Towards a *Wolbachia* Multilocus Sequence Typing system: discrimination of *Wolbachia* strains present in *Drosophila* species

Charalampos Paraskevopoulos[1], Seth R. Bordenstein[2], Jennifer J. Wernegreen[2], John H. Werren[3] and Kostas Bourtzis[1]

[1]Department of Environmental and Natural Resources Management, University of Ioannina, 2 Seferi St., 30100, Agrinio, Greece

[2]Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, The Marine Biological Laboratory, Woods Hole, MA 02543, USA

[3]Department of Biology, University of Rochester, Rochester, NY 14627, USA

Correspondence to: Kostas Bourtzis, kbourtz@cc.uoi.gr , +302641074114, fax. +302641039571

Running title: *Wolbachia* MLST

This work was partially supported by intramural funds of the University of Ioannina to K. Bourtzis, by grants to J.J. Wernegreen from the National Institutes of Health (R01 GM62626-01) and the NASA Astrobiology Institute (NNA04CC04A), and to J.H. Werren and J.J. Wernegreen from the National Science Foundation (EF-0328363). This work was performed in part while S.R. Bordenstein held a National Research Council Research Associateship Award at the Marine Biological Laboratory.

ABSTRACT

Among the diverse maternally inherited symbionts in arthropods, *Wolbachia* are the most common and infect over 20% of all species. In a departure from traditional genotyping or phylogenetic methods relying on single *Wolbachia* genes, the present study represents an initial Multilocus Sequence Typing (MLST) analysis to discriminate closely related *Wolbachia pipientis* strains, and additional data on sequence diversity in *Wolbachia*. We report new phylogenetic characterization of four genes (*aspC*, *atpD*, *sucB* and *pdhB*), and provide an expanded analysis of markers described in previous studies (16S rDNA, *ftsZ*, *groEL*, *dnaA* and *gltA*). MLST analysis of the bacterial strains present in sixteen different *Drosophila*-*Wolbachia* associations detected four distinct clonal complexes that also corresponded to maximum-likelihood identified phylogenetic clades. Among the sixteen associations analyzed, six could not be assigned to MLST clonal complexes and were also shown to be in conflict with relationships predicted by maximum-likelihood phylogenetic inferences. The results demonstrate the discriminatory power of MLST for identifying strains and clonal lineages of *Wolbachia* and provide a robust foundation for studying the ecology and evolution of this widespread endosymbiont.

INTRODUCTION

Bacteria belonging to the genus *Wolbachia* are a group of intracellular alpha proteobacteria (order Rickettsiales) that are maternally inherited and occur in numerous arthropod and filarial nematode species [4, 50, 54]. This group has attracted increasing interest for its induction of reproductive parasitism with implications for speciation [6], sex determination [27, 44], changes in sexual selection [32], and its applied significance for the control of arthropod and nematode pest populations of agricultural and medical importance [51, 58, 60].

Because phenotypic methods for bacterial strain determination cannot be applied to this intracellular bacterium, research efforts have focused on molecular systematic methods based on DNA sequencing and phylogenetics. To date, the 16S rDNA, *ftsZ*, *wsp*, *groEL*, *dnaA* and *gltA* genes have been characterized and used for phylogenetic studies showing that *Wolbachia* strains are clustered into eight divergent clades that are described as supergroups A-H [4, 5, 11, 12, 33, 40, 49, 56, 61]. Efficient methods for intragroup strain typing are more limited. The *Wolbachia* surface protein (*wsp*) gene has proved to be the fastest evolving and has been

extensively used for intragroup phylogenetic studies. However, *Wolbachia* is prone to high rates of recombination, especially within supergroups, and single gene phylogenetics are unreliable for resolving close relationships [2, 3, 7, 30, 55]. Taking a new approach to strain typing, Riegler and colleagues reported a number of polymorphic markers, such as size polymorphisms for IS5 insertion sites or minisatellites and the orientation of a chromosomal inversion, to detect and discriminate among five different *Wolbachia* variants present in *D. melanogaster* natural populations and laboratory stocks [43].

Given the observed presence of multiple, closely related *Wolbachia* strains in a single host group, we aimed to develop a Multilocus Sequence Typing (MLST; www.mlst.net) approach for *Wolbachia* strain discrimination that takes into account, and makes use of, data derived from multiple gene sequences. This method was first described by Maiden and colleagues [34] as a tool for strain discrimination of *Neisseria meningitis* strains and has since been used extensively to distinguish strains of pathogenic bacteria [18, 52, 62]. MLST is based on direct nucleotide sequencing of a target locus, to which a unique and arbitrary allele number is assigned. Upon sampling multiple target genes, the combination of allele numbers for each isolate is defined as the allelic profile. Each unique allelic profile is subsequently assigned an arbitrary number that is known as the Sequence Type (ST). Isolates that share at least five out of seven in their allelic profile can generally be considered as members of the same clonal complex [19]. MLST is also particularly appropriate for evolutionary studies, as variation is quantitative and the selection criteria for MLST loci often match criteria for phylogenetic markers. The most appropriate MLST loci are housekeeping loci, which: (i) are ubiquitous within the population; (ii) usually encode proteins which are essential for central metabolism; (iii) typically evolve at a moderate rate, (iv) are subject to purifying selection, although genes under positive selection have been used in the past for typing purposes (see below). The *ftsZ*, *groEL*, *dnaA* and *gltA* genes that have been used in *Wolbachia* phylogenetic studies are appropriate MLST loci, since these are housekeeping genes that meet all the criteria mentioned above. However, this is not the case for 16S rDNA and *wsp* genes. 16S rDNA evolves too slowly to be useful in MLST. At the other extreme, the hyper-variable *wsp* gene likely experiences positive selection [29] and recombination [2, 3, 30, 55], so that its utilization in an MLST scheme may be problematic. These caveats aside, 16S rDNA has been used in MLSTs as a reference point for other housekeeping genes [15] or even as a marker to differentiate strains [62].

Research on *Wolbachia* biology demands the ability to distinguish closely related strains. It is also important that researchers have a uniform nomenclature to refer to distinct strains, to avoid strain synonymy, widen the accessibility of *Wolbachia* literature, and provide a solid foundation for understanding the extent of horizontal transmission of *Wolbachia* between hosts and the evolution of phenotype switches within this labile endosymbiont. Despite these needs, it remains unclear whether an MLST scheme can be applied to distinguish *Wolbachia* or other maternally inherited bacteria. In this study, we report the characterization of four novel genetic markers (*aspC*, *atpD*, *sucB* and *pdhB*) and an expanded analysis of markers described in previous *Wolbachia* studies (16S rDNA, *ftsZ*, *groEL*, *dnaA* and *gltA*). Our primary objective is to evaluate all common genetic markers in *Wolbachia*, including 16S rDNA, in terms of the criteria for appropriate MLST loci. We then evaluate data from clearly appropriate MLST loci, aiming to demonstrate the value of this approach in discriminating closely related *Wolbachia pipientis* strains, from both A- and B-supergroups, infecting Drosophila species. More broadly, we contribute to a growing sequence database from which researchers can select the most informative genes for strain designation.

METHODS

***Wolbachia* isolates.** A total of sixteen *Wolbachia* strains were used for this study: *w*Mel, *w*MelCS and *w*Melpop harboured by *Drosophila melanogaster*; *w*Ri, *w*Ha, *w*Au, *w*No and *w*Ma harboured by *D. simulans*; *w*Dau harboured by *D. auraria*; *w*Sh harboured by *D. sechellia*; *w*Ana harboured by *D. ananassae*; *w*Tei harboured by *D. teissieri*; *w*Mau harboured by *D. mauritiana*; *w*Yak harboured by *D. yakuba*; *w*San harboured by *D. santomena*; and *w*Din harboured by *D. innubila*. All strains have been previously characterized by various molecular phylogenetic reports, as well as their abilities to induce cytoplasmic incompatibility, male killing or virulent phenotypes [8, 9, 17, 22, 25, 26, 28, 35, 36, 38-41, 59]. Bacterial DNA was extracted using the DNeasy Tissue Kit (Qiagen) according to the manufacturer's instructions.

Loci selection and primer design. Taking under consideration the two available *Wolbachia* genomes (*w*Mel [57] and *w*Bm [20]), the genomes of the closely related bacterial species of *Anaplasma marginale* [10] and *Rickettsia prowazekii* [1] as well as preliminary work on a larger set of candidate genes, four genes were selected for further analysis because they gave adequate PCR amplification for different *Wolbachia* supergroups: *sucB*

(dihydrolipoamide succinyltransferase), *aspC* (aspartate aminotransferase), *atpD* (ATP synthase) and *pdhB* (E1 component of the pyruvate dehydrogenase complex). Combined with five genes developed as markers in previous studies (16S rDNA, *dnaA, ftsZ, groEL* and *gltA*), a total of nine loci were sampled here. The slowly-evolving 16S rDNA gene was included as a reference point for comparison with other markers and to evaluate its performance in an MLST scheme. The average distance between these nine genes was 144,693bp (median of 138,406bp). *suc*B, *asp*C, *atp*D and *pdh*B amplifications and sequences were obtained using the following primers: sucB358F, 5'-AAAGGRACTGGYATGGGARG-3' and sucB981R, 5'-TGHGGAGGRTTWATWATCGG-3'; aspC49_F, 5'-ATYGCTGTRACYGATAAGGYAA-3'; aspC1134R, 5'-AGARGTWGCATAAGARATTCTRA-3'; aspC559F, 5'-GCRCCARTATTGCTTGARTATCC-3' and aspC578R, 5'-GATAYTCAAGCAATAYTGGYGCT-3'; atpD242F, 5'-ATAYAGTKCGTTGTATTGCTATG-3', atpD1210R, 5'-CWTCAGAYAGYTCATCCATAC-3', atpD653F, 5'-AAGGTAAYGATCTTTAYCAYGA-3' and atpD676R, 5'-TCRTGRTAAAGATCRTTACCTT-3'; pdhB86F, 5'-ARGAAGTTGCVGARTATSAWGG-3' and pdhB812R, 5'-GCAAAWRRCCAWCCTTCTTCTA-3'. Primers characterised in previous studies were used for amplifications and sequences of the 16S rDNA [40], *dnaA* [12], *ftsZ* [21], *groEL* and *gltA* [11] gene fragments.

PCR amplification and sequencing. Gene fragments were amplified with the following PCR conditions: 1μL DNA sample, 2μL of 10x reaction buffer (Promega), 0.8μL $MgCl_2$ (50mM), 0.1μL dNTP mix (25mM each), 1μL forward and reverse primer (10pmole μL^{-1}), 0.1μL (0.5 units) of Taq polymerase (Promega) and 14μL H_2O, total of 20μL PCR reaction volume. PCR amplification for *sucB*, *aspC*, *atpD* and *pdhB* was performed on MJ Research Thermal Cycler using the following thermal profiles: 1 cycle (94°C for 5 min), 35 cycles (94°C for 1 min, 55°C for 1 min, 72°C for 2 min) and 1 cycle (72°C for 10 min). PCR amplification for 16S rDNA, *ftsZ*, *dnaA*, *gltA* and *groEL* was performed as described previously [11, 12, 21, 40]. The Qiagen® Nucleotide Removal Kit was then used to purify the reaction products. Sequencing was performed by the MACROGEN company. Double-stranded sequencing was performed for at least three individuals from each strain. All individual gene sequence data for each strain were identical. The nucleotide sequences of the fragments of all *Wolbachia* genes and their alleles (132 sequences) reported in this study have been deposited in GenBank under accession numbers DQ235275-DQ235407.

Data analyses. Multiple sequence alignments were performed using BIOEDIT 7.0.4.1 software [23], confirmed by visual inspection, and edited manually. For each of the nine datasets, sequences with at least one nucleotide difference were assigned arbitrary allele numbers. Each *Wolbachia* isolate was then assigned an allelic profile designated by nine numbers (one per locus). The DnaSP (version 4.10.2) software [47] was used to calculate pairwise nucleotide diversity (π) and the proportion of segregating sites. The START program [31] was used to calculate the number of polymorphic sites, the % GC content and to perform several MLST statistical analyses testing for recombination and selection, such as the Sawyer's run test (10,000 trials), the index of association (I_A) the homoplasy test and the pairwise ratios of nonsynonymous substitutions to synonymous substitutions (d_N/d_S).

Before phylogenetic analyses, the model of sequence evolution for each dataset was estimated and selected via likelihood ratio tests in the program Modeltest 3.06 [42]. Maximum likelihood (ML) heuristic searches were performed using 500 random taxon addition replicates with tree bisection and reconnection (TBR) branch swapping. ML bootstrap support was determined using 500 bootstrap replicates, each using 10 random taxon addition replicates with TBR branch swapping. Searches were performed in parallel on a Beowulf cluster using custom software with PAUP version 4.0b10. To confirm the ML results, the concatenated dataset was also analysed with Bayesian analysis, using Mr.Bayes version 3.1.1 [45]. The likelihood model was set to the GTR with a proportion of the sites invariable and the rest drawn from a gamma distribution ("lset Nst=6 rates=invgamma"). Four independent runs were performed, each with 5,000,000 generations and four chains. Trees were sampled every 100 generations, resulting in 50,000 trees per run. The first 40% of these trees considered the 'burn-in' and discarded. Posterior probabilities were estimated from the consensus of the remaining 120,000 trees (30,000 for each of the four runs).

RESULTS AND DISCUSSION

Gene sequences and gene diversity. Diversity estimates and other sequence statistics are presented in Table 1. The size of the gene fragments analysed ranged from 402 bp (*dnaA*) to 972bp (*aspC*) with an average for all genes of 702 bp, and the percentage of gene lengths analysed ranged from 29.13% to 81.61% (54.99% average). Nucleotide diversity (π) ranged between 4.91% (*aspC*) to 0.69% (16S rDNA), with an average of 3.37%. The number of alleles per locus ranged from twelve (*pdhB*) to five (*dnaA* and 16S rDNA), with an average of seven.

The number of polymorphic sites per locus ranged from two (*aspC*, *dnaA*, *ftsZ*, 16S rDNA) to twenty (*sucB*), with an average of seven. G+C content ranged from 35.62% (*dnaA*) to 46.15% (16S rDNA), with an average of 38.42%. In terms of nucleotide diversity and number of polymorphic sites, the *aspC* and 16S rDNA genes exhibited the highest and lowest values respectively. This was not the case though for the number of alleles per locus, where only seven alleles were observed for the *aspC* locus.

Recombination and selection. The results of the statistical testing for recombination and selection are shown in Table 1. The Sawyer's run test did not determine any regions of sequence pairs sharing more consecutive identical polymorphic sites than expected by chance. In all cases the I_A significantly exceeded zero, indicating linkage disequilibrium due to constraints on recombination. Across all strains, I_A was found to be 1.81 and when calculated within supergroups A and B, to ascertain that the linkage disequilibrium is not a by-product of including both supergroups, it was 0.851 and 2.9 respectively. This indicates that some *Drosophila Wolbachia* have a clonal population structure, where recombination is relatively limited both within and between the supergroups. The homoplasy test could not be performed because it requires at least 10 informative sites per gene, whereas datasets obtained here did not exceed three informative sites per gene (data not shown). A d_N/d_S ratio ≤ 1 was observed for all loci, indicating that genes where not subject to positive selection. Based on these results, we infer that recombination is constrained within this set of *Wolbachia* strains and genes examined here, and that the few (if any) recombination events will not influence MLST results. Furthermore, genes are subject to purifying rather than positive selection.

Genetic relatedness of *Wolbachia* strains. Both the MLST dendrogram (Fig. 1a) constructed from the matrix of pairwise allelic differences among sixteen *Wolbachia* strains and the ML phylogenetic tree (Fig. 1b) of the concatenated data set (6319bp, 16 strains, 9 loci) identified two major clusters that correspond to supergroups B (*w*No, *w*Ma and *w*Mau) and A (the remaining 13 isolates). This was also confirmed by the allelic profile data, as the two supergroups differ at all nine loci (i.e., share no alleles). Within supergoup A, the MLST dendrogram and ML phylogeny share several groupings. These include: *w*Mel, *w*MelCS and *w*MelPop strains; *w*Dau, *w*Ri and *w*Ana strains; and *w*Tei, *w*Yak and *w*San strains. Notably, the ML tree groups together *w*Ha and *w*Sh, whereas the dendrogram does not. Within supergroup B, the only cluster and potential clonal complex observed in the MLST tree is the B_1 cluster, which consisted of the *w*Ma and *w*Mau strains. However within supergroup

B, relationships among the *w*No, *w*Ma and *w*Mau strains were poorly resolved and differ between the ML phylogeny and MLST dendrogram. Adopting the general criterion used in MLST studies of *Neisseria meningitides*, (when isolates share five out of seven alleles, they can be regarded as members of the same clonal complex) [34], a *Wolbachia* clonal complex is defined as a cluster of bacterial isolates sharing in this case at least six alleles out of a total of nine sampled. Under this criterion, four clonal complexes were detected: complex A_1 (*w*Mel, *w*MelCS and *w*MelPop), complex A_2 (*w*Dau and *w*Ana), complex A_3 (*w*Tei, *w*Yak and *w*San) and complex B_1 (*w*Ma and *w*Mau). The remaining six isolates (*w*Ri, *w*Sh, *w*Ha, *w*Din, *w*Au and *w*No) did not group with any of the above four clonal complexes.

Biological features of *Wolbachia* clonal complexes. The clonal complex A_1 consists of the *w*Mel, *w*MelCS and *w*MelPop strains that share alleles at seven loci. Those strains infect *D. melanogaster* natural populations and laboratory strains [9, 24, 39]. Based on our data these strains clearly share a common ancestor, share identical *wsp* sequences [61], but can be discriminated by sampling *dnaA* and *pdhB*. All three strains are known to be CI-inducers [9, 24, 36, 37], and the *w*MelPop strain has also been shown to reduce the life span of its hosts, *D. melanogaster* and *D. simulans* [36, 39]. It was recently shown that *Wolbachia* in *D. melanogaster* could be distinguished by size polymorphism of IS5 insertion sites, the VNTR loci, and the orientation of a chromosomal inversion [43].

The clonal complex A_2 consists of the *w*Ana and *w*Dau strains which also share alleles at seven loci. The *w*Ri strain, which has identical *wsp* gene sequences with *w*Ana and *w*Dau [61], was also a candidate of this complex as observed in the ML and MLST trees. However, despite the fact that it shared six alleles with *w*Dau, it shared only five with *w*Ana. The *w*Ana, *w*Dau and *w*Ri strains naturally infect *D. simulans*, *D. ananassae* and *D. auraria* respectively and are known to be CI-inducers [9, 26]. Based on our data these strains clearly share a common ancestor, but can be discriminated by using *sucB* and *pdhB* gene sequences.

The clonal complex A_3 consists of the *w*Yak, *w*Tei and *w*San strains, sharing alleles at six loci. The *w*Yak, *w*Tei and *w*San strains naturally infect *D. yakuba*, *D. teissieri* and *D. santomea* respectively and are all unable to induce CI in their native hosts, however, they are able to fully rescue the *w*Ri modification [59]. Based on our

data these strains share a common ancestor, have identical *wsp* sequences [59] and can be discriminated by using *ftsZ*, *gltA* and *aspC* gene sequences.

The *w*Au and *w*Din strains do not clearly group with any of the above complexes or clades in the ML tree (Fig. 1). These two strains share only five alleles (*sucB*, *ftsZ*, *pdhB*, *gltA* and 16S rDNA), so they cannot be regarded as members of the same clonal complex. The *w*Au and *w*Din strains naturally infect *D. simulans* and *D. innubila* respectively [17, 25]. The *w*Au strain is the only known *mod*$^-$ *resc*$^-$ *Wolbachia* strain. The *w*Din strain has been recently shown to induce male killing in its native host [17]. Notably, the *pdhB* gene, the locus that exhibited the highest number of alleles (twelve), was identical for *w*Au and *w*Din strains. By contrast, the two differ at *dnaA*, which had relatively few alleles (five). This exceptional pattern might be explained by lateral transfer of *pdhB* between *w*Au and *w*Din strains, a rare instance of gene transfer among these strains.

An unexpected result of our study was the finding that *w*Ha, which infects *D. simulans* [41], and *w*Sh, which infects *D. sechellia* [9, 22], do not belong to the same clonal complex, and in fact share only one allele (16S rDNA). Those two strains have identical *wsp* and 16S rDNA sequences [14] and group together on the ML phylogeny of concatenated data (Fig. 3) suggesting they are sister lineages. Furthermore, the close relationship of these strains is consistent with the similar CI properties of the two [14]. In this study, we identified several markers that discriminate the two strains; however, considering their similar biological features and the ML phylogeny, we can conclude that the MLST results do not accurately reflect the phylogenetic relationships of these two strains.

The clonal complex B_1 consists of *Wolbachia* strains *w*Ma and *w*Mau, sharing alleles at seven loci. The *w*Ma and *w*Mau strains naturally infect *D. simulans* and *D. mauritiana* respectively [22, 28, 46]. Neither strain induces CI [8, 13, 28, 38, 46, 53], however, both rescue the modification induced by the *w*No strain [8, 38]. The close relationship of these strains confirms previous studies of their evolutionary history [46]. They also have identical *wsp* sequences [13, 61], but can be discriminated by using *ftsZ*, and *pdhB* gene sequences. The *w*No strain is shown, in both the ML and MLST trees, as closely related to the *w*Ma and *w*Mau strains. Yet, interpretation of the allelic profile data shows that those three strains share alleles only in three genes (*aspC*, *atpD* and *gltA*). Several studies have suggested that *w*No, *w*Ma and *w*Mau are quite closely related strains which may only differ

in their ability to induce CI: *w*No is a *mod*$^{+}$ *resc*$^{+}$ strain while *w*Ma and *w*Mau have been considered as *mod*$^{-}$ *resc*$^{+}$ strains [8, 13, 22, 28, 38]. The MLST allelic profile data indicate that *w*No differs from the *w*Ma and *w*Mau strains raising the hypothesis that it should be considered as a distinct and rather unrelated strain. However, the ML tree and the fact that the *w*No imprint can be rescued by both *w*Ma and *w*Mau strains [8, 38] clearly show that all three strains are closely related.

MLST versus other typing methods. Several markers have already been, or are going to be, generated for *Wolbachia* through complete and ongoing genome sequencing projects [20, 48, 57], including chromosomal inversions, minisatellites and transposons. Such markers have already been used to type *Wolbachia* strains [16, 43]. Undoubtedly, markers such as transposons and phage genes can be informative in understanding *Wolbachia* biology and ecology; however they may not be abundant along strains spanning all known supergroups [20]. It has also to be noted that most of the above mentioned studies relied upon single-gene / single-sequence approaches to discriminate among strains. The MLST method is an obvious good alternative typing method, mainly due to its multilocus power, and offers the advantage that selection of target loci does not require the complete genome data for the strain examined. The present study shows that the development of a *Wolbachia* MLST approach is feasible and provides novel genetic markers useful for such an approach. One potential limitation to the use of MLST would be the case in which more than one type of *Wolbachia* infect a host species. In this case, it would be difficult to assign sequences to a particular strain; however, this issue could be resolved by designing single infection- specific primers. We also observed the high discriminatory power of the MLST, as it discriminated closely related strains that show very similar, sometimes identical, biological features. It is indeed a very sensitive method but also harbors potential limitation. The first one is that sequence data must be very accurate and confirmed by repeated sequencing of a given marker. The second limitation, which was also observed in the current study, is that allelic profile data might not reflect the phylogeny of the strains examined. The original sequence data can, however, be used in molecular phylogenetic studies.

In conclusion, the present study illustrates MLST as a potential uniform *Wolbachia* typing system, and in doing so, enriches the *Wolbachia* sequence database with novel genes, demonstrates the discriminatory power of closely related *Wolbachia* strains using multiple genes, and identifies clonal complexes of *Wolbachia* that either

reflect patterns of symbiont-host codivergence or lateral transfer of *Wolbachia* bacteria among ecologically-associated hosts.

ACKNOWLEDGMENTS

The authors would like to thank Laura Baldo for her constructive comments and suggestions. The *D. simulans* (*w*Ma) and *D. innubila* (*w*Din) were kindly provided by Prof. Bill Ballard and Prof. John Jaenike respectively.

LITERATURE CITED

1. Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, Podowski RM, Naslund AK, Eriksson AS, Winkler HH, Kurland CG (1998) The genome sequence of Rickettsia prowazekii and the origin of mitochondria. Nature 396:133-140.
2. Baldo L, Bordenstein S, Wernegreen JJ, Werren JH (2006) Widespread Recombination Throughout Wolbachia Genomes. Mol Biol Evol 23:437-449.
3. Baldo L, Lo N, Werren JH (2005) Mosaic nature of the wolbachia surface protein. J Bacteriol 187:5406-5418.
4. Bandi C, Anderson TJ, Genchi C, Blaxter ML (1998) Phylogeny of Wolbachia in filarial nematodes. Proc Biol Sci 265:2407-2413.
5. Bordenstein S, Rosengaus RB (2005) Discovery of a novel wolbachia supergroup in isoptera. Curr Microbiol 51:393-398.
6. Bordenstein SR, O'Hara FP, Werren JH (2001) Wolbachia-induced incompatibility precedes other hybrid incompatibilities in Nasonia. Nature 409:707-710.
7. Bordenstein SR, Wernegreen JJ (2004) Bacteriophage flux in endosymbionts (Wolbachia): infection frequency, lateral transfer, and recombination rates. Mol Biol Evol 21:1981-1991.
8. Bourtzis K, Dobson SL, Braig HR, O'Neill SL (1998) Rescuing Wolbachia have been overlooked. Nature 391:852-853.
9. Bourtzis K, Nirgianaki A, Markakis G, Savakis C (1996) Wolbachia infection and cytoplasmic incompatibility in Drosophila species. Genetics 144:1063-1073.

10. Brayton KA, Kappmeyer LS, Herndon DR, Dark MJ, Tibbals DL, Palmer GH, McGuire TC, Knowles DP, Jr. (2005) Complete genome sequencing of Anaplasma marginale reveals that the surface is skewed to two superfamilies of outer membrane proteins. Proc Natl Acad Sci U S A 102:844-849.

11. Casiraghi M, Bordenstein SR, Baldo L, Lo N, Beninati T, Wernegreen JJ, Werren JH, Bandi C (2005) Phylogeny of Wolbachia pipientis 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 151:4015-4022.

12. Casiraghi M, Werren JH, Bazzocchi C, Biserni A, Bandi C (2003) dnaA gene sequences from Wolbachia pipientis support subdivision into supergroups and provide no evidence for recombination in the lineages infecting nematodes. Parassitologia 45:13-18.

13. Charlat S, Le Chat L, Mercot H (2003) Characterization of non-cytoplasmic incompatibility inducing Wolbachia in two continental African populations of Drosophila simulans. Heredity 90:49-55.

14. Charlat S, Nirgianaki A, Bourtzis K, Mercot H (2002) Evolution of Wolbachia-induced cytoplasmic incompatibility in Drosophila simulans and D. sechellia. Evolution Int J Org Evolution 56:1735-1742.

15. Christensen H, Kuhnert P, Olsen JE, Bisgaard M (2004) Comparative phylogenies of the housekeeping genes atpD, infB and rpoB and the 16S rRNA gene within the Pasteurellaceae. Int J Syst Evol Microbiol 54:1601-1609.

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

17. Dyer KA, Jaenike J (2004) Evolutionarily stable infection by a male-killing endosymbiont in Drosophila innubila: molecular evidence from the host and parasite genomes. Genetics 168:1443-1455.

18. Enright MC, Knox K, Griffiths D, Crook DW, Spratt BG (2000) Molecular typing of bacteria directly from cerebrospinal fluid. Eur J Clin Microbiol Infect Dis 19:627-630.

19. Feil EJ, Maiden MC, Achtman M, Spratt BG (1999) The relative contributions of recombination and mutation to the divergence of clones of Neisseria meningitidis. Mol Biol Evol 16:1496-1502.

20. Foster J, Ganatra M, Kamal I, Ware J, Makarova K, Ivanova N, Bhattacharyya A, Kapatral V, Kumar S, Posfai J, Vincze T, Ingram J, Moran L, Lapidus A, Omelchenko M, Kyrpides N, Ghedin E, Wang S, Goltsman E, Joukov V, Ostrovskaya O, Tsukerman K, Mazur M, Comb D, Koonin E, Slatko B (2005) The Wolbachia genome of Brugia malayi: endosymbiont evolution within a human pathogenic nematode. PLoS Biol 3:e121.

21. Fukatsu T (1999) Acetone preservation: a practical technique for molecular analysis. Mol Ecol 8:1935-1945.

22. Giordano R, O'Neill SL, Robertson HM (1995) Wolbachia infections and the expression of cytoplasmic incompatibility in Drosophila sechellia and D. mauritiana. Genetics 140:1307-1317.

23. Hall TA (1999) BioEdit: a user-friendly biological sequence allignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95-98.

24. Hoffmann AA (1988) Partial cytoplasmic incompatibility between two Australian populations of Drosophila melanogaster. Entomol Exp Appl 48:61-67.

25. Hoffmann AA, Clancy D, Duncan J (1996) Naturally-occurring Wolbachia infection in Drosophila simulans that does not cause cytoplasmic incompatibility. Heredity 76 (Pt 1):1-8.

26. Hoffmann AA, Turelli M, Simmons GM (1986) Unidirectional incompatibility between populations of Drosophila simulans. Evolution 40:692-701.

27. Huigens ME, Stouthamer R (2003) Parthenogenesis associated with Wolbachia. In: K Bourtzis, T. Miller (eds), Insect Symbiosis. Boca Raton, FL: CRC Press, pp. 247-266.

28. James AC, Ballard JW (2000) Expression of cytoplasmic incompatibility in Drosophila simulans and its impact on infection frequencies and distribution of Wolbachia pipientis. Evolution Int J Org Evolution 54:1661-1672.

29. Jiggins FM, Hurst GD, Yang Z (2002) Host-symbiont conflicts: positive selection on an outer membrane protein of parasitic but not mutualistic Rickettsiaceae. Mol Biol Evol 19:1341-1349.

30. Jiggins FM, von Der Schulenburg JH, Hurst GD, Majerus ME (2001) Recombination confounds interpretations of Wolbachia evolution. Proc Biol Sci 268:1423-1427.

31. Jolley KA, Feil EJ, Chan MS, Maiden MC (2001) Sequence type analysis and recombinational tests (START). Bioinformatics 17:1230-1231.

32. Koukou K, Pavlikaki H, Kilias J, Werren JH, Bourtzis K, Alahiotis SN (2006) Influence of antibiotic treatment and Wolbachia curing on sexual isolation among Drosophila melanogaster cage populations. Evolution in press.

33. Lo N, Casiraghi M, Salati E, Bazzocchi C, Bandi C (2002) How many wolbachia supergroups exist? Mol Biol Evol 19:341-346.

34. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A 95:3140-3145.

35. McGarry HF, Pfarr K, Egerton G, Hoerauf A, Akue JP, Enyong P, Wanji S, Klager SL, Bianco AE, Beeching NJ, Taylor MJ (2003) Evidence against Wolbachia symbiosis in Loa loa. Filaria J 2:9.

36. McGraw EA, Merritt DJ, Droller JN, O'Neill SL (2002) Wolbachia density and virulence attenuation after transfer into a novel host. Proc Natl Acad Sci U S A 99:2918-2923.

37. McGraw EA, Merritt DJ, Droller JN, O'Neill SL (2001) Wolbachia-mediated sperm modification is dependent on the host genotype in Drosophila. Proc Biol Sci 268:2565-2570.

38. Mercot H, Poinsot D (1998) Wolbachia transmission in a naturally bi-infected Drosophila simulans strain from New Caledonia. Entomol Exp Appl 86:97-103.

39. Min KT, Benzer S (1997) Wolbachia, normally a symbiont of Drosophila, can be virulent, causing degeneration and early death. Proc Natl Acad Sci U S A 94:10792-10796.

40. O'Neill SL, Giordano R, Colbert AM, Karr TL, Robertson HM (1992) 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. Proc Natl Acad Sci U S A 89:2699-2702.

41. O'Neill SL, Karr TL (1990) Bidirectional incompatibility between conspecific populations of Drosophila simulans. Nature 348:178-180.

42. Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817-818.

43. Riegler M, Sidhu M, Miller WJ, O'Neill SL (2005) Evidence for a global Wolbachia replacement in Drosophila melanogaster. Curr Biol 15:1428-1433.

44. Rigaud T (1997) Inherited microorganisms and sex determination of arthropod hosts. In: SL O'Neill, A.A. Hoffmann and J.H. Werren (eds), Influential passengers. Oxford: Oxford University Press, pp. 81-101.

45. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

46. Rousset F, Solignac M (1995) Evolution of single and double Wolbachia symbioses during speciation in the Drosophila simulans complex. Proc Natl Acad Sci U S A 92:6389-6393.

47. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496-2497.

48. Salzberg SL, Hotopp JC, Delcher AL, Pop M, Smith DR, Eisen MB, Nelson WC (2005) Serendipitous discovery of Wolbachia genomes in multiple Drosophila species. Genome Biol 6:R23.

49. Schulenburg JH, Hurst GD, Huigens TM, van Meer MM, Jiggins FM, Majerus ME (2000) Molecular evolution and phylogenetic utility of Wolbachia ftsZ and wsp gene sequences with special reference to the origin of male-killing. Mol Biol Evol 17:584-600.

50. Stouthamer R, Breeuwer JA, Hurst GD (1999) Wolbachia pipientis: microbial manipulator of arthropod reproduction. Annu Rev Microbiol 53:71-102.

51. Taylor MJ (2000) Wolbachia bacteria of filarial nematodes in the pathogenesis of disease and as a target for control. Trans R Soc Trop Med Hyg 94:596-598.

52. van Loo IH, Heuvelman KJ, King AJ, Mooi FR (2002) Multilocus sequence typing of Bordetella pertussis based on surface protein genes. J Clin Microbiol 40:1994-2001.

53. Veneti Z, Clark ME, Zabalou S, Karr TL, Savakis C, Bourtzis K (2003) Cytoplasmic incompatibility and sperm cyst infection in different Drosophila-Wolbachia associations. Genetics 164:545-552.

54. Werren JH (1997) Biology of Wolbachia. Annu Rev Entomol 42:587-609.

55. Werren JH, Bartos JD (2001) Recombination in Wolbachia. Curr Biol 11:431-435.

56. Werren JH, Zhang W, Guo LR (1995) Evolution and phylogeny of Wolbachia: reproductive parasites of arthropods. Proc Biol Sci 261:55-63.

57. Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA, Martin W, Esser C, Ahmadinejad N, Wiegand C, Madupu R, Beanan MJ, Brinkac LM, Daugherty SC, Durkin AS, Kolonay JF, Nelson WC, Mohamoud Y, Lee P, Berry K, Young MB, Utterback T, Weidman J, Nierman WC, Paulsen IT, Nelson KE, Tettelin H, O'Neill SL, Eisen JA (2004) Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: a streamlined genome overrun by mobile genetic elements. PLoS Biol 2:E69.

58. Xi Z, Khoo CC, Dobson SL (2005) Wolbachia establishment and invasion in an Aedes aegypti laboratory population. Science 310:326-328.

59. Zabalou S, Charlat S, Nirgianaki A, Lachaise D, Mercot H, Bourtzis K (2004) Natural Wolbachia infections in the Drosophila yakuba species complex do not induce cytoplasmic incompatibility but fully rescue the wRi modification. Genetics 167:827-834.

60. Zabalou S, Riegler M, Theodorakopoulou M, Stauffer C, Savakis C, Bourtzis K (2004) Wolbachia-induced cytoplasmic incompatibility as a means for insect pest population control. Proc Natl Acad Sci U S A 101:15042-15045.

61. Zhou W, Rousset F, O'Neil S (1998) Phylogeny and PCR-based classification of Wolbachia strains using wsp gene sequences. Proc Biol Sci 265:509-515.

62. Zhu Y, Fournier PE, Eremeeva M, Raoult D (2005) Proposal to create subspecies of Rickettsia conorii based on multi-locus sequence typing and an emended description of Rickettsia conorii. BMC Microbiol 5:11.

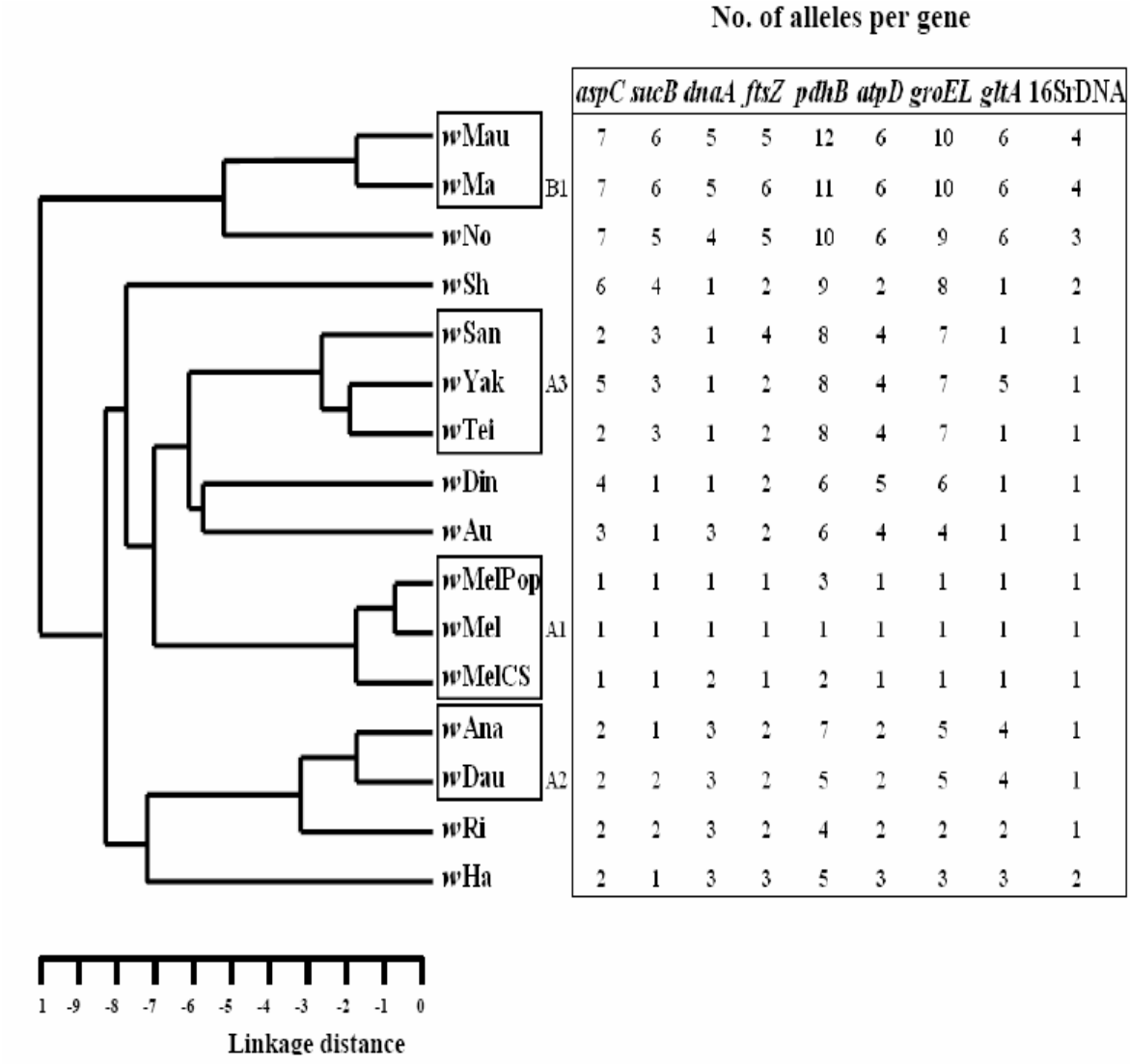
Figure Legend (a): The UPGMA dendrogram of the genetic relatedness among the sixteen *Wolbachia* strains, based on the pairwise differences of their allelic profile. The individually defined allele at each of the nine loci determined an allelic profile for each strain. The linkage distance is shown at the bottom. Strains sharing more than six alleles were characterized as potential members of the same clonal complex and are noted on the dendrogram with the corresponding labels, A_1-A_3 and B_1. **(b):** Phylogenetic tree based on a concatenated data set of nine genes. The maximum-likelihood (ML) best tree is shown with ML bootstrap values and Bayesian posterior probabilities (bold) assigned to their respective branches. The MLST-determined clonal complexes are noted on the tree with the corresponding labels, A_1-A_3 and B_1.

Table 1. Nucleotide sequence variation and statistical tests performed to evaluate recombination and selection in nine *Wolbachia* gene fragments

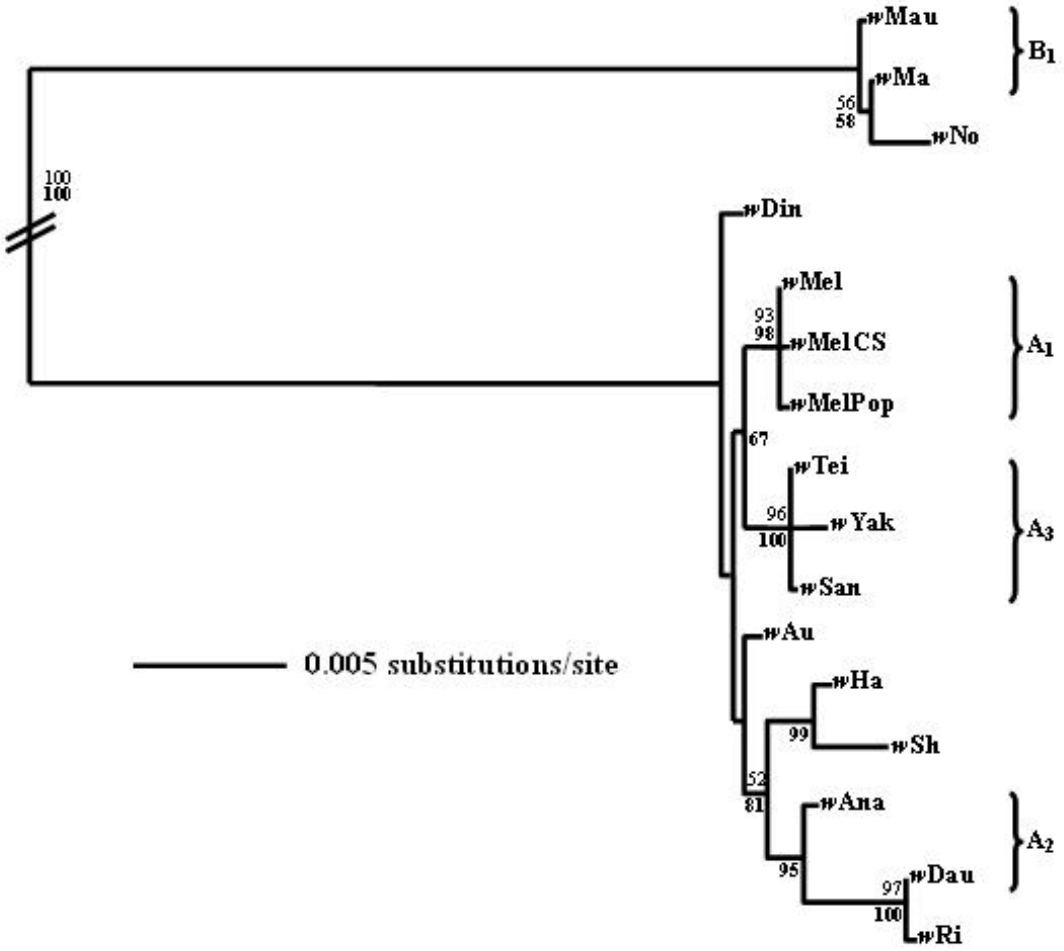
Gene	Fragment Size (bp)[a]	π[b]	No. of Alleles	No. of Polymorphic sites	Avg G+C content (%)	d_N/d_S ratio[c]	Sawyer's test SSCF / MCF[d]
aspC	972 (81.61)	0.04913 (4.91)	7	2	36.77	0.0000	0/0 (1)
sucB	535 (45.73)	0.04436 (4.44)	6	20	37.39	0.1422	1440/12(1)
dnaA	402 (29.13)	0.02898 (2.90)	5	2	35.62	0.1839	0/0 (1)
ftsZ	629 (52.68)	0.04475 (4.48)	6	2	38.69	0.0000	4/2 (1)
pdhB	741 (74.40)	0.03134 (3.13)	12	14	39.86	0.8655	51/2 (1)
atpD	927 (65.05)	0.03048 (3.05)	6	8	38.93	0.1727	106/5 (1)
16S rDNA	691 (47.79)	0.00689 (0.69)	5	2	46.15	-	0/0 (1)
groEL	795 (48.27)	0.04547 (4.55)	10	13	36.19	0.1528	351/7 (1)
gltA	627 (50.24)	0.02217 (2.22)	6	4	36.20	0.8779	0 /0 (1)
Avg	702 (54.99)	0.00337 (3.34)	7	7	38.42		

[a]percentage (%) coverage of the complete gene in parentheses; [b]π: pairwise nucleotide diversity based on the average of all pairwise comparisons, percentage (%) of π in parentheses; [c]pairwise ratio of nonsynonymous/synonymous substitutions, not applicable for the 16S rDNA gene; [d]SSCF: the sum of the squares of condensed fragments; MCF: maximum condensed fragment, *P* values in parentheses

(a)


No. of alleles per gene
aspC sucB dnaA ftsZ pdhB atpD groEL gltA 16SrDNA
wMau 7 6 5 5 12 6 10 6 4
wMa B1 7 6 5 6 11 6 10 6 4
wNo 7 5 4 5 10 6 9 6 3
wSh 6 4 1 2 9 2 8 1 2
wSan 2 3 1 4 8 4 7 1 1
wYak A3 5 3 1 2 8 4 7 5 1
wTei 2 3 1 2 8 4 7 1 1
wDin 4 1 1 2 6 5 6 1 1
wAu 3 1 3 2 6 4 4 1 1
wMelPop 1 1 1 1 3 1 1 1 1
wMel A1 1 1 1 1 1 1 1 1 1
wMelCS 1 1 2 1 2 1 1 1 1
wAna 2 1 3 2 7 2 5 4 1
wDau A2 2 2 3 2 5 2 5 4 1
wRi 2 2 3 2 4 2 2 2 1
wHa 2 1 3 3 5 3 3 3 2
1 -9 -8 -7 -6 -5 -4 -3 -2 -1 0
Linkage distance

(b)


wMau
wMa
wNo
B1
56
58
100
100
wDin
wMel
wMelCS
wMelPop
A1
93
98
67
wTei
wYak
wSan
A3
96
100
wAu
wHa
wSh
99
52
81
wAna
wDau
wRi
A2
95
97
100
0.005 substitutions/site