Multilocus Sequence Typing System for the Endosymbiont *Wolbachia pipientis*

Laura Baldo,1* Julie C. Dunning Hotopp,2 Keith A. Jolley,3 Seth R. Bordenstein,4 Sarah A. Biber,4 Rhitoban Ray Choudhury,5 Cheryl Hayashi,1 Martin C. J. Maiden,3 Hervé Tettelin,2 and John H. Werren5

Department of Biology, University of California, Riverside, California1; The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland2; The Peter Medawar Building for Pathogen Research, University of Oxford, Oxford, United Kingdom3; Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, The Marine Biological Laboratory, Woods Hole, Massachusetts4; and Department of Biology, University of Rochester, Rochester, New York5

Received 29 March 2006/Accepted 13 August 2006

The eubacterial genus *Wolbachia* comprises one of the most abundant groups of obligate intracellular bacteria, and it has a host range that spans the phyla Arthropoda and Nematoda. Here we developed a multilocus sequence typing (MLST) scheme as a universal genotyping tool for *Wolbachia*. Internal fragments of five ubiquitous genes (gatB, cox4, hepA, fbpA, and ftsZ) were chosen, and primers that amplified across the major *Wolbachia* supergroups found in arthropods, as well as other divergent lineages, were designed. A supplemental typing system using the hypervariable regions of the *Wolbachia* surface protein (WSP) was also developed. Thirty-seven strains belonging to supergroups A, B, D, and F obtained from singly infected hosts were characterized by using MLST and WSP. The number of alleles per MLST locus ranged from 25 to 31, and the average levels of genetic diversity among alleles were 6.5% to 9.2%. A total of 35 unique allelic profiles were found. The results confirmed that there is a high level of recombination in chromosomal genes. MLST was shown to be effective for detecting diversity among strains within a single host species, as well as for identifying closely related strains found in different arthropod hosts. Identical or similar allelic profiles were obtained for strains harbored by different insect species and causing distinct reproductive phenotypes. Strains with similar WSP sequences can have very different MLST allelic profiles and vice versa, indicating the importance of the MLST approach for strain identification. The MLST system provides a universal and unambiguous tool for strain typing, population genetics, and molecular evolutionary studies. The central database for storing and organizing *Wolbachia* bacterial and host information can be accessed at http://pubmlst.org/wolbachia/.

In the past 10 years there has been increasing interest in the maternally inherited *Wolbachia* endosymbionts because of their remarkably widespread distribution and significant impact on the ecology, evolution, and reproductive biology of their host species (46, 54). Approximately 20 to 75% of all insect species harbor *Wolbachia* (20, 55), as do many arachnids and terrestrial crustaceans (7, 11, 12, 40). Individual insects can be infected with multiple *Wolbachia* strains (24, 55, 58), and geographically distinct populations of the same species can harbor different strains (29, 39). Outside the phylum Arthropoda, high infection levels have also been detected in the vast majority of pathogenic filarial nematodes (4). Overall, the extraordinary infection frequency among insects alone places members of the genus *Wolbachia* among the most widespread intracellular bacteria described thus far (55, 56). *Wolbachia* strains are typically vertically transmitted within a species through the cytoplasm of eggs (maternal inheritance). However, the wide range of infected hosts cannot be explained by vertical transmission alone, and the ability of these bacteria to spread to new host species via horizontal transfer accounts for the pandemic distribution, although the mechanisms and patterns of interspecies transfer are not well understood (59).

The importance of *Wolbachia* strains ranges from their effects on the reproductive biology, ecology, and evolution of their hosts to their potential use in biological control of pest insects and biomedical applications (35, 54, 61). *Wolbachia* strains can extensively manipulate host reproduction by means such as inducing parthenogenesis, feminizing genetic males, killing male embryos, or causing cytoplasmic incompatibility between gametes (16, 45, 46). In filarial nematodes, the host dependence on *Wolbachia* for propagation has generated great interest in the medical field for targeting these bacteria in the treatment of filariasis (10, 38).

*Wolbachia* strains were first found in the mosquito species *Culex pipiens* and designated *Wolbachia pipientis* (15). These bacteria were subsequently found to be widely distributed in arthropods. However, additional species have not been named, pending improved understanding of the genetic variation of this genus (6, 26). At present, supergroup designations are routinely used to describe the major phylogenetic subdivisions of this bacterial group. So far, eight supergroups have been designated (supergroups A to H) primarily on the basis of data from the 16S rRNA, ftsZ, and wsp (*Wolbachia* surface protein [WSP]) genes (31, 59, 62). Most of the supergroups are found in arthropods (supergroups A, B, E, F, G, and H), and the majority of insect *Wolbachia* strains belong to supergroups A and B. Fully annotated genomes of the following two *Wolbachia* strains are available: wMel, a supergroup A strain from the arthropod host *Drosophila mela-

---

1 Corresponding author. Mailing address: Department of Biology, University of California, 900 University Avenue, Riverside, CA 92521. Phone: (951) 827-3841. Fax: (951) 827-4286. E-mail: laurab@ucr.edu.
2 Published ahead of print on 25 August 2006.
nogaster (60), and wBm, a supergroup D strain from the nematode host Brugia malayi (9). Despite increasing studies of Wolbachia and a growing list of infected arthropod species, researchers still lack a standard and reliable system for typing and quantifying strain diversity. A strain typing system was previously developed using the Wolbachia surface protein gene, wsp (62). However, evidence of extensive recombination (2, 57) and strong diversifying selection (1, 21) at this gene make it unsuitable for reliable strain characterization if it is used alone. Furthermore, recent detection of recombination both within and between Wolbachia genes (3) suggests that a single-locus approach to strain characterization may be misleading. A multigene typing approach has recently been used to type Wolbachia strains (32). This study focused on strains found in Drosophila hosts and demonstrated the utility of multigene typing for Wolbachia strains.

Here we developed a standard multilocus sequence typing (MLST) system for Wolbachia, which works broadly with Wolbachia strains found in diverse singly infected host species. This system includes a web-accessible central database and toolkit for storage, organization, and analysis of data. The data are fully portable between laboratories and allow extensive comparative analyses (22). This accurate strain typing system should provide the genetic framework for tracing the movement of Wolbachia globally and within insect communities and for associating Wolbachia strains with geographic regions, host features (e.g., ecology and phylogeny), and phenotypic effects on hosts.

Originally used for global epidemiology and surveillance of recombinant pathogenic bacteria (27), the MLST approach defines a strain as a sequence type (ST) (equivalent to a haplotype) on the basis of a combination of alleles (i.e., allelic profile) at a sample of housekeeping genes. Similarity relationships among strains are described using the nucleotide sequence identity of alleles at each locus. The use of combinations of alleles as molecular markers to genotype strains means that recombination is not an impediment to strain characterization, and allelic information can be used to readily detect recombination between genes and rates of recombination within a population (17). Phylogenetic and other clustering algorithms can be applied to individual or concatenated gene datasets for further analyses of strain relationships, although caution in interpretation of phylogenetic relationships is necessary due to recombination issues.

The MLST scheme developed for Wolbachia indexes variation at five conserved genes (ftsZ, gatB, coxA, hcpA, and pbpA). The PCR primers used robustly amplify loci of strains belonging to superfamilies A and B and potentially amplify loci of strains belonging to other superfamilies as well. Specific primers for amplification of superfamilies A and B strains coexisting on the same host individual are also described here. A profile database and an isolate database were generated to store and distribute the sequences and the strain and host information, respectively (22; http://pubmlst.org/wolbachia/). New strains and profiles can be submitted to the curator for inclusion in the databases, which thus represent a focal point for future storage and management of all Wolbachia strain data.

Given the current extensive use of wsp sequences for characterizing Wolbachia strains (25, 30, 36, 43, 53) and the suspected role of the Wolbachia surface protein in the host-symbiont interactions (1, 62), use of this gene as an additional optional strain marker is proposed. A WSP typing system was developed based on the amino acid sequences encoded by the four hypervariable regions (HVRs) of this gene (2), and a separate database for storing both nucleotide sequences of the wsp gene and amino acid sequences of single HVRs was generated. WSP typing is analogous to antigen protein typing used for pathogenic bacteria (33) and can complement the MLST information.

Thirty-seven strains representing superfamilies A, B, D, and F from diverse singly infected host species were typed using the MLST and WSP systems. MLST unambiguously identified, differentiated, and grouped genetically similar strains through comparisons of their allelic profiles. The MLST data revealed considerable diversity within the genus Wolbachia and also identified horizontal movement of strains between hosts and shifts in Wolbachia phenotypic effects on hosts (e.g., cytoplasmic and parthenogenogenesis induction).

MATERIALS AND METHODS

Choice of loci. The choice of the genes used for the MLST system was facilitated by an ongoing project to characterize genetic diversity in Wolbachia. A total of 46 unique loci were amplified from a subset of more than 50 Wolbachia strains belonging to superfamilies A, B, C, D, E, F, and H and sequenced. DNA was extracted from either a whole individual, the abdomen, or the gonads, depending on the size of the specimen, using a DNAeasy tissue kit (QIAGEN, Germantown, MD).

Amplicons were generated using standard PCR conditions with HotStarTaq (QIAGEN) according to the manufacturer’s recommendations and with each ~64-fold degenerate primer at a concentration of 0.5 µM. The forward and reverse primers had 5’ tags of M13 forward and reverse primer sequences, respectively. These tags served as anchors for the degenerate primers during amplification and were subsequently used for sequencing. Amplification reactions were initiated with 15 min of incubation at 95°C, followed by 50 cycles of 95°C for 15 s, 55°C for 30 s, and 72°C for 1 min. The amplification reaction mixtures (8 µl) were treated with 0.5 U shrimp alkaline phosphatase and 1.0 U exonuclease I (Amersham) in the buffer supplied. Sequencing reactions were performed at the J. Craig Venter Joint Technology Center, Rockville, MD (an affiliate of The Institute for Genomic Research [TIGR], Rockville, MD) with the M13 forward and reverse primers. Assembly was done with the TIGR assembler (47), and sequences were manually curated with the multiple-sequence editor Cloe (a multiple-sequence editor heavily integrated into the TIGR database). Alignments were initially generated using CLUSTALX (51) and then manually edited in MacClade 4.08 (http://www进化.net) and Bioclade.org (http://www进化.net) and then edited in MacClade 4.08 (http://www进化.net) and Bioclade.org (http://www进化.net) and finally edited in MacClade 4.08 (http://www进化.net) and Bioclade.org (http://www进化.net). Finally, sequences were manually curated with the multiple-sequence editor MacClade 4.08 (http://www进化.net) and Bioclade.org (http://www进化.net).

From the initial collection of 46 loci, five conserved genes were chosen for MLST on the basis of the following features: (i) presence throughout the sequenced Rickettsiales order; (ii) a single copy in the wMel genome (60); (iii) wide distribution in the wMel genome (Fig. 1); and (iv) evidence of strong stabilizing selection within the genus Wolbachia (average ratio of the number of nonsynonymous substitutions per nonsynonymous site to the number of synonymous substitutions per synonymous site [Ks/Ka] <1). These criteria conform to the standard locus requirements for an MLST system (27, 52). The five genes chosen for the MLST scheme are gatB, coxA, hcpA, ftsZ, and pbpA (Table 1). Only one of these genes, ftsZ, was used in previous studies (59, 32). Briefly, the gatB gene encodes aspartyl/glutamyl-tRNA amidotransferase subunit B that synthesizes charged tRNAs in organisms lacking certain tRNA synthetases; coxA encodes a catalytic subunit of cytochrome oxidase of the respiratory chain; hcpA encodes a functionally uncharacterized protein whose length and amino acid sequence are highly conserved in many bacteria, which is designated NCBI COG0217 (49, 50); ftsZ encodes a protein involved in bacterial cell division; and pbpA encodes fructose-biphosphatase aldolase that in Wolbachia is likely to be involved in gluconogenesis (8, 9). The surface protein WSP was included in this study as an additional marker for strain typing.

MLST and wsp primer design and amplification. For each gene, primers were designed manually using sequence alignments that were generated from the initial sample of more than 50 strains and, when possible, included outgroup sequences of Ehrlichia ruminantium Welgevonden, Ehrlichia chaffeensis Arkan-
sas, Anaplasma margarite St. Maries, Anaplasma phagocytophilum HZ, and
Neorickettsia sennetsu Miyayma. Primers (Table 1) were designed to be general across all arthropod Wolbachia supergroups but not to amplify any of the outgroups mentioned above. For \( wsp \), primers were designed to amplify most of the gene (546 bp) (Table 1), including the four HVRs.

PCR amplification was performed by using 20-\( \mu \)l mixtures containing each deoxynucleoside triphosphate at a concentration of 0.2 mM, 1.5 mM MgCl$_2$, each primer at a concentration of 0.5 \( \mu \)M, 0.5 U Taq DNA polymerase, 1× PCR buffer, and 1 \( \mu \)l DNA. Reactions were initiated by incubation at 94°C for 2 min, which was followed by 36 cycles of 94°C for 30 s, the optimal annealing temperature (see below) for 45 s, and 72°C for 1.5 min, a final elongation step at 70°C for 10 min, and a final hold at 4°C. The optimal PCR annealing temperature was 53°C for \( hcpA \), 54°C for both \( gatB \) and \( ftsZ \), 55°C for \( coxA \), and 59°C for \( fbpA \) and \( wsp \). PCR products were purified using Montage PCR centrifugal filter devices (Millipore).

**TABLE 1. MLST and \( wsp \) loci and primer features**

| Cluster category | Locus code (wMel) | Gene | Product | Primer Designation | Sequence (5′–3′) | Gene length (bp) | Amplified nucleotide range (bp) | MLST fragment size (bp) |
|------------------|------------------|------|---------|-------------------|-----------------|----------------|-----------------------------|------------------------|
| *α-Proteobacteria* | WD_0146 | gatB | Glutamyl-tRNA(Gln) amidotransferase, subunit B | gatB_F1 | GAKTTAAAYCGYCAGGGBTT | 1,425 | 421–891 | 369 |
| *Escherichia coli* | WD_0301 | coxA | Cytochrome c oxidase, subunit I | coxA_F1 | TTTGGRGRATYAACTTTATAG | 1,551 | 491–977 | 402 |
| *Escherichia coli* | WD_0484 | hcpA | Conserved hypothetical protein | hcpA_F1 | GAATATCAGTGTTGCAAACAGG | 741 | 91–605 | 444 |
| *Escherichia coli* | WD_0723 | ftsZ | Cell division protein | ftsZ_F1 | ATYATGGCARATATAAAGARTAG | 1,197 | 274–798 | 435 |
| *Rickettsiales* | WD_1238 | fbpA | Fructose-bisphosphate aldolase | fbpA_F1 | GTCGTTCCRCCTTGGGWTGAT | 900 | 241–749 | 429 |
| *Wolbachia* | WD_1063 | wsp | Outer surface protein | wsp_F1 | GTCCAAATRSTGATGARGAAAC | 714 | 85–688 | 546 |

$^a$ Clustering categories indicating the level of protein conservation.

$^b$ With respect to the wMel genome.
TABLE 2. Wolbachia strains analyzed

| Host species or subspecies (strain[s]) | Order       | Generic name | Supergroup | Host phenotype |
|---------------------------------------|-------------|--------------|------------|---------------|
| Acromis sparsa                        | Coleoptera  | Beetle       | A          |               |
| Aedes albopictus                      | Diptera     | Mosquito     | A          | CI            |
| Camponotus pennsylvanicus            | Hymenoptera | Ant          | A          |               |
| Drosophila bifasciata                 | Diptera     | Fruit fly    | A          | MK            |
| Drosophila innubila                   | Diptera     | Fruit fly    | A          | MK            |
| Drosophila melanogaster               | Diptera     | Fruit fly    | A          | CI            |
| Drosophila neotecestaea               | Diptera     | Fruit fly    | A          | CI            |
| Drosophila orientacea                 | Diptera     | Fruit fly    | A          | CI            |
| Drosophila simulans (wRi)             | Diptera     | Fruit fly    | A          | CI            |
| Drosophila simulans (wAu)             | Diptera     | Fruit fly    | A          | CI            |
| Ephestia kuehniella                   | Lepidoptera | Flour moth   | A          |               |
| Incesttermes snyderi                  | Isopitera   | Termite      | A          |               |
| Muscifurax uniraptor                  | Hymenoptera | Wasp         | A          | PI            |
| Nasonia giraulti (16.2)               | Hymenoptera | Wasp         | A          | CI            |
| Nasonia longicornis (2.1)             | Hymenoptera | Wasp         | A          | CI            |
| Nasonia vitripennis (12.1)            | Hymenoptera | Wasp         | A          | CI            |
| Solenopsis invicta                    | Hymenoptera | Ant          | A          |               |
| Acraea eponina                        | Lepidoptera | Butterfly    | B          | MK            |
| Acraea eponina                        | Lepidoptera | Butterfly    | B          |               |
| Armadillidium vulgare                 | Isopoda     | Isopod       | B          | FI            |
| Chelymorpha alternans                 | Coleoptera  | Beetle       | B          | CI            |
| Culex pipiens pipiens                 | Diptera     | Mosquito     | B          | CI            |
| Culex pipiens quinquefasciatus        | Diptera     | Mosquito     | B          | CI            |
| Drosophila simulans (wMa)             | Diptera     | Fruit fly    | B          | CI            |
| Drosophila simulans (wNo)             | Diptera     | Fruit fly    | B          | CI            |
| Encaaria formosa                      | Lepidoptera | Wasp         | B          | PI            |
| Ephestia kuehniella                   | Lepidoptera | Flour moth   | B          |               |
| Gryllus firmus                        | Orthoptera  | Cricket      | B          | CI            |
| Nasonia vitripennis (4.9)             | Lepidoptera | Wasp         | B          | CI            |
| Ostrinia scapulalis                   | Lepidoptera | Moth         | B          | MK            |
| Protocalliphora siutta (B1, 00-189)   | Diptera     | Blow fly     | B          |               |
| Teleogryllus taiwannemama             | Orthoptera  | Cricket      | B          | CI            |
| Tribolium confusus                    | Coleoptera  | Beetle       | B          | CI            |
| Trichogramma deion (TX)               | Lepidoptera | Wasp         | B          | PI            |
| Brugia malayi                         | Spirurida   | Nematode     | D          | M             |
| Folsomia candida                      | Collembola  | Springtail   | E          |               |
| Copiotermes acuinaciformis            | Isopitera   | Termite      | F          |               |
| Cimex lectularis                      | Heteroptera | Bed bug      | F          |               |
| Zootermopsis angusticollis            | Isopitera   | Termite      | F          |               |
| Cordylochernes scorpionoides          | Pseudoscorpion | Pseudoscorpion | MK  |
| Ctenocephalides felis                 | Siphonaptera | Dog flea     |               |               |

*a All hosts were infected with a single strain.

b MK, male killing; PI, parthenogenesis induction; CI, cytoplasmic incompatibility; FI, feminization induction; M, mutualism. The host phenotypes are based on previously published information.

central ST, which is identified as the group founder based on BURST analyses implemented in START2 (23).

When the complete set of the five MLST gene sequences cannot be obtained for a strain, single gene alleles and partial MLST allelic profiles can be submitted to the database. Partial data provide useful allele diversity information, allowing the profile database to grow. A strain cannot be assigned to an ST until full sequences, allelic profiles, and STs. The isolate database stores information about the strains and host strain biological information (e.g., taxonomy, natural history, collection date, and locality). Each strain is designated by an identification number and a code consisting of the first three letters of the host species name, underscore, a letter indicating the supergroup to which it belongs, underscore, and any further specific information about the strain needed for discrimination. For example, Dsim_A_wRi indicates the Wolbachia strain from D. simulans belonging to supergroup A found in Riverside County (CA) currently referred to as wRi. The isolate database allows storage of accession numbers for sequences other than the sequences for the MLST genes that are available for a specific strain. The WSP database stores both nucleotide and amino acid sequences for wsp alleles and HVRs. All of the databases are reciprocally linked and can be accessed at the Wolbachia MLST website (http://pubmlst.org/wolbachia/).

Analyses of genetic variation and recombination. For each locus, the number of variable sites (VI), G+C content, level of nucleotide diversity per site (Pi), and
Bayesian analyses were performed using MrBayes, v. 3.1.1 (18), with 10^6 generations to reconstruct phylogenies based on the concatenated MLST data set and the concatenated MLST gene data set and GTR + T model. Maximum likelihood inference and maximum parsimony were used to reconstruct phylogenies based on the concatenated MLST data set and wsp data. Bayesian analyses were performed using MrBayes, v. 3.1.1 (18), with 10^6 generations and a sample frequency of 100. The first 5,000 trees were discarded as burn-in, and 50% majority rule was applied to the remaining set of trees. Maximum parsimony bootstrap analyses were performed with PAUP v. 4.01 (48) using 1,000 replicates and 100 random-addition replicates per bootstrap replicate, and a 50% majority rule bootstrap was applied.

Tree reconstruction. Strict maximum likelihood (ML) inference was used to reconstruct phylogenies based on single and concatenated MLST genes (PAUP v. 4.01). The appropriate model of evolution was estimated with Modeltest v. 3.0, and the best likelihood score was evaluated with the Akaike information criterion (37). The models selected were GTR + T for gatB, wsp, and the concatenated MLST gene data set and GTR + I + G for coxA, hcpA, and fbpA. Bayesian maximum likelihood inference and maximum parsimony were used to reconstruct phylogenies based on the concatenated MLST data set and wsp data. Bayesian analyses were performed using MrBayes, v. 3.1.1 (18), with 10^6 generations and a sample frequency of 100. The first 5,000 trees were discarded as burn-in, and 50% majority rule was applied to the remaining set of trees. Maximum parsimony bootstrap analyses were performed with PAUP v. 4.01 (48) using 1,000 replicates and 100 random-addition replicates per bootstrap replicate, and a 50% majority rule bootstrap was applied.

A neighbor-joining tree was reconstructed from the 37 MLST allelic profiles using the START2 program (23). An unweighted pair group method with arithmetic mean dendrogram was constructed from the matrix of pairwise similarities for the 36 WSP profiles (START2 program). START2 and a suite of publicly available bioinformatics software for MLST data analysis can be found at the MLST home page (http://pubmlst.org).

Analysis of congruence between trees. We used the Shimodaira-Hasegawa test (42) to evaluate the significance of topological incongruence among the ML trees based on the five MLST genes and the concatenated data set. The lnL scores and significance values for nucleotide substitution were optimized separately for all topologies considered. Differences in lnL scores between the best ML topology and an alternative topology were calculated for each data set. The significance of these differences was measured using a bootstrap approach with RELL and full optimization for 1,000 replicates (PAUP v. 4.01).

Nucleotide sequence accession numbers. All MLST and wsp sequences generated in this study have been deposited in the GenBank database under accession numbers DQ842268 to DQ842286.

RESULTS

Primers. The MLST primers worked for 34 of 35 supergroup A and B strains tested (Table 2, first 35 strains). In only one case, Trichogramma deion (TX), did the ftsZ primer not amplify. This strain sequence was generated using the MLST primers with tags (http://pubmlst.org/wolbachia/). The wsp primers amplified the gene from all the supergroup A and B strains tested except the strain in T. deion. All sequences generated using MLST primers perfectly matched the corresponding sequences independently generated using the initial 64-fold degenerate primers. MLST and wsp primers consistently generated amplification products in all the laboratories where they were tested, and bidirectional sequencing of PCR products gave unambiguous results.

The primers were sufficiently degenerate to also amplify divergent strains not belonging to supergroups A and B, such as Wolbachia strains from the nematode B. malayi (supergroup D) and the bed bug Cimex lectularius (supergroup F). For other divergent strains the primers worked well for only some of the loci. All loci except ftsZ were amplified from Wolbachia strains in the termite Zootermopsis angusticollis (supergroup H) and the flea Ctenocephalides felis. All loci except cox1 were amplified from Wolbachia strains in the pseudoscorpion Cor-
As a likely result of recombination both within and outside the gene boundaries (intra- and intergenic recombination, respectively), some strains were extremely similar at one locus and highly divergent at another. This disparity in divergence was visualized by plotting the genetic distances for loci for pairs of strains that were identical at one or two loci (Fig. 2). The levels of genetic divergence at nonidentical loci estimated for pairs of strains that exhibited identity at $\textit{gatB}$ were remarkably high in some cases (Fig. 2). For instance, for $\textit{Ostrinia scapulalis}$ and $\textit{N. vitripennis}$ (B) there was about 11% divergence at $\textit{fbpA}$. Similar values were found for $\textit{O. scapulalis}$ and $\textit{Acraea encedon}$, for $\textit{Telogryllus taiwanemma}$ and $\textit{N. vitripennis}$ (B), for $\textit{T. taiwanemma}$ and $\textit{A. encedon}$, for $\textit{N. vitripennis}$ (B) and $\textit{Chelymorpha alternans}$, and for $\textit{N. vitripennis}$ (B) and $\textit{A. encedon}$. These values are significantly greater than the average level of divergence for this locus (2.7%) based on the pairs of strains with identity at another locus ($P < 0.01$). Conflicting relationships among strains could also be visualized (Fig. 2); $\textit{N. vitripennis}$ B and $\textit{C. alternans}$ were very similar based on their $\textit{coxA}$ alleles (0.2% divergence) but were very divergent based on their $\textit{fbpA}$ alleles (11.4%). Conversely, $\textit{N. vitripennis}$ (B) and $\textit{A. encedon}$ were very similar based on their $\textit{fbpA}$ alleles (0.2%) but were very divergent based on their $\textit{coxA}$ alleles (8.2%). These inversions of relationships indicate that there was recombination among divergent strains.

**Phylogenetic inference of relationships.** As expected for a data set affected by recombination, phylogenetic analyses for single and concatenated MLST loci revealed significant incongruence among loci (Table 4). Conflicts in tree topologies were inferred for all pairwise comparisons of the five genes. Indeed, there appeared to be no single phylogeny for a strain but multiple phylogenies for different fragments of the genome. As a result, the phylogenetic tree based on the concatenated sequences (Fig. 3) revealed strains that are closely related but

---

**TABLE 3. Genetic features and occurrence of recombination at the five MLST loci**

| Locus | Data set | No. of alleles | % of VI | Nucleotide diversity per site | G+C content (%) | $K_s/K_a$ | Recombination (MaxChi, $P < 0.01$) |
|-------|----------|----------------|---------|-------------------------------|-----------------|---------|----------------------------------|
| $\textit{gatB}$ | ABDF | 26 | 28.7 | 0.078 | 37.2 | 0.093 | Yes ($\textit{G. firmus}$) |
| A | 13 | 9.2 | 0.027 | 37.8 | 0.151 | No |
| B | 11 | 11.1 | 0.022 | 36.6 | 0.135 | No |
| $\textit{coxA}$ | ABDF | 27 | 28.4 | 0.083 | 39.3 | 0.078 | No |
| A | 11 | 4.5 | 0.015 | 39.9 | 0.099 | No |
| B | 14 | 17.4 | 0.053 | 38.7 | 0.072 | Yes ($\textit{P. sialis}$ [B1], $\textit{T. confusum}$, $\textit{E. formosa}$) |
| $\textit{hcpA}$ | ABDF | 31 | 36.5 | 0.081 | 37.0 | 0.140 | No |
| A | 13 | 12.6 | 0.029 | 37.7 | 0.090 | Yes ($\textit{A. albopictus}$, $\textit{A. sparsa}$) |
| B | 16 | 15.6 | 0.028 | 36.3 | 0.346 | No |
| $\textit{fbpA}$ | ABDF | 28 | 30.5 | 0.092 | 38.6 | 0.084 | Yes ($\textit{A. vulgare}$) |
| A | 12 | 9.0 | 0.034 | 38.7 | 0.085 | No |
| B | 15 | 20.1 | 0.063 | 38.7 | 0.089 | Yes ($\textit{C. alternans}$) |
| $\textit{ftsZ}$ | ABDF | 25 | 24.8 | 0.065 | 41.5 | 0.047 | No |
| A | 10 | 6.4 | 0.012 | 42.6 | 0.028 | No |
| B | 14 | 10.3 | 0.020 | 40.4 | 0.065 | No |

---

*Data sets A and B included 18 supergroup A strains and 17 supergroup B strains, respectively (Table 2).*

*The percentage of variable sites was estimated by considering indels in each data set variable sites.*

*Recombination inferences for data set ABDF refer to events that occurred between supergroups.*

*The organisms in parentheses are hosts of putative recombinant strains.*

---

**FIG. 2. Pairwise divergence at MLST loci.** Thirty-three strain pairs identical at one or two loci were used to show the variation in similarity across loci. The most striking pattern was found for strains identical at $\textit{gatB}$ (vertical bar); some of the strains showed remarkable divergence in the $\textit{fbpA}$ and $\textit{coxA}$ alleles, likely associated with recombination events. The arrows indicate examples of contradictory relationships inferred from the results for the five genes involving $\textit{N. vitripennis}$ (B), $\textit{C. alternans}$, and $\textit{A. encedon}$ strains (values are circled; see text for explanation).
also created phylogenetic artifacts that resulted from recombination between more divergent strains. For instance, members of the five clusters highlighted in the tree (Fig. 3) likely descended from a recent common ancestor, as reflected by strong grouping in the concatenated tree (both posterior probability and bootstrap values were >95%) and consistent phylogenetic clustering based on single-gene phylogenies (data not shown). However, the concatenated tree also revealed several clusters of strains whose relationships differ in single-locus phylogenies. For example, based on the concatenated data set, the *Encarsia formosa* and *P. sialia* (B1) strains formed a highly supported cluster, with both bootstrap and posterior probability values of ~100%. However, based on *coxA* phylogeny, the *E. formosa* strain clustered with the *Acraea eponina* strain (data not shown) and was separated from the *P. sialia* (B1) strain by a remarkable genetic distance (Fig. 2). Consistent with the effects of recombination, these two strains also did not cluster based on the *wsp* phylogeny (Fig. 4). Hence, the concatenated gene approach alone can be useful for identifying very closely related strains and to resolve major supergroups. However, it

![Diagram](image)

**FIG. 3.** Bayesian likelihood inference phylogeny based on the concatenated data set for the five MLST loci (37 strains, 2,079 bp). Groups of strains sharing at least three MLST alleles are highlighted. Arrows and asterisks indicate two examples of strain pairs whose predicted relationships are highly discordant with the *wsp* phylogeny (Fig. 4). Posterior probability (left) and parsimony bootstrap (right) values are indicated at major nodes if they were supported by both clustering algorithms.

### TABLE 4. Results of a Shimodaira-Hasegawa test of alternative tree topologies for the *Wolbachia* MLST genes

| Topology | *gotB* | *coxA* | *hcpA* | *fbpA* | *ftsZ* | Concatenated |
|----------|--------|--------|--------|--------|--------|--------------|
| Concatenated | 1,222.25*** | 1,445.83*** | 1,649.30* | 1,653.84 | 1,238.10 | 7,767.07 |
| *coxA* | 1,260.60*** | 1,328.74 | 1,751.42*** | 1,990.72*** | 1,310.56*** | 8,161.79*** |
| *hcpA* | 1,273.48*** | 1,532.01*** | 1,584.14 | 1,869.74*** | 1,280.41*** | 8,135.25*** |
| *fbpA* | 1,331.80*** | 1,552.98*** | 1,733.80*** | 1,959.66 | 1,318.52*** | 8,090.05*** |
| *ftsZ* | 1,300.73*** | 1,562.61*** | 1,845.66*** | 2,069.53*** | 1,208.86 | 8,561.70*** |

*a*, *P* < 0.05; ***, *P* < 0.01; ****, *P* < 0.001. The lowest (best) likelihood scores are indicated by boldface type.
cannot be used to interpret more distant phylogenetic relationships within a supergroup, even when clades are highly supported, because of artifacts resulting from recombination among genes. Combined analyses using allelic profiles, single-gene phylogenies, and concatenated gene phylogenies are advisable for any inference of strain relationships.

MLST inference. Based on the neighbor-joining tree constructed from the STs (Fig. 5), the majority (35/37) of the STs were unique to a strain, as were most of the alleles at a single locus. Nine of the 35 STs (Fig. 4) did not share any alleles with any other ST. This allelic diversity was probably due in part to the great genetic diversity of *Wolbachia* and in part to the sample of strains selected, which spanned a wide range of host species. No alleles were shared by STs of supergroup A and B strains. The following two pairs of strains among the pairs analyzed exhibited identity at all loci: the pair consisting of the *Culex pipiens* and *C. pipiens quinquefasciatus* strains (ST-9) and the pair consisting of the *N. vitripennis* (A) and *Muscidifurax uniraptor* strains (ST-23). *Culex* strains are identical at other chromosomal genes (13, 44), while the identity for the *N. vitripennis* (A) and *M. uniraptor* wasp strains was not reported previously.

By convention, strains that share at least three MLST alleles with a central ST are grouped in a complex (ST complex). Additional sequencing of 28 genes showed that members of an ST complex based on our five MLST genes were also consistently closely related based on the larger gene set, indicating the reliability of the MLST system for identifying closely related strains.

A total of three ST complexes were identified based on the criteria used (three shared alleles with a central ST) (Fig. 5). The ST-13 complex included supergroup A strains present in *Drosophila* species and the parasitoid wasp *N. longicornis* (A) (three shared alleles). The five strains differed by only five nucleotide polymorphisms based exclusively on the concatenated alignment of the five MLST genes (2,079 bases). The central ST (ST-13) was identified as the *Drosophila recens* ST based on BURST analysis. The other two complexes included only two strains each and were temporarily designated ST-a and ST-b, until additional STs that allow identification of the
central ST become available. ST-a included two supergroup B strains, the *D. simulans* wNo strain and the *D. simulans* wMa strain (three shared alleles, three polymorphic sites), and ST-b included two supergroup A strains, the *Drosophila orientacea* and *Drosophila neotestacea* strains (four shared alleles, one polymorphic site).

**WSP characterization.** The HVRs of the WSP protein were employed as an additional, optional marker to assess strain diversity. A new method for representing WSP sequence relationships is proposed here, which can be used in combination with standard distance-based methods. This method takes into account the extensive intragenic shuffling of the four HVRs (2) and can be used to readily detect allele shuffling based on unique identifiers for each HVR. Since the majority of amino acid changes among WSP sequences occur in the four HVRs, these motifs can be used as signatures for discrimination of the different WSP protein types (see Materials and Methods).

The sample of strains characterized by MLST was also typed by WSP, with the exception of the *T. deion* strain, for which amplification of *wsp* failed. Figure 6 shows the WSP profile, a combination of the four HVR amino acid haplotypes. For 36 strains, 33 unique WSP profiles were identified. Five groups of strains shared at least three HVR haplotypes and differed by only two amino acids in HVR3. Two of the three ST complexes (Fig. 5) were also very closely related based on their WSP profiles (Fig. 6).

**WSP inference and comparison to MLST.** Analogous to the ST-based neighbor-joining tree (Fig. 5), the relationships among WSP profiles could be schematically represented in a tree (Fig. 6). We noted that relationships inferred by MLST and WSP profiles were not necessarily concordant but were complementary. In some cases strains that were closely related based on their MLST profiles also had similar WSP profiles. Indeed, the two *C. pipiens* strains had identical ST (ST-9) (Fig. 5) and WSP (Fig. 6) profiles. The *M. uniraptor* and *N. vitripennis* (A) strains, which had the same ST (ST-23), shared three of four HVR haplotypes and differed by only two amino acids in HVR3. Two of the three ST complexes (Fig. 5) were also very closely related based on their WSP profiles (Fig. 6). Specifically, *D. simulans* wMa and *D. simulans* wNo (ST-a
As found with the MLST profiles, no identical HVRs were observed for the supergroup A and B strains, although some HVRs were closely related. Nine of 16 supergroup A strains had the same haplotype sequence in HVR1 (haplotype 1). Two strains, from the tortoise beetle Acromis sparsa and the ant Solenopsis invicta, had four unique HVR haplotype sequences. Both strains belonged to supergroup A, as supported by phylogenetic analyses based on MLST data (both the posterior probability and bootstrap values were 100) (Fig. 3) and MLST profiles (both strains exhibited identity at least at one allele with another supergroup A strain) (Fig. 5). However, based on the wsp gene phylogeny (Fig. 4), the strain from A. sparsa was excluded from both supergroup A and supergroup B and formed a divergent lineage, while the strain from S. invicta clustered significantly in supergroup B (both the posterior probability and bootstrap value were >90). Based on MaxChi analyses, the A. sparsa wsp sequence was the product of recombination between the supergroup A wsp-type sequence of Ephesia kuehniella (A) and the supergroup B-type sequence of E. formosa (P < 0.001). In the case of the Wolbachia strain from S. invicta, lateral transfer of the whole wsp gene from a supergroup B-type donor strain and recombination may explain the contradictory clustering inferred from the MLST and wsp data sets (Fig. 3 and 4).

Overall, there was strong incongruence between MLST- and wsp-based relationships for both subgroup and supergroup strain assignment, indicating the need for a multilocus approach.

DISCUSSION

Wolbachia strains are among the most common obligate intracellular bacteria (56), and by virtue of the diverse effects that they have on their hosts (46, 54) and their potential health and agricultural implications (35, 61), they represent an extremely important group. The MLST system described here allows universal characterization of Wolbachia strains that induce a variety of phenotypes in a diverse range of hosts. It provides a standard method for reliable strain typing, a unified nomenclature for Wolbachia, and a centralized and curated database for data storage and management. Accurate strain typing is an extremely powerful tool for investigating the biology and evolution of Wolbachia. Because of the indexing of a standard set of genes for typing Wolbachia strains and the common field host information in the isolate database, it is possible to relate any new submitted strain to an expanding data set. With the isolate database, host and strain information are partitioned into multiple, searchable fields. Thus, each field can be used to sort the data, and the frequency of a single field or a combination of fields can be calculated and graphically displayed (for instance, host geographical location versus specific allele frequencies). As the database grows, informative patterns can emerge and key questions can be addressed; such as (i) what is the global pattern of diversity of Wolbachia strains; (ii) are there particular taxa that have an unusually high number of infected species; (iii) how do Wolbachia strains move between host taxa, and what genetic changes occur when they shift hosts; (iv) how do the strains move locally and globally within arthropod communities (e.g., are there strains or alleles that are found in particular geographical regions, while
other strains or alleles have a global distribution); and (v) do particular closely related strains show shifts of phenotypes.

MLST is also an important source of sequence data for extensive comparative genetics, providing a platform for investigating molecular evolutionary processes in these intracellular bacteria.

Results reported in this study reveal that (i) the MLST primers work reliably for the two major supergroups of Wolbachia found in arthropods (supergroups A and B), as well as with other groups (however, additional flanking primers are provided at the MLST website); (ii) there is considerable allelic diversity in genes and among Wolbachia strains, with recombination occurring both within and between the MLST genes; (iii) MLST can identify and discriminate closely related strains using chromosomal genes; (iv) closely related strains detected by MLST (strains that share a minimum of three alleles in their MLST profiles) can show remarkable divergence at WSP as a result of recombination at the wsp gene; (v) strains with very similar WSP can have highly divergent MLST profiles; and (vi) strains closely related based on MLST can produce very different host effects.

**Discrimination and grouping of closely related strains.** One of the challenges in the study of Wolbachia is the difficulty encountered when workers attempt to infer similarities among strains. Because of horizontal genetic exchange, Wolbachia genes do not evolve in a bifurcating, tree-like fashion. Thus, groups of truly derived strains cannot be recovered with traditional cladistic methods. Previous studies have inferred phylogenetic relatedness among strains based on similarity or identity at one or two genes, for instance, ftsZ (59) or wsp (43, 62). The data presented here demonstrate that because of recombination, identity at single genes is not sufficient for reliable strain typing and that the high level of support for a clade in a concatenated phylogeny could be biased by the unequal divergence of loci, masking single-locus conflicts. Because MLST does not describe strain relationships in terms of the level of homology among sequences (i.e., phylogenetically) but describes them in terms of identity versus nonidentity among alleles (i.e., similarity), it avoids false inferences due to recombination issues. The ST tree in Fig. 2 provides a schematic representation of relationships among allelic profiles and can be combined with phylogenetic and other clustering algorithm analyses based on the sequences in individual or concatenated MLST data sets.

Based on the data set analyzed, only STs sharing three of five identical MLST loci (designated an ST complex) show consistent similarity at all the remaining loci. Strains identical at three alleles are likely to have a recent common ancestor and thus represent a phylogenetic cluster of recently diverged strains. Preliminary analyses of an additional 28 chromosomal genes for the members of three complexes (unpublished data) confirmed their consistent relatedness, indicating that MLST profiles can be used to predict similarity at other loci. Furthermore, no recombination was found between members of the same ST complex at MLST genes. In contrast, recombination at MLST genes was detected between strains identical at one or two MLST alleles. This indicates that grouping strains based on at least three shared alleles at MLST loci can form the basis of a reliable system for assessing actual similarity at chromosomal genes among strains. However, as more allelic profiles become available, it should be possible to test this observation more extensively.

Wolbachia researchers have used the gene wsp extensively to determine relationships among strains. However, using our data set, we found several instances in which strains closely related based on MLST data are very divergent based on WSP data and strains that have similar WSP differ substantially based on MLST data. For instance, members of the ST-13 complex showed extensive divergence in HVR4 of WSP, as a result of recombination. However, when phylogenetic methods were applied to this recombinant data set, a recombination signature in HVR4 was masked by a high level of similarity in the first three HVRs and the five strains apparently formed a strong cluster (posterior probability, 100%; bootstrap support, 100%) (Fig. 4). As an example of the opposite pattern, highly similar WSP sequences were shared by strains found in N. vitripennis (B) and E. kuehniella (B) (1% amino acid divergence), whose FbpA and HcpA protein sequences differ substantially (8 and 5% divergence, respectively). Such findings further support the conclusion that wsp alone is an unreliable predictor of either strain similarity or divergence.

Overall, the results show that the MLST system described here provides an excellent method for typing Wolbachia strains from diverse hosts and also for discriminating among strains in the same host species (e.g., D. simulans strains). It should be noted that identity in MLST genes does not mean that two strains are identical throughout their genomes. Even within a host species, where an ancestral strain can be derived from a relatively recent ancestor, variation within the genome can accumulate, particularly in quickly evolving or highly recombinant genes, such as mobile elements and/or ankyrin repeat domains. This has been shown for strains found in C. pipiens and Drosophila species (19, 44). Discrimination of substrains within the same ST requires the use of additional specific polymorphic markers (39, 44). The MLST database can accommodate this additional information through links that can be accessed through stored accession numbers.

**Detection and typing of multiple coinfecting strains.** Double Wolbachia infections are common in insects (24, 55), and only in some circumstances can they be separated experimentally (34). MLST primers for typing supergroup A and B Wolbachia strains from individuals that are doubly infected have been designed to address this issue. However, PCR-based systems like MLST and WSP typing are constrained for examination of double infections as the different alleles cannot always be reliably assigned to specific strains coinfecting the same individual. A preliminary analysis using a small sample of multiply and singly infected species gave positive results. Primer performance, however, should be verified on a larger scale. Use of the group-specific primers for MLST requires workers to assume that there have not been reciprocal gene exchanges between the strains. Therefore, MLST for typing hosts that are doubly infected with supergroup A and B Wolbachia strains should not be used unless reliable supergroup A- and B-specific alleles can be identified for each of the MLST genes. Allelic profiles for doubly infected individuals are given the status “candidate” STs in the MLST database until the strains can be experimentally separated (e.g., by partial antibiotic curing or segregation by other means).

In some cases host species are coinfectd with very closely
related strains that belong to the same supergroup. These strains can occur in double or multiple infections and are coded A1A2, B1B2, etc. Because of the high levels of genetic similarity among the strains, these multiple infections are normally hard to detect, and when they are detected, the specific amplification and correct assignment of sequences to one of the strains are often difficult. Such multiple infections may be more common than has been assumed, and providing an estimate of their global frequency in host species represents an important first step. Based on a preliminary screening of a small sample of *N. longicornis* strains doubly infected with supergroup A and/or B strains using MLST group-specific primers, the multiple peaks in the electropherograms of some of the MLST gene products can be read as a signature of the potential presence of multiple strain variants. However, additional analyses are necessary to verify the actual existence of multiple strains.

**Relating MLST to Wolbachia biology.** Even with the limited number of strains analyzed here, several patterns regarding the transmission of *Wolbachia* are apparent. Based on both MLST and WSP data, different host organisms were found to harbor very similar or identical *Wolbachia* strains; these organisms include the wasp species *N. vitripennis* (A) and *M. uniraptor*, the mosquito subspecies *C. pipiens pipiens* and *C. pipiens quinquefasciatus*; the fly species *D. melanogaster*, *D. simulans* wAu, *D. recens*, and *D. innubila*, as well as the wasp species *N. longicornis*; and the fly species *D. neotestacea* and *D. orientacea*. The strain similarities may have been due to recent horizontal transmission of *Wolbachia* across host species or, in the case of the closely related species *D. neotestacea* and *D. orientacea*, to codivergence and coevolution of *Wolbachia* strains and their hosts. Because genetically similar strains are often found in similar insect hosts, ecological interactions among hosts may have mediated such horizontal transfers. In particular, it appears that horizontal transmission of *Wolbachia* has occurred quite recently and frequently within the genus *Drosophila*.

Many of the reproductive phenotypes of the strains analyzed in this study remain unknown, and the possibility that there is a correlation between STs and/or WSP profiles and the strain phenotype cannot be eliminated yet. Based on the data set analyzed, genetically similar strains found in different host species were often associated with distinct host reproductive manipulations (Table 2). For example, strains found in *N. vitripennis* (A) and *M. uniraptor*, which have identical ST and very similar WSP profiles, are responsible for inducing cytoplasmic incompatibility (5) and parthenogenesis (45), respectively, in their hosts. Similarly, based on both MLST and WSP data the *D. simulans* wAu and wasp *N. longicornis* (A) strains are very closely related but are responsible for different patterns of cytoplasmic incompatibility (5, 29, 58). Based on these observations, it is now relevant to investigate the genetic changes in these closely related strains that may be involved in the phenotypic shift. In addition, as future studies clarify the specific role of wsp in host-parasite interactions, information on amino acid motifs in HVRs may prove to be useful. As more allelic profiles become available for strains that are found in hosts belonging to different taxa, occur in different geographic regions, and are subjected to different phenotypes, the potential of the MLST database for revealing ecological, evolutionary, and functional patterns should increase.

**ACKNOWLEDGMENTS**

We thank the following individuals for providing arthropods for this study: D. Bouchon, K. Christianson, K. Dyer, M. Hoy, S. Dobson, R. Hesse, D. Marshall, J. Janiak, F. Jiggins, N. Lo, J. Marshall, J. Raagon, T. Sasaki, D. Shoemaker, R. Stouthamer, M. Wade, D. Windsor, D. Zeh, J. Zeh, N. Ayoub, and M. Collin. We thank R. Nene, C. Westbrook, S. Sinkins, and R. Edwards for assistance with this project. This work was supported by funds from U.S. National Science Foundation grant EF-0528363 to J.H.W. Development of the database software was funded by the Wellcome Trust and the European Union.

**REFERENCES**

1. Baldo, L., J. D. Bartos, J. H. Werren, C. Bazzocchi, M. Casiraghi, and S. Panelli. 2002. Different rates of nucleotide substitutions in *Wolbachia* endosymbionts of arthropods and nematodes: arms race or host shifts? Parasitology 125:179–187.

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

3. Baldo, L., S. Bordenstein, J. J. Wernegreen, and J. H. Werren. 2006. Widespread recombinant throughout *Wolbachia* genomes. Mol. Biol. Evol. 23:437–449.

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

5. Bordenstein, S. R., and J. H. Werren. 1998. Effects of A and B *Wolbachia* and host genotype on interspecies cytoplasmic incompatibility in *Nasonia*. Genetics 148:1833–1844.

6. Casiraghi, M., S. R. Bordenstein, L. Baldo, N. Lo, T. Beninati, J. J. Wernegreen, J. H. Werren, C. Bandi. 2005. Phylogeny of Wasps (*Acari: Tetranychidae*) in Japan. Heredity 91:245–250.

7. 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.

8. Dunning Hotopp, J. C., M. Lin, R. Madupu, J. Crabtree, S. V. Angiuoli, J. Eisen, R. Seshadri, Q. Ren, M. A. T. R. Utterback, S. Smith, M. Lewis, H. Khouri, C. Zhang, H. Niu, Q. Lin, N. Ohashi, N. Zhi, W. Nelson, L. M. Brinkac, R. J. Dodson, M. J. Rosovitz, J. Sundaram, S. C. Daughtery, T. Davidsen, A. S. Durkin, M. Gwinn, D. H. Haft, J. D. Selengut, S. A. Sullivan, N. Zafrar, L. Zhou, F. Benahmed, H. Forberger, R. Halpin, S. Mulligan, J. Robinson, O. White, Y. Rikihisa, and H. Tettelin. 2006. Comparative genomics of emerging human ehrlichiosis agents. PLoS Genet. 2:e21.

9. Foster, J., M. Manatra, I. Kamal, J. Ware, K. Makarova, N. Ivanova, A. Bhattacharya, V. Kapatral, S. Kumar, J. Posfai, T. Vincez, J. Ingram, L. Moran, A. Lapidus, M. Omelchenko, N. Kyrpides, E. Ghedin, S. Wang, E. Goltsman, V. Joukov, O. Ostrovskaya, K. Tsukerman, M. Mazur, D. Comb, E. Koonin, and B. Saltko. 2005. *Wolbachia* genome of *Bretia malayi*: endosymbiont evolution within a human pathogenic nematode. PLoS Biol. 3:e21.

10. Gilbert, J., C. K. Nfon, B. L. Makepepe, L. M. Njomgeita, I. M. Hastings, K. M. Pfarr, A. Renz, V. N. Tanya, and A. J. Trees. 2005. Antibiotic chemotheraphy of onchocerciasis: in a bovine model, killing of adult parasites requires a sustained depletion of endosymbiotic bacteria (*Wolbachia* species). J. Infect. Dis. 192:1483–1493.

11. Goodacre, S. L., O. Y. Martin, C. F. Thomas, and G. M. Hewitt. 2006. *Wolbachia* and other endosymbiont infections in spiders. Mol. Ecol. 15:517–527.

12. Gottlieb, E., S. K. Olu, L. N. Khan, K. O. Iwu, and B. S. Oba. 2003. The *Wolbachia* endosymbiont genome of *N. felifi*: phylogenetic inferences and potential role in arbovirus transmission. PLoS Genet. 5:221.

13. 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.

14. Guimellaud, T., N. Pasteur, and F. Rouset. 1997. Contrasting levels of variability between cytoplasmic genomes and incompatibility types in the mosquito *Culex pipiens*. Proc. R. Soc. Lond. B 264:245–251.

15. 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.

16. Hertig, M., and S. B. Wolbach. 1924. Studies on rickettsia-like microorganisms in insects. J. Med. Res. 44:329–374.

17. Hoffmann, A. A., and M. Turelli. 1997. Cytoplasmic incompatibility in insects, pp. 42–80. In S. L. O’Neill, J. H. Werren, and A. Hoffman (ed.), *Influential passengers*. Oxford University Press, New York, N.Y.

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

19. Hueslenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755.

20. 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.

21. Jayaprakash, 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.
21. Jiggins, F. M., G. D. Hurst, and Z. Yang. 2002. Host-symbiont conflicts: positive selection on an outer membrane protein of parasitic but not mutualistic Rickettsiaceae. Mol. Biol. Evol. 19:1341–1349.
22. Jolley, K. A., M. S. Chan, and M. C. Maiden. 2004. mlstDBNet—distributed multi-locus sequence typing (MLST) databases. BMC Bioinformatics 5:1–86.
23. Jolley, K. A., E. J. Feil, M. S. Chan, and M. C. Maiden. 2001. Sequence type analysis and recombinational tests (START). Bioinformatics 17:1230–1231.
24. Kikuchi, Y., and T. Fukatsu. 2003. Diversity of Wolbachia endosymbionts in heteropteran bugs. Appl. Environ. Microbiol. 69:6082–6090.
25. Kyei-Poku, G. K., D. D. Colwell, P. Coghlin, B. Benkel, and D. K. Floate. 2005. On the ubiquity and phylogeny of Wolbachia in lice. Mol. Ecol. 14:285–294.
26. Lo, N., M. Casiraghi, E. Salati, C. Bazzocchi, and C. Bandi. 2005. Wolbachia infections and cytoplasmic incompatibility. Genetica 122:1428–1433.
27. Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, and M. C. Maiden. 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21:260–262.
28. Martin, D., C. Williamson, and D. Posada. 2005. RDP2: recombination approaches over likelihood ratio tests. Syst. Biol. 53:413–422.
29. Nirgianaki, A., G. K. Banks, D. R. Frohlich, Z. Veneti, H. R. Braig, T. A. Achtman, and D. R. E. Brownlie. 2004. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian methods. Sinauer Associates, Sunderland, Mass.
30. Posada, D., and T. J. Price. 2008. ProModel: network construction and phylogeny inference using a likelihood based network construction and pruning approach. Syst. Biol. 57:570–583.
31. Punja, Z. G., S. M. Rowley, and E. A. McGraw. 2004. Wolbachia pipientis in Australian spiders. Curr. Microbiol. 49:208–214.
32. Shimoda, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16:1114–1116.
33. Shoemaker, D. D., C. A. Machado, D. Molbo, J. H. Werren, D. M. Windsor, and E. A. Herre. 2002. The distribution of Wolbachia in fig wasps: correlations with host phylogeny, ecology and population structure. Proc. R. Soc. Lond. B 269:2257–2267.
34. 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.
35. Stouthamer, R., J. A. J. Breeuwer, R. F. Luck, and J. H. Werren. 1993. Molecular identification of microorganisms associated with parthenogenesis. Nature 361:68–68.
36. Stouthamer, R., J. A. J. Breeuwer, and G. D. Hurst. 1999. Wolbachia pipientis: microbial manipulator of arthropod reproduction. Annu. Rev. Microbiol. 53:71–102.
37. Sutton, G., O. White, M. Adams, and A. Kerlavage. 1995. TIGR Assembler: a new tool for assembling large shotgun sequencing projects. Genome Sci. Technol. 1:9–19.
38. Swoford, D. L. 2000. PAUP*: phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, Mass.
39. Tatusov, R. L., E. V. Koonin, and D. J. Lipman. 1997. A genomic perspective on protein families. Science 278:631–637.
40. Tatusov, R. L., N. D. Fedorova, J. D. Jackson, A. R. Jacobs, B. Kiryutin, E. V. Koonin, D. M. Krylov, R. Mazumder, S. L. Mekhedov, A. N. Nikolayskaya, B. S. Rao, S. Smirnov, A. V. Sverdlov, S. Vasudevan, Y. I. Wolf, J. J. Yin, and D. A. Natale. 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4:41.
41. 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.
42. Urwin, R., and M. C. Maiden. 2003. Multi-locus sequence typing: a tool for global epidemiology. Trends Microbiol. 11:479–487.
43. van Meer, M. M. M., J. Witteveeld, and R. Stouthamer. 1999. Phylogeny of the arthropod endosymbiont Wolbachia based on wsp gene. Insect Mol. Biol. 8:399–408.
44. Higgins, G. J. H. 1997. Biology of Wolbachia. Annu. Rev. Entomol. 42:587–609.
45. Werren, J. H., and D. M. Windsor. 2000. Wolbachia infection frequencies in insects: evidence of a global equilibrium? Proc. R. Soc. Lond. B 267:1277–1285.
46. Werren, J. H., D. M. Windsor, and L. R. Guo. 1995. Distribution of Wolbachia among neotropical arthropods. Proc. R. Soc. Lond. B 262:197–204.
47. Wittman, R., and J. J. Bier. 1960. Wolbachia strains using wsp gene sequences. Proc. R. Soc. Lond. 151:57–63.
48. Wu, M., L. V. Sun, J. Vamathevan, M. Riegler, R. Deboy, J. C. Brownlie, E. A. McGraw, W. Martin, C. Esser, N. Ahmadiannejad, C. Wiegand, R. Madupu, M. J. Beanan, L. M. Brinkac, S. C. Daugherty, A. S. Durkin, J. F. Kolonay, W. C. Nielsen, Y. Mohamoud, P. Lee, K. Berry, M. B. Young, T. Utterback, J. Weidman, W. C. Nierman, K. E. Nelson, H. Tettelin, S. L. O’Neill, and J. A. Eisen. 2004. Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: a streamlined genome overrun by mobile genetic elements. PLoS Biol. 2:69.
49. Zabalon, S., M. Riegler, M. Theodorakopoulou, C. Stauffer, C. Savakis, and K. Bourtzis. 2004. Wolbachia-induced cytoplasmic incompatibility as a means for insect pest population control. Proc. Natl. Acad. Sci. USA 101:15042–15045.
50. 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 265:509–515.