Hierarchical Analysis of Variation in the Mitochondrial 16S rRNA Gene Among Hymenoptera

James B. Whitfield and Sydney A. Cameron*†*

*Department of Entomology and †Department of Biological Sciences, University of Arkansas

Nucleotide sequences from a 434-bp region of the 16S rRNA gene were analyzed for 65 taxa of Hymenoptera (ants, bees, wasps, parasitoid wasps, sawflies) to examine the patterns of variation within the gene fragment and the taxonomic levels for which it shows maximum utility in phylogeny estimation. A hierarchical approach was adopted in the study through comparison of levels of sequence variation among taxa at different taxonomic levels. As previously reported for many holometabolous insects, the 16S data reported here for Hymenoptera are highly AT-rich and exhibit strong site-to-site variation in substitution rate. More precise estimates of the shape parameter (α) of the gamma distribution and the proportion of invariant sites were obtained in this study by employing a reference phylogeny and utilizing maximum-likelihood estimation. The effectiveness of this approach to recovering expected phylogenies of selected hymenopteran taxa has been tested against the use of maximum parsimony. This study finds that the 16S gene is most informative for phylogenetic analysis at two different levels: among closely related species or populations, and among tribes, subfamilies, and families. Maximization of the phylogenetic signal extracted from the 16S gene at higher taxonomic levels may require consideration of the base composition bias and the site-to-site rate variation in a maximum-likelihood framework.

Introduction

Selecting a gene for phylogenetic analysis requires matching the level of sequence variation to the desired taxonomic level of study. Several recent papers have focused on the identification of genes that are useful for phylogenetic analysis at different taxonomic levels (Brower and DeSalle 1994; Friedlander, Regier, and Mitter 1994; Graybeal 1994; Simon et al. 1994; Cho et al. 1995). For many of these genes, sequence data are available from a relatively small sample of taxa with roughly known divergence times. These studies permit estimates of sequence divergence rates (e.g., number of nucleotide substitutions or percentage of sequence divergence over time), providing information on the relative rate of change of a gene. However, estimates of sequence divergence rate calculated from a small sample of taxa may not be appropriate when applied more generally (Graybeal 1994), because unsampled lineages may differ in divergence rate. This problem will resolve itself as sequence data are collected from additional genes for increasingly larger numbers of taxa.

A few mitochondrial (mtDNA) genes have been studied extensively within recently diverged lineages of arthropods ($<$5 MYA). These genes (12S rRNA, 16S rRNA, cytochrome oxidase I) exhibit nearly the same divergence rate, which is linear with time and approximates 2.3% per Myr for silent sites (Brower 1994). However, when more anciently diverged lineages ($>$75 MYA) are compared, different mtDNA genes exhibit considerable variation in sequence divergence rate (Cummings, Otto, and Wakeley 1995), with some showing greater conservation than others. Furthermore, a number of constraints can influence variation in the rate of nucleotide substitution among sites within a gene (Wheeler and Honeycutt 1988; Mindell and Honeycutt 1990; Hillis and Dixon 1991; Kraus et al. 1992). These constraints may be of a general nature, such as variation in the rate of substitution by codon position in protein-coding genes or by secondary structural position in rRNA genes, or they may be lineage-specific (some taxa appear to evolve more slowly than others; e.g., DeSalle and Templeton 1988; Hasegawa and Kishino 1989).

Graybeal (1994) pointed out that any given gene's potential phylogenetic utility at a particular taxonomic level depends not only on the percentage of sequence divergence at that level, but also on the shape of the sequence divergence accumulation curve. For example, at a given observed divergence level, genes in which only a few sites are ''free to vary'' (*sensu* Palumbi 1989) will contain more superimposed changes (i.e., be more saturated with nucleotide substitutions) than those in which many sites are able to change. The pattern that emerges when few sites are free to vary is a sequence divergence accumulation curve that is strongly convex near the origin, then flattens out at a low level over the remainder of the distribution of divergence times as additional substitutions are superimposed and thus go unobserved. (This curve contrasts sharply with the more linear curve seen with recently diverged lineages or with genuinely highly conserved genes.) The low overall sequence divergence level for older divergences might suggest a strongly conserved gene appropriate for higher-level comparisons, when, in fact, the available variation may be useful only at lower taxonomic levels, among recently diverged taxa. An appreciation of the distribution of variable sites across the gene is therefore important in examining the phylogenetic utility of a gene for a particular taxonomic level.

Factors other than the distribution of rate variation among sites can determine the shape of the sequence divergence accumulation curve. For instance, in many holometabolous insects, including Hymenoptera and *Drosophila,* mtDNA exhibits a nucleotide composition

Key words: Hymenoptera, molecular phylogeny, insects, mitochondrial DNA.

Address for correspondence and reprints: James B. Whitfield, Department of Entomology, 321 Agriculture Building, University of Arkansas, Fayetteville, Arkansas 72701. E-mail: jwhitfie@comp.uark.edu.

Mol. Biol. Evol. 15(12):1728–1743. 1998
© 1998 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

which is strongly biased toward adenine and thymine (AT bias). For some groups, the mean percentage of AT can be higher than 80% (Cameron 1993; Crozier and Crozier 1993; Simon et al. 1994; Dowton and Austin 1994, 1997*a,* 1997*b*; Whitfield 1997). When the base composition is biased to that degree, obviously, the ratio of transversions (tv) to transitions (ti) increases, as does the probability of convergent substitutions to the more common bases. Both of these increases further reduce the ability to correctly estimate the number and proportion of hidden mutations and, hence, the ability to correct sequence divergence for hidden changes.

Fortunately, increasingly complex models of substitutional change continue to be incorporated into methods of correcting sequence divergence rates for saturation (Jukes and Cantor [1969] and the models that followed), ti/tv bias (Kimura 1980), base composition bias (Tajima and Nei 1984; Tamura 1992; Tamura and Nei 1993), and rate variation among sites (Yang 1995; Nielsen 1997). Thus, even though the functional constraints on rRNA, tRNA, and a large variety of protein-coding genes are not well understood, it is still possible to investigate the patterns of variation and phylogenetic utility of these genes in some detail.

In this paper, we examine the mitochondrial large-subunit (16S) rRNA gene for hierarchical patterns of sequence variation and divergence among a large and diverse array of hymenopteran insects (ants, bees, wasps, parasitoids, and sawflies). We address several aspects of 16S sequence variation: (1) the observed patterns of sequence divergence at different taxonomic levels; (2) how those patterns are affected by correcting the sequence divergences under different models; (3) the inferred ti : tv ratios at various hierarchical levels; (4) base composition bias within the Hymenoptera, and the potential effects of this bias on rates of change and phylogenetic informativeness; and (5) the distribution of variable sites across the surveyed gene and the position of those sites relative to the inferred secondary structure of the molecule.

For the vast majority of Hymenoptera, fossils are not available for accurate estimates of divergence times (Whitfield 1998). Thus, explicit plots of sequence divergence against time (as in Brower 1994; Graybeal 1994) are not possible. Instead, we have taken a hierarchical approach in this study, comparing sequence divergences for species within genera, for genera within tribes or subfamilies, for subfamilies within families, for families within superfamilies, for superfamilies within the order. We realize that taxonomic levels are necessarily subjective and somewhat arbitrarily determined among different taxa (although this problem should be minimized by confining comparisons to within a single order of animals), and that this subjectivity will introduce an unknown amount of variability into our analyses. What this approach lacks in determination of rate accuracy may, however, be counterbalanced by (1) the large and diverse number of our taxonomic comparisons and (2) the applicability of our methods without prior estimates of fossil-calibrated divergence times.

Prior Uses and Criticisms of the 16S Gene for Phylogeny

Sequences coding for the 16S rRNA gene have been used for estimating phylogenies over a notable range of taxonomic levels (see table 1 for a survey of studies involving insects). The existence of sites that are changing at widely differing rates within this single gene (Hillis and Dixon 1991; Simon et al. 1994) suggests that 16S sequences contain historical information that is useful at more than one level of phylogenetic divergence. In a recent review, Simon et al. (1994) suggested that data from 16S might be useful primarily for phylogenetic estimation at higher levels, because few sites were variable at lower levels among closely related species, and even some of those quickly saturated. In contrast, Engel and Schultz (1997), in a reanalysis of Cameron's (1991, 1993) data for estimating relationships among the corbiculate bees, suggested that *Apis* (honey bee) species relationships recovered from 16S sequences (Cameron et al. 1992) were strongly congruent with those inferred from morphological data.

In cases in which well-corroborated phylogenies are available, concordance of DNA-based results with well-researched phylogenies based on morphology and other evidence often provides a good test of the phylogenetic informativeness of molecular data (e.g., Friedlander et al. [1996] for a nuclear protein-coding gene; Smith [1989] for rRNA genes). In this context, attempts to use the 16S gene for estimation of relationships at higher taxonomic levels have met with mixed (although significant) success. Cameron (1993) obtained a tribal phylogeny of the corbiculate bees which conflicts with morphology-based phylogenies (Roig-Alsina and Michener 1993) but is fully concordant with results from other genes, including the nuclear large-subunit (28S) rRNA gene (Sheppard and McPheron 1991; unpublished data), the major opsin gene (Mardulyn and Cameron 1998), and mitochondrial cytochrome *b* (Koulianos et al. 1998). Dowton and Austin (1994) and Flook and Rowell (1997*a,* 1997*b*) used 16S sequence data to recover relationships among the superfamilies of Hymenoptera and Orthoptera, respectively, which were largely consistent with those based on morphology. Other 16S phylogenetic analyses of hymenopteran families (Dowton et al. 1997) and subfamilies (Whitfield 1997; Dowton, Austin, and Antolin 1998) have recovered relationships that are largely congruent with those based on morphology.

In these higher-level molecular studies, considerable phylogenetic ''noise'' is present, presumably due to the saturation of many of the variable sites at those levels. Such noise has often been nullified or reduced by incorporating compensatory calculations or weights into parsimony or maximum-likelihood (ML) analyses (Swofford et al. 1996). Greater knowledge of how noise accumulates with increasing divergence level and further elucidation of patterns of variation within the 16S gene will greatly assist in the extraction of meaningful phylogenetic signal from the sequence data (Dowton and Austin 1997*a*; Flook and Rowell 1997*a,* 1997*b*;

Table 1
A Selection of Published Phylogenetic Studies of Insect Taxa Using 16S Sequence Data

Taxon	Hierarchical Levels	References
Blattaria (entire order)	Superfamilies, families	Kambhampati (1995)
Coleoptera		
Gonioctena (Chrysomelidae)	Species	Mardulyn, Milinkovitch, and Pasteels (1997)
Ophraella (Chrysomelidae)	Species	Funk et al. (1995)
Cicindelidae	Species, Populations	Vogler and DeSalle (1993); Vogler et al. (1993*a,* 1993*b*)
Diptera		
Drosophilidae	Subgenera, species groups, species	DeSalle (1992*a,* 1992*b*); DeSalle et al. (1987)
Simuliidae	Sibling species	Xiong and Kocher (1993*a,* 1993*b*)
Homoptera		
Cicadellidae	Genera	Fang et al. (1993)
Hymenoptera		
Apis (Apidae)	Species	Cameron (1991); Cameron et al. (1992); Engel and Schultz (1997)
Microgastrine Braconidae	Genera	Mardulyn and Whitfield (1998)
Apidae	Tribes, genera	Cameron (1993)
Entire order	Superfamilies, families	Derr et al. (1992*a,* 1992*b*); Dowton and Austin (1994, 1997*a,* 1997*b*)
Proctotrupomorpha, Evaniomorpha	Families	Dowton et al. (1997)
Microgastroid Braconidae	Subfamilies	Whitfield (1997)
Braconidae	Subfamilies	Dowton, Austin, and Antolin (1998)
Lepidoptera		
Spodoptera and other Noctuidae	Populations, some species	Pashley and Ke (1992)
Orthoptera		
Entire order	Suborders, superfamilies	Flook and Rowell (1997*b*)
Caelifera	Superfamilies, families	Flook and Rowell (1997*a*)

Whitfield 1997). These are the goals of the analyses reported below.

Materials and Methods

Sources of Sequence Data

We examined 16S sequences that originated from five different phylogenetic studies: (1) an analysis of subfamily relationships within the hymenopteran lineage of microgastroid Braconidae (Whitfield 1997); (2) an analysis of relationships among the four tribes of corbiculate bees within the family Apidae *sensu* Roig-Alsina and Michener (1993) (Cameron 1993); (3) an analysis of hymenopteran relationships focusing on the superfamilies Evanioidea and Proctotrupoidea s. l. (Dowton et al. 1997); (4) a preliminary analysis of the phylogenetic utility of the 16S gene in the order Hymenoptera (Derr et al. 1992*a,* 1992*b*); and (5) a survey of hymenopteran relationships, with a special focus on the nonsawfly taxa or Apocrita (Dowton and Austin 1994). Our analyses utilize a 434-bp portion of the 3′ end of the 16S gene that was shared among each of the data sets from these studies. This portion of the gene corresponds to positions 13470–13894 in *Apis mellifera* (Crozier and Crozier 1993). A list of all 65 taxa examined in this analysis, along with their current classifications, GenBank accession numbers (when available), and source references, is provided in table 2.

Sequence Alignment

All sequences were entered unaligned into SeqApp, version 1.9a (Gilbert 1993), and checked for accuracy. Sequences from two braconid wasps (fig. 1*a* and *b*) and a bumble bee were fitted by hand to the 16S secondary-structure model of Gutell (1993). These, in turn, served as templates for aligning sequences within the superfamilies Ichneumonoidea and Apoidea, respectively, using CLUSTAL W (Thompson, Higgins, and Gibson 1994). These two blocks of aligned sequences within superfamilies were then aligned to one another. Finally, sequences from the remaining taxa were aligned to this set of aligned sequences using CLUSTAL W. A completely automated alignment of all 65 taxa using the default parameters within CLUSTAL W was used in a previous set of hierarchical analyses (SAC, 1996 meeting of the Society for the Study of Evolution/Society of Systematic Biologists). The automated alignment procedure resulted in general patterns of variability in the data that were virtually identical to those reported here. However, excluding considerations of secondary structure has been found to be less effective for phylogeny estimation for these taxa (Whitfield 1997). The patterns observed below account for secondary structure.

Hierarchical Comparisons

The taxa used at each taxonomic (hierarchical) level of comparison are given in table 3. Exemplars were selected so that each taxon is represented only once at the next lowest hierarchical level. For instance, to compare divergences at the level of genera within subfamilies, one exemplar was selected from each of one or more genera within a subfamily to represent that genus in the calculation of pairwise matrices. This procedure

Table 2
Taxa Examined in this Analysis, Along with Their Current Classifications and Sources

Superfamily	Family	Subfamily	Genus	Species	GenBank Accession No.	Source
Tenthredinoidea						
....	Tenthredinidae	Undetermined	Undetermined	Undetermined	Not submitted	Derr et al. (1992*a*)
	Pergidae	Perginae	*Perga*	*condei*	U06953	Dowton and Austin (1994)
	Pergidae	Phylacteophaginae	*Phylacteophaga*	*froggattii*	U06954	Dowton and Austin (1994)
Cephoidea	Cephidae	Cephinae	*Hartigia*	*trimaculata*	U06955	Dowton and Austin (1994)
Orussoidea	Orussidae	Orussinae	*Orussus*	*terminalis*	U06956	Dowton and Austin (1994)
Evanioidea	Evaniidae	Evaniinae	*Evania*	Undetermined	U06975	Dowton and Austin (1994)
	Gasteruptiidae	Hyptiogastrinae	*Eufoenus*	Undetermined	U06972	Dowton and Austin (1994)
	Gasteruptiidae	Gasteruptiinae	*Gasteruption*	Undetermined	U06974	Dowton and Austin (1994)
Trigonalyoidea ..	Trigonalyidae	Trigonalyinae	*Orthogonalys*	*pulchella*	U06973	Dowton and Austin (1994)
	Trigonalyidae	Trigonalyinae	*Poeciligonalys*	*costalis*	U06971	Dowton and Austin (1994)
Megalyroidea ...	Megalyridae	Megalyrinae	*Megalyra*	Undetermined	U39955	Dowton et al. (1997)
Ceraphronoidea						
....	Ceraphronidae	Ceraphroninae	*Aphanogmus*	Undetermined	U39949	Dowton et al. (1997)
	Megaspilidae	Megaspilinae	*Conostigmus*	Undetermined	U39951	Dowton et al. (1997)
Proctotrupoidea						
....	Pelecinidae	Pelecininae	*Pelecinus*	*polyturator*	U39956	Dowton et al. (1997)
	Proctotrupidae	Proctotrupinae	*Codrus*	Undetermined	U39950	Dowton et al. (1997)
	Proctotrupidae	Proctotrupinae	*Disogmus*	*areolator*	U39953	Dowton et al. (1997)
	Vanhorniidae	Vanhorniinae	*Vanhornia*	*eucnemidarum*	U06969	Dowton and Austin (1994)
	Roproniidae	Roproniinae	*Ropronia*	*garmani*	U06968	Dowton and Austin (1994)
	Heloridae	Helorinae	*Helorus*	Undetermined	U39954	Dowton et al. (1997)
	Diapriidae	Ambositrinae	*Diphoropria*	Undetermined	U39952	Dowton et al. (1997)
	Diapriidae	Diapriinae	*Spilomicrus*	Undetermined	U39957	Dowton et al. (1997)
Platygastroidea						
....	Scelionidae	Scelioninae	*Scelio*	*fulgidus*	U06964	Dowton and Austin (1994)
	Scelionidae	Telonominae	*Trissolcus*	*basalis*	U06962	Dowton and Austin (1994)
Cynipoidea	Figitidae	Anacharitinae	*Anacharis*	*zealandica*	U39948	Dowton et al. (1997)
	Ibaliidae	Ibaliinae	*Ibalia*	*leucospoides*	U06970	Dowton and Austin (1994)
Chalcidoidea	Aphelinidae	Aphelininae	*Aphytis*	*melinus*	U06965	Dowton and Austin (1994)
	Aphelinidae	Coccophaginae	*Encarsia*	*formosa*	U06966	Dowton and Austin (1994)
	Pteromalidae	Pteromalinae	*Pteromalus*	*puparum*	U06967	Dowton and Austin (1994)
Ichneumonoidea						
....	Ichneumonidae	Ichneumoninae	*Ichneumon*	*promissorius*	U06960	Dowton and Austin (1994)
	Ichneumonidae	Campopleginae	*Venturia*	*canescens*	U06961	Dowton and Austin (1994)
	Ichneumonidae	Pimplinae	*Xanthopimpla*	*stemmator*	Not submitted	Derr et al. (1992*a*, 1992*b*)
	Braconidae	Braconinae	*Digonogastra*	*kimballi*	Not submitted	Derr et al. (1992*a*, 1992*b*)
	Braconidae	Braconinae	*Bracon*	*hebetor*	U68145	Whitfield (1997)
	Braconidae	Ichneutinae	*Paroligoneurus*	Undetermined	U68148	Whitfield (1997)
	Braconidae	Agathidinae	*Alabagrus*	*stigma*	Not submitted	Derr et al. (1992*a*, 1992*b*)
	Braconidae	Meteorinae	*Meteorus*	*pulchricornis*	U68146	Whitfield (1997)
	Braconidae	Neoneurinae	*Neoneurus*	*mantis*	U68147	Whitfield (1997)
	Braconidae	Cheloninae	*Chelonus*	Undetermined	U68150	Whitfield (1997)
	Braconidae	Cheloninae	*Asogaster*	*argenitifrons*	U68145	Whitfield (1997)
	Braconidae	Cardiochilinae	*Taxoneuron*	*nigriceps*	U69151	Whitfield (1997)
	Braconidae	Miracinae	*Mirax*	*lithocolletidis*	U68152	Whitfield (1997)
	Braconidae	Microgastrinae	*Pholetesor*	*bedelliae*	U68153	Whitfield (1997)
	Braconidae	Microgastrinae	*Microgaster*	*canadensis*	U68154	Whitfield (1997)
	Braconidae	Microgastrinae	*Microplitis*	Undetermined	U68155	Whitfield (1997)
	Braconidae	Microgastrinae	*Cotesia*	*autographae*	U68156	Whitfield (1997)
	Braconidae	Microgastrinae	*Cotesia*	*congregata*	U68157	Whitfield (1997)
	Braconidae	Microgastrinae	*Cotesia*	*glomerata*	U06958	Dowton and Austin (1994)
	Braconidae	Microgastrinae	*Cotesia*	*orobenae*	U68158	Whitfield (1997)
	Braconidae	Microgastrinae	*Cotesia*	*rubecula*	U06959	Dowton and Austin (1994)
Vespoidea	Vespidae	Polistinae	*Polistes*	*versicolor*	Not submitted	Derr et al. (1992*a*)
	Formicidae	Myrmeciinae	*Myrmecia*	*forficata*	U06963	Dowton and Austin (1994)
Apoidea	Apidae	Apinae	*Xylocopa*	*virginica*	L22905	Cameron (1993)
	Apidae	Apinae	*Eufriesea*	*caerulescens*	L22904	Cameron (1993)
	Apidae	Apinae	*Eulaema*	*polychroma*	L22903	Cameron (1993)
	Apidae	Apinae	*Bombus*	*avinoviellus*	L22897	Cameron (1993)
	Apidae	Apinae	*Bombus*	*pennsylvanicus*	L22896	Cameron (1993)
	Apidae	Apinae	*Melipona*	*compressipes*	L22899	Cameron (1993)
	Apidae	Apinae	*Scaptotrigona*	*luteipennis*	L22900	Cameron (1993)
	Apidae	Apinae	*Trigona*	*hypogaea*	L22901	Cameron (1993)
	Apidae	Apinae	*Trigona*	*pallens*	L22902	Cameron (1993)
	Apidae	Apinae	*Apis*	*cerana*	L22892	Cameron (1993)
	Apidae	Apinae	*Apis*	*dorsata*	L22893	Cameron (1993)
	Apidae	Apinae	*Apis*	*florea*	L22894	Cameron (1993)
	Apidae	Apinae	*Apis*	*koschevnikovi*	L22895	Cameron (1993)
	Apidae	Apinae	*Apis*	*mellifera*	L22891	Cameron (1993)

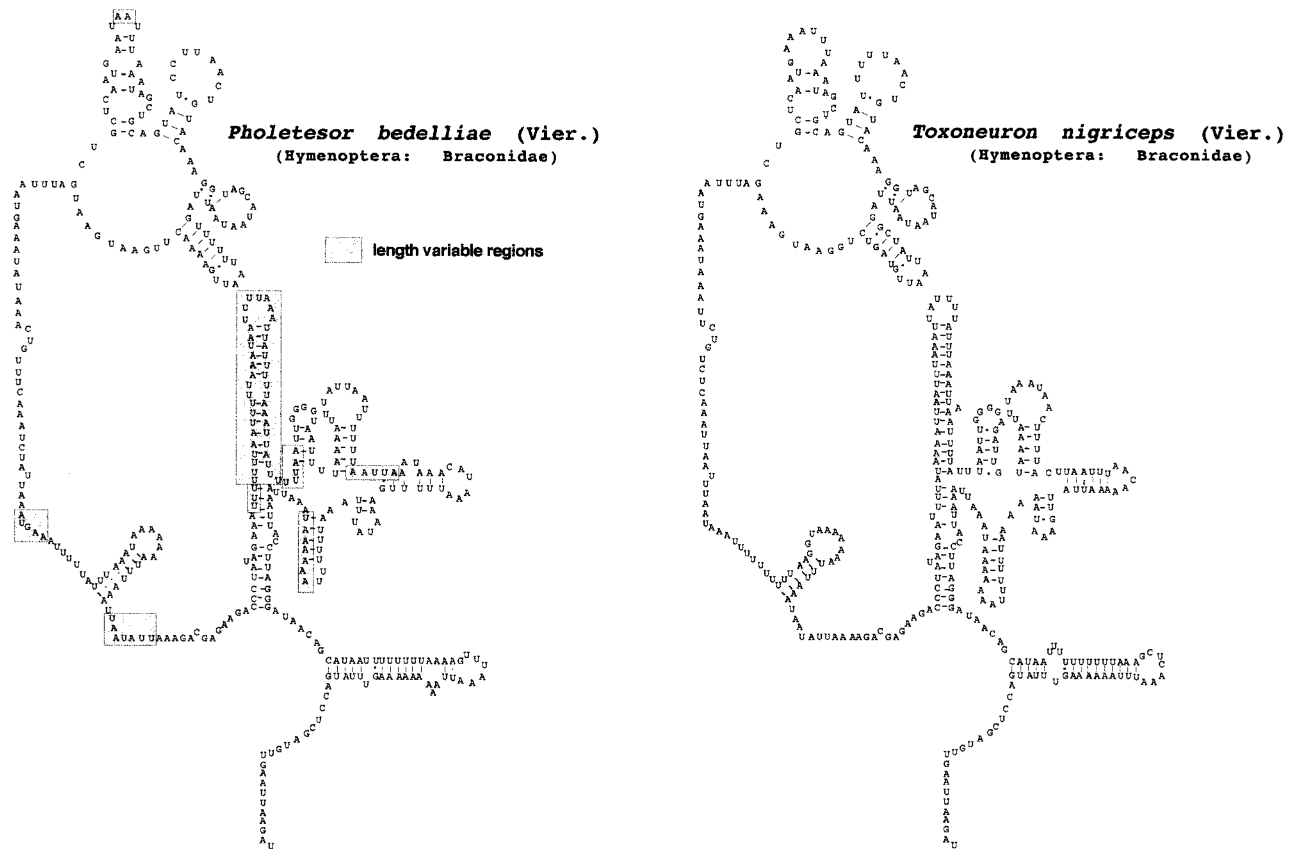

FIG. 1.—Secondary structures for portions of the 16S gene analyzed for two braconid wasps in different subfamilies, *Pholetesor bedelliae* (Viereck) and *Toxoneuron nigriceps* (Viereck), as fitted to the structural model of Gutell (1993).

eliminated unnecessary duplication of divergence calculations from sets of closely related species to other sets of closely related species, while still maintaining a relatively large sample size of comparisons. Initial calculations using the entire pairwise matrix indicated that choice of exemplar taxon had little effect on the results.

Obtaining Sequence Divergence Data and Statistics

Aligned sequences were reformatted appropriately and entered into MEGA, version 1.01 (Kumar, Tamura, and Nei 1993), for calculation of sequence statistics and measures of sequence divergence from pairwise matrices. For each pairwise comparison, we calculated the ti : tv ratio, the uncorrected sequence divergence (p-distance), and several corrected divergence estimates, including the Jukes and Cantor (1969), Kimura (1980) two-parameter, Tajima and Nei (1984), and Tamura and Nei (1993) models. In the Tamura-Nei divergence estimates, a gamma distribution was assumed using a shape parameter of 0.4, consistent with (but slightly higher than) estimates of the gamma shape parameter reported by Yang (1996) for insect 16S rRNA. Estimates of sequence divergence using the Hasegawa, Kishino, and Yano (1985; subsequently referred to as HKY85) and General Time Reversible (Yang 1994; subsequently referred to as GTR) models were obtained using PAUP* 4.0 beta version b1 (Swofford 1998), with the estimated parameters obtained as described in the following paragraph.

More precise estimates of site-to-site rate variation were obtained by applying ML analysis (PAUP* test version 4.0d54; Swofford 1997) to a relatively well corroborated reference tree topology for the 65 hymenopteran taxa (table 1; see discussion of the reference phylogeny below). A gamma distribution was initially inferred using four rate categories and the HKY85 model (Hasegawa, Kishino, and Yano 1985). The gamma shape parameter, the proportion of invariant sites, and ti : tv ratio were estimated from the data (with reference to the phylogeny), using the fast ML method (Rogers and Swofford 1998) to obtain starting branch lengths. Alternative methods for estimating site-to-site rate variation (e.g., Nielsen 1997) have been proposed, but these require divergence time or branch length information that is not currently available for Hymenoptera.

Using the reference phylogeny to consider subclades at several hierarchical levels, we implemented MacClade, version 3.06 (updated version of Maddison and Maddison 1992), to obtain site-to-site estimates of the actual number of changes inferred (results summarized in fig. 7). MacClade was also used with the reference phylogeny to obtain graphical depictions of the frequencies of changes from one nucleotide to another (results summarized in fig. 4).

The Reference Phylogeny

The higher-level phylogeny of Hymenoptera, as estimated from paleontological, morphological, and mo-

Table 3
Comparisons Used for Hierarchical Divergence Estimates

Among species within genera (n = 22 pairwise comparisons)
- Within *Apis*: *A. mellifera, A. cerana, A. dorsata, A. koschevnikovi, A. florea*
- Within *Bombus*: *B. avinoviellus, B. pennsylvanicus*
- Within *Cotesia*: *C. autographae, C. congregata, C. glomerata, C. orobenae, C. rubecula*
- Within *Trigona*: *T. hypogaea, T. pallens*

Among genera within subfamilies (n = 38 pairwise comparisons)
- Within Apinae: *Apis* (*mellifera*), *Bombus* (pennsylvanicus), *Eufriesea* (*caerulescens*), *Eulaema* (*polychroma*), *Trigona* (*hypogaea*), *Scaptotrigona* (*luteipennis*), *Melipona* (*compressipes*), *Xylocopa* (*virginica*)
- Within Braconinae: *Bracon* (*hebetor*), *Digonagastra* (*kimballi*)
- Within Cheloninae: *Chelonus* (sp.), *Ascogaster argentifrons*
- Within Microgastrinae: *Microplitis* (sp.), *Microgaster* (*canadensis*), *Cotesia* (*glomerata*), *Pholetesor* (*bedelliae*)
- Within Proctotrupinae: *Codrus* (sp.), *Disogmus* (*areolator*)
- Within Trigonalyinae: *Orthogonalys* (*pulchella*), *Poecilogonalys* (*costalis*)

Among subfamilies within families (n = 43 pairwise comparisons)
- Within Aphelinidae: *Encarsia formosa* (Coccophaginae), *Aphytis melinus* (Aphelininae)
- Within Braconidae: *Alabagrus stigma* (Agathidinae), *Ascogaster argentifrons* (Cheloninae), *Bracon hebetor* (Braconinae), *Cotesia glomerata* (Microgastrinae), *Meteorus pulchricornis* (Meteorinae), *Mirax lithocolletidis* (Miracinae), *Neoneurus mantis* (Neoneurinae), *Paroligoneurus* sp. (Ichneutinae), *Toxoneuron nigriceps* (Cardiochilinae)
- Within Diapriidae: *Diphoropria* sp. (Ambositrinae), *Spilomicrus* (Diapriinae)
- Within Gasteruptiidae: *Gasteruption* sp. (Gasteruptiinae), *Eufoenus* sp. (Hyptiogastrinae)
- Within Ichneumonidae: *Ichneumon promissorius* (Ichneumoninae), *Venturia canescens* (Campopleginae), *Xanthopimpla stemmator* (Pimplinae)
- Within Scelionidae: *Scelio fulgidus* (Scelioninae), *Trissolcus basalis* (Telonominae)

Among families within superfamilies (n = 21 pairwise comparisons)
- Within Ceraphronoidea: *Aphanogmus* sp. (Ceraphronidae), *Conostigmus* sp. (Megaspilidae)
- Within Chalcidoidea: *Aphytis melinus* (Aphelinidae), *Pteromalus puparum* (Pteromalidae)
- Within Cynipoidea: *Anacharis zealandica* (Figitidae), *Ibalia leucospoides* (Ibaliidae)
- Within Evanioidea: *Evania* sp. (Evaniidae), *Gasteruption* sp. (Gasteruptiidae)
- Within Proctotrupoidea: *Codrus* sp. (Proctotrupidae), *Helorus* sp. (Heloridae), *Pelecinus polyturator* (Pelecinidae), *Roproniа garmani* (Roproniidae), *Spilomicrus* sp. (Diapriidae), *Vanhornia eucnemidarum* (Vanhorniidae)
- Within Tenthredinoidea: Tenthredinidae sp., *Perga condei* (Pergidae)
- Within Vespoidea: *Myrmecia forficata* (Formicidae), *Polistes versicolor* (Vespidae)

Among superfamilies within Hymenoptera (n = 91 pairwise comparisons)
- Representing Apoidea: *Apis mellifera*
- Representing Cephoidea: *Hartigia trimaculata*
- Representing Ceraphronoidea: *Aphanogmus* sp.
- Representing Chalcidoidea: *Pteromalus puparum*
- Representing Cynipoidea: *Ibalia leucospoides*
- Representing Evanioidea: *Evania* sp.
- Representing Ichneumonoidea: *Ichneumon promissorius*
- Representing Megalyroidea: *Megalyra* sp.
- Representing Orussoidea: *Orussus terminalis*
- Representing Platygastroidea: *Scelio fulgidus*
- Representing Proctotrupoidea: *Codrus* sp.
- Representing Tenthredinoidea: *Phylacteophaga frogattii*
- Representing Trigonalyoidea: *Orthogonalys pulchella*
- Representing Vespoidea: *Polistes versicolor*

lecular data, was recently reviewed by Whitfield (1998). Using the consensus phylogeny from that review as a foundation, we constructed a composite phylogeny (fig. 2) using results from Cameron (1993) for Apidae, Whitfield (1997) for Braconidae, and Dowton and Austin (1994) and Dowton et al. (1997) for some family- and superfamily-level hymenopteran relationships. Relationships among exemplar taxa from these studies were used to reconstruct the tips of the tree. Because this reference phylogeny is based on data from multiple sources, it is likely to be relatively accurate, although it could differ in some minor details from a maximum-parsimony tree estimated in an actual combined analysis. A combined analysis is not possible at the present time, because major differences in taxon representation exist between studies. However, the reference topology (fig. 2) has the advantage of being largely corroborated by both molecular and morphological data. Therefore, rate parameters estimated with reference to this phylogeny should be relatively robust.

Testing the Estimated Parameters in Phylogeny Estimation

To determine the effects of the estimated ML parameters and their ability to recover the correct tree relative to parsimony analysis, two reduced sets of exemplar taxa were selected. One comprised species of bees within the tribe Apini, and the other comprised superfamily representatives. These were selected because relationships among the taxa have been well-corroborated from multiple studies (apine bees: Alexander 1991; Cameron 1991, 1993; Engel and Schultz 1997; superfamilies: see Whitfield 1998 for a review of the hymenopteran superfamily relationships based on molecular, morphological, and fossil data). "Expected" phylogenies could thus be specified for these well-corroborated groups. The 16S data were subjected to equally weighted maximum-parsimony analysis and two ML analyses: one using the HKY85 model (Hasegawa, Kishino, and Yano 1985) assuming a gamma distribution of among-site rate variation estimated from the empirical base frequencies and using estimates of the shape parameter and proportion of invariant sites from the analyses described above in *Obtaining Sequence Divergence Data and Statistics)*; and one using the GTR (Yang 1994) model (same site-to-site rate variation assumptions) after estimation of the general rate matrix in an initial run on the entire data set. Each analysis was run as a branch-and-bound search and repeated as a bootstrap analysis (heuristic search, 400 replications) using PAUP*. The percentage of clades correct (Hillis, Huelsenbeck, and Cunningham 1994) and the slightly more sensitive bootstrapped percentage of clades correct (Cunningham 1997) were calculated as measures of the ability to recover the expected phylogenies. Likelihood ratio tests (Huelsenbeck and Rannala 1997) were conducted on a series of analyses to determine which of the estimated parameters (alone or in combination) significantly improved the ML estimation.

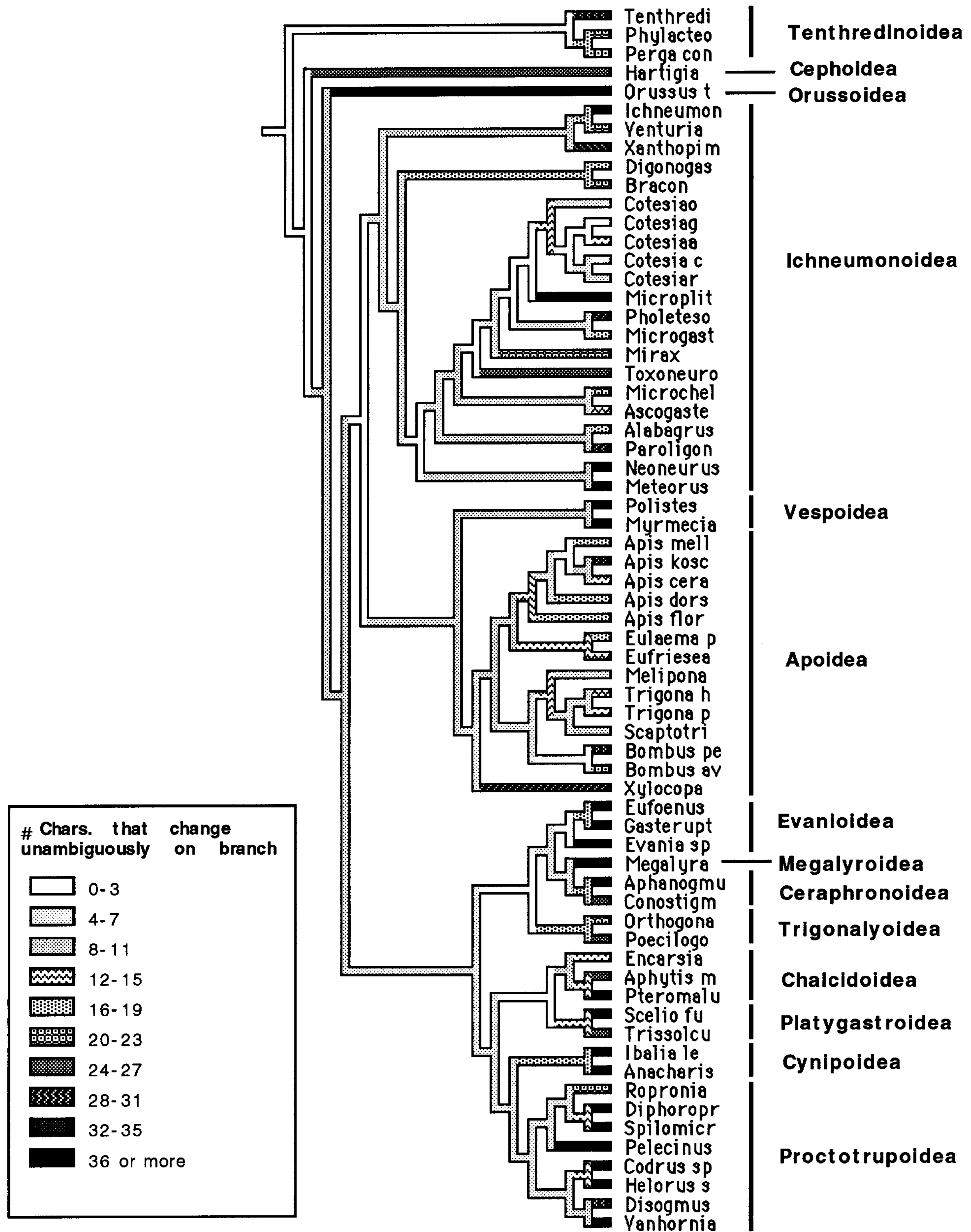

FIG. 2.—The reference phylogeny used to estimate the number of evolutionary changes. See text for origin of this phylogeny, and table 2 for complete names and classifications of the taxa.

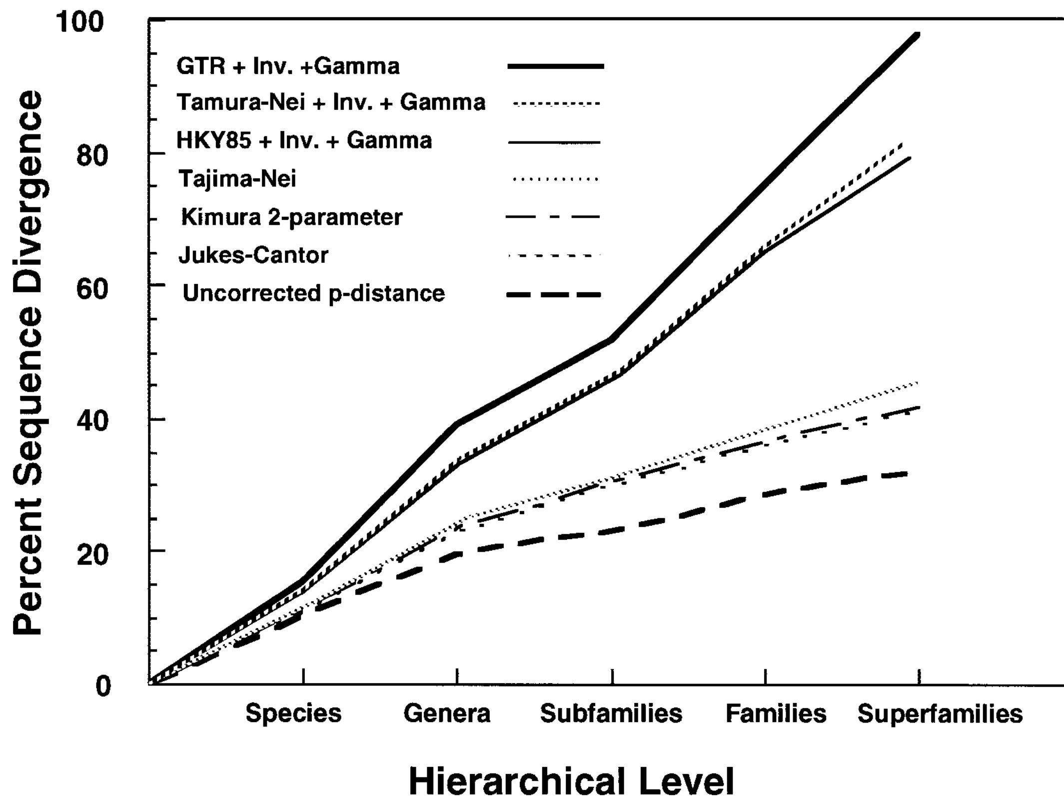

FIG. 3.—Sequence divergence, uncorrected and corrected using the models of Jukes and Cantor (1969), Tajima and Nei (1984), Kimura (1980), Tamura and Nei (1993), HKY85 (Hasegawa, Kishino, and Yano 1985), and GTR (Yang 1994) plotted against taxonomic level. See text for further explanation and interpretation of the specific models used.

Results

Nucleotide Divergences at Various Taxonomic Levels

Figure 3 depicts the mean pairwise percentages of nucleotide divergence among taxa at five hierarchical levels (species, genera, subfamilies, families, and superfamilies). The data are represented as uncorrected (raw p-distance) and corrected, applying the Jukes and Cantor (1969), Kimura (1980) two-parameter, Tajima and Nei (1984), Tamura and Nei (1993), HKY85, and GTR (Yang 1994) substitution models. All of these correction methods correct, to some degree, for multiple nucleotide replacements (saturation) at a site. The Jukes-Cantor model is the simplest in assuming equal base frequencies, ti : tv substitution rates, and no site-to-site rate variation. Correcting for transition bias with the Kimura two-parameter model resulted in divergence estimates virtually indistinguishable from those based on the Jukes-Cantor model, indicating that transition bias has relatively little effect on these data (see also below). Correcting for base composition bias using the Tajima-Nei model results in a small but distinguishable increase in estimated level of divergence at higher taxonomic levels. This increase is small despite the relatively strong AT bias in Hymenoptera. Applying the Tamura-Nei assumption of gamma-distributed (shape parameter α = 0.4) rates across sites and estimating the proportion of invariant sites from the data results in major increases in the estimated divergences at all levels above species. Finally, the HKY85 and GTR models, using the gamma shape parameter and proportion of invariant sites estimated from the data, also resulted in major increases in estimated divergence (with the HKY85 estimates strongly resembling the Tamura-Nei estimates). Golding (1983) noted that failure to account for site-to-site rate variation (when it is substantial) can result in an underestimation of the actual number of substitutions, clearly an influential factor with our 16S data. The estimated divergence levels among families and superfamilies are extremely high, nearing or exceeding 100%, clearly the result of superimposed changes at highly variable sites.

Nucleotide Composition Bias

It has previously been noted that the mtDNA of insects in general (Simon et al. 1994), and Hymenoptera in particular (Dowton and Austin 1997*a,* 1997*b*), exhibits a significantly larger proportion of A and T nucleotides as compared with C and G. Our findings from the large hymenopteran data set confirm these reports (table 4) both in magnitude and direction of the bias. AT content is highest in groups considered to be relatively recently diverged in the hymenopteran phylogeny (bees, chalcidoids, scelionids, and some endoparasitoid braconids). The base composition bias is obviously reflected in the substitution bias toward A's and T's at different hierarchical levels (fig. 4).

Transition/Transversion Bias

The uncorrected ti : tv ratios for Hymenoptera (fig. 5) are unusually low (especially for species-level com-

Table 4
Mean Percentages of A's and T's for 16S data from Hymenopteran Taxa

Superfamily	*N*	Mean % A+T
Tenthredinoidea	3	77.4
Cephoidea	1	78.5
Orussoidea	1	75.1
Evanioidea	3	80.6
Trigonalyoidea	2	82.1
Megalyroidea	1	82.1
Ceraphronoidea	2	81.4
Proctotrupoidea	8	83.1
Platygastroidea	2	85.9
Cynipoidea	2	83.6
Chalcidoidea	3	83.6
Ichneumonoidea	21	83.4
Vespoidea	2	81.5
Apoidea	14	81.2
Total	65	82.2

parisons) and *increase* with divergence time, contrary to expectations (Wakeley 1996). It is possible that we lack sufficient comparisons at the population and closely related species levels to detect the characteristic dominance of transitions at low taxonomic levels, but there is no indication that transitions predominate at any level in our comparisons (fig. 6). The overall ti : tv ratio estimated for all sequences using ML is also unusually low (0.28).

Site-to-Site Rate Variation

In our initial estimates of corrected nucleotide divergence among taxa at different taxonomic levels (fig. 3), we employed the Tamura-Nei model and assumed a gamma distribution of rates across sites and a shape parameter (α) of 0.4. The shape parameter value was relatively close to the value of 0.31 estimated from 16S data for 17 eukaryotes (Yang and Kumar 1996). Our subsequent estimate of α, using ML in conjunction with the reference phylogeny (fig. 2), resulted in the considerably higher value of 0.8728. The estimated proportion of invariant sites was 0.1281.

Length-Variable Regions

Several regions of the 434-bp fragment of the 16S molecule exhibit considerable length variation (indels), especially in higher-taxon comparisons. These length-variable regions (fig. 1) do not consistently correspond to any specific structural features of the molecule (e.g., loop regions), but do occur in the same positions across taxa within the Hymenoptera. They also correspond to those regions of figure 7 with the largest numbers of estimated changes. Some of these variable regions can be reconciled with the highly variable regions found in Orthoptera (Flook and Rowell 1997*a,* 1997*b*).

Hierarchical Accumulation of Variation

Figure 7 shows the estimated number of changes at each site within the 434-bp region for several taxonomic levels, optimized onto the reference phylogeny using maximum parsimony. From this, it is clear that the most highly variable sites are clumped in distribution and vary in magnitude at different taxonomic levels. For example, while variability may be relatively low for comparisons among subfamilies, it becomes enormous when different superfamilies are compared.

Efficacy of the Estimated Parameters in Phylogeny Estimation

Figures 8 and 9 depict results of analyses designed to assess which of three approaches comes closer to recovering the expected phylogenies (figs. 8*A* and 9*A*): (1) unweighted parsimony, (2) ML under the HKY85 model

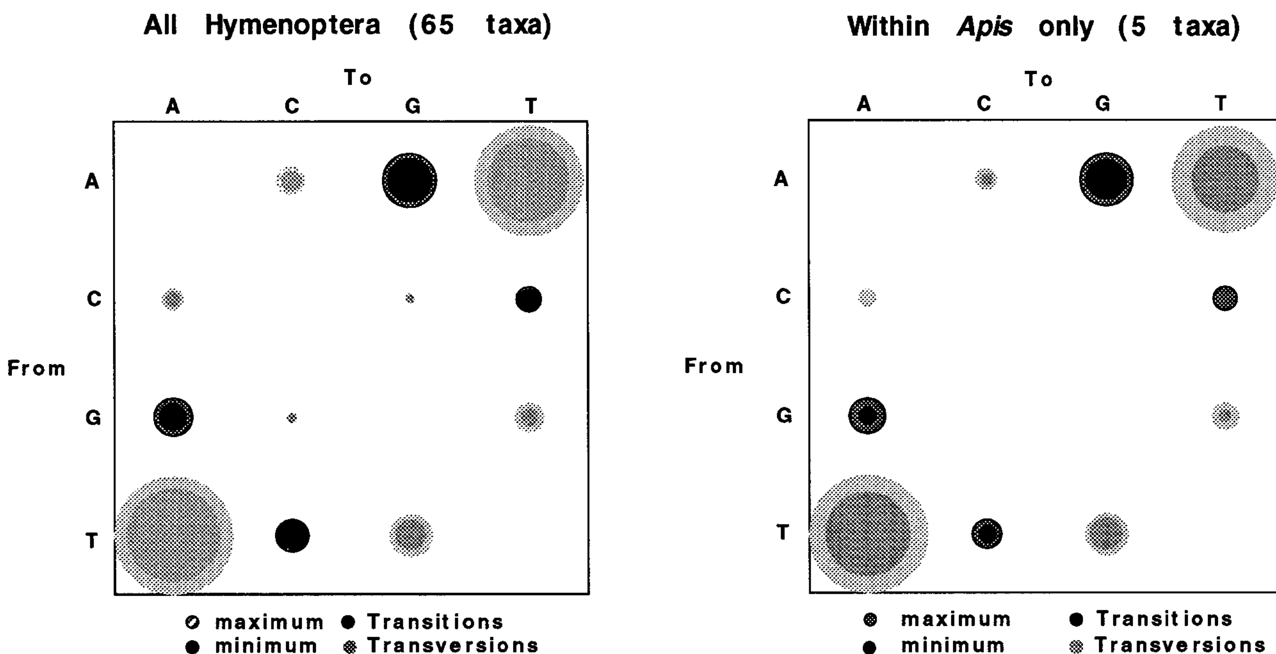

FIG. 4.—Relative frequency of types of base changes, as estimated under parsimony assumptions using MacClade, version 3.06 (Maddison and Maddison 1992) and the reference phylogeny in figure 2.

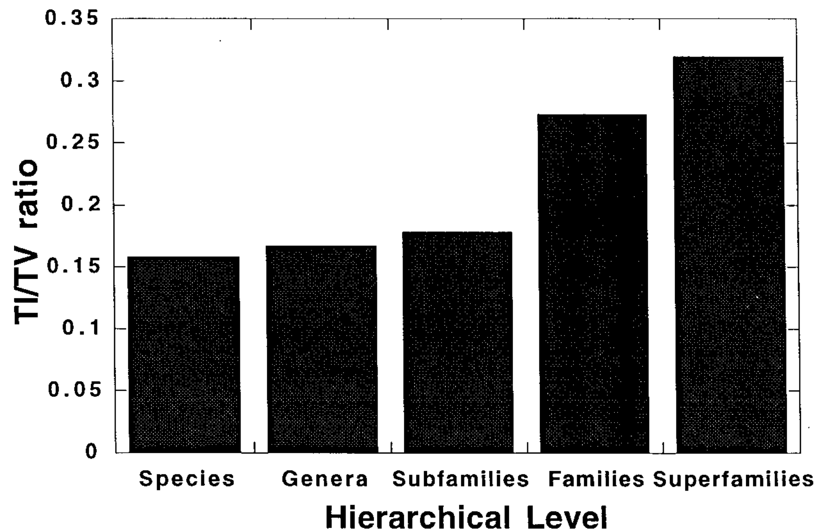

FIG. 5.—Uncorrected pairwise ti : tv ratio, plotted against taxonomic level of comparison within Hymenoptera.

of evolutionary change incorporating site-to site rate variation, or (3) ML under the GTR model (also similarly incorporating site-to-site rate variation). At both moderately low (fig. 8) and high (fig. 9) taxonomic levels, the more complex ML models incorporating our parameter estimates improve the ability of the analyses to recover the expected phylogeny (figs. 8*D* and 9*D*). However, at the higher taxonomic level, even the complex ML methods result in a tree with unacceptably low bootstrap support for most nodes (fig. 9*D*), suggesting that these data are not useful at the superfamily level or that taxon sampling needs to be more complete. Nevertheless, the use of additional data from other sources would be advisable to correctly estimate the entire phylogeny.

Likelihood ratio tests (Huelsenbeck and Rannala 1997) indicate that taking the empirical base frequencies into account significantly improves the fit of the ML model to the tree ($P \ll 0.01$), as do the models accounting for both base frequencies and site-to-site rate variation ($P \ll 0.01$). However, accounting for ti : tv ratio alone has little effect ($P > 0.05$).

Discussion

Unusual Features of Hymenopteran 16S Sequence Data

Hymenopteran mtDNA exhibits one the highest proportions of AT nucleotides of any organism yet measured (Cameron 1991, 1993; Cameron et al. 1992; Crozier and Crozier 1993; Simon et al. 1994; Dowton and Austin 1994, 1997*a,* 1997*b*; Whitfield 1997). A current hypothesis for this AT richness is that strand-specific compositional bias (a predominance of G→A transitions, perhaps the result of asymmetries in stem base-pairing capabilities) has led to an increase in A content, followed by an increase in T content (reviewed in Dowton and Austin 1997*b*). Dowton and Austin (1997*b*) showed that the AT content of 16S increases from the

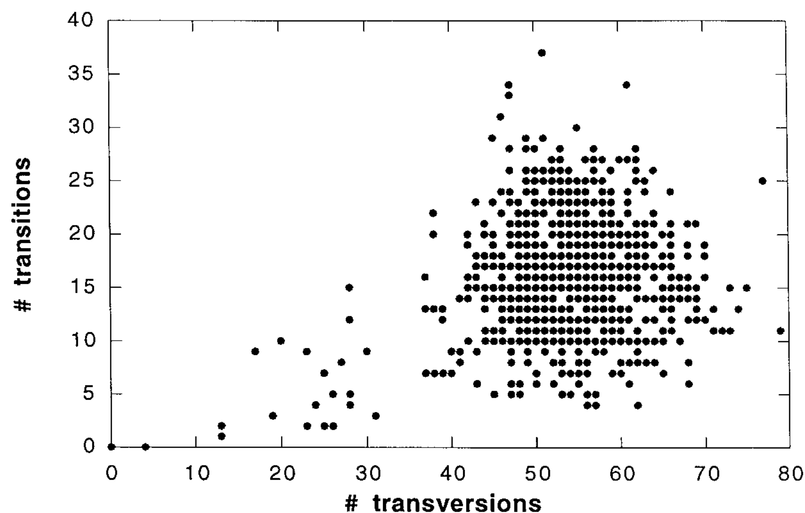

FIG. 6.—Observed number of transitions plotted against number of transversions from all pairwise comparisons within Hymenoptera. A large number of superimposed points are hidden in the cloud to the right.

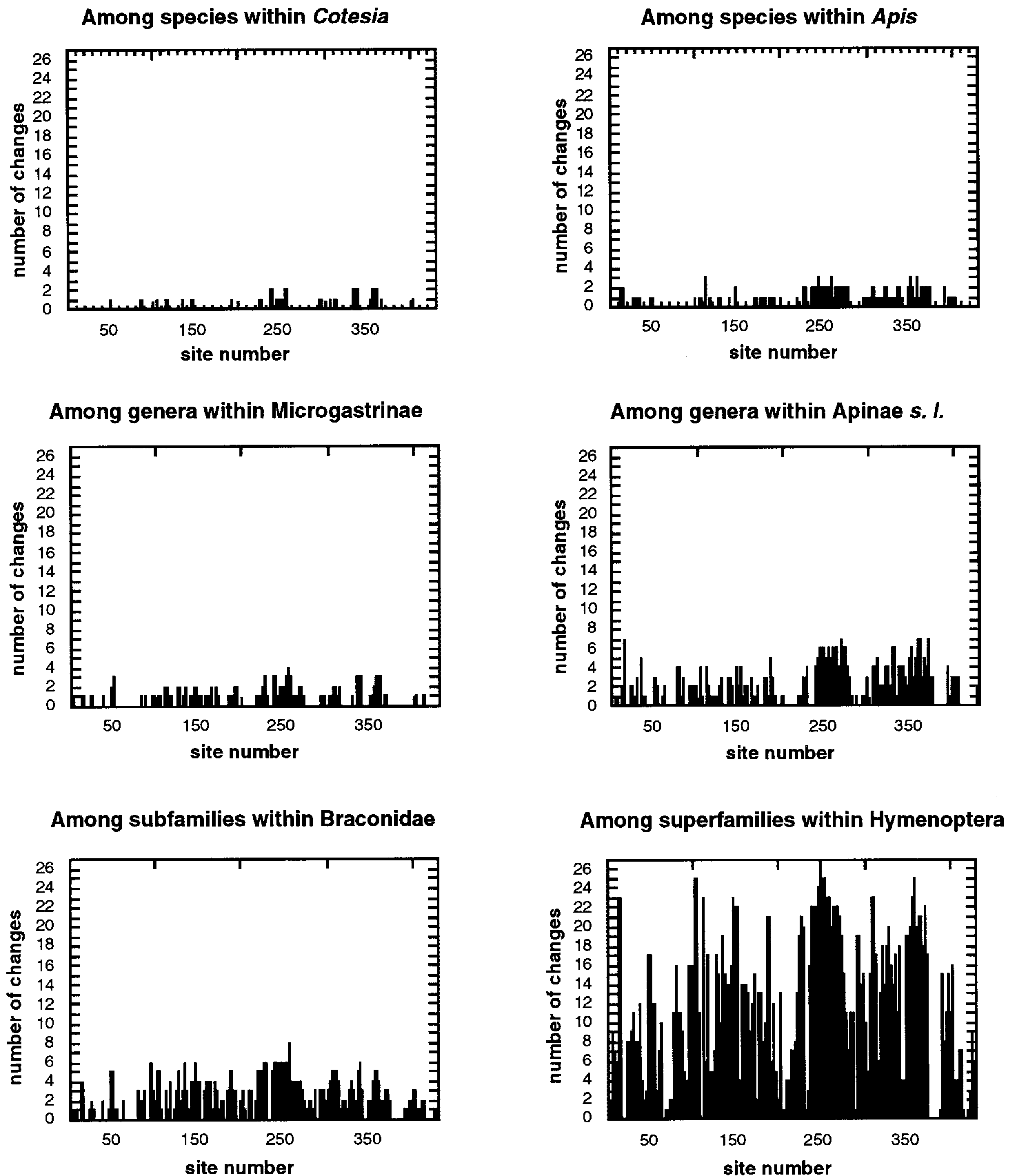

FIG. 7.—Number of changes by site at various taxonomic levels within Hymenoptera. The number of changes at each position was estimated using MacClade 3.06 (Maddison and Maddison 1992) and the reference phylogeny in figure 2.

base to the tips of the tree, suggesting that AT bias has continued to accumulate over time within lineages of Hymenoptera. Our data are consistent with their findings.

This high AT bias could explain, in part, the strikingly low ti : tv ratio observed at all taxonomic levels. Furthermore, the fact that the ti : tv ratio increases with increasing taxonomic levels (fig. 5) suggests that the AT bias changes among lineages (which it does to some degree; see table 4). Nonetheless, the ti : tv ratio is exceptionally low compared with those of other organisms, even after compensating for the AT bias using ML. Clearly, there must be other factors operating here which require further investigation.

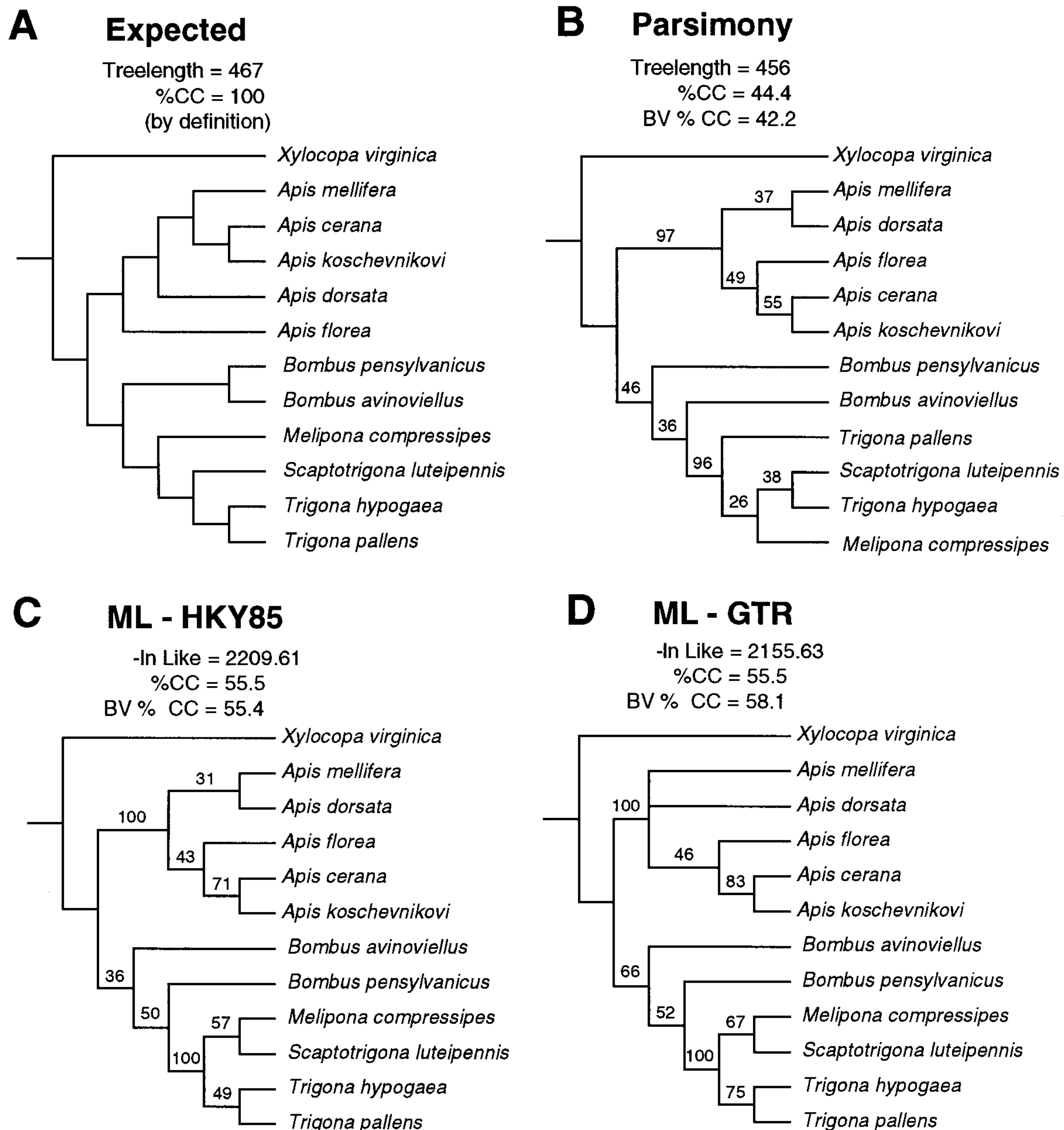

FIG. 8.—Performance of three phylogenetic estimation methods in recovering an expected phylogeny (*A*) for a subset of taxa (species, genera, and tribes of bees) in figure 2. The three methods compared are (*B*) unweighted maximum parsimony, (*C*) ML using the HKY85 (Hasegawa, Kishino, and Yano 1985) model with the parameters estimated in this study (base frequencies, ti : tv ratio, proportion of invariant sites, gamma shape parameter), and (*D*) ML using the GTR model (same site-to-site rate assumptions as in *C*). −ln Like = inverse log likelihood; % CC = percentage of clades correct (Hillis, Huelsenbeck, and Cunningham 1994); BV % CC = bootstrapped percentage of clades correct (Cunningham 1997). Numbers on branches represent bootstrap proportions (400 replications). Tree lengths and log likelihoods are based on the shortest trees from branch-and-bound searches using PAUP* (Swofford 1997–1998).

16S rRNA sequences in general possess unique patterns of length variation and site-to-site rate variation that are absent in sequences of protein-coding genes. These patterns are most conspicuous in comparisons of distantly related taxa. Moreover, our investigation reveals that the location of the most variable sites is consistent across a wide array of taxa. Knowledge of the locations of high variability reported here for 16S should be useful for future investigations of Hymenoptera and other insect groups.

Treatment of 16S Data in Phylogenetic Analysis

At least two aspects of AT-richness have important implications for phylogenetic analysis. First, it has the effect of reducing a majority of sites to two-state characters (A or T), thus increasing the potential for homoplasy. Second, branch lengths will tend to be underestimated unless the AT bias is considered. The simplest way to compensate for AT bias is to downweight AT transversions in parsimony analyses (Knight and Mindell 1993; Collins, Wimberger, and Naylor 1994; Dow-

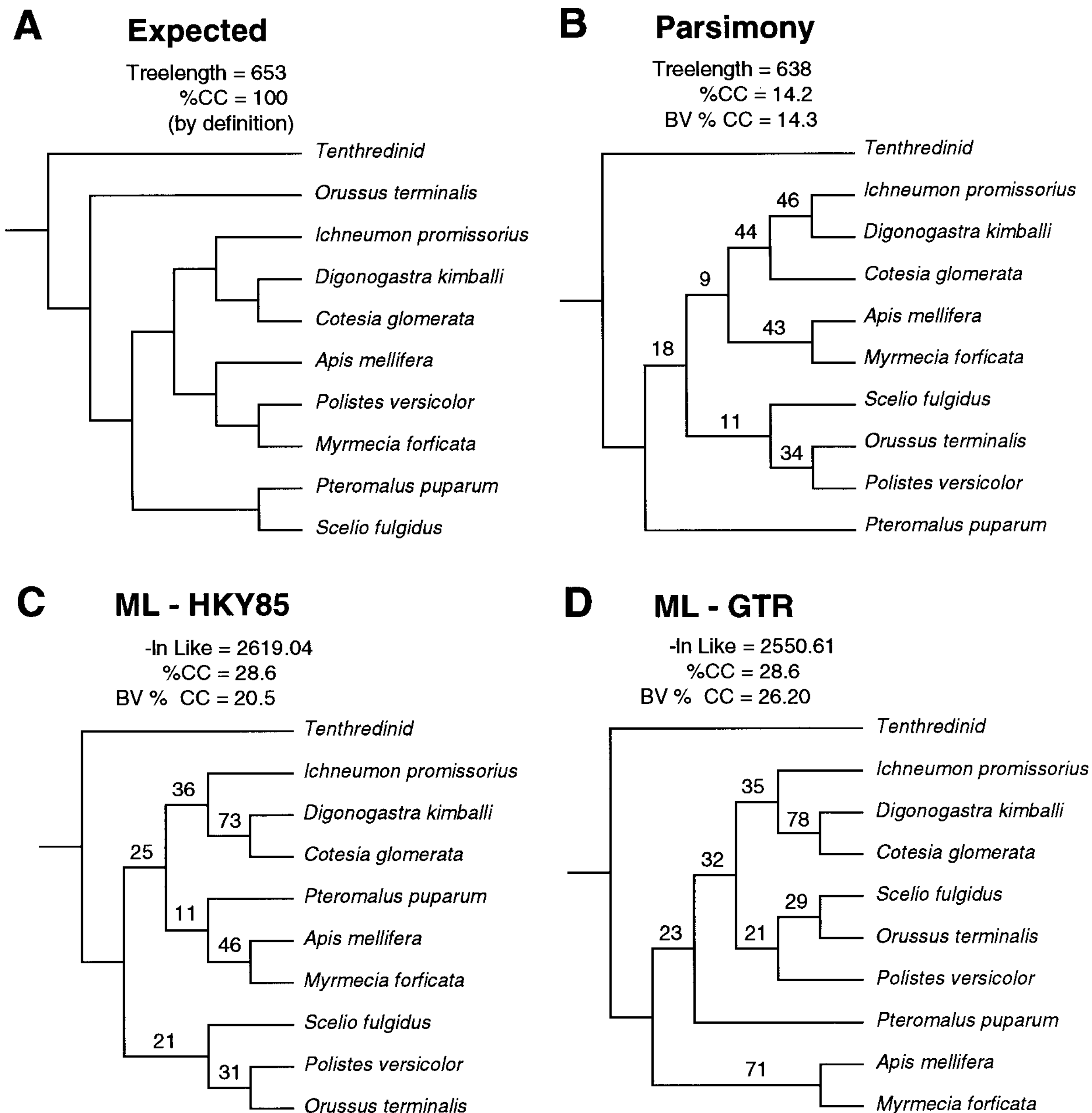

FIG. 9.—As figure 8, except that taxa are exemplars of superfamilies represented in figure 2.

ton and Austin 1997*a*). Differences among taxa in compositional bias can be compensated for by employing log-determinant (transformed) distances, as was done by Lockhart et al. (1994) for 16S data from honey bees. Although this method appears to successfully compensate for AT bias among lineages, we expect that a more informative general strategy will employ an ML approach, in which site-to-site rate variation can be considered simultaneously with compositional bias. The GTR (Yang 1994) model of sequence change appears to provide the best correction among those tried here, no doubt because it allows for AT transversions to be treated as an independent rate.

A number of phylogenetic studies have excluded the length-variable regions of 16S in some analyses (Cameron 1991; Dowton and Austin 1994; Dowton et al. 1997; Flook and Rowell 1997*a,* 1997*b*; Whitfield 1997), often with little effect on the outcome, although at higher levels, it appears that exclusion improves the signal : noise ratio (Flook and Rowell 1997*b*; Whitfield 1997). Alignment of length-variable regions to the 16S secondary structure (using the Gutell [1993] model) significantly increased the overall phylogenetic signal : noise ratio in at least one analysis (Whitfield 1997) and yielded a larger fraction of useful sequence data.

Recent work (reviewed in Yang 1996) has shown that when significant site-to-site rate variation exists in sequence data, it is important to account for this variation in order to obtain an accurate estimate of phylogeny. Rate variability across sites is more difficult to reconcile than compositional bias in phylogenetic analysis. Unfortunately, parsimony methods as currently imple-

mented do not deal effectively with rate variation across sites (Yang 1996). At present, ML methods incorporating the gamma distribution, as well as mixed-distribution models incorporating both gamma-distributed sites and estimates of the proportion of invariant sites, are the most effective and easily implemented methods to accommodate site-to-site rate variation.

Finally, it is evident at the level of distant family and superfamily comparisons that the number of estimated changes at many sites within 16S becomes so large as to render the gene phylogenetically useless at those taxonomic levels. Our analyses of the data at these levels (fig. 9*B–D*) indicates that the data are poor at recovering expected clades. Figure 7 provides a graphic depiction of why the gene cannot resolve basal divergences within the Hymenoptera (Dowton and Austin 1994).

In summary, our analyses of patterns of variation in 16S sequences of Hymenoptera suggest that the 16S gene may be useful for phylogenetic analysis across a relatively wide range of taxonomic levels, with the following caveats:

1. When using 16S sequences in phylogenetic analyses, the magnitude of base composition bias (typically AT bias) requires consideration, particularly for higher-taxon comparisons. One approach is to downweight AT transversions in parsimony analysis. However, compensation for this bias is probably best accomplished using ML methods, in which site-to-site rate variation can be simultaneously accommodated.
2. Site-to site variation in substitution rate is well documented for hymenopteran (and other animal) 16S sequence data, and should be incorporated into models for phylogenetic estimation, especially in studies of higher taxa (above the level of distantly related species). Attempts to estimate phylogeny using 16S data at higher taxonomic levels without considering site-to-site rate variation are likely to produce inaccurate estimations of branch lengths and perhaps even wrong topologies. We highly recommend estimating the shape parameter for the gamma distribution from the data, rather than employing previously published values, until a wider range of taxa have been investigated fully. Our data suggest a higher value for α than has previously been estimated for insects (0.87 compared with 0.3–0.4).
3. The 16S gene may be useful for estimating relationships among closely related species if a sufficient number of variable sites can be found. It clearly contains phylogenetically useful signal at the tribal/subfamily and close family levels. However, among genera and distantly related species groups, the highly variable sites appear to be saturated with substitutions, while too few of the conserved sites exhibit variation (Simon et al. 1994; Mardulyn and Whitfield 1998). Results of several phylogenetic studies suggest that the upper limit of utility for 16S is exceeded at the superfamily and subordinal levels (Dowton and Austin [1994] and Dowton et al. [1997] for Hymenoptera; Flook and Rowell [1997*a,* 1997*b*] for Orthoptera). This conclusion is strongly supported by our phylogenetic tests (especially fig. 9) and by the extremely high divergence levels among superfamilies (fig. 3).

Our analysis of the hierarchical utility of 16S nucleotide sequences for phylogeny estimation is somewhat limited in scope due to the intensive computational effort involved in summarizing such complex patterns. As future analyses continue to clarify patterns of variation within this and other genes, the task of matching the appropriate sequence data to specific evolutionary questions should become easier. At the same time, development of appropriate models of sequence change for each gene remains a critical step in phylogenetic analysis of DNA sequences. It is still a challenge to utilize such models in analyses incorporating multiple data sets, but this situation is likely to improve.

Acknowledgments

We thank Chris Simon, Karl Kjer, Mike Antolin, and Thomas Buckley for discussions about secondary structure and its incorporation into sequence alignment; Patrick Mardulyn, John Wakeley, and David Pollock for discussion of trends in our data and suggestions for their presentation; Paul Flook, Mark Dowton, and David Swofford for sharing manuscripts and preprints of relevant papers; and David Swofford for discussion of ML estimation of site-to-site rate heterogeneity and for prerelease versions of PAUP*. David Rand and two anonymous reviewers provided many useful comments on an earlier draft of the manuscript. This work was supported by NSF grants to J.B.W. (BSR-9111938) and S.A.C. (GER-9450117) and by USDA grant 9501893 to J.B.W.

LITERATURE CITED

ALEXANDER, B. A. 1991. Phylogenetic analysis of the genus *Apis* (Hymenoptera: Apidae). Ann. Entomol. Soc. Am. **84**: 137–149.

BROWER, A. V. Z. 1994. Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. Proc. Natl. Acad. Sci. USA **91**:6491–6495.

BROWER, A. V. Z., and R. DESALLE. 1994. Practical and theoretical considerations for choice of a DNA sequence region in insect molecular systematics, with a short review of published studies using nuclear gene regions. Ann. Entomol. Soc. Am. **87**:702–716.

CAMERON, S. A. 1991. A new tribal phylogeny of the Apidae inferred from mitochondrial DNA sequences. Pp. 71–87 *in* D. R. SMITH, ed. Diversity in the genus *Apis.* Westview Press, Boulder, Colo.

———. 1993. Multiple origins of advanced eusociality in bees inferred from mitochondrial DNA sequences. Proc. Natl. Acad. Sci. USA **90**:8687–8691.

CAMERON, S. A., J. N. DERR, A. D. AUSTIN, J. B. WOOLLEY, and R. A. WHARTON. 1992. The application of nucleotide sequence data to phylogeny of the Hymenoptera: a review. J. Hymenopt. Res. **1**:63–79.

CHO, S., A. MITCHELL, J. C. REGIER, C. MITTER, R. W. POOLE, T. P. FRIEDLANDER, and S. ZHAO. 1995. A highly conserved nuclear gene for low-level phylogenetics: elongation factor-

1∝ recovers morphology-based tree for heliothine moths. Mol. Biol. Evol. **12**:650–656.

Collins, T. M., P. H. Wimberger, and G. P. Naylor. 1994. Compositional bias, character-state bias, and character-state reconstruction using parsimony. Syst. Biol. **43**:482–496.

Crozier, R. H., and Y. C. Crozier. 1993. The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. Genetics **133**:97–117.

Cummings, M., S. Otto, and J. Wakeley. 1995. Sampling properties of DNA sequence data in phylogenetic analysis. Mol. Biol. Evol. **12**:814–822.

Cunningham, C. W. 1997. Can three incongruence tests predict when data should be combined? Mol. Biol. Evol. **14**: 733–740.

Derr, J. N., S. K. Davis, J. B. Woolley, and R. A. Wharton. 1992*a.* Variation and phylogenetic utility of the large ribosomal subunit of mitochondrial DNA from the insect order Hymenoptera. Mol. Phylogenet. Evol. **1**:136–147.

———. 1992*b.* Reassessment of the 16S rRNA nucleotide sequence from members of the parasitic Hymenoptera. Mol. Phylogenet. Evol. **1**:338–341.

DeSalle, R. 1992*a.* The phylogenetic relationships of flies in the family Drosophilidae deduced from mtDNA sequences. Mol. Phylogenet. Evol. **1**:31–40.

———. 1992*b.* The origin and possible time of divergence of the Hawaiian Drosophilidae: evidence from DNA sequences. Mol. Biol. Evol. **9**:905–916.

DeSalle, R., T. Freedman, E. M. Prager, and A. C. Wilson. 1987. Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila.* J. Mol. Evol. **26**:157–164.

DeSalle, R., and A. R. Templeton. 1988. Founder effects accelerate the rate of mitochondrial DNA evolution in Hawaiian Drosophila. Evolution **42**:1076–1084.

Dowton, M., and A. D. Austin. 1994. Molecular phylogeny of the insect order Hymenoptera: apocritan relationships. Proc. Natl. Acad. Sci. USA **91**:9911–9915.

———. 1997*a.* Evidence for AT-transversion bias in wasp (Hymenoptera: Symphyta) mitochondrial genes and its implications for the origin of parasitism. J. Mol. Evol. **44**:398–405.

———. 1997*b.* The evolution of strand-specific compositional bias. A case study in the hymenopteran mitochondrial 16S rRNA gene. Mol. Biol. Evol. **14**:109–112.

Dowton, M., A. D. Austin, and M. F. Antolin. 1998. Evolutionary relationships among the Braconidae (Hymenoptera: Ichneumonoidea) inferred from partial 16S rDNA gene sequences. Insect Mol. Biol. **7**:129–150.

Dowton, M., A. D. Austin, N. Dillon, and E. Bartowsky. 1997. Molecular phylogeny of the apocritan wasps with particular reference to the Proctotrupomorpha and Evaniomorpha. Syst. Entomol. **22**:245–255.

Engel, M. S., and T. R. Schultz. 1997. Phylogeny and behavior in honey bees (Hymenoptera: Apidae). Ann. Entomol. Soc. Am. **90**:43–53.

Fang, Q., W. C. Black Jr., H. D. Blocker, and R. F. Whitcomb. 1993. A phylogeny of New World *Deltocephalus*-like leafhopper (Homoptera: Cicadellidae) genera based on mitochondrial 16S ribosomal DNA sequences. Mol. Phylogenet. Evol. **2**:119–131.

Flook, P. K., and C. H. F. Rowell. 1997*a.* The phylogeny of the Caelifera (Insecta, Orthoptera) as deduced from mtrRNA gene sequences. Mol. Phylogenet. Evol. **8**:89–103.

———. 1997*b.* The effectiveness of mitochondrial RNA gene sequences for the reconstruction of the phylogeny of an insect order (Orthoptera). Mol. Phylogenet. Evol. **8**:177–192.

Friedlander, T. P., J. C. Regier, and C. Mitter. 1994. Phylogenetic information content of five nuclear gene sequences in animals: intitial assessment of character sets from concordance and divergence studies. Syst. Biol. **43**:511–525.

Friedlander, T. P., J. C. Regier, C. Mitter, and D. L. Wagner. 1996. A nuclear gene for higher level phylogenetics: phosphoenolpyruvate carboxykinase tracks Mesozoic-age divergences within Lepidoptera (Insecta). Mol. Biol. Evol. **13**:594-604.

Funk, D. J., D. J. Futuyma, G. Orti, and A. Meyer. 1995. Mitochondrial DNA sequences and multiple data sets: a phylogenetic study of phytophagous beetles (Chrysomelidae: *Ophraella*). Mol. Biol. Evol. **12**:627–640.

Gilbert, D. 1993. SeqApp. Version 1.9a. Distributed by the author, Department of Biology, Indiana University, Bloomington.

Golding, G. B. 1983. Estimates of DNA and protein sequence divergence: an examination of some assumptions. Mol. Biol. Evol. **1**:125–142.

Graybeal, A. 1994. Evaluating the phylogenetic utility of genes: a search for genes informative about deep divergences among vertebrates. Syst. Biol. **43**:174–193.

Gutell, R. R. 1993. Comparative studies of RNA: inferring higher-order structure from patterns of sequence variation. Curr. Opin. Struct. Biol. **3**:313–322.

Hasegawa, M., and H. Kishino. 1989. Heterogeneity of tempo and mode of mitochondrial DNA evolution among mammalian orders. Jpn. J. Genet. **64**:243-258.

Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22**:160–174.

Hillis, D. M., and M. T. Dixon. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. Q. Rev. Biol. **66**:411–453.

Hillis, D. M., J. P. Huelsenbeck, and C. W. Cunningham. 1994. Application and accuracy of molecular phylogenies. Science **264**:671–676.

Huelsenbeck, J. P., and B. Rannala. 1997. Phylogenetic methods come of age: testing hypotheses in an evolutionary context. Science **276**:227–232.

Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules. Pp. 21–132 *in* H. N. Munro, ed. Mammalian protein metabolism. Academic Press, New York.

Kambhampati, S. 1995. A phylogeny of cockroaches and related insects based on DNA sequence of mitochondrial ribosomal RNA genes. Proc. Natl. Acad. Sci. USA **92**:2017–2020.

Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.

Knight, A., and D. P. Mindell. 1993. Substitution bias, weighting of DNA sequence evolution, and the phylogenetic position of Fea's viper. Syst. Biol. **42**:18-31.

Koulianos, S., R. Schmid-Hempel, D. W. Roubik, and P. Schmid-Hempel. 1998. Phylogenetic relationships within the Apinae (Hymenoptera) and the evolution of eusociality. J. Evol. Biol. **11** (in press).

Kraus, F., L. Jarecki, M. M. Miyamoto, S. M. Tanhauser, and P. J. Laipis. 1992. Mispairing and compensational changes during the evolution of mitochondrial ribosomal RNA. Mol. Biol. Evol. **9**:770–774.

Kumar, S., K. Tamura, and M. Nei. 1993. MEGA: molecular evolutionary genetics analysis. Version 1.01. Pennsylvania State University, University Park.

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.

MADDISON, W. P., and D. R. MADDISON. 1992. MacClade: analysis of phylogeny and character evolution. Version 3. Sinauer, Sunderland, Mass.

MARDULYN, P., and S. A. CAMERON. 1998. The major opsin in bees (Insecta: Hymenoptera): a promising nuclear gene for higher level phylogenetics. Mol. Phylogenet. Evol. (in press).

MARDULYN, P., M. C. MILINKOVITCH, and J. M. PASTEELS. 1997. Phylogenetic analyses of DNA and allozyme data suggest that *Gonioctena* leaf beetles (Coleoptera: Chrysomelidae) experienced convergent evolution in their history of host-plant family shifts. Syst. Biol. **46**:699–725.

MARDULYN, P., and J. B. WHITFIELD. 1998. Phylogenetic signal in the COI, 16S, and 28S genes for inferring relationships among genera of Microgastrinae (Hymenoptera: Braconidae): evidence of a high rate of diversification in this group of parasitoids. Mol. Phylogenet. Evol. (in press).

MINDELL, D. P., and R. L. HONEYCUTT. 1990. Ribosomal RNA in vertebrates: evolution and phylogenetic applications. Ann. Rev. Ecol. Syst. **21**:541–566.

NIELSEN, R. 1997. Site-by-site estimation of the rate of substitution and the correlation of rates in mitochondrial DNA. Syst. Biol. **46**:346–353.

PALUMBI, S. R. 1989. Rates of molecular evolution and the fraction of nucleotide positions free to vary. J. Mol. Evol. **29**:180–187.

PASHLEY, D. P., and L. D. KE. 1992. Sequence evolution in mitochondrial ribosomal and ND-1 genes in Lepidoptera: implications for phylogenetic analyses. Mol. Biol. Evol. **9**: 1061–1075.

ROGERS, J. S., and D. L. SWOFFORD. 1998. A fast method for approximating maximum likelihoods of phylogenetic trees from nucleotide sequences. Syst. Biol. **47**:7–89.

ROIG-ALSINA, A., and C. D. MICHENER. 1993. Studies of the phylogeny and classification of long-tongued bees (Hymenoptera: Apoidea). Univ. Kans. Sci. Bull. **55**:124–162.

SHEPPARD, W. S., and B. A. MCPHERON. 1991. Ribosomal DNA diversity in Apidae. Pp. 89–102 *in* D. R. SMITH, ed. Diversity in the genus *Apis.* Westview Press, Boulder, Colo.

SIMON, C., F. FRATI, A. BECKENBACH, B. CRESPI, H. LIU, and P. FLOOK. 1994. Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. **87**:651–701.

SMITH, A. B. 1989. RNA sequence data in phylogenetic reconstruction: testing the limits of its resolution. Cladistics **5**: 321–344.

SWOFFORD, D. L. 1997/1998. PAUP*—phylogenetic analysis using parsimony (* and other methods). Test versions 4.0d54–d64 used with permission of author; beta version released officially 1998 by Sinauer, Sunderland, Mass.

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.

TAJIMA, F., and M. NEI. 1984. Estimation of evolutionary distance between nucleotide sequences. Mol. Biol. Evol. **1**: 269–285.

TAMURA, K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. Mol. Biol. Evol. **9**:678-687.

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.

THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. CLUSTAL W: improving the sensititvity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.

VOGLER, A. P., and R. DESALLE. 1993. Phylogeographic patterns in coastal North American tiger beetles, *Cicindela dorsalis,* inferred from mitochondrial DNA sequences. Evolution **47**:1192–1202.

VOGLER, A. P., R. DESALLE, T. ASSMANN, C. B. KNISLEY, and T. D. SCHULTZ. 1993*a.* Molecular population genetics of the endangered tiger beetle *Cicindela dorsalis* (Coleoptera: Cicindelidae). Ann. Entomol. Soc. Am. **86**:142–152.

VOGLER, A. P., C. B. KNISLEY, S. B. GLUECK, J. M. HILL, and R. DESALLE. 1993*b.* Using molecular and ecological data to diagnose endagered populations of the puritan tiger beetle *Cincindela puritana.* Mol. Ecol. **2**:375–383.

WAKELEY, J. 1996. The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. TREE **11**:158–163.

WHEELER, W. C., and R. L. HONEYCUTT. 1988. Paired sequence difference in ribosomal RNAs: evolutionary and phylogenetic implications. Mol. Biol. Evol. **5**:90–96.

WHITFIELD, J. B. 1997. Molecular and morphological data suggest a single origin of the polydnaviruses among braconid wasps. Naturwissensshaften **84**:502–507.

———. 1998. Phylogeny and evolution of host-parasitoid interactions in Hymenoptera. Ann. Rev. Entomol. **43**:129–151.

XIONG, B., and T. D. KOCHER. 1993*a.* Intraspecific variation in sibling species of *Simulium venustum* and *Simulium verecundum* complexes (Diptera: Simuliidae) revealed by the sequence of the mitochondrial 16S rRNA gene. Can. J. Zool. **71**:1202–1206.

———. 1993*b.* Phylogeny of sibling species of *Simulium venustum* and *S. verecundum* (Diptera: Simuliidae) based on sequences of the mitochodrial large subunit rRNA gene. Mol. Phylogenet. Evol. **4**:293–303.

YANG, Z. 1994. Estimating the pattern of nucleotide substitution. J. Mol. Evol. **39**:105-111.

———. 1995. Evaluation of several methods for estimating phylogenetic trees when substitution rates differ over nucleotide sites. J. Mol. Evol. **40**:689–697.

———. 1996. Among-site rate variation and its impact on phylogenetic analysis. TREE **11**:367–372.

YANG, Z., and S. KUMAR. 1996. Approximate methods for estimating the pattern of nucleotide subtitution and the variation of substitution rate among sites. Mol. Biol. Evol. **13**: 650–659.

DAVID M. RAND, reviewing editor

Accepted September 4, 1998