Phylogenetic relationships of the Wolbachia of nematodes and arthropods

Citation for published version:
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 Pathogens, vol. 2, no. 10, e94, pp. 887-899. https://doi.org/10.1371/journal.ppat.0020094

Digital Object Identifier (DOI):
10.1371/journal.ppat.0020094

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
PLoS Pathogens

Publisher Rights Statement:
This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Phylogenetic Relationships of the Wolbachia of Nematodes and Arthropods

Katelyn Fenn1*, Claire Conlon1*, Martin Jones1, Michael A. Quail2, Nancy E. Holroyd2, Julian Parkhill2, Mark Blaxter1*

1 Institutes of Evolutionary Biology and Immunology and Infection Research, University of Edinburgh, Edinburgh, United Kingdom, 2 The Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom

Wolbachia are well known as bacterial symbionts of arthropods, where they are reproductive parasites, but have also been described from nematode hosts, where the symbiotic interaction has features of mutualism. The majority of arthropod Wolbachia belong to clades A and B, while nematode Wolbachia mostly belong to clades C and D, but these relationships have been based on analysis of a small number of genes. To investigate the evolution and relationships of Wolbachia symbionts we have sequenced over 70 kb of the genome of wOvo, a Wolbachia from the human-parasitic nematode Onchocerca volvulus, and compared the genes identified to orthologues in other sequenced Wolbachia genomes. In comparisons of conserved local synteny, we find that wBm, from the nematode Brugia malayi, and wMel, from Drosophila melanogaster, are more similar to each other than either is to wOvo. Phylogenetic analysis of the protein-coding and ribosomal RNA genes on the sequenced fragments supports reciprocal monophyly of nematode and arthropod Wolbachia. The nematode Wolbachia did not arise from within the A clade of arthropod Wolbachia, and the root of the Wolbachia clade lies between the nematode and arthropod symbionts. Using the wOvo sequence, we identify a lateral transfer event whereby segments of the Wolbachia genome were inserted into the Onchocerca nuclear genome. This event predated the separation of the human parasite O. volvulus from its cattle-parasitic sister species, O. ochengi. The long association between filarial nematodes and Wolbachia symbionts may permit more frequent genetic exchange between their genomes.

Introduction

Wolbachia are alphaproteobacteria that live intracellularly in a range of animal hosts [1]. Wolbachia belong to the Anaplasmataceae in the Rickettsiales, a diverse group of intracellular symbionts. In other Rickettsiales, the symbiosis is usually parasitic or pathogenic, and many of these bacteria cause significant human and veterinary disease problems. Rickettsiales have also been identified as symbionts of arthropods, and are implicated in causing reproductive manipulations in their hosts similar to those of Wolbachia. (See below; we note that our knowledge of these bacteria is likely to have a severe ascertainment bias, as disease-causing pathogens are obvious and important, whereas innocuous or even beneficial interactors, and free-living species, will be missed. In this context it is informative that unbiased surveys of ecosystems using PCR amplification of conserved genes are turning up rickettsia-like bacteria in many unexpected situations [2].)

In arthropods, where they were first discovered, Wolbachia are the causative agents of a number of fascinating reproductive manipulations [3]. These manipulations serve to promote the survival of infected female arthropods, which pass the Wolbachia vertically to their offspring. A range of phenotypes are caused by Wolbachia infection in arthropods, including killing or feminisation of genetic males, induction of parthenogenetic reproduction in haplo-diploid females, and induction of reproductive incompatibility between individuals that do not have the same infection status. The prevalence of Wolbachia in current arthropod faunas is very high [4,5]; this is due to rare but successful horizontal transfer of the infection between taxa, and is likely to play a role in speciation. Selective sweeps caused by introgression of new Wolbachia strains have strongly shaped mitochondrial population genetics [6], and genomic conflict between the bacterium and the nuclear genome may promote reproductive isolation [7]. There is limited congruence between host and bacterial phylogenies in the arthropod system.

Most arthropod Wolbachia derive from two relatively closely related clades, called A and B [1]. The only formally named Wolbachia is W. pipiensis from the mosquito Culex pipiens, but divergence between the major clades is similar to that observed between species in other bacterial genera [8]. Variant arthropod Wolbachia have been described, from springtails, termites, and spiders, that define additional, more deeply separated clades (E, F, and G) [8,9]. Resolution of the relationships of these additional clades is currently poor. However, Wolbachia “infections” are not limited to the
Arthropoda. Parasitic filarial nematodes of the Onchocercidae, including several major human pathogens, harbour intracellular Wolbachia [10–12]. No other nematodes are currently known to harbour Wolbachia [13], though other nematode–bacterial symbioses are common. In the onchocercids, the Wolbachia can be divided into two major clades, C and D [14], which, unlike the arthropod Wolbachia clades, show phylogenetic congruence with their hosts [15]. Thus, closely related filarial nematodes have closely related Wolbachia, and the association between nematode and bacterium appears to be one of long-term (>100 million years), stable, vertical transmission. The Wolbachia of one filarial species, Mansonia ozzardi, has been placed by analysis of a small number of genes in clade F with termite and weevil isolates.

Analysis of the relationship between the nematodes and their symbionts has revealed that they are likely to be mutualists [16]. Killing the bacteria with tetracycline affects nematode growth, moulting, fecundity, and lifespan [17,18]. In arthropods, in most cases, tetracycline treatment yields cured, healthy hosts, and related parasitic nematodes that do not harbour Wolbachia are unaffected by tetracycline treatment [18]. This feature of nematode–Wolbachia interaction has led to trialling of tetracycline antibiotics for treatment of human filariases, with very positive results [19–22].

In the Rickettsiales and Wolbachia, therefore, where the intracellular habit is ancestral, there has either been a loss of the parasitic or pathogenic phenotype in the nematode Wolbachia or evolution of novel parasitic mechanisms in the arthropod Wolbachia. Previous analyses of Wolbachia phylogeny, and of the relationships of the genus to other Rickettsiales, have been based on very few genes (the Wolbachia surface protein wsp, cell-division protein ftsZ, citrate synthase gInA, groEL chaperone, and small subunit ribosomal RNA [16S] genes) [1,14,15,23]. These analyses were equivocal concerning the deeper structure of the Wolbachia, and could not resolve the placement of the root of the genus; clades E, F, and G are significantly under-sampled. A major limiting factor has been the inferred length of the branches leading to the outgroup taxa. As the genome sequences have generally been chosen for their ability to resolve within-clade, between-isolate relationships, they are not suited to robust resolution of the deeper relationships of Wolbachia. Studies on yeasts and other taxa have shown that extended, multigene datasets can often provide robust resolution when individual constituent genes cannot [24].

Given that clades A and B are very closely related, two possibilities seem most likely. The first is that the nematode symbionts and the arthropod parasites form two distinct radiations (i.e., the tree has the form [outgroup[A,B],[C,D]]); Tree 1 of Figure 1). The second is that one of the nematode symbiont clades (most probably clade C, found in Onchocerca species and close relatives) arises basal to the other clades (i.e., the tree has the form [outgroup[C,D],[A,B]]); Tree 2). A final possibility is that nematode Wolbachia arose from within the arthropod-infecting clades (Tree 4). Trees 1 and 4 have been implicit in many discussions of Wolbachia evolution, possibly because of the historical accident that arthropod Wolbachia were the first to be identified, and are the more widely studied. We have generated genome sequence from a clade C Wolbachia, wOvo from the human parasite Onchocerca volvulus, and here analyse it along with genome sequence from the Wolbachia of Drosophila melanogaster (wMel) (clade A), Wolbachia from Brugia malayi (wBm) (nematode, clade D), and a series of anaplasmatacean outgroups to re-examine this question. We find that the root of Wolbachia is robustly placed between clades A and [C and D], and thus that the mutualist nematode symbionts likely arose from parasitic or pathogenic ancestors. The close coevolution of nematodes and their Wolbachia is underlined by the discovery of a segment of the Wolbachia genome translocated to the O. volvulus nuclear genome.

Figure 1. Hypotheses of Wolbachia Relationships
While we analysed seven taxa, they can be treated as if they were four: outgroups (Anaplasma and Ehrlichia), clade A Wolbachia (wMel, wAna, and wSim), clade D Wolbachia (wBm), and clade C Wolbachia (wOvo). There are thus three possible placements of the root of Wolbachia: (1) Tree 1 [outgroups[A,C,D]], (2) Tree 2 [outgroups[C,D,A]], and (3) Tree 3 [outgroups[D,C,A]]. As clade A included more than one taxon, trees with clade A paraphyletic are also possible. In practice only one such arrangement was found (Tree 4; [outgroups[A’,A”,C,D]]), and may have arisen from analysis of paralogous genes.
DOI: 10.1371/journal.ppat.0020094.g001
Results

Five Segments of the Genome of *Wolbachia* from *O. volvulus*

Twenty-seven primer pairs derived from a range of putative genes from *Wolbachia* from *O. volvulus* (*wOvo*) were tested and yielded 11 probes (Table 1). Five of these identified positive clones in the *O. volvulus* genomic libraries, and the inserts of these clones were amplified by long-range PCR and sequenced (Table 2). The total unique sequence length of the segments is 70,830 bp, representing 6.5% of the estimated 1.1 Mb of the *wOvo* genome [25]. The proportion of the sequenced segments made up of guanine and cytosine bases (GC%) ranged from 31.8% to 35.38% with a mean value of 32.9%. The average GC% of *wBm*, *wMel*, and *R. prowazekii* (67.4%, 81%, and 76%, respectively). This corresponds to a gene density of 0.72 protein-coding genes/kb, which is comparable to *wBm* and *R. prowazekii* (both 0.75 functional genes/kb) but much less than *wMel* (0.94 functional genes/kb). If the genome of *wOvo* is similar in size to those of *wMel* and *wBm*, it is estimated to have approximately 800 functional genes, like *wBm* (which has 806) [26], but many fewer than *wMel* (1,270) [27].

Functional annotation was possible for the majority of the 51 protein-coding genes [26,27] (Table 3). Six are *Wolbachia*-specific, having no orthologue in any of the alphaproteobacterial genomes examined, or elsewhere. These include *Wolbachia* surface protein and five conserved hypothetical proteins. As these genes are present only in *Wolbachia*, they may encode proteins involved in the particular symbiotic biology of the bacteria. One gene, *OW2-L*, is *wOvo*-specific: no function can be ascribed by similarity. A partial pseudogene similar to an ATP-dependent caseinolytic protease ATP-binding subunit, *Clp P*, was identified (Figure 3). An orthologous *Clp P* gene is intact in *wMel* [27], is degraded in *wBm* [26], and is missing from *R. prowazekii* [28]. While it is possible that there is another copy of *Clp P* in the *wOvo* genome, this seems unlikely given the synteny of *wBm* and *wOvo*. *Clp P* acts as a molecular chaperone, and when in complex with the protease *Clp P* (*Clp AP*) it recognises and targets proteins for degradation. *Clp X*, another *Clp* regulator, is distinct from *Clp P*, and also forms complexes with *Clp P* (*Clp XP*) [29]. Both *Clp P* and *Clp X* are present in the genomes of *wBm*, *wMel*, and

Table 1. PCR Primers Used

| Primer Use | Primer Name | Primer Sequence |
|------------|-------------|-----------------|
| Primers used to generate probes | `Ov_wsp_F1` | `TTTTATGGCTGGTGGTAGTGC` |
| | `Ov_wsp_R1` | `TGCTTAGAGGATCTGCAATGTAGC` |
| | `AI111287_F1` | `TGCCTGAAAGCACCATATACC` |
| | `AI111287_R1` | `CTGCTGTGACAGGGATTAGC` |
| | `AI261653_F1` | `CAGGGCCTGCAAACACCG` |
| | `AI261653_R1` | `TCAAGGCTCTCCACACCTACCC` |
| | `AW330455_F1` | `GGCTGAGAGCTCAGAAAGTGGG` |
| | `AW330455_R1` | `CAACACCAGCTGATATTTTG` |
| | `AW351423_F1` | `GGTATGAAAGCACAACCTACCAG` |
| | `AW351423_R1` | `TTTGGTAAAAAGGTCGACTGAGC` |
| Lambda vector primers for long-range PCR and sequencing | `L_FIX_EXP_F1` | `GAGCCTCTAACCTTGACTATAGGGGC` |
| | `L_FIX_EXP_R1` | `CTCCTAAGGGTGACTGAGC` |
| | `Lambda_FIX_T3seq2` | `CACTAAAGGGTGACTGAGC` |
| Primers for PCR and sequencing of wOvo HtrA | `pBACE3.6_T7` | `TAATACGACTTATAGGG` |
| | `wOvo_Htra_F1` | `CGTACCTTCAACAAAATATAG` |
| | `wOvo_Htra_F2` | `GTAACCTTGATTAATTCG` |
| | `wOvo_Htra_R1` | `GCTAAAGGATTATAGGC` |
| Primers used to identify the Wolbachia nuclear insert in Onchoerca sp. | `TATA_F` | `GTTCCGACTTATAGGG` |
| | `TATA_R` | `TTTCCCGAGGTCAAAGGAC` |
| | `TATA_phys` | `TGTCCGACTTATAGGG` |
| | `TATA_OWC` | `CCTACCGTGAAACGACAT` |
| | `Phos` | `TGTGTTGTGCATCTCTTATG` |
| | `OWAC` | `CTCAATCTAAACACCATGCT` |

DOI: 10.1371/journal.ppat.0020094.t001

Table 2. Genome Sequence from wOvo

| Fragment Name | Length (bp) | %GC | Number of Protein-Coding Genes | Other Genes Identified |
|---------------|-------------|-----|-------------------------------|-----------------------|
| OW1           | 12,024      | 32.16 | 11                            | 115 rRNA              |
| OW2           | 11,498      | 31.80 | 10                            | 16S rRNA              |
| OW3           | 15,081      | 35.38 | 8                             | 235 and 5S rRNA      |
| OW4           | 17,997      | 32.07 | 12                            |                       |
| OW5           | 14,230      | 32.91 | 10                            | 1 pseudogene          |

DOI: 10.1371/journal.ppat.0020094.t002
Table 3. Genes Identified in Genomic Sequence from wOvo, and Their Support for Different Phylogenies

| wOvo Co-ordinates Gene Name | Gene Annotation | Comparator Taxa | Neighbour Joining | MrBayes | Maximum Likelihood | Alignment Length |
|-----------------------------|----------------|-----------------|------------------|---------|-------------------|-----------------|
| wOvo Co-ordinates Gene Name | Gene Annotation | Comparator Taxa | Neighbour Joining | MrBayes | Maximum Likelihood | Alignment Length |
|-----------------------------|----------------|----------------|------------------|---------|-------------------|-----------------|
| OW1_A 838-2223              | Amidophosphoribosyltransferase | Y Y Y Y Y Y Y | 71 1 100 | 1 | 0.071 | 475 |
| OW1_B 2658-3242             | NifU-related protein | Y Y Y Y Y Y Y | 97 2 100 | 1 | 0.377 | 443 |
| OW1_C 3197-4580             | Histidyl-tRNA synthetase | Y Y Y Y Y Y Y | 57 1 100 | 1 | 0.172 | 475 |
| OW1_D 4497-4994             | Hypothetical protein | Y - - - - - - | 90 2 100 | 1 | 0.453 | 178 |
| OW1_E 4987-5880             | RNA polymerase sigma-32 factor | Y Y Y Y Y Y Y | 84 2 100 | 2 | 0.225 | 322 |
| OW1_F 6163-6880             | Wolbachia surface protein | Y Y Y Y Y Y Y | - - - - - - | 1 | 0.172 | 443 |
| OW1_G 7185-7517             | Conserved hypothetical protein | Y Y Y Y Y Y Y | 80 3 100 | 1 | 0.144 | 542 |
| OW1_H 7521-9005             | DnaX/DNA polymerase III tau subunit | Y Y Y Y Y Y Y | 100 2 100 | 2 | 0.181 | 1,766 |
| OW1_I 9370-10023            | Glutathione S-transferase | Y Y Y Y Y Y Y | 92 1 100 | 1 | 0.117 | 242 |
| OW1_J 10025-10673           | O-methyltransferase | Y Y Y Y Y Y Y | 93 1 82 | 1 | 0.217 | 221 |
| OW1_K 11229-11924           | Hypothetical protein | Y Y Y Y Y Y Y | 73 4 | 1 | 0.182 | 542 |
| OW2_A 92-451                | 5′ truncated Smr family protein | Y Y Y Y Y Y Y | 97 1 56 | 1 | 0.182 | 245 |
| OW2_B 8494-1457             | Hypothetical protein | Y Y Y Y Y Y Y | 54 1 100 | 1 | 0.217 | 221 |
| OW2_C 1575-3088             | 16S ribosomal RNA gene | Y Y Y Y Y Y Y | 72 2 100 | 1 | 0.181 | 1,766 |
| OW2_D 3231-3608             | Succinate dehydrogenase b-type cytochrome subunit (SdhC) | Y Y Y Y Y Y Y | 59 1 92 | 1 | 0.092 | 136 |
| OW2_E 4056-5303             | Wolbachia-specific hypothetical protein | Y Y Y Y Y Y Y | 82 2 99 | 1 | 0.411 | 126 |
| OW2_F 6430-7659             | Hypothetical protein | Y Y Y Y Y Y Y | 94 1 99 | 1 | 0.399 | 281 |
| OW2_G 7601-8333             | Hypothetical protein | Y Y Y Y Y Y Y | 98 2 100 | 2 | 0.166 | 418 |
| OW2_H 8338-9600             | Hypothetical protein | Y Y Y Y Y Y Y | 56 1 100 | 1 | 0.418 | 414 |
| OW2_I 10465-11946           | Wolbachia-specific hypothetical protein | Y Y Y Y Y Y Y | 66 2 83 | 1 | 0.325 | 3,208 |
| OW2_J 11229-11924           | Hypothetical protein | Y Y Y Y Y Y Y | 68 1 100 | 1 | 0.082 | 415 |
| OW3_A 3381-4454             | Hypothetical protein | Y Y Y Y Y Y Y | 97 1 100 | 1 | 0.158 | 610 |
| OW3_B 4464-6293             | Ribonucleoside-diphosphate reductase alpha chain | Y Y Y Y Y Y Y | 60 1 63 | 1 | 0.375 | 164 |
| OW3_C 6351-6761             | Holliday junction resolvase | Y Y Y Y Y Y Y | 92 2 | 1 | 0.439 | 621 |
| OW3_D 7080-8732             | Excinuclease ABC subunit | Y Y Y Y Y Y Y | 96 1 100 | 1 | 0.076 | 718 |
| OW3_E 8729-10834            | Excinuclease ABC subunit | Y Y Y Y Y Y Y | 97 2 88 | 1 | 0.326 | 283 |
| OW3_F 11735-12481           | Hypothetical protein | Y Y Y Y Y Y Y | 84 3 | 1 | 0.308 | 262 |
| OW3_G 12629-15067           | Hypothetical protein | Y Y Y Y Y Y Y | 100 2 100 | 2 | 0.249 | 862 |
| OW3_H 1359-2072             | Hypothetical protein | Y Y Y Y Y Y Y | 50 2 | 100 | 2 | 0.107 | 277 |
| OW4_A 2072-2851             | Exodeoxyribonuclease III | Y Y Y Y Y Y Y | 93 2 | 98 | 2 | 0.354 | 281 |
| OW4_B 3167-4564             | Wolbachia-specific hypothetical protein | Y - - Y - - - | - - - - - - | 1 | 0.108 | 513 |
| OW4_C 5062-7269             | Wolbachia-specific hypothetical protein | Y Y Y Y Y Y Y | 100 2 100 | 2 | 0.156 | 293 |
| OW4_D 10568-11570c          | Serine protease (HtrA) | Y Y Y Y Y Y Y | 98 2 | 100 | 2 | 0.156 | 293 |
| wOvo Gene Name | Co-Ordinates | Gene Annotation | Comparator Taxa<sup>a</sup> | Neighbour Joining | MrBayes | Maximum Likelihood | Alignment Length<sup>c</sup> |
|---------------|--------------|-----------------|-----------------------------|-------------------|---------|-------------------|-----------------------------|
|               |              |                 | wMel | wAna | wSim | wBm | E. rum. | E. can. | A. mar. | Tree Supported<sup>d</sup> | Bootstrap Support (%) | Tree Supported<sup>d</sup> | Posterior Probability (%) | Preferred Tree<sup>d</sup> | SH Test p-Value |
| OW4_G         | 11573–12607c | Protease subunit (HflK) | Y | Y | Y | Y | Y | Y | Y | 2 | 100 | 2 | 100 | 1 | 0.396 | 371 |
| OW4_H         | 12625–12837c | Wolbachia-specific hypothetical protein | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | 2 | 2 | 100 | 1 | 0.396 | 371 |
| OW4_I         | 13026–14845 | Wolbachia-specific hypothetical protein | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | 2 | 2 | 100 | 1 | 0.396 | 371 |
| OW4_J         | 14649–14866 | Divalent cation tolerance protein | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | 2 | 2 | 100 | 1 | 0.396 | 371 |
| OW4_K         | 15067–15618 | Hypothetical protein | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | 2 | 2 | 100 | 1 | 0.396 | 371 |
| OW4_L         | 16658–17956c | Heat shock protein (GroEL) | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | 2 | 2 | 100 | 1 | 0.396 | 371 |
| OW5_A         | 1204–2712   | NADH-ubiquinone oxidoreductase subunit | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | 2 | 2 | 100 | 1 | 0.396 | 371 |
| OW5_B         | 2745–3650c  | Hypothetical protein | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | 2 | 2 | 100 | 1 | 0.396 | 371 |
| OW5_C         | 3643–4374c  | Geranyltransferrase | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | 2 | 2 | 100 | 1 | 0.396 | 371 |
| OW5_D         | 4526–4798   | Acyl carrier protein | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | 2 | 2 | 100 | 1 | 0.396 | 371 |
| OW5_E         | 4807–6060c  | 3-oxoacyl-synthase II | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | 2 | 2 | 100 | 1 | 0.396 | 371 |
| OW5_F         | 6368–7297   | Conserved hypothetical protein | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | 2 | 2 | 100 | 1 | 0.396 | 371 |
| OW5_G         | 7360–8835c  | Glutamyl-tRNA amidotransferase subunit A | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | 2 | 2 | 100 | 1 | 0.396 | 371 |
| OW5_H         | 9537–10368c | Ribonuclease III | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | 2 | 2 | 100 | 1 | 0.396 | 371 |
| OW5_J         | 10326–11423c| Dehydrogenase | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | 2 | 2 | 100 | 1 | 0.396 | 371 |
| OW5_K         | 12317–13213c| Fructose-bisphosphate aldolase | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | 2 | 2 | 100 | 1 | 0.396 | 371 |
| OW5_L         | 13339–13536c| 5’ truncated ClpA pseudogene | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | 2 | 2 | 100 | 1 | 0.396 | 371 |

<sup>a</sup>“c” indicates a gene on the complementary strand.
<sup>b</sup>A. mar., A. marginale; E. can., E. canis; E. rum., E. ruminantium.
<sup>c</sup>The alignment length is in amino acids for protein-coding genes and in bases for the ribosomal RNA genes.
<sup>d</sup>Trees supported as illustrated in Figure 1. Entries in red have significant support (i.e., >70% bootstrap support under NJ and >90% posterior probability under Bayesian analysis).
<sup>e</sup>No analysis was possible because the gene is specific to Wolbachia.
<sup>f</sup>For OW1-G and OW5-D, R. prowazekii and R. typhi orthologues were used as outgroups.
<sup>g</sup>Not analysed because wBm orthologue is a pseudogene.
U, unresolved polytomy.
DOI: 10.1371/journal.ppat.0020094.t003
It has been reported that ClpAP and ClpXP have distinct substrate specificities in that ClpXP binds only substrate proteins that contain a recognition signal [30]. The benefits for mutualist Wolbachia of not having ClpA are unclear, as ClpA is the more generalist subunit, dealing with proteins damaged by heat shock and starvation.

A second wOvo serine protease subunit, identified as HtrA, was found in fragment OW4 (gene OW4-E). A HtrA from wOvo has been reported previously [31], but OW4-E differs from the published sequence, particularly in the 3′ half of the gene. Resequencing of wOvo HtrA from O. volvulus genomic DNA yielded the same sequence as OW4-E. No fragments or sequences corresponding to the published HtrA were recovered. Alignment of OW4-E and other alphaproteobacterial HtrA genes and the published sequence revealed many single base changes and several indel events that change the frame of the translated protein with respect to other HtrA sequences. The 3′ end of the published “wOvo” HtrA is, however, identical to wBm HtrA, while the 5′ end is nearly identical to OW4-E: it is likely to be an artefactual fusion between wOvo and wBm genes, with some indel sequencing errors also.

**Synteny Comparisons between wOvo, wBm, and wMel**

The arrangement of genes in the five fragments of the wOvo genome was compared to the sequenced genomes of other Wolbachia and Anaplasmataceae. None of the five wOvo fragments was fully syntenic with either fully sequenced Wolbachia (Figure 2). Fragment OW2 differed from wBm only in the presence of a wOvo-specific coding sequence (OW2-I). The other wOvo fragments had two or three rearrangements compared to wBm. Comparison to wMel identified between two and four rearrangements per fragment. Overall, wMel and wBm were more similar to each other in the compared regions, sharing many gene order structures compared to wOvo. Of the five instances where rearrangements compared to wOvo differ between wMel and wBm, wBm is more like wOvo in four (Figure 2). In the fifth (in OW4), gene OW4-L is inverted, but still linked, in wMel, while it is unlinked (but in the same transcriptional orientation) in wBm. None of the gene arrangements specific to wOvo, wBm, or wMel were
found in the other Anaplasmataceae genomes surveyed (unpublished data).

Phylogenetic Analyses of Wolbachia Based on 46 Genes

We identified putative orthologues for the genes identified on the wOvo fragments from the complete and partial genomes of wBm, wMel, Wolbachia from D. ananassae (wAna), Wolbachia from D. simulans (wSim), Ehrlichia canis, E. ruminantium, and Anaplasma marginale. For each gene, we collected all homologues from all sequenced genes from alphaproteobacteria, constructed alignments, and analysed these phylogenetically using the neighbour joining (NJ) algorithm. For the set of target taxa (see Table 3) we selected those homologues that were robustly defined as orthologous to the wOvo genes.

For two proteins (OW1-G and OW5-D) no orthologues were identified in Ehrlichia or Anaplasma, and for these we selected orthologues from R. typhi and R. prowazekii as outgroups. Calculation of the distance from each wOvo protein to that of E. canis, compared to its wMel or wBm orthologue, showed that there was no obvious long branch artefact that might artificially associate two of the three Wolbachia, and that the set of genes analysed embody a wide range of evolutionary rates (Figure 4). The gene set is thus suited to analysis of both local and deep phylogenetic problems [24].

Each alignment of orthologues was then subjected to phylogenetic analysis using NJ, maximum likelihood (ML), and Bayesian ML models. The use of multiple methods of analysis is of utility in the identification of sequences or
clades that behave differently or aberrantly under one method compared to others. The Bayesian ML analytical method is generally recognised to be very effective in dealing with biases in sequence alignments, though it is not foolproof [32]. NJ, as it effectively reduces all signal to a single pairwise difference, is most liable to systematic error. Under NJ, 28 of the 44 protein-coding genes yielded support (bootstrap values > 70%) for a close relationship between wMel and wBm to the exclusion of wOvo (i.e., Tree 2 of Figure 1; Table 4). Two genes supported Tree 1 and one Tree 3; the other genes did not yield phylogenies with >70% bootstrap support for any of the trees. Under Bayesian ML, only 15 of the individual proteins supported Tree 2 with significant posterior probability (>90%), while 11 supported Tree 1. Tree 3 was supported under Bayesian analysis of the same protein, OW1-G, that yielded Tree 3 in the NJ analysis. We note that we had to use Rickettsia outgroups for this gene as no orthologues were identified in Ehrlichia or Anaplasma genomes, and that this may have resulted in a long branch artefact. The ribosomal RNA genes yielded support for Tree 2 in NJ and Bayesian ML analyses, though the support was low. Surprisingly, despite the strong support for distinct trees by both methods for many genes, Shimodaira-Hasegawa (SH) tests found no cases in which there was a significant difference in support for Trees 1 or 2 (Table 3).

Bayesian ML analyses were also carried out on a concatenated alignment of 42 protein-coding genes (excluding those lacking Anaplasma and Ehrlichia outgroups) using two models of protein evolution. The first used a single rate for all the sequences, while the second, more realistic model allowed each protein to evolve with its own rate multiplier. The second model was significantly better (harmonic meanLnL partitioned = −121,745.01; unpartitioned = −122,039.86; Bayes factor ≈ e^{294} ≈ 10^{127}). Using a single rate yielded Tree 1, a result that might be expected considering the relative lengths of the proteins supporting Tree 1 versus Tree 2 (Table 3). A SH test showed highly significant support for Tree 1 (p = 0.003). Analysis using the partitioned model yielded Tree 1 with high posterior probabilities at all nodes (Figure 4). Although Bayesian ML analysis can overestimate support for trees, this result was found in multiple independent analyses.

Identification of a Lateral Gene Transfer Event from Wolbachia to the Nematode Nuclear Genome

Comparison of the sequenced wOvo genomic fragments to available O. volvulus DNA sequences identified a segment of O. volvulus genomic DNA that had significant nucleotide sequence identity to two distinct genes in wOvo (Figure 5). A 5,074-bp EcoRI fragment of O. volvulus genomic DNA had been isolated and sequenced because it contained a TATA box–binding protein gene (GenBank accession L13731) [33]. The TATA box–binding protein gene is located from residues ~2290 to 3500 of the fragment, but a full-length coding sequence was not predicted previously [33]. We resurveyed this sequence, identifying a likely 5′ trans-splice acceptor site at bases 2096 to 2101 and an initiation ATG at 2105 to 2107. The ~2 kb upstream of this trans-splice acceptor site are free of obvious coding features and have no BLASTx matches in public databases (unpublished data). We identified a region of 104 bases (from position 182 to 384 of L13731) that was 63% identical to wOvo OW4-C (Figure 6). There are three insertions (totalling four bases) and one deletion (of one base) in L13731 compared to wOvo OW4-C. Immediately following this section in L13731 is a stretch of 205 bases (385 to 589) that is 84% identical to wOvo OWJ-2 (with two insertions, of one base and 13 bases, and one deletion of one

Table 4. Summary of Support under Different Models of Phylogenetic Inference

| Mode of Inference | Number of Genes Supporting Tree 1 | Number of Genes Supporting Tree 2 | Number of Genes Supporting Tree 3 |
|-------------------|----------------------------------|----------------------------------|----------------------------------|
| NJ<sup>a</sup>    | 8 (1 with support > 90%)         | 32 (17 with support > 90%)       | 3 (0 with support > 90%)         |
| MrBayes<sup>b</sup> | 19 (10 with support > 95%)       | 19 (14 with support > 95%)       | 4 (3 with support > 95%)         |
| ML                | 30                               | 14                               | 0                                |

<sup>a</sup>For NJ analysis, the number of genes that yielded each tree with bootstrap support greater than 90% is given in parentheses.

<sup>b</sup>For Bayesian analysis, the number of genes that yielded each tree with posterior probability greater than 95% is given in parentheses.

DOI: 10.1371/journal.ppat.0020094.g004
of the putative insertion fragment (Figure 6A) revealed five nuclear insertion and confirmed their presence. Sequencing vector species [34]. We surveyed the genomes of *Wolbachia* *O. volvulus*, overlapping that of the putative insertion in the nuclear genome, or from the copy using primers designed to be able to amplify either from the source of amplifying the putative insertion from our independent sequenced from a genomic DNA clone it was possible that it their genomic DNA derived from Mali. As the fragment was 5A).

![Figure 5](PLoS Pathogens | www.plospathogens.org October 2006 | Volume 2 | Issue 10 | e940895)

**Figure 5. Wolbachia** Relationships Inferred from 42 Protein-Coding Genes

(A) Phylogram of consensus tree with inferred distances based on the ultimate model parameters inferred in MrBayes, and rooted by the *Anaplasma* and *Ehrlichia* outgroups.

(B) Cladogram showing Bayesian support for each node. DOI: 10.1371/journal.ppat.0020094.g005

Neither of the wOvo-like segments in L13731 has a complete open reading frame because of the indel differences. Both of these wOvo genes have orthologues in wMel and wBm, but the region of wOvo OW4-C that is similar to L13731 is very divergent from the other *Wolbachia* genes (not shown). Alignment of the wMel and wBm orthologues of wOvo OWJ-2 (a predicted phosphomannomutase) to L13731 shows that the *O. volvulus* nuclear fragment is more similar to wOvo than it is to either of the two *Wolbachia* (Figure 5A).

The nematodes from which Li and Donelson [33] prepared their genomic DNA derived from Mali. As the fragment was sequenced from a genomic DNA clone it was possible that it was a cloning artefact. This possibility was excluded by firstly amplifying the putative insertion from our independent source of *O. volvulus* specimens (from Ghana), and secondly by identifying an orthologous insertion in the genome of the related cattle parasite *O. ochengi*. We carried out PCR assays using primers designed to be able to amplify either from the putative insertion in the nuclear genome, or from the copy resident in the wOvo genome. We were able to amplify, and confirm by sequencing (Figure 6A), the presence of the wOvo-like segments upstream of the *O. volvulus* TATA box-binding protein gene (Figure 6B). *O. volvulus* is one of a group of onchocercid species endemic in Africa. It is known to be close phylogenetically to *O. ochengi*, a cattle parasite that has a range overlapping that of *O. volvulus*, with which it shares some vector species [34]. We surveyed the genomes of *O. ochengi* for *Wolbachia* from *O. ochengi* (wOoc) genes and the putative nuclear insertion and confirmed their presence. Sequencing of the putative insertion fragment (Figure 6A) revealed five single base pair differences from *O. volvulus*. We were unable to confirm that the insertions were close to the *O. ochengi* TATA box-binding protein gene (unpublished data). By surveying the emerging genome sequence data for the filarial parasite *B. malayi*, we were able to identify a TATA box-binding protein gene, the orthologue of the *O. volvulus* gene, but did not find any significant sequence similarity to the wOvo gene fragments in the region upstream of this gene, and, indeed, did not identify any possible nuclear insertions of sequence similar to the five wOvo genome segments isolated in the *B. malayi* whole genome shotgun.

**Discussion**

The Genome of wOvo

The sequenced segments yielded 70 kb of genome sequence for wOvo. Additional rounds of screening failed to yield further wOvo fragments, and construction of *Wolbachia*-enriched genomic libraries was unsuccessful. It would be very informative to complete the wOvo genome and we are continuing to investigate routes to this end.

**Relationships of Wolbachia** Revealed by Sequence Phylogenetics and Synteny

We analysed the sequence of the genes encoded in the five wOvo fragments for phylogenetic signal, as for these we could identify credible orthologues in outgroup taxa. For individual genes, the signal was mixed, but biased towards Tree 2 of Figure 1. However, under ML models, none of the individual genes gave strong support to either of Trees 1 or 2. We identified no particular functional annotation to separate those genes supporting Tree 1 from those supporting Tree 2 (Table 3). As combining genes can yield resolution of phylogenetic problems, by summing the minor signal present in each gene such that it was detectable above the background noise of homoplasy [24], we generated a concatenated alignment of 42 of the wOvo proteins and their orthologues. Analysis of this concatenated alignment using unpartitioned or partitioned (more realistic, given the variation in inferred rates between genes; Figure 4) models yielded robust support for Tree 1, i.e., [outgroups[\[wAna,wSim\],\[wBm,wBm\]]] (Figure 5). Notably, the shortest inferred internal branch in the phylogeny was that linking the last common ancestor of all *Wolbachia* and the last common ancestor of the nematode (clade C and D) *Wolbachia*. The length of this branch compared to neighbouring ones in the phylogeny may explain the difficulty in robustly recovering a distinct phylogeny with more limited datasets. As genes from clade B *Wolbachia* are consistently very closely related to those from clade A rather than from clades C or D [1,35], we predict that inclusion of clade B in the analysis would yield a tree [outgroups[\[C,D\],[A]]] (Figure 5).

Conserved gene arrangements (synteny) can be used to infer phylogenetic relationships between genomes. The wOvo fragments share some local synteny with both wMel and wBm. When breakage of local synteny occurs, two features are apparent. Firstly, wBm and wMel are more similar to each other than either is to wOvo. Secondly, wBm is closer to wOvo than is wMel, as wMel has several unique rearrangements. Comparison to the outgroup genomes was uninformative because of the high levels of rearrangement that have taken place in *Wolbachia* genomes since they last shared a common
ancestor with Anaplasmataceae [26,27]. Mapping of these changes in synteny onto the phylogeny derived from the sequence data suggests that the wOvo genome has undergone many more rearrangements since the last common ancestor of the three Wolbachia we have analysed than have either wBm or wMel.

We fully recognise that we have not been able to analyse with the larger dataset the more enigmatic and rarely described clades of Wolbachia, clades E, F, and G [8,9]. Current data suggest that clades E, F, and G arise basal to [A,B], but have not clearly resolved the pattern of branching compared to C and D [8,23]. We note that the standard three genes used for within-Wolbachia phylogenetics, wsp, ftsZ, and 16S ribosomal RNA, may not be the best set for analysing deeper relationships in the genus. Thus, wsp is essentially restricted to Wolbachia, while ftsZ has a high rate of evolutionary change, and is possibly subject to long branch artefacts. The ribosomal RNA genes yield Tree 1, though with relatively low NJ bootstrap support (66% for 23S and 5S, and 72% for 16S; Table 3). The addition of groEL and gltA genes to the analysis was unable to place the root with certainty [23]. Our sample of genes with a wide range of evolutionary rates has yielded strong support for one of the competing models. It will be very informative to utilise an expanded set of genes such as those sampled here to address the question of the relationships of the E, F, and G clades to the better known A, B, C, and D organisms.

The Evolution of Symbiotic Phenotypes in Wolbachia

As a whole, the Rickettsiales have lifestyles that involve intracellular replication in a eukaryotic host cell, and the outgroups analysed here have parasitic or pathogenic lifestyles. The support for Tree 1 suggests that the ancestor of all extant Wolbachia was probably an intracellular pathogen or parasite. Our analyses suggest that this intracellular pathogen was then tamed by, or evolved beneficial symbiotic relationships with, its nematode hosts, but evolved towards specific reproductive parasitism in the arthropod-infecting clade A (and B) strains. A single transfer of an ancestral Wolbachia to
an onchocercid nematode host is most likely. The nematode *Wolbachia* have apparently coevolved with their hosts through strictly vertical descent, while the arthropod strains have undergone frequent (on an evolutionary timescale) horizontal transfers or host captures, while also maintaining themselves on a life-cycle timescale by vertical transmission. As arthropod *Wolbachia* are parasites, it is possible for individuals and populations to lose their infections. Importantly, it is also evident that nematodes can lose their *Wolbachia*, as *Wolbachia*-negative nematode species are nested within clades of infected taxa [16]. There is a correlation between the presence of *WO* phage in *Wolbachia* genomes [36] and the parasitic phenotype, and thus *WO* phage and/or genes transduced by *WO* phage may underpin parasitic manipulations [37]. There were no *WO* phage–like elements in the *wOvo* genome segments analysed.

### Lateral Transfer of *Wolbachia* Genetic Material to the *O. volvulus* Nuclear Genome

Serendipitously, we identified two short fragments of *Wolbachia* genes in one of the few segments of the *O. volvulus* genome to have been sequenced. Transfer of *Wolbachia* genetic material into the host nuclear DNA has been noted previously, in the adzuki bean beetle, *Callosobruchus chinesis*, where a reasonably large segment of *Wolbachia* DNA has been inserted into the X chromosome [38]. The adzuki bean beetle insertion is not thought to be expressed.

The sequenced *O. volvulus* segment incorporates the gene for a TATA box–binding protein and a region 2 kb upstream. In this upstream region we detected two short segments that have significant pairwise identity to *wOvo* OW2-1 and to *wOvo* OW4-C. We confirmed that the putative insertion was present in *O. volvulus* genomic DNA (and was not therefore a cloning artefact) by isolating it by specific PCR from an independent source of *O. volvulus*. Neither fragment is a complete gene, and both have been subject to mutational accumulation such that the open reading frames are no longer intact. The two genes do not lie beside each other in the bacterial chromosome. The insertional fragment is not unique to *O. volvulus*, as it is also present in the cattle onchocercid, *O. ochengi*. *O. ochengi* is very closely related to *O. volvulus*, and indeed *O. volvulus* in humans is thought to represent a recent host capture by, and vicariant speciation of, onchocercids of ungulates. No homologous insertion was detected in the partial genome sequence of *B. malayi*, but the orthologous TATA box–binding protein gene was identified. Examination of the region between the *B. malayi* TATA box–binding protein gene and the next gene upstream identified no sequences with significant similarity to the putative *Wolbachia* insertions (unpublished data). We also used PCR to screen for the insertion in the deer onchocercid *O. flexuosa*. *O. flexuosa* is interesting because it appears to lack *Wolbachia* entirely (as determined by PCR screens and electron microscopy) [39]. Identification of an insertional relic of *Wolbachia* would bolster suggestions that this species has lost its symbiont. However, we were unable to amplify any insertion fragments from *O. flexuosa* (unpublished data), leaving the question of symbiont loss unanswered.

Nuclear integration of fragments of other cytoplasmic genomes, such as the mitochondrial and chloroplast genomes, is relatively common, but no plausible integrants of *wBm* were detected in the near-complete *B. malayi* genome [26]. Whether acquisition of *Wolbachia* genes by the host plays any part in host evolution remains conjectural. Similarly, the *Wolbachia* could capture host genes, but none of the sequenced genomes contain genes with signatures of animal, rather than alphaproteobacterial, origin.

### Materials and Methods

**Selection of *wOvo* probes and identification of *wOvo* genomic gene sequences.** Probes were prepared from the plasmid sequenced *wOvo* genes, including the 16S ribosomal RNA gene, *wOvo* hsp60, and others identified in the *O. volvulus* EST (expressed sequence tag) programme [40,41] (Table 1). Probes were labelled with alpha32P dCTP by oligo-primed synthesis. *O. volvulus* libraries in lambda phage, gifts of John Donelson [33] and Steve Williams, were plated on bacterial lawns, and the lifts were prepared for Southern hybridisation using standard methods. Initial hybridisations used a mix of probes from several genes. After autoradiography, positive plaques were identified by gene-specific PCR, and purified by dilution and repropbing. Inserts were isolated by long-range PCR using lambda vector primers, and end sequenced. End-probes were generated and used to reprobe plaque lifts. Primer sequences are given in Table 1.

**Sequencing and annotation.** Long-range PCR products were sequenced by standard shotgun methods at the Wellcome Trust Sanger Institute, and assembled using standard methods. The insert sequences were completed with a combination of directed sequencing of selected plasmid subclones, and primer walking. One clone insert proved to be a chimera of human and *Wolbachia* DNA; the human segment was identified by its sequence identity to human genomic sequence, and was removed from the analysis. Genes were identified and annotated in the *wOvo* genome segments using Artemis [42]. The Artemis comparative tool [43] and phobos [44] were used to display and investigate synteny relationships with the *wBm* [26] and *wMel* [27] genomes. A putative *wOvo* HtrA serine protease (GenBank accession AAP79877) similar to OW4-E had been published previously [31]. To test if *wOvo* has more than one HtrA gene or if the difference was due to technical error, primers (see Table 1) were designed within the OW4-E 5′ and 3′ extragenic regions. Multiple PCR and sequencing reactions were performed according to standard procedures using *O. volvulus* genomic DNA. The sequences were aligned and a consensus sequence was obtained. To assess the possible function of the *wOvo-specific* genes, *wAna*, SignapP v2.0 [45] and phobos [44] were used to identify a possible signal peptide and a probable cellular location.

**Phylogenetic analysis.** For phylogenetic analysis, particularly since we wanted to identify the root of the *Wolbachia* clade, it was essential to analyse alignments of orthologous sequences, and to exclude paralogues. Each protein-coding gene in *wOvo* was used to search *wAna* and *wSim* EST [45] a custom database of alphaproteobacterial proteins extracted from EMBL and GenBank to identify homologues. In addition, homologues were identified from the complete and partial genomes of *wBm*, *wMel*, *wAna*, *wSim* [46,47], *A. marginale* [48], *E. ruminantium* [49], and *E. canis*. For each *wOvo* protein, a multiple alignment was constructed using ClustalW [50] and subjected to NJ analysis in PHYLLIP (using character difference) [51]. From the resulting phylogenies we identified orthologous genes from the seven complete and partial genomes. Importantly, we excluded paralogues from genomes where an orthologue was absent. These paralogues were not the best scoring match in the selected genome, but by phylogenetic analysis we were clearly not orthologous to the *wOvo* query. The *wAna* and *wSim* genomes were assembled from whole genome shotgun reads “contaminating” those generated for the nuclear genome projects of their host species, and are incomplete. Profiling *wAna* identified several genes that are present in other bacterial genomes but are duplicated (or partially duplicated) in the *wAna* assembly. We interpret these to be due to either misassemblies or the presence of two closely related *Wolbachia* genomes in *D. ananassae*. If one whole genome sequence shotgun survey includes DNA from two distinct *Wolbachia*, the genes we selected for subsequent analysis may be selected stochastically from two distinct genomes, but the close relationship implied by comparison of the “duplicated” segments in the assembly (≥99% identity) means that they can effectively be considered a single taxon.
For the 44 proteins with matches, and the 106 and 238/55 ribosomal RNA genes, we realigned each evo sequence with its orthologues. The alignments are available as Protocol S1 online. The protein gene alignments were combined with DNA alignments using Tree-Puzzle 5.1 (http://www.tree-puzzle.de) using accurate (slow) parameter estimation. Since, for many genes, one of the trees was the one selected by ML, this test is more appropriate than the Kishino-Hasegawa likelihood ratio test, which assumes that trees be specified a priori. For protein-coding genes, amino acid alignments were analysed under the JTT model with gamma rate variation (four categories) and a proportion of invariant sites. For RNA genes, DNA alignments were analysed under the HKY model with gamma rate variation (four categories) and a proportion of invariant sites. For each gene, two independent runs were executed for 1,000,000 generations, and sampled every 1,000 generations, with default prior and Markov chain parameters. After visual confirmation of stationarity, the first 10% of saved trees were discarded as burn in. The significance of the difference in support for the two credible alternative hypotheses was tested for each gene using a likelihood ratio test, p-Values were calculated using the SH test as implemented in Tree-Puzzle 5.1 (http://www.tree-puzzle.de) using accurate (slow) parameter estimation. Since, for many genes, one of the trees was the one selected by ML, this test is more applicable than the Kishino-Hasegawa likelihood ratio test, which assumes that trees be specified a priori. For protein-coding genes, amino acid alignments were analysed under the JTT model with gamma rate variation (four categories) and a proportion of invariant sites. Rokas et al. [24] have shown that the use of large datasets, employing many genes with varying rates, is effective in recovering “correct” phylogenies when single-gene analyses fail to do so. Bayesian analyses of the concatenated alignment of 42 protein-coding genes was carried out under two models. In the first model, all genes shared a fixed rate JTT model of protein evolution with gamma rate variation approximated by four rate categories and a proportion of invariant sites. For RNA genes, DNA alignments were analysed under the HKY model with gamma rate variation (four categories) and a proportion of invariant sites.

ACKNOWLEDGMENTS

We thank Achim Hoerauf for O. volvulus material, Sandy Trees for O. ochengi material, and Norbert Brattig for O. fexuxa material. O. volvulus genomic libraries were provided by John Donelson and Steve Willard (Columbia University). Graham Stone provided valuable comments on the manuscript. We acknowledge the assistance of the Wellcome Trust Sanger Institute core sequencing and informatics groups.

Author contributions. KK, CC, MJ, MAQ, JP, and MB conceived and designed the experiments. KK, CC, MJ, MAQ, NEH, and JP performed the experiments. KF, CC, MJ, JP, and MB analyzed the data and wrote the paper.

Funding. This project was funded by the Wellcome Trust. Edinburgh computing facilities were funded by the National Environment Research Council.

Competing interests. The authors have declared that no competing interests exist.

Supporting Information

Protocol S1. Multiple Sequence Alignments of evoOvo Proteins and tRNAs Used in the Analysis of Wolbachia Relationships.

The data are in NEXUS format. Found at: DOI: 10.1371/journal.ppat.0020094.sd001 (414 KB TXT).

References

1. Werren JH, Zhang W, Guo LR (1995) Evolution and phylogeny of Wolbachia: Reproductive parasites of arthropods. Proc Biol Sci 261: 55–63.
2. Yammin C, Petroni G, Verni F, Rosati G (2005) A bacterium belonging to the Rickettsiaceae family inhabits the cytoplasm of the marine ciliate Diophrys appendiculata (Ciliophora, Hypotrichia). Microb Ecol 49: 445–452.
3. Werren JH (1997) Biology of Wolbachia. Annu Rev Entomol 42: 587–609.
4. Werren JH, Guo L, Windsor DW (1995) Distribution of Wolbachia among neotropical arthropods. Proc Biol Sci 262: 197–200.
5. Jiggins FM, Bentley JK, Majerus ME, Hurst GD (2001) How many species are infected with Wolbachia? Cryptic sex ratio distorters revealed to be common by intensive sampling. Proc Biol Sci 268: 1125–1126.
6. Hurst GD, Jiggins FM (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: The effects of inherited symbionts. Proc Biol Sci 272: 1525–1534.
7. Gottoh T, Noda H, Fujita T, Iwade K, Higo Y, et al. (2005) Wolbachia and nuclear-nuclear interactions contribute to reproductive incompatibility in the spider mite Panonychus ulmi (Acari: Tetanychidae). Hereditas 148: 237–246.
8. Lo N, Casiraghi M, Salati E, Bazzocchi C, Bandi C (2002) How many Wolbachia superfamilies exist? Mol Biol Evol 19: 541–546.
9. Rowe MJ, Raven RJ, McGraw EA (2004) Wolbachia pipientis in Australian spiders. Curr Microbiol 49: 208–214.
10. McLaren DJ, Worms MJ, Laurence BR, Simpson MG (1975) Microorganisms in filarial larvae (Nematoda). Trans R Soc Trop Med Hyg 69: 591–594.
11. Freidzon M, Battermann H (1985) The fine structure of Ovary in filarial worms. IV. The heads of the cord-like structures of the testes. Zentbl Entwicklungsme 34: 122–128.
12. Sironi M, Bandi C, Sacchi L, Di Sacco B, Damiani G, et al. (1995) Molecular evidence for a close relative of the arthropod endosymbiont Wolbachia in a filarial worm. Mol Biochem Parasitol 74: 881–893.
randomized, double-blind clinical trial of a 3-week course of doxycycline plus albendazole and ivermectin for the treatment of *Wuchereria bancrofti* infection. Clin Infect Dis 42: 1081–1089.

25. Casiraghi M, Bordenstein SR, Baldo L, Lo N, Beninati T, et al. (2005) Phylogeny of *Wolbachia pipientis* based on gtfA, 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 (Pt 12): 4015–4022.

26. Foster J, Granata M, Kamal I, Ware J, Makarova K, et al. (2005) The *Wolbachia* genome of *Brugia malayi*: Endosymbiont evolution within a human pathogenic nematode. PLoS Biol 3: e121. DOI: 10.1371/journal.pbio.0030121.

27. Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, et al. (2004) Phylogenomics of the reproductive parasite *Wolbachia pipientis* aMe: A streamlined genome overun by mobile genetic elements. PLoS Biol 2: e90. DOI: 10.1371/journal.pbio.0020069.

28. Andersson SG, Zomorodipour A, Andersson JO, Sichler-Jo, Alsmark UC, et al. (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature 396: 133–140.

29. Chandu D, Nandi D (2004) Comparative genomics and functional roles of the ATP-dependent proteases Lon and Clp during cytosolic protein degradation. Res Microbiol 155: 710–719.

30. Sharma S, Hoskins JR, Wickner S (2004) Binding and degradation of heterodimeric substrates by ClpAP and ClpXP. J Biol Chem 280: 5449–5455.

31. Jolodar A, Fischer P, Buttner DW, Brattig NW (2004) Neutrophil accumulation around *Onchocerca volvulus* microfilariae: dependence on *Wolbachia* endosymbiont transferred to X chromosome of host insect. Proc Natl Acad Sci U S A 99: 14280–14285.

32. Rutherford K, Parkhill J, Crowe J, Hornsell T, Rice P, et al. (2000) Artemis: Sequence visualization and annotation. Bioinformatics 16: 944–945.

33. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340: 785–795.

34. Nakai K, Horton P (1999) PSORT: A program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem Sci 24: 34–36.

35. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402.

36. Salzberg SL, Dunning Hotopp JC, Delcher AL, Pop M, Smith DR, et al. (2005) MerenSAT: Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species. Genome Biol 6: 492.

37. Salzberg SL, Dunning Hotopp JC, Delcher AL, Pop M, Smith DR, et al. (2005) Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species. Genome Biol 6: R23.

38. Claydon KL, Kappmeyer LS, Herndon DR, Dark MJ, Tibbals DL, et al. (2005) Complete genome sequencing of *Anaplasm marginale* reveals that the surface is skewed to two superfamilies of outer membrane proteins. Proc Natl Acad Sci U S A 102: 844–849.

39. Collins NE, Liebenberg J, de Villiers EP, Bratton KA, Loun E, et al. (2005) The genome of the heartwater agent *Ehrlichia ruminantium* contains multiple tandem repeats of actively variable copy number. Proc Natl Acad Sci U S A 102: 838–843.

40. Thompson JD, Higgins DG (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680.

41. Feenstra J (2004) PHYLIP: Phylogeny Inference Package, version 3.6 [computer program]. Seattle: Department of Genome Sciences and Department of Biology, University of Washington.

42. Swofford DL (2001) PAUP*: Phylogenetic Analysis using Parsimony * and other methods [computer program]. Sunderland (Massachusetts): Sinauer Associates.

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