e., to identify fragments and highly divergent copies that may have been missed

before), we generated a library of *Wolbachia* IS sequences based on the IS elements detected as described above. Next, BlastN searches against the four *Wolbachia* genomes were performed using as queries the aforementioned *Wolbachia* IS library. BlastN searches were performed with default parameters without low-complexity region filter, using a minimal subject size of 40 bp, minimal similarity of 75%, maximal *e* value of 0.05, reward of 2, and penalty of 3. To eliminate potentially redundant or overlapping IS matches, the positions of each IS copy were picked up and compared with all others for each genome. This procedure allowed us to identify 18 cases of IS copies split in two parts by nested IS insertions; each of the disrupted IS copies was counted as a single insertion event (3 in *w*Mel, 6 in *w*Ri, and 9 in *w*Pel). Overall, this analysis yielded a total of 511 candidate IS copies from the four *Wolbachia* genomes. We discarded 13 candidates as false positives (e.g., *DnaA* mistakenly assigned to IS21 and *XerC/D* recombinases mistakenly assigned to IS91 in all four genomes). Thus, the final data set consisted of 498 validated IS copies.

Each IS copy was assigned to an IS family by TBlastX searches against ISFinder (Siguier, Perochon, et al. 2006). The sequences of IS copies assigned to the same IS family were aligned using ClustalW as implemented in the software Bioedit ver 7.0 (Hall 1999), followed by manual adjustments. Due to high sequence divergence, some IS sequences could not be aligned to each other within some IS families. Therefore, we defined groups within IS families as IS nucleotide sequences that can reliably be aligned to each other within groups but cannot be aligned with sequences from other groups. As a quality control, BlastN searches were performed using all IS sequences as queries against all four *Wolbachia* genomes. For all queries, the returned matches exclusively comprised copies from the query IS group, thereby confirming the validity of the defined IS groups. *Wolbachia* IS group names were assigned based on ISFinder best IS group matches in TBlastX searches. *Wolbachia* IS groups with no functional representative (which thus could not be deposited in ISFinder) were named as follows: IS family name, followed by "-w" (for *Wolbachia*) and a specific upper-case letter. For example, IS110-wA represents a *Wolbachia* IS group from the IS110 family, which has no known representative in ISFinder.

Structure and Nucleotide Divergence of IS Elements

To investigate IS structure, for each IS group, IS sequences along with 500 bp of 5′ and 3′ flanking genomic sequences were aligned. These alignments were used to identify IS boundaries and to determine terminal inverted repeats and direct repeats generated upon insertion, whenever present. Transposase genes were identified through open reading frame detection using the NCBI online tool ORFFinder (http://www.ncbi.nlm.nih.gov/) and the FSFinder software (Moon et al. 2004). For each IS group with at least two alignable copies in a given genome, we calculated pairwise nucleotide divergence between copies with the MEGA ver 4.0 software based on observed nucleotide substitutions (Kumar et al. 2008).

Orthology Analyses

To identify IS copies inserted at orthologous genomic sites between genomes, we performed BlastN searches (minimal subject size of 40 bp, minimal similarity of 75%, maximal *e* value of 0.05, reward of 2, and penalty of 3) using as queries 300 bp of upstream and downstream genomic regions flanking each IS copy. For each IS locus, orthologous flanking regions from queried genomes were aligned and compared to identify orthologous IS insertions. However, for many IS loci, BlastN searches yielded no, partial, or multiple matches in queried genomes, thereby preventing reliable identification of orthologous IS insertion sites. Therefore, the final data set used for orthology analyses exclusively consisted of IS loci with unambiguous matches for both flanking sequences in queried genomes.

Simulations of IS Evolutionary Dynamics

Scenarios of IS dynamics are based on the simulation of a set of evolving IS sequences in a haploid genome. Simulations were designed to allow qualitative comparisons between scenarios and thus only roughly represent the biological complexity of IS evolution. Four major processes of IS evolution were considered: acquisition by horizontal transfer, copy number expansion from a resident copy, degradation through random substitutions, and loss through deletion. IS elements were represented by sequences of 300 numbers ranging in value from 0 to 63. Each number represented a hypothetical codon, with three codons representing stop codons (as in the bacterial genetic code). Each initial IS sequence did not carry any stop codon and was considered as functional.

All simulations counted a fixed number of generations. At each generation, IS sequences were allowed to mutate or be deleted at fixed constant rates. We arbitrarily chose 10,000 generations and a mutation rate of 15×10^{-6} mutation per codon to produce an average pairwise divergence of 30% between the oldest copies at the end of the simulations and keep our simulations manageable in terms of computational time. Mutations changed the value of the mutated codon to another random value. If the new value corresponded to a stop codon, the IS sequence irreversibly moved from functional to nonfunctional status. Four deletion rates were tested: equal to the mutation rate, 10 or 100 times slower than the mutation rate, and equal to 0 (no deletion).

Two types of copy number expansions were implemented: 1) instantaneous expansion (or burst), in which a horizontally transferred (and functional) copy was duplicated several times in the genome at once and 2) slow expansion, in which

a resident (and functional) copy was regularly selected at random and duplicated once in the genome. Horizontal transfers were simulated by adding a random functional copy from a source genome to the target genome. The source genome was a reservoir of 10 copies of the initial functional IS sequence evolving in parallel with the target genome. IS copies in the source genome underwent frequent random bursts, thus providing an unlimited reservoir of functional elements for future horizontal transfers.

Five different scenarios (1–5) were implemented, all designed to provide *n* IS copies:

Scenario 1 (single ancient burst): A single horizontal transfer at generation 1 immediately followed by an instantaneous expansion to *n* copies at generation 2.

Scenario 2 (single recent burst): A single horizontal transfer at generation 9,900 immediately followed by an instantaneous expansion to *n* copies at generation 9,901.

Scenario 3: (slow expansion): A single horizontal transfer at generation 1 followed by one copy duplication every 10,000/*n* generations.

Scenario 4 (two recent bursts): Two independent horizontal transfers at generation 9,900, each immediately followed by instantaneous expansions of *n*/2 copies at generation 9,901.

Scenario 5 (ancient and recent bursts): Two independent horizontal transfers at generations 1 and 9,900, each immediately followed by instantaneous expansions of *n*/2 copies at generations 2 and 9,901, respectively.

After the simulation ended, pairwise codon divergence between all functional and nonfunctional copies was calculated on the whole sequence length. Each distribution in figure 1 represents the pooled pairwise divergence distribution from 23 simulations, each of which has an expected final number of elements equal to the size of 1 of the 23 *Wolbachia* IS groups comprising at least two alignable IS copies (supplementary table S1, Supplementary Material online).

IS Survey across *Wolbachia* Strains

We assessed the presence or absence of 17 IS groups in a panel of 22 diverse *Wolbachia* strains from the A, B, and G supergroups, available from a previous study (Cordaux et al. 2008). Supplementary table S2, Supplementary Material online, provides details on the *Wolbachia* strains. The single *Wolbachia* infection status of each of the 22 samples was confirmed by polymerase chain reaction (PCR) amplification and sequencing of two to three chromosomal markers (*wsp*, 16S rRNA, and *GroE*) (Cordaux et al. 2008). The 17 IS groups were selected from the three sequenced genomes from the A and B supergroups (i.e., *w*Mel, *w*Ri, and *w*Pel) based on the occurrence of at least one potentially functional IS copy and/or several full-length copies. For each IS group, within-IS specific oligonucleotide primer pairs were designed to amplify 499- to 706-bp long fragments, using the program Primer3 (Rozen and Skaletsky 2000). PCR amplification, separation, and visualization were performed using a standard protocol (Cordaux et al. 2006, 2008). The D.mel, D.sim, and Slab DNA samples corresponding to the *w*Mel, *w*Ri, and *w*Pel *Wolbachia* strains, respectively, were used as positive controls (supplementary table S2, Supplementary Material online). Water controls were used in all PCR assays. PCR conditions for each IS group, including primer sequences and expected PCR product sizes, are shown in supplementary table S3, Supplementary Material online. To confirm the results, all PCR amplifications were performed twice independently, and PCR fragments were sequenced, as previously described (Cordaux et al. 2001). For each IS group, sequences obtained from PCR fragments were aligned with Bioedit, and pairwise nucleotide divergence between each pair of sequences was calculated with MEGA ver 4.0 based on observed nucleotide substitutions (sequence alignments are available upon request).

To infer the number of IS group acquisitions and losses, the distribution of each of the 17 IS groups was mapped onto a phylogeny of the 22 tested *Wolbachia* strains (Cordaux et al. 2001, 2008; Lo et al. 2007). For each IS group, we favored the most parsimonious scenario, that is, the scenario requiring the smallest number of acquisitions and losses to explain the distribution of the IS group according to *Wolbachia* strain phylogenetic relationships. For IS groups with two or more equiparsimonious scenarios, we conservatively favored the scenario minimizing the number of acquisitions.

Results and Discussion

Abundance and Distribution of IS Elements in *Wolbachia*

We used three independent and complementary strategies to identify IS elements in the four completely sequenced *Wolbachia* genomes from the *w*Mel, *w*Ri, *w*Pel, and *w*Bm strains (see "Materials and Methods"). This analysis revealed an IS copy number (at least 40 bp in length) per genome ranging from 52 in *w*Bm to 171 in *w*Ri (table 1). This is considerable given that prokaryotic genomes generally carry relatively few IS copies as illustrated by a survey of 262 genomes that identified a median number of 12 IS copies (range 0–342) per genome (Touchon and Rocha 2007). Overall, IS copies account for up to 11% (~160 kb) of *Wolbachia* genomes (table 1). Such IS genomic coverage exceeds that described in most other prokaryotic genomes, which is generally below 3% (Siguier, Filee, et al. 2006), except in a few rare cases such as *Shigella dysenteriae*, *Orientia tsutsugamushi*, or *Sulfobolus solfataricus* where it can reach >10% (Brugger et al. 2004; Yang et al. 2005; Cho et al. 2007; Filee et al. 2007; Nakayama et al. 2008).

Wolbachia IS elements encompass a total of 11 IS families (table 1), out of the ~20 major recognized IS families (Chandler and Mahillon 2002; Siguier, Filee, et al. 2006;

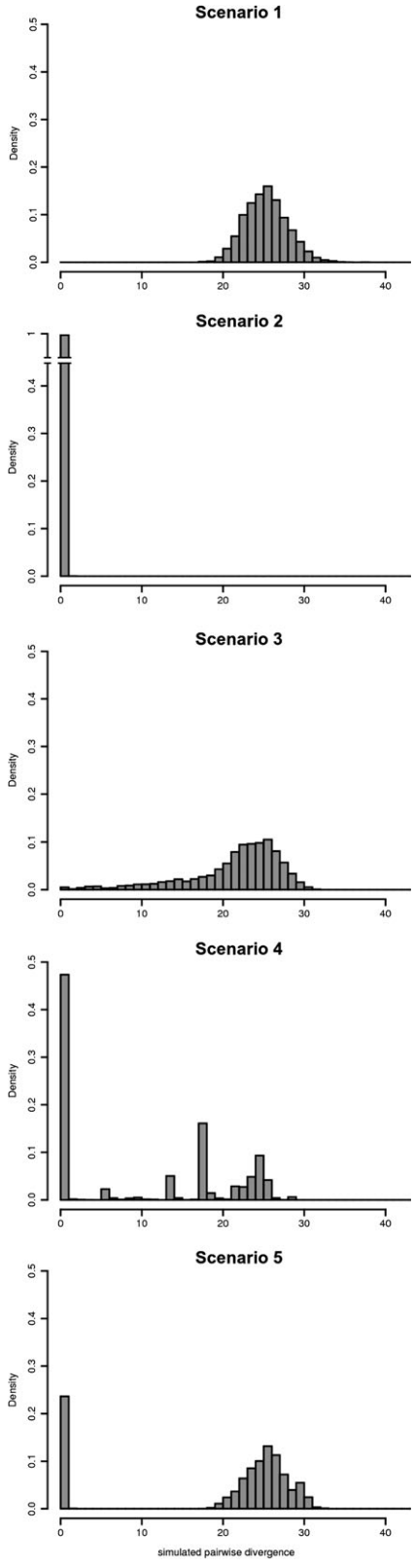

Fig. 1.—Frequency distribution of simulated pairwise IS divergence in a haploid genome under five models of IS dynamics. Scenario 1: single ancient burst; scenario 2: single recent burst; scenario 3: slow expansion; scenario 4: two independent recent bursts; and scenario 5: ancient and recent bursts. Each distribution represents the pooled pairwise divergence distribution from 23 simulations, each of which has an expected final number of elements equal to the size of 1 of the 23 *Wolbachia* IS groups comprising at least two alignable IS copies.

Siguier, Perochon, et al. 2006). Except for the IS6 and IS200/605 families that are specific to *w*Pel, all other IS families (9/11 or 82%) are shared by at least three of the four *Wolbachia* genomes (table 1). However, there are significant differences in the distribution of IS families among the various *Wolbachia* genomes (chi-square test, $P < 10^{-16}$) (table 1). Because IS families contain large numbers of heterogeneous IS types (Chandler and Mahillon 2002; Siguier, Perochon, et al. 2006), we refined our analysis by classifying all *Wolbachia* IS copies into 1 of 33 IS groups (see "Materials and Methods"). The group-level analysis confirmed the family-level analysis, in that most IS groups (27/33 or 82%) are shared by multiple genomes. However, the most frequent IS groups per genome are largely specific to each genome, and globally, the distribution of IS groups is significantly different among the various *Wolbachia* genomes (chi-square test, $P < 10^{-16}$) (supplementary table S1, Supplementary Material online).

Overall, these results demonstrate that *Wolbachia* genomes qualitatively carry IS elements from the same families and groups, but they substantially differ in the families and groups that mostly contributed to their IS genomic landscapes. This is particularly striking for the *w*Mel and *w*Ri genomes, which are phylogenetically closely related (Wu et al. 2004; Klasson et al. 2009). But the two genomes display very different IS profiles in terms of copy number, genomic coverage, and most frequent families and groups, despite the fact that *w*Mel and *w*Ri virtually possess IS elements from the same families and groups (table 1; supplementary table S1, Supplementary Material online).

Diversity of Potentially Functional IS Copies in *Wolbachia*

To investigate the causes of differential IS abundance and distribution among *Wolbachia* genomes, we searched for potentially functional IS copies, defined as full-length copies with intact transposase genes. This analysis revealed that the *w*Mel, *w*Ri, and *w*Pel genomes possess 20–64 potentially functional IS copies (table 2). Each of these *Wolbachia* genomes possesses at least one potentially functional copy from six different IS groups. Overall, there is an important diversity of potentially functional IS copies in *Wolbachia*, belonging to 14 different IS groups from 9 different IS families. Not surprisingly, IS groups exhibiting the highest numbers of potentially functional copies within individual *Wolbachia* genomes are also the IS groups with the highest overall copy numbers. Altogether, these results suggest that multiple IS copy number expansions have taken place during recent

Table 1
Distribution of IS Elements in Four Completely Sequenced *Wolbachia* Genomes

IS Family	*w*Bm[a]	*w*Mel[a]	*w*Ri[a]	*w*Pel[a]
IS3		18 (17)	15 (9)	2 (1)
IS4	3 (6)	13 (12)	12 (7)	5 (3)
IS5	7 (13)	26 (25)	30 (17)	26 (15)
IS6				11 (6)
IS110	21 (40)	18 (17)	39 (23)	6 (4)
IS200/605				5 (3)
IS256	1 (2)	3 (3)	2 (1)	32 (19)
IS481	3 (6)	4 (4)	12 (7)	5 (3)
IS630	3 (6)	7 (7)	12 (7)	15 (9)
IS982	1 (2)	2 (2)	3 (2)	54 (32)
IS66	13 (25)	14 (13)	46 (27)	9 (5)
Total IS copy number	52	105	171	170
Total IS family number	8	9	9	11
IS density (copies/Mb)	48	83	118	115
IS genomic coverage (bp)	28,188	76,886	159,767	123,823
IS genomic proportion	2.6	6.1	11.0	8.4

[a]The first number relates to observed copy number. The number in brackets indicates the proportion of the IS family among all IS copies in the genome.

Wolbachia evolution as a result of IS group-specific and genome-specific expansions.

The ISWen2 group in *w*Ri provides an excellent example of a group-specific expansion in a specific *Wolbachia* genome (fig. 2*a*). ISWen2 copies were identified only in the closely related *w*Mel and *w*Ri genomes. Despite the fact that both genomes carry at least one potentially functional ISWen2 copy, *w*Ri carries as many as 23 ISWen2 copies, whereas *w*Mel carries only 3 copies. To investigate the recent amplification dynamics of the ISWen2 group in *w*Ri, we performed a phylogenetic analysis using the median-joining network approach, as implemented in the Network ver 4.5.1.6 software (Bandelt et al. 1999; Cordaux, Hedges, et al. 2004). The ISWen2 group network displays a star-like structure in which 35% of the copies fall in the central, most likely ancestral, node (I in fig. 2*b*). Interestingly, several peripheral nodes in the network are not directly connected to the central node (III.1, III.2, and III.3 in fig. 2*b*) or encompass

Table 2
Distribution of Potentially Functional IS Copies Inserted in Four *Wolbachia* Genomes

IS Family	IS Group	*w*Bm	*w*Mel	*w*Ri	*w*Pel
IS4	ISWen1		3		
IS5	ISWpi1		13	20	
IS110	ISWen2		1	16	
	ISWpi12				3
	ISWpi13			1	
	ISWpi14		1		
IS200/605	ISW1				2
IS256	ISWpi15		1		6
IS481	ISWpi2			5	
	ISWpi4		1	1	
IS630	ISWpi11				4
	ISWpi10				1
IS982	ISWpi16				44
IS66	ISWen3			21	
Number of potentially functional IS copies		0	20	64	60
Proportion (%)		0	19	37	35
Number of nonfunctional IS copies		52	85	107	110
Proportion (%)		100	81	63	65

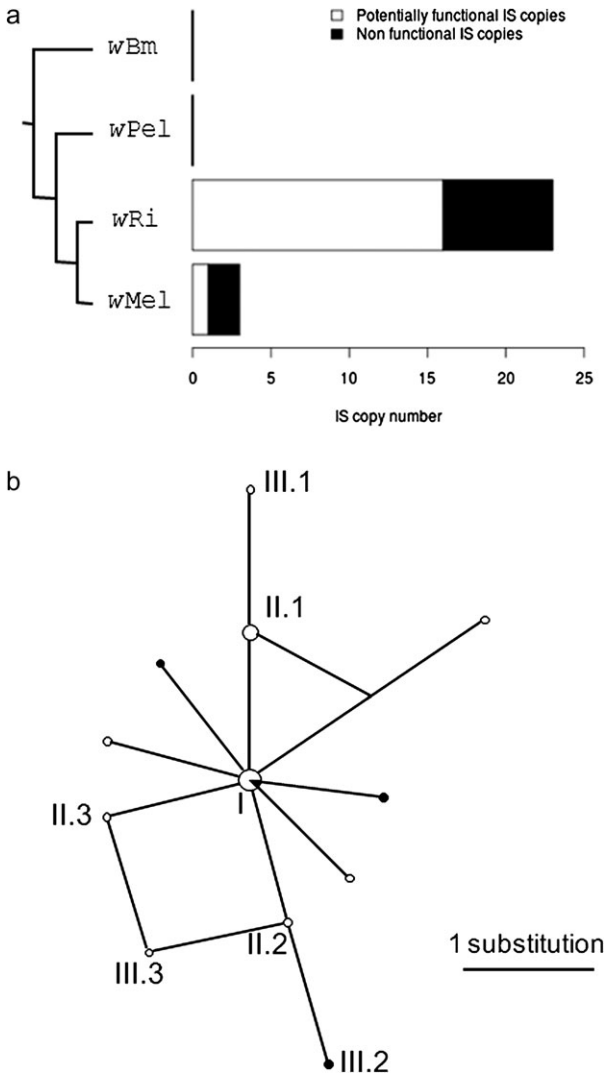

Fig. 2.—Expansion of the ISWen2 group in the *w*Ri genome. (*a*) Copy number of the ISWen2 group in four completely sequenced *Wolbachia* genomes. Branch lengths of the phylogenetic tree are arbitrary. (*b*) Median-joining network of the 20 full-length ISWen2 copies from the *w*Ri genome. Circles denote IS sequence types (nodes). Nodes discussed in the main text were labeled I, II.1–3, and III.1–3. Node size is proportional to IS copy number: $n = 1$ for all nodes except nodes I ($n = 7$) and II.1 ($n = 3$). Lines denote substitution steps, with a one-step distance being indicated in the lower right corner. Potentially functional and nonfunctional copies are shown in white and black, respectively.

Table 3
Insertion Presence/Absence Polymorphism Patterns of 44 IS Copies in the *w*Mel and *w*Ri *Wolbachia* Genomes

IS Family	IS Group	IS Copies Specific to *w*Mel	IS Copies Specific to *w*Ri	IS Copies Shared by *w*Mel and *w*Ri
IS3	IS3			12
IS4	IS4-wB			6
IS5	IS903	1		
	IS1031			1
	ISWpi1	6	4	
IS110	IS1111			1
	ISWen2		4	
	ISWpi12			1
	IS110-wA			1
IS481	ISWpi2		1	
	ISWpi4			1
IS630	ISWpi11	1		
IS982	ISWpi16			1
IS66	ISWen3		2	1
Number of IS copies		8	11	25
Proportion of IS copies that are potentially functional (%)		75	91	4

several ISWen2 sequences (II.1 in fig. 2*b*). Given that bacterial IS elements generally exhibit strong cis preference for transpositional activity (i.e., preferential interaction of a transposase with the element from which it is expressed) (Chandler and Mahillon 2002; Nagy and Chandler 2004) and assuming that homoplasy is negligible at this phylogenetic depth, we conclude that the ISWen2 expansion in *w*Ri may have been mediated by at least 3 IS copies (at least one copy from nodes I and II.1 and the copy at the node II.2) and at most 11 copies (from nodes I, II.1, II.2, and II.3). These results suggest that 15–55% of the copies may have contributed to ISWen2 expansion during recent *w*Ri evolution. It is generally thought that multiple copies may contribute to the expansion of DNA transposon families (Deininger and Batzer 1993; Robertson 2002). To our knowledge, our analysis provides the first quantitative estimate of the proportion of "source" copies that may have contributed to the expansion of a DNA transposon family.

Intense and Global IS Transpositional Activity

We have previously shown that one IS group (ISWpi1) is widespread among *Wolbachia* strains, but individual ISWpi1 copies are inserted at given genomic loci in a single or very few closely related *Wolbachia* strains (Cordaux 2008; Cordaux et al. 2008). Such insertion presence/absence polymorphism patterns demonstrate intense ISWpi1 transpositional activity during recent *Wolbachia* evolution (Cordaux 2008; Cordaux et al. 2008). To investigate whether this evolutionary trend can be extended to other IS groups, we searched for IS insertion presence/absence polymorphisms at orthologous sites for all full-length IS elements inserted in the *w*Mel and *w*Ri genomes. We found 19 IS copies specifically inserted in *w*Mel or *w*Ri out of 44 unambiguously orthologous loci identified between *w*Mel and *w*Ri (table 3). The 19 polymorphic insertions encompass six different IS groups (including ISWpi1) from five different IS families, indicating that multiple IS groups have been transpositionaly active during recent *Wolbachia* evolution. The fact that 84% (16/19) of polymorphic insertions versus only 4% (1/25) of shared insertions by *w*Mel and *w*Ri are potentially functional IS copies (table 3) further corroborates the recent origin of the polymorphic IS copies (i.e., they have not resided in the genomes long time enough as to accumulate inactivating mutations). This result also suggests that IS transpositional activity may be ongoing in these *Wolbachia* genomes.

Exceptional Amount of Nonfunctional Copies in *Wolbachia* Genomes

The analysis of IS copy structure revealed that the *w*Mel, *w*Ri, and *w*Pel *Wolbachia* genomes contain several tens of potentially functional IS copies (table 2). However, these copies only account for a small fraction of all IS copies inserted in *Wolbachia* genomes. In fact, 71% of all IS elements inserted in *Wolbachia* genomes are nonfunctional (table 2). Nonfunctional IS copies are defined here as full-length copies with pseudogenized transposase genes or non–full-length copies (i.e., truncated copies and fragments). In the most extreme case, all 52 IS copies inserted in the *w*Bm genome are nonfunctional (Cordaux 2009). Given that IS elements are usually considered to be of recent origin in bacterial genomes and subject to rapid turnover (Wagner 2006; Wagner et al. 2007; Rocha 2008), the occurrence of so many disrupted and degraded IS elements in *Wolbachia* genomes is all the more surprising.

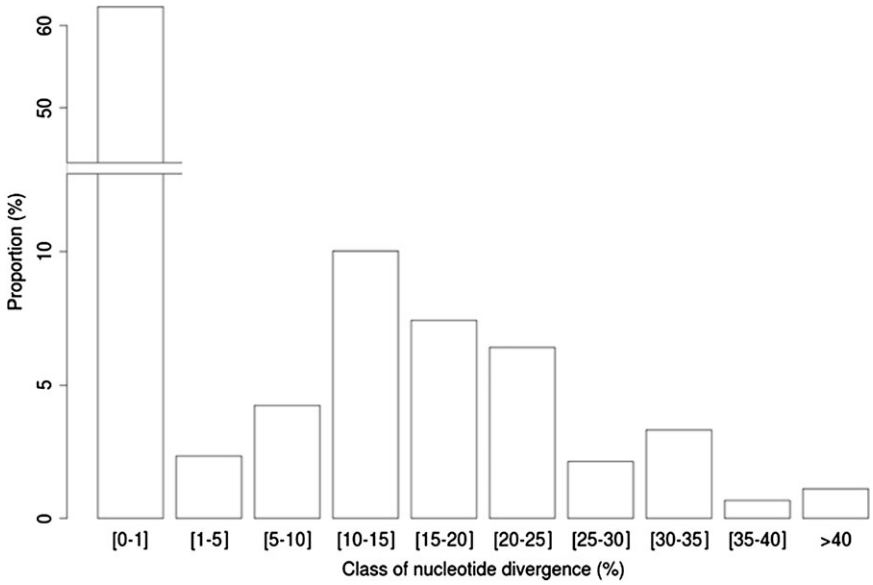

Fig. 3.—Frequency distribution of pairwise IS nucleotide divergence for four *Wolbachia* genomes. IS copies from 23 IS groups comprising at least two alignable IS copies are considered ($n = 454$). The distribution is based on a total of 4,312 pairwise comparisons (169 for *w*Bm, 379 for *w*Mel, 2152 for *w*Pel, and 1612 for *w*Ri).

Investigating Long-term Evolutionary Dynamics of IS Elements

The many degraded IS copies in *Wolbachia* genomes offer a unique opportunity to directly investigate the long-term evolutionary dynamics of bacterial TEs. We analyzed 454 copies from the 23 IS groups comprising at least two copies with alignable sequences from the four *Wolbachia* genomes. For each IS group, we calculated intragenomic nucleotide divergence between pairs of IS copies. The majority of pairwise comparisons exhibited <1% nucleotide divergence (fig. 3). This is consistent with a recent origin of these IS copies and the high copy numbers of recently expanded IS groups, which generates many pairwise comparisons with no or very low divergence. In addition, we found that in the four *Wolbachia* genomes, 22–47% of the pairwise comparisons displayed at least 10% nucleotide divergence (fig. 3). This indicates that *Wolbachia* genomes contain an important amount of ancient IS copies that are witnesses of past IS expansions during *Wolbachia* evolution.

Interestingly, the distribution of IS copies is bimodal, with a first peak corresponding to identical or nearly identical IS copies (<1% divergence) and a second peak at 10–15% divergence (fig. 3). Importantly, this pattern holds when *Wolbachia* genomes are analyzed separately (supplementary fig. S1, Supplementary Material online). This demonstrates that the global bimodal pattern cannot be ascribed to an artifact due to pooling data from multiple genomes that would exhibit different individual distribution patterns. To explore the evolutionary causes of this bimodal distribution, we simulated the evolution of an IS population in a haploid (bacterial) genome under different scenarios. Four major processes of IS evolution were considered: acquisition by horizontal transfer, copy number expansion from a resident copy, degradation through random substitutions, and loss (see "Materials and Methods"). These processes were combined to test five different evolutionary scenarios differing in the tempo of IS acquisitions and bursts.

The first two scenarios simulated a single IS acquisition immediately followed by a sudden burst, at the start (scenario 1) or near the end (scenario 2) of the simulation. We observed a single peak in both simulations, with high divergence and variance for the scenario of ancient IS acquisition and burst (scenario 1) and low divergence and variance for the scenario of recent IS acquisition and burst (scenario 2) (fig. 1). Next, scenario 3 simulated a constant but low transpositional activity for the duration of the simulation, following an initial IS copy acquisition. The resulting distribution showed a single peak at high divergence skewed toward lower divergence (fig. 1). Finally, scenarios 4 and 5 each simulated two independent IS acquisitions and bursts: Both were recent in scenario 4 (i.e., corresponding to scenario 2 repeated twice), and one was ancient and one recent in scenario 5 (i.e., combining scenarios 1 and 2). Scenario 5 is analogous to the model of recurrent horizontal transfers and bursts proposed in the literature (Wagner 2006). The distribution pattern resulting from two recent IS acquisitions bursts (scenario 4) showed multiple peaks with low variance at irregular divergence levels (fig. 1). Interestingly, scenario 5 was the only one that displayed a clear bimodal distribution with a flat peak at high divergence (corresponding to the ancient IS acquisition and burst) and

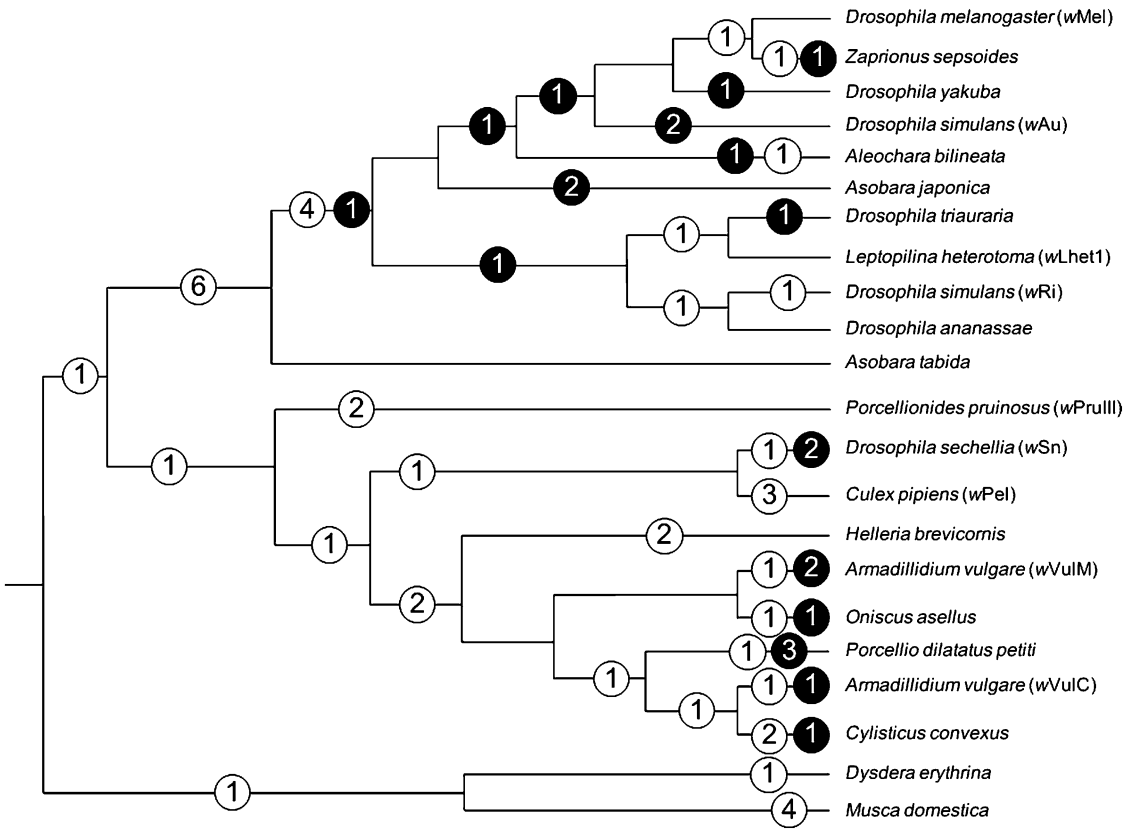

Fig. 4.—History of IS group acquisitions and losses in 22 *Wolbachia* strains. The most parsimonious distribution of acquisitions (white circles) and losses (black circles) of 17 IS groups according to the phylogenetic relationships of 22 *Wolbachia* strains is shown. Numbers of acquisitions and losses are indicated in the circles. Phylogenetic relationships between *Wolbachia* strains are adapted from Cordaux et al. (2001, 2008) and Lo et al. (2007); branch lengths of the phylogenetic tree are arbitrary. *Wolbachia* strains are named after the host species from which they were isolated.

a straight peak at low divergence (corresponding to the recent IS acquisition and burst) (fig. 1), as observed for *Wolbachia* IS elements (fig. 3). Note that higher deletion rates resulted in lower amounts of ancient copies in scenario 5, but the second peak corresponding to ancient copies was apparent whatever the deletion rate (supplementary fig. S2, Supplementary Material online).

Overall, the bimodal distribution indicates that IS transposition activity in *Wolbachia* genomes was not constant over time. Instead, most *Wolbachia* IS elements may have been generated in at least two major periods of intense transpositional activity, including an ancient expansion embodied by ~15% divergent IS copies and a very recent (and perhaps ongoing) expansion corresponding to identical or nearly identical IS copies. Our simulations also emphasize that multiple horizontal IS acquisitions are required to explain the observed distribution pattern.

Frequent IS Horizontal Transfers among *Wolbachia* Strains

To test whether *Wolbachia* IS dynamics is intimately linked with frequent horizontal transmission among strains, as suggested by our simulations above and by evidence from the ISWpi1 group (Cordaux et al. 2008), we screened a panel of 22 diverse *Wolbachia* strains (supplementary table S2, Supplementary Material online) for the presence of 17 IS groups using group-specific PCR detection assays and verification by sequencing (supplementary table S4, Supplementary Material online). To evaluate the reliability of our PCR assays, we compared PCR amplification results in our *w*Mel, *w*Ri, and *w*Pel DNA samples with genome sequence predictions. We obtained results in agreement with expectations in 47 of 51 combinations tested (i.e., 17 IS groups in three reference strains). For the four cases in which conflicting results were recorded (i.e., no PCR result, although an amplification is predicted based on presence of at least one copy of the tested IS group in the genome sequence), average sequence divergence between IS copies among genomes was high in all cases (>13%). Therefore, we conclude that our PCR detection assays are generally highly reliable, at least for copies with low to moderate divergence, hence enabling us to confidently detect recent events of IS horizontal transfers that may have occurred between *Wolbachia* strains.

By mapping IS group distribution onto a phylogeny of the *Wolbachia* strains, we inferred the most parsimonious scenario of IS group acquisitions and losses during the

evolutionary history of the 22 *Wolbachia* strains under investigation (fig. 4). This analysis revealed that all investigated *Wolbachia* genomes (encompassing three different supergroups) possess IS elements from at least three different IS groups. Our large-scale screening thus demonstrates that IS elements are broadly albeit patchily distributed among *Wolbachia* strains. Thus, IS represents a general feature of *Wolbachia* strains, not merely a characteristic of the few sequenced genomes.

Our results also indicate that the presence of the 17 IS groups in the 22 *Wolbachia* strains requires at least 44 independent acquisitions at this level of resolution (fig. 4). This is most likely a very conservative estimate because 1) we favored the scenario minimizing the number of acquisitions when several equiparsimonious scenarios were possible for particular IS groups; 2) we assumed that individual *Wolbachia* strains or monophyletic groups of *Wolbachia* strains possessing a given IS group resulted from a single ancestral acquisition; however, multiple independent acquisitions could also explain such distribution patterns, as previously shown for ISWpi1 (Cordaux et al. 2008); and 3) a larger screening of *Wolbachia* strains for IS group presence might uncover additional acquisition events.

Such a patchy distribution strongly suggests that horizontal transfers of IS copies occur frequently in *Wolbachia*. This is further substantiated by the fact that >60% of the pairwise comparisons of IS sequences obtained by sequencing of PCR fragments displayed nucleotide divergence <1% (supplementary fig. S3, Supplementary Material online). As this is lower than the divergence between most of the analyzed *Wolbachia* strains, such IS groups are unlikely to have been vertically inherited from a common ancestor. Based on patchy distribution of IS groups and generally high similarity of IS sequences between strains, we conclude that horizontal transmission is a major determinant of the current IS distribution in *Wolbachia* strains.

The frequent horizontal transmission of IS elements across *Wolbachia* strains poses the question of the underlying mechanisms of these transfers. The intracellular confinement of bacterial endosymbionts is generally thought to limit exchange of genetic material with other bacterial populations or species (Wernegreen 2002; Moran et al. 2008; Moya et al. 2008). However, a distinguishing feature of *Wolbachia* endosymbionts is their propensity to switch between arthropod hosts (Vavre et al. 1999; Cordaux et al. 2001). Such dynamics favors the occasional co-occurrence of divergent *Wolbachia* strains within the same host cells, either stably or transiently (Vavre et al. 1999; Bordenstein and Wernegreen 2004; Verne et al. 2007). In addition to physical proximity, exchange of genetic material between *Wolbachia* strains might be facilitated by the presence of bacteriophages in many *Wolbachia* endosymbionts (Bordenstein and Wernegreen 2004; Braquart-Varnier et al. 2005; Tanaka et al. 2009) that might serve as shuttles for transferring IS elements among strains. We identified a total of 15 IS elements from 9 different IS groups, including 11 potentially functional copies, inserted in the 12 prophages integrated in the *w*Mel, *w*Ri, and *w*Pel genomes (table 4). However, it is unclear whether these IS copies inserted in prophage genomes following phage integration into *Wolbachia* genomes or the IS elements were already present in bacteriophage genomes and were imported in *Wolbachia* genomes during bacteriophage genome integration. Nevertheless, a potentially functional ISWpi12 copy is inserted in the genome of the active bacteriophage WOcauB2 of the *w*CauB *Wolbachia* strain (Tanaka et al. 2009), whereas bacteriophage genomes most generally lack IS elements (Leclercq and Cordaux 2011). This is consistent with the notion that bacteriophages might be able to shuttle IS elements between *Wolbachia* strains.

Table 4
Distribution of IS Copies Inserted in Prophage Regions of Three *Wolbachia* Genomes

IS Family	IS Group	*w*Mel[a]	*w*Ri[a]	*w*Pel[a]
IS5	IS1031			1 (0)
	ISWpi1	1 (1)	3 (3)	
IS110	ISWen2	1 (1)	2 (2)	
	ISWpi12	1 (0)		1 (1)
	ISWpi13		1 (1)	
	ISWpi14	1 (1)		
IS256	ISWpi15			1 (0)
IS630	ISWpi10		1 (0)	
IS982	ISWpi16			1 (1)
Number of IS copies		4 (3)	7 (6)	4 (2)
Number of prophage regions		3	4	5

[a]Number of potentially functional IS copies shown in brackets.

Conclusions

Our analyses highlighted the patchy distribution of IS groups in *Wolbachia* genomes. The identification of multiple IS groups experiencing independent copy number expansions in different *Wolbachia* genomes is notable because it suggests that IS expansions may occur simultaneously in different genomes (i.e., *w*Mel, *w*Ri, and *w*Pel) through a global activation of transposition. This synchronicity may be linked to the high rate of recent IS horizontal transfers we identified in *Wolbachia* strains. Nevertheless, the evolutionary success of IS families and groups within genomes is highly variable, indicating that horizontal transfer is a necessary but not sufficient condition to IS proliferation. The apparently stochastic loss or success of individual IS families or groups within bacterial strains following import by horizontal transfer may be the result of a complex interplay between various parameters, such as IS intrinsic transpositional efficiency, cellular factors involved in transpositional control, and genomic environment (Chandler and Mahillon 2002; Nagy and

Chandler 2004; Cerveau et al. 2011). However, such targeted effects can hardly explain a global activation of transposition simultaneously involving multiple IS families and groups. This suggests that population-level effects may also play a role in the evolutionary dynamics of bacterial IS elements.

Remarkably, our results show that *Wolbachia* genomes contain an important archive of past IS evolution, as the vast majority of *Wolbachia* IS copies actually are more or less severely degraded. The rich IS fossil record buried in *Wolbachia* genomes provides direct empirical evidence for a long-term evolutionary dynamics of IS elements following a scenario of cyclic bursts of transposition separated by periods of relative transpositional quiescence as previously suggested based on the analysis of exclusively recent IS copies (Wagner 2006; Wagner et al. 2007). This raises the question whether the abundance of IS fossils is specific to *Wolbachia* genomes or it may be a general, albeit overlooked, feature of prokaryote genomes. Unfortunately, IS annotation is rarely optimal in completely sequenced prokaryotic genomes, and currently, it is often limited at best to identification of potentially functional transposase genes (Varani et al. 2011). Therefore, it is possible that many other prokaryote genomes carry an abundance of IS relics, but they cannot be detected using standard annotation procedures. In any event, our detailed analysis of IS elements in *Wolbachia* bacteria shows that comprehensive TE annotations have the potential to uncover unexpected patterns of prokaryote genome evolution. Indeed, the identification of an IS fossil genomic record in *Wolbachia* demonstrates that IS elements are not always of recent origin, contrary to the conventional view of TE evolution in prokaryote genomes.

Supplementary Material

Supplementary tables S1–S4 and supplementary figures S1–S3 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

We are grateful to Mylène Weill, Fabrice Vavre, Hervé Merçot, and Denis Poinsot for providing samples. We thank Mick Chandler and Patricia Siguier for access to ISsaga prior to publication and discussions and Daniel Guyonnet for technical assistance. This research was funded by a Young Investigator ATIP Award from the Centre National de la Recherche Scientifique (CNRS) to R.C., a European Research Council Starting Grant (FP7/2007-2013 grant 260729 EndoSexDet) to R.C., and the Research Interdisciplinary Programme "Infectious Diseases and Environment" from the CNRS. N.C. was supported by a PhD fellowship from the French Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche. S.L. was supported by a postdoctoral fellowship from the CNRS.

Literature Cited

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 16(1):37–48.

Bichsel M, Barbour AD, Wagner A. 2010. The early phase of a bacterial insertion sequence infection. Theor Popul Biol. 78(4):278–288.

Bordenstein SR, Reznikoff WS. 2005. Mobile DNA in obligate intracellular bacteria. Nat Rev Microbiol. 3(9):688–699.

Bordenstein SR, Wernegreen JJ. 2004. Bacteriophage flux in end (Wolbachia): infection frequency, lateral transfer, and recombination rates. Mol Biol Evol. 21(10):1981–1991.

Bouchon D, Cordaux R, Grève P. 2008. Feminizing Wolbachia and the evolution of sex determination in isopods. In: Bourtzis K, Miller T, editors. Insect Symbiosis. Insect Symbiosis, Vol. 3. Boca Raton (FL): Taylor and Francis Group LLC. p. 273–294.

Braquart-Varnier C, Greve P, Felix C, Martin G. 2005. Bacteriophage WO in Wolbachia infecting terrestrial isopods. Biochem Biophys Res Commun. 337:580–585.

Brugger K, Torarinsson E, Redder P, Chen L, Garrett RA. 2004. Shuffling of Sulfolobus genomes by autonomous and non-autonomous mobile elements. Biochem Soc Trans. 32(Pt 2):179–183.

Cerveau N, Leclercq S, Bouchon D, Cordaux R. 2011. Evolutionary dynamics and genomic impact of prokaryote transposable elements. In: Pontarotti P, editor. Evolutionary biology: concepts, biodiversity, macroevolution and genome evolution. Berlin (Germany): Springer. p. 291–312.

Chandler M, Mahillon J. 2002. Insertion sequences revisited. In: Craig NL, Craig R, Gellert M, Lambowitz AM, editors. Mobile DNA II. Washington (DC): ASM Press. p. 305–366.

Cho NH, et al. 2007. The Orientia tsutsugamushi genome reveals massive proliferation of conjugative type IV secretion system and host-cell interaction genes. Proc Natl Acad Sci U S A. 104(19):7981–7986.

Cordaux R. 2008. ISWpi1 from Wolbachia pipientis defines a novel group of insertion sequences within the IS5 family. Gene 409(1–2):20–27.

Cordaux R. 2009. Gene conversion maintains nonfunctional transposable elements in an obligate mutualistic endosymbiont. Mol Biol Evol. 26(8):1679–1682.

Cordaux R, Batzer MA. 2009. The impact of retrotransposons on human genome evolution. Nat Rev Genet. 10(10):691–703.

Cordaux R, Bouchon D, Greve P. 2011. The impact of endosymbionts on the evolution of host sex-determination mechanisms. Trends Genet. 27:332–341.

Cordaux R, Hedges DJ, Batzer MA. 2004. Retrotransposition of Alu elements: how many sources? Trends Genet. 20(10):464–467.

Cordaux R, Lee J, Dinoso L, Batzer MA. 2006. Recently integrated Alu retrotransposons are essentially neutral residents of the human genome. Gene 373:138–144.

Cordaux R, Michel-Salzat A, Bouchon D. 2001. Wolbachia infection in crustaceans: novel hosts and potential routes for horizontal transmission. J Evol Biol. 14(2):237–243.

Cordaux R, Michel-Salzat A, Frelon-Raimond M, Rigaud T, Bouchon D. 2004. Evidence for a new feminizing Wolbachia strain in the isopod Armadillidium vulgare: evolutionary implications. Heredity 93(1):78–84.

Cordaux R, et al. 2008. Intense transpositional activity of insertion sequences in an ancient obligate endosymbiont. Mol Biol Evol. 25(9):1889–1896.

Deininger PL, Batzer MA. 1993. Evolution of retroposons. Evol Biol. 27:157–196.

Emerman M, Malik HS. 2010. Paleovirology—modern consequences of ancient viruses. PLoS Biol. 8(2):e1000301.

Feschotte C, Pritham EJ. 2007. DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet. 41:331–368.
Filee J, Siguier P, Chandler M. 2007. Insertion sequence diversity in archaea. Microbiol Mol Biol Rev. 71(1):121–157.
Foster J, et al. 2005. The Wolbachia genome of Brugia malayi: endosymbiont evolution within a human pathogenic nematode. PLoS Biol. 3(4):e121.
Gilbert C, Feschotte C. 2010. Genomic fossils calibrate the long-term evolution of hepadnaviruses. PLoS Biol. 8(9):e1000495.
Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser. 41:95–98.
Kapitonov VV, Jurka J. 2003. Molecular paleontology of transposable elements in the Drosophila melanogaster genome. Proc Natl Acad Sci U S A. 100(11):6569–6574.
Khan H, Smit A, Boissinot S. 2006. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. Genome Res. 16(1):78–87.
Klasson L, et al. 2008. Genome evolution of Wolbachia strain wPip from the Culex pipiens group. Mol Biol Evol. 25(9):1877–1887.
Klasson L, et al. 2009. The mosaic genome structure of the Wolbachia wRi strain infecting Drosophila simulans. Proc Natl Acad Sci U S A. 106(14):5725–5730.
Kumar S, Nei M, Dudley J, Tamura K. 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform. 9(4):299–306.
Kuo CH, Ochman H. 2010. The extinction dynamics of bacterial pseudogenes. PLoS Genet. 6(8):e1001050.
Lander ES, et al. 2001. Initial sequencing and analysis of the human genome. Nature 409(6822):860–921.
Lawrence JG, Ochman H, Hartl DL. 1992. The evolution of insertion sequences within enteric bacteria. Genetics 131(1):9–20.
Leclercq S, Cordaux R. 2011. Do phages efficiently shuttle transposable elements among prokaryotes? Evolution. Advance Access published September 13, 2011, doi:10.1111/j.1558-5646.2011.01395.x.
Leclercq S, Giraud I, Cordaux R. 2011. Remarkable abundance and evolution of mobile group II introns in Wolbachia bacterial endosymbionts. Mol Biol Evol. 28(1):685–697.
Lo N, et al. 2007. Taxonomic status of the intracellular bacterium Wolbachia pipientis. Int J Syst Evol Microbiol. 57(Pt 3):654–657.
Moon S, Byun Y, Kim HJ, Jeong S, Han K. 2004. Predicting genes expressed via −1 and +1 frameshifts. Nucleic Acids Res. 32(16):4884–4892.
Moran NA, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet. 42:165–190.
Moran NA, Plague GR. 2004. Genomic changes following host restriction in bacteria. Curr Opin Genet Dev. 14(6):627–633.
Moya A, Pereto J, Gil R, Latorre A. 2008. Learning how to live together: genomic insights into prokaryote-animal symbioses. Nat Rev Genet. 9(3):218–229.
Nagy Z, Chandler M. 2004. Regulation of transposition in bacteria. Res Microbiol. 155(5):387–398.
Nakayama K, et al. 2008. The whole-genome sequencing of the obligate intracellular bacterium Orientia tsutsugamushi revealed massive gene amplification during reductive genome evolution. DNA Res. 15(4):185–199.
Pace JK 2nd, Feschotte C. 2007. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. Genome Res. 17(4):422–432.
Parkhill J, et al. 2003. Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica. Nat Genet. 35(1):32–40.
Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large genomes. Bioinformatics 21(Suppl 1):i351–i358.
Qiu N, et al. 2010. Prevalence and diversity of insertion sequences in the genome of Bacillus thuringiensis YBT-1520 and comparison with other Bacillus cereus group members. FEMS Microbiol Lett. 310(1):9–16.
Robertson HM. 2002. Evolution of DNA transposons in eukaryotes. In: Craig NL, Craig R, Gellert M, Lambowitz AM, editors. Mobile DNA II. Washington (DC): ASM Press. p.1093–1110.
Rocha EP. 2008. Evolutionary patterns in prokaryotic genomes. Curr Opin Microbiol. 11(5):454–460.
Rozen S, Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol. 132:365–386.
Saridaki A, Bourtzis K. 2010. Wolbachia: more than just a bug in insects genitals. Curr Opin Microbiol. 13(1):67–72.
Sawyer SA, et al. 1987. Distribution and abundance of insertion sequences among natural isolates of Escherichia coli. Genetics 115(1):51–63.
Schnable PS, et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. Science 326(5956):1112–1115.
Siguier P, Filee J, Chandler M. 2006. Insertion sequences in prokaryotic genomes. Curr Opin Microbiol. 9(5):526–531.
Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 34(Database issue):D32–D36.
Tanaka K, Furukawa S, Nikoh N, Sasaki T, Fukatsu T. 2009. Complete WO phage sequences reveal their dynamic evolutionary trajectories and putative functional elements required for integration into the Wolbachia genome. Appl Environ Microbiol. 75(17):5676–5686.
Touchon M, Rocha EP. 2007. Causes of insertion sequences abundance in prokaryotic genomes. Mol Biol Evol. 24(4):969–981.
Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M. 2011. ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. Genome Biol. 12(3):R30.
Vavre F, Fleury F, Lepetit D, Fouillet P, Bouletreau M. 1999. Phylogenetic evidence for horizontal transmission of Wolbachia in host-parasitoid associations. Mol Biol Evol. 16(12):1711–1723.
Verne S, Johnson M, Bouchon D, Grandjean F. 2007. Evidence for recombination between feminizing Wolbachia in the isopod genus Armadillidium. Gene 397(1–2):58–66.
Wagner A. 2006. Periodic extinctions of transposable elements in bacterial lineages: evidence from intragenomic variation in multiple genomes. Mol Biol Evol. 23(4):723–733.
Wagner A, Lewis C, Bichsel M. 2007. A survey of bacterial insertion sequences using IScan. Nucleic Acids Res. 35(16):5284–5293.
Wernegreen JJ. 2002. Genome evolution in bacterial endosymbionts of insects. Nat Rev Genet. 3(11):850–861.
Werren JH, Zhang W, Guo LR. 1995. Evolution and phylogeny of Wolbachia: reproductive parasites of arthropods. Proc R Soc Lond Ser B Biol Sci. 261(1360):55–63.
Wu M, et al. 2004. Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: a streamlined genome overrun by mobile genetic elements. PLoS Biol. 2(3):E69.
Xing J, et al. 2004. Alu element mutation spectra: molecular clocks and the effect of DNA methylation. J Mol Biol. 344(3):675–682.
Yang F, et al. 2005. Genome dynamics and diversity of Shigella species, the etiologic agents of bacillary dysentery. Nucleic Acids Res. 33(19):6445–6458.

Associate editor: Emmanuelle Lerat