Microbiology (2005), **151**, 4015–4022 **DOI** 10.1099/mic.0.28313-0

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

M. Casiraghi,[1]† S. R. Bordenstein,[2] L. Baldo,[1]‡ N. Lo,[3] T. Beninati,[1] J. J. Wernegreen,[2] J. H. Werren[4] and C. Bandi[1]

Correspondence
N. Lo
nathan@usyd.edu.au

[1]Dipartimento di Patologia Animale, Igiene e Sanità Pubblica Veterinaria, Sezione di Patologia Generale e Parassitologia, Università degli Studi di Milano, Milano, Italy

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

[3]School of Biological Sciences, The University of Sydney, NSW 2006, Australia

[4]Department of Biology, University of Rochester, Rochester, NY, USA

Current phylogenies of the intracellular bacteria belonging to the genus *Wolbachia* identify six major clades (A–F), termed 'supergroups', but the branching order of these supergroups remains unresolved. Supergroups A, B and E include most of the wolbachiae found thus far in arthropods, while supergroups C and D include most of those found in filarial nematodes. Members of supergroup F have been found in arthropods (i.e. termites), and have previously been detected in the nematode *Mansonella ozzardi*, a causative agent of human filariasis. To resolve the phylogenetic positions of *Wolbachia* from *Mansonella* spp., and other novel strains from the flea *Ctenocephalides felis* and the filarial nematode *Dipetalonema gracile*, the authors generated new DNA sequences of the *Wolbachia* genes encoding citrate synthase (*gltA*), heat-shock protein 60 (*groEL*), and the cell division protein *ftsZ*. Phylogenetic analysis confirmed the designation of *Wolbachia* from *Mansonella* spp. as a member of the F supergroup. In addition, it was found that divergent lineages from *Dip. gracile* and *Cte. felis* lack any clear affiliation with known supergroups, indicating further genetic diversity within the *Wolbachia* genus. Finally, although the data generated did not permit clear resolution of the root of the global *Wolbachia* tree, the results suggest that the transfer of *Wolbachia* spp. from arthropods to nematodes (or vice versa) probably occurred more than once.

Received 30 June 2005
Revised 12 September 2005
Accepted 12 September 2005

INTRODUCTION

The genus *Wolbachia* encompasses obligately intracellular bacteria that are cytoplasmically transmitted in arthropods and filarial nematodes (Werren, 1997; Bandi *et al.*, 2001; Bordenstein *et al.*, 2003). Unlike most other obligately intracellular bacteria, the genus *Wolbachia* forms a monophyletic clade comprising both mutualistic and parasitic lineages that showcase the diversity of symbiotic associations. In arthropods, *Wolbachia* spp. are commonly known as reproductive parasites, since they distort host reproductive strategies to selfishly enhance their maternal transmission into the next generation (Werren, 1997). *Wolbachia* spp. have a remarkably high prevalence in some arthropod groups (Jeyaprakash & Hoy, 2000). In contrast, the *Wolbachia* lineages infecting filarial nematodes are beneficial

†Present address: Dipartimento di Biotecnologie e Bioscienze, Università degli Studi di Milano Bicocca, Milano, Italy.

‡Present address: Department of Biology, University of California, 900 University Avenue, Riverside, CA, USA.

Abbreviations: BI, Bayesian inference; ML, maximum-likelihood; TBR, tree bisection and reconnection.

The phylogeny inferred from the concatenated dataset (*gltA, groEL, ftsZ*), rooted with two outgroup species (*Anaplasma marginale* and *Ehrlichia ruminantium*), is shown in Supplementary Fig. S1, available with the online version of this paper.

The GenBank/EMBL/DDBJ accession numbers for the nucleotide sequences reported in this paper are listed in Table 1.

symbionts required for host fertility and larval viability (Bandi *et al.*, 2001).

The gene phylogenies of the genus *Wolbachia* have shown the existence of six major clades (A–F), which, in the absence of a formal species description, have been named 'supergroups' (see Lo *et al.*, 2002, and references therein). The monophyly of some of these supergroups is not firmly established (i.e. supergroup F; Lo *et al.*, 2002), and their relationships are not resolved. Supergroups A and B include most of the parasitic *Wolbachia* spp. thus far found in arthropods (Werren *et al.*, 1995). Supergroups C and D include the majority of the *Wolbachia* spp. found in filarial nematodes (Bandi *et al.*, 1998) (though some nematodes have been found to lack *Wolbachia* spp.; Bordenstein *et al.*, 2003). The E supergroup encompasses *Wolbachia* spp. from primitively wingless insects, the springtails (Collembola) (Vandekerckhove *et al.*, 1999; Czarnetzki & Tebbe, 2004). Members of supergroup F are known to infect arthropods (termites), and recent studies suggest that they also infect the filarial parasite *Mansonella ozzardi* (Casiraghi *et al.*, 2001a; Lo *et al.*, 2002). More recently, a new supergroup, named G, has been proposed for *Wolbachia* spp. of certain Australian spiders (Rowley *et al.*, 2004). Divergent lineages have been detected in fleas (Fischer *et al.*, 2002; Gorham *et al.*, 2003), and in the filarial nematode *Dipetalonema gracile* (Casiraghi *et al.*, 2004); however, they have not been labelled new supergroups, since only 16S rRNA has been obtained. In summary, the overall diversity and phylogeny of the genus *Wolbachia* is increasingly informative, but also incomplete, with possible new hosts (*Mansonella* spp.) and new lineages (i.e. in *Dip. gracile* and fleas) still requiring confirmation.

In this report, we address two key issues facing the molecular taxonomy of the genus *Wolbachia*. First, we verify the presence of supergroup F in the genus *Mansonella*, and in termites with additional host species and *Wolbachia* genes. Supergroup F is of particular interest for understanding the origins and evolutionary relationships of *Wolbachia* spp. because it is the only supergroup reported to infect both nematodes and arthropods (Lo *et al.*, 2002). If the supergroup's dual host range is confirmed, it could represent a clear and relatively recent transfer of *Wolbachia* spp. between the nematode and arthropod phyla. The apparent novelty of this group's host range is tentative, and based upon only a single gene analysis of 16S rDNA sequences spanning a few termite species, and microfilariae from one nematode species, *M. ozzardi*. With the small possibility of a false-positive result in *Mansonella* due to cross-contamination with *Wolbachia*-infected microarthropods (i.e. mites), it is necessary to expand the taxon and gene sampling for this supergroup. We do so by developing new primers and PCR protocols for the genes encoding citrate synthase (*gltA*), heat-shock protein 60 (*groEL*) and the cell division protein *ftsZ* of *Wolbachia*. Second, we investigate the phylogenetic placement of *Wolbachia* from the cat flea (*Ctenocephalides felis*) and the nematode *Dip. gracile* – two lineages that may be genetically distinct from the other supergroups. In investigating both these issues, we contribute to the reconstruction of the overall phylogeny of the *Wolbachia* genus, and extend the available *Wolbachia* gene dataset (28 new *gltA* sequences; 30 *groEL*; 2 *ftsZ*).

METHODS

DNA isolation, amplification and sequencing. A total of 25 insect and 15 filarial nematode species harbouring *Wolbachia pipientis* were included in the study (see Table 1). DNA extraction from arthropod specimens was performed using either a standard phenol/chloroform procedure (Sambrook & Russell, 2001), or the DNeasy Tissue Kit (Qiagen). For nematode specimens, crude DNA preparations were obtained through proteinase K treatment, according to Bandi *et al.* (1994). All arthropod and nematode specimens used in this study were infected with a single strain of *Wolbachia*.

gltA was amplified and sequenced using the following primers: WgltAF1, 5′-TAC GAT CCA GGG TTT GTT TCT AC-3′; WgltARev1, 5′-CTC ATT AGC TCC ACC GTG TG-3′; WgltARev2, 5′-CAT TTC ATA CCA CTG GGC AA-3′. These primers were designed on the basis of homologous regions of *gltA* among *Anaplasma marginale* (AF304146), *Ehrlichia ruminantium* (AF304140) and *W. pipientis* from *Drosophila melanogaster* (AE017260). *groEL* was amplified and sequenced using the following primers: WgroF1, 5′-GGTGAGCA-GTTGCAAGAAGC-3′; WgroRev1, 5′-AGATCTTCCATCTTGAT-TCC-3′; WgroF1deg, 5′-GGT GAG CAG TT(GA) CA(GA) (CG)AA GC-3′; WgroR1d, 5′-AG(GA) TCT TCC AT(CT) TT(AG) ATT CC-3′; WgroR2d, 5′-GGT ATT (AG)TC TTT AGT (AG)A(TC) TTT AAC-3′; WgroR4d, 5′-TTT (AG)AC AGC ATC AGC AAT-3′. Primers were designed on the basis of *groEL* regions conserved among the *Wolbachia* spp. from *Onchocerca volvulus*, *Litomosoides sigmodontis*, *Dirofilaria immitis*, *Brugia malayi* and *Dro. melanogaster* (Y09416; AF409113; AJ558023; AF373870; AE017257). *ftsZ* and 16S rDNA were amplified and sequenced using the primers and conditions described in Casiraghi *et al.* (2001b, 2004).

PCR amplifications were performed in 20–50 μl volumes, under the following final conditions: 1 × Eppendorf buffer including 1·5 mM $MgCl_2$, 0·2 μM of each dNTP, 1 μM each of forward and reverse primers, and 0·5 units MasterTaq (Eppendorf). The thermal profiles we used were: (1) *gltA*, 94 °C 45 s, 52 °C 45 s, and 72 °C 90 s, for 40 cycles; (2) *groEL*, 94 °C 45 s, 60 °C 45 s, 72 °C 80 s, for 5 cycles, and 94 °C 45 s, 55 °C 45 s, and 72 °C 80 s, for 34 cycles; (3) *ftsZ*, 94 °C 30 s, 60 °C 45 s, 72 °C 90 s, for 5 cycles, and 94 °C 30 s, 57 °C 45 s, and 72 °C 90 s, for 34 cycles.

PCR products were purified and sequenced bidirectionally on an ABI 3730 or 3310 automated sequencer using Big Dye v2.0 or v3.0 (Applied Biosystems). The 60 new gene sequences generated in this study were deposited in the EMBL database (see Table 1 for accession numbers).

The *gltA*, *groEL* and *ftsZ* sequences were not obtained from all the taxa included in this study, mainly due to the scarcity of certain specimens, and amplification/sequencing problems for some of the species examined. For example, despite attempts with various primer combinations and PCR conditions, we were unable to amplify *Wolbachia groEL* and *ftsZ* sequences from *Folsomia candida* and *Dip. gracile*, respectively. Based on their high divergence at other loci (see Results), these genes may have accumulated substitutions in the regions of the primers tested.

Sequence alignment and phylogenetic analysis. DNA alignments were based on translated proteins, and were performed

Table 1. Host taxonomic details and *Wolbachia* sequence accession numbers

Twenty-five arthropod and 15 nematode host species are listed in the table. Accession numbers in bold represent new sequence data generated for this study.

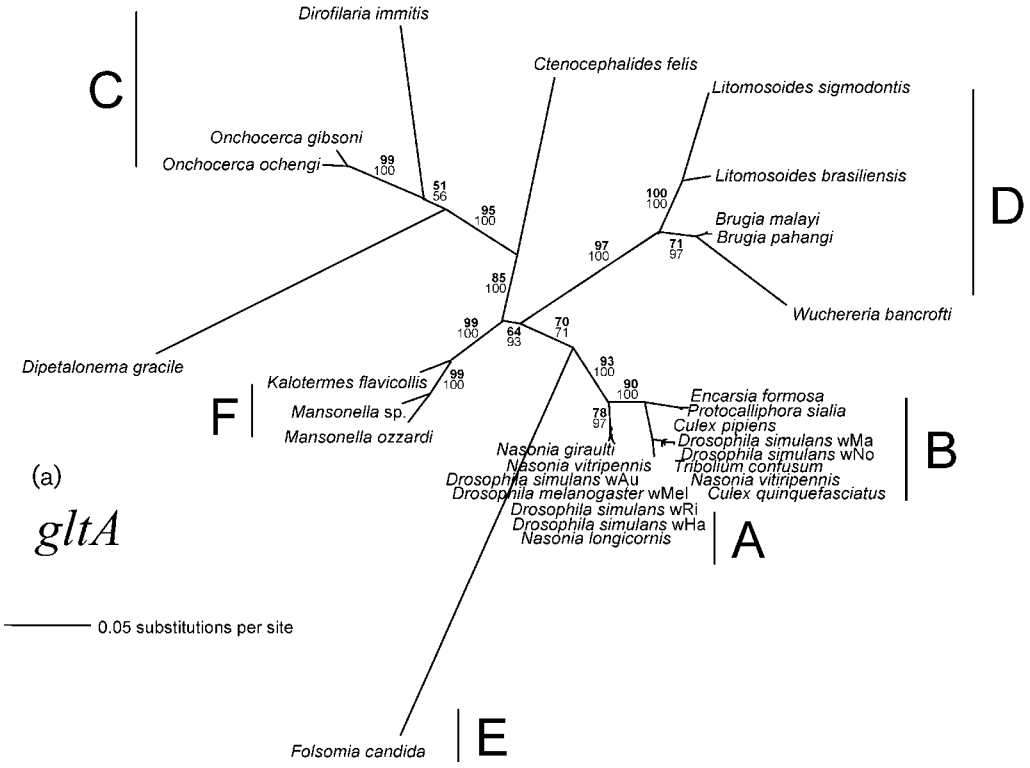
Host			Accession nos for *Wolbachia*:		
Phylum	**Order**	**Species**	***gltA***	***groEL***	***ftsZ***
Arthropoda	Hymenoptera	*Asobara tabida*	–	**AY714809**	–
	Dictyoptera	*Coptotermes lacteus*	–	**AJ627385**	–
	Dictyoptera	*Coptotermes acinaciformis*	–	**AJ627384**	–
	Diptera	*Culex pipiens*	**AY714785**	–	–
	Diptera	*Culex quinquefasciatus*	**AY714789**	**AY714804**	–
	Siphonaptera	*Ctenocephalides felis*	**AJ609650**	**AJ609659**	**AJ628415**
	Diptera	*Drosophila melanogaster* wMel	AE017260	AE017257	U28189*
	Diptera	*Drosophila simulans* wAu	**AY714792**	**AY714807**	AY227739
	Diptera	*Drosophila simulans* wRi	**AY714791**	**AY714806**	U28178
	Diptera	*Drosophila simulans* wHa	**AY714790**	**AY714805**	AY508998
	Diptera	*Drosophila simulans* wMa	**AY714786**	**AY714799**	AY508999
	Diptera	*Drosophila simulans* wNo	**AY714787**	**AY714800**	AY509001
	Hymenoptera	*Encarsia formosa*	**AY714783**	**AY714797**	U28196
	Collembola	*Folsomia candida*	**AJ609649**	–	AJ344216
	Dictyoptera	*Kalotermes flavicollis*	**AJ609651**	**AJ609660**	AJ292345
	Hymenoptera	*Leptopilina australis*	–	**AY714802**	–
	Hymenoptera	*Mellitobia digitata*	–	**AY714808**	–
	Dictyoptera	*Microcerotermes* sp.	–	**AJ628411**	AJ292346
	Hymenoptera	*Nasonia giraulti* 16.2 RV2D	**AY714793**	**AY714810**	U28182
	Hymenoptera	*Nasonia longicornis* 2.1 IV7D	**AY714794**	**AY714811**	–
	Hymenoptera	*Nasonia vitripennis* 12.1 R511D	**AY714795**†	**AY714812**†	U28188†
	Hymenoptera	*Nasonia vitripennis* 4.9 R511D	**AY714782**‡	**AY714796**‡	U28205‡
	Hymenoptera	*Protocalliphora sialia* 00-189	**AY714788**	**AY714801**	U28202
	Coleoptera	*Tribolium confusum*	**AY714784**	**AY714798**	U28194
	Hymenoptera	*Trichogramma cordubensis*	–	**AY714803**	–
Nematoda	Spirurida	*Brugia malayi*	**AJ609643**	AF373870	AJ010269
	Spirurida	*Brugia pahangi*	**AJ609642**	**AJ609654**	AJ010270
	Spirurida	*Dipetalonema gracile*	**AJ609648**	**AJ609658**	–
	Spirurida	*Dirofilaria immitis*	**AJ609641**	AJ558023	AJ010272
	Spirurida	*Dirofilaria repens*	–	**AJ609653**	AJ010273
	Spirurida	*Litomosoides brasiliensis*	**AJ609646**	**AJ609655**	–
	Spirurida	*Litomosoides hamletti*	–	**AJ609656**	–
	Spirurida	*Litomosoides sigmodontis*	**AJ609645**	AF409113	AJ010271
	Spirurida	*Mansonella ozzardi*	**AJ609647**	**AJ609657**	–
	Spirurida	*Mansonella* sp.	**AJ628413**	**AJ628412**	**AJ628414**
	Spirurida	*Onchocerca gibsoni*	**AJ609639**	**AJ609652**	AJ010267
	Spirurida	*Onchocerca ochengi*	**AJ609640**	–	AJ010266
	Spirurida	*Onchocerca gutturosa*	–	–	AJ010268
	Spirurida	*Onchocerca volvulus*	–	Y09416	AJ276501
	Spirurida	*Wuchereria bancrofti*	**AJ609644**	–	AF081198

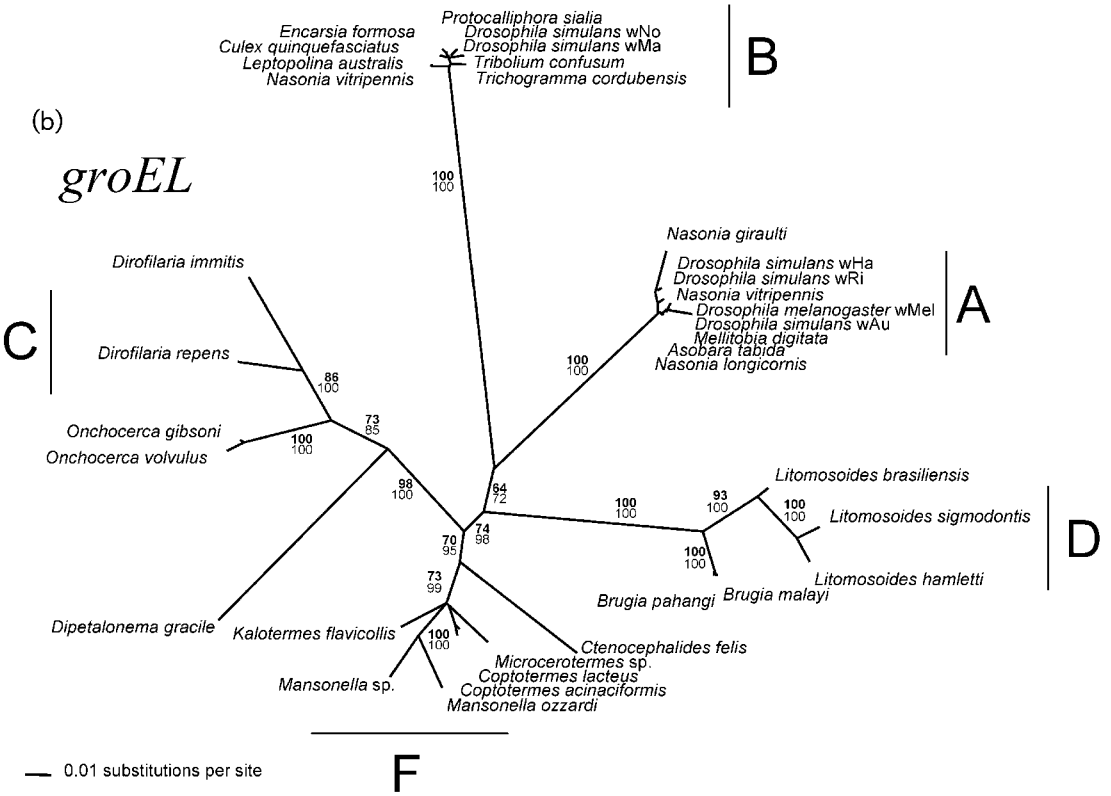
*Strain BMH6′.

†Supergroup A sequence (see Fig. 1).

‡Supergroup B sequence (see Fig. 1).

manually in MacClade 4.05. A total of four straightforward alignments were generated, with very few gaps. Total lengths of alignments were 639 bp for *gltA*, 876 bp for *groEL*, 735 bp for *ftsZ*, and 2250 bp for a concatenated alignment that included taxa for which two out of three gene sequences were available. Each of the alignments has been deposited in the EBI alignment databases, with the following accession numbers: ALIGN_000922 (*gltA*), ALIGN_000923 (*groEL*) and ALIGN_000920 (*ftsZ*).


Dirofilaria immitis
Ctenocephalides felis
Litomosoides sigmodontis
Onchocerca gibsoni
Onchocerca ochengi
Litomosoides brasiliensis
Brugia malayi
Brugia pahangi
Wuchereria bancrofti
Dipetalonema gracile
Kalotermes flavicollis
Mansonella sp.
Mansonella ozzardi
Encarsia formosa
Protocalliphora sialia
Culex pipiens
Drosophila simulans wMa
Drosophila simulans wNo
Nasonia girault
Nasonia vitripennis
Tribolium confusum
Drosophila simulans wAu
Nasonia vitripennis
Drosophila melanogaster wMel
Culex quinquefasciatus
Drosophila simulans wRi
Drosophila simulans wHa
Nasonia longicornis
Folsomia candida
C
D
F
B
A
E
(a)
gltA
0.05 substitutions per site


Encarsia formosa
Culex quinquefasciatus
Leptopolina australis
Nasonia vitripennis
Protocalliphora sialia
Drosophila simulans wNo
Drosophila simulans wMa
Tribolium confusum
Trichogramma cordubensis
B
(b)
groEL
Nasonia girault
Dirofilaria immitis
Drosophila simulans wHa
Drosophila simulans wRi
Nasonia vitripennis
Drosophila melanogaster wMel
Drosophila simulans wAu
Mellitobia digitata
Asobara tabida
Nasonia longicornis
A
C
Dirofilaria repens
Onchocerca gibsoni
Onchocerca volvulus
Litomosoides brasiliensis
Litomosoides sigmodontis
Litomosoides hamletti
D
Brugia pahangi
Brugia malayi
Dipetalonema gracile
Kalotermes flavicollis
Ctenocephalides felis
Microcerotermes sp.
Coptotermes lacteus
Coptotermes acinaciformis
Mansonella sp.
Mansonella ozzardi
0.01 substitutions per site
F

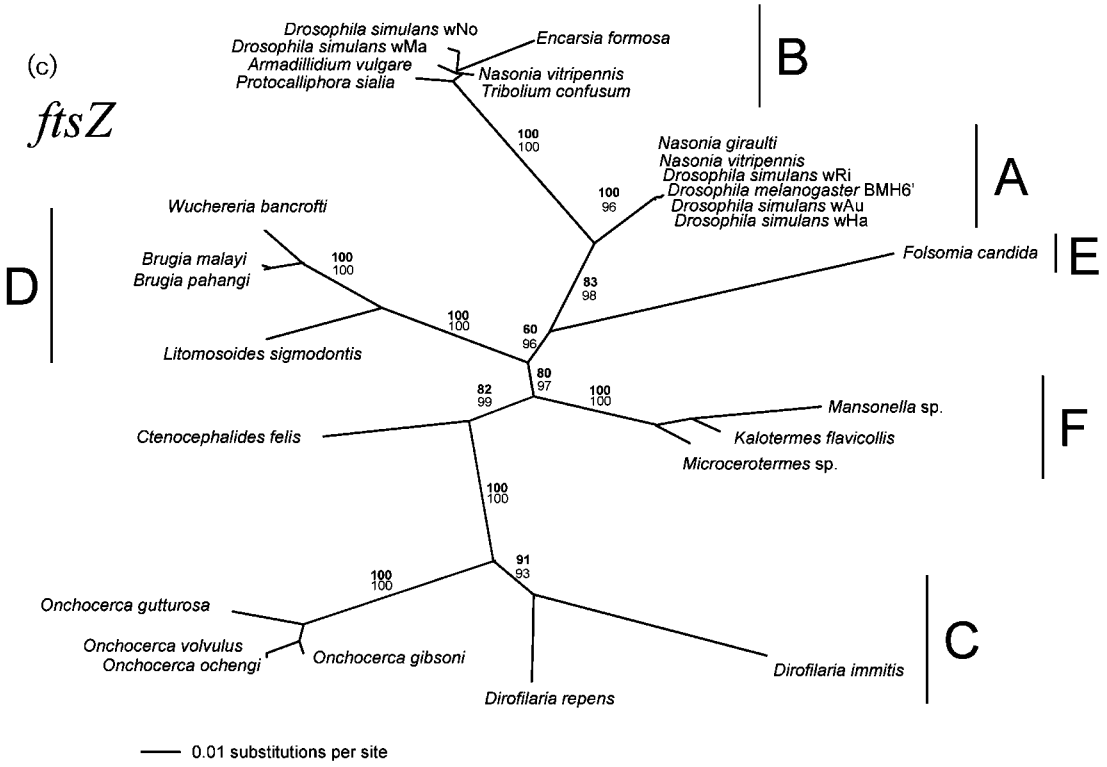


Fig. 1. Phylogenetic reconstructions based on individual analyses of (a) *gltA*, (b) *groEL* and (c) *ftsZ*. Taxa include representatives of supergroups A–F, *Wolbachia* spp. of *Cte. felis*, and *Wolbachia* spp. of *Dip. gracile*. Data for *groEL* of *F. candida*-derived isolates and *ftsZ* of *Dip. gracile*-derived isolates were not obtained (see text). For each dataset, phylogenies based on ML and BI were topologically identical. The phylogenies presented reflect ML branch lengths. Nodes are marked by their ML bootstrap values (top number, in bold) estimated from 100 bootstrap replicates, each of which was based on 10 random starting trees, and Bayesian posterior probabilities (bottom number) based on two independent runs of three million generations each. For clarity of presentation, the figure excludes confidence values for some recent divergences within supergroups.

Phylogenetic analyses were conducted using maximum-likelihood (ML) and Bayesian inference (BI) methods. For ML, the appropriate models of sequence evolution for each dataset were estimated via likelihood ratio tests in the program modeltest 3.06 (Posada & Crandall, 1998). The models selected were GTR+I+G for *gltA*, GTR+I+G for *groEL*, GTR+G for *ftsZ*, and GTR+I+G for the combined dataset. ML heuristic searches were performed using 100 random taxon addition replicates, with tree bisection and reconnection (TBR) branch swapping. ML bootstrap support was determined using 100 bootstrap replicates, each using 10 random taxon addition replicates with TBR branch swapping. Searches were performed in parallel on a Beowulf cluster using custom software with PAUP version 4.0b10 (Swofford, 2002). ML analysis of the combined dataset was performed with and without the outgroup taxa *A. marginales* and *E. ruminantium* to estimate rooted and unrooted phylogenies, respectively.

Initial BI analyses were performed using MrBayes 3.0 (Ronquist & Huelsenbeck 2003), with simultaneous estimation of genealogies and DNA sequence parameters. In the analyses of the *gltA*, *groEL*, *ftsZ* and combined datasets, these initial analyses ran for 100 000 generations, with trees sampled every 100 generations. The first 500 of the 1000 sampled trees were considered the 'burn-in', and were discarded. From the remaining 500 trees, 50 % majority-rule consensus trees were generated. For each single-gene dataset, topological agreement and overall similarity of posterior probabilities among five independent runs indicated adequate convergence and mixing. As for the ML analyses, the combined dataset was analysed with and without the presence of outgroups.

Final BI analyses for each single-gene dataset and the combined dataset (without outgroups) consisted of two independent runs with 3 000 000 generations, and four chains per run, using Mr Bayes version 3.1.1 (Ronquist & Huelsenbeck, 2003). 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'). When applicable (see below), the site-specific rates model was allowed to vary among data partitions (prset ratepr=variable). Trees were sampled every 100 generations, resulting in 30 000 trees per run (60 000 per analysis consisting of two independent runs). The first third of these trees was considered the 'burn-in', and was discarded. Posterior probabilities were estimated from the consensus of the remaining 40 000 trees. Partitioning of datasets by codon positions or (in the case of combined dataset) by gene did not affect the resulting topology, and had minimal effects on posterior probabilities. Posterior probabilities presented here (Figs 1 and 2) are those from the 3 000 000-generation BI analyses performed as described above, in which data were partitioned by first and second positions in one partition, and third positions in a second partition.

RESULTS AND DISCUSSION

Host range and taxonomic positioning of supergroup F

The filarial nematode *Mansonella* sp. included in this study represents a different species than the human parasite *M. ozzardi* sampled previously (Casiraghi *et al.*, 2001a). We found that *Wolbachia* spp. of the two *Mansonella* species group together and, as shown previously for *M. ozzardi* (Casiraghi *et al.*, 2001a), also group with *Wolbachia* spp. from termites (see Figs 1 and 2, and text below). In addition, new *Wolbachia groEL* sequences from the termites *Coptotermes lacteus* and *Coptotermes acinaciformis* confirmed both the presence of *Wolbachia* spp. in termites, and their phylogenetic proximity to *Wolbachia* spp. of *Mansonella* spp. In all analyses, *Wolbachia* sequences from termites and the two *Mansonella* spp. consistently clustered together, forming a separate clade from the other supergroups. This pattern confirms the existence of the F supergroup of the genus *Wolbachia* and its dual host range spanning nematodes (*Mansonella*) and arthropods (termites). In addition to *Mansonella* spp. and termites, the F supergroup appears to encompass *Wolbachia* spp. of other hosts, including the human bed bug (*Cimex lectulatis*), the cliff swallow bug (*Oeciacus vicarius*), and the beetle *Rhinocyllus conicus* (Rasgon & Scott, 2004; Bandi *et al.*, 1997; Lo *et al.*, 2002).

In previous PCR screens for *Wolbachia* spp. in *Mansonella* spp. (Casiraghi *et al.*, 2001a; Lo *et al.*, 2002), we used microfilariae of *M. ozzardi* (i.e. 250-μm-long juveniles) collected in South America from human blood. To rule out the possibility that previously identified *Wolbachia* spp. arose from some other contaminant of the blood, here we used adult specimens of *Mansonella* spp. (3–5 cm long) from Sika deer (*Cervus nippon*). To rule out the possibility that termite *Wolbachia* spp. actually infect nematodes that might infest the termites, we performed a PCR screen of the four termite species using general primers for nematode 18S rDNA. No evidence was found for the presence of nematodes within the termites (data not shown).

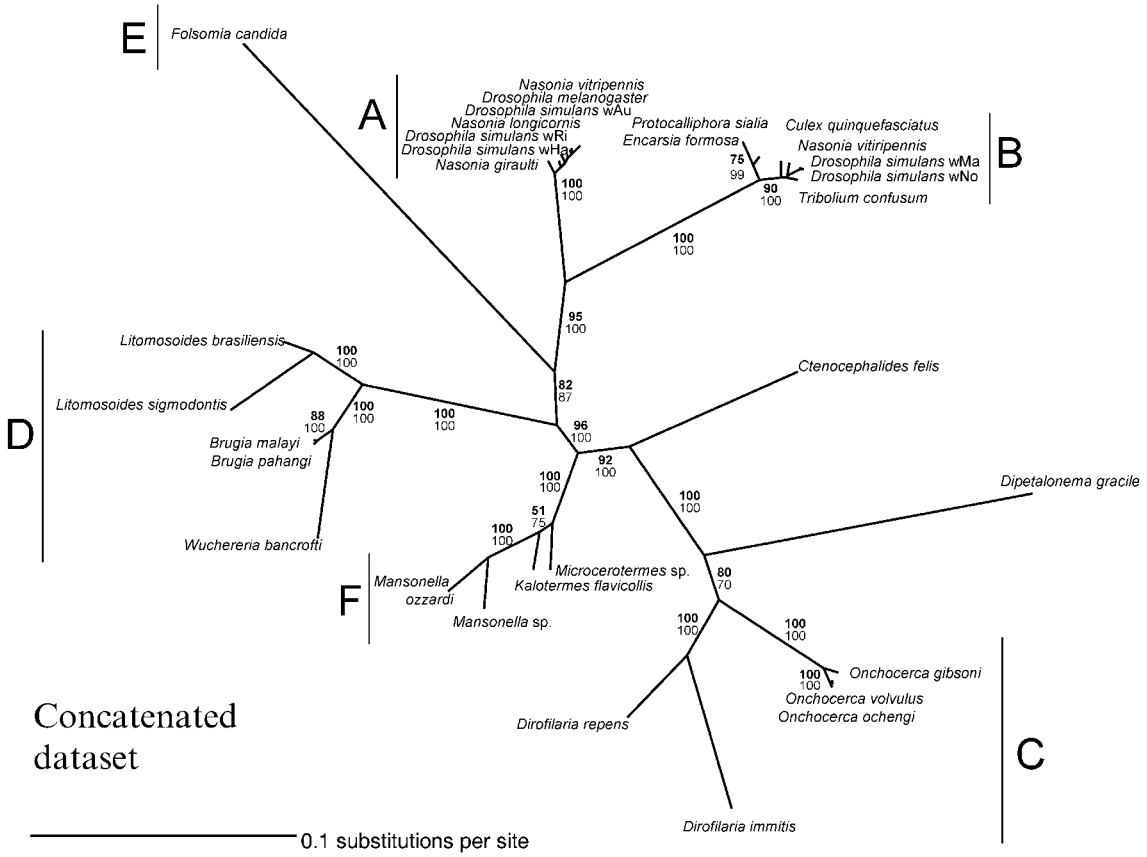


Fig. 2. *W. pipientis* phylogeny based on the concatenated datasets of *gltA*, *groEL* and *ftsZ* gene sequences (2250 bp). This tree topology was inferred using both ML and BI, and the phylogeny presented reflects ML branch lengths. Nodes are marked by their ML bootstrap values (top number, in bold) estimated from 100 bootstrap replicates, each of which was based on 10 random starting trees, and Bayesian posterior probabilities (bottom number) based on two independent runs of three million generations each (bottom number). For clarity of presentation, the figure excludes confidence values for some recent divergences within subgroups.

Phylogeny of *Wolbachia* genus supergroups

Below we present phylogenetic relationships among *Wolbachia* genus supergroups, and discuss the implications of the F supergroup's association with both insects and nematodes. Phylogenies based on single-gene datasets were generally well resolved (Fig. 1a–c). For any given gene, ML and BI analyses gave identical topologies. Multiple BI runs for each dataset resulted in identical trees, and similar posterior probabilities, indicating convergence among runs. The *gltA* tree clearly distinguishes supergroups A–F (Fig. 1a), in agreement with previous studies (Werren *et al.*, 1995; Bandi *et al.*, 1998; Lo *et al.*, 2002). Supergroups A and B include *Wolbachia* spp. from arthropods only, while known members of supergroups C and D are restricted to filarial nematodes. *Wolbachia* spp. from the Collembolan *F. candida* represent a divergent lineage, named supergroup E by Vandekerckhove *et al.* (1999). As discussed above, supergroup F comprises representatives of filarial nematodes (*Mansonella* spp.) and the termite *Kalotermes flavicollis*. Notably, bacteria from the filarial nematode *Dip. gracile* and the flea *Cte. felis* represent two divergent branches. The long branch leading to the *Wolbachia* spp. of *Dip. gracile* is the sister lineage to supergroup C. Phylogenies based on *groEL* (Fig. 1b) and *ftsZ* (Fig. 1c) also distinguish major supergroups (A–D and F in the *groEL* tree, and A–F in the *ftsZ* tree). As in the *gltA* tree, the *groEL* and *ftsZ* phylogenies indicate that *Wolbachia* spp. from *Dip. gracile* and *Cte. felis* are two separate branches that are quite divergent from known subgroups, and termites are found to group with *Mansonella* spp.

Because there is only one representative in each of the two novel lineages, there are insufficient data to determine whether *Wolbachia* spp. of *Dip. gracile* and *Cte. felis* represent new supergroups. Further sampling and gene sequencing within taxonomic groups related to these two organisms (e.g. Kikuchi & Fukatsu, 2003; Czarnetzki & Tebbe, 2004) may help to decide the status of these lineages. In addition to *Cte. felis*, numerous other flea species also host *Wolbachia* spp. However, based on a relatively short fragment of 16S rDNA (Gorham *et al.*, 2003), these *Wolbachia* spp. do not form a single cluster signifying a coherent supergroup of flea associates. The results found by Gorham *et al.* (2003) require further investigations involving other gene sequences to determine their supergroup status.

The results obtained using the concatenated dataset are consistent with those based on single-gene analyses, with the *Wolbachia* spp. from *Dip. gracile* showing a distant but well-supported grouping with members of supergroup C (Fig. 2). In addition, most datasets support the position of *Wolbachia* spp. from *Cte. felis* just outside the *Dip. gracile* + supergroup C clade, within a major clade also encompassing the F and D supergroups. The exception is the *groEL* phylogeny, in which *Cte. felis* apparently falls just outside the F subgroup, rather than the *D. gracile* + supergroup C clade. It is possible this discrepancy is an artefact of excluding *F. candida* from the *groEL* analysis. In sum, with respect to the relationships among the main lineages, the phylogeny of the unrooted tree of the concatenated dataset was found to be identical to that of *gltA* and *ftsZ*, and, with the exception noted above, to that of *groEL* phylogeny.

The phylogeny inferred from the concatenated dataset was rooted with two outgroup species (*A. marginale* and *E. ruminantium*) provides no clear insight into the root of the global tree of diverse *Wolbachia pipientis* strains (see Supplementary Fig. S1 with the online version of this paper). While the 'best' rooted tree under ML and BI criteria suggests that supergroup B is the earliest-branching lineage within the genus *Wolbachia*, support for this hypothesis is very weak (e.g. 54 % bootstrap value). Using a Shimodaira–Hasegawa test (Shimodaira & Hasegawa, 1999), we compared the relative support for 18 topologies in which the root of the tree (leading to the two outgroups) was placed as a sister group to each of the different supergroups, as well as to *Wolbachia* spp. from *Dip. gracile* and *Cte. felis* (see Lo *et al.*, 2002). These phylogenies were statically indistinguishable, with likelihood values ranging from −14621·82 to −14627·88, and *P* values ranging from 0·916 to 0·148. That is, in no case was one root placement significantly more likely than any other. The rooting of the *Wolbachia pipientis* tree remains problematic. Since outgroups are extremely divergent from *Wolbachia*, they have not been useful in resolving the basal relationships among supergroups (for details see Lo *et al.*, 2002).

As stated above, supergroup F is of particular interest regarding the macroevolution of *Wolbachia*. This clustering of *Wolbachia* spp. from arthropod and nematode hosts suggests that an independent horizontal transfer of the bacterium between these host phyla might have occurred much more recently than ~100 million years ago, a proposed date for the first transfer based on the number of substitutions between supergroups A versus D (Bandi *et al.*, 1998). In the absence of a stable root for the overall tree of the *Wolbachia* genus, the original host group and interaction type (mutualistic or parasitic) of *Wolbachia* remains uncertain, as does the direction of host switching between phyla.

ACKNOWLEDGEMENTS

We thank Odile Bain and Coralie Martin (Muséum National d'Histoire Naturelle France) for samples of *Dip. gracile*, *Litomosoides hamletti* and *Litomosoides brasiliensis*, Shigeniko Uni (Osaka City University Medical School) for *Mansonella* sp., Vittorio Parisi (Università degli Studi di Parma) for *F. candida*, and John W. McCall (The University of Georgia) for *Cte. felis*. We are most grateful to C. Bazzocchi for technical assistance. We thank P. H. Degnan for superb technical assistance, including DNA sequencing. This work was partially supported by grants to C. B. (MIUR-FIRB RBAU01TA3W), J. J. W. from the National Institutes of Health (R01 GM62626-01) and the NASA Astrobiology Institute (NNA04CC04A), and J. H. W. and J. J. W. from the National Science Foundation (EF-0328363). This work was performed while S. R. B. held a National Research Council Research Associateship Award at the Marine Biological Laboratory. N. L. was supported by the Australian Research Council.

REFERENCES

Bandi, C., Damiani, G., Magrassi, L., Grigolo, A., Fani, R. & Sacchi, L. (1994). Flavobacteria as intracellular symbionts in cockroaches. *Proc Biol* **Sci257**, 43–48.

Bandi, C., Sironi, M., Nalepa, C. A., Corona, S. & Sacchi, L. (1997). Phylogenetically distant intracellular symbionts in termites. *Parassitologia* **39**, 71–75.

Bandi, C., Anderson, T. J. C., Genchi, C. & Blaxter, M. L. (1998). Phylogeny of *Wolbachia* in filarial nematodes. *Proc Biol Sci* **265**, 2407–2413.

Bandi, C., Trees, A. J. & Brattig, N. W. (2001). *Wolbachia* in filarial nematodes: evolutionary aspects and implications for the pathogenesis and treatment of filarial diseases. *Vet Parasitol* **98**, 215–238.

Bordenstein, S. R., Fitch, D. H. A. & Werren, J. H. (2003). Absence of *Wolbachia* in nonfilariid nematodes. *J Nematol* **35**, 266–270.

Casiraghi, M., Favia, G., Cancrini, G., Bartoloni, A. & Bandi, C. (2001a). Molecular identification of *Wolbachia* from the filarial nematode *Mansonella ozzardi*. *Parasitol Res* **87**, 417–420.

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

Casiraghi, M., Bain, O., Guerrero, R., Martin, C., Pocacqua, V., Gardner, S. L., Franceschi, A. & Bandi, C. (2004). Mapping the presence of *Wolbachia pipientis* on the phylogeny of filarial nematodes: evidence for symbiont loss during evolution. *Int J Parasitol* **34**, 191–203.

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

Fischer, P., Schmetz, C., Bandi, C., Bonow, I., Mand, S., Fischer, K. & Buttner, D. W. (2002). *Tunga penetrans*: molecular identification of *Wolbachia* endobacteria and their recognition by antibodies against proteins of endobacteria from filarial parasites. *Exp Parasitol* **102**, 201–211.

Gorham, C. H., Fang, Q. Q. & Durden, L. A. (2003). *Wolbachia* endosymbionts in fleas (Siphonaptera). *J Parasitol* **89**, 283–289.

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

Kikuchi, Y. & Fukatsu, T. (2003). Diversity of *Wolbachia* endosymbionts in heteropteran bugs. *App Environ Microbiol* **69**, 6082–6090.

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

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

Rasgon, J. L. & Scott, T. W. (2004). Characterization of *Wolbachia* symbionts infecting *Cimex lectularis* L. and *Oeciacus vicarius* Horvath (Hemiptera: Cimicidae). *J Med Entomol* **41**, 1174–1178.

Ronquist, F. & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.

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

Sambrook, J. & Russell, D. W. (2001). *Molecular Cloning, a Laboratory Manual*, 3rd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

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

Swofford, D. L. (2002). PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods), version 4. Sunderland, MA: Sinauer Associates.

Vandekerckhove, T. M. T., Watteyne, S., Willems, S., Swings, J. G., 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 the *Wolbachia* taxonomy. *FEMS Microbiol Lett* **180**, 279–286.

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

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