Widespread Recombination Throughout *Wolbachia* Genomes

Laura Baldo, Seth Bordenstein,† Jennifer J. Wernegreen,† and John H. Werren‡*

*Department of Biology, University of California, Riverside; †Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, The Marine Biological Laboratory, Woods Hole; and ‡Department of Biology, University of Rochester

Evidence is growing that homologous recombination is a powerful source of genetic variability among closely related free-living bacteria. Here we investigate the extent of recombination among housekeeping genes of the endosymbiotic bacteria *Wolbachia*. Four housekeeping genes, *gltA*, *dnaA*, *ftsZ*, and *groEL*, were sequenced from a sample of 22 strains belonging to supergroups A and B. Sequence alignments were searched for recombination within and between genes using phylogenetic inference, analysis of genetic variation, and four recombination detection programs (MaxChi, Chimera, RDP, and Geneconv). Independent analyses indicate no or weak intragenic recombination in *ftsZ*, *dnaA*, and *groEL*. Intragenic recombination affects *gltA*, with a clear evidence of horizontal DNA transfers within and between divergent *Wolbachia* supergroups. Intergenic recombination was detected between all pairs of genes, suggesting either a horizontal exchange of a genome portion encompassing several genes or multiple recombination events involving smaller tracts along the genome. Overall, the observed pattern is compatible with pervasive recombination. Such results, combined with previous evidence of recombination in a surface protein, phage, and IS elements, support an unexpected chimeric origin of *Wolbachia* strains, with important implications for *Wolbachia* phylogeny and adaptation of these obligate intracellular bacteria in arthropods.

Introduction

Parasexual processes, such as transformation, conjugation, and transduction, play important roles in the evolution of many bacterial groups. While the acquisition of new sets of genes (e.g., pathogenicity genes) via horizontal gene transfer have been widely documented (Medigue et al. 1991; Groisman and Ochman 1996; Hacker et al. 1997; Patil and Sonti 2004; Lesic and Carniel 2005; Ochman, Lerat, and Daubin 2005), the extent and the significance of homologous recombination, that is, the asymmetrical process that replaces a sequence in a recipient genome with homologous DNA from a donor genome, are more controversial. Detecting foreign sequences in a recipient genome is often confounded by high nucleotide similarity between the two recombining sequences (e.g., when recombination occurs among close relatives) or the obscuring of recombination signatures by subsequent mutations. Nevertheless such "localized sex" (as termed by Maynard Smith, Dowson and Spratt 1991) has significant implications (Doolittle 1999; Ochman, Lawrence, and Groisman 2000; Koonin, Makarova, and Aravind 2001; Jain et al. 2002; Boucher et al. 2003; Lawrence and Hendrickson 2003). By mixing genomes, recombination challenges the traditional view of a single evolutionary history for individual bacterial strains, thus confounding molecular phylogenetic reconstructions (Holmes, Urwin, and Maiden 1999; Feil and Spratt 2001; Spratt, Hanage, and Feil 2001) and their derived inferences. Homologous recombination also represents a powerful engine for accelerating genome evolution and adaptation by orthologous replacement and the generation of new, recombinant alleles (Feavers et al. 1992; Maiden 1993; Holmes, Urwin, and Maiden 1999).

Because similar sequences are more likely to escape the mismatch correction system, closely related strains or highly similar stretches of DNA generally undergo successful recombination events (i.e., the recombinant sequence is retained by the recipient genome and vertically inherited; Majewski and Cohan 1999). Indeed, data from multilocus enzyme electrophoresis (Maynard Smith et al. 1993) and from multilocus sequence typing (MLST) subsequently (Maiden et al. 1998) indicate that recombination rates among conspecific isolates can be extremely high in some bacterial species (Holmes, Urwin, and Maiden 1999). For example, to date recombination has been extensively documented among the disease agents of vertebrates, such as bacteria in the genera *Neisseria* (Gibbs et al. 1989; Holmes, Urwin, and Maiden 1999; Howell-Adams and Seifert 2000), *Anaplasma* (Brayton et al. 2002; Meeus et al. 2003), *Leptospira* (Haake et al. 2004), and *Borrelia* (Cadavid et al. 1994; Rich, Sawyer, and Barbour 2001). Recombination has been found to affect both housekeeping and rapidly evolving genes (such as surface proteins). For surface proteins with related antigenic functions, recombination is responsible for the appearance of new phenotypic variants that are likely needed for adaptation and evolution in parasitic interactions. It remains unclear whether surface proteins are hot spots for localized recombination or whether their high variability makes recombination events more detectable.

Among certain intracellular bacteria that are predominantly vertically inherited, recombination is apparently absent or rare (Moran 1996; Tamas et al. 2002). This may reflect a reduced exposure to distinct gene pools and the loss of many recombinase genes (J. O. Andersson and S. G. E. Andersson 1999; Moran and Wernegreen 2000; Wernegreen 2002; Dale et al. 2003). In addition, weak selective pressures for diversification expected in relatively stable and protected intracellular environments may not promote recombination occurrence (Frank, Amiri, and Andersson 2002). Concordant with this view, recently published genomes of the maternally transmitted endosymbionts of insects *Buchnera aphidicola* (mutualist of aphids), *Wigglesworthia glossinidia brevipalpis* (mutualist of tsetse flies), and *Blochmannia* (mutualist of ants) indicate highly stable genomes and a pattern compatible with little or no recombination (Shigenobu et al. 2000; Tamas et al. 2002; Akman et al. 2002; Gil et al. 2003; Van Ham et al. 2003; Degnan, Lazarus, and Wernegreen 2005). In

Key words: homologous recombination, *Wolbachia*, symbiosis, reproductive parasitism.

E-mail: laurab@ucr.edu.

Mol. Biol. Evol. 23(2):437–449. 2006
doi:10.1093/molbev/msj049
Advance Access publication November 2, 2005

© The Author 2005. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved.
For permissions, please e-mail: journals.permissions@oxfordjournals.org

addition, the genomes of many obligate intracellular parasites preferentially show a reduction in genes with mobile DNA functions in comparison to facultative intracellular bacteria (Bordenstein and Reznikoff 2005).

Exceptions to the above pattern are found in the endosymbiont *Wolbachia*. Unlike other vertically inherited endosymbionts found in invertebrates, *Wolbachia* shows a peculiar set of features: (1) it is a generalist in host use, (2) it usually establishes parasitic interactions with the host, involving manipulations of the host's reproductive biology, (3) its genome shows an unusually high level of repetitive and mobile DNA elements, and (4) it does recombine.

Wolbachia is an alpha-proteobacterium (order Rickettsiales) with a host range that spans about 20%–75% of insect species (Werren, Windsor, and Guo 1995; Jeyaprakash and Hoy 2000; Werren and Windsor 2000), some crustaceans (isopods; Cordaux, Michel-Salzat, and Bouchon 2001), chelicerata (spiders and mites; Gotoh, Noda, and Hong 2003; Rowley, Raven, and McGraw 2004), and filarial nematodes (Bandi et al. 1998). Accounting for their pandemic infection is the ability of *Wolbachia* to shift host species (i.e., a phenomenon known as horizontal transmission), besides the vertical transmission through eggs typically found within a single-host species (Werren, Zhang, and Guo 1995). In insects, *Wolbachia* is principally a manipulator of the host reproduction. Different strains can induce a sperm-egg incompatibility, parthenogenesis, feminization of genetic males, or killing male embryos in their hosts (for reviews see Werren 1997; Stouthamer, Breeuwer, and Hurst 1999). Due to the remarkably high level of genetic divergence within *Wolbachia*, the genus is currently divided into six taxonomic groups, termed supergroups A–F (Werren, Zhang, and Guo 1995; Bandi et al. 1998; Vandekerckhove et al. 1999; Lo et al. 2002; Czarnetzki and Tebbe 2004). New supergroups have been recently added to this major subdivision, although the nomenclature is not yet completely resolved (Rowley, Raven, and McGraw 2004; Bordenstein and Rosengaus 2005; Casiraghi et al. 2005).

The published genome of *Wolbachia* from the arthropod host *Drosophila melanogaster* (*w*Mel) has shown a pattern compatible with horizontal DNA transfer and recombination (Wu et al. 2004; Foster et al. 2005). Specifically, the large fraction of prophage sequences and insertion elements (IS) suggest uptake of mobile DNA (Wu et al. 2004; Bordenstein and Reznikoff 2005). In addition, several studies in different insect species specify lateral transfer of phage WO-B between coinfecting *Wolbachia* strains (Masui et al. 2000; Bordenstein and Wernegreen 2004). Moreover, intragenic recombination appears to be common in extrachromosomal DNA, including phage and transposable elements (Bordenstein and Wernegreen 2004; Duron et al. 2005), and comparisons of *Wolbachia* genomes *w*Mel and *w*Bm (from the nematode host *Brugia malayi*) reveal numerous inversion and translocation events (Foster et al. 2005).

So far, only two examples of recombination within *Wolbachia* chromosomal genes have been documented: the surface protein *wsp* (Werren and Bartos 2001; Reuter and Keller 2003; Keller et al. 2004; Malloch and Fenton 2005; Baldo, Lo, and Werren 2005) and to a lesser extent the cell division protein *ftsZ* (Jiggins et al. 2001; Jiggins 2002). However, there is evidence of a strong diversifying selection on the surface protein gene *wsp* (Baldo, Lo, and Werren 2005), and, therefore, this gene appears not informative concerning overall patterns of recombination in standard protein coding genes of *Wolbachia*. Furthermore, because *wsp* undergoes extensive intragenic recombination, it cannot be used as a reliable reference for intergenic recombination to other genes within *Wolbachia*.

To date, no housekeeping genes besides *ftsZ* have been screened for recombination in *Wolbachia* nor have previous studies distinguished between recombination events within and between genes. The housekeeping core of the *Wolbachia* genome is still assumed to be stably inherited through vertical transmission and thus to provide reliable information for reconstructing strain relationships (Jiggins et al. 2002; Lo et al. 2002; Luchetti et al. 2004; Rasgon and Scott, 2004). However, a robust evaluation of the impact of recombination within and between *Wolbachia* genes is imperative for understanding the origins and evolution of new genotypes and avoiding false inferences regarding the similarity/divergence among strains and their ecological features. To clarify whether housekeeping genes undergo recombination, here we investigate the extent of recombination within and between four housekeeping genes of *Wolbachia* (*dnaA*, *groEL*, *ftsZ*, and *gltA*). Based on the published sequence of *Wolbachia* from *D. melanogaster* (Wu et al. 2004), these genes are located in different portions of the chromosome (fig. 1) and should give a broader picture of the role of recombination than those of prior studies using *wsp* and/or *ftsZ* (Jiggins et al. 2001; Jiggins 2002; Baldo, Lo, and Werren 2005). Results based on comparative sequence analyses show extensive recombination within *gltA* (intragenic recombination) and between the different housekeeping genes (intergenic recombination). DNA exchange has occurred both within and across *Wolbachia* lineages. As a result, remarkable phylogenetic conflict occurs across the four gene phylogenies, invalidating attempts to reconstruct strain relationships due to the chimeric nature of the genomes.

Materials and Methods

Strain Isolation and Sequencing of *dnaA*, *ftsZ*, *groEL*, and *gltA*

Strains selected for the analyses are listed in table 1. DNA was extracted from whole insects or gonads using the DNAeasy Tissue Kit (Qiagen, Germantown, Md.). Strain infectious status was confirmed using 16S rDNA and *wsp* primers specific for supergroups A and B (Werren, Zhang, and Guo 1995; Zhou, Rousset, and O'Neill 1998). Four housekeeping genes, *gltA*, *dnaA*, *groEL*, and *ftsZ*, were selected based on their conserved protein function and distribution across the *w*Mel genome (see map in fig. 1, Wu et al. 2004). An internal fragment of each gene was sequenced.

groEL

The gene codes for the heat shock protein HSP60. It has been used in early phylogenetic studies of *Wolbachia* (Masui, Sasaki, and Ishikawa 1997; Anderson and Karr

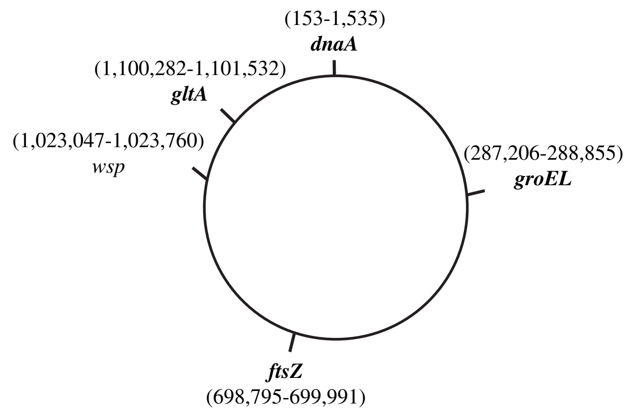

FIG. 1.—Circular map of *Wolbachia* chromosome showing the location of the four housekeeping genes, *dnaA*, *groEL*, *ftsZ*, *gltA* (in boldface), plus *wsp*. Gene order and base pair ranges refer to *w*Mel genome (Wu et al. 2004).

2001). Sequences were amplified using general primers for single-infected species: WgroF1d (5′-GGT GAG CAG TTR CAR SAA GC-3′) and WgroRev1d (5′-AGR TCT TCC ATY TTR ATT CC-3′). For double-infected species (i.e., *Nasonia longicornis* RV2 and *Nasonia giraulti* IV7, see table 1) we used the following three combinations of primers: WgroF1d/WgroBR2 [5′-AAT GCT TCA CCT TCA ACA TCT-3′] (general forward/B-specific reverse); WgroBF1 [5′-AAT TAG YAA GCC ATA TGG WGG-3′]/ WgroRev1d (B-specific forward/general reverse); and WgroBF2 [5′-CAG AGG TYA CAA AGG ATG GC-3′]/ WgroBR2 (B-specific forward/B-specific reverse). Polymerase chain reaction (PCR) cycling conditions were the following for all couples of primers: 94°C for 2 min; 38 cycles at 94°C for 30 s, 55°C for 45 s, 72°C for 1 min 30 s; 70°C for 10 min and then held at 4°C. Our final *groEL* sequence alignment corresponded to region 136–829 of the *w*Mel *groEL* gene (*w*Mel accession number NC_002978).

ftsZ

This gene codes for a protein involved in the cell division. It has been used in earlier phylogenetic (Werren, Zhang, and Guo 1995) and recombination studies (Jiggins et al. 2002) of *Wolbachia*. We amplified only single-infected species using the general primers ftsZunif/ftsZunir (Lo et al. 2002). PCR cycling conditions were as follows: 94°C for 1 min; six cycles at 94°C for 30 s, 59°C for 45 s, 72°C for 1 min 30 s; 32 cycles at 94°C for 30 s, 57°C for 45 s, 72°C for 1 min 30 s; 70°C for 10 min and then held at 4°C. B sequences from the two double-infected species (i.e., *N. giraulti* RV2 and *N. longicornis* IV7) were retrieved from GenBank. Our final *ftsZ* sequence alignment corresponded to region 331–917 of the *w*Mel *ftsZ* gene.

dnaA

The gene codes for a protein involved in the cell replication. We amplified only single-infected species using the following general primers: dnaA2F [5′-ACA ATT GGT TAT ATC AGC TG-3′]/dnaA2R [5′-TAC ATA GCT ATT TGY CTT GG-3′]. PCR cycling conditions were as follows: 94°C for 2 min; 38 cycles at 94°C for

Table 1
List and Features of the 22 *Wolbachia* Strains Analyzed

Order	Host Species Name	Strain Code[a]	Infection Type	Accession Number[b]			
				dnaA	*gltA*	*groEL*	*ftsZ*
Hymenoptera	*Nasonia giraulti*	RV2	B	AJ512660 (NY)	**DQ266527**	**DQ266414**	U28203
Hymenoptera	*Nasonia longicornis*	IV7	B	AJ512659 (Idaho)	**DQ266528**	**DQ266415**	U28204
Ortoptera	*Teleogryllus taiwanemma*		B	**DQ266516**	**DQ266529**	AB002286	**DQ266427**
Hymenoptera	*Tribolium confusum*		B	AJ512661	**AY714784**	**AY714798**	U28194
Hymenoptera	*Nasonia vitripennis*	4.9[c]	B	AJ512658 (LbII)	**AY714782**	**AY714796**	U28205 (LbII)
Diptera	*Drosophila simulans*	*w*No	B	**DQ266389**	**AY714787**	**AY714800**	**DQ266426**
Diptera	*Caudra cautella*		B	**DQ266517**	**DQ266530**	AB081646	U28207
Diptera	*Protocalliphora* sp.	0–026	B	**DQ266518**	**DQ266531**	**DQ266416**	**DQ266425**
Lepidoptera	*Encarsia formosa*		B	**DQ266519**	**AY714783**	**AY71479**	U28196
Lepidoptera	*Acraea eponina*		B	**DQ266520**	**DQ266532**	**DQ266417**	AJ271200
Diptera	*Drosophila innubila*		B	**DQ266521**	**DQ266411**	**DQ266418**	**DQ266424**
Diptera	*Protocalliphora* sp.	0–024	A1	**DQ266522**	**DQ266412**	**DQ266419**	**DQ266422**
Diptera	*Protocalliphora* sp.	0–134	A2	**DQ266523**	**DQ266413**	**DQ266420**	**DQ266423**
Hymenoptera	*Camponotus vafer*		A	**DQ266394**	**DQ266396**	**DQ266397**	**DQ266388**
Hymenoptera	*Camponotus sayi*		A	**DQ266393**	**DQ266395**	**DQ266398**	**DQ266387**
Diptera	*D. simulans*	*w*Ha	A	**DQ266391**	**AY714790**	**AY714805**	AY508998
Hymenoptera	*N. giraulti*	16.2[d]	A	**DQ266524**	**AY714793**	**AY714810**	U28182
Hymenoptera	*N. longicornis*	2.1[e]	A	**DQ266525**	**AY714794**	**AY714811**	U28187
Hymenoptera	*N. vitripennis*	12.1[f]	A	**DQ266526**	**AY714795**	**AY714812**	U28188(LbII)
Diptera	*D. simulans*	*w*Au	A	**DQ266390**	**AY714792**	**AY714807**	AY227739
Diptera	*D. melanogaster*	CS	A	AJ512656(BMH6)	NC002978(*w*Mel)	**DQ266421**	X71906
Diptera	*D. simulans*	*w*Ri	A	**DQ266392**	**AY714791**	**AY714806**	U28178

[a] Strain code refers to sequences generated in this study (accession number in boldface) and was indicated if known.
[b] Strain code (in brackets) is indicated if different from that used in this study.
[c] Segregated from A-B infection of strain ''R511''.
[d] Segregated from A-B infection of strain ''RV2''.
[e] Segregated from A-B infection of strain ''IV7''.
[f] Segregated from A-B infection of strain ''R511''.

30 s, 50°C for 45 s, 72°C for 1 min 30 s; 70°C for 10 min and then held at 4°C. B sequences from double-infected species were retrieved from GenBank. Our final *dnaA* sequence alignment corresponded to region 844–1197 of the *w*Mel *dnaA* gene.

gltA

The gene codes for the citrate synthase protein, an enzyme that catalyzes the first step in the citric acid cycle. Sequences from single-infected species were amplified with the following general primers: WgltAF1 [5′-TAC GAT CCA GGG TTT GTT TCT AC-3′]/WgltARev2 [5′-CAT TTC ATA CCA CTG GGC AA-3′]. B sequences from double-infected species were obtained using the following couple of primers: WgltAgrpBF1 [5′-GCA ATA GCA AAA GTT CCT G-3′]/WgltARev2 (i.e., B-specific forward/general reverse). PCR cycling conditions were: 94°C for 2 min; 38 cycles at 94°C for 30 s, 55°C for 45 s, 72°C for 1 min 30 s; 70°C for 10 min and then held at 4°C. Our final *gltA* sequence alignment corresponded to region 165–1152 of the *w*Mel *gltA* gene.

Sequencing was performed directly from PCR products using a BigDye v2.0 or v3.0 terminator sequencing kit and an ABI 3700 or 3730*xl* automated sequencer. Sequence accession numbers are listed in table 1.

Evolutionary Analyses of Sequences

Alignments were generated using ClustalX (Thompson et al. 1997) and modified in BioEdit vs 7.0.1 (Hall 1999). Straightforward alignments were obtained for *dnaA* (354 bp), *groEL* (694 bp), *ftsZ* (587 bp), and *gltA* (989 bp). Analyses of the genetic divergence were performed using DNAsp vs 4.10.2 (Rozas et al. 2003). The nucleotide diversity, π, was estimated using equation (10.5) of Nei (1987). The average number of nucleotide substitutions per site between populations (D_{xy}), was estimated using equation (10.20) in Nei (1987). The synonymous substitution distances (K_s) between genes were estimated with Jukes and Cantor corrections (Nei and Gojobori 1986, eq. 1–3). The statistical significance of synonymous distance between genes at single supergroups was evaluated using Mann-Whitney test based on pairwise comparisons of the mean K_s for each strain with respect to the whole supergroup strain set. For K_s distance between supergroups within the same gene, the Mann-Whitney test was also applied based on K_s values given by all pairwise divergences between data set A and B.

Phylogenetic Analyses

Maximum likelihood (ML) methods were used to infer phylogenetic relationships for each of the four genes. Prior to ML analyses, a DNA substitution model for each data set was selected using Modeltest v3.06 (Posada and Crandall 1998) and the Akaike information criterion. The following models were selected for each of the single-gene analyses: *gltA* (K81uf + I + Γ), *dnaA* (K81uf + Γ), *groEL* (K81uf + Γ), and *ftsZ* (TIM + Γ). ML heuristic searches were performed using 100 random taxon-addition replicates with tree bisection and reconnection (TBR) and branch swapping. ML bootstrap support was determined using 100 bootstrap replicates, each using 10 random taxon-addition replicates with TBR branch swapping. Searches were performed in parallel on a Beowulf cluster using a clusterpaup program (A.G. McArthur) and PAUP version 4.0b10 (Swofford 2003). The ML best trees are presented as midpoint rooted in figure 2.

We tested the significance of topological differences in phylogenetic trees of figure 2 using the Shimodaira-Hasegawa (SH) test (Shimodaira and Hasegawa 1999). The SH test compares the likelihood score (−lnL) of a given data set across its ML tree versus the −lnL of that data set across alternative topologies, which in this case are the ML phylogenies for other data sets. The differences in the −lnL values are evaluated for statistical significance using bootstrap (1,000 replicates) based on resampling estimated log-likelihood (RELL) method and the more extensive full optimization (PAUP version 4.0b10). These two approaches yielded similar results.

Detection of Recombination

Alignments were screened for evidence of recombination by using a set of four nonparametric detection programs (methods that do not assume population genetic models and/or estimate the population recombination rate): RDP, MaxChi, Chimera, and Geneconv (Posada and Crandall 2001). The programs are implemented in RDP2 program (Martin, Williamson and Posada 2005), which also provides a window interface and enables automated analyses of a nucleotide sequence alignment using all programs concomitantly (for details see Posada and Crandall, 2001; Martin and Rybicki 2000). All programs are local methods that search for regions in a set of aligned DNA sequences delimited by putative recombination breakpoints. The programs have already been used for detecting recombination in a variety of data sets (Evans et al. 2005; Tsaousis et al. 2005). MaxChi (Martin, Williamson and Posada 2005) uses a sliding-window approach to identify, for every possible sequence pair in the alignment, significant discrepancies inferred by the two partitions of the window. A chi-square value is calculated as an expression on the difference in the number of variable sites on either side of the central partition. Chimera (Martin, Williamson and Posada 2005) is a modification of MaxChi2, using triplets of sequences instead of pairs and a maximum-match chi-square statistic. Geneconv is an extension of the method of Sawyer (1989). It compares sequence pairs and gives, as a global permutation P value, the proportion of permuted alignments for which some fragment has a higher score than the observed fragment. The RDP method (Martin and Rybicki 2000) is a phylogenetic method that uses discordant branching patterns given by adjacent sequences along an alignment to infer recombination.

General recombination settings for all programs were the following: sequences were considered linear, the highest acceptable P value cut-off was set to 0.01, a Bonferroni correction was applied, consensus daughter sequences were found, breakpoints were then polished, and only recombination events detected by more than one program were listed. Each program was run on a single gene alignment using a step size of 5 nt and a window size of variable sites

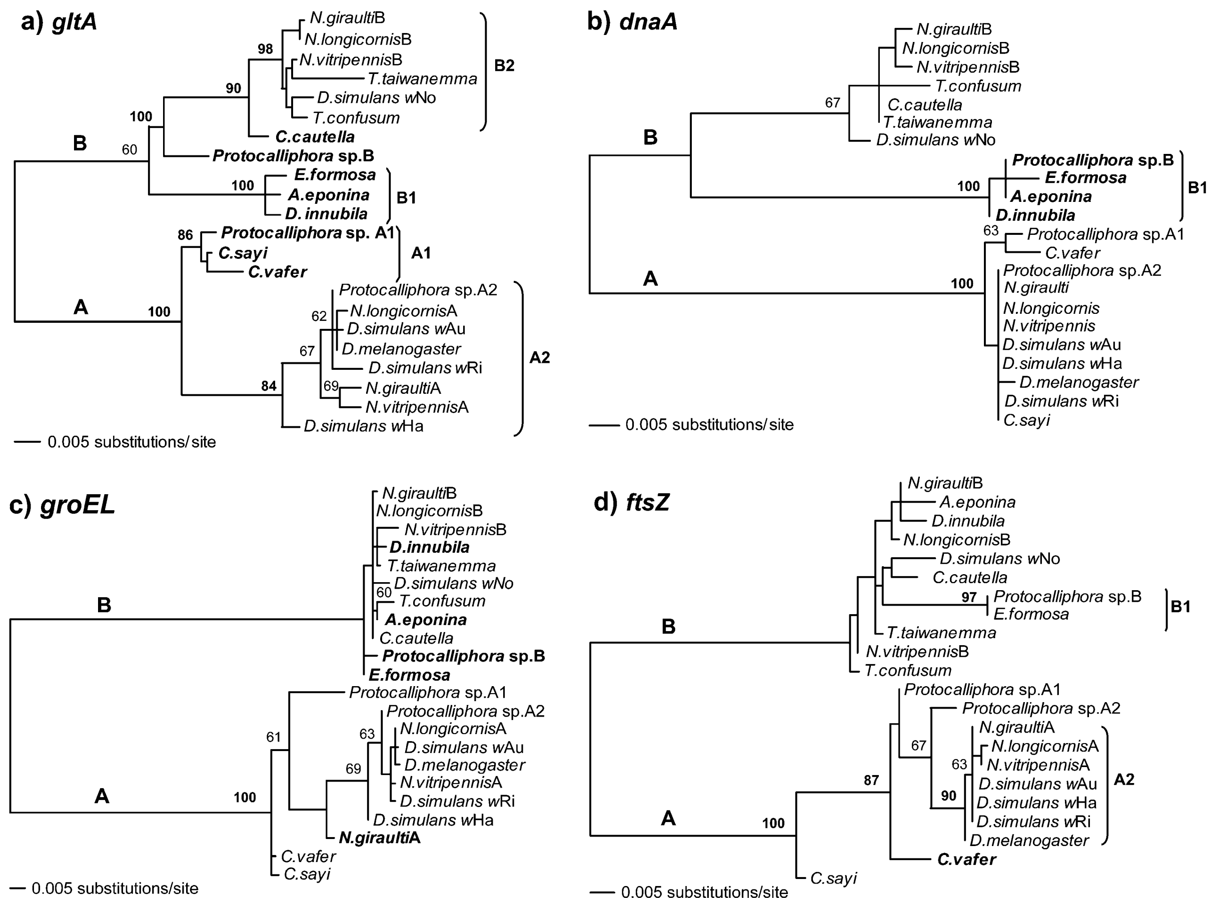

FIG. 2.—ML phylogeny of *gltA*, *dnaA*, *groEL*, and *ftsZ* (*a*–*d*) based upon the same sample of strains. Putative recombinant strains are shown in boldface. Bootstrap values lower than 60 are omitted. Clades supported by bootstrap values higher than 80 (in boldface) are labeled as A1, A2, B1, and B2.

(VI) set to different values (i.e., 10, 20, 30, and 40), in order to avoid misidentification of recombination due to differences in the intragenic variability shown by the four genes. Specific settings for each program were as follows. For RDP, no reference sequence was selected, and percentage of identity between recombinant sequences was set from 0 to 100. For MaxChi we included gaps, and 1,000 permutations were generated. For Chimera, 1,000 permutations were performed. For Geneconv we scanned sequence triplets, treating each indel as a polymorphism and setting the g-scale set to 0.

Results

We sequenced four housekeeping genes (*gltA*, *dnaA*, *groEL*, and *ftsZ*) for the same subset of 22 *Wolbachia* strains (11 strains from supergroup A and 11 from supergroup B) and explored signatures of recombination using (1) phylogenetic reconstructions, (2) statistical analyses of the genetic divergence, and (3) specific methods to detect recombination breakpoints from sequence alignments.

Phylogenetic Incongruence Among the Four Genes

An ML phylogenetic analysis was performed on each of the four genes (fig. 2*a*–2*d*). The main observation is that all four phylogenies clearly differ in both tree topology and branch lengths. While all trees clearly identify and support the existence of the two major supergroups A and B, branch lengths separating the two supergroups as well as relationships among strains within single supergroups are often highly discordant. Indeed, the likelihood-based SH test for alternative tree topologies (table 3) supports striking discordances among topologies of the four phylogenies. For example, highly significant incongruence between tree topology and data set, as well as for the reciprocal comparison (SH test, $P < 0.01$) have been detected for the following couples of genes: *dnaA* and *groEL*, *dnaA* and *ftsZ*, and *ftsZ* and *gltA*. In particular, the *gltA* data set is incompatible with tree topologies given by the other three genes. The overall picture indicates that no one gene tree topology significantly describes evolution of all the analyzed data sets. The tree topology based on the concatenated data set (illustrated in fig. 3) gives the greatest phylogenetic congruence (i.e., it is congruent with the *dnaA*, *groEL*, and *gltA* data sets and significantly discordant only with *ftsZ*, $P < 0.0001$, table 3). However, recombination bias could result in high support for the wrong tree.

All four phylogenies show remarkable conflicts within supergroup B (fig. 2). The *dnaA* phylogeny (fig. 2*b*) strongly supports a bifurcation of supergroup B into two distinct clades, B1 (P = 100) and the clade including the remaining strains. The genetic divergences between the two clades (π = 3.7%) and between B1 and A (4.2%) are in fact comparable. Similarly, the *gltA* phylogeny

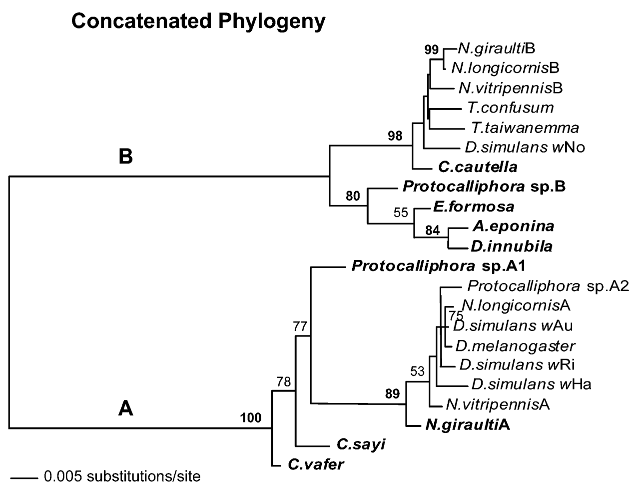

FIG. 3.—ML phylogeny based upon a concatenated alignment of the four genes. Bootstrap values of 80 or higher are in boldface. Putative recombinant strains are given in boldface (see also fig. 2 and 4).

strongly supports the existence of two major B clades, B1 (P = 100) and B2 (P = 98). However, in this case significantly high bootstrap values exclude *Protocalliphora* sp.B from clade B1 (P = 100) and *Caudra cautella* from clade B2 (P = 98), placing the two strains in an intermediate position between the two clades. *ftsZ* phylogeny shows one significant clade composed of *Protocalliphora* sp.B and *Encarsia formosa* (P = 97) which is concordant with relationships supported by *dnaA* for the two strains, but clearly in conflict with *gltA*, which places *Protocalliphora* sp.B in a distinct cluster with respect to *E. formosa*. *groEL* shows a remarkably high homogeneity among sequences within the supergroup without any strongly supported clades. This probably reflects its high conservation at synonymous level (K_s is only 2.8%). Overall within supergroup B, placement of strains *Protocalliphora* sp.B, *E. formosa*, *Acraea eponina*, *Drosophila innubila*, and *C. cautella* across the four phylogenies appear ambiguous, leading to conflicting inferences of the relationships among the four strains.

Supergroup A shows levels of topological similarity among the four genes that are greater than for supergroup B. However, clear examples of incongruence can be observed. For example, *gltA* phylogeny indicates a bipartition of the supergroup in two significant clades separated by a long genetic distance with high bootstrap confidences for clade A1 (P = 86) and clade A2 (P = 84) (see fig. 2*a*). Clades A1 and A2 are clearly in conflict with *ftsZ* phylogenetic inference where *Camponotus sayi* is excluded from clade A1 (P = 87), and *Protocalliphora* sp.A2 branches outside clade A2 (P = 90) (fig. 2*d*). Overall, within supergroup A, clear conflicts occur in the phylogenetic placement of strains *C. sayi*, *Camponotus vafer*, and *Protocalliphora* sp.A1 and A2.

Comparative Analysis of Genetic Diversity

The pattern of synonymous divergence (K_s) of supergroups A and B differs remarkably across the four genes (table 2, fig. 2). Indeed, *dnaA* and *groEL* show opposite patterns of genetic variability in the two supergroups. *dnaA* sequences are highly homogenous within supergroup A (K_s is only 0.6%) and remarkably divergent within supergroup B (16.5%), whereas *groEL* sequences are divergent within supergroup A (7.1%) and more homogeneous within supergroup B (2.8%) (see fig. 2*b* and 2*c* and table 2). Synonymous divergence within both supergroups is significantly different between the two genes (Mann-Whitney test, $P < 0.0001$). In contrast, *gltA* and *ftsZ* (fig. 2*a* and 2*d*) show highly similar patterns and values of K_s between genes: 5.9% and 6.4% for *ftsZ* and *gltA*, respectively, within group A, and 6.3% at both genes within group B (differences between genes are not significant, $P > 0.5$).

The two supergroups could show different rates of synonymous substitutions within a gene because of a possible bias in the strain sampling (e.g., due to sampling of closely related strains in one supergroup vs. more divergent strains in the other supergroup). However, in this case the pattern of variation between supergroups is expected to be similar across the four genes. Alternatively, G + C content, overlapping reading frames or local differences in mutation rate could account for the observed pattern. G + C content (table 2) is quite homogeneous across supergroups and it is unlikely responsible for the observed incongruence. Based on *w*Mel annotation (A-type genome), no overlapping reading frames are present at any of the four genes (Wu et al. 2004). However, because *Wolbachia* genomes from supergroup B have not been annotated yet (sequencing of *Wolbachia* B strain hosted in *Culex pipiens quinquefasciatus* is in progress at the Sanger Institute), a synteny between A and B-type genomes cannot be inferred at the moment.

Genetic distances separating supergroups A and B (see fig. 2 and table 2) also greatly differ across genes. Specifically, K_s between supergroups given for *gltA* (24.1%) is significantly lower than that inferred by *dnaA* and *ftsZ* (45.3% and 43.2%, respectively, Mann-Whitney test $P < 0.0001$), and it is about one-third of that estimated from *groEL* (67.4%, $P < 0.0001$). Interestingly, K_s between supergroups A and B based on *gltA* (24.1%) is even lower than the estimated K_s between clades B1 and B2 within supergroup B based on *dnaA* (30.0%) (see below).

These differences suggest distinct evolutionary histories for these genes. Such a heterogeneous pattern of genetic

Table 2
Genetic Divergence of the Four Genes

	gltA			*dnaA*			*groEL*			*ftsZ*		
	A	B	A-B	A	B	A-B	A	B	A-B	A	B	A-B
K_s (JC)[a]	6.4	6.3	24.1 ± 1.1	0.6	16.5	45.3 ± 2.7	7.1	2.8	67.4 ± 3.6	5.9	6.3	43.2 ± 3.3
G + C	0.347	0.343		0.362	0.345		0.351	0.355		0.421	0.405	

[a] Average percentage of K_s (based on Jukes and Cantor method) estimated within single data sets (A and B) and between the two data sets (A-B) for each of the four genes.

Table 3
Results of SH Test of Alternative Tree Topologies for Genes in the A and B *Wolbachia*

	Data Set				
Tree Topology	*dnaA*	*groEL*	*ftsZ*	*gltA*	Concatenated
dnaA	**787.35**	1796.47***	1525.09***	2647.99***	6987.45***
groEL	906.24***	**1689.78**	1475.50**	2698.55***	7047.96***
ftsZ	900.34***	1733.23*	**1416.69**	2716.29***	7045.00***
gltA	806.45	1730.70**	1516.54***	**2473.39**	6755.39
Concatenated	791.06	1717.82	1499.52***	2489.32	**6725.46**

NOTE.—Values denote the likelihood score (−lnL) of a given data set across its own ML tree (in boldface) as well as across each of the alternative trees. Significant differences between −lnL scores of the same data set across its own ML tree and alternative trees (i.e., the three other gene trees plus the concatenated tree) were calculated based on the full optimization criteria implemented in the SH test.

*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

variation could reflect differential rates of nucleotide substitutions either between supergroups within a single gene or within the same supergroup across different genes. However, this hypothesis would imply independent rates of evolution for each supergroup at each gene for the same subset of strains. An alternative hypothesis, also supported by a strong conflict in phylogenetic signal given by the four genes described above, is that lateral DNA transfer and recombination are responsible for a reassortment of the genetic variability in the data set, thus leading to the observed patterns of genetic divergence.

Tests for Recombination

To further support the above conclusion and clarify the pattern of recombination, we performed specific genetic analyses on single and combined data sets (A, B, and A-B) using different recombination detection programs. Because recombination can be intragenic and/or intergenic, these events were measured in separate analyses of the data sets.

Intragenic Recombination

Results given by the four recombination detection programs are only partially concordant. A previous evaluation of their performance using both simulated and empirical data (Posada and Crandall, 2001; Posada 2002) indicated that MaxChi is the most powerful, followed by Chimera, Geneconv, and RDP. We will thus utilize principally the information on recombination inferred by MaxChi. Table 4 summarizes results given by the program when single gene data sets were analyzed. Use of different window sizes gave concordant results for all programs with respect to recombination inference. We note that because the analyzed data set is unlikely to include the true donor and recipient strains of the recombination event, a possible misidentification of the "strains" involved in the recombination events described below can occur. Nevertheless, our aim was to infer significant recombination when it was present and to identify potential recombinant breakpoints and parental "sequence" types.

As reported in table 4, MaxChi infers no intragenic recombination in *dnaA* and *ftsZ*, either within supergroups A and B or between the two, and in *groEL* within B and between A and B. The same pattern is given by Chimera, Geneconv, and RDP. In contrast, two programs out of five, that is, MaxChi and Chimera, infer recombination in *groEL* within A (for both methods $P < 0.01$). Both programs indicate a single recombination event occurred at the region between nucleotide positions 3 and 111 of the alignment, involving the sequences of *N. giraulti* A (putative recombinant), *C. vafer*, and *Drosophila simulans* wHa (putative parental sequences). Indeed, removing *N. giraulti* A from the alignment and repeating the analyses failed to detect any recombination in this gene. According to this finding, the phylogeny based on *groEL* places *N. giraulti* in an intermediate position between the two putative parental sequences, discordant with the other three gene phylogenies, suggesting that this observed phylogenetic position is probably due to the recombination within the gene (fig. 2).

Analysis of Recombination in gltA. The most remarkable pattern of recombination is evident for *gltA*, where significant recombination was detected by MaxChi, as well as by Chimera and Geneconv, within both A and B, and by all programs between A and B ($P < 0.001$). To characterize recombination in *gltA*, we performed an analysis of the pattern of polymorphisms along the gene, highlighting putative breakpoints, gene regions, and strains involved in the horizontal DNA transfers. For simplicity, only major recombination events between supergroups will be discussed.

Figure 4 shows the pattern of polymorphic sites along the *gltA* alignment. Unique polymorphisms are not shown. The pattern is consistent with the occurrence of multiple

Table 4
Test for Intragenic Recombination Showing Results of MaxChi for Single Genes

Gene	Data Set[a]	MaxChi Results
gltA	A	Yes**
	B	Yes**
	A-B	Yes**
dnaA	A	—
	B	—
	A-B	—
groEL	A	Yes*
	B	—
	A-B	—
ftsZ	A	—
	B	—
	A-B	—

[a] A-B results indicate significant recombination between sequences of the two data sets.

*$P < 0.01$ and **$P < 0.001$ (where P is the highest acceptable probability value of recombination occurrence).

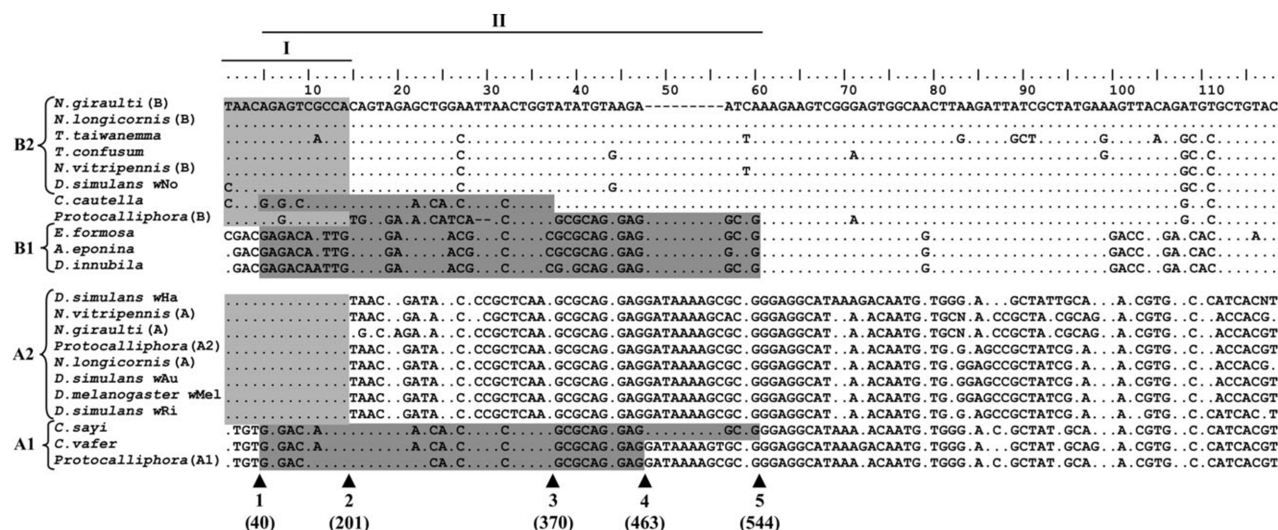

FIG. 4.—Alignment of polymorphic sites in *gltA* sequences. Unique polymorphisms were removed. Groups of sequences labeled A1, A2, B1, and B2 are based on phylogeny of figure 2*a*. I and II indicate the two major recombinant regions as detected by MaxChi. Similar color-coded nucleotides highlight sequences recombinant between supergroups A and B (see *Results*). Arrows indicate the five major breakpoints with the corresponding nucleotide position in the alignment.

recombination events. Two major recombinant regions have been detected and labeled I and II. Similar color-coded nucleotides highlight region and putative sequences involved in the single recombination event between supergroups A and B. The two regions partially overlap, suggesting that the same gene portions were exchanged in different recombinant tracts during distinct recombination events. Region I occurs between the first nucleotide and breakpoint 2 and involves the first 201 bp of the alignment. Within this 200-bp region, seven B-group sequences (*N. giraulti*, *N. longicornis*, *Teleogryllus taiwanemma*, *Tribolium confusum*, *Nasonia vitripennis*, *D. simulans* *w*No, and *Protocalliphora* sp.B) are nearly identical (only 5-bp differences) with eight A-group sequences (*D. simulans* *w*Ha, *N. vitripennis*, *N. giraulti*, *Protocalliphora* sp.A2, *N. longicornis*, *D. simulans* *w*Au, *D. melanogaster* *w*Mel, and *D. simulans* *w*Ri). Among the five polymorphisms, four are nonsynonymous. The average number of nucleotide differences per site between the two groups, D_{xy}, is only 0.35%. In contrast, the two groups show a D_{xy} of 8.75% at the nucleotide range 202–989 of *gltA*, 8.82% based on *dnaA*, 8.63% based on *ftsZ*, and 12.92% based on *groEL* (similarly, D_{xy} between supergroups A and B is about 9% based on both *ftsZ* and *dnaA*, and 12% based on *groEL*). Considering an expected mean D_{xy} of 9.78%, the two groups show a significantly much higher similarity at the first region of *gltA* (chi-square probability < 0.005). A recombination event between A and B sequences of *gltA* likely accounts for the reduced variance between supergroups at this region. An alternative explanation is that this remarkable nucleotide similarity reflects a strong pressure for conservation at the nucleotide level due, for example, to a functional role of the region as a protein-binding site. However, the region is not conserved among all sequences, and among the 21 polymorphic sites present in the divergent sequences, that is, those belonging to groups B1 and A1, 10 are nonsynonymous. Moreover, no significant secondary structures of the corresponding coded mRNA have been detected (folding was predicted using mfold Web server, Zuker 2003). Based on the *w*Mel annotation (Wu et al. 2004), the complementary strand of this region is noncoding, thus excluding a pressure for conservation at the complementary strand. All this strongly indicates that the region has been horizontally transferred between sequence representatives of the two supergroups. Region II occurs between breakpoints 1 and 5 and involves sites 40–544 of the alignment. MaxChi indicates sequence of *C. sayi* (from supergroup A) as a putative recombinant between A and B1 strains. Indeed, the strain shares a significantly high nucleotide similarity with B1 sequences at this region, while after breakpoint 5 it becomes very similar to A sequences. Region II can be visually partitioned in further segments, indicating striking discordant relationships among A and B sequences: (1) at segment between breakpoints 1 and 3, strains of clade A1 plus *Protocalliphora* sp.B and *C. cautella* and strains of clade B1 are indicated as putative recombinant; (2) at the segment between breakpoints 3 and 4, all sequences from group A and B1 plus *Protocalliphora* sp.B are nearly identical (with only one polymorphism present in *D. innubila*); and (3) at the segment between 4 and 5, *C. sayi* shares the 9-bp gap with all B sequences, and it is indicated as a putative recombinant with sequences from group B1. Noticeably, the sequence of *Protocalliphora* sp.B appears as a clear recombinant between clades B1 and B2 with a recombination breakpoint falling at site 544 (breakpoint 5 in fig. 4). The accuracy of the estimated parental versus recombinant sequences depends on the correct detection of the ancestral genotypes.

The complex pattern of recombination mentioned above clearly indicates that some sequence portions have been involved in multiple recombination events and that some of the strains have alternatively played both the recombinant and parental role in different events (i.e., as

a donor and recipient genome). This makes it extremely difficult to reconstruct the specific recombination history. Overall, the heterogeneous pattern of similarity/divergence shown by all the 22 sequences, each of them sharing regions of high nucleotide similarity alternatively with representatives of both supergroups along the gene, is consistent with extensive DNA transfer and recombination within and between supergroups of *Wolbachia gltA*.

Intergenic Recombination

We tested for intergenic recombination among the four genes by applying recombination programs to concatenated alignments of all possible gene pairs. To avoid problems related to the intragenic recombination signal in *gltA* and *groEL* we proceeded as follows: (1) in concatenated alignments including *groEL*, the recombinant strain *N. giraulti* A was removed, and (2) when *gltA* was analyzed in pair with another gene, only the alignment portion after breakpoint 5 was analyzed (see fig. 4), thus excluding the major intragenic recombination region. Data sets A and B were analyzed separately. In any case, recombination is inferred in a concatenated alignment, a breakpoint is expected to fall in the region between the two genes. Results given by MaxChi are shown in table 5.

Within data set A, two out of five programs (MaxChi and Chimera) detected intergenic recombination between *groEL* and *ftsZ*. *Camponotus vafer* is indicated as a putative recombinant between *Protocalliphora* sp.A1 and *C. sayi*. Within data set B, all programs infer intergenic recombination between *gltA* and *dnaA*, indicating *Protocalliphora* sp.B as recombinant. Between *dnaA* and *groEL*, three out of five programs infer significant recombination, indicating *Protocalliphora* sp.B, *E. formosa*, *A. eponina*, and *D. innubila* are recombinants. As expected, in all cases each program indicates one breakpoint falling in the region between the two genes, with the recombinant unit encompassing one of the two gene alignments. Following removal of the recombinant sequences from the above concatenated alignments of gene pairs, the programs failed to infer recombination.

Overall, according to the phylogenetic discordances described above, within supergroup A, strains belonging to clade A1 are involved in recombination between *groEL* and *ftsZ*. Within supergroup B, strains belonging to clade B1 plus *Protocalliphora* sp.B have undergone recombination in the genomic region between *dnaA* and *groEL* (see fig. 1). In both cases, the second recombination breakpoint likely falls somewhere in the chromosome between *ftsZ* and *gltA* genes. Current data cannot distinguish between a single event involving the two large regions and several recombination events affecting multiple regions. However, extensive intragenic recombination in *gltA* suggests a pattern consistent with a highly mosaic genome structure of the analyzed strains.

Table 5
Test for Intergenic Recombination Showing Results of MaxChi Run on Combined Data Sets of Each Pair of Genes

Data Set	Combined Genes[a]	MaxChi Results	Putative Recombinant Strains[b]
A	*gltA-dnaA*	—	
	dnaA-groEL	—	
	groEL-ftsZ	Yes**	*Camponotus vafer*
B	*gltA-dnaA*	Yes**	*Protocalliphora* sp.B
	dnaA-groEL	Yes*	*Protocalliphora* sp.B *D. innubila*, *E. formosa* *A. eponina*
	groEL-ftsZ	—	

[a] The pairs of genes respects their spatial order on the chromosome based on *w*Mel (see fig. 1).

[b] Putative recombinants are based on MaxChi results.

* $P < 0.01$ and ** $P < 0.001$ (where P is the highest acceptable probability value of recombination occurrence).

Discussion

Despite increasing studies on recombination in *Wolbachia* (e.g., within *wsp* and *ftsZ* genes, phages, and IS elements), it remained unclear whether *Wolbachia* housekeeping genes typically recombine, and if so, to what extent. Implications are large: strain similarities and phylogenetic relationships, as well as all derived inferences on *Wolbachia* genome adaptation and evolution, often rely on this gene category. To clarify this major issue, here we tested for intragenic and intergenic recombination in *Wolbachia* of supergroups A and B using four common housekeeping genes (*gltA*, *dnaA*, *ftsZ*, and *groEL*) and 22 randomly selected strains. A variety of analyses, based on phylogenetic reconstructions, rates of variation within and among genes, and statistical methods for detecting recombination are all concordant. Overall, they indicate that intergenic recombination occurred between all pairs of genes in at least one of the two supergroups. Intragenic recombination likely occurred within supergroup A in *groEL* and, to a major extent, within and between supergroups in *gltA*. While intragenic recombination was previously demonstrated at *ftsZ* (Jiggins 2002), our analyses did not show any evidence for recombination among *ftsZ* alleles. Differences between our and previous results could be related to the different sample of strains analyzed or different method for detecting recombination. Applying the same method used by Jiggins (2002) for *ftsZ*, that is, LDhat (McVean 2001), to the four data sets used here, we found patterns consistent with two other programs that we used (MaxChi, and Chimera). The only exception is for *ftsZ*, where recombination was detected in the A-B data set contrary to the results of nonparametric programs (results not shown).

As expected for recombining loci, phylogenetic reconstructions varied among the genes sampled. No single pair of gene tree topologies is concordant due to intergenic and/or intragenic recombination events. As a result, no tree topology among the four shown in figure 2 can be considered representative of "strain relationships." Overall, we expect that when intergenic recombination events occur (i.e., whole gene sequences are recombined) sets of mutations are moved onto different background genomes and, consequently, phylogenetic homoplasies among bacterial strains are simulated. A likely effect is a reduction of the expected gene variance, which makes gene sequences falsely appearing more homogeneous, and can lead to star-like trees. This could explain, for example, the greater similarity among B sequences at synonymous sites of *groEL* compared to their

divergence at the other three loci. Indeed, intergenic recombination has been detected between *dnaA* and *groEL* supergroup B. In contrast, by creating mosaic gene sequences, intragenic recombination breaks apart potential phyletic clusters and so increases the number of divergent lineages in a tree. For example, within *gltA*, intragenic recombination between supergroups (involving *Protocalliphora* sp.B and *C. cautella*) results in placement of these two strains into two divergent lineages, while they group together at *dnaA* (see fig. 2*a–b*). Interestingly, the recombination between supergroups does not seem to change supergroup assignation of the recombinant strains (i.e., in the case of the recombinant strains belonging to clusters B1 and A1 in fig. 2*a*), either because the phylogenetic contribution of the recombinant portion is not relevant or the recombination signal has been obscured by mutations. However, the entire branching pattern of the supergroup phylogeny could be seriously misleading. Overall, inferring a phylogeny from a data set affected by recombination can lead to (1) misinterpretation of phylogenetic relationships, (2) incorrect inference of the mutation rate across sequences, and (3) false distribution of the branch lengths in a tree (Martin 1999).

This study also highlights the high potential of recombination in shaping genetic diversity. This is apparent in the highly heterogeneous patterns of genetic divergence at the four genes. The discordance in the synonymous divergence within and between supergroups indicated by the four genes appears to be related, at least in part, to recombination events. Considering the relevant genetic divergence existing between supergroups (ranging from 43.3% to 67.2% based on *ftsZ*, *dnaA*, and *groEL*), sequence transfer from one supergroup into another, through intragenic shuffling, is expected to (1) increase the genetic diversity within the recipient supergroup and to (2) reduce the estimated divergence between supergroups. As a result, evolutionary distances between strains and supergroups can be mistaken. In principle, differential mutation rates for the two supergroups at the same gene could also account for part of the observed discrepancies; indeed, the relative contributions of mutation and recombination cannot be quantified with confidence. Whether recombination has a role in affecting rates of substitution has still to be clarified.

Previous evidence for recombination at *ftsZ*, *wsp*, phages, and IS elements (Jiggins et al. 2001; Jiggins 2002; Bordenstein and Wernegreen 2004; Baldo, Lo, and Werren 2005; Duron et al. 2005) coupled with our findings of recombination within and between housekeeping genes illustrate high levels of genome flux in *Wolbachia*. The physical distance among the genes sampled here (fig. 1) argues against a single hot spot for recombination. Rather, homologous recombination in *Wolbachia* appears widespread relative to gene function and location. Moreover, the inferred intergenic recombination suggests transfer events that involve either large DNA portions, encompassing one or several genes, or several small sequence tracts across the genome. Both scenarios clearly imply that several A and B *Wolbachia* strains are likely to be chimeras, raising concern about the reliability of current phylogenetic reconstructions as well as the very existence of genetically cohesive *Wolbachia* strains. If present, a conserved "core" of genes not affected by recombination would be ideal for developing strain phylogenies. Its possible characterization represents a future challenge for study of *Wolbachia* genome diversity and evolution, together with the development of a MLST system.

Outside of *Wolbachia*, no cases of extensive recombination have been reported for vertically inherited endosymbiotic bacteria. Given the unique features of *Wolbachia* among strict endosymbionts of invertebrates (e.g., being a parasite and a generalist in host use), we expect that *Wolbachia* "sexuality" could be crucial for survival and the adaptation of these bacteria to their arthropod hosts. Specifically, we expect that recombinant alleles that provide some advantages to the strain can be rapidly fixed in a population and subsequently spread through horizontal transmission of *Wolbachia* to new hosts. The existence of a common gene pool for *Wolbachia* could also act as a reinforcement of the clade genetic identity and be partly responsible for the high divergence of *Wolbachia* from the currently known outgroups (Wu et al. 2004).

The biology of *Wolbachia* involves intimate associations with host cells and transmission via host cytoplasm. However, unlike the obligate mutualists mentioned above, *Wolbachia* also show extensive horizontal transfer between often highly different host species (e.g., insects in different orders) and dramatic changes in phenotypic effects on hosts, such as cytoplasmic incompatibility, parthenogenesis induction, and male killing (Werren 1997). Recombination may play an important role in phenotypic and host liability of these bacteria, although this has not yet been demonstrated. Indeed, so far little is known about specific interactions (i.e., genes and cellular pathways) between *Wolbachia* and their hosts, even if insights into the topic are increasing (e.g., Iturbe-Ormaetxe et al. 2005; Sinkins et al. 2005). We posit that recombinant genes are involved in bacterial-host interactions and therefore show both the signatures of recombination and selection. Future work will reveal to what extent this is true.

Acknowledgments

This work was supported by the grants from National Science Foundation (EF-0328363) to J.H.W. and grants from the National Institutes of Health (R01 GM62626-01), National Science Foundation (DEB 0089455), and the NASA Astrobiology Institute (NNA04CC04A) to J.J.W. This work was performed while S.B. held a National Research Council Research Associates Award at the Marine Biological Laboratory.

Literature Cited

Akman, L., A. Yamashita, H. Watanabe, K. Oshima, T. Shiba, M. Hattori, and S. Aksoy. 2002. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. Nat. Genet. **32**:402–407.

Anderson, C. L., and T. L. Karr. 2001. *Wolbachia*: evolutionary novelty in a rickettsial bacteria. Evol. Biol. **1**:10.

Andersson, J. O., and S. G. E. Andersson. 1999. Insights into the evolutionary process of genome degradation. Curr. Opin. Genet. Dev. **9**:664–671.

Baldo, L., N. Lo, and J. H. Werren. 2005. Mosaic nature of *wsp* (*Wolbachia* surface protein). J. Bacteriol. **187**:5406–5418.

Bandi, C., C. G. Anderson, C. Genchi, and M. L. Blaxter. 1998. Phylogeny of *Wolbachia* in filarial nematodes. Proc. R. Soc. Lond. B Biol. Sci. **265**:2407–2413.

Bordenstein, S. R., and W. R. Reznikoff. 2005. Mobile DNA in obligate intracellular bacteria. Nat. Rev. Microbiol. **3**:688–699.

Bordenstein, S. R., and R. B. Rosengaus. 2005. Discovery of a novel *Wolbachia* supergroup in Isoptera. Curr. Microbiol. (in press).

Bordenstein, S. R., and J. J. Wernegreen. 2004. Bacteriophage flux in endosymbionts (*Wolbachia*): infection frequency, lateral transfer, and recombination rates. Mol. Biol. Evol. **21**:1981–1991.

Boucher, Y., C. J. Douady, R. T. Papke, D. A. Walsh, M. E. Boudreau, C. L. Nesbo, R. J. Case, and W. F. Doolittle. 2003. Lateral gene transfer and the origins of prokaryotic groups. Annu. Rev. Genet. **37**:283–328.

Brayton, K. A., G. H. Palmer, A. Lundgren, J. Yi, and A. F. Barbet. 2002. Antigenic variation of *Anaplasma marginale msp2* occurs by combinatorial gene conversion. Mol. Microbiol. **43**:1151–1159.

Cadavid, D., D. D. Thomas, R. Crawley, and A. G. Barbour. 1994. Variability of a bacterial surface protein and disease expression in a possible mouse model of systemic *Lyme borreliosis*. J. Exp. Med. **179**:631–642.

Casiraghi, M., S. R. Bordenstein, L. Baldo, N. Lo, T. Beninati, J. J. Wernegreen, J. H. Werren, and C. Bandi. 2005. Phylogeny of *Wolbachia* based on *gltA*, *groEL* and *ftsZ* gene sequences: clustering of arthropod and nematode symbionts in the F supergroup and evidence for further diversity in the *Wolbachia* tree. Microbiology (in press).

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

Czarnetzki, A. B., and C. C. Tebbe. 2004. Detection and phylogenetic analysis of *Wolbachia* in Collembola. Environ. Microbiol. **6**:35–44.

Dale, C., B. Wang, N. Moran, and H. Ochman. 2003. Loss of DNA recombinational repair enzymes in the initial stages of genome degeneration. Mol. Biol. Evol. **20**:1188–1194.

Degnan, P. H., A. B. Lazarus, and J. J. Wernegreen. 2005. Genome sequence of *Blochmannia pennsylvanicus* indicates parallel evolutionary trends among bacterial mutualists of insects. Genome Res. **15**:1023–1033.

Doolittle, W. F. 1999. Lateral genomics. Trends Genet. **15**: M5–M8.

Duron, O., J. Lagnel, M. Raymond, K. Bourtzis, P. Fort, and M. Weill. 2005. Transposable element polymorphism of *Wolbachia* in the mosquito *Culex pipiens*: evidence of genetic diversity, superinfection and recombination. Mol. Ecol. **14**:1561–1573.

Evans, B. J., D. B. Kelley, D. J. Melnick, and D. C. Cannatella. 2005. Evolution of RAG-1 in polyploid clawed frogs. Mol. Biol. Evol. **22**:1193–1207.

Feavers, I. M., A. B. Heath, J. A. Bygraves, and M. C. Maiden. 1992. Role of horizontal genetic exchange in the antigenic variation of the class 1 outer membrane protein of *Neisseria meningitidis*. Mol. Microbiol. **6**:489–495.

Feil, E. J., and B. G. Spratt. 2001. Recombination and the population structures of bacterial pathogens. Annu. Rev. Microbiol. **55**:561–590.

Foster, J., M. Ganatra, Kamal, I. et al. (26 co-authors). 2005. The Wolbachia genome of Brugia malayi: endosymbiont evolution within a human pathogenic nematode. PLoS Biol. **3**:e121.

Frank, A. C., H. Amiri, and S. G. Andersson. 2002. Genome deterioration: loss of repeated sequences and accumulation of junk DNA. Genetica **115**:1–12.

Gibbs, C. P., B. Y. Reimann, E. Schultz, A. Kaufmann, R. Haas, and T. F. Meyer. 1989. Reassortment of pilin genes in *Neisseria gonorrhoeae* occurs by two distinct mechanisms. Nature **338**:651–652.

Gil, R., F. J. Silva, E. Zientz et al. (13 co-authors). 2003. The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. Proc. Natl. Acad. Sci. USA **100**:9388–9393.

Gotoh, T., H. Noda, and X. Y. Hong. 2003. *Wolbachia* distribution and cytoplasmic incompatibility based on a survey of 42 spider mite species (Acari: Tetranychidae) in Japan. Heredity **91**:208–216.

Groisman, E. A., and H. Ochman. 1996. Pathogenicity islands: bacterial evolution in quantum leaps. Cell **87**:791–794.

Haake, D. A., M. A. Suchard, M. M. Kelley, M. Dundoo, D. P. Alt, and R. L. Zuerner. 2004. Molecular evolution and mosaicism of leptospiral outer membrane proteins involves horizontal DNA transfer. J. Bacteriol. **186**:2818–2828.

Hacker, J., G. Blum-Oehler, I. Muhldorfer, and H. Tschape. 1997. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. Mol. Microbiol. **23**:1089–1097.

Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. **41**:95–98.

Holmes, E. C., R. Urwin, and M. C. Maiden. 1999. The influence of recombination on the population structure and evolution of the human pathogen *Neisseria meningitidis*. Mol. Biol. Evol. **16**:741–749.

Howell-Adams, B., and H. S. Seifert. 2000. Molecular models accounting for the gene conversion reactions mediating gonococcal pilin antigenic variation. Mol. Microbiol. **37**: 1146–1158.

Iturbe-Ormaetxe, I., G. R. Burke, M. Riegler, and S. L. O'Neill. 2005. Distribution, expression, and motif variability of ankyrin domain genes in *Wolbachia pipientis*. J. Bacteriol. **187**:5136–5145.

Jain, R., M. C. Rivera, J. E. Moore, and J. A. Lake. 2002. Horizontal gene transfer in microbial genome evolution. Theor. Popul. Biol. **61**:489–495.

Jeyaprakash, A., and M. A. Hoy. 2000. Long PCR improves *Wolbachia* DNA amplification: *wsp* sequences found in 76% of sixty-three arthropod species. Insect Mol. Biol. **9**: 393–405.

Jiggins, F. M. 2002. The rate of recombination in *Wolbachia* bacteria. Mol. Biol. Evol. **19**:1640–1643.

Jiggins, F. M., J. K. Bentley, M. E. Majerus, and G. D. Hurst. 2002. Recent changes in phenotype and patterns of host specialization in *Wolbachia* bacteria. Mol. Ecol. **11**:1275–1283.

Jiggins, F. M., J. H. von Der Schulenburg, G. D. Hurst, and M. E. Majerus. 2001. Recombination confounds interpretations of *Wolbachia* evolution. Proc. R. Soc. Lond. B Biol. Sci. **268**:1423–1427.

Keller, G. P., D. M. Windsor, J. M. Saucedo, and J. H. Werren. 2004. Reproductive effects and geographical distributions of two *Wolbachia* strains infecting the Neotropical beetle, *Chelymorpha alternans* Boh. (Chrysomelidae, Cassidinae). Mol. Ecol. **13**:2405–2420.

Koonin, E. V., K. S. Makarova, and L. Aravind. 2001. Horizontal gene transfer in prokaryotes: quantification and classification. Annu. Rev. Microbiol. **55**:709–742.

Lawrence, J. G., and H. Hendrickson. 2003. Lateral gene transfer: when will adolescence end? Mol. Microbiol. **50**:739–749.

Lesic, B., and E. Carniel. 2005. Horizontal transfer of the high-pathogenicity island of *Yersinia pseudotuberculosis*. J. Bacteriol. **187**:3352–3358.

Lo, N., M. Casiraghi, E. Salati, C. Bazzocchi, and C. Bandi. 2002. How many *Wolbachia* supergroups exist? Mol. Biol. Evol. **19**:341–346.

Luchetti, A., B. Mantovani, M. L. Fioravanti, and M. Trentini. 2004. *Wolbachia* infection in the newly described

Ecuadorian sand flea, *Tunga trimamillata*. Exp. Parasitol. **108**:18–23.

Maiden, M. C. 1993. Population genetics of a transformable bacterium: the influence of horizontal genetic exchange on the biology of *Neisseria meningitidis*. FEMS Microbiol. Lett. **112**:243–250.

Maiden, M. C., J. A. Bygraves, E. Feil et al. (13 co-authors). 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. USA **95**:3140–3145.

Majewski, J., and F. M. Cohan. 1999. DNA sequence similarity requirements for interspecific recombination in Bacillus. Genetics **153**:1525–1533.

Malloch, G., and B. Fenton. 2005. Super-infections of *Wolbachia* in byturid beetles and evidence for genetic transfer between A and B super-groups of *Wolbachia*. Mol. Ecol. **14**:627–637.

Martin, D., and E. Rybicki. 2000. RDP: detection of recombination amongst aligned sequences. Bioinformatics **16**: 562–563.

Martin, D., C. Williamson, and D. Posada. 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics **21**:260–262.

Martin, W. 1999. Mosaic bacterial chromosomes: a challenge en route to a tree of genomes. Bioessays **21**:99–104.

Masui, S., S. Kamoda, T. Sasaki, and H. Ishikawa. 2000. Distribution and evolution of bacteriophage WO in *Wolbachia*, the endosymbiont causing sexual alterations in arthropods. J. Mol. Evol. **51**:491–497.

Masui, S., T. Sasaki, and H. Ishikawa. 1997. *groE*-homologous operon of *Wolbachia*, an intracellular symbiont of arthropods: a new approach for their phylogeny. Zool. Sci. **14**: 701–706.

Maynard Smith, J., C. G. Dowson, and B. G. Spratt. 1991. Localized sex in bacteria. Nature **349**:29–31.

Maynard Smith, J., N. H. Smith, M. O'Rourke, and B. G. Spratt. 1993. How clonal are bacteria? Proc. Natl. Acad. Sci. USA **90**:4384–4388.

McVean, G. A. T. 2001. LDhat. Department of Statistics, University of Oxford, Oxford, United Kingdom.

Medigue, C., T. Rouxel, P. Vigier, A. Henaut, and A. Danchin. 1991. Evidence for horizontal gene transfer in *Escherichia coli* speciation. J. Mol. Biol. **222**:851–856.

Meeus, P. F. M., K. A. Brayton, G. H. Palmer, and A. F. Barbet. 2003. Conservation of a gene conversion mechanism in two distantly related paralogues of *Anaplasma marginale*. Mol. Microbiol. **47**:633–643.

Moran, N. A. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. USA **93**: 2873–2878.

Moran, N. A., and J. J. Wernegreen. 2000. Lifestyle evolution in symbiotic bacteria: insights from genomics. Trends Ecol. Evol. **15**:321–326.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.

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

Ochman, H., E. Lerat, and V. Daubin. 2005. Examining bacterial species under the specter of gene transfer and exchange. Proc. Natl. Acad. Sci. USA **1**:6595–6599.

Patil, P. B., and R. V. Sonti. 2004. Variation suggestive of horizontal gene transfer at a lipopolysaccharide (lps) biosynthetic locus in *Xanthomonas oryzae* pv. oryzae, the bacterial leaf blight pathogen of rice. BMC Microbiol. **4**:40.

Posada, D. 2002. Evaluation of methods for detecting recombination from DNA sequences: empirical data. Mol. Biol. Evol. **19**:708–717.

Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics **14**:817–818.

———. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc. Natl. Acad. Sci. USA **98**:13757–13762.

Rasgon, J. L., and T. W. Scott. 2004. Phylogenetic characterization of *Wolbachia* symbionts infecting *Cimex lectularius* L. and *Oeciacus vicarius* Horvath (Hemiptera: Cimicidae). J. Med. Entomol. **41**:1175–1178.

Reuter, M., and Keller, L. 2003. High levels of multiple *Wolbachia* infection and recombination in the ant *Formica exsecta*. Mol. Biol. Evol. **20**:748–753.

Rich, S. M., S. A. Sawyer, and A. G. Barbour. 2001. Antigen polymorphism in *Borrelia hermsii*, a clonal pathogenic bacterium. Proc. Natl. Acad. Sci. USA **98**:15038–15043.

Rowley, S. M., R. J. Raven, and E. A. McGraw. 2004. *Wolbachia pipientis* in Australian spiders. Curr. Microbiol. **49**:208–214.

Rozas, J., J. C. Sanchez-DelBarrio, X. Messeguer, R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19**:2496–2497.

Sawyer, S. 1989. Statistical tests for detecting gene conversion. Mol. Biol. Evol. **6**:526–538.

Shigenobu, S., H. Watanabe, M. Hattori, Y. Sakaki, and H. Ishikawa. 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. Nature **407**:81–86.

Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference: Mol. Biol. Evol. **16**:1114–1116.

Sinkins, S. P., T. Walker, A. R. Lynd, A. R. Steven, B. L. Makepeace, H. C. Godfray, and J. Parkhill. 2005. *Wolbachia* variability and host effects on crossing type in Culex mosquitoes. Nature **436**:257–260.

Spratt, B. G., W. P. Hanage, and E. J. Feil. 2001. The relative contributions of recombination and point mutation to the diversification of bacterial clones. Curr. Opin. Microbiol. **4**:602–606.

Stouthamer, R., J. A. J. Breeuwer, and G. D. Hurst. 1999. *Wolbachia pipientis*: microbial manipulator of arthropod reproduction. Annu. Rev. Microbiol. **53**:71–102.

Swofford, D. L. 2003. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0b 10. Sinauer Associates, Sunderland, Mass.

Tamas, I., L. Klasson, B. Canback, A. K. Naslund, A. S. Eriksson, J. J. Wernegreen, J. P. Sandstrom, N. A. Moran, and S. G. Andersson. 2002. 50 million years of genomic stasis in endosymbiotic bacteria. Science **296**:2376–2379.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **24**:4876–4882.

Tsaousis, A. D., D. P. Martin, E. D. Ladoukakis, D. Posada, and E. Zouros. 2005. Widespread recombination in published animal mtDNA sequences. Mol. Biol. Evol. **22**:925–933.

Vandekerckhove, T. M. T., S. Watteyne, S. Willems, J. G. Swings, J. Mertens, and M. Gillis. 1999. Phylogenetic analysis of the 16S rDNA of the cytoplasmic bacterium *Wolbachia* from the novel host *Folsomia candida* (Hexapoda, Collembola) and its implications for the *Wolbachia* taxonomy. FEMS Microbiol. Lett. **180**:279–286.

Van Ham, R. C., J. Kamerbeek, C. Palacios et al. (16 co-authors). 2003. Reductive genome evolution in *Buchnera aphidicola*. Proc. Natl. Acad. Sci. USA **100**:581–586.

Wernegreen, J. J. 2002. Genome evolution in bacterial endosymbionts of insects. Nat. Rev. Genet. **3**:850–861.

Werren, J. H. 1997. Biology of *Wolbachia*. Annu. Rev. Entomol. **42**:587–609.

Werren, J. H., and J. D. Bartos. 2001. Recombination in *Wolbachia*. Curr. Biol. **11**:431–435.

Werren, J. H., D. Windsor, and L. R. Guo. 1995. Distribution of *Wolbachia* among neotropical arthropods. Proc. R. Soc. Lond. B Biol. Sci. **262**:197–204.

Werren, J. H., and D. M. Windsor. 2000. *Wolbachia* infection frequencies in insects: evidence of a global equilibrium? Proc. R. Soc. Lond. B Biol. Sci. **267**:1277–1285.

Werren, J. H., W. Zhang, and L. R. Guo. 1995. Evolution and phylogeny of *Wolbachia*: reproductive parasites of arthropods. Proc. R. Soc. Lond. B Biol. Sci. **261**:55–63.

Wu, M., L. V. Sun, J. Vamathevan et al. (30 co-authors). 2004. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. PLoS Biol. **2**:E69.

Zhou, W., F. Rousset, and S. L. O'Neill. 1998. Phylogeny and PCR-based classification of *Wolbachia strains* using *wsp* gene sequences. Proc. R. Soc. Lond. B Biol. Sci. **265**: 509–515.

Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. **31**:3406–3415.

William Martin, Associate Editor

Accepted October 19, 2005