THE ROYAL SOCIETY

doi 10.1098/rspb.2001.1656

Recombination confounds interpretations of *Wolbachia* evolution

Francis M. Jiggins[1*], J. Hinrich G. von der Schulenburg[2], Gregory D. D. Hurst[3] and Michael E. N. Majerus[1]

[1]*Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK*
[2]*Abteilung für Evolutionsbiologie, Institut für Spezielle Zoologie, Westfälische Wilhelms-Universität Münster, Hüfferstrasse 1, 48149 Münster, Germany*
[3]*Department of Biology, University College London, The Galton Laboratory, 4 Stephenson Way, London NW1 2HE, UK*

Wolbachia are vertically transmitted bacteria known from arthropods and nematode worms, which are maintained in host populations because they either physiologically benefit infected individuals or parasitically manipulate their reproduction. The different manipulation phenotypes are scattered across the *Wolbachia* phylogeny, suggesting that there have been multiple evolutions of similar phenotypes. This conclusion relies on the assumption of an absence of recombination between bacterial strains, so that the gene used to reconstruct the phylogeny reflects the evolutionary history of the genes involved in the trait. We tested for recombination by reconstructing the phylogeny of two *Wolbachia* genes from seven B-subdivision strains. The two genes produced mutually incompatible topologies, indicating that these lineages are subject to genetic recombination. This means that many evolutionary patterns inferred from *Wolbachia* phylogenies must be re-evaluated. Furthermore, recombination may be an important feature both in the evolution of the manipulation phenotypes and avoidance of Müller's ratchet. Finally, we discuss the implications of recombination for attempts to genetically engineer *Wolbachia* for use in the control of crop pests and human pathogens.

Keywords: *Wolbachia*; recombination; phylogeny; *wsp*; *ftsZ*; cytoplasmic incompatibility

1. INTRODUCTION

Wolbachia are a genus of inherited bacterial symbionts known from most of the major arthropod groups and filarial nematodes. They have attracted considerable attention due to their intriguing ability to manipulate arthropod reproduction, and consequently drive evolutionary change in their host species. The four classes of reproductive manipulation that have been recorded are parthenogenesis induction (hymenopteran hosts), male-killing, the feminization of genetic males (lepidopteran and isopod hosts) and cytoplasmic incompatibility (Hurst *et al*. 1999; O'Neill *et al*. 1992; Rousset *et al*. 1992; Stouthamer *et al*. 1993). These have all evolved because *Wolbachia* bacteria are maternally inherited; the first three increase the number or fitness of infected females at the expense of males, while cytoplasmic incompatibility decreases the fitness of uninfected females in the population. *Wolbachia* have also been described in nematode worms that show mutualistic phenotypes, and they therefore cover the entire spectrum of symbiotic interactions (Bandi *et al*. 1999).

Because of this diversity of phenotypes, reconstruction of the *Wolbachia* phylogeny can provide insights into the evolution of symbiotic associations. Much research has centred on the phylogenetic analysis of three *Wolbachia* genes. Initially, phylogenies inferred from a 16S rDNA sequence demonstrated that bacteria causing the different reproductive phenotypes form a monophyletic group: the genus *Wolbachia* (O'Neill *et al*. 1992; Rousset *et al*. 1992; Stouthamer *et al*. 1993). However, this gene provided little phylogenetic resolution within this group, and subsequently the faster-evolving cell-cycle gene *ftsZ* and then the surface protein *wsp* gene were used for phylogenetic analysis (Werren *et al*. 1995; Zhou *et al*. 1998). The *wsp* gene is the most variable and informative of the three.

One of the most striking patterns to emerge from analysis of the *ftsZ* and *wsp* genes is that the different reproductive phenotypes are scattered throughout the bacterial phylogeny (Von der Schulenburg *et al*. 2000; Zhou *et al*. 1998). This suggests either that the effect is host specific, or that there may have been multiple evolutions of at least three out of the four phenotypes within *Wolbachia*. However, this conclusion is based on the observation that the same phenotypes appear on disparate branches of the *Wolbachia* phylogenetic tree, and it implicitly assumes that the evolutionary history of the gene which is used to reconstruct the phylogeny reflects that of the genes involved in the reproductive manipulation. That is to say, the approach assumes that there is no recombination between *Wolbachia* lineages.

We have tested this assumption by sequencing both the *ftsZ* and *wsp* genes from seven *Wolbachia* strains, adding them to other published sequences, and investigating whether the phylogenies of the two genes are congruent. When the genes' phylogenies were reconstructed they proved to be mutually incompatible, and our results suggest that this has resulted from recombination between *Wolbachia* strains.

2. MATERIAL AND METHODS

(a) *Strains used*

Sequences were isolated from seven strains of *Wolbachia*. Two were closely related male-killers from the ladybird beetle *Adalia*

*Author for correspondence (fmj1001@mole.bio.cam.ac.uk).

© 2001 The Royal Society

bipunctata (strains Y and Z), which occur within the same host population but do not 'double' infect the same host individuals (Hurst *et al*. 1999). Two further strains were also male-killers found within a single population of the butterfly *Acraea encedon* (strains U and T) (Jiggins *et al*. 2000*b*). Two more *Wolbachia* were isolated from the butterflies *Acraea equitoria* and *Acraea eponina*. The seventh infection is associated with cytoplasmic incompatibility in the mosquito *Aedes albopictus*.

(b) *Sequencing the wsp and ftsZ genes*

Some of the *ftsZ* and *wsp* sequences from these *Wolbachia* strains have already been described, and the remainder were sequenced for this study. DNA was extracted from the insects by phenol-chloroform extraction (Sambrook *et al*. 1989). PCR was performed on these extractions using *Wolbachia*-specific *wsp* primers (81F, 691R; Zhou *et al*. 1998) and polymerase with proofreading activity (Expand High Fidelity PCR system; Boehringer–Mannheim Ltd, Indianapolis, IN, USA). The *ftsZ* gene was then amplified from the same templates with B-subdivision *Wolbachia*-specific primers (*ftsZ*Bf, *ftsZ*Br; Werren *et al*. 1995). Thermal cycling conditions have been described previously (Hurst *et al*. 1999). Prior to the addition of template DNA, all PCR reactions were irradiated with 150 mJ of ultraviolet light in a UV Stratalinker 2400 (Stratagene Ltd, La Jolla, CA, USA), in order to cross-link any contaminant DNA.

The bacterial genes were sequenced from two independent lines of all the hosts, except for *Aedes albopictus*, where a single line was used. The PCR product was purified using Microcon-50 Microconcentrators (Amicon Ltd, Beverly, MA, USA). In the case of the *wsp* gene, both strands were then sequenced in totality directly from the PCR product, using the PCR primers. The purified *ftsZ* PCR product was cloned before sequencing. The PCR product was ligated into a T-tailed vector (pGEM-T vector system, Promega Ltd, Madison, WI, USA) and transformed into *Escherichia coli* DH5α, which was grown for only 45 min prior to plating in order to minimize duplication of clones. White colonies, which contain an insert, were picked and grown, and the plasmids subsequently purified using the Wizard Minipreps Purification System (Promega Ltd). The *fts*Z inserts were sequenced using the M13 vector primers and internal sequencing primers. Sequencing was performed via cycle sequencing using the ABI PRISM Big Dye Terminator Cycle Sequencing Kit (Perkin Elmer, Inc., Boston, MA, USA), and the results visualized on an ABI 377 DNA sequencer (Perkin Elmer, Inc.).

(c) *Testing for double B-group* Wolbachia *infections*

Analysis based on a comparison of *ftsZ* and *wsp* sequences that were isolated from the same host could be confounded if the insects were infected with more than one B-subdivision *Wolbachia*. We checked for the presence of double infection by (i) looking for double peaks following direct sequencing of the *wsp* gene, (ii) checking that all of the *ftsZ* clones had the same sequence, and (iii) confirming that there were no restriction fragment length polymorphisms (RFLPs) in the PCR products.

When checking for RFLPs, we digested the *wsp* PCR product from all the strains with the enzymes *Rca*I and *Hind*III, which each cut about half of the B-group *wsp* sequences. In addition, some populations of *Aedes albopictus* are double infected, and only one of the bacterial strains has a cutting site for *Eco*RV in the *ftsZ* gene (O'Neill *et al*. 1997). This probably represents the double infection with an A- and B-group *Wolbachia* that has been reported elsewhere (Werren *et al*. 1995). In order to check that the insect we used was not double infected with two B-group *Wolbachia*, we amplified the *ftsZ* gene from *Aedes albopictus* (as above) and digested it with *Eco*RV, alongside controls which either contained the cutting site (*Acraea encedon* strain U) or lacked the cutting site (*Acraea encedon* strain T).

(d) *Phylogenetic analysis*

The newly isolated sequences were added to the alignment of Von der Schulenburg *et al*. (2000), excluding some highly similar sequences from *Trichogramma* species so as to enhance the speed of tree reconstruction. The phylogenetic analysis was performed with the program PAUP* using the maximum-likelihood criterion (Swofford 1998). We used the GTRG (general time reversible with gamma-distributed rate heterogeneity across sites) (*wsp*) and GTRGI (GTRG with consideration of the proportion of invariable sites) (*ftsZ*) models of sequence evolution that, in a previous analysis, most accurately reflected the evolution of the two genes. The rate matrix for six substitution types, the shape of the gamma distribution for four rate categories and the proportion of invariable sites (*ftsZ* only) were all estimated from an unweighted maximum-parsimony tree (Von der Schulenburg *et al*. 2000). Taking account of these parameter estimates, the tree was then reconstructed using a heuristic search based on nearest-neighbour interchanges. The robustness of the inferred topology was assessed by performing 100 non-parametric bootstrap replicates, using the same settings as above. These results were confirmed by also generating a parsimony-based bootstrapped phylogeny from the same data.

We then tested directly whether the phylogenetic signals of the two genes are mutually incompatible using the parsimony-based partition homogeneity test (Farris *et al*. 1994). This test compares the summed lengths of both genes' maximum-parsimony trees with the lengths of trees generated from random partitions of the combined sequences of both genes. Ten thousand replicates of this process were performed on the seven pairs of sequences that we obtained, together with an A-group outgroup.

The partition homogeneity test is liable to be confounded if the genes have different evolutionary dynamics, therefore we also tested whether the two genes had conflicting phylogenetic signals using a likelihood ratio test (Huelsenbeck & Bull 1996). This test compares the maximum likelihood, L_0, under the null hypothesis that the two genes share the same topology, with the maximum likelihood, L_1, under the alternative hypothesis that the genes have different topologies. L_0 was calculated using the GTRG model in program PAML v. 3.0*b* (Yang 1997) by forcing the two genes to share the same topology, but allowing them to differ in all other respects (branch lengths, six substitution rates, base frequencies and the gamma rate distribution). L_1 was then calculated by summing the maximum likelihoods of topologies estimated from the two genes in isolation. From this, we calculated the test statistic

$$\delta = 2(\ln L_1 - \ln L_0). \quad (1)$$

The significance of δ was determined by parametric bootstrapping. One hundred replicate data sets of each gene were simulated using the program Seq-Gen v. 1.2 (Rambaut & Grassly 1997) under the null hypothesis (i.e. the GTRG model parameters were estimated for each gene assuming that they share the same topology but may differ in all other respects). The test statistic δ was then recalculated for each replicated pair of sequences and compared with the value of δ obtained from the original data.

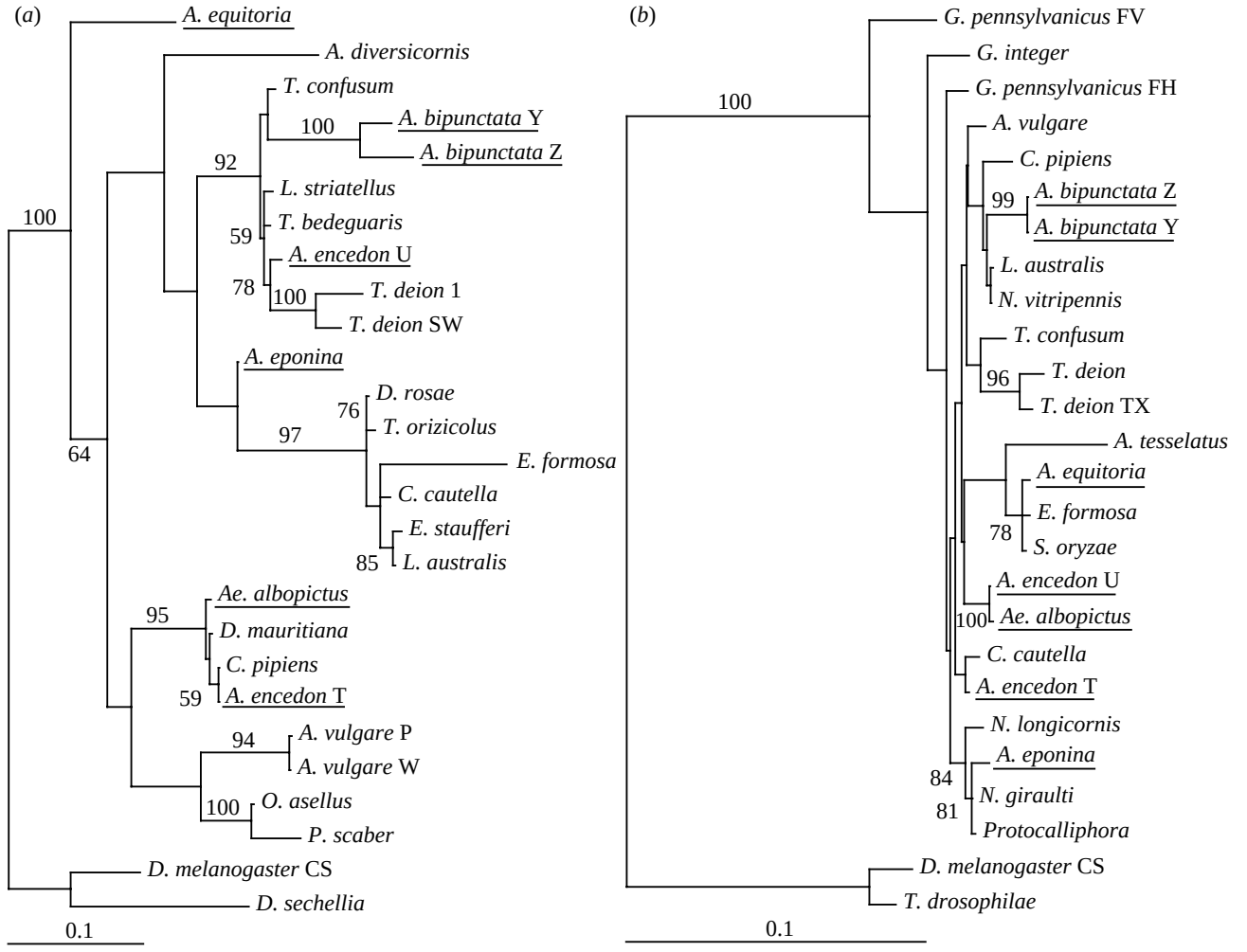

Figure 1. The maximum-likelihood phylogenies of the *Wolbachia* genes (*a*) *wsp* and (*b*) *fts*Z. Where both genes have been isolated from the same singly infected host specimen, the name of the host is underlined. The branches show the results of 100 bootstrap replicates.

3. RESULTS

(a) *The sequences and absence of double infections*

Unique sequences were obtained from each of the seven *Wolbachia* strains (accession numbers AJ130714–7, AJ271198–201, AJ271194–5, AF020059, AJ130892 and U28206). The *wsp* sequences did not contain double peaks, and none of the *fts*Z clones (four or more clones per strain) differed by more than one nucleotide in sequence, which is consistent with the error rates given by the manufacturer for the polymerases. The two sequences from *Aedes albopictus* were found to be identical to those previously published, and the two *Adalia bipunctata* male-killers had identical *fts*Z sequences.

Aside from the sequence evidence, we did not find any evidence for RFLPs. When the *wsp* PCR products were digested with *Rca*I or *Hind*III, there were no RFLPs. The same was true when *fts*Z was cut with *Eco*RV.

(b) *Incompatible phylogenies and recombination*

The maximum-likelihood phylogeny of these bacterial strains and other published sequences are shown in figure 1 (*fts*Z parameter estimates: nucleotide frequencies, A = 0.34, C = 0.14, G = 0.24, T = 0.27; gamma shape parameter, $\alpha = 0.12$; substitution rates A–C = 1.55, A–G = 7.51, A–T = 0.38, C–G = 0.83, C–T = 14.07, G–T = 1; *wsp* parameter estimates: nucleotide frequencies, A = 0.29, C = 0.14, G = 0.22, T = 0.33; gamma shape parameter, $\alpha = 0.45$; substitution rates, A–C = 4.96, A–G = 10.79, A–T = 1.17, C–G = 2.27, C–T = 11.74, G–T = 1). The maximum-parsimony tree had both a similar topology and similar levels of bootstrap support for the different clades. There is a striking conflict between the two phylogenies in the relationship of *Acraea encedon* strain U, *Acraea encedon* strain T and *Aedes albopictus*. The *fts*Z gene provides strong bootstrap support for *Aedes albopictus* and *A. encedon* strain U being monophyletic and relatively distant from *Acraea encedon* strain T. In contrast, the *wsp* gene places *Aedes albopictus* near to *Acraea encedon* strain T, and distant from *Acraea encedon* strain U, again with strong bootstrap support. The hypothesis that the two genes have different evolutionary histories is supported by both the partition homogeneity test ($p < 0.0001$) and the likelihood ratio test, where the value of δ calculated from the replicate data sets was in all cases less than that calculated from the original data ($\delta = 200$; $p < 0.01$).

4. DISCUSSION

Wolbachia have been the focus of an explosion of research interest in recent years, with many researchers using molecular phylogenies to investigate evolutionary patterns within this clade. These analyses rely implicitly on the assumption that the phylogeny inferred from a single gene reflects the true evolutionary history of the entire genome. However, we have shown that phylogenies inferred from different genes are mutually incompatible. The assumption that gene and genome phylogenies are concordant is therefore untenable.

Our methodology depends on the fact that the insects that we studied were infected by a single bacterial strain. In addition to the evidence we present, further support for the absence of double infection comes from earlier breeding experiments on three of the male-killing strains (*Adalia bipunctata* strains Y and Z, and *Acraea encedon* strain U) in which these infections segregate perfectly with the male-killing trait, which is not expected if there is dual infection. Additionally, an identical pair of *ftsZ* and *wsp* sequences to that found in *Acraea encedon* strain U is also found in *Acraea encedana* (Jiggins *et al*. 2000*a*). This further supports the hypothesis that these two sequences, which are central to our analysis, are derived from a single bacterial strain.

There are two possible explanations for the incompatible tree topologies. First, they may result from recombination between *Wolbachia* strains. Second, the genes may have identical evolutionary histories, but features of their evolutionary dynamics may have caused an incorrect phylogeny to be inferred. Factors that can be positively misleading during phylogenetic reconstruction include substitution rate heterogeneity between lineages, base composition bias, and synonymous codon-usage bias. However, a detailed assessment of these factors has been undertaken for both of these genes, and none of them was found to apply to the *ftsZ* or *wsp* sequences of either *Aedes albopictus* or *Acraea encedon* strain U (Von der Schulenburg *et al*. 2000). Taken together with the extremely strong support that both parsimony- and likelihood-based analyses provide for the *ftsZ* and *wsp* genes of these taxa belonging to different clades, it is reasonable to conclude that there has been genetic recombination between *Wolbachia* lineages.

Homologous recombination is known to be widespread in bacteria, and can occur by a number of mechanisms including transformation, transduction (DNA exchange via a phage vector) and conjugation. In *Wolbachia*, recombination is likely to require that the bacteria come into close contact either transitorily, for instance during infection of hosts and their parasitoids, or in a stable co-infection of the same host.

Reconstruction of the evolution of specific traits from unrelated gene sequences is therefore unreliable and should only be attempted using genes that are directly involved in the expression of these traits. An important consequence of this observation is that the different reproductive phenotypes of *Wolbachia* can no longer be assumed to have arisen on multiple occasions. Furthermore, other conclusions based on *Wolbachia* phylogenies must be treated with caution. While it is indisputable that *Wolbachia* have moved horizontally between hosts (the arthropods diverged *ca*. 450 million years (Myr) ago, as compared with the *Wolbachia* that diverged only *ca*. 60 Myr ago), analysis of phylogenies cannot be used to investigate patterns of horizontal transmission, since horizontal transfer of a portion of the genome by recombination could easily be mistaken for horizontal transfer of the entire bacterium. For example, figure 1 suggests that the *wsp* gene of *Acraea encedon* strain U may be derived by recombination from a *Trichogramma* egg parasitoid.

Our analysis has only looked for recombination within the B subgroup of *Wolbachia*. However, there is some evidence that recombination may also occur between the A and B subdivisions of the genus, as diagnostic PCR tests based on different genes sometimes place the same *Wolbachia* infection in different subgroups (Werren *et al*. 1995). Recombination is perhaps unlikely in the C and D groups of *Wolbachia* found in nematodes as double infections are unknown, and horizontal transmission is rare.

Recombinant forms may be particularly likely to be observed when investigating genes involved in host–parasite coevolution, as it is the recombinants at these sites that may be most likely to spread by selection (as is the case for genes involved in antibiotic resistance). For this reason, the *wsp* gene, which encodes a surface protein, is especially ill-suited to its role in phylogenetic reconstruction, because the ratio of synonymous to non-synonymous substitutions suggests that it is under strong positive selection in some lineages, and is likely to be involved in host–symbiont interactions (Von der Schulenburg *et al*. 2000).

Recombination may be an important feature of *Wolbachia* evolution. For example, it may allow *Wolbachia* to escape the accumulation of mildly deleterious alleles by Müller's ratchet, which may be a particular problem for symbionts that pass through a population bottleneck in each generation (Moran 1996). Similarly, recombination may increase the rate of adaptive evolution, for instance during the evolution of novel cytoplasmic incompatibility crossing types.

The discovery of recombination may also have practical applications in attempts to exploit *Wolbachia* to control pests and human diseases. Recombination may allow the genetic modification of *Wolbachia*, but it may also pose problems for such efforts, since engineered genes may move between *Wolbachia* strains. Furthermore, if the genes inserted into the *Wolbachia* genome are costly, then recombination may cause them to be lost from the bacterial genome. Key questions still remain to be answered: at what rate do co-infections lead to recombination, and what is the mechanism by which this occurs?

We thank Nick Goldman for discussing our phylogenetic analysis. Scott O'Neill kindly supplied genomic DNA from *Aedes albopictus* and Chris Jiggins collected *Acraea equitoria*. F.M.J. was funded by a BBSRC studentship and an Emmanuel College (University of Cambridge) Research Fellowship, J.H.G.v.d.S. by a Training and Mobility of Researchers fellowship from the EU and G.D.D.H. by a BBSRC D. Phillips' Research Fellowship.

REFERENCES

Bandi, C., McCall, J. W., Genchi, C., Corona, S., Venco, L. & Sacchi, L. 1999 Effects of tetracycline on the filarial worms *Brugia pahangi* and *Dirofilaria immitis* and their bacterial endosymbionts Wolbachia. *Int. J. Parasitol.* **29**, 357–364.

Farris, J. S., Kallersjo, M., Kluge, A. G. & Bult, C. 1994 Testing significance of incongruence. *Cladistics* **10**, 315–319.

Huelsenbeck, J. P. & Bull, J. J. 1996 A likelihood ratio test to detect conflicting phylogenetic signal. *System. Biol.* **45**, 92–98.

Hurst, G. D. D., Jiggins, F. M., Von der Schulenberg, J. H. G., Bertrand, D., West, S. A., Goriacheva, I. I., Zakharov, I. A., Werren, J. H., Stouthamer, R. & Majerus, M. E. N. 1999 Male-killing *Wolbachia* in two species of insect. *Proc. R. Soc. Lond.* B **266**, 735–740.

Jiggins, F. M., Hurst, G. D. D., Dolman, C. E. & Majerus, M. E. N. 2000*a* High prevalence male-killing *Wolbachia* in the butterfly *Acraea encedana*. *J. Evol. Biol.* **13**, 495–501.

Jiggins, F. M., Hurst, G. D. D., Von der Schulenburg, J. H. G. & Majerus, M. E. N. 2000*b* Two male-killing *Wolbachia* strains coexist within a population of the butterfly *Acraea encedon*. *Heredity* **86**, 161–166.

Moran, N. A. 1996 Accelerated evolution and Müller's ratchet in endosymbiotic bacteria. *Proc. Natl Acad. Sci. USA* **93**, 2873–2878.

O'Neill, S., Pettigrew, M., Sinkins, S., Braig, H., Andreadis, T. & Tesh, R. 1997 *In vitro* cultivation of *Wolbachia pipientis* in an *Aedes albopictus* cell line. *Insect Mol. Biol.* **6**, 33–39.

O'Neill, S. L., Giordano, R., Colbert, A. M. E., Karr, T. L. & Robertson, H. M. 1992 16S ribosomal-RNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatability in insects. *Proc. Natl Acad. Sci. USA* **89**, 2699–2702.

Rambaut, A. & Grassly, N. C. 1997 Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput. Applic. Biosci.* **13**, 235–238.

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

Sambrook, J., Fritsch, E. F. & Maniatis, T. 1989 *Molecular cloning: a laboratory manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Stouthamer, R., Breeuwer, J. A. J., Luck, R. F. & Werren, J. H. 1993 Molecular identification of microorganisms associated with parthenogenesis. *Nature* **361**, 66–68.

Swofford, D. L. 1998 *PAUP*: phylogenetic analysis using parsimony (*and other methods)*, v. 4. Sunderland, MA: Sinauer Associates.

Von der Schulenburg, J. H. G., Hurst, G. D. D., Huigens, T. M. E., Van Meer, M. M. M., Jiggins, F. M. & Majerus, M. E. N. 2000 Molecular evolution and phylogentic utility of *Wolbachia ftsZ* and *wsp* gene sequences with special reference to the origin of male-killing. *Mol. Biol. Evol.* **17**, 584–600.

Werren, J. H., Zhang, W. & Guo, L. R. 1995 Evolution and phylogeny of *Wolbachia*: reproductive parasites of the arthropods. *Proc. R. Soc. Lond.* B **261**, 55–71.

Yang, Z. H. 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Applic. Biosci.* **13**, 555–556.

Zhou, W. F., Rousset, F. & O'Neill, S. 1998 Phylogeny and PCR based classification of *Wolbachia* strains using *wsp* gene sequences. *Proc. R. Soc. Lond.* B **265**, 509–515.