GBE

Phylogenomics and Analysis of Shared Genes Suggest a Single Transition to Mutualism in *Wolbachia* of Nematodes

Francesco Comandatore[1], Davide Sassera[1], Matteo Montagna[1], Sujai Kumar[2,8], Georgios Koutsovoulos[2], Graham Thomas[2], Charlotte Repton[2], Simon A. Babayan[3,9], Nick Gray[3], Richard Cordaux[4], Alistair Darby[5], Benjamin Makepeace[6], and Mark Blaxter[2,7,*]

[1]DIVET, Università degli Studi di Milano, Milano, Italy

[2]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom

[3]Centre for Immunity, Infection and Evolution, University of Edinburgh, Edinburgh, United Kingdom

[4]Université de Poitiers, UMR CNRS 7267 Ecologie et Biologie des Interactions, Equipe Ecologie Evolution Symbiose, Poitiers Cedex, France

[5]Centre for Genomics Research, Institute of Integrative Biology, The University of Liverpool, Liverpool, United Kingdom

[6]Institute of Infection and Global Health, University of Liverpool, Liverpool, United Kingdom

[7]Edinburgh Genomics, University of Edinburgh, Edinburgh, United Kingdom

[8]Present address: Department of Zoology, University of Oxford, Oxford, United Kingdom

[9]Present address: Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow & Moredun Research Institute, Glasgow, United Kingdom

*Corresponding author: E-mail: mark.blaxter@ed.ac.uk.

Accepted: August 14, 2013

Data deposition: Sequence fastq files for *Litomosoides sigmodontis* have been submitted to the European Read Archive with accession ERP001496. Assemblies of wLs and wDi have been deposited in ENA with accessions PRJEB4155 and PRJEB4154 respectively. The ortholog clustering data and alignment files used in the analyses have been deposited with DataDryad with accession doi:10.5061/dryad.4nt8m.

Abstract

Wolbachia, endosymbiotic bacteria of the order Rickettsiales, are widespread in arthropods but also present in nematodes. In arthropods, A and B supergroup *Wolbachia* are generally associated with distortion of host reproduction. In filarial nematodes, including some human parasites, multiple lines of experimental evidence indicate that C and D supergroup *Wolbachia* are essential for the survival of the host, and here the symbiotic relationship is considered mutualistic. The origin of this mutualistic endosymbiosis is of interest for both basic and applied reasons: How does a parasite become a mutualist? Could intervention in the mutualism aid in treatment of human disease? Correct rooting and high-quality resolution of *Wolbachia* relationships are required to resolve this question. However, because of the large genetic distance between *Wolbachia* and the nearest outgroups, and the limited number of genomes so far available for large-scale analyses, current phylogenies do not provide robust answers. We therefore sequenced the genome of the D supergroup *Wolbachia* endosymbiont of *Litomosoides sigmodontis*, revisited the selection of loci for phylogenomic analyses, and performed a phylogenomic analysis including available complete genomes (from isolates in supergroups A, B, C, and D). Using 90 orthologous genes with reliable phylogenetic signals, we obtained a robust phylogenetic reconstruction, including a highly supported root to the *Wolbachia* phylogeny between a (A + B) clade and a (C + D) clade. Although we currently lack data from several *Wolbachia* supergroups, notably F, our analysis supports a model wherein the putatively mutualist endosymbiotic relationship between *Wolbachia* and nematodes originated from a single transition event.

Key words: *Wolbachia*, phylogenomics, mutualism, *Litomosoides sigmodontis*, endosymbiosis.

Introduction

Bacteria of the order Rickettsiales have an intracellular lifestyle and are involved in a variety of associations with eukaryotic hosts, from protists to vertebrates. These bacteria present distinctive genomic features that are likely to be driven by their intracellular lifestyle, including genome size and gene content

reduction, distorted nucleotide composition, and rapid gene evolution (Darby et al. 2007; Renvoise et al. 2011). *Wolbachia pipientis* is one of the most studied members of the Rickettsiales. Symbiotic associations with *Wolbachia* are widespread in arthropods, but have also been identified in nematodes: the animal-parasitic filarial nematodes and the plant-parasitic nematode *Radopholous similis* (Bandi et al. 1998; Werren et al. 2008; Haegeman et al. 2009). The molecular diversity within the single nominal species *Wolbachia pipientis* (Lo et al. 2007) has been used to define a series of 13 supergroups (monophyletic clades; labeled alphabetically, A to N) that show different lifestyles and host ranges (Doudoumis et al. 2012). The A and B supergroups were the first to be described (Werren et al. 1995), followed by the C and D (Bandi et al. 1998). These four are also the most widely investigated *Wolbachia* supergroups. The A and B supergroup strains are associated with arthropods, whereas C and D are associated with filarial nematodes.

In arthropods, *Wolbachia* normally have a patchy distribution among species and populations, and infection is generally associated with alterations of host reproduction, such as parthenogenesis, killing of male embryos, feminization of genetic males, and cytoplasmatic incompatibility (Werren et al. 2008; Cordaux et al. 2011). In a few cases, *Wolbachia* has been demonstrated to be essential for the reproduction of the arthropod host (Starr and Cline 2002; Pannebakker et al. 2007). All *Wolbachia* lineages are vertically inherited (from mother to offspring), but horizontal transmission is evident between hosts for numerous strains of the A and B supergroups. The phylogenies of A and B supergroup *Wolbachia* do not track their hosts' phylogenies, suggesting frequent host switching (Werren et al. 1995).

The characteristics of the symbiosis are different in filarial nematodes, where available evidence indicates that the symbionts are beneficial to their hosts. *Wolbachia* usually have 100% prevalence in positive species (Taylor et al. 2005; Ferri et al. 2011), and are strictly vertically inherited, with phylogenies largely congruent with that of their hosts (Bandi et al. 1998; Casiraghi et al. 2001). In addition, they appear to be essential for host survival, as *Wolbachia* elimination with tetracyclines harms the host (Bandi et al. 1999; Hoerauf et al. 1999). Supergroup C and D *Wolbachia* have smaller genomes, and fewer genes, than the parasitic supergroup A and B *Wolbachia* (Foster et al. 2005; Werren et al. 2008) as would be expected from closer integration of host and symbiont genomes. Comparative metabolic reconstruction from the genomes of sequenced *Wolbachia* from filarial nematodes has not revealed an unequivocal signal of the essential symbiotic partnership. Currently favoured models include heme and riboflavin biosynthesis (Foster et al. 2005; Godel et al. 2012), but energy provisioning and immunomodulatory models may be more realistic (Darby et al. 2012).

The origins of the mutualistic relationships of C and D supergroup *Wolbachia* with filarial nematodes are of particular interest. *Wolbachia* have evolved from intracellular symbionts (Rickettsiales), and the closest related taxa are generally considered to be pathogens, such as the arthropod-infecting A and B supergroups. Are filarial-infecting mutualists monophyletic, implying a single origin of mutualism, or has mutualism arisen independently multiple times? Are the filarial *Wolbachia* more closely related to A or B supergroups? Several studies have highlighted the critical importance, and difficulty, of rooting *Wolbachia* supergroup phylogeny to the solution of this question (Lo et al. 2002, 2007; Fenn et al. 2006; Bordenstein et al. 2009). Two well-known artifacts likely explain the difficulty of obtaining a well-resolved phylogeny: long-branch attraction (LBA), caused by the large distances to the nearest outgroup taxa *Anaplasma* spp. and *Erhlichia* spp., and a basal, star-like evolutionary radiation of the genus *Wolbachia* (Bordenstein et al. 2009). Fenn et al. (2006), analyzing 42 protein-coding genes from five taxa in A, C, and D supergroups, proposed rooting *Wolbachia* between A and (C + D). Bordenstein et al. (2009) used 21 protein-coding genes from 18 *Wolbachia* taxa, representing the A, B, C, D, E, F, and H supergroups, but did not find unequivocal support for the position of the root, and suggested that reliable resolution of the *Wolbachia* phylogenetic tree would require improved taxon and gene sampling. It is becoming clear that careful selection of loci before analysis is key to robust and believable resolution of many phylogenetic questions when multigene data sets are used (Salichos and Rokas 2013). In particular, coanalysis of genes with different underlying patterns of substitution, horizontal gene transfer, acquisition by hybridization, and hidden paralogy can confound strong signal within data sets.

We have determined the genome sequence of an additional supergroup D *Wolbachia* from the filarial nematode *Litomosoides sigmodontis*. Here, we revisit the selection of single-copy orthologs from completely sequenced genomes for *Wolbachia* phylogenomics, and use an extended gene data set to develop a robust hypothesis of *Wolbachia* relationships.

Materials and Methods

Litomosoides sigmodontis DNA was extracted from nematodes grown in gerbils (*Meriones unguiculatus*) as previously described (Diagne et al. 1990). Short-insert paired-end libraries with 300 and 600 bp inserts were prepared by the GenePool Genomics Facility and sequenced on the Illumina HiSeq2000 with V3 reagents. Reads were corrected using SOAPec (Luo et al. 2012), digitally normalized using khmer (Brown et al. 2012), and preliminary assemblies produced using velvet (Zerbino and Birney 2008). These assemblies were screened for *Wolbachia*-derived sequence using taxon-annotated GC%-coverage plots (Kumar and Blaxter 2011) and the 18 likely *Wolbachia*-derived contigs and their reads selected for stringent reassembly using ABySS (Simpson et al. 2009) (using

a kmer of 83, default coverage cutoff and a minimum of 3 read pairs to join contigs). Joins in the assembly that had low coverage were validated using polymerase chain reaction (PCR). The assembly (wLs.2.0) is available through http://litomosoides.nematod.es (last accessed September 5, 2013).

The genomes of 11 *Wolbachia* strains and 4 outgroups were retrieved from the databases (see fig. 1). For wDi from the nematode *Dirofilaria immitis*, we reassembled the genome (Godel et al. 2012) using improved informatic routines, extracting additional read data from the raw genome sequence for *D. immitis*. The new assembly is improved (in that it has many fewer contigs). The contiguity of this new assembly (wDi.2.2) was verified by directed PCR and is available from http://dirofilaria.nematod.es (last accessed September 5, 2013).

Ortholog detection was performed using OrthoMCL 1.4 with default settings (Chen et al. 2006). All sequences of each putative orthologous cluster were automatically annotated (using BLASTP with an E-value cutoff of 10^{-5}) against the Clusters of Orthologous Groups (COG) database (Tatusov et al. 2000). An orthologous cluster was selected for subsequent analyses if all sequences of the cluster were coherently annotated by comparison to the COG database. Orthologous clusters containing members from all 16 genomes and lacking within-genome duplicates were selected, and amino acid sequences aligned using Muscle (Edgar 2004) with default settings. Nucleotide alignments were obtained by retrotranslation of these amino acid alignments. The Pairwise Homoplasy Index (PHI) and MaxChi were calculated for each nucleotide alignment with PhiPack (Bruen et al. 2006) with 1,000 permutations and window dimensions of 30, 60, and 100 bases. To detect potential recombination events, recombination analyses were repeated, for each alignment, considering the sequences of the strains belonging to the A + B, C + D, and A + B + C + D supergroup sets. Alignments presenting no evidence of recombination were subjected to mutational saturation analysis with Xia's method (Xia and Xie 2001).

Poorly aligned positions and divergent regions of nucleotide and amino acid alignments were eliminated with Gblocks (Castresana 2000), allowing gap positions (-b5 = all option). Nucleotide and amino acid alignments for the 90 genes were concatenated. Phylogenetic reconstructions were estimated on nucleotide and on amino acid unpartitioned concatenates with Maximum Likelihood (ML) and Bayesian methods using models chosen using jModelTest (Darriba et al. 2012) and Prottest3 (Darriba et al. 2011). The best-fit model for unpartitioned analyses of the concatenated nucleotide alignment was GTR, whereas for unpartitioned analyses of amino acid alignment JTT was identified as optimal. The models selected for partitioned analyses of the alignments are given in supplementary table S3, Supplementary Material online. The GTR model identified as best-fitting for 30 of 90 nucleotide alignments was used for ML and Bayesian analyses for all partitions. The amino acid alignment was split into five partitions, grouping all genes that shared the best model among those implemented in MrBayes, and ML and Bayesian analyses were performed on this partitioned concatenate using the best-fit model for each partition.

ML phylogenetic analyses were executed with 1,000 rapid bootstrap replicates within RaxML 7.2.8 (Stamatakis et al. 2008). Bayesian phylogenetic analyses were carried out on unpartitioned concatenates with MrBayes 3.2 (Ronquist et al. 2012) on the web-based Bioportal (Kumar et al. 2009). Bayesian Markov chain Monte Carlo analyses were implemented in two parallel analyses, each composed of one cold and five incrementally heated chains that were run for 10 million generations. Trees were sampled every 1,000 generations and burn-in fraction was calculated according to lnL stationary analyses.

Gene presence–absence information for wUni (*Wolbachia* endosymbiont of *Muscidifurax uniraptor*) was removed from all orthologous clusters before performing gene presence–absence analysis, because the wUni genome is not yet complete (Klasson et al. 2009). The ortholog presence–absence matrix was derived from the ortholog cluster data and used to calculate the Bray–Curtis dissimilarity matrix. The Bray–Curtis dissimilarity index evaluates the gene fraction not shared between two taxa with the formula $1 - ([2*(A \cap B)]/A + B)$, where A and B represent the gene sets of the two taxa. Fingerprint Analysis with Missing Data (Schlüter and Harris 2006) was used to perform UPGMA analysis and the heatmap was drawn with R.

Results

Litomosoides sigmodontis is an onchocercid filarial parasite of cotton rats (Hoffmann et al. 2000). The *L. sigmodontis Wolbachia*, wLs, was assembled from data generated as part of the ongoing *L. sigmodontis* genome project (Koutsovoulos G, Kumar S, Babayan SA, Blaxter M, unpublished data; see http://litomosoides.nematod.es, last accessed September 5, 2013). The wLs genome assembly was generated by identifying contigs in initial genome assemblies that contained *Wolbachia* genes, extracting the raw data that mapped to these contigs and performing independent assembly. The genome was refined through cycles of additional read identification and assembly and validation of some joins by PCR. The raw data have been submitted to INSDC databases under project accession ERP001496.

We retrieved whole genome-derived gene data for 11 additional *Wolbachia* strains and four outgroup species from ENA (see Materials and Methods). The genes used in phylogenomic analyses were selected from sets of orthologs, identified as reciprocal best BLAST hits and validated through comparison with the COG database. A subset of orthologs present in all taxa was identified. Ortholog sets showing evidence of paralog duplication, evidence of recombination

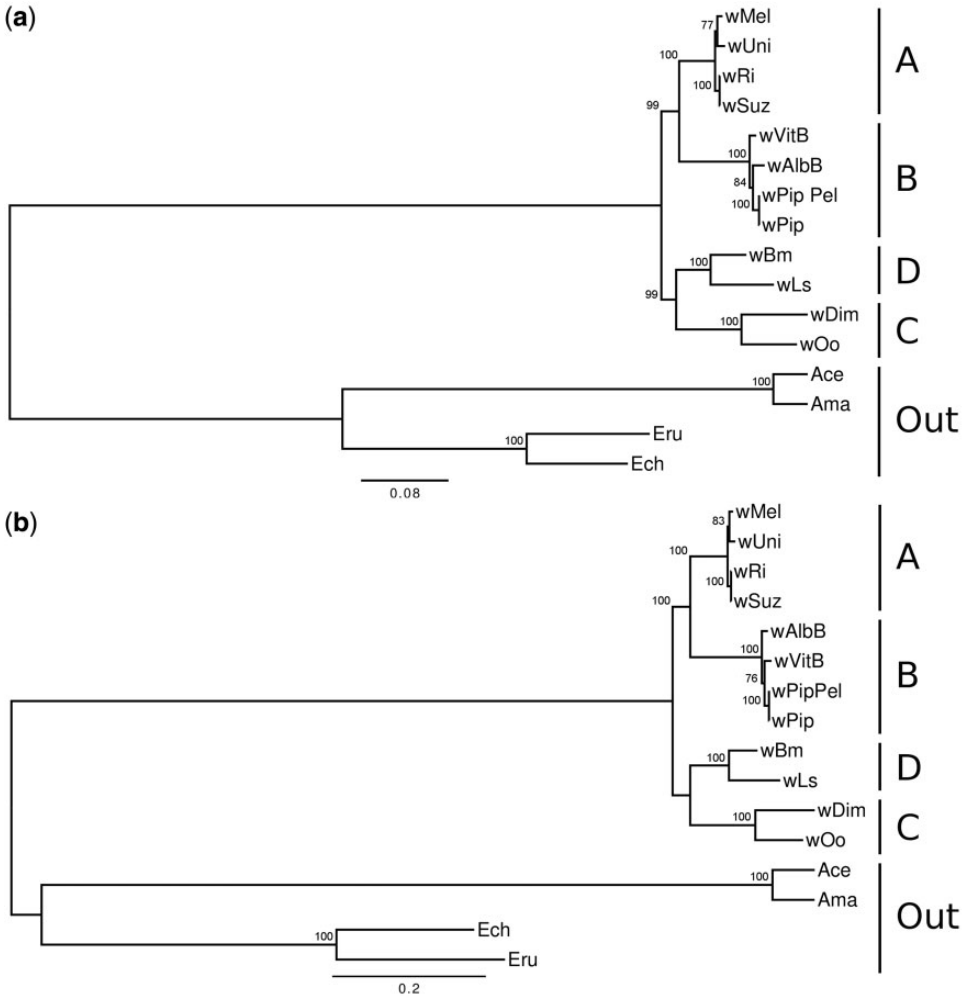

Fig. 1.—Phylogenomic analysis of *Wolbachia*. Phylogenetic trees generated with RaxML based on amino acid (*A*) and nucleotide (*B*) partitioned concatenates. ML bootstrap values are reported above each node of the trees. The corresponding trees generated with MrBayes, showing completely congruent topologies and posterior probability of 1 for each node, are reported in supplementary figure S1, Supplementary Material online. The strains analyzed are *Wolbachia* endosymbiont of *Drosophila melanogaster*, wMel; *Wolbachia* endosymbiont of *Drosophila simulans*, wRi; *Wolbachia* endosymbiont of *Drosophila suzukii*, wSuz; *Wolbachia* endosymbiont of *Muscidifurax uniraptor*, wUni; *Wolbachia* endosymbiont of *Culex quinquefasciatus* JHB, wPip; *Wolbachia* endosymbiont of *Culex quinquefasciatus* Pel, wPip Pel; *Wolbachia* endosymbiont of *Nasonia vitripennis*, wVitB; *Wolbachia* endosymbiont of *Aedes albopictus*, wAlbB; *Wolbachia* endosymbiont of *Brugia malayi*, wBm; *Wolbachia* endosymbiont of *Onchocerca ochengi*, wOo; *Wolbachia* endosymbiont of *Dirofilaria immitis*, wDi; *Anaplasma centrale* str. Israel; *Anaplasma marginale* str. Florida; *Ehrlichia chaffeensis* str. Arkansas; *Ehrlichia ruminantium* str. Gardel. Letters A, B, C, and D indicate *Wolbachia* supergroup memberships.

between genomes, or evidence of nucleotide substitution saturation were removed. We identified 1,677 ortholog clusters, 1,519 of which were coherently annotated. The sixteen bacterial genomes shared 390 of the 1,519 gene clusters, 341 of which presented no evidence of duplication. Of these 341 clusters, 126 showed no evidence of recombination, and, of these, 90 showed no evidence of nucleotide substitution saturation. These 90 clusters were retained and used for phylogenomic analysis (see supplementary table S1, Supplementary Material online, for a complete list of these 90 genes). Maximum likelihood (ML; fig. 1) and Bayesian phylogenetic inference (supplementary fig. S1, Supplementary Material online) using nucleotide and amino acid alignments of these 90 genes differed only in the relative position of the strains wAlbB and wVitB within supergroup B. Other than this disagreement, all nodes received high joint support. Importantly, the length of the branches between the three genera *Wolbachia*, *Anaplasma*, and *Ehrlichia* were reasonably homogeneous and did not suggest the presence of LBA artifacts. Analysis of individual alignments showed that (C + D) monophyly was supported by 48 of the 90 loci, and (A + B) monophyly was supported by 43 loci (see supplementary information, Supplementary Material online). None of the nodes in the catenated analysis is supported by a low

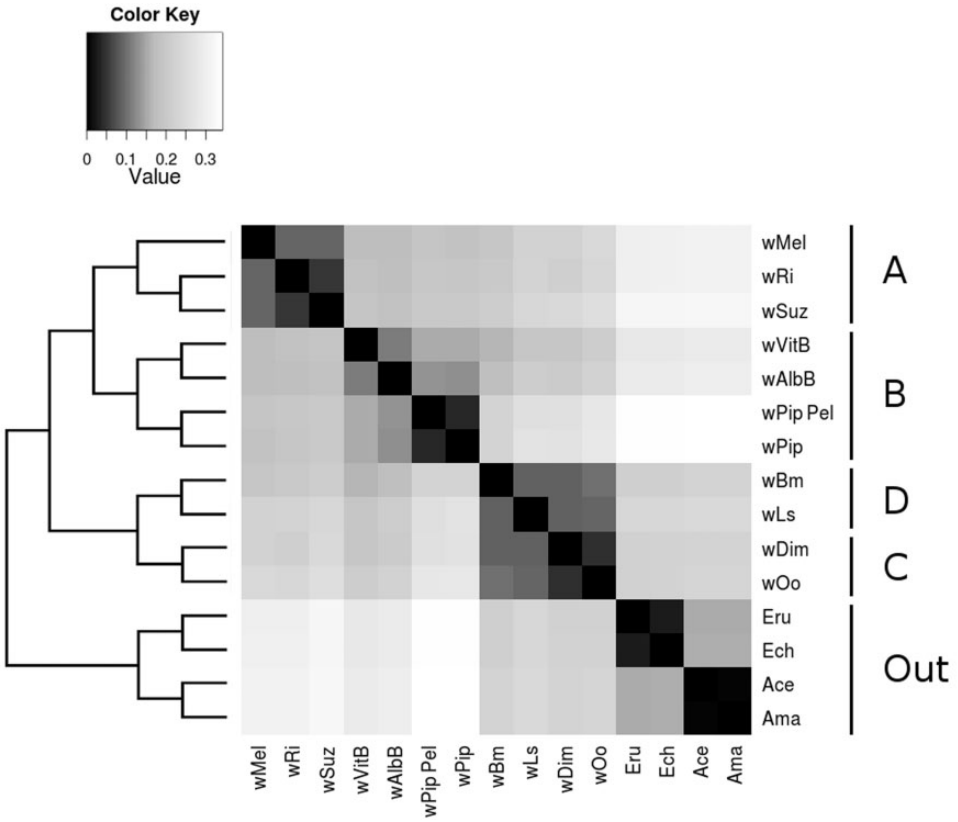

Fig. 2.—Gene presence–absence analysis of *Wolbachia* genomes. An UPGMA tree (left) was inferred based on the Bray–Curtis dissimilarity matrix calculated on the presence–absence matrix of genes in the examined genomes. The heatmap to the right of the tree represents the values of the Bray–Curtis dissimilarity matrix. Strain abbreviations are as given in figure 1.

number of individual genes, and the majority rule consensus of the individual locus phylogenies is the same as the concatenated analysis. The analyses supported a root placement between the A and B supergroups and the C and D supergroups, yielding a monophyletic filarial mutualist clade.

A presence–absence matrix was constructed from the 1,519 coherently annotated orthologous clusters from the complete *Wolbachia* genomes (i.e., excluding wUni; see Materials and Methods). From this matrix, a pairwise Bray–Curtis dissimilarity matrix was calculated, and this dissimilarity matrix was subjected to UPGMA phenetic analysis (fig. 2). The phenetic analysis was congruent with the sequence-based phylogenomic analyses, linking the A and B and the C and D supergroups.

Discussion

Previously published molecular phylogenetic reconstructions have not revealed the number of independent transitions to mutualism in filarial nematode *Wolbachia* (Casiraghi et al. 2005; Fenn et al. 2006; Bordenstein et al. 2009). This has been due to limited phylogenetic signal present in few loci (Casiraghi et al. 2005; Bordenstein et al. 2009), a lack of genomic data from a representative diversity of strains (Fenn et al. 2006), and LBA due to extreme divergence from the nearest outgroup taxa (Bordenstein et al. 2009). We sequenced a new supergroup D genome, wLs of *L. sigmodontis* and collated a 90-gene phylogenomic data set using a custom pipeline designed to remove all loci likely to contain phylogenetic noise. Salichos and Rokas (2013) have recently explored issues of data incongruity in phylogenomic analyses, using a deep phylogeny of yeasts as their model. We concur with their proposals to eliminate rigorously from consideration loci with abberations in phylogenetic signature, assessed independently of the derivation of the phylogeny. We note that our phylogenetic question is less problematic than their model, with fewer taxa overall and some unquestioned groupings (such as the monophyly of clades A, B, C, and D). Thus, we reduced our original set of more than 400 putative single copy orthologs to a core set of 90 genes with validated behavior. Only 3 of the 21 genes of the Bordenstein set (Bordenstein et al. 2009) passed the stringent assessment for inclusion in our database. ML and Bayesian analyses of nucleotide and amino acid alignments yielded congruent topologies with high statistical support. Importantly, the branch lengths observed between the two genera in the outgroup (*Anaplasma*

and *Ehrlichia*) was comparable to the branch length observed between the outgroup clade (*Anaplasma* + *Ehrlichia*) and the *Wolbachia* clade, suggesting the absence of LBA effects on the phylogenies. Individual locus phylogenies tended to support this hypothesis. Our phylogenies provide strong evidence, with high statistical support, for the monophyletic origin of arthropod (A and B) and nematode (C and D) *Wolbachia* strains (fig. 1 and supplementary fig. S1, Supplementary Material online). Phenetic analysis of gene presence–absence data also supported this set of relationships.

In summary, our analyses demonstrate that arthropod (A and B) and nematode (C and D) *Wolbachia* originated after the split of an ancestral lineage. We cannot determine whether this ancestral lineage was associated with nematodes, arthropods, or another host group, and we cannot derive conclusions on the nature of the symbiosis (parasitic vs. mutualistic) of this ancestor. Considering the phylogenetic position of *Wolbachia* within the order Rickettsiales, we can reasonably infer that it was an intracellular bacterium. A monophyletic origin of the C and D supergroups is congruent with the idea that the characteristics shared by these two supergroups (strict association with the host, strict vertical transmission, and evidence for a beneficial contribution to host biology) originated only once during evolution, in the lineage that led to the *Wolbachia* of filarial nematodes. As noted by authors in previous studies, analysis of *Wolbachia* diversity is compromised by partial sampling across the known supergroups (Casiraghi et al. 2005; Fenn et al. 2006; Bordenstein et al. 2009). We have been able to use complete genome data from only four supergroups and eagerly await emerging data for strains from other supergroups. Of particular interest will be genomic data from supergroup F strains, as these are reported to infect both arthropods and filarial nematodes (Lefoulon et al. 2012). The placement of supergroup F strains in phylogenies is variable, but they are often associated with supergroup C and D (Lefoulon et al. 2012). An exciting possibility is that supergroup F is sister taxon to C and D, and this may represent the lifestyle of the last common C and D ancestor.

The origin of the relationship between nematodes and *Wolbachia* is interesting from both evolutionary and medical standpoints. Nematode *Wolbachia* represent important targets for the treatment of human and animal filariases (Slatko et al. 2010). Our analyses suggest that an endosymbiotic relationship with nematodes is a plesiomorphic character of the (C + D) clade. In this scenario, an ancestral *Wolbachia* strain invaded the first filarid host, evolved a mutualistic association that included strict vertical inheritance, and the C and D supergroups originated through ancient host lineage divergence. Under this model, all these *Wolbachia* strains probably share common metabolic traits that underpin their mutualistic relationship with the filarial hosts. This in turn suggests the possible presence of common anti-*Wolbachia* pharmacological targets for the control of their pathogenic filarial hosts.

Supplementary Material

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

Acknowledgments

This work was supported by postgraduate fellowships awarded to S.K. and G.K. from the School of Biological Sciences, the University of Edinburgh, and to G.S. from the Wellcome Trust. Genome sequencing was supported by an award from the EU FP7 programme (EU Specific International Cooperation Action [SICA] reference 242131 Enhanced Protective Immunity Against Filariasis) awarded to Prof. David Taylor. R.C. was supported by an European Research Council Starting Grant (FP7/2007-2013, grant 260729 EndoSexDet).

Literature Cited

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

Bandi C, et al. 1999. Effects of tetracycline on the filarial worms *Brugia pahangi* and *Dirofilaria immitis* and their bacterial endosymbionts Wolbachia. Int J Parasitol. 29:357–364.

Bordenstein SR, et al. 2009. Parasitism and mutualism in *Wolbachia*: what the phylogenomic trees can and cannot say. Mol Biol Evol. 26: 231–241.

Brown C, Howe A, Zhang Q, Pyrkosz A, Brom T. 2012. A reference-free algorithm for computational normalization of shotgun sequencing data, arXiv:1203.4802.

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 122:93–103.

Casiraghi M, et al. 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.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 17:540–552.

Chen F, Mackey AJ, Stoeckert CJ Jr, Roos DS. 2006. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res. 34:D363–D368.

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

Darby AC, Cho NH, Fuxelius HH, Westberg J, Andersson SG. 2007. Intracellular pathogens go extreme: genome evolution in the Rickettsiales. Trends Genet. 23:511–520.

Darby AC, et al. 2012. Analysis of gene expression from the *Wolbachia* genome of a filarial nematode supports both metabolic and defensive roles within the symbiosis. Genome Res. 22:2467–2477.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27:1164–1165.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. 9:772.

Diagne M, Petit G, Liot P, Cabaret J, Bain O. 1990. The filaria Litomosoides galizai in mites; microfilarial distribution in the host and regulation of the transmission. Ann Parasitol Hum Comp. 65:193–199.

Doudoumis V, et al. 2012. Detection and characterization of *Wolbachia* infections in laboratory and natural populations of different species of tsetse flies (genus *Glossina*). BMC Microbiol. 12(Suppl 1):S3.

Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113.

Fenn K, et al. 2006. Phylogenetic relationships of the *Wolbachia* of nematodes and arthropods. PLoS Pathog. 2:e94.

Ferri E, et al. 2011. New insights into the evolution of *Wolbachia* infections in filarial nematodes inferred from a large range of screened species. PLoS One 6:e20843.

Foster J, et al. 2005. The Wolbachia genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. PLoS Biol. 3: e121.

Godel C, et al. 2012. The genome of the heartworm, *Dirofilaria immitis*, reveals drug and vaccine targets. FASEB J. 26:4650–4661.

Haegeman A, et al. 2009. An endosymbiotic bacterium in a plant-parasitic nematode: member of a new *Wolbachia* supergroup. Int J Parasitol. 39:1045–1054.

Hoerauf A, et al. 1999. Tetracycline therapy targets intracellular bacteria in the filarial nematode *Litomosoides sigmodontis* and results in filarial infertility. J Clin Invest. 103:11–18.

Hoffmann W, et al. 2000. *Litomosoides sigmodontis* in mice: reappraisal of an old model for filarial research. Parasitol Today. 16:387–389.

Klasson L, et al. 2009. The mosaic genome structure of the Wolbachia wRi strain infecting *Drosophila simulans*. Proc Natl Acad Sci U S A. 106: 5725–5730.

Kumar S, Blaxter ML. 2011. Simultaneous genome sequencing of symbionts and their hosts. Symbiosis 55:119–126.

Kumar S, et al. 2009. AIR: a batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses. BMC Bioinformatics 10:357.

Lefoulon E, et al. 2012. A new type F *Wolbachia* from Splendidofilariinae (Onchocercidae) supports the recent emergence of this supergroup. Int J Parasitol. 42:1025–1036.

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

Lo N, et al. 2007. Taxonomic status of the intracellular bacterium *Wolbachia pipientis*. Int J Syst Evol Microbiol. 57:654–657.

Luo R, et al. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1:18.

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

Renvoise A, Merhej V, Georgiades K, Raoult D. 2011. Intracellular Rickettsiales: insights into manipulators of eukaryotic cells. Trends Mol Med. 17:573–583.

Ronquist F, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 61: 539–542.

Salichos L, Rokas A. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 497:327–331.

Schlüter PM, Harris SA. 2006. Analysis of multilocus fingerprinting data sets containing missing data. Mol Ecol Notes. 6:569–572.

Simpson JT, et al. 2009. ABySS: a parallel assembler for short read sequence data. Genome Res. 19:1117–1123.

Slatko BE, Taylor MJ, Foster JM. 2010. The *Wolbachia* endosymbiont as an anti-filarial nematode target. Symbiosis 51:55–65.

Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. Syst Biol. 57:758–771.

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

Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 28:33–36.

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

Werren JH, Baldo L, Clark ME. 2008. *Wolbachia*: master manipulators of invertebrate biology. Nat Rev Microbiol. 6:741–751.

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

Xia X, Xie Z. 2001. DAMBE: software package for data analysis in molecular biology and evolution. J Hered. 92:371–373.

Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821–829.

Associate editor: John McCutcheon