Parasitism and Mutualism in *Wolbachia*: What the Phylogenomic Trees Can and Cannot Say

Seth R. Bordenstein,†[1] *Charalampos Paraskevopoulos,*‡[1] *Julie C. Dunning Hotopp,*§‖
Panagiotis Sapountzis,‡ *Nathan Lo,*¶[2] *Claudio Bandi,*# *Hervé Tettelin,*§‖ *John H. Werren,*** and
Kostas Bourtzis‡[1]

*Josephine Bay Paul Center, Marine Biological Laboratory, Woods Hole, MA; †Department of Biological Sciences, Vanderbilt University; ‡Department of Environmental and Natural Resources Management, University of Ioannina, Agrinio, Greece; §J. Craig Venter Institute, Rockville, MD; ‖Institute for Genome Sciences, University of Maryland School of Medicine; ¶School of Biological Sciences, The University of Sydney, New South Wales, Australia; #Dipartimento di Patologia Animale, Igiene e Sanità Pubblica Veterinaria, Università di Milano, Milano, Italy; and **Department of Biology, University of Rochester, Rochester, NY

Ecological and evolutionary theories predict that parasitism and mutualism are not fixed endpoints of the symbiotic spectrum. Rather, parasitism and mutualism may be host or environment dependent, induced by the same genetic machinery, and shifted due to selection. These models presume the existence of genetic or environmental variation that can spur incipient changes in symbiotic lifestyle. However, for obligate intracellular bacteria whose genomes are highly reduced, studies specify that discrete symbiotic associations can be evolutionarily stable for hundreds of millions of years. *Wolbachia* is an inherited obligate, intracellular infection of invertebrates containing taxa that act broadly as both parasites in arthropods and mutualists in certain roundworms. Here, we analyze the ancestry of mutualism and parasitism in *Wolbachia* and the evolutionary trajectory of this variation in symbiotic lifestyle with a comprehensive, phylogenomic analysis. Contrary to previous claims, we show unequivocally that the transition in lifestyle cannot be reconstructed with current methods due to long-branch attraction (LBA) artifacts of the distant *Anaplasma* and *Ehrlichia* outgroups. Despite the use of 1) site-heterogenous phylogenomic methods that can overcome systematic error, 2) a taxonomically rich set of taxa, and 3) statistical assessments of the genes, tree topologies, and models of evolution, we conclude that the LBA artifact is serious enough to afflict past and recent claims including the root lies in the middle of the *Wolbachia* mutualists and parasites. We show that different inference methods yield different results and high bootstrap support did not equal phylogenetic accuracy. Recombination was rare among this taxonomically diverse data set, indicating that elevated levels of recombination in *Wolbachia* are restricted to specific coinfecting groups. In conclusion, we attribute the inability to root the tree to rate heterogeneity between the ingroup and outgroup. Site-heterogenous models of evolution did improve the placement of aberrant taxa in the ingroup phylogeny. Finally, in the unrooted topology, the distribution of parasitism and mutualism across the tree suggests that at least two interphylum transfers shaped the origins of nematode mutualism and arthropod parasitism. We suggest that the ancestry of mutualism and parasitism is not resolvable without more suitable outgroups or complete genome sequences from all *Wolbachia* supergroups.

Introduction

Anton de Bary coined the term "symbiosis" in his 1879 monograph as the living together of dissimilar organisms (de Bary 1879). By not specifically referring to beneficial or parasitic associations, this definition set the stage for understanding species interactions without constraint. Symbiosis as defined today encompasses all forms of species interactions because there are no universal principles that differentiate the mechanisms of mutualism and parasitism (Hentschel et al. 2000). Symbionts and hosts experience frequent transitions between different lifestyles, and mutualists and parasites can use similar genetic machinery for mediating parasitism and mutualism (Gargas et al. 1995; Ewald 2004; Sachs and Simms 2006).

One possible exception to this plastic view of symbiosis occurs in obligate intracellular (or endosymbiotic) bacteria that exclusively replicate inside host cells. These bacteria occur in diverse eukaryotic hosts and form parasitic and mutualistic interactions that can be evolutionarily stable

for hundreds of millions of years. A prevailing view is that such endosymbiotic lifestyles become irreversible as the bacteria streamline their gene content, thereby limiting the evolutionary potential for encoding alternative lifestyles (Moran and Wernegreen 2000).

Here, we present an evolutionary analysis of mutualism and parasitism in *Wolbachia pipientis* endosymbionts, widespread intracellular bacteria of arthropods, and filarial nematodes. *Wolbachia* evolved from a ~400-My-old clade of gram-negative, aerobic, α-proteobacteria that encompass obligatory intracellular, vertebrate pathogens and arthropod infections of the genera *Rickettsia*, *Ehrlichia*, *Anaplasma*, *Orientia*, *Neorickettsia*, and *Midichloria*. Despite the clade's ancient intracellular association, small genome size (0.9–1.6 Mb) and dependence on intracellular replication, the genus *Wolbachia* evolved labile lifestyles, primarily as reproductive parasites in arthropods and mutualists in filarial nematodes.

In arthropods, the reproductive parasites distort sex ratios and sexual reproduction strategies to gain a maternal transmission advantage (Werren 1997; Stouthamer et al. 1999). These sexual alterations include parthenogenesis, feminization, male killing, and cytoplasmic incompatibility, some of which are implicated in driving the evolution of new mechanisms of host sex determination (Rousset et al. 1992; Normark 2003; Negri et al. 2006), alternative modes of sexual selection (Jiggins et al. 2000), and incipient species (Bordenstein et al. 2001; Jaenike et al. 2006; Koukou et al. 2006). In rare cases, arthropod hosts have evolved

codependencies with reproductive parasites to the point where the *Wolbachia* are essential to host fertility (Starr and Cline 2002; Pannebakker et al. 2007). In contrast to the arthropods, antibiotic curing experiments suggest that in nematodes, *Wolbachia* infections are primarily beneficial to nematode fertility and larval development (Taylor et al. 2005). Further, the *Wolbachia* genome sequence from the filariid *Brugia malayi* suggests that these mutualists contribute essential compounds such as nucleotides, heme, and riboflavin to the host nematodes (Foster et al. 2005).

The major lifestyle differences in *Wolbachia* notably associate with discrete phylogenetic supergroups that differ at bacterial protein-coding genes and typically adhere to the criteria of greater than 3% divergence at the 16S rDNA gene (Lo et al. 2007). Thus, these lifestyle transitions within the *Wolbachia*-invertebrate endosymbiosis have occurred in a time frame potentially amenable to assessing the evolutionary trajectory of mutualism and parasitism through molecular phylogenetic analysis. The distinct lineages include the arthropod reproductive parasites in supergroups A and B and the nematode mutualists of supergroups C and D. Other diagnostic differences include a 200-kb smaller genome (Foster et al. 2005), complete vertical transmission (Casiraghi et al. 2001), and no recombination in the nematode C and D *Wolbachia*, whereas the arthropod A and B *Wolbachia* contain higher fractions of mobile DNA (Wu et al. 2004; Bordenstein and Reznikoff 2005), horizontally transfer between host species (Werren et al. 1995), and undergo high levels of recombination throughout the genome (Baldo et al. 2006). Prior detection of recombination in the A and B supergroups was based on a rich taxonomic sampling in these two groups and from strains known to coinfect the same hosts. The majority of *Wolbachia* supergroups are less prone to superinfection, and many of their functions remain uncharacterized. These taxa include supergroup E from primitively wingless insects, the springtails (Collembola) (Vandekerckhove et al. 1999; Lo et al. 2002; Czarnetzki and Tebbe 2004), supergroup F from termites, weevils, true bugs, and filarial nematodes (Casiraghi et al. 2001; Lo et al. 2002; Rasgon and Scott 2004), supergroup H from termites (Bordenstein and Rosengaus 2005), and three other divergent lineages that have not been labeled supergroups including those from the flea *Ctenocephalides canis* (Casiraghi et al. 2005), the filarial nematode *Dipetalonema gracile* (Casiraghi et al. 2005), and the pseudoscorpion *Cordylochernes scorpioides* (Zeh et al. 2005; Zeh JA and Zeh DW 2006). One other supergroup, G, has been reported to occur in spiders (Rowley et al. 2004) and a nonfilarial nematode (Tsai et al. 2007); we have excluded this group from our analysis as it may represent an artificial clade due to recombination within the gene used for phylogenetic analysis (Baldo and Werren 2007).

The utility of the *Wolbachia* endosymbiosis to assess directional shifts in parasitism and mutualism has not gone unnoticed. Several phylogenetic studies, including our own, have highlighted the utility of rooting the *Wolbachia* phylogenetic supergroups to polarize the evolutionary trajectory of changes in parasitism and mutualism (Lo et al. 2002, 2007; Fenn and Blaxter 2006). Despite attempts to reconstruct the rooted molecular phylogeny, limited gene and taxon sampling has yielded different phylogenetic results. Studies claim to either have positioned the root in the middle of the arthropod and nematode *Wolbachia* with no apparent direction in the evolution of the lifestyles (Anderson and Karr 2001; Fenn et al. 2006) or have expressed caution over certain rooting positions based on statistically weak phylogenetic support (Bandi et al. 1998; Lo et al. 2002; Bordenstein and Rosengaus 2005; Casiraghi et al. 2005).

The problems in resolving the root of the *Wolbachia* tree under these varied conditions are most likely due to one of two reasons: sampling error associated with too few genes in the alignment or systematic error from long-branch attraction (LBA) artifacts of the distantly related *Anaplasma* and *Ehrlichia* outgroup taxa. In general, if the outgroup branch is long enough from the ingroup, it can result in severe model violations due to the systematic error of multiple substitutions occurring per site, that is, mutational saturation (Jeffroy et al. 2006). This violation can lead to artificial but highly supported rootings. Although not previously recognized, all models used to reconstruct the *Wolbachia* phylogeny to date have failed to address this systematic error. Similar to maximum parsimony (MP), both maximum likelihood (ML) and Bayesian inference methods are not immune to LBA artifacts, especially under conditions of poor taxon sampling and poor gene or site selection with large data sets.

Elimination or reduction of systematic error associated with LBA artifacts can be achieved by three factors: 1) better taxon sampling because it enables increased detection of multiple substitutions, 2) rigorous selection of sites or genes to increase the ratio of phylogenetic:nonphylogenetic signal, and 3) probabilistic inference models that account for across-site heterogeneities to reduce model misspecifications (Lartillot et al. 2007). In our attempt to determine the ancestry of reproductive parasitism and mutualism in the rooted *Wolbachia* phylogeny, we take into account all three factors for a comprehensive evaluation of whether this endosymbiotic transition can be correctly reconstructed.

Materials and Methods
Gene and Taxon Sampling

Primary sequences were obtained from three sources: 1) published gene sequences (Baldo et al. 2006; Paraskevopoulos et al. 2006), 2) genome sequences of *Wolbachia* strains *w*Mel (Wu et al. 2004) (AE017196) and *w*Bm (Foster et al. 2005) (AE017321), *Ehrlichia chaffeensis* Arkansas (Hotopp et al. 2006) (CP000236), *Ehrlichia ruminantium* Welgevonden (Collins et al. 2005) (CR767821), *Anaplasma phagocytophilum* HZ (Hotopp et al. 2006) (CP000235), and *Anaplasma marginale* St Maries (Brayton et al. 2005) (CP000030), and 3) direct sequencing of polymerase chain reaction (PCR) products for which methods have been previously published (Baldo et al. 2006). All new sequences were deposited in GenBank under accession numbers: FJ390143–FJ390370. The average length of each trimmed sequence used in the alignments was 542 bp, and the average taxon representation per ortholog was 15 of 18 *Wolbachia* taxa. The 21 loci include WD_0001 (*dnaA*), WD_0160 (*nuoG*), WD_0183, WD_0198, WD_0203 (*atpD*), WD_0237, WD_0301 (*coxA*), WD_0307 (*groEL*), WD_0359, WD_0413 (*aspS*),
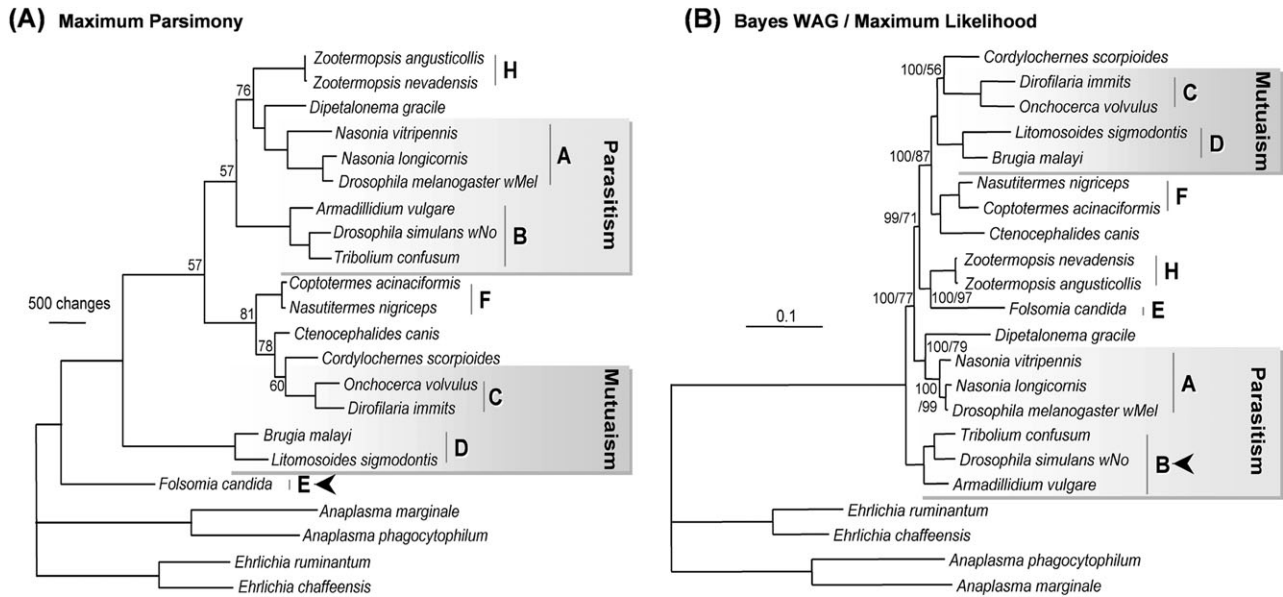
FIG. 1.—Phylogenomic trees showing the relationships of the *Wolbachia* supergroups and the placements of the root (arrow head symbol) from the *Anaplasma* and *Ehrlichia* outgroups to the *Wolbachia* ingroup. The taxon names of the *Wolbachia* ingroup denote those of the host species. Only support values below 100% are shown. (*A*) MP tree based upon concatenated nucleotide sequences of 21 protein-coding genes. (*B*) Common topology reconstructed from a Bayesian analysis of the amino acid translation of the concatenated proteins and a ML analysis of the concatenated genes. Support values above the nodes indicate Bayesian posterior probabilities/ML bootstrap values. The scale bar indicates the distances in substitutions per nucleotide.

WD_0473 (*pdhB*), WD_0484 (*hcpA*), WD_0544 (*sucB*), WD_0560 (*nuoD*), WD_0723 (*ftsZ*), WD_0791, WD_0976, WD_1005, WD_1029 (*aspC*), WD_1151 (*gltA*), and WD_1170 (*fabK*). A second data set was derived from previous analyses (Fenn and Blaxter 2006).

Sequence Nomenclature and Alignments

Wolbachia are named according to their host species and are labeled with a capital letter denoting the supergroup association. Primary and translated amino acid sequences were aligned in ClustalX and edited manually in MacClade 4.08. All indels and hypervariable regions were removed for a final alignment length of 11.9 kb per taxon.

Phylogenomic Analyses

PhyloBayes, MrBayes, and ML methods were used to infer phylogenomic relationships. PhyloBayes 2.1c and 2.3 analyses were performed with the Jones-Taylor-Thorton (JTT), Whelan and Goldman (WAG), general time reversible (GTR) models, and the category amino acid site-heterogenous mixture model to suppress tree artifacts associated with LBA (Lartillot and Philippe 2004; Lartillot et al. 2007). For all PhyloBayes analyses, at least two independent runs were performed with default PhyloBayes conditions until a maxdiff value <0.15 was achieved to ensure chain equilibration. The first 100 points were discarded as burn-in, and the posterior consensus was computed on the remaining trees. Data sets were also analyzed with WAG amino acid model in MrBayes. Runs were one million generations with tree sampling every 100 generations, and

a consensus tree was built on the last 7,000 trees. Posterior probabilities were determined by constructing a 50% majority rule tree of all the sampled trees. Finally, concatenated nucleotide sequences were analyzed with ML and MrBayes. Trees were reconstructed based on all nucleotides and only the first two codon positions to examine potential artifacts associated with third-codon positions. Prior to ML analyses, a DNA substitution model for each data set was selected using Modeltest v3.06 and the Akaike information criterion. The models selected for the first two codon positions and all nucleotides were the tranversional model with a gamma distribution + invariant sites (TVM + I + G) and the GTR model with a gamma distribution + invariant sites (GTR + I + G), respectively. ML heuristic searches were performed using 100 random taxon addition replicates with tree bisection and reconnection branch swapping. ML bootstrap support was determined using 100 bootstrap replicates, each using 10 random taxon addition replicates. Searches were performed in parallel on a Beowulf cluster using a clusterpaup program and PAUP version 4.0b10. Bayesian nucleotide analyses were performed with the same running conditions as the protein analyses.

Statistical Assessment of the Root

We tested the significance of topological differences in phylogenetic trees with different rooting positions using the Shimodaira–Hasegawa (SH) test (Shimodaira and Hasegawa 1999). The SH test compares the likelihood score ($-\ln L$) of a given data set across its ML tree with the $-\ln L$ of that data set across alternative topologies, which in this case are the ML phylogenies for other data sets. The differences in the $-\ln L$ values are evaluated

for statistical significance using bootstrap (1,000 replicates) based on RELL sampling and the more extensive full optimization (PAUP version 4.0b10). These two approaches yielded similar results.

Posterior Predictive Analyses

We used the predefined, posterior predictive tests, -div and -comp, as implemented in PhyloBayes. These statistics measure the site-specific amino acid diversity and the compositional heterogeneity among taxa, respectively. In the latter case, a compositional Z score is produced along with a P value to determine significance. All genes with taxon showing a Z score greater than 2 always had P values less than 0.05. In these cases, the genes were excluded to make a new data set when noted.

Detection of Recombination (Φ_W)

Alignments of individual and concatenated genes both with and without outgroups were screened for significant levels of recombination using the pairwise homoplasy index (PHI) or PHItest (Bruen et al. 2006) implemented in SplitsTree4 under default conditions (Huson and Bryant 2006). As a positive control, an alignment from the highly recombining *Wolbachia* surface protein (*wsp*) gene was included in this analysis. In comparison to other tests, Max χ^2 and neighbor similarity score (NSS), the PHItest reduces the frequency of false positives when confounding processes such as substitution rate heterogeneity affect the data set, especially under conditions of large sample sizes. It has been applied to varied data sets including viruses, mitochondria, fungi, and *Wolbachia* endosymbionts (Bruen et al. 2006).
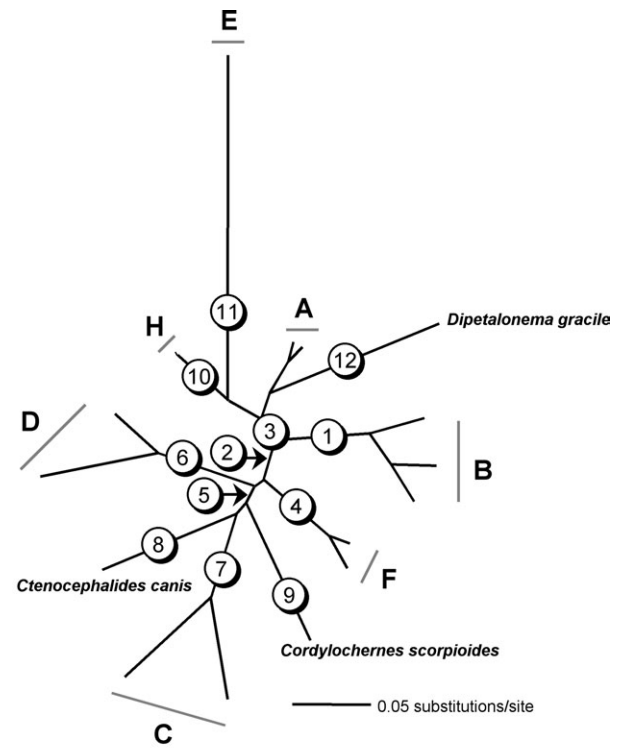
Results and Discussion
Rooting Position Using Standard Reconstruction Methods

We obtained primary sequences directly from the major *Wolbachia* supergroups A–F, H, and several divergent lineages from PCR products or public databases. These taxa cover the major supergroups/lifestyles and, when possible, two to three lineages within each of those supergroups. The final data set included 18 *Wolbachia* ingroup taxa and 4 outgroup species. The supermatrix consisted of a total of 21 protein-coding genes and up to 261 kb of sequence with indels or hypervariable regions removed to reduce nonphylogenetic signal in the data.

MP (DNA and Amino Acid)

Figure 1*A* shows the MP reconstruction of the rooted *Wolbachia* tree, based on the final concatenated matrix of 11,919 nt. The root is positioned at the base of supergroup E from the host springtail *Folsomia candida*. The symbiotic function of this lineage has not yet been characterized. The next earliest branching lineage in this tree is the mutualistic *Wolbachia* from the filarial nematodes *B. malayi* and *Lito-*

Summary of Results from Shimodaira-Hasegawa Tests			
Root Position	−ln L	Diff −ln L	P-value
1	90893.9	(best)	
2	90902.6	8.7	0.449
3	90905.2	11.4	0.291
4	90906.3	12.4	0.261
5	90907.3	13.4	0.220
6	90909.1	15.2	0.182
7	90919.5	25.6	0.035
8	90919.5	25.6	0.035
9	90916.1	22.2	0.060
10	90920.3	26.5	0.019
11	90914.9	21.1	0.090
12	90920.1	26.3	0.025

FIG. 2.—The unrooted ML tree showing the evolutionary relationships of the *Wolbachia* ingroup based on the concatenated alignment of 21 protein-coding genes and a tabular summary of an analysis of 12 different rooting positions. The numbers on the internal branches correspond to the rooting positions analyzed with the SH test. The lowest (best) −ln L scores for each data set are shown in descending order followed by the −ln L difference between the best score and the other data sets. Significance levels are based on full optimization. Taxon labels at the end of each terminal node refer to host species or to letters that correspond to the same supergroup letters in figure 1.

mosoides sigmodontis. Although there is 100% bootstrap support for the root's placement, there is reason to doubt the supergroup E rooting in the MP analysis. MP inference methods as mentioned above can be inferior to probabilistic models such as ML. In particular, the susceptibility of MP to LBA artifacts can artificially lead to the placement of the divergent outgroup branch next to the longest branch in the unrooted, ingroup tree. Consistent with this artifact, the longest ingroup branch of the *Wolbachia* tree is supergroup E from *F. candida* (see fig. 2). It is the single, divergent representative of supergroup E, and it has a different positioning in the unrooted MP phylogeny. The different evolutionary relationships in the rooted and unrooted MP

trees are a further indicator of LBA artifacts. Finally, a separate MP analysis of the translated amino acid sequences also places the root in supergroup E with weak posterior probability support of 67% and a highly unresolved tree (data not shown).

Bayes and ML (DNA and Amino Acid)

ML and Bayesian inference methods are generally preferred over MP for analyzing real sequence data because they incorporate models of evolution and are less sensitive to LBA artifacts (Felsenstein 1981). Figure 1*B* shows the Bayesian reconstruction of the rooted *Wolbachia* tree based on the 3,972 amino acid translation using the WAG model of protein sequence evolution. The most striking difference between this Bayesian tree and the MP tree is that the rooting position changed from the base of supergroup E to supergroup B—the reproductive parasites of arthropods (100% posterior probability support). The JTT and GTR models of protein-sequence evolution also yielded the identical topology with support for the root of supergroup B at 92% and 98%, respectively. Finally, the topology from the Bayesian amino acid analysis parallels the ML topology based on the entire nucleotide data set, which yielded 77% bootstrap support for the root at the base of the parasitic supergroup B.

When taken together, these analyses support reproductive parasitism as the ancestral association of the *Wolbachia* symbiosis. Although the probabilistic inferences models appear to be more robust than MP against LBA artifacts and it would be tempting to conclude that supergroup B is the root, a deeper assessment suggests this conclusion would be erroneous.

We also note caution in interpreting the unusual placement of the *D. gracile* nematode and *C. scorpioides* arthropod branches. *Dipetalonema gracile* is surprisingly positioned as a sister taxa to the arthropod supergroup A, and similarly *C. scorpioides* is positioned next to the filarial nematode supergroup C. These phylogenomic relationships between arthropod and nematode *Wolbachia* would suggest recurrent instances of interphylum host transfers of *Wolbachia*. In fact, some of these peculiar evolutionary relationships are contradicted by analyses under more complex inference models used below. However, to ensure that these patterns could not be explained by template contamination between arthropod and nematode DNAs, diagnostic PCRs using conserved rDNA primers confirmed the absence of contaminating arthropod DNA in the *D. gracile* template and the absence of nematode DNA in the *C. scorpioides* template.

Assessment of the Support for the Supergroup B Root
Codon Position

The ML bootstrap support for the rooting in the reproductive parasites is 77% when the nucleotide data set is analyzed with all three codon positions but 53% when the first and second codon positions are analyzed. This reduction is interesting because the removal of rapidly evolving, third-codon positions could potentially reduce the nonphyloge-

netic signal in the analysis and thereby increase support for the root. Alternatively, the reduction could be due to a decrease in the number of base pairs analyzed from 11,919 to 7,946 bp.

Single-Gene Trees

Of the 21 single-gene nucleotide alignments analyzed by ML with all three codon positions, only 6 produced trees with >50% bootstrap support for any rooting position. Thus, there is an overall lack of support concerning the rooting when each gene is taken separately. Of the six with >50% bootstrap support, only three placed the root at the base of supergroup B (67%, 58%, and 57% support), and the other three showed similarly weak support for alternative rootings, including the bases of supergroup C (53%), supergroup E (56%), and lineage *C. scorpioides* (52%). Finally, of the remaining 15 gene trees with less than 50% bootstrap support for the root, only 4 of these placed the root basally in supergroup B parasites. These variable results indicate a significant nonphylogenetic signal among the individual genes that could be reinforced as systematic error when using concatenated data sets.

Likelihood-Based Test of Topologies

If supergroup B is the earliest branching lineage of *Wolbachia*, then we predict that there will be significant topological differences between this rooting position and topologies with alternative rooting positions. In contrast, if the supergroup B rooting is not strongly supported, then we predict that there will be no topological differences to trees with different outgroup positions. Using an SH test (Shimodaira and Hasegawa 1999), we compared the relative support of the concatenated ML best tree (i.e., root in supergroup B, fig. 1*B*) with itself and 11 other topologies in which the root of the tree (leading to the four outgroups) was placed as a sister group to each of the different supergroups as well as to *Wolbachia* spp. from *D. gracile*, *C. canis*, and *C. scorpioides*. All 12 topologies were statically indistinguishable after a Bonferroni correction (fig. 2, corrected *P* value = 0.004); 8/12 were statistically indistinguishable without a Bonferroni correction, with no overall clustering of significant *P* values and taxonomic supergroups. Therefore, even with a framework of a taxonomically rich, multigenic data set and likelihood-based inference models, reproductive parasitism in supergroup B is not significantly more likely to be the root (fig. 1*B*) than other competing positions. Therefore, we are still left with the question of whether the analyses can accurately resolve the root of the *Wolbachia* clade using the closest known outgroups.

An Unresolved and Unresolvable Rooting Position

Multigenic analyses are predicated on the assumption that the concatenation of many genes can lead to an improved species tree. Systematic artifacts are the biggest potential problems associated with large genetic data sets owing to model violations and/or poor gene and taxon
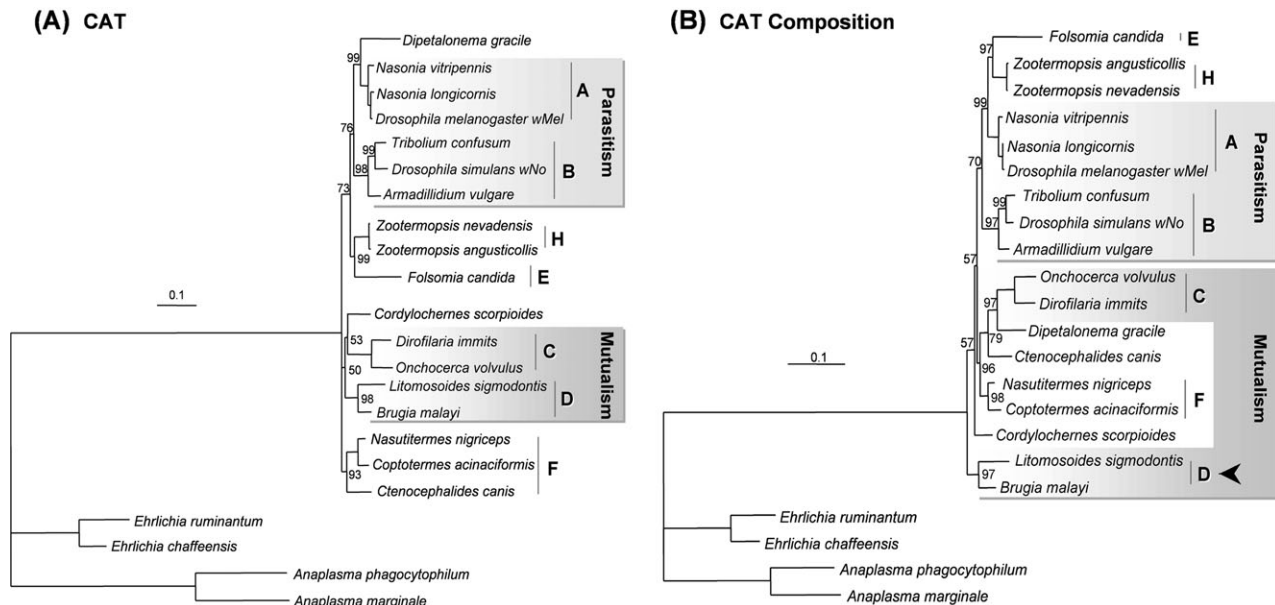
FIG. 3.—Phylogenomic trees inferred from the site-heterogenous CAT mixture model. Only support values below 100% are shown. (*A*) PhyloBayes CAT tree based upon the concatenated 21 proteins. (*B*) PhyloBayes CAT tree following removal of six genes with compositionally deviant taxa. Support values above the nodes indicate Bayesian posterior probabilities. The scale bars indicate the distances in substitutions per nucleotide.

sampling. For instance, the use of site-homogenous substitution models under an ML and Bayesian framework may not be resourceful enough to overcome the LBA phenomena associated with the *Wolbachia* tree. Further, extensive levels of gene exchange could also lead to artifacts and an unresolved species tree (Schierup and Hein 2000; Posada and Crandall 2002). To clarify the roles of recombination and rate heterogeneity on the rooting position, we took two approaches.

Recombination

First, we used the program PHItest (Bruen et al. 2006) to determine the presence or absence of recombination within each of the nucleotide alignments of the 21 gene data set. An evaluation of the performance of several recombination programs using both simulated and empirical data found that PHItest effectively determines recombination under diverse conditions and performs markedly better than Max χ^2 and NSS at avoiding false positives of recombination under models of substitution rate heterogeneity (Bruen et al. 2006). As the different *Wolbachia* and outgroup lineages can vary in their branch lengths, it is especially important to control for artifacts associated with rate heterogeneity. Recombination within and between the A and B *Wolbachia* supergroups has been reported previously using Max χ^2, and the A and B *Wolbachia* frequently coinfect the same arthropod hosts (Baldo et al. 2006). However, recombination and superinfection appear to be absent or much less frequent in other *Wolbachia* supergroups (Jiggins 2002; Bordenstein and Wernegreen 2004).

PHItest inferred recombination in only 1 of the 21 genes following a Bonferroni correction (WD0976, putative NADH dehydrogenase, $P < 0.001$). No recombination was evident when outgroups were included, and only one other gene showed evidence of recombination, without a Bonferroni correction (WD0359, helicase, $P = 0.027$). As a positive control, we included an alignment of the highly recombining *wsp* gene from *Wolbachia* and found a highly significant level of recombination (WD1063, $P < 0.0001$), thus confirming the usefulness of the program. We conclude that recombination does not significantly affect our data set and that the genes can be concatenated for phylogenomic analysis.

At least two reasons explain why recombination is absent in the same genes previously reported to recombine: 1) taxonomic sampling from *Wolbachia* supergroups other than A and B, many of which are not known to undergo recombination and 2) a test statistic, Φ_W, that reduces false-positive detections of recombination. Using PHItest, we inferred significant recombination in the same two data sets previously used to infer recombination with Max χ^2, *gltA*, and *groEL* (Baldo et al. 2006). This result indicates that diverse taxon sampling schemes or the use of different taxa from the same supergroups can mitigate the signal of recombination in certain portions of the *Wolbachia* phylogeny.

PhyloBayes CAT

Second, we used the CAT inference model as implemented in the PhyloBayes software package to attempt to overcome at least some of the risks of systematic errors in our analyses (Lartillot and Philippe 2004). Briefly, in contrast to models using a single empirical substitution matrix, CAT incorporates site-specific features of protein evolution by using a mixture model in which the amino acid replacement pattern at each position of the alignment can be described by a different substitution process and thus a unique subset of the 20 amino acids. Recent studies

emphasize that accounting for across-site heterogeneities in the amino acid replacement process is crucial for a better model fit and importantly can alleviate LBA (Lartillot et al. 2007; Rodriguez-Ezpeleta et al. 2007). Thus, the CAT model is an ideal method to assess the robustness of the *Wolbachia* phylogeny, including those trees from previous analyses claiming to have positioned the root between the mutualists and parasites (Fenn et al. 2006).

The CAT model was applied to the amino acid alignment previously studied in figure 1*B* using WAG, JTT, and GTR site-homogenous models. A less resolved tree was recovered that included an exceedingly long branch to the outgroup, which in turn did not resolve the evolutionary relationships of the base of the *Wolbachia* ingroup (fig. 3*A*). Although the tips of the ingroup tree were well resolved with high posterior probability support, the unresolved base of the ingroup seems to reflect the model's ambiguity in positioning the root. This conclusion is supported by the observation that the uncertainty in the tree disappears when the CAT model was applied to the same data set without the divergent outgroups. Posterior probabilities for each node of the unrooted tree then ranged confidently between 93% and 100%, with the exception of the branch from supergroup C leading to *C. scorpioides* (posterior probability is 57%, data not shown). This branch is also weakly supported in the rooted tree of figure 3*A*.

Analysis of Model Violations

The CAT model was confirmed to be a better statistical fit to the data set than the other site-homogenous models by examining two statistics that account for systematic artifacts, the mean number of substitutions per site (diversity) and the level of convergences and reversions (homoplasy). Diversity accounts for how well a model predicts the site-specific biochemical pattern of the data set. For our data set, the mean number of substitutions per site was 2.06. Under WAG, JTT, and GTR models, however, the posterior predicted number of substitutions per site was higher at 2.18, 2.19, and 2.16, respectively. In contrast, under CAT, the value was 2.05, a much closer prediction to the observed value. If the posterior predictive values are significantly different from the observed ones from the data, then the model of amino acid evolution underestimates diversity and can be more prone to false reconstructions. Because CAT and other models such as the site-homogenous WAG model principally differ in their amino acid replacement process, differences in the posterior predictive analyses indicate whether the amino acid replacement process directly alters the tree reconstructions. As expected, there were unequivocal significant differences between the observed and predicted diversity values for WAG, JTT, and GTR ($P < 0.0001$), whereas those for CAT were not significantly different ($P = 0.59$). Further, the posterior predicted mean number of homoplasies per site for CAT, 2.01, does not deviate from the observed number, 1.98. In contrast, the other three models underestimate the level of saturation as measured by homoplasy. For example, WAG has a much lower posterior number, 0.80, and a more deviating observed value of 0.95. These findings indicate that the CAT site-heterogenous model correctly accommodates
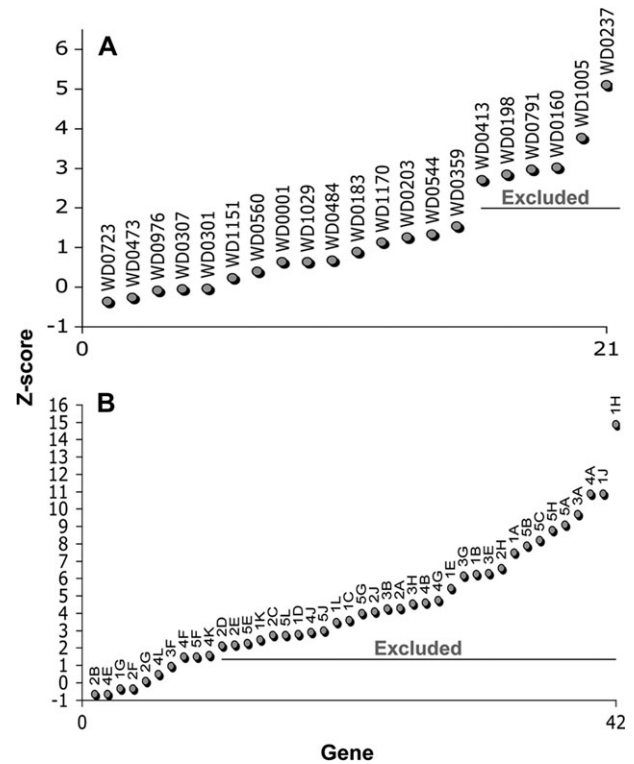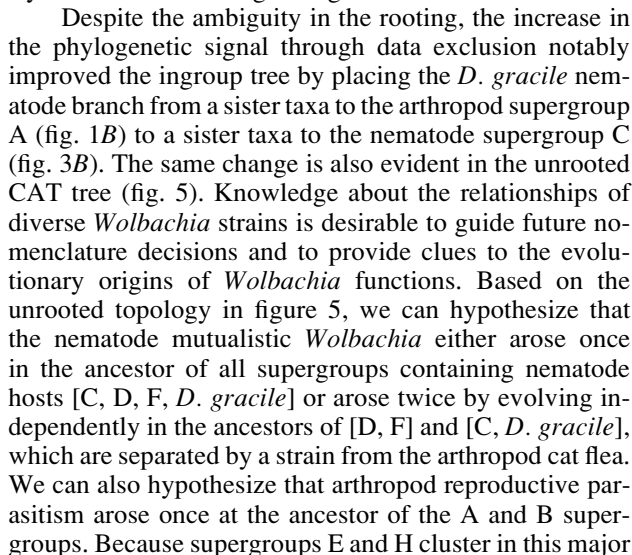

FIG. 4.—Summary plot of *Z* scores for each gene's amino acid alignment in ascending order for (*A*) the current data set and (*B*) a previously published data set (Fenn et al. 2006). Protein alignments with *Z* scores greater than 2 have taxa that show significant compositional deviation (*P* value < 0.05) and are excluded in downstream analyses. Labels in each plot correspond to the respective gene labels used in each data set.

the saturation patterns of the amino acid sequences. Consequently, it is more likely to offer accurate phylogenetic predictions, especially in the case where the root is not resolvable with current data (fig. 3*A*).

Data Exclusion to Reduce Systematic Error

An additional approach to reduce ambiguity and LBA artifacts in the tree reconstruction is to remove the sequences that most clearly violate some hypotheses of the model. This can have the effect of increasing the ratio of phylogenetic to nonphylogenetic signal in the data set. Such an approach is complementary to using the more sophisticated inference models of sequence evolution. Because fast-evolving genes by definition are more likely to accumulate multiple substitutions, we used the Phylo-Bayes software package to identify and remove the six genes with taxa that show significant compositional deviation (*Z* score > 2.0, *P* value < 0.05). The ascending *Z* scores for each gene are shown in fig. 4*A*. As figure 3*B* shows, the resulting tree under the CAT model, with deviant genes removed, yields an increase in resolution of the basal evolutionary relationships, a rooting placement at the base of the mutualists, and a shortening of the outgroup branch in comparison to the poorly resolved tree in figure 3*A*. However, despite the mild increase in resolution that suggests symbiotic lifestyle progressed from the

FIG. 5.—Unrooted phylogenomic tree inferred using the site-heterogenous CAT mixture model after excluding 6 compositionally deviant protein sequences of the original 21 sequences. Support values at the nodes indicate Bayesian posterior probabilities. The scale bar indicates the distance in substitutions per nucleotide.

mutualistic supergroup D *Wolbachia* to the later originating reproductive parasites, the root in the mutualists remains weakly supported at 57% posterior probability support, and the reconstruction cannot be interpreted as a more accurate tree. In contrast, under WAG, JTT, and GTR, the reconstructions yield strong support (99–100%) for the same rooting placement as CAT; this marked difference in support is a good example of when site-homogenous models can overpredict support for these basal relationships. WAG, JTT, and GTR models are more sensitive to saturation and attraction artifacts, further indicating their lack of reliability for which *Wolbachia* lifestyle is the earliest originating.

Despite the ambiguity in the rooting, the increase in the phylogenetic signal through data exclusion notably improved the ingroup tree by placing the *D. gracile* nematode branch from a sister taxa to the arthropod supergroup A (fig. 1*B*) to a sister taxa to the nematode supergroup C (fig. 3*B*). The same change is also evident in the unrooted CAT tree (fig. 5). Knowledge about the relationships of diverse *Wolbachia* strains is desirable to guide future nomenclature decisions and to provide clues to the evolutionary origins of *Wolbachia* functions. Based on the unrooted topology in figure 5, we can hypothesize that the nematode mutualistic *Wolbachia* either arose once in the ancestor of all supergroups containing nematode hosts [C, D, F, *D. gracile*] or arose twice by evolving independently in the ancestors of [D, F] and [C, *D. gracile*], which are separated by a strain from the arthropod cat flea. We can also hypothesize that arthropod reproductive parasitism arose once at the ancestor of the A and B supergroups. Because supergroups E and H cluster in this major

group of the tree [A, B, E, H], we can further hypothesize that these two taxa are or were ancestrally reproductive parasites. Based on this analysis and excluding the *C. scorpioides* lineage due to poor support in the tree, we conclude that there were at least two interphylum transfers of *Wolbachia* between nematodes and arthropods to explain the radiation of this endosymbiosis.

Reconsideration of the Previously Published Root

The results above demonstrate that caution is warranted concerning the placement of the root, even under cases of strong root support. A recent study using 42 protein-coding genes derived from five *Wolbachia* genomes and three outgroups (Fenn et al. 2006) gave a ML phylogeny where the root was placed between the reproductive parasites (supergroup A) and the mutualists (supergroups C and D) with 100% posterior probability support. However, the data set was noted to be missing the other taxonomic supergroups now included in our analysis. Further, support values assess sampling effects but cannot indicate whether the tree is accurate. Thus, the previous published root may appear correct, as had several of our own reconstructions, but could be subject to systematic artifacts. If the previously published rooting is due to nonphylogenetic signal in some of the data set, then we predict that there will be a significant fraction of genes with compositionally deviant taxa and that removing these genes would have a marked effect on the tree accuracy.

Using the same taxa from that study (Fenn et al. 2006), we constructed protein-sequence alignments for each of
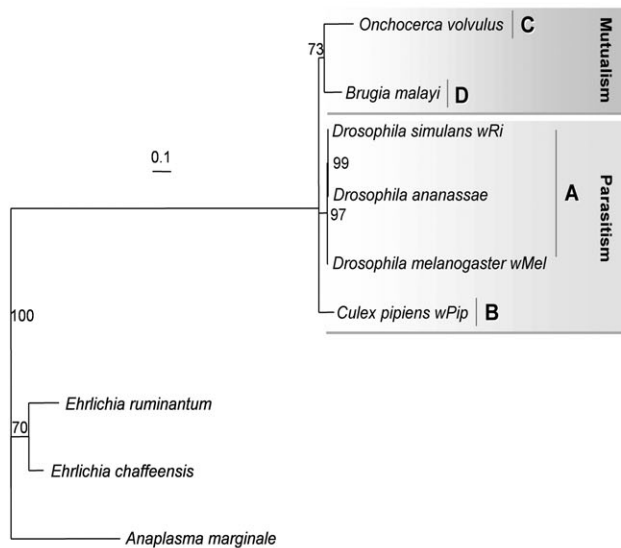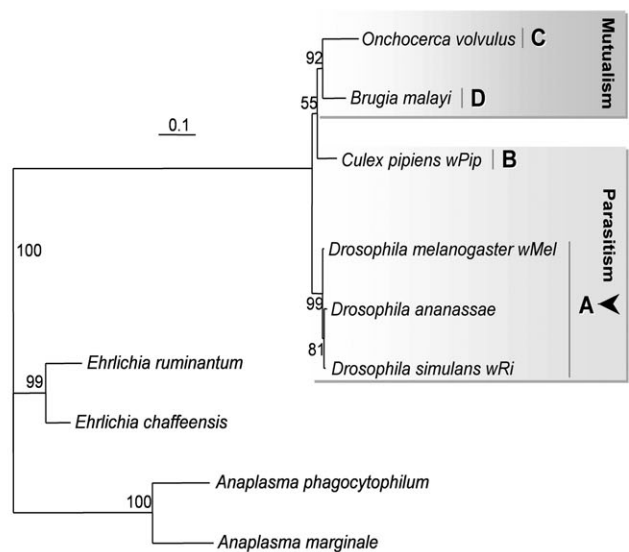
(A) CAT Composition (Fenn et al. 2006)

(B) CAT Composition (Combined)

FIG. 6.—Phylogenomic trees inferred from the site-heterogenous CAT mixture model following removal of genes with compositionally deviant taxa. (*A*) PhyloBayes CAT tree based upon 10 concatenated proteins from a previous data set (Fenn et al. 2006) that did not have compositionally deviant taxa. (*B*) PhyloBayes CAT tree based upon 25 combined proteins from the previous and current protein sequence data in this paper. Only Bayesian posterior probabilities values below 100% are shown.

the 42 genes. To expand the taxon diversity in this data set, we also added newly available genes from the supergroup B *w*Pip *Wolbachia* genome sequence to the alignments (Klasson et al. 2008). Using the PhyloBayes Z-score analysis for compositional heterogeneity, we identified that the overwhelming majority of genes had compositionally deviant taxa that could mislead the phylogenomic inferences. Indeed, only 10 of the original 42 genes in that data set were found to have Z scores below the default cutoff of 2.0, $P < 0.05$ (fig. 4*B*). Following data removal and tree reconstruction under CAT, we demonstrate in figure 6*A* that there is no defined, rooting position and the branch to the three outgroups is fatally long. Thus, composition variability has severely misled the tree accuracy reported previously. These results further illustrate the importance of examining nonphylogenetic signal that can mislead phylogenomic analyses.

To help increase the number of genes analyzed under the same criteria established above, we combined the compositionally homogenous genes from our data set with those of the previously published data set (Fenn et al. 2006) for a total of 25 protein-coding genes. We continued to use the same ingroup and outgroup taxa, with the exception that we added a fourth outgroup, *A. phagocytophilum*, because the whole-genome sequence was available. Following a new Z-score analysis, four genes were discarded: 2B, 2F, 2G, and WD0359. Under the CAT inference method, these 21 genes yielded a tree that polarizes the ancestor of *Wolbachia* in the reproductive parasites, supergroup A (fig. 6*B*). As demonstrated in all our analyses, the CAT model continues to caution the rooting placement in this particular clade or for that matter any clade. The root has a 55% posterior probability support. Therefore, the benefit of using the data and models that

can recover the most phylogenetic signal is interestingly the warning they send against confidently placing the root anywhere in the *Wolbachia* tree.

Conclusion

We presented an extensive, evolutionary analysis of mutualism and parasitism in *W. pipientis* endosymbionts, widespread intracellular bacteria of arthropods and filarial nematodes. The utility of the *Wolbachia* endosymbiosis to assess transitional steps between parasitism and mutualism in obligate intracellular bacteria has been well recognized (Lo et al. 2002, 2007; Bordenstein and Rosengaus 2005; Fenn and Blaxter 2006). Within the genus, there are four genetically defined lineages that can be discretely clustered into clades of mutualists that enhance filarial nematode fertility or viability (supergroups C and D) and clades of mostly reproductive parasites that distort sexual reproduction of their arthropod hosts to enhance their own transmission through the matriline (supergroups A and B). Despite attempts to reconstruct the ancestry of mutualism and parasitism in this bacterium, only one large-scale phylogenomic data set has been performed and claimed to position the root between the arthropod A and B parasites and the nematode C and D mutualists (Fenn et al. 2006).

We demonstrated in the analyses here that exploiting an amino acid mixture model and excluding mutationally saturated genes are crucial aspects of correctly accounting for the systematic error associated with the exceedingly long branch lengths between the last common ancestor of the ingroup and that of the outgroup genera. Indeed, a sequential look at the MP, ML/Bayesian, and CAT trees

shows a progressively, elongating outgroup branch. As the quality of the inference methods improve, the systematic biases become more evident in the length of the branch to the outgroup. Empirical matrices for protein sequence evolution including WAG, JTT, and GTR and probabilistic inference methods such as ML and Bayesian ML all failed to resolve the root upon a deeper assessment. In some cases, high statistical ML and Bayesian support for the rooting placement could have been interpreted as discovering the accurate tree and closing the book on the transition in symbiotic lifestyle. However, each of these models and implementations can be subject to systematic errors because they assume that the underlying model of sequence evolution is the same across the whole sequence and that there is no or weak mutational saturation. This assumption can lead to erroneous, but highly supported reconstructions, especially in genome-scale data sets with poor gene or taxon sampling. Phylogenomics must emphasize quality rather than quantity. Phylogenetic artifacts that lurk in the background of single-gene analyses can be compounded in large data sets such as the ones used here. Indeed, our subsequent analyses revealed that LBA artifacts grossly misled the phylogenetic accuracy of our data set and a previously published one (Fenn et al. 2006). After correctly accounting for the source of these LBA artifacts, the rooting placement returned to its rightful, ambiguous placement. In addition, recombination did not affect the global analysis and the accuracy of the ingroup tree improved once compositionally deviant genes were excluded. The *D. gracile* nematode lineage did improve from an aberrant positioning near the supergroup A *Wolbachia* in arthropods to a more accurate placement with strong support near the other nematode *Wolbachia* in supergroup C.

Even with the most taxonomically rich, phylogenomic data set of the *Wolbachia* clade, careful gene selection, and site-heterogenous inference models, LBA artifacts preclude an accurate rooting of the *Wolbachia* tree. The findings suggest that all past reconstructions using the Anaplasmataceae outgroups are fatally flawed from LBA artifacts. It should be stressed that the evolutionary relationships of the *Wolbachia* parasites and mutualists remain both unresolved and unresolvable until more suitable outgroup taxa or more taxonomic characters from full genome sequences of all the major *Wolbachia* supergroups become available.

Literature Cited

Anderson CL, Karr TL. 2001. *Wolbachia*: evolutionary novelty in a rickettsial bacteria. BMC Evol Biol. 1:10.

Baldo L, Bordenstein S, Wernegreen JJ, Werren JH. 2006. Widespread recombination throughout *Wolbachia* genomes. Mol Biol Evol. 23:437–449.

Baldo L, Werren JH. 2007. Revisiting *Wolbachia* supergroup typing based on WSP: spurious lineages and discordance with MLST. Curr Microbiol. 55:81–87.

Bandi C, Anderson TJ, Genchi C, Blaxter ML. 1998. Phylogeny of *Wolbachia* in filarial nematodes. Proc Biol Sci. 265:2407–2413.

Bordenstein S, Rosengaus RB. 2005. Discovery of a novel *Wolbachia* super group in Isoptera. Curr Microbiol. 51:393–398.

Bordenstein SR, O'Hara FP, Werren JH. 2001. *Wolbachia*-induced incompatibility precedes other hybrid incompatibilities in *Nasonia*. Nature. 409:707–710.

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

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

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 USA. 102:844–849.

Bruen TC, Philippe H, Bryant D. 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics. 172:2665–2681.

Casiraghi M, Anderson TJ, Bandi C, Bazzocchi C, Genchi C. 2001. A phylogenetic analysis of filarial nematodes: comparison with the phylogeny of *Wolbachia* endosymbionts. Parasitology. 122(Pt 1):93–103.

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.

Collins NE, Liebenberg J, de Villiers EP, et al. (21 co-authors). 2005. The genome of the heartwater agent *Ehrlichia ruminantium* contains multiple tandem repeats of actively variable copy number. Proc Natl Acad Sci USA. 102: 838–843.

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

de Bary A. 1879. The phenomenon of symbiosis. Strassburg: Verlag Von Karl J. Trubner.

Ewald PW. 2004. Evolution of virulence. Infect Dis Clin North Am. 18:1–15.

Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol. 17:368–376.

Fenn K, Blaxter M. 2006. *Wolbachia* genomes: revealing the biology of parasitism and mutualism. Trends Parasitol. 22:60–65.

Fenn K, Conlon C, Jones M, Quail MA, Holroyd NE, Parkhill J, Blaxter M. 2006. Phylogenetic relationships of the *Wolbachia* of nematodes and arthropods. PLoS Pathog. 2:e94.

Foster J, Ganatra M, Kamal I, et al. (25 co-authors). 2005. The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. PLoS Biol. 3:e121.

Gargas A, DePriest PT, Grube M, Tehler A. 1995. Multiple origins of lichen symbioses in fungi suggested by SSU rDNA phylogeny. Science. 268:1492–1495.

Hentschel U, Steinert M, Hacker J. 2000. Common molecular mechanisms of symbiosis and pathogenesis. Trends Microbiol. 8:226–231.

Hotopp JC, Lin M, Madupu R, et al. (39 co-authors). 2006. Comparative genomics of emerging human ehrlichiosis agents. PLoS Genet. 2:e21.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 23:254–267.

Jaenike J, Dyer KA, Cornish C, Minhas MS. 2006. Asymmetrical reinforcement and *Wolbachia* infection in *Drosophila*. PLoS Biol. 4:e325.

Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006. Phylogenomics: the beginning of incongruence? Trends Genet. 22:225–231.

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

Jiggins FM, Hurst GD, Majerus ME. 2000. Sex-ratio-distorting *Wolbachia* causes sex-role reversal in its butterfly host. Proc Biol Sci. 267:69–73.

Klasson L, Walker T, Sebaihia M, et al. (11 co-authors). 2008. Genome evolution of *Wolbachia* strain *wPip* from the *Culex pipiens* group. Mol Biol Evol. 25:1877–1887.

Koukou K, Pavlikaki H, Kilias G, Werren JH, Bourtzis K, Alahiotis SN. 2006. Influence of antibiotic treatment and *Wolbachia* curing on sexual isolation among *Drosophila melanogaster* cage populations. Evol Int J Org Evol. 60:87–96.

Lartillot N, Brinkmann H, Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol Biol. 7(Suppl 1):S4.

Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol. 21:1095–1109.

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

Lo N, Paraskevopoulos C, Bourtzis K, O'Neill SL, Werren JH, Bordenstein SR, Bandi C. 2007. Taxonomic status of the intracellular bacterium *Wolbachia* pipientis. Int J Syst Evol Microbiol. 57:654–657.

Moran NA, Wernegreen JJ. 2000. Lifestyle evolution in symbiotic bacteria: insights from genomics. Trends Ecol Evol. 15:321–326.

Negri I, Pellecchia M, Mazzoglio PJ, Patetta A, Alma A. 2006. Feminizing *Wolbachia* in *Zyginidia pullula* (Insecta, Hemiptera), a leafhopper with an XX/X0 sex-determination system. Proc Biol Sci. 273:2409–2416.

Normark BB. 2003. The evolution of alternative genetic systems in insects. Annu Rev Entomol. 48:397–423.

Pannebakker BA, Loppin B, Elemans CP, Humblot L, Vavre F. 2007. Parasitic inhibition of cell death facilitates symbiosis. Proc Natl Acad Sci USA. 104:213–215.

Paraskevopoulos C, Bordenstein SR, Wernegreen JJ, Werren JH, Bourtzis K. 2006. Toward a *Wolbachia* multilocus sequence typing system: discrimination of *Wolbachia* strains present in *Drosophila* species. Curr Microbiol. 53:388–395.

Posada D, Crandall KA. 2002. The effect of recombination on the accuracy of phylogeny estimation. J Mol Evol. 54:396–402.

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

Rodriguez-Ezpeleta N, Brinkmann H, Roure B, Lartillot N, Lang BF, Philippe H. 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. Syst Biol. 56:389–399.

Rousset F, Bouchon D, Pintureau B, Juchault P, Solignac M. 1992. *Wolbachia* endosymbionts responsible for various alterations of sexuality in arthropods. Proc Biol Sci. 250:91–98.

Rowley SM, Raven RJ, McGraw EA. 2004. *Wolbachia* pipientis in Australian spiders. Curr Microbiol. 49:208–214.

Sachs JL, Simms EL. 2006. Pathways to mutualism breakdown. Trends Ecol Evol. 21:585–592.

Schierup MH, Hein J. 2000. Consequences of recombination on traditional phylogenetic analysis. Genetics. 156:879–891.

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

Starr DJ, Cline TW. 2002. A host parasite interaction rescues *Drosophila* oogenesis defects. Nature. 418:76–79.

Stouthamer R, Breeuwer JA, Hurst GD. 1999. *Wolbachia* pipientis: microbial manipulator of arthropod reproduction. Annu Rev Microbiol. 53:71–102.

Taylor MJ, Bandi C, Hoerauf A. 2005. *Wolbachia* bacterial endosymbionts of filarial nematodes. Adv Parasitol. 60:245–284.

Tsai KH, Huang CG, Wang LC, Yu YW, Wu WJ, Chen WJ. 2007. Molecular evidence for the endosymbiont *Wolbachia* in a non-filaroid nematode, *Angiostrongylus cantonensis*. J Biomed Sci. 14:607–615.

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

Werren JH. 1997. Biology of *Wolbachia*. Annu Rev Entomol. 42:587–609.

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

Wu M, Sun LV, Vamathevan J, et al. (29 co-authors). 2004. Phylogenomics of the reproductive parasite *Wolbachia* pipientis *wMel*: a streamlined genome overrun by mobile genetic elements. PLoS Biol. 2:E69.

Zeh DW, Zeh JA, Bonilla MM. 2005. *Wolbachia*, sex ratio bias and apparent male killing in the harlequin beetle riding pseudoscorpion. Heredity. 95:41–49.

Zeh JA, Zeh DW. 2006. Male-killing *Wolbachia* in a live-bearing arthropod: brood abortion as a constraint on the spread of a selfish microbe. J Invertebr Pathol. 92:33–38.