The Royal Society

Phylogeny and PCR-based classification of *Wolbachia* strains using *wsp* gene sequences

Weiguo Zhou[1,2], François Rousset[1,3] and Scott O'Neill[1*]

[1]*Section of Vector Biology, Department of Epidemiology and Public Health, Yale University School of Medicine, 60 College Street, New Haven, CT 06520, USA*
[2]*Institute of Genetics, Fudan University, Shanghai 200433, People's Republic of China*
[3]*Laboratoire Génétique et Environnement, Institut des Sciences de l'Evolution, Université de Montpellier II, 34095 Montpellier, France*

Wolbachia are a group of intracellular inherited bacteria that infect a wide range of arthropods. They are associated with a number of different reproductive phenotypes in their hosts, such as cytoplasmic incompatibility, parthenogenesis and feminization. While it is known that the bacterial strains responsible for these different host phenotypes form a single clade within the α-Proteobacteria, until now it has not been possible to resolve the evolutionary relationships between different *Wolbachia* strains. To address this issue we have cloned and sequenced a gene encoding a surface protein of *Wolbachia* (*wsp*) from a representative sample of 28 *Wolbachia* strains. The sequences from this gene were highly variable and could be used to resolve the phylogenetic relationships of different *Wolbachia* strains. Based on the sequence of the *wsp* gene from different *Wolbachia* isolates we propose that the *Wolbachia pipientis* clade be initially divided into 12 groups. As more sequence information becomes available we expect the number of such groups to increase. In addition, we present a method of *Wolbachia* classification based on the use of group-specific *wsp* polymerase chain reaction (PCR) primers which will allow *Wolbachia* isolates to be typed without the need to clone and sequence individual *Wolbachia* genes. This system should facilitate future studies investigating the distribution and biology of *Wolbachia* strains from large samples of different host species.

Keywords: cytoplasmic incompatibility; parthenogenesis; feminization; *Wolbachia*; *wsp*

1. INTRODUCTION

Wolbachia are maternally inherited intracellular rickettsia-like bacteria known to infect a wide range of arthropods. Recent surveys indicate that around 16% of all insect species may be infected with *Wolbachia*, making it one of the most ubiquitous endosymbionts described to date (Werren *et al.* 1995*a*). Infections with this agent have been associated with various reproductive abnormalities in the host, including cytoplasmic incompatibility (CI) in a variety of arthropod species, parthenogenesis in wasps, and feminization of genetic males in an isopod species (Werren 1997). The ability of *Wolbachia* to modify the reproductive success of its host enables it to increase in frequency in host populations without the need for horizontal transmission.

Through comparative analysis of 16S *rRNA* gene sequences it has been established that CI, parthenogenesis and feminization-inducing bacteria form a monophyletic clade in the α-Proteobacteria, being most closely related to the *Ehrlichia* assemblage (O'Neill *et al.* 1992; Rousset *et al.* 1992*b*; Stouthamer *et al.* 1993). However, the slow evolutionary rate of the 16S *rRNA* gene has not made it possible to adequately resolve a fine-scale phylogeny of *Wolbachia* strains with these sequences. More recently, the faster evolving cell-cycle gene *ftsZ* has been used to improve the phylogenetic resolution within the *Wolbachia* clade (Werren *et al.* 1995*b*). While it has been possible to discriminate broad *Wolbachia* groupings with both 16S *rRNA* and *ftsZ* sequences, neither gene has provided sufficient information to adequately resolve the relationships between individual *Wolbachia* strains which display different reproductive phenotypes (Rousset *et al.* 1992*a*; Werren *et al.* 1995*b*).

In this paper we report the cloning and sequencing of the *wsp* gene (Braig *et al.* 1997) from a number of representative *Wolbachia* strains and show that this gene is evolving at a much faster rate than any other previously reported *Wolbachia* gene. In addition we show that phylogenetic analysis based on *wsp* gene sequences results in an improved phylogenetic resolution of the *Wolbachia pipientis* assemblage. This resolution can be used to divide *Wolbachia* into a number of subgroups which appear to predict the reproductive phenotype of female *Drosophila* hosts. Furthermore we propose a system for naming and typing *Wolbachia* strains based on *wsp* gene sequences which obviates the need to clone and sequence genes from individual *Wolbachia* isolates to broadly classify *Wolbachia*.

2. METHODS

(a) *Insect strains*

The various insect strains used in this study and their source are listed in table 1.

*Author for correspondence (scott.oneill@yale.edu).

Received 3 December 1997 *Accepted* 16 December 1997

Table 1. *Insect species and strains used*

species	strain (supplier)	phenotype
Aedes albopictus	Houston, Texas	CI
Ephestia (Cadra) cautella	Gainesville (USDA), Florida	CI
Culex quinquefasciatus	Gainesville, Florida	CI
Culex pipiens	ESPRO - Tunisia	CI
Drosophila auraria	17.8	CI
Drosophila melanogaster	Aubiry 253 (A. Fleuriet)	CI weak
Drosophila melanogaster	Cairns (A. Hoffmann)	?
Drosophila melanogaster	Canton-S (P. Holden)	none
Drosophila melanogaster	Harwich (P. Holden)	none
Drosophila melanogaster	yw^{67c23}	CI weak
Drosophila sechellia	S-9	CI
Drosophila simulans	Coffs Harbour S-20 (A. Hoffmann)	none
Drosophila simulans	Hawaii	CI
Drosophila simulans	Riverside (M. Turelli)	CI
Drosophila simulans	DSW(Mau) (R. Giordano)	none
Drosophila simulans	Noumea (C. Biémont)	CI
Glossina austeni	Sth Africa (S. Aksoy)	?
Glossina m. centralis	Nairobi (S. Aksoy)	?
Glossina m. morsitans	Bristol (S. Aksoy)	?
Laodelphax striatellus	Yunnan Province, China	CI
Muscidifurax uniraptor	California	parthenogenesis
Nasonia vitripennis	Sweden E13 (R. Stouthamer)	CI
Phlebotomus papatasi	Israel	?
Tagosodes orizicolus	Costa Rica (A. Espinoza)	CI
Tribolium confusum	U. Vermont (L. Stevens)	CI
Trichogramma deion	Texas 223 (R. Stouthamer)	parthenogenesis

(b) *PCR amplification*

DNA was most commonly extracted using the crude STE boiling method (O'Neill *et al.* 1992) from either whole adult insects, abdomens or pupae. In some cases where this method did not produce DNA of sufficient quality for reliable PCR amplification we extracted DNA using the Holmes–Bonner method (Holmes & Bonner 1973) or a CTAB method (Guillemaud *et al.* 1997). The naturally *Wolbachia*-infected *Drosophila simulans* (Riverside) strain (DSR) and tetracycline treated DSR (DSRT) were used as positive and negative controls respectively.

Polymerase chain reactions (PCR's) were done in 20 μl reaction volumes: 13.5 μl dd H_2O, 2 μl 10 × buffer (Promega), 2 μl 25 mM $MgCl_2$, 0.5 μl dNTPs (10 mM each), 0.5 μl 20 μM forward and reverse primer and 1 unit of *Taq* DNA polymerase (Promega). PCR amplification was done under the following thermal profile: 94 °C 1 min, 55 °C 1 min and 72 °C 1 min per cycle for 35 cycles. In total 10 μl of PCR product was run on a 1% agarose gel to determine the presence and size of the amplified DNA.

General *wsp* primers were used as previously described (Braig *et al.* 1997): *wsp* 81F (5′TGG TCC AAT AAG TGA TGA AGA AAC) and *wsp* 691R (5′ AAA AAT TAA ACG CTA CTC CA) which were shown to be able to amplify the *wsp* gene fragment from all the *Wolbachia* strains tested in this paper. These primers amplify a DNA fragment ranging from 590 to 632 bp depending on the individual *Wolbachia* strain.

(c) *Cloning and sequencing*

For cloning, PCR products were incubated for an additional 90 min at 72 °C after 35 cycles of amplification. Then 1 μl of the PCR reaction was directly ligated into pGEM-T vector (Promega) without further purification in a 10 μl reaction overnight at 15 °C. At least three independent clones were sequenced for each *Wolbachia* strain to identify polymerase errors. Consensus sequences were generated from these multiple clones and used in further analysis. For insects which had previously been reported to be infected with multiple *Wolbachia* strains e.g. *D. simulans* Noumea, *Ephestia cautella* and *Aedes albopictus* (Rousset & Solignac 1995; Sinkins *et al.* 1995; Werren *et al.* 1995*b*) transformants were first screened with A (136F/691R) and B (81F/522R) group-specific primers (table 3) to quickly select clones for sequencing.

(d) *Phylogenetic analysis*

Partial *wsp* gene sequences from 28 strains of *Wolbachia* were aligned using the clustal algorithm followed by manual modifications based on the amino acid translation of the different genes. A 41 bp region (positions 519–559) corresponding to the third hypervariable region of the gene (Braig *et al.* 1997) was deleted from the analysis because it could not be aligned with confidence. The resulting alignment included 565 bases of which 205 were considered informative by parsimony criteria. This alignment has been deposited in the EMBL alignment database and is available by FTP from `ftp://ftp.ebi.ac.uk/pub/databases/embl/align/` under accession number DS32273. The data set was analysed by maximum parsimony using PAUP 3.1 (Swofford 1993). Branch-and-bound searches were done and the resulting trees were midpoint-rooted in the absence of a suitable outgroup. Bootstrap analysis was done with 500 replications. The same data set was also analysed by maximum likelihood using PHYLIP 3.57c (Felsenstein 1995) to search for the tree with the highest likelihood. This analysis was done by using a transition−transversion ratio of 2.0 and the assumption of one substitution rate.

3. RESULTS AND DISCUSSION

By using the general primers (81F, 691R) a fragment of the *wsp* gene was amplified from 28 *Wolbachia* strains.

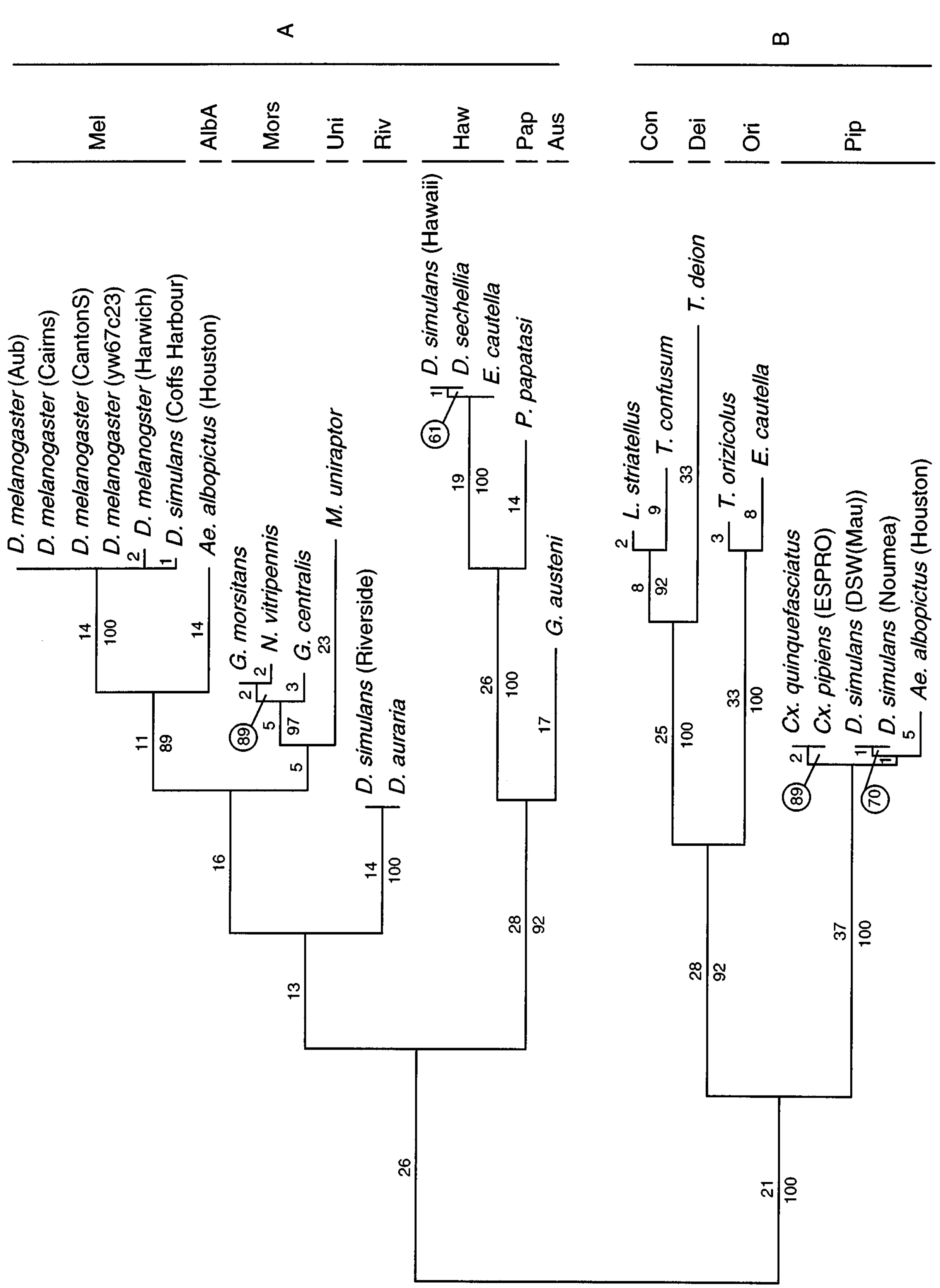

Figure 1. One of four most parsimonious trees generated from a branch and bound search of aligned *wsp* sequences (tree length = 472; CI = 0.64). Tree shown is midpoint rooted. Branch lengths, as determined from PAUP table of linkages, are labelled above branches and bootstrap values (500 replicates) are labelled below branches. Bootstrap values less than 50 are not shown. Taxa are labelled as the host from which the *Wolbachia* strain was isolated.

These primers were only able to amplify fragments from infected insects and not from uninfected hosts. A comparison of the sequences amplified from different taxa showed that they were up to 23% divergent, which is almost ten times greater than the divergence present in previously published 16S *rRNA* sequences of *Wolbachia* (O'Neill *et al.* 1992; Rousset *et al.* 1992*a*) and, to our knowledge, clearly makes it the fastest evolving *Wolbachia* gene yet described.

Table 2. *Sequences of* wsp *specific primers*

(Primers are numbered based on the *w*Ri *wsp* gene sequence corresponding to the 5′ base. Position 1 is equivalent to the first base of the coding region of the *w*Ri *wsp* gene.)

primer	sequence
81F	5′-TGG TCC AAT AAG TGA TGA AGA AAC
136F	5′-TG AAA TTT TAC CTC TTT TC
165F	5′-TGG TAT TAC AAA TGT AGC
169F	5′-ATT GAA TAT AAA AAG GCC ACA GAC A
172F	5′-ACC TAT AAG AAA GAC AAG
173F	5′-CCT ATA AGA AAG ACA ATG
178F	5′-AAA GAA GAC TGC GGA TAC
181F	5′-GAA GAC TGC AGA TAC TGC
183F	5′-AAG GAA CCG AAG TTC ATG
202F	5′-AAA (AG)GA TAG TCC CTT AAC
207F	5′-AGT GAT TAC AGT CCA TTG
211F	5′-CCA TCT TTT CTA GCT GGA
212[a]F	5′-GGA TAG T(AC)C CTT AA(AC) AAG
217F	5′-TT TAT AGC TGG TGG TGT
308F	5′-TTA AAG ATG TAA CAT TTG
328F	5′-CCA GCA GAT ACT ATT GCG
484R	5′-TTT GAT CAT TCA CAG CGT
522R	5′-ACC AGC TTT TGC TTG ATA
531R	5′-ATA (GA)CT (GA)AC ACC AGC TCT
691R	5′-AAA AAT TAA ACG CTA CTC CA

[a]Numbered to the nearest equivalent *w*Ri position for primers without a *w*Ri homologous base.

This is consistent with the observation that *wsp* is a single copy gene coding for an outer membrane protein of *Wolbachia* (Braig *et al.* 1997) and these genes are generally highly variable (Goward *et al.* 1993).

Wsp gene sequences provide many more informative characters with which to determine evolutionary relationships between strains. Figure 1 shows one of four most parsimonious trees obtained by analysis of aligned *wsp* sequences. The topologies of the four most parsimonious trees only differed from each other with respect to small changes within the *Pip* and *Mel* groups (figure 1). Bootstrap analysis showed strong support for the outer branches of the tree. However, deep nodes within the A group were not strongly supported. The *Wolbachia* strain infecting *Muscidifurax uniraptor* appears to account for much of this uncertainty. The position of this taxon within the tree was quite sensitive to small changes in the alignment of the data set. In addition, maximum likelihood analysis of the same data set produced a tree with the highest likelihood which had an identical topology to the one presented in figure 1 with the exception of the *M. uniraptor Wolbachia*. In this case the *M. uniraptor* symbiont formed a deep branch basal to the *G. austeni* symbiont. While the high level of variability in the *wsp* gene makes it ideal for strain typing and diagnostics, it raises some problems for phylogenetic analysis. This is particularly obvious in the deep nodes of the A group which show weak support from bootstrapping analysis. This uncertainty may well be caused by the high level of variability between different A group sequences. As such these nodes will need to be confirmed in future studies.

Surface proteins of intracellular parasites of mammals including related α-Proteobacteria like *Anaplasma* are subject to strong positive selection presumably in response to immune recognition by their host (Endo *et al.* 1996). It might be hypothesized that the *wsp* gene of *Wolbachia* may be subject to similar selection pressure which might compromise its value as a phylogenetic tool. We calculated synonymous substitution rates (d_S) and non-synonymous rates (d_N) between *wsp* gene fragments amplified from *Wolbachia* infecting *D. simulans* (Riverside) and *Culex pipiens* using the method developed by Nei & Gojobori (1986). The d_S and d_N values were 0.324 and 0.157, respectively. Surface proteins of bacterial parasites which are considered to be under strong selection have been reported to have $d_N : d_S$ ratios greater than one, as is the case for the *Anaplasma marginale msp 1* α-gene (Endo *et al.* 1996) However, the calculated value for the entire *wsp* gene was significantly lower (0.485). While this does not indicate that the *wsp* gene is not under selection, it does not appear to be under the same selection pressure as other described parasites. Considering that *Wolbachia* can be artificially transferred between distantly related hosts (Braig *et al.* 1994) it could be hypothesized that the intracellular environment experienced by different *Wolbachia* strains may have a small influence on its physiology. Indeed, *Wolbachia* may be able to stabilize its own microenvironment within the vacuole that typically encloses it.

Synonymous and non-synonymous rates were also calculated for the previously sequenced *ftsZ* gene (Werren *et al.* 1995*b*) from the *Wolbachia* strains infecting *Culex pipiens* and *D. simulans* (Riverside) using the same method ($d_S = 0.413$, $d_N = 0.056$). The synonymous rate as calculated by d_S for both genes was different, which might be a reflection of codon usage bias between these different *Wolbachia* genes. We calculated codon usage for the *wsp* gene as well as for the *ftsZ* and *dnaA* (Bourtzis *et al.* 1994) genes. The *wsp* gene did show significant codon usage bias when compared to these other genes (Fisher's exact test for contingency tables for each amino acid, global test by Fisher's combination of probabilities test (Sokal & Rohlf 1981), $p = 0.0039$). This bias can be mainly attributed to a high use of CGU (Arg), AAU (Asn), GGU (Gly) and to a lesser extent ACU (Thr) in the *wsp* sequence. This in turn may be related to the high expression levels of the *wsp* gene (Braig *et al.* 1997). These same codons are also over-represented in highly expressed *E. coli* genes (Sharp & Li 1987).

Considering that *wsp* sequences may be under selection, then it is unclear to what extent *wsp* based analyses may be confounded by convergence. If selection has not biased the phylogeny then the tree presented in figure 1 should be consistent with trees generated using other genes. Previous phylogenetic analyses have only been able to resolve a limited number of broad *Wolbachia* strain groupings, designated A and B (Werren *et al.* 1995*b*) (or I and II (Stouthamer *et al.* 1993)) and two subgroups within the A group based on *ftsZ* sequences (Werren *et al.* 1995*b*). The *wsp* analysis does support previous observations of the A and B groups (Rousset *et al.* 1992*b*; Werren *et al.* 1995*b*). Furthermore, in accordance with earlier work with the *ftsZ* gene, the B group *Wolbachia* show much greater total *wsp* gene sequence divergence than the A group (22% compared with 14% maximum). However, analysis of *wsp* sequences revealed a number of distinct *Wolbachia* clades within both the A and B groups which have not been resolved previously. A total of eight potential groups

Table 3. Wolbachia *group nomenclature and diagnostic primers for discriminating different groups*

Wolbachia group	supergroup	host and associated *Wolbachia* strain (reference strain is bolded)	GenBank accession number	forward primer	reverse primer	expected size of PCR product
	A Group			136F	691R	556 bp
Mel		***Drosophila melanogaster (yw^{67C23}) w*****Mel**	AF020072	308F	691R	405 bp
		D. melanogaster (Aubiry 253) *w*Mel	AF020063			
		D. melanogaster (Canton-S) *w*MelCS	AF020065			
		D. melanogaster (Cairns) *w*MelCS	AF020064			
		D. melanogaster (Harwich) *w*MelH	AF020066			
		Drosophila simulans (Coffs Harbour) *w*Cof	AF020067			
AlbA		***Aedes albopictus w*****AlbA**	AF020059	328F	691R	379 bp
	A: Mel+AlbA			172F	691R	541 bp
Mors		***Glossina morsitans w*****Mors**	AF020079	173F	691R	516 bp
		*Nasonia vitripennis w*VitA	AF020081			
		*Glossina centralis w*Cen	AF020078			
Riv		***Drosophila simulans*** **(*Riverside*) *w*Ri**	AF020070	169F	691R	523 bp
		*Drosophila auraria w*Ri	AF020062			
	A: Mors+Ri			81F	484R	404 bp
Uni		***Muscidifurax uniraptor w*****Uni**	AF020071	207F	691R	493 bp
Haw		***Drosophila simulans*** **(*Hawaii*) *w*Ha**	AF020068	178F	691R	581 bp
		*Drosophila sechellia w*Ha	AF020073			
		Ephestia cautella (A) *w*CauA	AF020075			
Pap		***Phlebotomus papatasi w*****Pap**	AF020082	181F	691R	506 bp
Aus		***Glossina austeni w*****Aus**	AF020077	165F	691R	506 bp
	A: Uni+Ha+ Pap+Aus			81F	531R	460 bp
	B Group			81F	522R	442 bp
Con		***Tribolium confusum w*****Con**	AF020083	202F	691R	488 bp
		*Laodelphax striatellus w*Stri	AF020080			
Dei		***Trichogramma deion w*****Dei**	AF020084	217F	691R	463 bp
	B: Con+Dei			212F	691R	485 bp
Pip		***Culex pipiens w*****Pip**	AF020061	183F	691R	501 bp
		*Culex quinquefasciatus w*Pip	AF020060			
		Drosophila simulans (*mauritiana*) *w*Ma	AF020069			
		Drosophila simulans (Noumea) *w*No	AF020074			
		*Aedes albopictus w*AlbB	AF020059			
CauB		***Ephestia cautella w*****CauB**	AF020076	211F	691R	466 bp
		*Tagosodes orizicolus w*Ori	AF020085			

could be recognized within the A group and four within the B group using the sequenced strains. To determine the extent to which the analysis presented in figure 1 is representative of a true *Wolbachia* phylogeny sequences from other *Wolbachia* genes will be needed which are evolving at a faster rate than the currently sequenced *fts*Z and 16S *rRNA* genes.

The *wsp* gene is a very useful tool for typing different *Wolbachia* strains. The large variability in the gene makes it possible to design specific PCR primers which can recognize both individual strains of *Wolbachia* or groups of *Wolbachia* strains. Examples of a number of such primers and their application are presented in tables 2 and 3. By using a combination of these primers in PCR reactions it is possible to quickly assign an unknown *Wolbachia* strain to a particular group. This approach will allow for different *Wolbachia* strains to be rapidly typed without the need to individually clone and sequence genes from all new isolates. We envisage that this will become increasingly useful as *Wolbachia* are proving to be such a widespread group of parasites that it will be unrealistic to individually characterize all isolates by sequence analysis. The use of such a system will allow for large-scale studies on *Wolbachia* strain distribution and natural history which until now have not been possible. However, to efficiently use this form of *Wolbachia* classification a naming system is needed for *Wolbachia* groups and supergroups. We propose that, for typing purposes, each named *Wolbachia* group be defined by the *wsp* sequence similarity of its members, which should generally be greater than 97.5% identical. While this similarity value is by definition arbitrary, it should allow for a relatively fine-scale system of *Wolbachia* strain grouping. We propose that individual *Wolbachia* strains be given a unique name if they have a sequenced *wsp* gene which differs from other sequenced *wsp* genes and that these names follow the already accepted abbreviation style *w*Host (Rousset & de Stordeur 1994). The exception to this convention would be to give *Wolbachia* variants with identical *wsp* sequences unique names only if they show different phenotypic effects in hosts. For example, the *Wolbachia* strains *w*No and *w*Ma which naturally infect different *Drosophila* species have identical *wsp* sequences but different reproductive phenotypes in males when transferred into *D. simulans*, as such they are given unique names (table 3).

Wolbachia groups would then be named according to a reference *Wolbachia* strain within the group. For example the *Wolbachia* group which contains the *w*Mel *Wolbachia* strain would be known as the *Mel* group. The work presented in this paper describes 12 such groups, eight from the A group *Wolbachia* and four from the B group (table 3). As more strains are sequenced the number of *Wolbachia* groups can be expected to increase and be accommodated within this nomenclature.

The names and reference strains for the *Wolbachia* groups presented in this paper are shown in table 3. For each group, specific primer sets were designed based on *wsp* sequence data. The specificity of the primer sets was tested by doing PCR on samples from each group. The results confirmed that the primer sets are diagnostically specific to each group without cross-reacting with other groups, using the *Wolbachia* strains described in this paper. In addition supergroup primers were designed with less specificity which could amplify DNA from a number of specific groups (table 3). The use of supergroup primers allows for the rapid classification of an unknown *Wolbachia* strain to a specific group in a more efficient way than randomly amplifying an isolate with all the available group-specific primer sets. An unknown sample would be first amplified with more general primers to determine which group-specific primers should then be used.

The use of *wsp* gene sequences as diagnostic and evolutionary tools will have immediate impact on studies on the biology of *Wolbachia*. For example, the ability to determine degrees of relatedness between different strains of *Wolbachia* infecting *Drosophila* can be used to successfully predict the ability of different infected *Drosophila* females to successfully rescue the CI imprint of males carrying related *Wolbachia* strains; females being able to rescue the imprint of males from the same *Wolbachia* group in the absence of density effects between strains. This has proven especially useful in being able to predict that females from *Drosophila* 'non-expressor' or A crossing types (Turelli & Hoffmann 1995; Hoffmann *et al.* 1996) can rescue the sperm modification generated by related 'expressor' males from the same group (Bourtzis *et al.* 1997). We would anticipate that a *wsp* based classification will also prove to be largely predictive with regard to CI rescue for *Wolbachia* strains infecting other species. The ability to use *wsp* sequences to predict other phenotypes such as parthenogenesis may be limited. As in previous phylogenetic analyses, *wsp* analysis confirms that the parthenogenesis phenotype is polyphyletic within the *Wolbachia*. This in turn suggests a potential host component to the expression of this phenotype. In addition, the *wsp* gene can provide tools to individually track multiple infections within individual hosts to study segregation and competition effects which until now could only be achieved with a limited number of distantly related *Wolbachia* strains using techniques based on 16S or *ftsZ* sequences. The usefulness of this system will ultimately depend on expanding the number of *Wolbachia* variants sequenced beyond what is presented in this paper so that a thorough system of *Wolbachia* group-specific primer sets can be defined.

We thank H. Braig, K. Bourtzis, J. Werren, M. van Meer and A. Hoffmann for critical comments on the manuscript. We also thank S. Aksoy, H. Braig, C. Biémont, K. Bourtzis, T. Carde, A. Espinoza, A. Fleuriet, P. Holden, R. Stouthamer, M. Turelli and L. Stevens for supplying insect strains. This work was supported by grants from the National Institutes of Health (AI 34355, AI 40620), the McKnight foundation, the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases, and a NATO research grant to F.R.

REFERENCES

Bourtzis, K., Dobson, S. L., Braig, H. R. & O'Neill, S. L. 1998 *Wolbachia* that only rescue. *Nature.* (In the press.)

Bourtzis, K., Nirgianaki, A., Onyango, P. & Savakis, C. 1994 A prokaryotic *dnaA* sequence in *Drosophila melanogaster*: *Wolbachia* infection and cytoplasmic incompatibility among laboratory strains. *Insect Molec. Biol.* **3**, 131–142.

Braig, H. R., Guzman, H., Tesh, R. B. & O'Neill, S. L. 1994 Replacement of the natural *Wolbachia* symbiont of *Drosophila simulans* with a mosquito counterpart. *Nature* **367**, 453–455.

Braig, H. R., Zhou, W., Dobson, S. & O'Neill, S. L. 1998 Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia*. (Submitted.)

Endo, T., Ikeo, K. & Gojobori, T. 1996 Large-scale search for genes on which positive selection may operate. *Molec. Biol. Evol.* **13**, 685–690.

Felsenstein, J. 1995 PHYLIP: phylogeny inference package, version 3.57c: Felsenstein, J., Department of Genetics, University of Washington, Seattle.

Goward, C. R., Scawen, M. D., Murphy, J. P. & Atkinson, T. 1993 Molecular evolution of bacterial cell-surface proteins. *Trends Biochem. Sci.* **18**, 136–140.

Hoffmann, A. A., Clancy, D. & Duncan, J. 1996 Naturally-occurring *Wolbachia* infection in *Drosophila simulans* that does not cause cytoplasmic incompatibility. *Heredity* **76**, 1–8.

Holmes, D. S. & Bonner, J. 1973 Preparation, molecular weight, base composition and secondary structure of giant nuclear ribonucleic acid. *Biochemistry* **12**, 2330–2338.

Nei, M. & Gojobori, T. 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitution. *Molec. Biol. Evol.* **3**, 418–426.

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

Rogers, S. O. & Bendich, A. J. 1988 Extraction of DNA from plant tissues. *Pl. Molec. Biol. Man.* **A6**, 1–10.

Rousset, F. & de Stordeur, E. 1994 Properties of *Drosophila simulans* strains experimentally infected by different clones of the bacterium *Wolbachia*. *Heredity* **72**, 325–331.

Rousset, F. & Solignac, M. 1995 Evolution of single and double *Wolbachia* symbioses during speciation in the *Drosophila simulans* complex. *Proc. Natn. Acad. Sci. USA* **92**, 6389–6393.

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

Rousset, F., Vautrin, D. & Solignac, M. 1992*b* Molecular identification of *Wolbachia*, the agent of cytoplasmic incompatibility in *Drosophila simulans*, and variability in relation with host mitochondrial types. *Proc. R. Soc. Lond.* B **247**, 163–168.

Sharp, P. M. & Li, W. 1987 The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential application. *Nucleic Acids Res.* **15**, 1281–1295.

Sinkins, S. P., Braig, H. R. & O'Neill, S. L. 1995 *Wolbachia* superinfections and the expression of cytoplasmic incompatibility. *Proc. R. Soc. Lond.* B **261**, 325–330.

Sokal, R. R. & Rohlf, F. J. 1981 *Biometry.* New York: W. H. Freeman.

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

Swofford, D. L. 1993 PAUP: phylogenetic analysis using parsimony, version 3.1. Champaign, IL: Center for Biodiversity, Illinois Natural History Survey.

Turelli, M. & Hoffmann, A. A. 1995 Cytoplasmic incompatibility in *Drosophila simulans*: dynamics and parameter estimates from natural populations. *Genetics* **140**, 1319–1338.

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

Werren, J. H., Guo, L. R. & Windsor, D. W. 1995*a* Distribution of *Wolbachia* in neotropical arthropods. *Proc. R. Soc. Lond.* B **262**, 197–204.

Werren, J. H., Zhang, W. & Guo, L. R. 1995*b* Evolution and phylogeny of *Wolbachia*: reproductive parasites of arthropods. *Proc. R. Soc. Lond.* B **261**, 55–63.