Molecular Evolution and Phylogenetic Utility of *Wolbachia ftsZ* and *wsp* Gene Sequences with Special Reference to the Origin of Male-Killing

J. Hinrich G. v. d. Schulenburg,[*1] *Gregory D. D. Hurst,*† *Ties M. E. Huigens,*‡ *Marnix M. M. van Meer,*‡ *Francis M. Jiggins,** *Michael E. N. Majerus**

*Department of Genetics, University of Cambridge, Cambridge, England; and †Department of Biology, University College London, London, England; and ‡Department of Plant Science, Laboratory of Entomology, Wageningen Agricultural University, Wageningen, the Netherlands

A detailed assessment of the evolution and phylogenetic utility of two genes, *ftsZ* and *wsp*, was used to investigate the origin of male-killing *Wolbachia*, previously isolated from the ladybird *Adalia bipunctata* and the butterfly *Acraea encedon*. The analysis included almost all available sequences of B-group *Wolbachia* and two outgroup taxa and showed that (1) the two gene regions differ in phylogenetic utility, (2) sequence variation is here correlated with phylogenetic information content, (3) both genes show significant rate heterogeneity between lineages, (4) increased substitution rates are associated with homoplasy in the data, (5) *wsp* sequences of some taxa appear to be subject to positive selection, and (6) only a limited number of clades can be inferred with confidence due to either lack of phylogenetic information or the presence of homoplasy. With respect to the evolution of male-killing, the two genes nevertheless seemed to provide unbiased information. However, they consistently produce contradictory results. Current data therefore do not permit clarification of the origin of this behavior. In addition, *A. bipunctata* was found to be a host to two recently diverged strains of male-killing *Wolbachia* that showed increased substitution rates for both genes. Moreover, the *wsp* gene, which codes for an outer membrane protein, was found to be subject to positive selection in these taxa. These findings were postulated to be the product of high selection pressures due to antagonistic host-symbiont interactions in this ladybird species. In conclusion, our study demonstrates that the results of a detailed phylogenetic analysis, including characterization of the limitations of such an approach, can serve as a valuable basis for an understanding of the evolution of *Wolbachia* bacteria. Moreover, particular features of gene evolution, such as elevated substitution rates or the presence of positive selection, may provide information about the dynamics of *Wolbachia*-host associations.

Introduction

The genus *Wolbachia* constitutes a group of intracellular and maternally inherited bacteria from a large variety of arthropod and some nematode hosts (Werren 1997; Bandi et al. 1998). These bacteria have attracted scientific interest due to their ability to manipulate host reproduction, leading to distinct phenotypic effects in the host, such as cytoplasmic incompatibility (Hoffmann and Turelli 1997), induction of parthenogenesis (Stouthamer 1997), feminization of genetic males (Rigaud 1997), and male-killing (Hurst et al. 1999*a*). In general, the different *Wolbachia*-host systems seem to cover the whole spectrum of symbiotic interactions, including mutualism, commensalism, and parasitism (e.g., Werren 1997; Bourtzis et al. 1998; Mercot and Poinsot 1998; Vavre, Girin, and Bouletreau 1999). These associations therefore represent a valuable source for our general understanding of the dynamics and evolution of symbiosis. In addition to detailed information on the factors which characterize these associations, the utilization of this source requires knowledge of the evolutionary history of different *Wolbachia* behaviors. This knowledge may be derived from knowledge of the phylogeny of these organisms. Various studies have employed such a comparative approach, using phylogenetic analysis of bacterial DNA sequences such as those of the 16S ribosomal RNA and the *ftsZ* and *wsp* genes (e.g., Werren, Zhang, and Guo 1995; Bandi et al. 1998; Bouchon, Rigaud, and Juchault 1998; Zhou, Rousset, and O'Neill 1998; Van Meer, Witteveldt, and Stouthamer 1999). However, results obtained have shown inconsistencies. Although the data all provide support for four major clades, *Wolbachia* groups A–D, the phylogenetic relationships within these groups, particularly within groups A and B, have shown differences when different genes have been analyzed. The reliability of the different molecular markers therefore requires assessment.

Various sources of error have been described for phylogenetic analysis of DNA sequences, even in cases in which the data can be assumed to be infinite and character homology can be ascertained. Extremely low or high substitution rates might result in paucity of usable information or high levels of homoplasy, respectively. In both cases, tree reconstruction methods, however sophisticated, are unlikely to uncover the underlying phylogenetic signal (e.g., Wägele 1996; Yang 1998*a*). Furthermore, systematic errors are introduced if the assumed substitution model of the tree reconstruction method is not consistent with the evolutionary dynamics of the analyzed sequences. This can be avoided through characterization of the factors that shape the evolution of the studied sequences and subsequent incorporation of this information into substitution models. For instance, consideration of the presence and extent

[1] Present address: Abteilung für Evolutionsbiologie, Institut für Spezielle Zoologie, Westfälische Wilhelms-Universität Münster, Münster, Germany.

Key words: *Wolbachia*, male-killing, phylogenetic analysis, substitution rate heterogeneity, positive selection.

Address for correspondence and reprints: J. Hinrich G. v. d. Schulenburg, Abteilung für Evolutionsbiologie, Institut für Spezielle Zoologie, Westfälische Wilhelms-Universität Münster, Hüfferstrasse 1, 48149 Münster, Germany. E-mail: hschulen@uni-muenster.de.

Mol. Biol. Evol. 17(4):584–600. 2000

of substitution rate heterogeneity across sites can significantly improve the outcome of phylogenetic analysis (Yang 1996). However, it is not always possible to accommodate the various factors in the available tree reconstruction methods. Base composition and synonymous codon usage biases between evolutionary lineages produce hierarchical structure in the data which is unrelated to phylogeny and might thus be positively misleading during tree reconstruction (e.g., Lockhart et al. 1994; He and Haymer 1995; Foster and Hickey 1999). Consideration of among-site and between-lineage variation of selectional constraints as defined by nonsynonymous/synonymous substitution (d_N/d_S) ratios can significantly increase the maximum likelihood of a given tree topology, indicating their importance in phylogenetic analysis (Yang 1998*b*). In addition, substitution rate heterogeneity between lineages, although taken into account in most tree reconstruction methods, may still lead to ''long branch attraction'' (Felsenstein 1978; Hendy and Penny 1989). These examples illustrate the dilemma of molecular systematics: the evolutionary dynamics of the studied sequences are likely to affect the results of phylogenetic analyses regardless of the method used. It is thus reasonable to conclude that it is the utility of the molecular marker to meet the assumptions of the available tree reconstruction methods which ascertains unambiguous inference of the phylogeny of the studied organisms.

Our study, therefore, focuses on a detailed exploration of the evolution of two genes, *ftsZ* and *wsp*, and their utility for phylogenetic analysis of *Wolbachia* bacteria. We give special consideration to the origin of male-killing *Wolbachia* which have recently been isolated from a beetle, *Adalia bipunctata* (Coccinellidae), and a butterfly, *Acraea encedon* (Nymphalidae). Phylogenetic analysis of *ftsZ* and *wsp* sequences indicated that these taxa belong to different evolutionary lineages within the B-group *Wolbachia* clade. However, monophyly of these genes could not be rejected with absolute certainty (Hurst et al. 1999*a*). Using an extended data set for both genes, we here reconsider this problem. First, a detailed assessment of the phylogenetic utility of the two genes is attempted via characterization of phylogenetic information content and signal consistency and of the evolution of the genes with special reference to putatively inference-biasing factors. Second, phylogenetic tree reconstruction is performed in consideration of this information. Third, monophyly of male-killing is specifically tested with the method of Kishino and Hasegawa (1989) (the KH test). Finally, the results obtained are discussed with respect to the utility of these genes for the inference of the evolution of *Wolbachia.*

Materials and Methods

Almost all available and unique *ftsZ* and *wsp* sequences of B-group *Wolbachia* and two outgroup taxa of different A-group lineages were included in the analysis (table 1).

DNA sequences were aligned by eye, taking into account the coding structure of the genes. Two regions of the *wsp* gene could not be aligned with confidence (alignment positions 217–255 and 523–585), even when translated amino acid sequences were considered. As positional homology of the aligned nucleotides is essential for correct phylogeny reconstruction (e.g., Cerchio and Tucker 1998), these ambiguous positions were excluded from subsequent analyses. The complete sequence alignments have been deposited in the EMBL alignment database under accession numbers DS38176 and DS40027.

Assessment of phylogenetic utility was mainly concerned with a characterization of differences between taxa. The first part involved analysis of the extent and consistency of hierarchical structure in the data as a measure of phylogenetic information content and signal consistency. This was studied via the phenomenological analysis of split-supporting positions (PASSP) as described in Wägele (1996), Wägele and Rödding (1998), and Schulenburg, Englisch, and Wägele (1999), using a preliminary version of the program PHYSID (available from J. W. Wägele, e-mail: johann.w.waegele@rz.ruhr-uni-bochum.de). This method relies on one of the basic concepts of phylogenetic systematics *sensu* Hennig (1966), which predicts that members of a monophylum exclusively share a number of identical nucleotide states due to identity by descent. Three categories of character state distributions in support of a putative monophylum can be discerned. Symmetrical split-supporting positions are defined to show two different nucleotide states at a given alignment position, one for the functional ingroup and a different one for the corresponding outgroup. Asymmetrical split-supporting positions bear one nucleotide state in the functional ingroup and at least two in the corresponding outgroup which all differ from that of the functional ingroup. ''Noisy'' asymmetrical split-supporting positions are additionally allowed to show a defined degree of analogies between ingroup and outgroup. An analysis of such character state distributions along the sequence alignment consequently allows identification of putatively monophyletic groups without any further transformation of the data. It needs to be emphasized that this method cannot replace phylogenetic tree estimation based on the reconstruction of the dynamics of sequence evolution. However, it does permit characterization of simple character state distributions regardless of assumptions about the evolution of the studied sequences.

For our analysis, the numbers of split-supporting positions were calculated for both groups of a particular split and subsequently set against each other, resulting in a spectrum of split-supporting positions. The permitted level of analogies for noisy positions was set at 25% per alignment position and taxon for both groups of a particular split. The group indicated by a larger number of supporting positions was considered the functional ingroup. Phylogenetic information content was then assessed from the number of supported groups and the extent to which these groups were supported. Phylogenetic signal consistency was indicated by the compatibility of the supported clades. In addition, an analysis of split compatibility was used to identify noisy se-

Table 1
***Wolbachia* Strains Used for Phylogenetic Analysis**

ftsZ Gene		*wsp* Gene	
Taxon[a]	GenBank[b]	Taxon[a]	GenBank[b]
Acraea encedon	AJ130892	*A. encedon*	AJ130716
Adalia bipunctata	AJ130717	*A. bipunctata*, strain Y[c]	AJ131714
Aedes albopictus	U28206	*A. bipunctata*, strain Z[c]	AJ130715
Aramigus tesselatus	U28193	*A. albopictus*	AF020059
Armadillidium vulgare	U28208	*Apoanagyrus diversicornis*	AF071916
Cadra cautella	U28207	*A. vulgare*, strain Portugal (P)	AJ269474
Culex pipiens	U28209	*A. vulgare*, strain Wageningen (W)	AF071917
Drosophila melanogaster, strain Canton S	X71906	*C. cautella*	AF020076
Encarsia formosa	U28196	*C. pipiens*	AF020061
Gryllus integer	U83105	*Diplolepis rosae*	AF071922
Gryllus pennsylvanicus, strain Forrest Home (FH)	U28195	*D. melanogaster*, strain Canton S	AF020065
G. pennsylvanicus, strain Fairfax Virginia (FV)	U83099	*Drosophila mauritiana*	AF020069
Leptopilina australis	U28210	*Drosophila sechellia*	AF020073
Nasonia giraulti	U28203	*E. formosa*	AF071918
Nasonia longicornis	U28204	*Eretmocerus staufferi*	AF071919
Nasonia vitripennis	U28205	*Laodelphax striatellus*	AF020080
Protocalliphora sp.	U28202	*L. australis*	AF071920
Sitophilus oryzae	U28197	*Oniscus asellus*	AJ269475
Tribolium confusum (1)	U28194	*Porcellio scaber*	AJ269476
T. confusum (2)	U97351	*Torymus bedeguaris*	AF071915
Trichogramma brevicapillum	U28198	*T. confusum*	AF020083
Trichogramma cordubensis	U28200	*T. deion* (1)	AF020084
Trichogramma deion (1)	U59696	*T. deion*, strain SW436 (SW)	AF071925
T. deion, strain Mt	U28199	*Trichogramma kaykai*, strain JT6-3 (JT)	AF071924
T. deion, strain TX	U28201	*T. kaykai*, strain LC110 (LCB)	AF071927
Trichopria drosophilae	U28190	*Trichogramma nubilale*	AF071926
		Trichogramma sibericum	AF071923
		Tagosedes orizicolus	AF020085

[a] *Wolbachia* are represented by their respective host species.

[b] GenBank accession number.

[c] The *A. bipunctata*–*Wolbachia* strains A and B (Hurst et al. 1999*a*) were renamed Y and Z, respectively, to avoid confusion with nomenclature for *Wolbachia* subgroups.

quences. Each supported group was compared with all other identified clades. For all pairwise comparisons which were incompatible, we identified the minimum number of taxa which had to be excluded from one of the clades to restore compatibility. Following the parsimony criterion, sequences of these taxa were assumed to represent a source of homoplasy in the data.

The second part of the phylogenetic utility assessment consisted of a characterization of evolutionary processes which may produce hierarchical structure in the data unrelated to organismal phylogeny and may thus bias tree estimation. In particular, evidence of substitution rate heterogeneity between lineages was sought, and the presence of selectional constraints was tested, as identified by base composition, synonymous codon usage, and d_N/d_S ratio biases among sequences.

The presence of rate heterogeneity between evolutionary lineages was analyzed using the two-cluster test as implemented in the program LINTRE (Takezaki, Rzhetsky, and Nei 1995) and a likelihood ratio test as suggested by Felsenstein (1981), and Yang, Goldman, and Friday (1995). For the two-cluster test, neighbor-joining tree reconstruction and statistical analysis were performed assuming the substitution model of Tamura and Nei (1993) with rate heterogeneity across sites. The α shape parameter of rate heterogeneity across sites used was estimated with the program PUZZLE, version 4.0 (Strimmer and von Haeseler 1996). For the second approach, PUZZLE was employed to estimate trees with and without the assumption of a molecular clock using the HKY85 substitution model (Hasegawa, Kishino, and Yano 1985) and consideration of eight across-site rate heterogeneity categories. Likelihood scores for these trees were subsequently compared with the likelihood ratio test. If the likelihood of the clocklike tree was significantly worse, the minimum number of taxa which had to be excluded from the data to obtain insignificant likelihood differences was determined. For this test, we specifically excluded taxa with comparatively long branch lengths in the inferred maximum-likelihood (ML) tree to identify sequences which may evolve with increased rates.

Homogeneity of base composition was tested with a standard χ^2 test. Analysis was performed on complete sets and subsets of the data. The latter included averaged values for each well-supported clade (bootstrap values $\geq$50) of the later inferred ML trees to take account of a possible phylogenetic bias inherent in the complete data sets. Synonymous codon usage bias was calculated with the program MEA (Moriyama 1998) using the effective number of codons (ENC) (Wright 1990), which has been shown to produce less biased results than alternative methods (e.g., Comeron and Aguade 1998). As ENC values are correlated with the GC content at silent third codon positions, deviations in the usage of synonymous codons between taxa were visualized via the

Nc-plot as suggested by Wright (1990). d_N/d_S ratios were inferred for all pairwise comparisons with the method of Nei and Gojobori (1986) (the NG method) using the program MEGA (Kumar, Tamura, and Nei 1993). Alignment gaps were excluded in pairwise comparisons only. Variation in d_N/d_S ratios (particularly if it included sequences under diversifying selection; d_N/d_S ratios >1) were here considered to represent a putative source of hierarchical structure in the data which may confound phylogenetic analysis.

In the final part of the phylogenetic utility assessment, the presence of variation across sites was analyzed for a selected number of factors. We specifically looked at phylogenetic information content in comparison to the distribution of nucleotide variation. As the true phylogeny of B-group *Wolbachia* included in our study must be considered unknown, it is not possible to determine which alignment positions contain unbiased or misleading phylogenetic information. We therefore employed PASSP to identify those positions which exclusively supported the partition between A- and B-group *Wolbachia* and the splits that were compatible with all other indicated partitions. The distribution of such putatively informative positions then served to test whether distinct and consistent phylogenetic information was restricted to particular regions within the genes or evenly present throughout the whole sequence alignments. Under the assumption of no bias across sites, the distribution of putatively informative positions should be correlated with that of the variable sites, because only variable positions contribute to the hierarchical structure of the data and thus directly provide phylogenetic information. Hence, the number of putatively informative and variable positions was calculated along the genes using a sliding-window approach. Correlation probabilities between the respective distributions were then estimated from nonoverlapping windows, which can be assumed to be mutually independent, by applying the z* transformation for dealing with small sample size ($n < 50$; Sokal and Rohlf 1995, chapter 15). Values for overlapping windows (shifted by one position at a time) served to graphically illustrate the resulting profiles.

In addition, as the *wsp* gene codes for an outer membrane protein of which some sites may be directly involved in symbiont-host interactions (see below), differences in d_N and d_S rates were assessed along *wsp* sequences to test whether specific regions evolve under positive selection. Rate variation was estimated with the NG method, as before, and a sliding-window following the procedures used by Alvarez-Valin, Jabbari, and Bernardi (1998). As rates calculated from significantly fewer than 30 codons may be unreliable due to high stochastic errors, and as the *wsp* sequence alignment contained sequences with continuous gaps of up to five codons, averaged d_N and d_S rates were obtained for all pairwise comparisons using a window size of 90 bp. Each comparison thus always included at least 25, and in almost all cases exactly 30, codons. Such a strategy only produces values for less than six independent (=nonoverlapping) windows because *wsp* sequences comprise a maximum of only 171 codons. We therefore did not attempt to calculate correlation probabilities, but instead obtained profiles from overlapping windows (shifted by one codon at a time) for a graphical illustration of rate variation across sites.

The aligned sequences were subsequently subjected to phylogenetic analysis using the program PAUP*, versions 4.0.0d55 and 4.0b1, written by D. L. Swofford. Tree estimation was attempted with the ML method. For both data sets, overall base frequencies are not equal and substitution rates vary across sites (see *Results*). In addition, a preliminary ML analysis with the program PUZZLE indicated that the transition/transversion (ts/tv) ratio is clearly different from 1. For ML estimation, we therefore employed the HKY85 model or the general time reversible (GTR) model (e.g., Lanave et al. 1984) with consideration of gamma-distributed rate heterogeneity across sites (Γ) and, in addition, a proportion of invariable positions ($\Gamma+I$). In each case, all model parameters were estimated from the data using a starting tree topology which had previously been inferred with unweighted maximum parsimony (U-MP). These parameter estimates and the starting tree topology were then employed for a heuristic search based on branch-swapping by nearest-neighbor interchanges (NNIs) and tree bisection and reconnection (TBR). In addition, for phylogeny reconstruction, we employed U-MP to assess the effects of gaps, which cannot be taken into account during ML analysis. Two extremes were compared for which each gap at a particular alignment position was treated as either missing data or a fifth base. U-MP was performed using a heuristic search based on the NNI branch-swapping algorithm. Robustness of the resulting tree topologies was assessed via nonparametric bootstrapping (Felsenstein 1985) using the same settings as above and 100 and 1,000 replicates for ML and U-MP analyses, respectively.

Monophyly of male-killing *Wolbachia* was specifically tested via the KH test (Kishino and Hasegawa 1989) as implemented in PAUP*. For both data sets, we used ML, as outlined above, to estimate trees with the topological constraint of monophyletic male-killing *Wolbachia*. The KH test was thereafter employed to compare the likelihood scores obtained with those of the optimal ML trees. For the *wsp* gene only, the same strategy was used to test the monophyly of the *Wolbachia* from *A. encedon* and *Aedes albopictus,* which was well supported by *ftsZ* sequences (see below).

Results and Discussion

General

The *ftsZ* data set consists of 26 sequences and 969 alignment positions, of which 183 show variability (18.89%). Between 94 and 108 nucleotide differences are found in pairwise comparisons between B-group *Wolbachia* and the outgroup (10.08%–11.15%), whereas 1–57 positions differ within the B group (0.11%–6.03%). Twenty-eight sequences are included in the *wsp* sequence alignment, which contains 515 positions, with 231 of these being variable (44.85%). Here, nucleotide differences are observed at 72–108 positions between

B-group and outgroup taxa (13.98%–20.97%) and at 1–109 positions within the B group (0.19%–21.17%). The *wsp* gene thus shows about twice as much variability as the *ftsZ* gene. In addition, maximum *ftsZ* sequence divergence within the B-group *Wolbachia* clade is much lower than that between the B-group and the two outgroup taxa. In contrast, these values are approximately equal for the *wsp* gene, indicating substitutional saturation between A- and B-group *wsp* sequences.

Phylogenetic Information Content and Signal Consistency

Twelve putatively monophyletic groups are indicated by at least four split-supporting positions for the *ftsZ* data set (fig. 1*A*). One group is identified by a much larger number of positions than all other groups (about 10% of the alignment positions) and separates *Wolbachia* taxa of the A and B groups (split 1). Only three groups are supported by more than 1% of the alignment positions (splits 1–3). In addition, splits 2, 9, and 12 show compatibility with all other indicated clades. The three best supported partitions (splits 1, 2, and 3) are compatible with each other and with splits 4, 7, 9, and 12 (fig. 1*A* and table 2). An analysis of the compatibility between supported groups indicates that *ftsZ* sequences of some *Wolbachia* represent putative sources of homoplasy. In particular, that of *Gryllus pennsylvanicus* FV has to be excluded most often from pairwise comparisons to restore compatibility. Phylogenetic signal inconsistency is additionally associated with *Wolbachia* from *Aramigus tesselatus, A. bipunctata, Gryllus integer, Trichogramma* sp., and the two A-group taxa (table 2).

In comparison to the *ftsZ* data set, *wsp* sequences support a larger number of clades (fig. 1*B*). In total, 22 partitions are identified by at least four split-supporting positions (0.78% of the alignment positions). Eight of these are supported by more than 1% of the alignment sites (splits 1–7 and 14). The largest number of split-supporting positions are found for the two A-group *Wolbachia* and thus indicates the partition which separates taxa of *Wolbachia* groups A and B (split 1), in agreement with results inferred from *ftsZ* sequences. In addition, the *wsp* data set supports seven splits which are compatible with all other identified clades (splits 2–4, 11, 14, 15, and 19). The largest set of groups which are intercompatible additionally includes splits 1, 6, 7, 10, and 21. Incompatibility between clades can most often be reversed by exclusion of *wsp* sequences of the two *Wolbachia* from *Armadillidium vulgare.* Other putative sources of phylogenetic signal inconsistency are the sequences of the A-group *Wolbachia,* those of *Apoanagyrus diversicornis, Encarsia formosa,* and possibly those of *Porcellio scaber, Oniscus asellus,* and *Trichogramma* sp. (table 3).

For both data sets, the outgroup is indicated by a larger number of positions than the B-group *Wolbachia* (split 1 for both data sets), although support for both should be the same. This is likely to be due to the fact that the B-group *Wolbachia* clade contains a much larger number of taxa than the outgroup, such that independent substitutions in the various included B-group lineages have masked the original signal for this clade. This effect might be particularly pronounced for the *wsp* data set, for which sequence divergence among B-group taxa is high. However, both data sets also support groups which include outgroup taxa and various B-group *Wolbachia* (split 4 for the *ftsZ* gene; splits 8, 12, and 18 for the *wsp* data set). Although these two observations taken together suggest that A-group *Wolbachia* have originated within the B-group clade, phylogenetic analysis of 16S rDNA or *ftsZ* sequences of a larger variety of taxa uniformly show A- and B-group *Wolbachia* to represent monophyletic sister groups (Bandi et al. 1998; unpublished data). The outgroup and some B-group taxa are thus likely to share homoplasious character states, as already indicated from the analysis of incompatible splits. Such homoplasies could result from analogous character state evolution in the respective lineages or represent ancestral states which have been exclusively retained in the taxa of the indicated clade. In the latter case, plesiomorphies could have been lost once or multiple times in the remaining taxa of the B group, indicating that these taxa are either monophyletic or polyphyletic, respectively. Such alternatives may be discerned in consideration of the results of phylogenetic tree reconstruction.

In conclusion, the analysis of split-supporting positions in the two gene regions suggests that the general level of sequence variation is correlated with the level of phylogenetic information content, as *wsp* sequences provide a higher proportion of support for a larger number of splits than the *ftsZ* gene. For both data sets, phylogenetic signal inconsistency is associated with roughly the same number of taxa.

Substitution Rate Heterogeneity Between Lineages

For the *ftsZ* gene, two clades are identified by the two-cluster test to show a significantly increased substitution rate. The first consists of *Wolbachia* from *Trichogramma* sp. ($P < 0.01$), while the second also includes that from *A. vulgare* ($P < 0.05$). As the tree topology, inferred with the distance-based method in LINTRE, shows differences from the later-estimated ML tree, the two-cluster test was repeated, using the ML topology. A significantly increased substitution rate is here indicated for three clusters, as illustrated in figure 4*A*. The likelihood ratio test confirms that *ftsZ* sequences do not evolve with a uniform substitution rate because the tree estimated with the assumption of a molecular clock produces a significantly smaller likelihood ($2\Delta l = 79.16$, $P_{\nu=24} < 0.01$). Trees calculated with or without the molecular-clock assumption produce no significant differences if any of the following three sets of taxa are excluded: *Wolbachia* from *G. pennsylvanicus* FV, *G. integer*, and *Trichogramma* sp. ($2\Delta l = 22.93$, $P_{\nu=17} = 0.152$); those from *G. pennsylvanicus* FV, *Trichogramma* sp., and *A. bipunctata* ($2\Delta l = 26.13$, $P_{\nu=17} = 0.072$); or those from *G. pennsylvanicus* FV, *Trichogramma* sp., and *A. tesselatus* ($2\Delta l = 23.20$, $P_{\nu=17} = 0.143$) (fig. 4*A*).

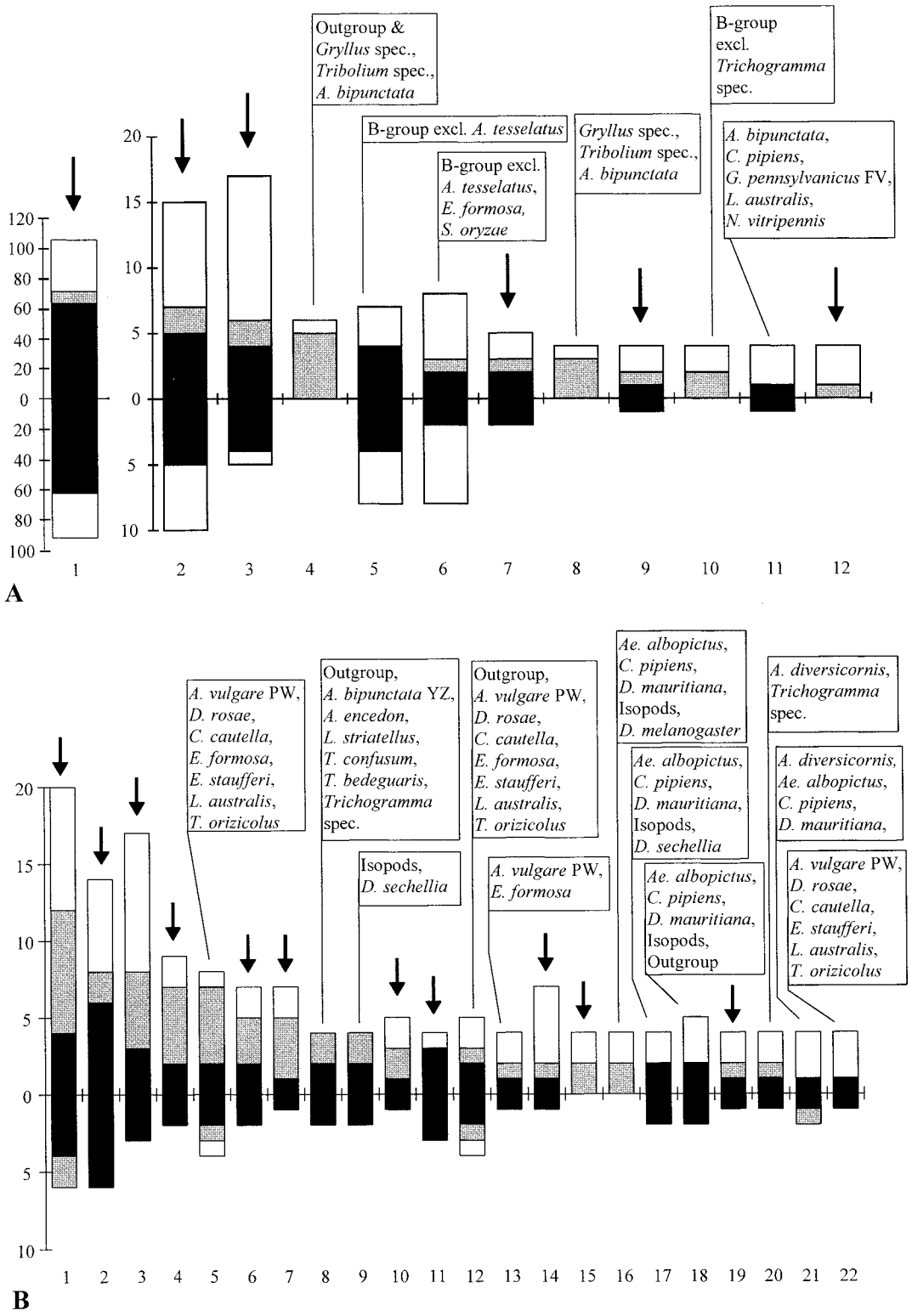

FIG. 1.—Spectrum of split-supporting positions for (*A*) the *ftsZ* gene and (*B*) the *wsp* gene, as identified with the program PHYSID (Wägele 1996; Wägele and Rödding 1998). The X-axis denotes the split between ingroup and outgroup. Column height corresponds to the absolute number of split-supporting positions. The black, gray, and white parts of columns indicate symmetrical, asymmetrical, and 'noisy' split-supporting positions, respectively. Numbers below columns denote partitions which are supported by at least four split-supporting positions. Arrows point to ingroups which are compatible with the maximum-likelihood (ML) tree as indicated in figure 4. Taxa for ingroups which are incompatible with the ML tree are given above columns (*Wolbachia* are represented by their respective host species).

Table 2
Incompatible Splits and the Minimum Number of Taxa Excluded to Restore Compatibility for the *ftsZ* Gene Data Set

Incompatible Splits[a]	Taxa Excluded[b]
3 ↔ 6; 3 ↔ 8; 3 ↔ 11; 7 ↔ 11; 8 ↔ 11	*Gryllus pennsylvanicus* FV
3 ↔ 5	*G. pennsylvanicus* FV or *Aramigus tesselatus*
3 ↔ 10	*G. pennsylvanicus* FV or *Trichogramma* sp.[c]
4 ↔ 11	*G. pennsylvanicus* FV and *Adalia bipunctata*
6 ↔ 7; 7 ↔ 8	*G. pennsylvanicus* FV and *Gryllus integer*
7 ↔ 10	*G. pennsylvanicus* FV and *G. integer* or *Trichogramma* sp.
4 ↔ 5; 4 ↔ 6; 4 ↔ 10	A-group taxa
5 ↔ 7	*A. tesselatus*
5 ↔ 10	*A. tesselatus* or *Trichogramma* sp.
6 ↔ 10	*Trichogramma* sp.

[a] Split numbers are identical to those given in figure 1*A*.
[b] Minimum number of taxa which need to be excluded from one of the pairwise compared splits to restore compatibility.
[c] The *Wolbachia* of the different *Trichogramma* species were considered as one entity, as their sequences share a high degree of similarity.

For the *wsp* data set, the two-cluster test was again performed twice, using the distance-based and the later-inferred ML tree topologies. Increased substitution rates are consistently identified for the following seven lineages: the *Wolbachia* from *E. formosa*; from *P. scaber*; from the isopod species (*A. vulgare* PW, *O. asellus, P. scaber*); from the isopod species and *Drosophila mauritiana, A. albopictus,* and *Culex pipiens*; from *A. bipunctata*; from *Trichogramma deion* (1); and from *Trichogramma* sp. (fig. 4*B*). Other clades are indicated for only one of the topologies, and these clades always include some of the above-mentioned taxa. Differences in the results here correlate with differences between the two tree topologies. However, not all lineages within these clades are likely to evolve with an elevated substitution rate, as is apparent from branch lengths in the ML tree (fig. 4*B*). The outcome of the two-cluster test thus seems to be biased by lineages with unusually high substitution rates, such that clades which include both lineages with unusually high rates and lineages with normal or decreased rates may still be identified to evolve with significantly increased substitution rates. Such a bias may also have been produced for the *ftsZ* data set, for which sequences of the symbionts from *Tribolium confusum* or *A. vulgare* are associated with comparatively short branches but are in each case indicated in combination with those from *Trichogramma* sp. to show elevated substitution rates (see above and fig. 4*A*).

The likelihood ratio test provides additional evidence for *wsp* substitution rate heterogeneity between lineages, as the likelihood of the clocklike tree is significantly worse ($2\Delta l = 128.02$, $P_{\nu=26} < 0.01$). The following minimum number of taxa have to be excluded from the data set to obtain likelihood scores that are no

Table 3
Incompatible Splits and the Minimum Numbers of Taxa Excluded to Restore Compatibility for the *wsp* Gene Data Set

Incompatible Splits[a]	Taxa Excluded
5 ↔ 7; 5 ↔ 9; 5 ↔ 16; 5 ↔ 17; 5 ↔ 18; 7↔ 12; 7 ↔ 22; 9 ↔ 22; 16 ↔ 22; 17 ↔ 22; 18 ↔ 22	*Armadillidium vulgare* PW[b]
7 ↔ 13; 10 ↔ 13; 10 ↔ 22; 13 ↔ 16; 13 ↔ 17; 13 ↔ 18; 9 ↔ 13; 13 ↔ 22	*A. vulgare* PW or *Encarsia formosa*
12 ↔ 16	*A. vulgare* PW and *Drosophila melanogaster* CS
12 ↔ 17	*A. vulgare* PW and *Drosophila sechellia*
12 ↔ 18	*A. vulgare* PW and A-group taxa
9 ↔ 12	*A. vulgare* PW and *D. sechellia* or *Porcellio scaber* and *Oniscus asellus*
1 ↔ 9; 1 ↔ 17; 8 ↔ 9; 8 ↔ 17; 9 ↔ 16	*D. sechellia*
1 ↔ 16; 8 ↔ 16	*D. melanogaster* CS
8 ↔ 12; 8 ↔ 18	A-group taxa
16 ↔ 17	*D. melanogaster* CS or *D. sechellia*
20 ↔ 21	*Apoanagyrus diversicornis*
6 ↔ 20; 8 ↔ 20	*A. diversicornis* or *Trichogramma* sp.[b]

[a] Split numbers are identical to those given in figure 1*B*.
[b] The different *Wolbachia* of *A. vulgare* (*A. vulgare* PW) and also of the *Trichogramma* species were treated as one entity. Otherwise, see comments in table 2.

longer significantly different between trees estimated with or without the molecular clock: *Wolbachia* from *A. bipunctata*, *P. scaber*, *Trichogramma* sp., and *A. diversicornis* ($2\Delta l = 20.48$, $P_{\nu=16} = 0.199$). Moreover, if the *Wolbachia* from *E. formosa* or *A. vulgare* are excluded instead of that from *A. diversicornis*, then the trees are still significantly different at the 5% level ($2\Delta l = 30.19$, $P_{\nu=16} = 0.017$; and $2\Delta l = 29.04$, $P_{\nu=15} = 0.016$, respectively) (fig. 4*B*).

In spite of some differences in the results obtained, the two methods clearly indicate that both genes show significant substitution rate heterogeneity between lineages, with some of them being consistently identified in all tests to be subject to accelerated substitution rates. In this context, it is interesting to note that for both genes, all B-group *Wolbachia* sequences that represent putative sources of homoplasy are also found to show increased substitution rates (tables 2 and 3 and fig. 4). This suggests that increased substitution rates in particular lineages have given rise to homoplasy.

Presence of Selectional Constraints

DNA sequences of B-group *Wolbachia* show little variation in overall base composition for the *ftsZ* (32.9%–34.2% A, 13.7%–15.4% C, 23.8%–25.6% G, 25.5%–28.3% T) and the *wsp* gene (27.0%–30.6% A, 13.1%–16.1% C, 21.1%–23.5% G, 32.1%–35.2% T). Base composition heterogeneity is not significant for these genes, either if complete data sets are considered or if subsets of the data are analyzed which include averaged values for the well-supported clades of the later-inferred ML trees (results not shown). Synonymous codon usage bias is more variable between taxa for the *wsp* gene than for the *ftsZ* gene. In particular, *wsp* sequences of the symbionts from *A. encedon* and *Laodelphax striatellus* produce comparatively small ENC values (37.81 and 37.72, respectively), whereas the ENC value for *P. scaber* is comparatively high (50.45). Nevertheless, both genes show variation in synonymous codon usage which is close to a distribution of no selectional constraints (fig. 2). In addition, all inferred ENC values are higher than those observed for strongly selected genes in previous studies (e.g., Wright 1990; Moriyama and Powell 1997). Heterogeneity in base composition or synonymous codon usage between lineages therefore represents an unlikely source of hierarchical structure in the two data sets.

Almost all d_N/d_S ratios for pairwise-compared *ftsZ* sequences were less than 0.4 (322 out of 325 comparisons). Only the comparison between the two *Wolbachia* strains from *T. confusum* produced a d_N/d_S ratio of infinity. As these two sequences only differ at one nonsynonymous position, the estimated d_N/d_S ratio is likely to be unreliable. *FtsZ* sequences consequently do not seem to be subject to positive selection or variation of selectional constraints. In contrast, d_N/d_S ratios inferred for the *wsp* data set show considerable variation. Six of the 378 d_N/d_S ratios calculated have values less than 0.2, 105 were between 0.2 and 0.4, 204 were between 0.4 and 0.6, 41 were between 0.6 and 0.8, 13 were between 0.8 and 1.0, and 9 were greater than 1.0. Note that the NG method represents a conservative approach as to the detection of d_N/d_S ratios >1, as it tends to overestimate d_S and underestimate d_N (e.g., Li 1993). On the other hand, for almost all comparisons with d_N/d_S ratios >1, standard deviations for d_N or d_S are high, indicating that the inferred values are subject to stochastic error (table 4). It must therefore remain uncertain whether a larger number of cases are subject to positive selection or some of the indicated comparisons produce d_N/d_S ratios which are not significantly greater than 1, respectively. The results nevertheless show that *wsp* sequences are generally subject to variation in selectional constraints between lineages, with some of them possibly being under positive selection.

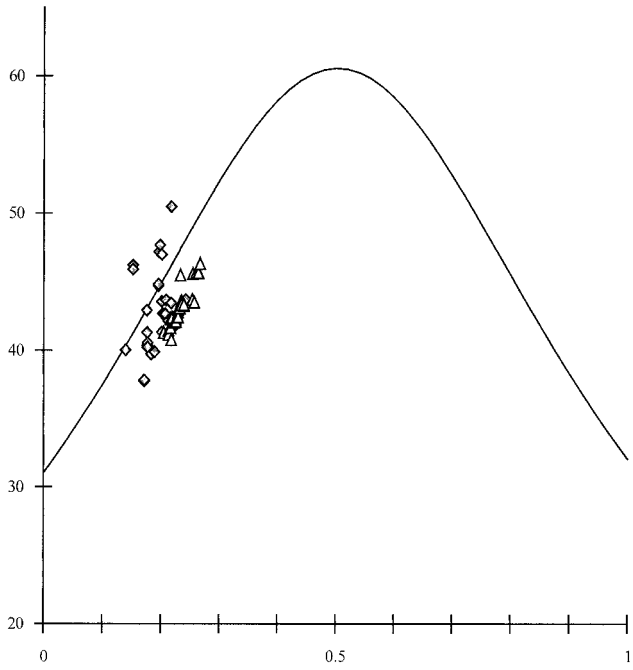

FIG. 2.—Nc plot of synonymous codon usage bias for *ftsZ* (triangles) and *wsp* (diamonds) sequences. The X-axis indicates GC content at silent third codon positions, and the Y-axis indicates ENC values. ENC values were calculated as outlined in Wright (1990) with the program MEA (Moriyama 1998). The continuous curve represents the relationship between ENC and GC content at silent third codon positions under no selectional constraints (Wright 1990).

Variation Across Sites

For a selected number of factors, we assessed the distribution of variation across sites. In particular, an estimate of phylogenetic information along the two genes was obtained via identification of the alignment positions which exclusively provide distinct support for the partition between A- and B-group *Wolbachia* (split 1 for both data sets; fig. 1) and the splits that were compatible with all other indicated partitions (splits 2, 9, and 12 and 2–4, 11, 14, 15, and 19 for the *ftsZ* and *wsp* genes, respectively; fig. 1). The distribution of these positions was compared with nucleotide variation across sites. The results obtained show the numbers of both variable and putatively informative positions to vary

Table 4
Pairwise Comparison of *wsp* Sequences Producing d_N/d_S Ratios >1

Taxa Compared	d_N/d_S[a]	D_N[b]	SD_{d_N}[c]	d_s[d]	SD_{d_S}[c]
Adalia bipunctata Y, *A. bipunctata* Z	4.2970	0.0709	0.0144	0.0165	0.0117
Acraea encedon, Torymus bedeguaris	2.2738	0.0191	0.0072	0.0084	0.0084
Laodelphax striatellus, T. bedeguaris	1.6386	0.0136	0.0061	0.0083	0.0083
Cadra cautella, Eretmocerus staufferi	1.3354	0.0219	0.0078	0.0164	0.0116
C. cautella, Leptopilina australis	1.3354	0.0219	0.0078	0.0164	0.0116
Apoanagyrus diversicornis, Tribolium confusum	1.1581	0.1355	0.0207	0.1170	0.0332
T. bedeguaris, Trichogramma deion (1)	1.1455	0.0748	0.0147	0.0653	0.0241
A. diversicornis, T. bedeguaris	1.0560	0.1376	0.0209	0.1303	0.0353
E. staufferi, L. australis	∞	0.0054	0.0038	0	0

NOTE.—All values are calculated with the method of Nei and Gojobori (1986) using the program MEGA (Kumar, Tamura, and Nei 1993).
[a] Nonsynonymous/synonymous substitution.
[b] Nonsynonymous substitution rate.
[c] Standard deviation.
[d] Synonymous substitution rate.

along *ftsZ* and *wsp* sequences (fig. 3*A* and *B*). Both distributions were significantly correlated in the two genes, irrespective of whether the analysis was based on all ($P < 0.01$ for both data sets), only first and second ($P < 0.01$ for both data sets), or only third codon positions ($P < 0.01$ for the *ftsZ* gene and $0.01 < P < 0.05$ for the *wsp* gene). Therefore, phylogenetic information content appears to represent a direct function of nucleotide variation across sites and should thus be considered unbiased.

It is nevertheless interesting to note that the *wsp* gene contains three highly variable regions, of which only a comparatively small proportion of sites bear distinct phylogenetic information. Intriguingly, variation in these regions is primarily contained in first and second codon positions, suggesting that it is associated with positive selection. For two of these regions, this is confirmed by an analysis of d_N and d_S rate variation across *wsp* sequences. In these cases, average d_N values inferred from all pairwise comparisons were either clearly higher or almost as high as those obtained for d_S (fig. 3*C*). Similar results were produced if the analysis included only data from B-group *Wolbachia* (results not shown) or for those comparisons which generated d_N/d_S ratios >1 for complete *wsp* sequences (fig. 3*D*). In the latter case, d_N rates, in fact, exceeded d_S rates throughout almost the entire length of the *wsp* gene. However, differences between these rates were clearly most pronounced in the two specific regions. In conclusion, two *wsp* gene regions generally appear to be subject to a bias in selectional constraints across *Wolbachia* taxa. Such constraints may cause analogous character state evolution. Although phylogenetic information across sites is, in general, indicated to be unbiased, these two regions can therefore not be excluded to provide a higher proportion of misleading information.

Tree Estimation

Phylogenetic utility assessment indicated that the two genes differed with respect to their overall information content. In addition, they both showed substitution rate heterogeneity between lineages, with increased rates being, in most cases, associated with the presence of homoplasy in the data. Complete *wsp* sequences (as present in the alignment used for phylogenetic analysis) also seemed to be subject to positive selection between some *Wolbachia*. As the presence of increased substitution rates or positive selection may bias phylogenetic tree reconstruction, affected lineages should be excluded to maximize reliability of results. However, this would apply to a rather large number of taxa, including the male-killing *Wolbachia* of *A. bipunctata*, whose origin and evolution is to be investigated. Phylogenetic trees were thus estimated from all sequences included, and the reliability of the position of ''problematic'' taxa was assessed a posteriori. Finally, two regions within the *wsp* gene appeared to be positively selected across lineages. As they may contain a higher proportion of biased phylogenetic information, tree estimation was repeated on *wsp* sequences excluding these regions (positions 70–117 and 331–384 of the alignment used for phylogenetic analysis, corresponding to positions 70–117 and 370–423 of the original alignment submitted to the EMBL database).

Phylogenetic inference is primarily based on the ML method, which has been shown to produce consistent results across a relatively broad range of conditions (e.g., Swofford et al. 1996). For each of the data sets, identical tree topologies are obtained with the different substitution models. The more complex models generally produce higher likelihood scores (table 5; results shown only for the complete data sets). Consideration of a proportion of invariable positions, in addition to gamma-distributed rate heterogeneity across sites, significantly improves likelihood estimates inferred from the *ftsZ* gene (HKY85+Γ+I vs. HKY85+Γ; GTR+Γ+I vs. GTR+Γ; see also insignificant difference between GTR+Γ and HKY85+Γ+I). The GTR+Γ+I model seemed to provide the most realistic representation of the evolutionary dynamics inherent in *ftsZ* sequences, as the likelihoods for all other models were significantly worse. For both *wsp* data sets, consideration of invariable sites did not result in significantly increased likelihood scores (HKY85+Γ+I vs. HKY85+Γ and GTR+Γ+I vs. GTR+Γ). Here, the data appeared to be best described by the GTR+Γ model, for which the like-

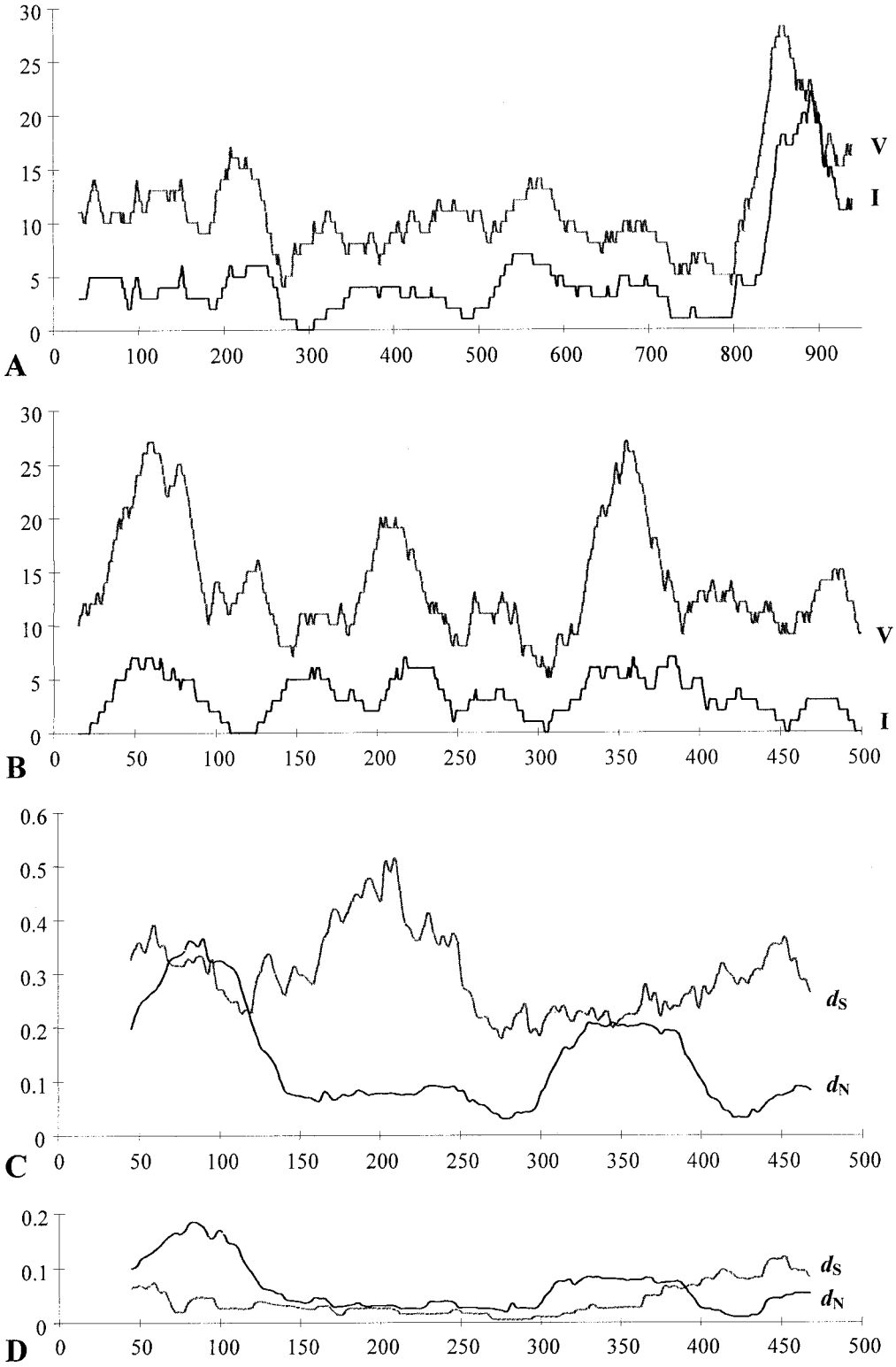

FIG. 3.—Distribution of the number of variable (V) and putatively informative positions (I) along (*A*) *ftsZ* and (*B*) *wsp* sequences and variation in non synonymous (d_N) and synonymous rates (d_S) along the *wsp* gene inferred from the average of (*C*) all comparisons and (*D*) only those which produced d_N/d_S ratios >1 for complete sequences (table 4). Putatively informative positions refer to alignment sites which exclusively support the split between ingroup and outgroup (split 1 for both data sets) and the partitions which are all-compatible (splits 2, 4, and 9 and 2–4, 11, 14, 15, and 19 for the *ftsZ* and *wsp* genes, respectively; fig. 1). Profiles in *A* and *B* are calculated from overlapping windows of 60 and 30 bp for the *ftsZ* and *wsp* genes, respectively. Average d_N and d_S rates are estimated from pairwise comparisons with Nei and Gojobori's (1986) method using MEGA (Kumar, Tamura, and Nei 1993) in overlapping windows of 90 bp (=30 codons). The X-axis describes the position along the genes from the 5′ end to the 3′ end. Values are given for the central positions of the respective windows.

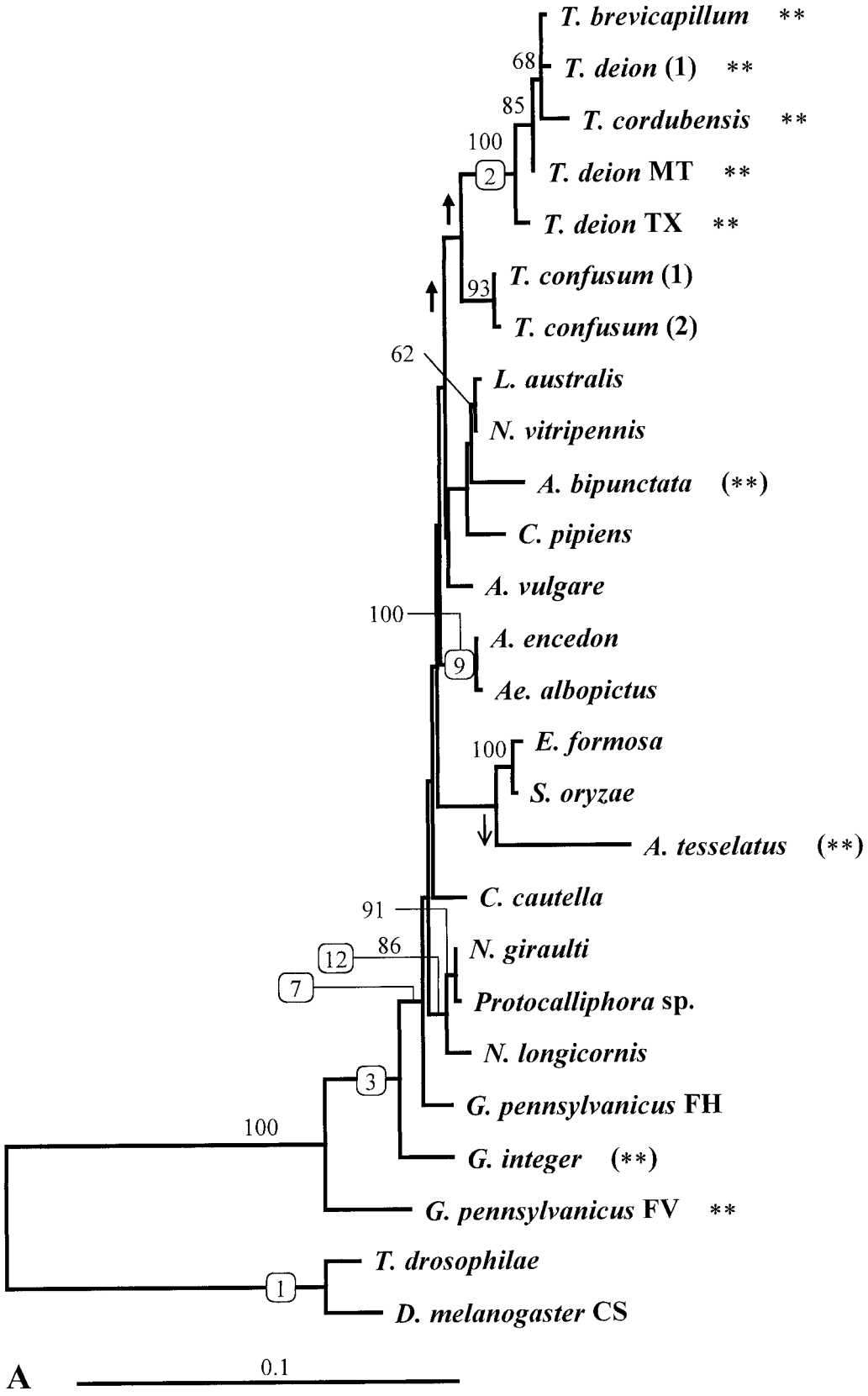

FIG. 4.—Phylogeny of B-group *Wolbachia* as inferred from maximum-likelihood (ML) analysis on (*A*) *ftsZ* and (*B*) *wsp* sequences. ML estimation was performed with PAUP*, versions 4.0b1 and 4.0.0d55, written by D. L. Swofford, using a heuristic search with branch-swapping by NNI and assuming the GTR+Γ+I or GTR+Γ substitution model for the *ftsZ* or *wsp* gene, respectively. Trees were rooted with *Drosophila melanogaster* CS and *Trichopria drosophilae* for the *ftsZ* data set and with *D. melanogaster* CS and *Drosophila sechellia* for the *wsp* gene. Branches are drawn in proportion to the estimated number of nucleotide substitutions per site (see bar in the bottom left corner). Circled numbers on branches refer to groups which are supported by split-supporting positions as indicated in figure 1. Bootstrap values ≥50, inferred from 100

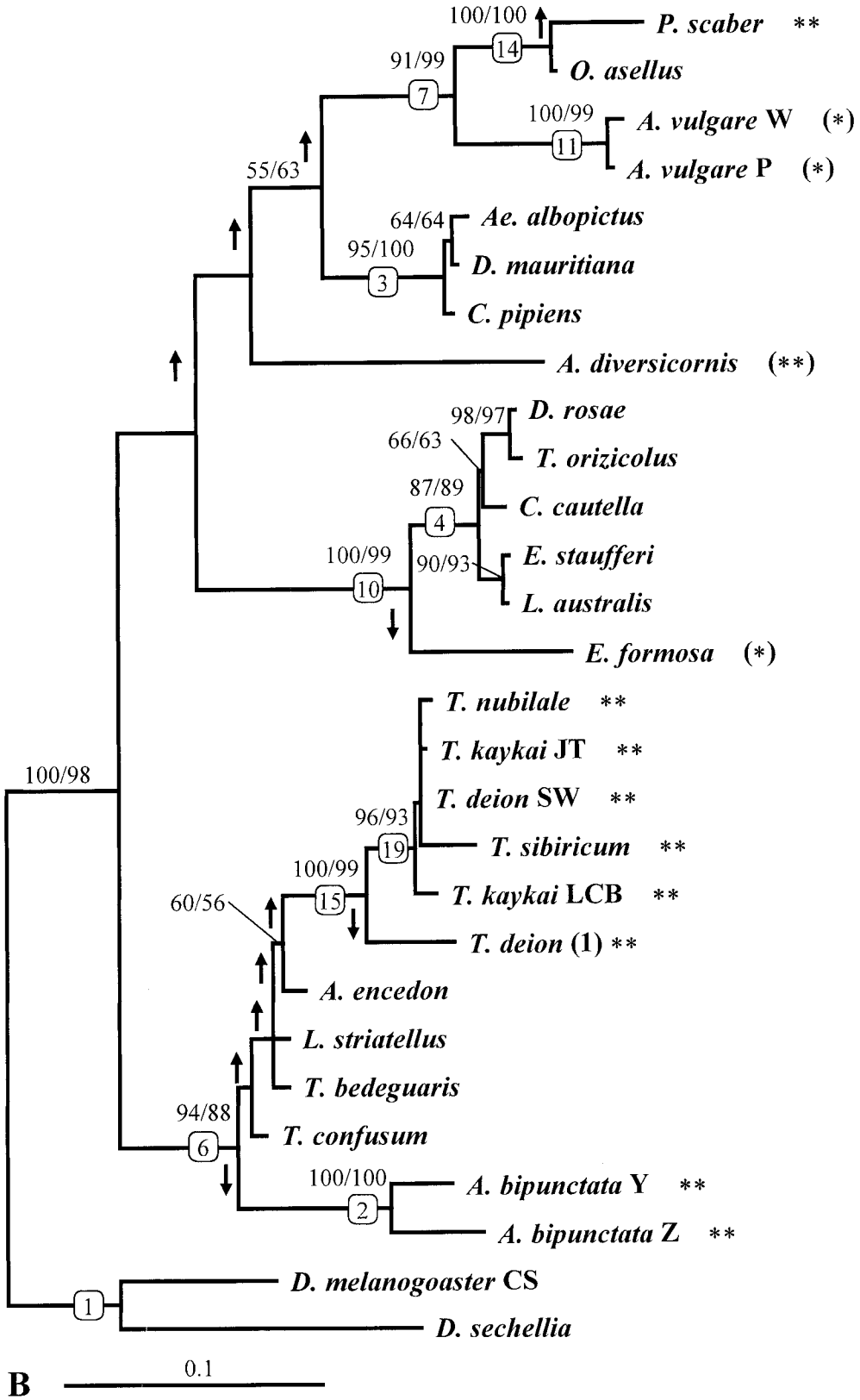

FIG. 4 *(Continued)* replicates, are given above branches. For the *wsp* gene, values before and after slashes were obtained from complete and reduced data sets (excluding the two positively selected regions), respectively. Arrows point to the clades with increased substitution rates as identified by the two-cluster test (Takezaki, Rzhetsky, and Nei 1995). Arrows in bold type and in arrows in normal type refer to the 1% and 5% significance levels, respectively. Stars indicate taxa which were shown to have elevated substitution rates by the method based on the likelihood ratio test. One and two stars refer to those taxa which had to be excluded from the data set to obtain likelihood differences between trees estimated with or without the assumption of a molecular clock that were no longer significant at the 5% and 1% levels, respectively. If different sets of a minimum number of taxa could be excluded, then stars for those taxa which were not included in all different sets are given in parentheses.

Table 5
Comparison of Different Substitution Models via the Likelihood Ratio Test

	HKY85 + Γ	HKY85 + Γ + I	GTR + Γ	GTR + Γ + I
ftsZ gene				
l[a]	−2,933.853	−2,926.394	−2,923.378	−2,916.438
$2\Delta l_{HKY85+\Gamma+I}$[b]		14.918 ($P\ \nu = 1 = 0.0001$)	20.95 ($P_{\nu=4} = 0.0003$)	34.83 ($P_{\nu=5} < 0.0001$)
$2\Delta l_{GTR+\Gamma}$			6.032 ($P_{\nu=3} = 0.1101$)	19.912 ($P_{\nu=4} = 0.0005$)
$2\Delta l_{GTR+\Gamma+I}$				13.88 ($P_{\nu=1} = 0.0002$)
wsp gene				
l	−3,173.274	−3,171.932	−3,148.916	−3,147.949
$2\Delta l_{HKY85+\Gamma+I}$		2.684 ($P_{\nu=1} = 0.1014$)	48.716 ($P_{\nu=4} < 0.0001$)	50.650 ($P_{\nu=5} < 0.0001$)
$2\Delta l_{GTR+\Gamma}$			46.032 ($P_{\nu=3} < 0.0001$)	47.966 ($P_{\nu=4} < 0.0001$)
$2\Delta l_{GTR+\Gamma+I}$				1.934 ($P_{\nu=1} = 0.1643$)

NOTE.—Substitution models used included the HKY85 (Hasegawa, Kishino, and Yano 1985) or GTR model (e.g., Lanave et al. 1984) with consideration of gamma-distributed rate heterogeneity across sites (Γ) and, in addition, a proportion of invariable positions (Γ + I). Maximum-likelihood estimation on either *ftsZ* or complete *wsp* sequences was performed with PAUP*, versions 4.0b1 and 4.0.0d55, written by David L. Swofford.

[a] Ln likelihood.

[b] Twice the likelihood difference between models (rows versus columns); the probability P of obtaining $2\Delta l$ under the hypothesis of no difference between models is calculated assuming a χ^2 distribution (e.g., Yang, Goldman, and Friday 1995).

lihood inferred was significantly higher than those of the simpler models but not significantly lower than that of GTR+Γ+I. The importance of invariable sites for reconstruction of *ftsZ* but not *wsp* gene evolution is likely to be due to the former bearing a smaller proportion of positions that are variable. The GTR+Γ+I and GTR+Γ models were therefore used for detailed ML analysis on *ftsZ* and *wsp* sequences, respectively. It is worth noting that simpler models, such as HKY85+Γ, produced identical topologies, the same number of clades supported by high bootstrap values, and the same results when hypotheses were assessed with the KH test. Consistency in the results generated with the different substitution models thus indicated that the phylogenetic signal, as detectable with the ML method, was rather robust in the data sets.

In detail, ML analysis of *ftsZ* sequences always produced three optimal tree topologies with identical likelihood scores. These trees only differed with respect to the relationships between symbionts from *Trichogramma brevicapillum, Trichogramma cordubensis,* and *T. deion* (1). In addition, ML-based bootstrapping provided support for 10 clades (bootstrap values ≥50), and the inferred topologies also included six lineages which were indicated by split-supporting positions (fig. 4*A*). Consideration of indels, as assessed with U-MP, resulted in one major difference in tree topology. If alignment gaps were treated as missing data, then the *Wolbachia* from *A. tesselatus* was represented as the sister taxon to all other B-group *Wolbachia* (bootstrap support of 56; results not shown). This partition is also indicated by PASSP (split 5, fig. 1*A*). However, if single gaps were given a weight of one (gaps = fifth base), then the tree showed *G. pennsylvanicus* FV as the sister taxon to all other B-group *Wolbachia,* in agreement with the results of ML and PASSP (figs. 1*A* and 4*A*).

Moreover, monophyly of all B-group *Wolbachia* excluding those from *A. bipunctata, T. confusum,* and *Gryllus* sp. is indicated. This clade, which uniquely shares a 9-bp deletion, was not inferred using ML, although it was indirectly indicated by split-supporting positions. Here, the corresponding outgroup was supported by six positions, all within the 9-bp indel region (split 4, fig. 1*A*). Note that for PASSP, alignment gaps are not considered ''positive'' character states. These six positions thus referred to alignment sites for which the *Wolbachia* of the A group, *A. bipunctata, T. confusum,* (1+2) and *Gryllus* sp., bore identical nucleotide states, including a defined degree of ''noise,'' whereas the remaining B-group *Wolbachia* all showed alignment gaps. Scarcity and positional conservation of deletions in the *ftsZ* gene suggests that those gaps at particular alignment positions are homologous and should therefore represent valid character states for phylogenetic analysis (Werren, Zhang, and Guo 1995; Bandi et al. 1998). It needs to be reiterated, though, that this particular clade not only is exclusively defined through the presence of the 9-bp deletion, but is also incompatible with other supported groups (table 2 and figs. 1*A* and 4*A*). The correct tree topology at this level can therefore only be inferred if it is known whether the 9-bp deletion or the nucleotide substitutions which support alternative clades are more likely to have a single origin. This is difficult to assess, as the likelihood of deletions and nucleotide substitutions cannot be directly compared using a simple model of sequence evolution. In addition, common ancestry of the 9-bp deletion is not supported by phylogenetic analysis of the *wsp* gene (fig. 4*B*). The exact phylogenetic relationships at this level should therefore currently be considered unresolved.

In conclusion, phylogenetic tree estimation confirms the results of PASSP. The *ftsZ* gene generally does not provide high resolution of the phylogeny of B-group *Wolbachia.* However, some clades are consistently identified by split-supporting positions and tree reconstruction methods (bootstrap support of ≥50). These include the *Wolbachia* from *Trichogramma* sp.; from *Nasonia girault, Nasonia longicornis,* and *Protocalliphora* sp.; and from *A. albopictus* and *A. encedon* (figs. 1*A* and 4*A*). Unambiguous phylogenetic information therefore seems to be conserved for at least some closely related taxa.

The *wsp* gene generally proved to be more informative than *ftsZ* sequences. Results are identical for the

two *wsp* data sets, although exclusion of the positively selected regions generally led to lower bootstrap support of the inferred clades (fig. 4*B*). Therefore, these two specific regions did not appear to bear a higher proportion of misleading information. In both cases, ML analysis generated 15 optimal tree topologies. Differences between topologies only refer to the relationships between the different *Trichogramma* species, excluding *T. deion* (1). Consideration of indels did not affect tree reconstruction (results not shown). The inferred phylogenetic trees thus contained 11 clades which were indicated by PASSP, including almost all well-supported splits (figs. 1*B* and 4*B*). Seventeen clades were also supported by high bootstrap values. Little support (bootstrap values <50) was found for those parts of the trees characterized by short branch-lengths, indicating a lack of phylogenetic information. In addition, bootstrap analysis suggested that the relationships of the major B-group *Wolbachia* lineages cannot be resolved with absolute certainty, although they were related to each other along branches of considerable length in the optimal ML trees. The data should therefore generally contain sufficient information at this level. However, phylogenetic signal inconsistency was already indicated here by PASSP. Four of the major B-group lineages in the ML trees were identified by split-supporting positions (splits 3, 6, 7, and 10). Their relationships to each other was indirectly indicated via support for the respective outgroups (splits 8, 12, and 18). These splits were supported by about the same number of positions but were incompatible with each other (table 3). This suggests, first, that the original signal for a subordinate B-group *Wolbachia* clade has been masked, whereas that for the corresponding outgroup has been conserved, and, second, that the more derived major B-group *Wolbachia* clades additionally bear analogies with the outgroup.

In this context, it is interesting to note that previous analyses of *wsp* sequences which included a smaller number of B-group *Wolbachia* indicated different relationships of the major B-group lineages and of those clades with short branches in the ML tree of figure 3*B* regardless of the tree reconstruction method used (Zhou, Rousset, and O'Neill 1998; Hurst et al. 1999*a*; Van Meer, Witteveldt, and Stouthamer 1999; unpublished data). These differences in tree topology generally support our observation that *wsp* sequences do not contain sufficient unambiguous information at these levels. In fact, the different branching order of the major B-group *Wolbachia* lineages in some of these studies is most likely to be due to ''long-branch attraction,'' as clades with long branches are shown to cluster together. This is not the case with the extended data set used in our study. The addition of taxa therefore seems to improve phylogenetic resolution. However, the ''long-branch problem'' may not have been resolved in all cases. For instance, the symbionts from *G. pennsylvanicus* FV or *A. bipunctata* were inferred from *ftsZ* and *wsp* sequences, respectively, to represent the most basal lineages of well-supported clades (fig. 4). The inferred phylogenetic positions of these taxa, which both show increased substitution rates in the respective genes, could thus be a consequence of long-branch attraction to the respective outgroups. In these cases, the results of phylogenetic tree reconstruction must be considered to be ambiguous.

Finally, complete *wsp* sequences produced d_N/d_S ratios >1 between closely related taxa with the exception of two comparisons which included symbionts from *A. diversicornis*, *Torymus bedeguaris*, and *T. confusum* (table 4 and fig. 4*B*). Values greater than 1 indicate diversifying selection along the evolutionary lineage connecting the respective taxa. Convergent character state evolution due to positive selection is not expected between such taxa, although analogies might arise to unaffected organisms. Positive selection on *wsp* sequence evolution is thus unlikely to have a major effect on phylogenetic tree reconstruction in those cases in which d_N/d_S ratios >1 are produced between closely related organisms. However, it may produce biased results for the symbionts of *A. diversicornis*, *T. bedeguaris*, or *T. confusum*. Note that for these taxa, exact inference of phylogenetic relationships may also be hampered by either lack of sufficient information (*T. bedeguaris*, *T. confusum*) or increased substitution rates (*A. diversicornis*) (see above).

Origin and Evolution of Male-Killing *Wolbachia*

Although *ftsZ* sequences generally did not provide high resolution of the relationships of B-group *Wolbachia*, the inferred phylogenetic trees consistently showed the male-killing *Wolbachia* to belong to different lineages. In particular, the male-killer of *A. encedon* was found to form a monophyletic group with the symbiont of *A. albopictus* by tree reconstruction methods and split-supporting positions (figs. 1*A* and 4*A*). Their *ftsZ* sequences showed 1 nucleotide difference, whereas those of the male-killing *Wolbachia* from *A. encedon* and *A. bipunctata* differed at 22 positions and with respect to the presence of the 9-bp deletion. For the *wsp* data set, the male-killers of *A. bipunctata* and *A. encedon* seemed to belong to different lineages, although within the same clade. In contrast to the *ftsZ* gene, *wsp* sequences of the *Wolbachia* from *A. encedon* and *A. albopictus* fell into distantly related clades (fig. 4*B*). They showed 67 nucleotide differences, whereas those of the symbionts from *A. encedon* and *A. bipunctata* varied at 44–47 positions.

As the two data sets apparently produced incongruent results, we specifically tested monophyly of male-killing *Wolbachia* using the KH test (Kishino and Hasegawa 1989). For the *ftsZ* gene, the tree topology which considered male-killer monophyly was significantly worse than the optimal ML tree. This was not the case for the *wsp* gene. Moreover, the *wsp* data set clearly rejected monophyly of the *Wolbachia* from *A. encedon* and *A. albopictus*, which is supported by *ftsZ* sequences (table 6). Such consistently identified discrepancies between the two genes may be a consequence of one of the following factors:

1. The phylogenetic positions of the symbionts from *A. encedon*, *A. bipunctata*, or *A. albopictus* have not been correctly inferred due to increased substitution

Table 6
Results of the Kishino-Hasegawa (KH) Test

Null Hypothesis[a]	l[b]	Δl[c]	SD	T[d]	P[e]
ftsZ gene					
(*Acraea encedon, Adalia bipunctata*)	−2,944.642	28.204	9.697	2.908	0.0037
wsp gene					
(*A. encedon, A. bipunctata*)	−3,162.041	13.125	6.583	1.875	0.0614
(*A. encedon, Aedes albopictus*)	−3,309.145	160.229	23.847	6.6833	<0.0001

[a] Maximum-likelihood (ML) tree estimates were obtained for the null hypotheses with the GTR + Γ + I (*ftsZ* gene) or the GTR + Γ substitution model (*wsp* gene), using the indicated topological constraints.

[b] Ln likelihood.

[c] Ln likelihood difference from the optimal ML tree.

[d] Test statistic.

[e] Probability of getting a more extreme *T* under the assumption of no difference between null hypothesis and the optimal ML tree according to the KH test as implemented in PAUP*.

rates in at least one of the genes. This may be the case for the male-killers of *A. bipunctata*. However, increased substitution rates have not been indicated for the *Wolbachia* from either *A. encedon* or *A. albopictus* and thus cannot explain the differences between genes regarding monophyly of these two taxa.

2. *ftsZ* and *wsp* sequences have not been isolated for the same, but for distantly related symbiont strains from at least one of the above-mentioned host species.

3. Double infections with distantly related symbionts in a single specimen of one of these hosts could have resulted in the exchange of genetic material between strains through recombination.

In the latter two cases, *ftsZ* and *wsp* sequences isolated from the same host or even the same symbiont may have different phylogenetic origins. At least for the male-killing symbionts, hypothesis 2 can be excluded, as their sequences have been isolated from the same host specimens. Sequencing of independent clones in each case has not revealed the presence of more than one symbiont strain per host specimen (Hurst et al. 1999*a*; unpublished data). This cannot be excluded for most of the other taxa, including *A. albopictus.* It is also interesting to note that double infections of host species have been recorded for a variety of arthropods, including *A. bipunctata* and *A. albopictus* (e.g., Werren, Zhang, and Guo 1995; Hurst et al. 1999*a*). However, the different strains of any of these hosts are either very closely related or belong to different *Wolbachia* subgroups (A+B), whereas distantly related taxa of the same subgroup must be expected if the above-described hypotheses are to explain the observed inconsistencies.

Although current data are thus insufficient for a clarification of the origin of the male-killing behavior within the genus *Wolbachia*, they nevertheless provide insights into the evolution of the male-killing *Wolbachia* from *A. bipunctata.* These are found to bear identical *ftsZ* but different *wsp* sequences (Hurst et al. 1999*a*). *Adalia bipunctata* is therefore host to two different strains of male-killing *Wolbachia* which diverged recently, as they are only indicated by the faster- evolving *wsp* gene. Interestingly, divergence of strains concurs with diversification of *wsp* sequences in response to positive selection, as indicated by the unusually high d_N/d_S ratio of 4.297. Both genes, moreover, show increased substitution rates.

We here propose that these observations have a biological meaning. The phenomenon of male-killing entails an evolutionary conflict of ''interest'' between symbiont and host. As male-killing bears a cost for the host organism, selection should favor suppressers of the male-killing behavior and/or resistance against the infection in general. This, in turn, is expected to produce a response from the symbiont, subsequently resulting in an evolutionary arms race between symbiont and host (Hurst, Atlan, and Bengtsson 1996; Hurst, Hurst, and Majerus 1997). Selection pressures involved in such a scenario should be particularly pronounced in *A. bipunctata*, which is host to at least four different male-killing bacteria, including the two *Wolbachia* strains, which all coinhabit the Moscow population (Hurst et al. 1999*a*, 1999*b*; unpublished data). Increased substitution rates may therefore reflect a generally elevated rate of evolution in these organisms in response to antagonistic symbiont-host interactions. In addition, positive selection on the *wsp* gene, which codes for an outer membrane protein (Braig et al. 1998), suggests that the WSP protein plays an important role in symbiont-host interactions, in accordance with previous findings about the function of positively selected membrane proteins of parasites and viruses (Endo, Ikeo, and Gojobori 1996).

Consequences for Investigations of the Evolution of *Wolbachia*

Molecular phylogenies currently represent the primary source of information about the evolution of *Wolbachia* bacteria. They have been employed to study host-symbiont associations, the origin of symbiont behavior, and symbiont divergence dates (e.g., Werren, Zhang, and Guo 1995; Bandi et al. 1998; Bouchon, Rigaud, and Juchault 1998; Schilthuizen and Stouthamer 1998; Van Meer, Witteveldt, and Stouthamer 1999). Such analyses crucially depend on the validity of the inferred phylogeny. The results of our study demonstrate that two of the available *Wolbachia* gene regions show different qualities as to their utility for phylogenetic reconstruction. In particular, lack of phylogenetic information content, or the presence of homoplasies due to

substitution rate heterogeneity between lineages, confounds phylogenetic analysis of *ftsZ* and *wsp* sequences at different phylogenetic levels. Phylogenetic utility of the only other previously analyzed *Wolbachia* gene, 16S rDNA, is expected to show similar restrictions. It bears less variation than the genes considered here (Bandi et al. 1998; Bouchon, Rigaud, and Juchault 1998) and must thus be assumed to be less informative. Therefore, the phylogenetic relationships of *Wolbachia* bacteria can currently be inferred with confidence only for a limited number of clades. In addition, the presence of significant substitution rate heterogeneity between lineages in both *ftsZ* and *wsp* gene sequences forbids reliable estimation of divergence dates and also limits the applicability of a simple sequence-based classification system such as that proposed for the *wsp* gene (Zhou, Rousset, and O'Neill 1998).

In spite of such restrictions, the available information does allow specific hypotheses about the evolution of these organisms to be tested. For instance, our results consistently indicate at least two independent transitions to the parthenogenesis-inducing behavior within B-group *Wolbachia*. In addition, ''colonization'' of the different *Trichogramma* species or of the isopod hosts is shown in each case to have a common origin. Moreover, horizontal transfer is confirmed as a mode of symbiont transmission, at least for those cases in which *Wolbachia* from distantly related insect hosts are consistently found to belong to the same well-supported clade (see also Stouthamer 1997; Werren 1997). More detailed insights into the evolution of *Wolbachia* may be obtained in the future by the addition of taxa to ''break up'' long branches and an analysis of faster-evolving DNA regions to resolve the relationships of taxa for which current data lack sufficient phylogenetic information. In addition, for a reliable reconstruction of *Wolbachia* phylogenies, it is also important to evaluate whether and how often recombination contributed to symbiont genome evolution and may therefore confound phylogenetic inferences.

Finally, the above-hypothesized causes of elevated substitution rates and selectional constraints in the *Wolbachia* from *A. bipunctata* should also apply to other taxa. Accelerated substitution rates may thus serve as a means to test for the persistence of antagonistic host-symbiont interactions. In addition, the presence of positive selection on the evolution of complete *wsp* sequences, as indicated between some taxa by d_N/d_S ratios >1, suggests a possible role of the respective surface proteins in the symbiont's interaction with the host. In this context, it is worth reiterating that two specific regions within the *wsp* gene appear to be subject to positive selection across all symbionts. This may indicate that WSP proteins are generally involved in *Wolbachia*-host interactions and, in particular, that these two regions represent the primary location of such interactions. d_N/d_S ratios >1 are then only found for complete sequences in those cases in which selectional constraints are very strong, e.g., in response to a pronounced antagonism in symbiont-host relationships, as here suggested for *A. bipunctata*. In fact, such an antagonism should be associated with transitions to a sex ratio biasing behavior, as the distortion of a 1:1 sex ratio is assumed to produce conflict between the causative agent and host (Hurst, Atlan, and Bengtsson 1996). Interestingly, a general increase in the substitution rate of *wsp* sequences is almost exclusively found in such bacteria, e.g., the *Wolbachia* of *A. bipunctata* (male-killing), *A. vulgare* (feminization), *A. diversicornis*, *E. formosa*, and *Trichogramma* sp. (all parthenogenesis induction). However, such observations are not entirely consistent between *Wolbachia* genes. Moreover, although d_N/d_S ratios >1 for complete *wsp* sequences are observed between some of the sex ratio distorters, they are also inferred from comparisons which include cytoplasmic incompatibility-inducing bacteria (e.g., *Wolbachia* from *Cadra cautella*, *L. striatellus*, and *T. confusum*). Assessment of the origin of increased substitution rates or the presence of positive selection thus requires further testing, using sequence data from a larger variety of taxa and more detailed information on the exact nature of symbiont-host interactions at both the phenotypic and the molecular levels.

Acknowledgments

We are grateful to Nick Goldman, Laurence Hurst, Richard Stouthamer, Wolfgang Wägele, Ziheng Yang, and two anonymous reviewers for valuable comments on this work. We also thank Etsuko Moriyama and Wolfgang Wägele for providing computer programs and John Huelsenbeck for advice on PAUP*. J.H.G.v.d.S. was funded by a TMR fellowship of the EU, G.D.D.H. by a BBSRC D. Phillips fellowship, and F.M.J. by a BBSRC studentship.

LITERATURE CITED

ALVAREZ-VALIN, F., K. JABBARI, and G. BERNARDI. 1998. Synonymous and nonsynonymous substitutions in mammalian genes: intragenic correlations. J. Mol. Evol. **46**:37–44.

BANDI, C., T. J. C. ANDERSON, C. GENCHI, and M. L. BLAXTER. 1998. Phylogeny of *Wolbachia* in filarial nematodes. Proc. R. Soc. Lond. B Biol. Sci. **265**:2407–2413.

BOUCHON, D., T. RIGAUD, and P. JUCHAULT. 1998. Evidence for widespread *Wolbachia* infection in isopod crustaceans: molecular identification and host feminization. Proc. R. Soc. Lond. B Biol. Sci. **265**:1081–1090.

BOURTZIS, K., S. L. DOBSON, H. R. BRAIG, and S. L. O'NEILL. 1998. Rescuing *Wolbachia* have been overlooked . . . Nature **391**:852–853.

BRAIG, H. R., W. G. ZHOU, S. L. DOBSON, and S. L. O'NEILL. 1998. Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia pipientis*. J. Bacteriol. **180**:2373–2378.

CERCHIO, S., and P. TUCKER. 1998. Influence of alignment on the mtDNA phylogeny of Cetacea: questionable support for a Mysticeti/Physeteroidea clade. Syst. Biol. **47**:336–344.

COMERON, J. M., and M. AGUADE. 1998. An evaluation of measures of synonymous codon usage bias. J. Mol. Evol. **47**:268–274.

ENDO, T., K. IKEO, and T. GOJOBORI. 1996. Large-scale search for genes on which positive selection may operate. Mol. Biol. Evol. **13**:685–690.

FELSENSTEIN, J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. Syst. Zool. **27**:401–410.

———. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17**:368–376.

———. 1985. Confidence limits on phylogenies—an approach using the bootstrap. Evolution **39**:783–791.

FOSTER, P. G., and D. A. HICKEY. 1999. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. J. Mol. Evol. **48**:284–290.

HASEGAWA, M., H. KISHINO, and T. YANO. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **21**:160–174.

HE, M., and D. S. HAYMER. 1995. Codon bias in actin multigene families and effects on the reconstruction of phylogenetic relationships. J. Mol. Evol. **41**:141–149.

HENDY, M. D., and D. PENNY. 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. **38**:297–309.

HENNIG, W. 1966. Phylogenetic systematics. University of Illinois Press, Urbana, Ill.

HOFFMANN, A. A., and M. TURELLI. 1997. Cytoplasmic incompatibility in insects. Pp. 42–80 *in* S. L. O'NEILL, A. A. HOFFMANN, and J. H. WERREN, eds. Influential passengers. Oxford University Press, Oxford, New York, Tokyo.

HURST, G. D. D., L. D. HURST, and M. E. N. MAJERUS. 1997. Cytoplasmic sex-ratio distorters. Pp. 125–154 *in* S. L. O'NEILL, A. A. HOFFMANN, and J. H. WERREN, eds. Influential passengers. Oxford University Press, Oxford, New York, Tokyo.

HURST, G. D. D., F. M. JIGGINS, J. H. G. V. D. SCHULENBURG, D. BERTRAND, I. I. GORIACHEVA, I. A. ZAKHAROV, J. H. WERREN, R. STOUTHAMER, and M. E. N. MAJERUS. 1999*a*. Male killing *Wolbachia* in two species of insect. Proc. R. Soc. Lond. B Biol. Sci. **266**:735–740.

HURST, G. D. D., J. H. G. V. D. SCHULENBURG, T. M. O. MAJERUS, D. BERTRAND, I. A. ZAKHAROV, J. BAUNGAARD, W. VOLKL, R. STOUTHAMER, and M. E. N. MAJERUS. 1999*b*. Invasion of one insect species, *Adalia bipunctata*, by two different male-killing bacteria. Insect Mol. Biol. **8**:133–139.

HURST, L. D., A. ATLAN, and B. O. BENGTSSON. 1996. Genetic conflicts. Q. Rev. Biol. **71**:317–364.

KISHINO, H., and M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. **29**:170–179.

KUMAR, S., K. TAMURA, and M. NEI. 1993. MEGA: molecular evolutionary genetic analysis. Version 1.0. Pennsylvania State University, University Park.

LANAVE, C., G. PREPARATA, C. SACCONE, and G. SERIO. 1984. A new method for calculating evolutionary substitution rates. J. Mol. Evol. **20**:86–93.

LI, W.-H. 1993. Unbiased estimation of the rate of synonymous and nonsynonymous substitution. J. Mol. Evol. **36**:96–99.

LOCKHART, P. J., M. A. STEEL, M. D. HENDY, and D. PENNY. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol. Biol. Evol. **11**:605–612.

MERCOT, H., and D. POINSOT. 1998. . . . and discovered on Mount Kilimanjaro. Nature **391**:853.

MORIYAMA, E. N. 1998. Molecular evolutionary analysis package (MEA). Department of Ecology and Evolutionary Biology, Yale University, New Haven, Conn.

MORIYAMA, E. N., and J. R. POWELL. 1997. Codon usage bias and tRNA abundance in Drosophila. J. Mol. Evol. **45**:514–523.

NEI, M., and T. GOJOBORI. 1986. Simple methods for estimating the number of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.

RIGAUD, T. 1997. Inherited microorganisms and sex determination of arthropod hosts. Pp. 81–101 *in* S. L. O'NEILL, A. A. HOFFMANN, and J. H. WERREN, eds. Influential passengers. Oxford University Press, Oxford, New York, Tokyo.

SCHILTHUIZEN, M., and R. STOUTHAMER. 1998. Distribution of *Wolbachia* among the guild associated with the parthenogenetic gall wasp *Diplolepis rosae*. Heredity **81**:270–274.

SCHULENBURG, J. H. G. V. D., U. ENGLISCH, and J. W. WAGELE. 1999. Evolution of ITS1 rDNA in the Digenea (Platyhelminthes: Trematoda): 3′ end sequence conservation and its phylogenetic utility. J. Mol. Evol. **48**:2–12.

SOKAL, R. R., and F. J. ROHLF. 1995. Biometry. 3rd edition. W. H. Freeman, New York.

STOUTHAMER, R. 1997. *Wolbachia*-induced parthenogenesis. Pp. 102–124 *in* S. L. O'NEILL, A. A. HOFFMANN, and J. H. WERREN, eds. Influential passengers. Oxford University Press, Oxford, New York, Tokyo.

STRIMMER, K., and A. VON HAESELER. 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. **13**:964–969.

SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, and D. M. HILLIS. 1996. Phylogenetic inference. Pp. 407–514 *in* D. M. HILLIS, C. MORITZ, and B. K. MABLE, eds. Molecular systematics. 2nd edition. Sinauer, Sunderland, Mass.

TAKEZAKI, N., A. RZHETSKY, and M. NEI. 1995. Phylogenetic test of the molecular clock and linearized trees. Mol. Biol. Evol. **12**:823–833.

TAMURA, K., and M. NEI. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. **10**:512–526.

VAN MEER, M. M. M., J. WITTEVELDT, and R. STOUTHAMER. 1999. Phylogeny of the arthropod endosymbiont *Wolbachia* based on the *wsp* gene. Insect Mol. Biol. **8**:399–408.

VAVRE, F., C. GIRIN, and M. BOULETREAU. 1999. Phylogenetic status of a fecundity-enhancing *Wolbachia* that does not induce thelytoky in Trichogramma. Insect Mol. Biol. **8**:67–72.

WÄGELE, J. W. 1996. First principles of phylogenetic systematics, a basis for numerical methods used for morphological and molecular characters. Vie Milieu **46**:125–138.

WÄGELE, J. W., and F. RÖDDING. 1998. A priori estimation of phylogenetic information conserved in aligned sequences. Mol. Phylogenet. Evol. **9**:358–365.

WERREN, J. H. 1997. Biology of *Wolbachia*. Annu. Rev. Entomol. **42**:587–609.

WERREN, J. H., W. ZHANG, and L. R. GUO. 1995. Evolution and phylogeny of *Wolbachia*—reproductive parasites of arthropods. Proc. R. Soc. Lond. B Biol. Sci. **261**:55–63.

WRIGHT, F. 1990. The effective number of codons used in a gene. Gene **87**:23–29.

YANG, Z. 1996. Among-site variation and its impact on phylogenetic analyses. Trends Ecol. Evol. **11**:367–371.

———. 1998*a*. On the best evolutionary rate for phylogenetic analysis. Syst. Biol. **47**:125–133.

———. 1998*b*. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. **15**:568–573.

YANG, Z., N. GOLDMAN, and A. FRIDAY. 1995. Maximum likelihood trees from DNA sequences: a peculiar statistical problem. Syst. Biol. **44**:384–399.

ZHOU, W. G., F. ROUSSET, and S. O'NEILL. 1998. Phylogeny and PCR-based classification of *Wolbachia* strains using *wsp* gene sequences. Proc. R. Soc. Lond. B Biol. Sci. **265**:509–515.

HOWARD OCHMAN, reviewing editor

Accepted December 14, 1999