Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 2699–2702, April 1992
Microbiology

16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects

(*Wolbachia pipientis*/rickettsia/α-Proteobacteria/PCR)

SCOTT L. O'NEILL*†, ROSANNA GIORDANO*, ANGELA M. E. COLBERT*, TIMOTHY L. KARR‡, AND HUGH M. ROBERTSON*

Departments of *Entomology and ‡Biochemistry, University of Illinois, Urbana, IL 61801

Communicated by Irwin C. Gunsalus, December 24, 1991

ABSTRACT **Bacterial endosymbionts of insects have long been implicated in the phenomenon of cytoplasmic incompatibility, in which certain crosses between symbiont-infected individuals lead to embryonic death or sex ratio distortion. The taxonomic position of these bacteria has, however, not been known with any certainty. Similarly, the relatedness of the bacteria infecting various insect hosts has been unclear. The inability to grow these bacteria on defined cell-free medium has been the major factor underlying these uncertainties. We circumvented this problem by selective PCR amplification and subsequent sequencing of the symbiont 16S rRNA genes directly from infected insect tissue. Maximum parsimony analysis of these sequences indicates that the symbionts belong in the α-subdivision of the Proteobacteria, where they are most closely related to the *Rickettsia* and their relatives. They are all closely related to each other and are assigned to the type species *Wolbachia pipientis*. Lack of congruence between the phylogeny of the symbionts and their insect hosts suggests that horizontal transfer of symbionts between insect species may occur. Comparison of the sequences for *W. pipientis* and for *Wolbachia persica*, an endosymbiont of ticks, shows that the genus *Wolbachia* is polyphyletic. A PCR assay based on 16S primers was designed for the detection of *W. pipientis* in insect tissue, and initial screening of insects indicates that cytoplasmic incompatibility may be a more general phenomenon in insects than is currently recognized.**

The phenomenon of cytoplasmic incompatibility has been observed in a diverse array of insect species from five orders (1). Incompatible crosses are characterized by abortive karyogamy (2–5), which for most insect species leads to early embryonic death. In the case of the haploid/diploid sex determining Hymenoptera however, incompatibility is not lethal but, instead, results in a distorted sex ratio biased toward the haploid sex (2, 5, 6). The molecular basis of cytoplasmic incompatibility is unknown, but it has been demonstrated that the genetic determinants are maternally inherited (7, 8) and correlate with the presence of rickettsia-like bacterial endosymbionts in the arthropods' gonad tissue (9–13). In most cases the incompatible cross is that between infected males and uninfected females, although in some species reciprocal crosses between infected individuals are also incompatible, yielding multiple crossing types within and between species (4, 5, 9). The postmating reproductive isolation and sex ratio distortions generated by cytoplasmic incompatibility have intriguing evolutionary implications and potential for the genetic manipulation of arthropod populations of economic and public health importance (14–17).

The bacterium associated with incompatibility in the mosquito *Culex pipiens* was first observed and described as *Wolbachia pipientis* by Hertig (18). While incompatibility was first characterized in *Culex pipiens* (15), it has since been reported in a number of other insects (1). The relationship between the symbionts of these other insects and *W. pipientis* has been unclear, as has the taxonomic position of *W. pipientis* relative to other described members of the genus *Wolbachia* and, indeed, their phylogenetic position within the eubacteria. This uncertainty is largely due to the fastidious nature of these bacteria, which has prevented them from being cultured outside their insect hosts (19). The 16S rRNA molecule is now widely recognized and used as a conservative macromolecule that allows phylogenetic placement of bacterial species (20). It has recently been used to resolve the phylogenetic positions of a variety of bacteria placed in the family Rickettsiaceae (21, 22). Using 16S rRNA primers specific to eubacteria, we have selectively amplified, cloned, and sequenced symbiont 16S rRNA genes from total genomic DNA extracted from insect ovary tissue.§ This approach has enabled us to resolve these issues by maximum parsimony analysis of sequences of the 16S rRNA genes of the bacteria associated with incompatibility. Representative symbionts from six insect species were analyzed as well as symbionts from two lines of *Drosophila simulans* known to possess differing cytoplasmic incompatibility crossing types (4).

The data generated from this study have also been used to develop an assay for determining infection status of individual insects. Specific PCR primers from the variable V1 and V6 regions of the 16S gene were designed to detect *W. pipientis* infection in insect tissue. Initial studies using this assay in a survey of pest insect species not previously known to display incompatibility show that this bacterium, and likely, therefore, the phenomenon of cytoplasmic incompatibility, is much more pervasive in insects than is currently recognized.

MATERIALS AND METHODS

Insect Strains. The following insects were used: *Culex pipiens* collected from Champaign, IL; *Drosophila simulans* Riverside (12); *Drosophila simulans* Hawaii (2); *Hypera postica* western strain (10); *Aedes albopictus* collected from East St. Louis, IL; *Tribolium confusum* BIV (1); and *Ephestia cautella* lab strain from the U.S. Department of Agriculture/Agriculture Research Station, Gainesville, FL.

Phylogenetics. Insect ovary tissue was dissected into 50 μl of STE (100 mM NaCl/10 mM Tris Cl, pH 8.0/1 mM EDTA, pH 8.0), homogenized with a clean (DNA-free acid-washed) sterile polypropylene pestle, and incubated with 2 μl of proteinase K (10 mg/ml) for 30 min at 37°C, followed by 5 min

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

†Present address: Department of Epidemiology and Public Health, Yale University School of Medicine, P.O. Box 3333, New Haven, CT 06510.

§The sequences reported in this paper have been deposited in the GenBank data base (accession nos. X61767, X61768, X61769, X61770, X61771, X62247, and X62248).

at 95°C. Samples were briefly centrifuged in a microcentrifuge, and 1 μl of the supernatant was used as the template in subsequent PCRs. PCR conditions included 2.5 mM $MgCl_2$, all four dNTPs (each at 200 μM), and 400 nM primer concentration. A temperature profile of 95°C for 1 min, 55°C for 1 min, and 72°C for 1 min was used for 35 cycles with the following eubacterial specific primers: 5'-GCTTAACACATGCAAG, *Escherichia coli* positions 45–61 forward; 5'-CCATTGTAGCACGTGT, *Escherichia coli* positions 1242–1227 reverse. These primers amplified an ≈1180-base-pair (bp) fragment of the gene from all insect samples. In addition the following primers were used to amplify a larger region from *Culex pipiens* tissue: 5'-AYTTTGAGAGTTTGATCCTG, *Escherichia coli* positions 2–21 forward (where Y is a pyrimidine); 5'-ACGGGCAGTGTGTACAAGACC, *Escherichia coli* positions 1406–1386 reverse.

Appropriate controls with naturally uninfected and tetracycline-cured lines of insects showed that PCR amplification only occurred in lines that were known to be carrying the symbiont of interest, thus only the DNA from the symbiont was amplified and not that of other contaminating bacteria. PCR products were cloned directly into the pCR2000 plasmid vector (TA cloning kit; Invitrogen, San Diego) and three to six clones from each insect population were sequenced using Sequenase Version 2.0 (United States Biochemical). A consensus sequence was generated for each species from these multiple clones to overcome *Taq* polymerase errors. Sequences were manually aligned with assistance from the preexisting alignments of the Ribosomal Database Project (23) and subjected to maximum parsimony analysis using PAUP Version 3.0 for the Macintosh (24).

PCR Assay. DNA extraction and PCR conditions were the same as outlined above except that annealing was carried out at 52°C instead of 55°C and the number of cycles was reduced to 30. Primers designed to be specific for *W. pipientis* were 5'-TTGTAGCCTGCTATGGTATAACT, which is in the variable V1 region and corresponds to *Escherichia coli* positions 76–99 forward, and 5'-GAATAGGTATGATTTTCATGT, which is the reverse complement of the variable V6 region and corresponds to *Escherichia coli* positions 1012–994 reverse. The following primers specific for insect mitochondrial 12S rRNA were used as controls to check for the quality of each DNA extraction: 5'-AAACTAGGATTAGATACCCTATTAT and 5'-AAGAGCGACGGGCGATGTGT, known as 12SAI and 12SBI, respectively (25).

RESULTS AND DISCUSSION

Nearly the entire 16S rRNA gene (1325 bp) was sequenced from *W. pipientis*, the symbiont of the mosquito *Culex pipiens*, which is the type species for the genus *Wolbachia* (18). For the other six insect strains or species, 623–691 bases of the 5' region of the symbiont 16S rRNA gene were sequenced. Variation among the six clones sequenced from *Culex pipiens* tissue was 0.19%. Similarly, variation between the clones sequenced from each of the other insect strains ranged from 0.04% to 0.20%. This variation might reflect intercistronic variation, assuming there are multiple cistrons as is the case for most bacteria, or might result from *Taq* polymerase infidelity. Of the clones displaying variation from the consensus at a given site however, none showed further compensating substitutions in other regions of the gene that would correct for secondary structure modifications of stems induced by the initial substitution. Also, none of the variable sites was shared by more than one clone at any given position. Furthermore, the amount of variation observed is in the known range for *Taq* polymerase incorporation error rates (26, 27). These observations suggest that most of the observed variation is due to *Taq* polymerase infidelity with intercistronic variation being a minor component at most. This conclusion is in agreement with findings from related bacteria in which intercistronic variation is minimal (28). In addition, of the 21 total intrapopulation variable positions, only three sites (one from *Culex pipiens*, one from *Tribolium confusum*, and one from *Aedes albopictus*) possessed substitutions that could have been interpreted as potential synapomorphies in the later analysis. Therefore, intrapopulation variation was considered to be of negligible influence on the interpretation of the sequence data and a consensus sequence for each population was used in the phylogenetic analysis.

Maximum parsimony analysis of the nearly full-length sequence of *W. pipientis* from *Culex pipiens* tissue, aligned with representative eubacteria, revealed that *W. pipientis* is situated in the α-subdivision of the Proteobacteria (29) (Fig. 1). This subdivision contains most of the purple nonsulfur photosynthetic bacteria and many nonphotosynthetic endosymbionts of eukaryotic cells, including the mitochondria (30). The nearest relatives of *W. pipientis* are *Anaplasma marginale*, *Ehrlichia risticii*, and the *Rickettsia* spp., all arthropod-born pathogens of mammals (21, 22). This placement with the rickettsia confirms many observations, primarily using the electron microscope, that these endosymbionts are rickettsia-like (9–13). Closer examination of the relationship between *W. pipientis* and these rickettsial relatives, whose sequences can be aligned with more confidence, confirmed that *Anaplasma marginale* represents the sister group of *W. pipientis* (Fig. 1). They have, however, diverged significantly as indicated by their deep branchings in the tree. For example, although *W. pipientis* aligns most closely with these mammalian pathogens, it is not pathogenic to mammals (19).

The only other member of the genus *Wolbachia* for which 16S rRNA sequence data are available is the tick symbiont *Wolbachia persica* (21). It is clear from comparison of the 16S sequences of *W. persica* and *W. pipientis* that the two bacteria are not closely related, *W. persica* belonging in the γ-subdivision of the Proteobacteria (21). If the genus *Wolbachia* is constrained to be monophyletic prior to parsimony analysis, the shortest tree that can be generated is 85 steps longer than the tree of 2165 steps presented in Fig. 1. Clearly, the genus *Wolbachia* as currently defined is polyphyletic, and since *W. pipientis* is the type species for the genus, *W. persica* should be removed from the genus *Wolbachia*.

Partial 16S rRNA sequences were obtained for symbionts of the beetles *Hypera postica* and *Tribolium confusum*, the mosquito *Aedes albopictus*, the moth *Ephestia cautella*, and two bidirectionally incompatible strains of *Drosophila simulans* originally taken from Hawaii and Riverside, California (4). The sequences from these symbionts show them to be a monophyletic assemblage with divergences ranging from 0.2% to 2.6% for the 623-bp 5' region of the gene (Fig. 2). The relevance of 16S rRNA divergence to species status in bacteria is a relatively new issue and no set rules are established (31). Some species, which have been described and generally accepted as distinct, exhibit low divergence (<2%) in 16S rRNA sequence. Examples include the *Bacillus thuringiensis* group (32) and the *Rickettsia* spp. shown in Fig. 2. In others, subspecies show divergence bordering on 3%, for example, in the genus *Fusobacterium* (33). In the absence of any other differentiating characters, the similarity in ecology and phenotypic effect that these *Wolbachia* symbionts have on their hosts provisionally warrants that they be classified as members of the same species, *W. pipientis*. Similarly, closely related sequences have been obtained independently from the symbionts of parasitic wasps of the genus *Nasonia* (35), confirming that the bacteria responsible for cytoplasmic incompatibility in different hosts are all closely related. We conclude that cytoplasmic incompatibility is not a general physiological response to infection by a diverse range of bacteria, rather it is associated with a specialized bacterium that infects a wide range of different arthropod hosts.

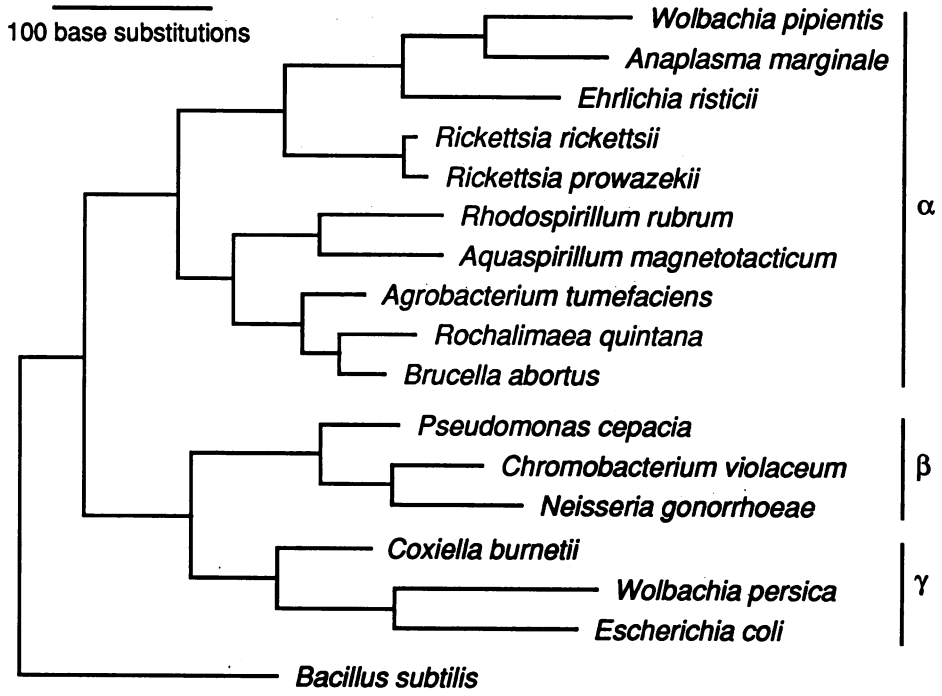

FIG. 1. Phylogenetic tree derived from maximum parsimony analysis of the 16S rRNA genes from *W. pipientis* (taken from *Culex pipiens* mosquitoes) (GenBank accession no. X61768) and other representative proteobacteria from the α, β, and γ subdivisions. The Gram-positive bacterium *Bacillus subtilis* was used as the outgroup. The 1325 bases determined for *W. pipientis* were manually aligned to the corresponding regions from the other bacteria. This tree is the most parsimonious generated by the Heuristic algorithm of PAUP Version 3.0. Deletion of the variable regions V1–V9 from the alignment did not alter the relationships. A more restricted analysis was performed to examine more closely the relationship of *Wolbachia* and its rickettsial relatives. In this case the *Rickettsia* spp. were used as the outgroup and the topology shown in this figure was confirmed by the Exhaustive algorithm of PAUP Version 3.0, yielding a tree of 459 steps with *Anaplasma marginale* as the sister group of *W. pipientis* versus a tree of 466 steps to make *Ehrlichia risticii* the sister group.

Data from the 16S rRNA gene provide a limited number of characters that are informative for the phylogeny of the strains of *W. pipientis*, so the topology of the *Wolbachia* branches of the tree presented in Fig. 2 is not very robust. However, aspects of the tree appear to be informative. For example, the symbionts of the fruitfly *Drosophila simulans* are more closely related to those of the moth *Ephestia cautella* than they are to the symbionts of the other dipterans, *Culex pipiens* and *Aedes albopictus*. This lack of congruence of the phylogenies of the *Wolbachia* and their insect hosts can be examined more rigorously by constraining the symbionts from hosts of different insect families and orders to be monophyletic prior to parsimony analysis. By using *Anaplasma marginale* as the outgroup for this closer comparison, the most parsimonious tree that can be generated under these constraints is 33 steps long, 11 steps longer than the most parsimonious unconstrained topology in Fig. 2. This analysis suggests that the symbiont has been acquired more than once by different insects, further justifying the classification of these bacteria as members of the same species and arguing against naming the symbionts after the insect hosts they infect (34). Thorough analysis of the strains of *W. pipientis* will require additional data from a more rapidly evolving gene sequence.

The observation that incompatibility is associated with a specific group of bacteria and the acquisition of sequence

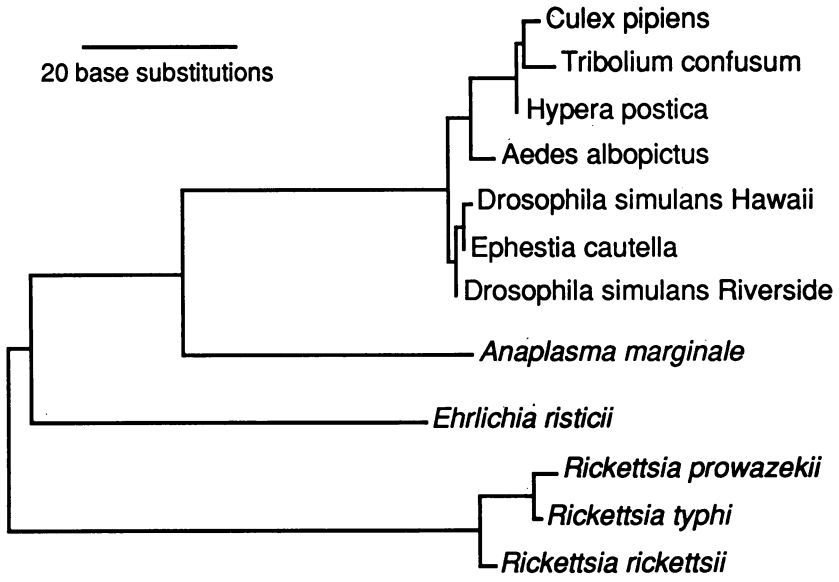

FIG. 2. Phylogenetic tree derived from maximum parsimony analysis of 623 bases sequenced from the 5′ region of the 16S rRNA gene of symbionts from the following insect hosts: *Hypera postica* (GenBank accession no. X62248), *Tribolium confusum* (X62247), *Culex pipiens* (X61768), *Aedes albopictus* (X61767), *Drosophila simulans* Riverside (X61770), *Drosophila simulans* Hawaii (X61769), and *Ephestia cautella* (X61771), aligned with the corresponding regions of the gene from *Ehrlichia risticii*, *Anaplasma marginale*, and the *Rickettsia* spp. The *Rickettsia* spp. were defined as the outgroup, and this tree is one of four most parsimonious trees generated by the Branch-and-Bound algorithm of PAUP Version 3.0, the other three differing in details of the arrangement of the most closely related *Wolbachia* strains. The names of the insect hosts are not italicized to indicate that these are not considered to be specific names for the *Wolbachia*.

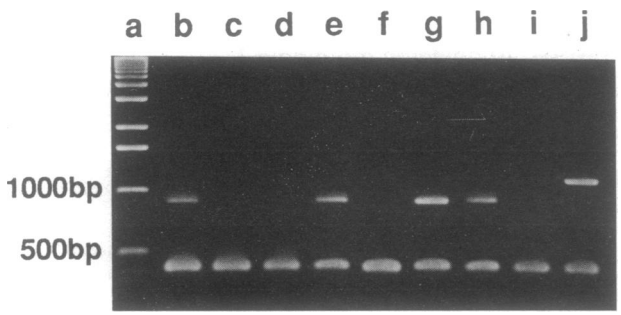

FIG. 3. PCR assay for *Wolbachia* infection using the specific 76–99 forward and 1012–994 reverse primers in lanes b–h, and the specific 76–99 forward primer combined with the general eubacterial 1242–1227 reverse primer in lanes i and j. Mitochondrial 12S rRNA primers were also used in separate PCRs as a control for the quality of the DNA extraction (band at 400 bp). For the purpose of this figure the PCR products from both reaction mixtures were combined and electrophoresed on a 1.2% agarose gel, stained with ethidium bromide, and visualized with UV. Lanes: a, BRL 1-kilobase ladder; b, *Drosophila simulans* Riverside (naturally infected) (12); c, *Drosophila simulans* Riverside (tetracycline-cured strain) (4); d, *Drosophila simulans* Watsonville (naturally uninfected) (12); e, *Hypera postica* western strain (naturally infected) (10); f, *Hypera postica* eastern strain (naturally uninfected) (10); g, *Culex pipiens* (naturally infected); h, *Aedes albopictus* (naturally infected); i, *Tribolium confusum* BI (naturally uninfected) (1); j, *Tribolium confusum* BIV (naturally infected) (1).

data from variable regions of the 16S gene have allowed design of species-specific oligonucleotide primers that distinguish *W. pipientis* from other members of the α-Proteobacteria for which sequence data are currently available. These primers have been used in PCRs to determine reliably the infection status of infected, uninfected, and tetracycline-cured strains of *Drosophila simulans, Tribolium confusum, Hypera postica, Culex pipiens*, and *Aedes albopictus*. DNA extracted from single ovaries was adequate for detection (Fig. 3). The application of this assay to DNA extractions from ovaries of various other insect species has revealed a number of cases of infection in insects that were previously unknown to harbor such symbionts. Examples include the rice moth *Corcyra cephalonica*, the angoumois grain moth *Sitotroga cerealella*, the western corn rootworm *Diabrotica virgifera virgifera*, the black carpet beetle *Attagenus unicolor*, the apple maggot fly *Rhagoletis pomonella*, the blueberry maggot fly *Rhagoletis mendax*, and the Caribbean fruit fly *Anastrepha suspensa* (R.G., unpublished data). Whether these insects also display cytoplasmic incompatibility is unknown.

The observed diversity of the known host range of this bacterium and a growing number of examples of infection suggest that the phenomenon of cytoplasmic incompatibility is likely to be more pervasive among arthropods than currently suspected. The finding that cytoplasmic incompatibility in different insects is associated with a single species of bacterium suggests that the mechanism of incompatibility will prove to be similar in different insect species.

We gratefully acknowledge the assistance provided by the following people: C. R. Woese for generously providing some of the primers used in PCR amplification and sequencing; F. Adams, T. Hsiao, S. Kambhampati, T. Lyttle, L. Munstermann, B. Novak, M. Turelli, and M. Wade for supplying insect material; H. Breeuwer, E. Stackebrandt, D. Stahl, D. Stock, C. R. Woese, and J. Werren for helpful discussion and comments on the manuscript. This work was supported by grants from the National Science Foundation (DMB 88-18889 to H.M.R.) and the U.S. Department of Agriculture (AG 91-37302-6766 to H.M.R. and T.L.K.) and used the facilities of the Center for Molecular Phylogeny sponsored by the University of Illinois Natural History Museum.

1. Stevens, L. & Wade, M. J. (1990) *Genetics* **124,** 367–372.
2. Ryan, S. L. & Saul, G. B., II (1968) *Mol. Gen. Genet.* **103,** 29–36.
3. Jost, E. (1971) *Can. J. Genet. Cytol.* **13,** 237–250.
4. O'Neill, S. L. & Karr, T. L. (1990) *Nature (London)* **348,** 178–180.
5. Breeuwer, J. A. J. & Werren, J. H. (1990) *Nature (London)* **346,** 558–560.
6. Richardson, P. M., Holmes, W. P. & Saul, G. B. (1987) *J. Invertebr. Pathol.* **50,** 176–183.
7. Laven, H. (1959) *Cold Spring Harbor Symp. Quant. Biol.* **24,** 166–173.
8. Hoffmann, A. A. & Turelli, M. (1988) *Genetics* **119,** 435–444.
9. Yen, J. H. & Barr, A. R. (1973) *J. Invertebr. Pathol.* **22,** 242–250.
10. Hsiao, C. & Hsiao, T. H. (1985) *J. Invertebr. Pathol.* **45,** 244–246.
11. O'Neill, S. L. (1989) *J. Invertebr. Pathol.* **53,** 132–134.
12. Binnington, K. C. & Hoffmann, A. A. (1989) *J. Invertebr. Pathol.* **54,** 344–352.
13. Louis, C. & Nigro, L. (1989) *J. Invertebr. Pathol.* **54,** 39–44.
14. Curtis, C. F., Brooks, G. D., Ansari, M. A., Grover, K. K., Krishnamurthy, B. S., Rajagopalan, P. K., Sharma, L. S., Singh, D., Singh, K. R. P. & Yasuno, M. (1982) *Entomol. Exp. Appl.* **31,** 181–190.
15. Laven, H. (1967) *Nature (London)* **216,** 383–384.
16. Rao, T. R. (1974) *J. Commun. Dis.* **6,** 57–72.
17. Curtis, C. F. (1976) *Bull. W.H.O.* **53,** 107–119.
18. Hertig, M. (1936) *Parasitology* **28,** 453–490.
19. Kreig, N. R. & Holt, J. G. (1984) *Bergey's Manual of Systematic Bacteriology* (Williams & Wilkins, Baltimore).
20. Woese, C. R. (1987) *Microbiol. Rev.* **51,** 221–271.
21. Weisburg, W. G., Dobson, M. E., Samuel, J. E., Dasch, G. A., Mallavia, L. P., Baca, O., Mandelco, L., Sechrest, J. E., Weiss, E. & Woese, C. R. (1989) *J. Bacteriol.* **171,** 4202–4206.
22. Weisburg, W. G., Barnes, S. M., Pelletier, D. A. & Lane, D. J. (1991) *J. Bacteriol.* **173,** 697–703.
23. Olsen, G. J., Larsen, N. & Woese, C. R. (1991) *Nucleic Acids Res.* **19,** 2017–2021.
24. Swofford, D. L. (1990) PAUP: Phylogenetic Analysis Using Parsimony, Version 3.0. (Illinois Natural History Survey, Champaign).
25. Simon, C., Franke, A. & Martin, A. (1991) in *Molecular Techniques in Taxonomy*, eds. Hewitt, G. M., Johnston, A. W. B. & Young, J. P. W. (Springer, Berlin), pp. 329–355.
26. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239,** 487–491.
27. Erlich, H. A., Gelfand, D. & Sninsky, J. J. (1991) *Science* **252,** 1643–1651.
28. Dryden, S. C. & Kaplan, S. (1990) *Nucleic Acids Res.* **18,** 7267–7277.
29. Stackebrandt, E., Murray, R. G. E. & Trüper, H. G. (1988) *Int. J. Syst. Bacteriol.* **38,** 321–325.
30. Yang, D., Oyaizu, H., Olsen, G. J. & Woese, C. R. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4442–4447.
31. Murray, R. G. E., Brenner, D. J., Colwell, R. R., de Vos, P., Goodfellow, M., Grimont, P. A. D., Pfennig, N., Stackebrandt, E. & Zavarzin, G. A. (1990) *Int. J. Syst. Bacteriol.* **40,** 213–215.
32. Ash, C., Farrow, J. A. E., Dorsch, M., Stackebrandt, E. & Collins, M. D. (1991) *Int. J. Syst. Bacteriol.* **41,** 343–346.
33. Lawson, P. A., Gharbia, S. E., Shah, H. N., Clark, D. R. & Collins, M. D. (1991) *Int. J. Syst. Bacteriol.* **41,** 347–354.
34. Leu, S. C., Li, J. K. & Hsiao, T. H. (1989) *J. Invertebr. Pathol.* **54,** 248–259.
35. Breeuwer, J. A. J., Stouthamer, R., Barns, S. M., Pelletier, D. A., Weisburg, W. G. & Werren, J. H. (1992) *Insect Mol. Biol.* **1,** in press.