ACADEMIC
PRESS

Simultaneous analysis of the basal lineages of Hymenoptera (Insecta) using sensitivity analysis

Susanne Schulmeister,[a,*] Ward C. Wheeler,[b] and James M. Carpenter[b]

[a] *Institute of Zoology and Anthropology, University of Göttingen, Berliner Strasse 28, Göttingen 37073, Germany*
[b] *Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024-5192, USA*

Abstract

The first simultaneous analysis of molecular and morphological data of basal hymenopterans that includes exemplars from all families is presented. DNA sequences (of approximately 2000–2700 bp for each taxon) from the nuclear genes 18S and 28S and the mitochondrial genes 16S and CO1 have been sequenced for 39 taxa (four outgroup taxa, 29 symphytans, and six apocritans). These DNA sequences and 236 morphological characters from Vihelmsen [Zool. J. Linnean Soc. 131 (2001) 393] were analyzed separately as well as simultaneously. All analyses were performed on unaligned sequences, using the optimization alignment (= direct optimization) method. Sensitivity analysis sensu Wheeler [Syst. Biol. 44 (1995) 321] was applied by analyzing the data under nine different combinations of analysis parameter values. The superfamily level relationships of basal hymenopterans as proposed by Vilhelmsen [Zool. J. Linnean Soc. 131 (2001) 393] and Ronquist et al. [Zool. Scr. 28 (1999) 13] are mostly confirmed, except that Pamphilioidea is the sister group to Tenthredinoidea s.l. and that Anaxyelidae (i.e., *Syntexis libocedrii*) and Siricidae are supported as a monophyletic group, partly reestablishing the traditional concept of Siricoidea. The resulting hypothesis that best represents the combined evidence from morphology and DNA sequences is (Xyeloidea (Tenthredinoidea s.l. Pamphilioidea) (Cephoidea (Siricoidea (Xiphydrioidea (Orussidae Apocrita))))), with Siricoidea = Anaxyelidae + Siricidae. The phylogenetic system within Tenthredinoidea s.l., derived from the combined evidence, is (Blasticotomidae (Tenthredinidae including Diprionidae (Cimbicidae (Argidae Pergidae)))).

Introduction

"Symphyta" or sawflies have long been assumed to be paraphyletic, even before the advent of phylogenetic systematics: Börner (1919) depicted a tree of the presumed evolution of Hymenoptera in which the sawflies (without Orussidae) are shown as paraphyletic. Orussidae was even placed within Apocrita. Two years earlier, Rohwer and Cushman (1917) had already claimed that Orussidae "stands intermediate between" the other sawflies and the apocritans. Sawflies constitute the basalmost branches of the tree of Hymenoptera, hence the descriptive term "basal hymenopterans," which is preferred by those who want to avoid attaching a name to a presumably paraphyletic group. Members of this paraphyletic group can easily be distinguished from higher

hymenopterans (Apocrita) by a prominent symplesiomorphy, viz., the lack of the wasp–waist (a constriction between the first and second abdominal segments, which is a synapomorphy of Apocrita). A summary of previous hypotheses on the phylogenetic relationships of basal hymenopterans is presented under Background below.

The first extensive morphological character matrices for exemplars from all 14 sawfly and some apocritan families have been published recently (Vilhelmsen, 1997a, 2001), which presented the opportunity for the first simultaneous analysis of DNA sequences and morphology covering all families of sawflies. We sequenced approximately 2000–2700 bp of four different genes (16S, 18S, 28S, and CO1) for 39 of the 44 taxonomic units that were used in Vilhelmsen's analysis (2001). The species used for sequencing were not necessarily the same species as those used by Vilhelmsen (2001), but the sawflies were always from the same genera, except that *Onycholyda* was sequenced instead

*Corresponding author.

E-mail address: sschulm@gwdg.de (S. Schulmeister).

of *Pamphilius*. The apocritan exemplars were from the same genera or at least families as those used to study the morphology, except that sequences from Gasteruptionidae and Evaniidae were combined with the morphology of *Aulacus*. Details on the taxon sampling are given in Appendix A.

The molecular data were analyzed simultaneously with Vilhelmsen's (2001) morphological data, as well as separately. To examine the stability of our phylogenetic results, we performed a sensitivity analysis sensu Wheeler (1995) by repeating the analyses with nine different sets of values for the analysis parameters.

It should be mentioned that we do not take groups referred to as families (or superfamilies) as equivalent in any sense; they are only arbitrary taxonomic levels. For convenience, we refrain from indicating this further in the text, e.g., by the use of quotation marks. Also, if we speak of, for example, "all families of basal hymenopterans," this is shorthand for "all groups of basal hymenopterans arbitrarily classified as families" or even listing the names of all 14 taxa. Note that the discussion in the present paper deals exclusively with extant taxa; for example, "monophyly of Pamphiliidae" means "monophyly of the crown group of Pamphiliidae."

Background

Systematics of basal hymenopterans

Ross (1937, p. 37) presented a tree in which there are four groups of sawflies depicted as branching off from the stem leading to Apocrita: first Tenthredinoidea, second Megalodontoidea, then Siricoidea, and finally Cephoidea—as sister taxon to Apocrita. His Tenthredinoidea is divided into two groups: Tenthredinidae, Diprionidae, and Cimbicidae on one side, and Argidae and Pergidae on the other. (Blasticotomidae is missing in his tree.) His Megalodontoidea comprises Pamphiliidae, Megalodontesidae, and Xyelidae, and his Siricoidea contains Siricidae, Anaxyelidae, Xiphydriidae, and Orussidae.

Rasnitsyn (1969, 1988) hypothesized a paraphyletic Xyelidae, with Macroxyelinae as sister taxon to Tenthredinoidea s.l. and Xyelinae as sister group to all other Hymenoptera, the extant taxa of which would be grouped as (Pamphilioidea (Cephidae (Anaxyelidae (Siricidae (Xiphydriidae (Orussidae Apocrita)))))). However, when the characters from Rasnitsyn's (1988) paper were later coded into a cladistic data matrix (Ronquist et al., 1999), the analysis resulted in a monophyletic "Hymenoptera excluding Xyelidae;" the monophyly of Xyelidae remained uncertain. In this reanalysis, Tenthredinoidea is the sister taxon to all other nonxyelid hymenopterans, the relationships of which are congruent with the hypothesis of Rasnitsyn (1988)

mentioned above. Rasnitsyn (1988) introduced the name **Vespina** for the Orussidae + Apocrita clade; this name will be used in the present paper.

Within Tenthredinoidea s.str., Rasnitsyn (1988) proposed the relationships (Blasticotomidae ((Argidae Pergidae) (Cimbicidae Tenthredinidae))). This was confirmed by the cladistic reanalysis (Ronquist et al., 1999). Rasnitsyn's (1988) Tenthredinidae includes Diprionidae. This concept of the name Tenthredinidae will be termed **Tenthredinidae sensu Rasnitsyn** in the present paper, to distinguish it from the usual concept of Tenthredinidae.

Oeser (1961) was the first to apply phylogenetic systematic argumentation schemes (Hennig, 1950, 1966) to Hymenoptera, but his conclusions were limited because they were derived only from the morphology of the female ovipositor. In Königsmann's (1976, 1977, 1978a,b) cladogram, also based on phylogenetic systematics, Hymenoptera is divided into two monophyletic groups: all sawflies with the exception of Cephidae ("Symphyta s.str.") and Cephidae + Apocrita. The cenchri (metanotal patches for the attachment of the front wings) are a potential synapomorphy for Symphyta s.str.; potential synapomorphies for Cephidae and Apocrita are a slight wasp-waist, fusion of gonostipes and harpe in the male genitalia (also occurring in Orussidae), and the phallotrema opening apically.

Gibson (1985, 1993) also hypothesized Xyelidae as the sister group to all other hymenopterans, as well as proposing the relationships (Siricidae (Xiphydriidae (Orussidae Apocrita))); he elevated Xiphydriidae and Orussidae to superfamilies. Synapomorphies for Orussidae and Apocrita were also found by Johnson (1988), Whitfield et al. (1989), and Basibuyuk and Quicke (1995). Contrary to Rasnitsyn (1969, 1988), but in agreement with the later reanalysis (Ronquist et al., 1999), Gibson (1993) found evidence for the monophyly of nonxyelid hymenopterans. Characters analyzed by Heraty et al. (1994) support the relationships (Tenthredinoidea s.l. (Cephidae (Anaxyelidae Siricidae (Xiphydriidae Orussidae Apocrita)))).

In a cladistic analysis of basal hymenopterans, Wei and Nie (1997) coded 56 morphological characters for the two subfamilies of Xyelidae, for the other 13 sawfly families, and for Apocrita. The data were analyzed with different treatments, resulting in a number of trees with little similarity to one another. Each of the two unweighted analyses gave three most-parsimonious cladograms, only one of which is shown in their paper. A summary of this tree is (Xyelidae (Tenthredinoidea (Pamphilioidea (Cephidae ((Anaxyelidae Siricidae) (Xiphydriidae (Orussidae Apocrita))))))). However, it can be seen from their English abstract that Wei and Nie (1997) prefer a different hypothesis. That hypothesis (Wei and Nie, 1997 see their classification on p. 41, top) has Cephidae as a sister group to Apocrita, and

Pamphilioidea, Siricidae, Anaxyelidae, Xiphydriidae, and Orussidae united in a monophyletic group, which corresponds to their trees 3, 5, and 6 (Wei and Nie, 1997, p. 40) obtained from weighted analyses.

The most thorough and extensive work on the morphology and systematics of the extant basal lineages of Hymenoptera was provided by Vilhelmsen (1996, 1997a,b, 1999a,b, 2000a,b,c, 2001). For his first cladistic analysis (Vilhelmsen, 1997a), he used 98 morphological characters compiled from the literature as well as those from his studies of the preoral cavity and antennal bases (Vilhelmsen, 1996, 1997b). It resulted in the hypothesis (Macroxyelinae Xyelinae (Tenthredinoidea s.l. (Pamphilioidea (Cephidae (Anaxyelidae (Siricidae (Xiphydriidae (Orussidae Apocrita)))))))). Later (Vilhelmsen, 1999a, 2001), he added characters from his studies on the head and prothorax (Vilhelmsen, 1999b, 2000b), the metathoracic–abdominal region (Vilhelmsen, 2000a), and the ovipositor (Vilhelmsen, 2000c). The most recent analysis (Vilhelmsen, 2001) covers 236 morphological characters: 40 from Rasnitsyn (1988) as coded by Ronquist et al. (1999), 14 from Heraty et al. (1994), 12 from Basibuyuk and Quicke (1995, 1997, 1999), nine from Königsmann (1976, 1977), nine from Gibson (1985, 1986, 1993, 1999), eight from Yuasa (1922), five from Johnson (1988), four from Quicke et al. (1994), three from Whitfield et al. (1989), three from Kristensen (1991), two from Goulet (1993), one from Oeser (1961), one from Paulus (1979), one from Mickoleit (1966, 1969), and 124 from his own studies (see above). They were coded for six outgroup taxa, 32 sawflies from all 14 families, and six apocritans. The parsimony analysis, treating some characters as additive (and without any weighting), produced two most-parsimonious trees of length 880, which supported Xyelidae as monophyletic. Other than that, the consensus tree supports the major relationships from his first analysis as described above. The relationships within Tenthredinoidea s.l. are (Blasticotomidae (*Athalia Tenthredo Dolerus* (*Nematus* (Diprionidae (Cimbicidae (Pergidae Argidae)))))).

A parsimony analysis that treats all characters from Vilhelmsen (2001) as nonadditive (NONA, amb-) produces eight most-parsimonious trees of 853 steps, which disagree mainly on the most basal sawfly group (either Xyelidae or Tenthredinoidea s.l.) and on whether the Anaxyelidae or Siricidae is the sister group to the Xiphydriidae + Vespina clade. This indicates that the morphological evidence is somewhat weak in these areas of the tree.

Derr et al. (1992a,b) sequenced the DNA of representatives from across the Hymenoptera, but their sample of 10 taxa and one gene was too small to permit any reasonable phylogenetic conclusions. Considering that dipterans were used as outgroup taxa and only two symphytans were sequenced, the "result of a monophyletic Symphyta relative to Apocrita" is not really "sur-

prising." Dowton and Austin (1994) analyzed 16S rDNA sequences of two dipterans, seven symphytans, and 24 apocritans, but found that this gene by itself was unable to resolve the phylogeny of basal hymenopterans. In a later study, Dowton and Austin (1997) analyzed 16S and CO1 sequences of three dipterans, one lepidopteran, seven sawfly species, and three apocritans. Using the dipterans as outgroups, the analysis resulted in the cladogram ((Lepidoptera *Macroxyela*) Pergidae (Cephidae (Xiphydriidae (Orussidae (Siricidae (Apocrita)))))).

The first simultaneous analysis of the entire group Hymenoptera (Carpenter and Wheeler, 1999) included 36 species from 19 families. The morphology (from Ronquist et al., 1999) was coded as the presumed ground-plan for the families. A fragment of the 28S rDNA gene was sequenced for all 36 species, and fragments of 18S and CO1 were added for different subsets of species. The analysis suffers from a bias toward Apocrita: only three sawflies were included, one of which was used as the root.

Simultaneous analysis

Simultaneous analysis (= combined analysis) is the analysis of multiple data sets combined in one data matrix. The earliest simultaneous analysis of molecular and nonmolecular data was published (as far as we know) by Mickevich and Johnson (1976). Kluge (1989) introduced the terms "character congruence" and "total evidence" for combined analysis. The term "simultaneous analysis" was first used by Nixon and Carpenter (1996), who presented a review of the early literature on the topic. Recent reviews and discussions of simultaneous analysis and related topics have been presented, among others, by Brower et al. (1996), Bull et al. (1993), Chippindale and Wiens (1994), de Queiroz et al. (1995), Larson (1998), Miyamoto and Fitch (1995), and Nixon and Carpenter (1996).

Simultaneous analysis, or total evidence, is the method of choice for analyzing different data sets for a given sample of taxa: it maximizes explanatory power for all the data (Kluge and Wolf, 1993) and, as reviewed by Nixon and Carpenter (1996), it maximizes parsimony by seeking the hypothesis which is best supported by all the data. The justification for the application of parsimony is that it minimizes ad hoc hypotheses of homoplasy (Farris, 1983), and ad hoc assumptions of homoplasy can be minimized for all data only within the framework of simultaneous analysis. Application of consensus techniques to the results of independent analyses of multiple data sets (partitioned analysis), as well as the so-called "conditional combinability" (not combining if data sets are "heterogeneous"), does not measure the **relative support** (in terms of synapomorphies) from the different data sources. For example, even if one data set shows 10 synapomorphies for group A + B and the other data set contains three synapo-

morphies for group B + C, the consensus approach will be unresolved, even though the support for the first alternative is much stronger. Simultaneous analysis takes not only this relative support into account, but also potential **hidden support** ("secondary signals," see Gatesy et al., 1999). For example, one data set might contain five synapomorphies for group D + E and the second data set five synapomorphies for E + F. Looking at the separate results, it would seem that the strength of support is equal for the two alternatives. However, in a simultaneous analysis it could be discovered that the second data set actually contains four synapomorphies for D + E in addition to the five for E + F and that, hence, the total support for D + E from all the data is much larger than that for E + F.

Advocates of partitioned analysis clearly underestimate the effect of sampling error on small data sets. If the results from two (or more) data sets are incongruent, this is often not the result of different "histories," but an artifact of the small amount of data used to address the problem (e.g., Cummings et al., 1995). (In other words, if the trees from two different types of data are incongruent, they might have been congruent if more data of each type had been sampled.) It can, however, also be the result of systematic bias, but this cannot be objectively determined.

Direct optimization

The direct optimization (= optimization alignment) method (Wheeler, 1996) is a maximum-parsimony algorithm that can process unaligned molecular sequences in addition to morphological and aligned molecular data. Like other maximum-parsimony algorithms, direct optimization strives to find the shortest cladogram by determining the lengths of many different topologies, but unlike other optimization algorithms, it works with unaligned sequences (which may be of unequal length). A multiple alignment is created neither before nor during the analysis. However, pairwise comparisons are performed during the so-called down-pass of the optimization process: at each internal node of a given topology, direct optimization reconstructs the single (preliminary) hypothetical sequence of the node in question by optimizing the state and number of the bases, deleting or adding them where necessary. In other words, direct optimization is an extension of "generalized optimization" (Sankoff and Rousseau, 1975) to include insertion and deletion events in addition to base substitutions. In this way, indels are treated as processes rather than observations (see Wheeler, 1996). The cladogram length is the sum of the costs for all hypothesized nucleotide substitution and insertion/deletion (indel) events. Extensive explanations of the procedure are given by Wheeler (1996). During the up-pass of the optimization process, the preliminary (and often ambiguous) hypo-

thetical ancestral sequences are transformed into final hypothetical ancestral sequences. This way, a potentially unique scheme of positional homologies is hypothesized for each examined topology. (The positional homologies can be output after the analysis in what is called an implied alignment.) Herein lies the advantage of direct optimization over conventional methods. In conventional analysis, where the alignment procedure is performed prior to the analysis, the alignment topology can be quite different from the cladogram that is the final result of the analysis, which can lead to the possibility that the parsimony analysis is based on a suboptimal alignment. This is one reason why the result of a direct optimization analysis is usually more parsimonious than a conventional analysis of the same data (Wheeler, 2000).

Direct optimization and simultaneous analysis

True multiple alignment in which all sequences are aligned simultaneously (Sankoff and Cedergren, 1983) and by which the globally optimal solution can be found is too costly in terms of computational time to be used for more than a few short sequences (at least nowadays). Therefore, a multiple alignment is normally generated by aligning sequences pairwise, in an order that is outlined in a so-called guide-tree or alignment topology that was determined by the sequence data themselves. But even true multiple alignment requires a topology to determine (and minimize) the alignment cost. The order in which the pairwise alignments are performed, i.e., the alignment topology, has an influence on the positional homology statements (i.e., the multiple alignment), which in turn have an influence on the outcome of the phylogenetic analysis.

In conventional analysis (in which the alignments are separate from the analysis), a multiple sequence alignment is generated for each gene separately. These individually generated alignments are used not only for partitioned analysis, but also for simultaneous analysis of all the genes together as well as simultaneous analysis of molecular and morphological data. This means that the alignment topologies used for the alignments of the molecular data in a simultaneous analysis are based only on the information provided by the individual genes. There is a logical inconsistency in using multiple alignments derived from individual data partitions in a simultaneous analysis. To perform a logically consistent simultaneous analysis, it is necessary that not only the parsimony analysis be based on all the data, but the generation of the alignments as well.

Direct optimization is the only currently available solution to the problem of including the alignment procedure into the simultaneous analysis framework. Here, a unique scheme of positional homologies (a kind of multiple alignment) is created for each examined topology during the tree search (see above). Hence, the

length of the shortest simultaneous analysis cladogram is based on positional homology schemes that were generated specifically for this particular topology.

The conventional procedure of treating the alignment procedure and the parsimony analysis as two consecutive steps is not just logically inconsistent—it can also lead to a suboptimal solution for the simultaneous analysis. If the alignment topologies of the separate genes are different from the simultaneous analysis cladogram—which is often the case—it can lead to alignments that can, in turn, produce a simultaneous cladogram that is longer than if the simultaneous analysis uses homology schemes specifically tailored for the cladogram that is the result of the simultaneous analysis.

Wheeler (2000) showed that simultaneous analysis cladograms generated with direct optimization are much shorter than simultaneous analysis cladograms generated from the same data using conventional analysis. For three simultaneous analyses of molecular data, the lengths of the cladograms were reduced on average by 13.8%. Even if it is taken into account that direct optimization analyses of the individual genes were already shorter than conventional analyses of the individual genes (see above)—on average by 12.2%—there is still an enhancement of the relative performance in simultaneous analysis (in this case 1.6%). For a simultaneous analysis of molecular and morphological data, the length of the cladograms of the individual genes was reduced by 13.1% on average by the use of direct optimization, while the length of the simultaneous analysis cladogram was reduced by 17.5%, i.e., an additional 4.4%!

Sensitivity analysis

To derive a cladogram from DNA data, whether via multiple alignment or direct optimization, specific numerical values must be assigned for the alignment and analysis parameters commonly called gap cost or indel (= insertion and deletion) cost, transversion cost, and transition cost. They represent character state transformation weights; the indel cost specifies the weight that is given to an insertion or deletion event and the transversion and transition costs are the weights given to transformations of one base into another base. These parameters are often expressed as ratios: the gap:change cost ratio, which specifies the cost of an indel event relative to that of a transversion, and the transversion:transition cost ratio. The two ratios can also be expressed as a single gap:transversion:transition cost ratio; a parameter set of a gap:transversion ratio of 2:1 in combination with a transversion:transition ratio of 2:1 could, for instance, be written as a gap:transversion:transition ratio of 4:2:1 (2:2:1 would be incorrect, as done for example in Giribet and Ribera (2000); 2:2:1 would imply a gap:transversion ratio of 1:1 in combination with a transversion:transition ratio of 2:1).

These parameter values are needed for the alignment process as well as the phylogenetic analysis. Usually, they are chosen rather arbitrarily and only one set of parameter values is used for the analysis. However, the parameter values can have a strong influence on the outcome of a cladistic analysis. Whether a group appears as monophyletic or paraphyletic depends in many cases on the parameter values. If only a single phylogenetic analysis is performed with one set of parameter values, it remains completely unknown how much the result is dependent on these arbitrarily chosen values.

To examine the sensitivity of a cladogram to the alignment and analysis parameters and to remove the arbitrariness of the choice of these values, Wheeler (1995) introduced the concept of a sensitivity analysis. It starts by selecting a number of parameter sets, i.e., the combinations of values for indel-to-transversion ratio and transversion-to-transition ratio, that are to be examined in the sensitivity analysis. The molecular data are analyzed separately and simultaneously with the morphological data for the selected parameter sets. To choose among the parameter sets, Wheeler (1995) suggested the use of congruence as an external criterion. From the lengths of the single and combined trees, the character incongruence between the individual data partitions is calculated for each parameter set (see below). The combination of parameter values that minimizes the incongruence (hence, maximizes the congruence) is chosen as the best. The simultaneous analysis cladogram that has been calculated with the optimal parameter set is presented as the best phylogenetic hypothesis, because it is this hypothesis that minimizes the extra homoplasy forced upon the data through combination in a simultaneous analysis.

If a data set is analyzed with only one parameter set, it is overlooked that the resulting clades might strongly depend on the employed indel cost and transversion:transition ratio. One of the merits of a sensitivity analysis sensu Wheeler lies in the determination of alternative phylogenetic hypotheses inherent in the data and the assessment of their stability with respect to the analysis parameters. It can be examined under which of the parameter sets a given group of taxa comes out as monophyletic, which can be graphically represented in a so-called sensitivity plot. In this way, a sensitivity analysis can discern between robust clades (those that appear under most or all of the parameter sets) and unstable clades (those that appear only under one or few parameter regimes). Or in other words, the robustness of a clade in sensitivity analysis shows how stable a clade is to the relaxation of the optimality criterion, which is thus analogous to decay indices in parsimony analysis. Stability is here a kind of support measure.

Since morphological characters are not influenced by transition, transversion, and indel costs, clade sensitivity to the analysis parameters can be observed directly only

for molecular data. But in simultaneous analysis, the clade sensitivity caused by the molecular data can be compensated for by the morphological data: if the morphological support for a clade is very strong, it can appear in all simultaneous trees even if it is present in only a few or none of the molecular trees. Comparing the clade sensitivity of the molecular trees to that of the simultaneous trees can therefore provide information about the strength of the morphological support in some cases.

Congruence

There are basically two alternatives to determine congruence between several data sets for the same taxon sample. The congruence among the data themselves can be measured, which is character congruence, or the congruence between the most-parsimonious topologies resulting from the separate data sets can be assessed, which is termed topological or taxonomic congruence. (The expressions character congruence and taxonomic congruence are also used to mean simultaneous analysis and the consensus approach, respectively.)

If several data sets are analyzed simultaneously, the amount of character incongruence or homoplasy present among all the data can be distinguished into that occurring within the separate data sets (within-data-set incongruence) and the amount of additional homoplasy that results solely from combining the different data sets (between-data-sets incongruence). The latter can be calculated with the incongruence length difference (ILD; Mickevich and Farris, 1981, pp. 366–367, Farris et al., 1995). It is defined as the length of the simultaneous tree minus the sum of the lengths of the individual trees.

To make the ILD values derived from analyses with different parameter values comparable, the ILD needs to be standardized in some way. This can be done by dividing it by the total length of the simultaneous tree (Mickevich and Farris, 1981). The problem with this measure for the use in sensitivity analysis is that if the separate trees already contained much of the possible homoplasy, there was little homoplasy that could be added by combining the data, so that the ILD would be low (Goloboff, personal communication). To avoid this problem, Wheeler and Hayashi (1998) proposed to standardize the ILD by dividing it by the maximum possible between-data-set incongruence:

$$RILD = (\text{Length of simultaneous tree}$$
$$- \text{Sum of lengths of separate trees})/$$
$$(\text{Maximum length of simultaneous tree}$$
$$- \text{Sum of lengths of separate trees})$$
$$= \left(Lst - \sum Li\right)/\left(maxLst - \sum Li\right).$$

Materials and methods

Taxa

The taxon sample of Vilhelmsen (1999a, 2001) was matched as closely as possible to facilitate combination of data matrices and to avoid unnecessary inflation of the operational taxonomic units (OTUs). Appendix A lists the species used to generate the morphological and the molecular data matrices and the names of the OTUs. Most of the sawfly genera in Vilhelmsen's (1999a, 2001) analyses were sequenced here, with the following exceptions. The blasticotomids *Blasticotoma* and *Paremphytus* and the diprionid *Monoctenus* are missing from the present study and *Onycholyda* was substituted for *Pamphilius*, which belongs to the same tribe (Pamphiliini). Two outgroups, Psocodea and Neuroptera, were also not included.

Morphology

The morphological data matrix used here is the one from Vilhelmsen (2001) except for the omission of the five taxa mentioned above.

DNA

Most specimens were preserved in 95% ethanol, some were pinned and dried (but had been killed in ethanol). Total genomic DNA was extracted by overnight incubation of a tissue sample (thorax or leg muscles or ovaries) in a solution of guanadium isothyocyanate and 0.14 M β-mercaptoethanol followed by a standard phenol/chloroform extraction and ethanol precipitation. After the DNA was dried, it was resuspended in water.

Target genes were amplified by polymerase chain reaction (PCR), if necessary in several overlapping pieces. Sequences of primers are given in Appendix B. A typical PCR procedure for 16S, 18S, and 28S was an initial denaturation at 96 °C for 1 min and then 30–40 cycles of denaturation (15 s at 96 °C), annealing (15 s at 49 °C), and extension (15 s at 72 °C). For the CO1 gene, the denaturing and annealing temperatures were lowered to 94 and 46 °C, respectively. PCR and cycle sequencing were done in a Perkin–Elmer 9700 GeneAmp PCR System. PCR products were GeneCleaned with glassmilk (GeneClean II kit; Bio 101). Sequencing was performed by the dideoxy termination method with dye-labeled terminators using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq DNA polymerase and run on the ABI Prism 377 DNA sequencer and ABI Prism 3700 DNA analyzer (Perkin–Elmer). Complementary strands were combined and edited with the computer program Sequencher 3.1 (Gene Codes).

Our 16S sequence corresponds to positions 13,480–13,850 of the *Apis mellifera* sequence (Crozier and Cro-

zier, 1993), the 18S sequence to positions 600–1582 of the *Drosophila melanogaster* sequence (Tautz et al., 1988), the 28S sequence to positions 4067–4607 of the *D. melanogaster* sequence (Tautz et al., 1988), and the CO1 sequence to positions 1861–2904 of the *A. mellifera* sequence (Crozier and Crozier, 1993). The 16S partition contains approximately 252 informative positions ($\approx$59%), the 18S partition contains approximately 105 bp ($\approx$12%), the 28S partition contains approximately 116 bp ($\approx$15%), and the CO1 partition contains approximately 550 bp ($\approx$53%), together approximately 1023 bp ($\approx$32%). (These numbers were determined with Winclada (version 0.99.9, Nixon, 2001) from alignments generated by ClustalX (version 1.5b, Thompson et al., 1997) with default options. The exact number of informative positions depends on the alignment and is hence variable between analyses.) Each taxon is represented by all four genes. For the origin of the sequences see Appendix C.

Cladistic analysis

The sequences were first aligned with the computer program MALIGN (Wheeler and Gladstein, 1991–1998, 1994) and then split into subfragments, to speed up the analysis and to prevent clearly homologous (i.e., invariant) regions from being treated as non-homologous. For the position of these subfragments relative to the sequences of *Apis* and *Drosophila* and the alignment of Whiting et al. (1997) see Appendix D. Although all four genes were sequenced for all taxa, some subfragments are missing for a few taxa; Appendix E provides details.

For the phylogenetic analysis, the gaps were removed from the subfragments to obtain unaligned fragments of the genes. The unaligned sequences were analyzed by using the "direct optimization" or "optimization alignment" method developed by Wheeler (1996) and implemented in the computer program POY (Gladstein and Wheeler, 1997) (available at ftp.amnh.org/pub/molecular/poy). Due to the mechanics of the method, the costs for indels, transversions, and transitions necessarily have to be identical for alignment and analysis. Therefore, both the alignment and analysis parameters are varied and tested in a sensitivity analysis using the computer program POY.

The morphological data matrix (Vilhelmsen, 2001) was also analyzed with POY, in the case of the simultaneous trees together with the DNA sequences. Some of the characters were treated as additive (ordered), as described in Vilhelmsen (2001). In the combined analyses, morphological character state transformations were assigned the same cost as the indel cost.

Leading and trailing gaps in DNA sequences were weighted one-half internal gaps for homology purposes but this length contribution was then subtracted (-noleading). SPR and TBR branch swapping, tree fusing, and tree drifting (10 rounds based on SPR and another

10 on TBR) were performed with 16 multiple random addition sequences. This was run on a 256×500 MHz Pentium III CPU cluster at the AMNH. (POY command line: -fusemaxtrees 10 -holdmaxtrees 25 -repintermediate -multirandom -norandomizeoutgroup -controllers 16 -noleading -seed -1 -slop 0 -checkslop 5 -parallel -jobspernode 2 -multibuild 16 -buildspr -treefusespr -nodiscrepancies -buildmaxtrees 2 -noapproxbuild -maxtrees 2 -fitchtrees -treefuse -fusemingroup 5 -fuselimit 25 -random 16 -driftspr -numdriftspr 10 -numdriftchanges 30 -drifttbr -numdrifttbr 10 -fuseafterreplicates).

A sensitivity analysis sensu Wheeler (1995), as outlined under Background, was performed on the data. The chosen parameter sets are the nine possible combinations of the gap:transversion cost ratios 1:1, 2:1, and 4:1 and the transversion:transition cost ratios 1:1, 2:1, and 4:1. For each parameter set, the individual genes, the combined molecular data, and the combined molecular and morphological data were analyzed. (In the following, "molecular tree" means the (consensus of the) most-parsimonious tree(s) of a simultaneous analysis of all four genes under one parameter set.) From the lengths of these trees (as well as the maximum length of all the data), three RILD measures were calculated (Appendix F).

Results

Morphology

The analysis of our morphological matrix (which is the matrix of Vilhelmsen (2001) minus five taxa, see above) with all characters weighted equally, and some characters treated as additive (as specified in Vilhelmsen, 2001), generated one most-parsimonious tree of 835 steps (calculated with NONA, POY, and PAUP), which is shown in Fig. 1.

Sensitivity analysis

The four genes were analyzed separately and simultaneously for each of the nine parameter sets. A simultaneous analysis of the morphological and molecular data was also performed for each parameter set. The tree lengths resulting from these 54 analyses are given in Appendix F. The length of the morphological tree (835 steps) was multiplied by a weight of 2, 4, 8, or 16 (equal to the indel weight) to calculate the incongruence. Appendix F shows the resulting RILD values. The best RILD value for the molecular analyses was obtained from the 1:1:1 parameter set, the best RILD values for the simultaneous analyses of all data was obtained from the 4:1:1 set.

Molecules

The genes were analyzed separately and together, each under all nine parameter sets. The lengths of the
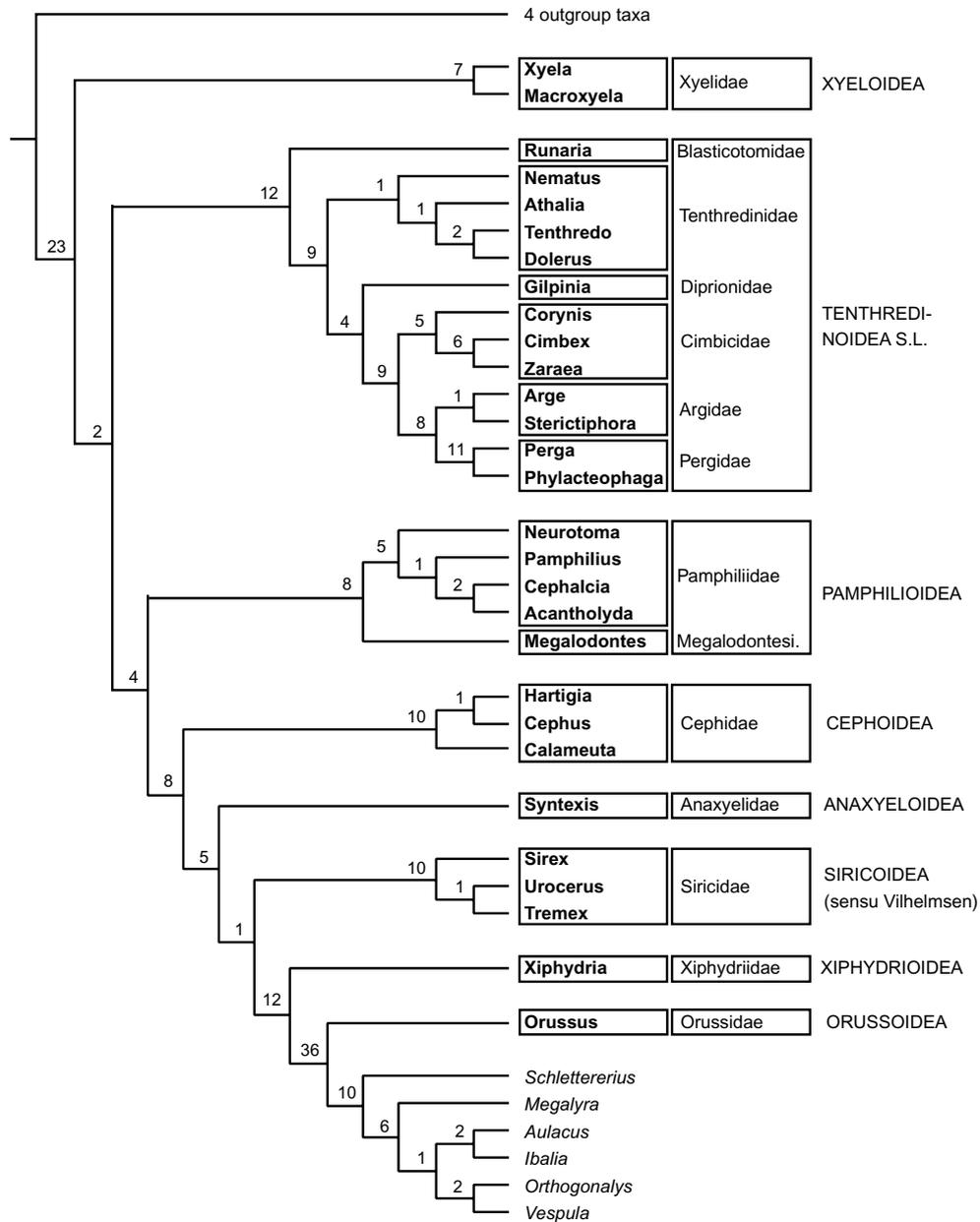
Fig. 1. The most-parsimonious tree that resulted from partly additive analysis of the morphological data matrix of Vilhelmsen (2001), after the removal of five taxa (see text). Numbers are Bremer values determined with POY. Names of genera are shown on the left; those belonging to the same family are framed. Family names are given in the middle; those belonging to the same superfamily are framed. Superfamily names shown on the right are those proposed by Vilhelmsen (2001). Note that his Siricoidea does not correspond to the concept of Siricoidea used in the present paper. Apocritan taxa are in italics and basal hymenopterans are in bold face type.

resulting most-parsimonious trees are given in Appendix F. The most-parsimonious solutions of the separate analyses of the four individual genes analyzed with the parameter regime 1:1:1 (which led to the optimal Molecular RILD) are shown in Fig. 2. The most-parsimonious results generated from simultaneous analyses of all molecular data are shown in Fig. 3 for all nine parameter sets. The implied alignment, which shows the positional homologies implied by the simultaneous analysis 4:1:1, is available on the *Cladistics* website. A majority-rule consensus tree showing those clades that

were present in five or more of the nine molecular trees of Fig. 3 is shown in Fig. 4.

Simultaneous analysis

The most-parsimonious cladograms resulting from simultaneous analyses of all data under all nine parameter sets are shown in Fig. 5. The information from these nine simultaneous trees is summarized in a majority-rule consensus tree (Fig. 6A) and a strict consensus tree (Fig. 6B). The sensitivity plots in Fig. 6A show in which of
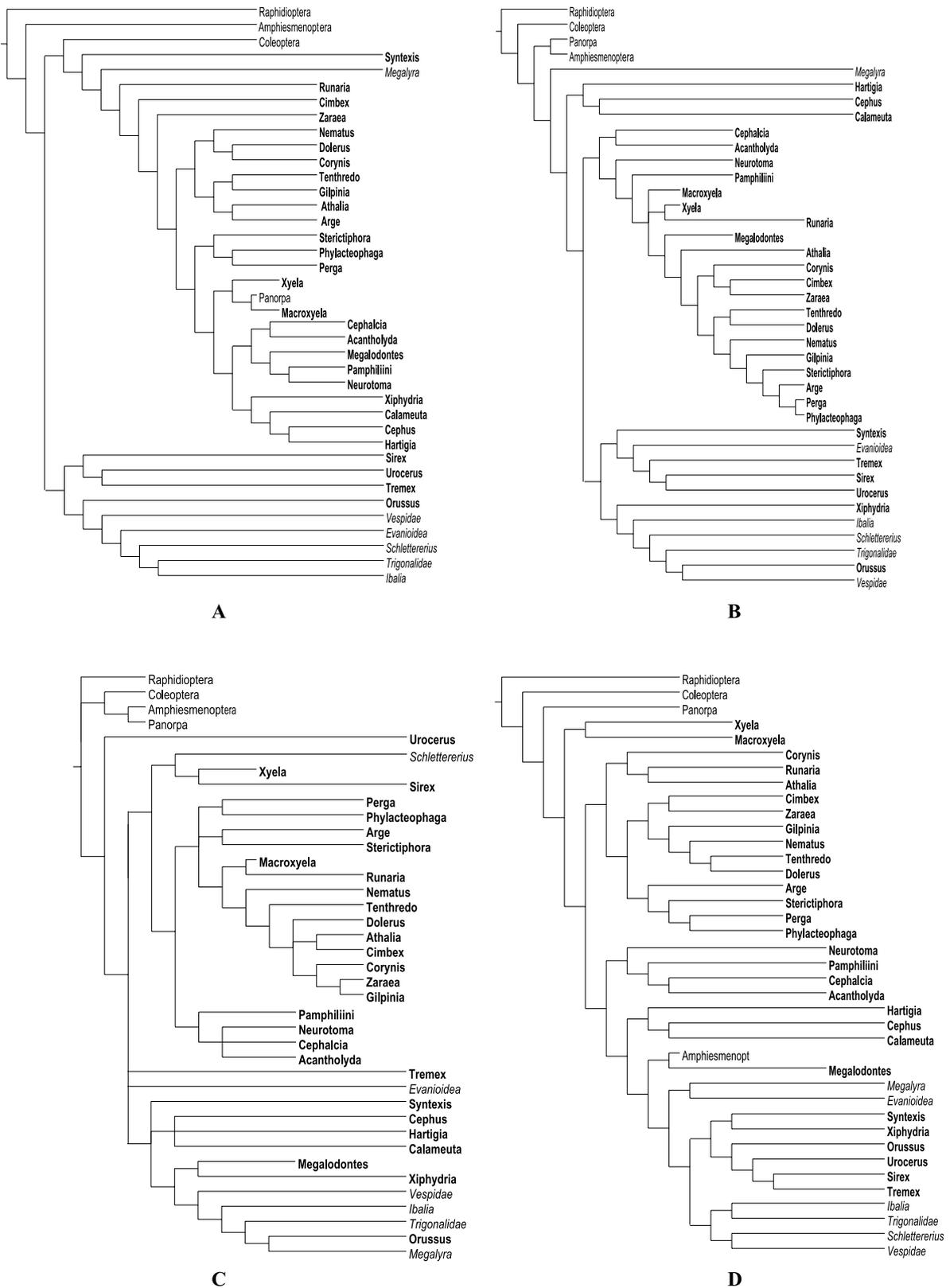
Fig. 2. The most-parsimonious trees that resulted from the analyses of the four individual genes with an indel:transversion:transition ratio of 1:1:1. Sawflies are in bold face type, apocritans are in italics, and outgroup taxa are in regular type. All trees were rooted on Raphidioptera. (A) 16S rDNA; (B) 18S rDNA, consensus; (C) 28S rDNA, consensus; (D) cytochrome oxidase 1.
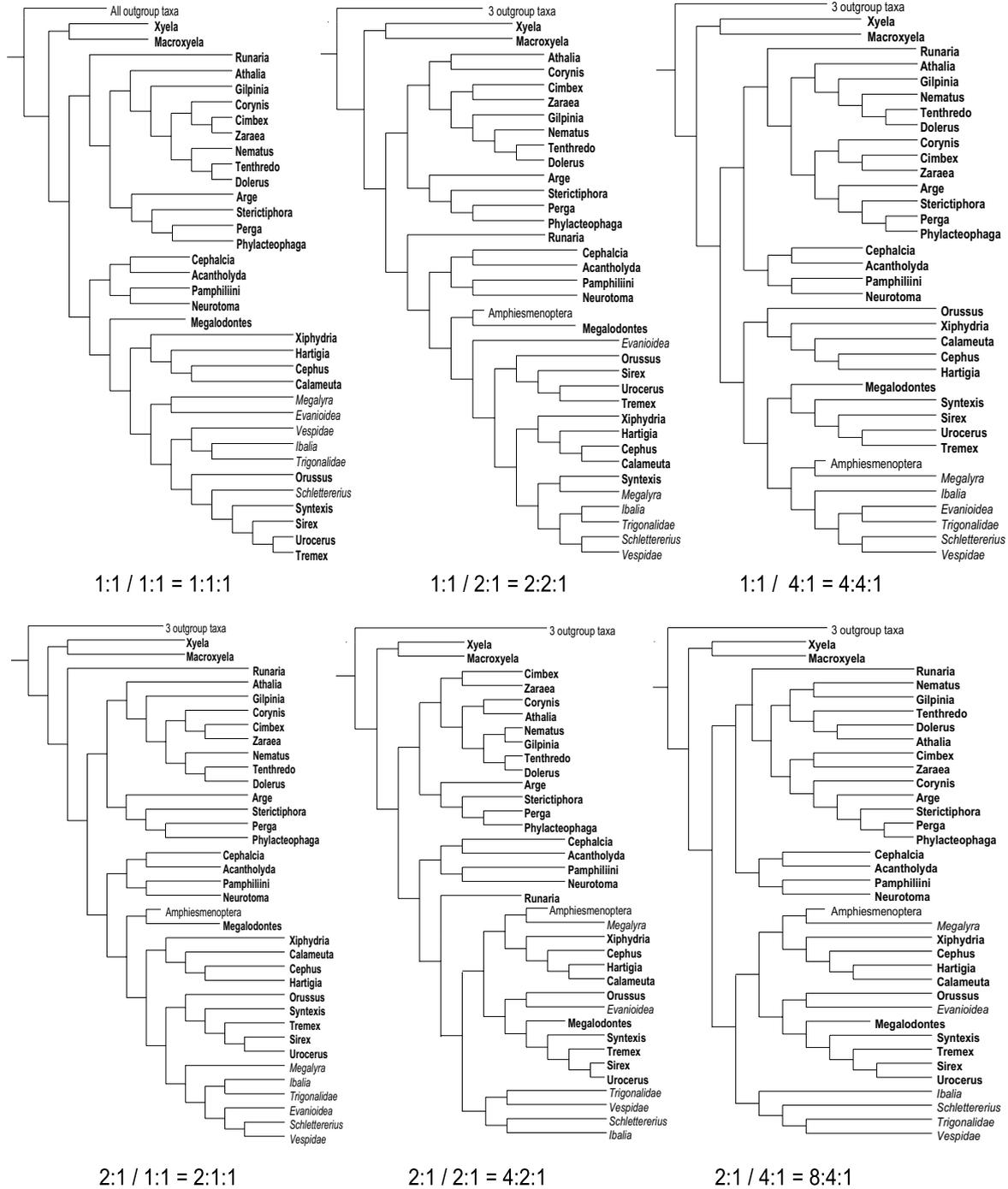
Fig. 3. The most-parsimonious trees that resulted from the analysis of all molecular data under different parameter sets. The numbers below each cladogram represent indel-to-transversion ratio, transversion-to-transition ratio, and indel-to-transversion-to-transition ratio. Sawflies are in bold face type, apocritans are in italics, and outgroup taxa are in regular type. All trees were rooted on Raphidioptera + Coleoptera. The 4:1:1 cladogram is the strict consensus of two equally parsimonious cladograms, which are shown on the bottom of the next page.

the nine parameter sets the clades came out as monophyletic (dark squares). The simultaneous analysis in which morphological transformations and insertion/deletion events were weighted four times base substitutions (transversions equal to transitions) exhibited the lowest value of character incongruence between the data

(Appendix F). The most-parsimonious cladogram obtained from this analysis is shown in the bottom left of Fig. 5. The final hypothesis, shown in Fig. 7, corresponds to the 4:1:1 cladogram, except that intrafamiliar relationships have been omitted, because our taxon sampling does not allow conclusions at this level (see below).

4:1 / 1:1 = 4:1:1   4:1 / 2:1 = 8:2:1   4:1 / 4:1 = 16:4:1
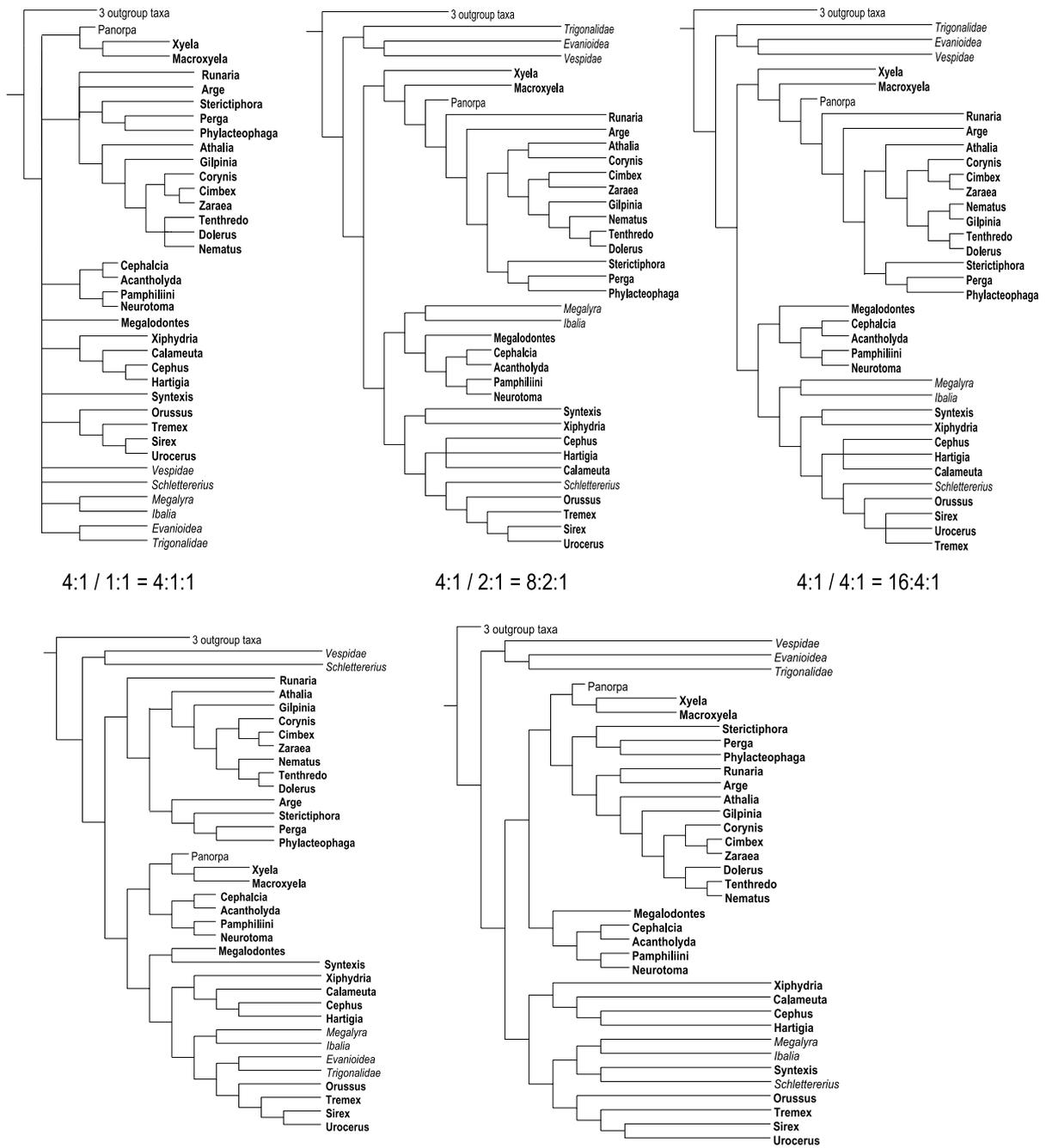
Fig. 3. (*continued*)

Discussion

Names for higher taxa

As mentioned under Background, we follow Rasnitsyn (1988) in using the name **Vespina** for the Orussidae + Apocrita clade. This should not be taken to mean that we agree with the nomenclatural system followed in that paper, which among other things substitutes the name "Vespida" for Hymenoptera. Rasnitsyn's system is part of an effort to standardize (and typify) names for taxa above the family group level, which is beyond the limits of the current International Code of Zoological Nomenclature. While we see merit in standardized (and typified) names at higher levels, any attempt to impose such a system unilaterally, outside of the international organizations that govern the codes of biological nomenclature, will certainly fail, as for example happened with Shipley's (1904) proposal to standardize insect ordinal names.

Raphidioptera, Coleoptera, Panorpa

Xyela
Macroxyela
XYELOIDEA

Runaria
Gilpinia
Nematus
Tenthredo
Dolerus
Athalia
Corynis
Cimbex
Zaraea
Arge
Sterictiphora
Perga
Phylacteophaga

TENTHREDINOIDEA S. L.

Neurotoma
Pamphiliini
Cephalcia
Acantholyda

Amphiesmenopt.
Megalodontes
Hartigia
Cephus
Calameuta
Xiphydria
Syntexis
Sirex
Urocerus
Tremex
Orussus
Schlettererius
Megalyra
Evanioidea
Ibalia
Trigonalidae
Vespidae

SIRICOIDEA

1:1:1	2:2:1	4:4:1
2:1:1	4:2:1	8:4:1
4:1:1	8:2:1	16:4:1

Fig. 4. Molecular stability tree: a majority-rule consensus (> 50%) of the molecular trees in Fig. 3 showing those clades that are present in five or more of the nine parameter sets of all molecular data analyzed together. The dark fields in the sensitivity plots indicate those parameter sets in which the respective group came out as monophyletic. Order of parameter sets as in Fig. 3 (see box at the bottom left of this figure).

The Cephidae + Anaxyelidae + Siricidae + Xiphydrii-dae + Vespina clade is very well supported in the simultaneous analyses and shows a number of morphological synapomorphies that are unique and unreversed within Hymenoptera (see below). To avoid this unwieldy term in the future, we decided to name this clade and propose the name **Unicalcarida** (from *unus*, lat.: one, only, sole, and *calcar*, lat.: spur). A prominent synapomorphy of this group is the reduction of the posterior apical spur on the front tibia (Basibuyuk and Quicke, 1995). It is reduced completely in almost all species of this clade, with the exception that some spe-

cies of the genera *Xiphydria* and *Orussus* have a very small second spur and some species of Agaonidae and Ichneumonoidea also show a second spur (Basibuyuk and Quicke, 1995). But the ancestor of this clade most probably had only one apical protibial spur. In entomology, the word calcar is commonly used for any tibial spur (see, for example, Kenneth (1960) and von Kéler (1963)). However, some hymenopterists (e.g., Brothers, 1975; Rasnitsyn, 1988; Basibuyuk and Quicke, 1995) use the word calcar exclusively for a spur that is modified in a certain way into a cleaning spur. Depending on which of these usages of the word calcar is preferred, the name
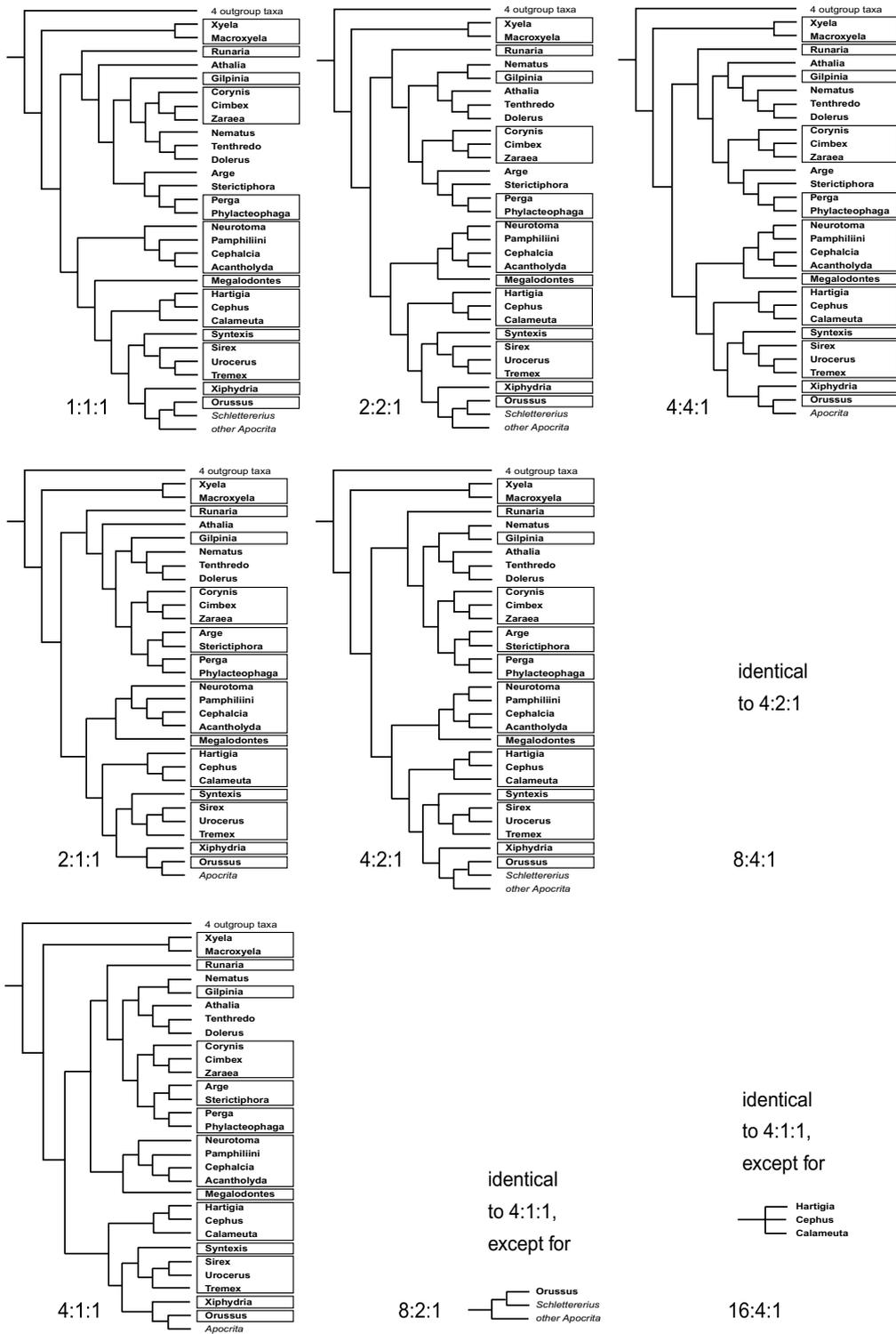
Fig. 5. The most-parsimonious trees that resulted from simultaneous analyses of the morphological data (Vilhelmsen, 2001) and the molecular data under different parameter regimes.

Unicalcarida can be interpreted either as "those having only one apical protibial spur" or "those having only a calcar and no other spur on the front tibia."

The Anaxyelidae + Siricidae clade is supported by the molecular as well as the combined evidence (see below).

Since each clade of basal hymenopterans which branches off from the main stem leading to Apocrita is usually assigned superfamily status, the Anaxyelidae + Siricidae clade is considered a superfamily in the present paper and accordingly called **Siricoidea**.
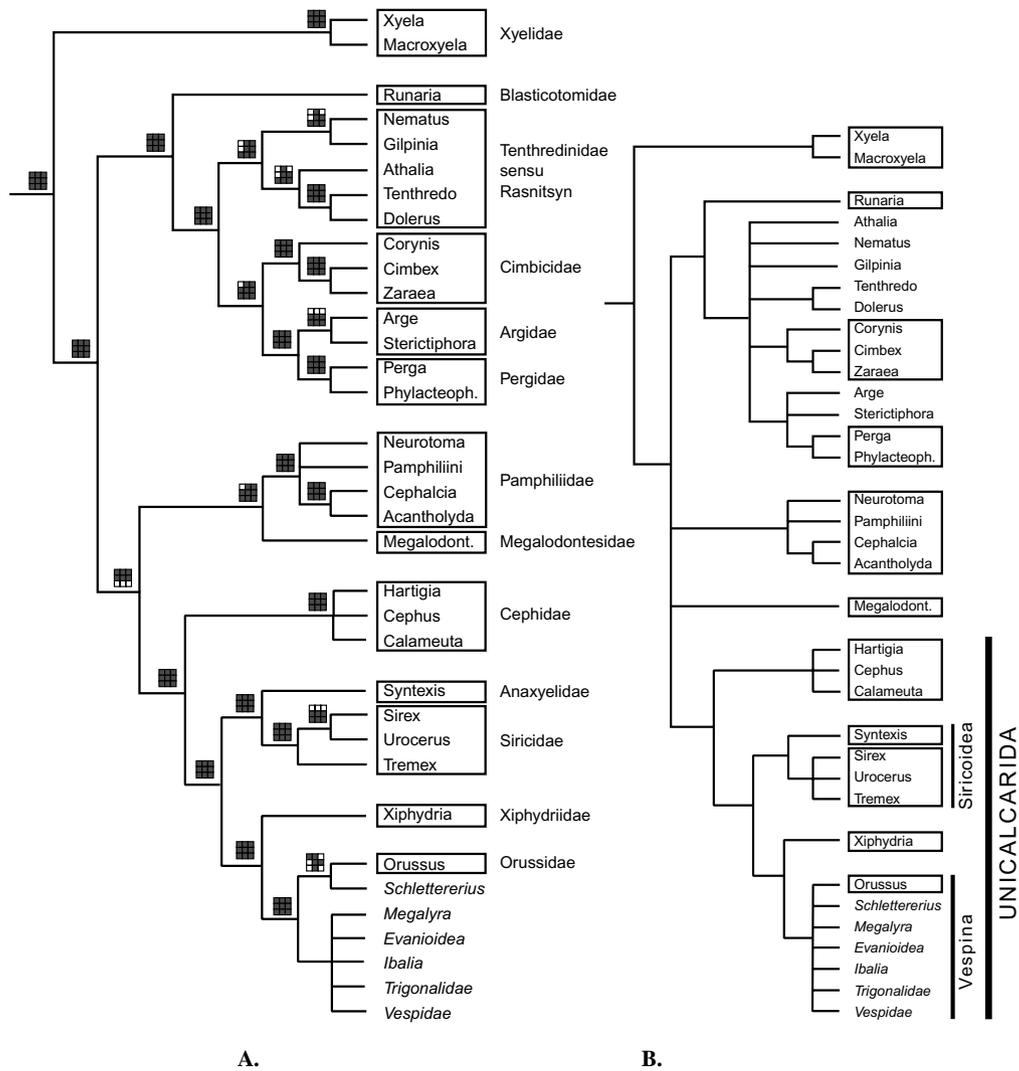
Fig. 6. Trees summarizing the cladograms in Fig. 5. (A) Stability tree showing those clades that are present in at least six of the nine simultaneous trees. Sensitivity plots as in Fig. 4. (B) Strict consensus. Names of genera belonging to the same family are framed. Tenthredinidae sensu Rasnitsyn is Tenthredinidae including Diprionidae.

Morphology

The relationships resulting from our reduced morphological data matrix (Fig. 1) differ from those resulting from the complete data matrix (Vilhelmsen, 2001) only in one point: while the tenthredinid taxa *Athalia*, *Nematus*, *Tenthredo*, and *Dolerus* are paraphyletic in the tree of Vilhelmsen (2001), they are monophyletic in our tree (Fig. 1).

The best-supported taxa in the morphological tree, as seen by the Bremer values in Fig. 1, are Vespina (36), Hymenoptera (23), Xiphydriidae + Vespina (12), Tenthredinoidea s.l. (12), Cephidae (10), and Siricidae (10). The weakest support, apart from intrafamiliar relationships and those within Vespina, is for Tenthredinidae (1), Argidae (1), Hymenoptera-Xyelidae (2),

and the Siricidae + Xiphydriiidae + Vespina clade (1). The weakness of support for the last three clades is also shown by the fact that they are not present in some of the eight trees resulting from nonadditive analysis of Vilhelmsen's, 2001 data matrix. The weakness of support for Tenthredinidae is also demonstrated by the fact that they come out as paraphyletic in the additive as well as nonadditive analyses of the complete morphological data set (Vilhelmsen, 2001, pp. 427 and 430).

Molecules

The separate analyses of the four genes are depicted in Fig. 2. The cladograms resulting from analysis of the combined molecular evidence under the nine different
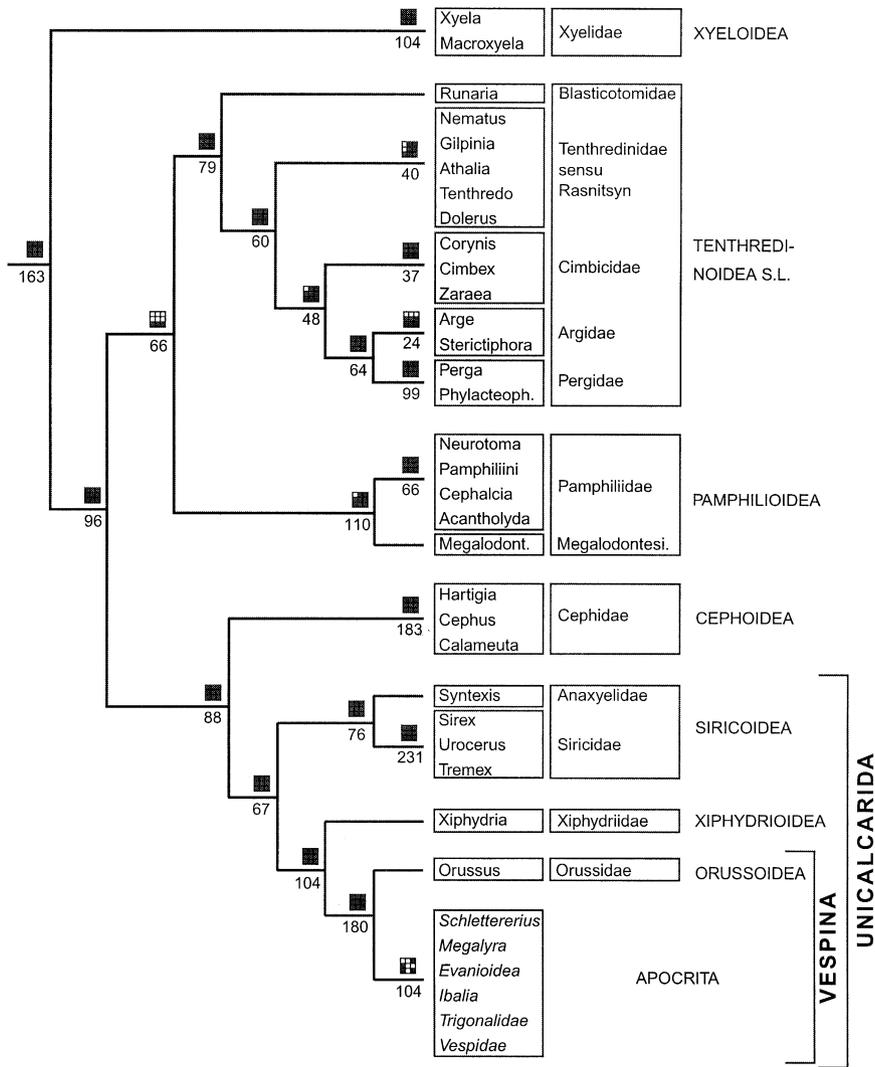
Fig. 7. Final hypothesis of the relationships of the basal lineages of Hymenoptera. The tree is the most-parsimonious tree resulting from the most congruent simultaneous analysis (4:1:1) (Fig. 5), except that the relationships within the sawfly families and within Vespina have been omitted. Tenthredinidae sensu Rasnitsyn is Tenthredinidae including Diprionidae.

parameter regimes are depicted in Fig. 3. To facilitate comparison, the outgroup taxa, and the exemplars belonging to Xyeloidea, Tenthredinoidea s.l., Pamphilioidea, and Unicalcarida (which are all well supported in the simultaneous analyses) have been aligned in the trees. The 4:1:1 analysis resulted in two most-parsimonious trees, which are shown at the bottom of Fig. 3 (continued).

The cladograms shown in Fig. 3 are substantially different. Many groups appear only in a few trees; i.e., they are highly unstable. To show those clades that are relatively stable, a majority-rule consensus tree was made from the nine cladograms resulting from the nine molecular analyses. This particular kind of majority-rule consensus tree, which summarizes cladograms resulting from repeated analyses of the same data with different parameter values, is here called a stability tree to dis-

tinguish it from the usual majority-rule consensus tree, which is made from equally most-parsimonious trees resulting from a single analysis. In this case, the stability tree (Fig. 4) shows those clades that are present in at least five of the nine molecular trees. Apparently, it shows a strong lack of resolution.

The only clades that are present in all molecular trees are *Cimbex* + *Zaraea*, *Perga* + *Phylacteophaga* (Pergidae), *Sterictiphora* + *Perga* + *Phylacteophaga*, Pamphiliinae, Cephalciinae, Pamphiliidae, Cephidae, and Siricidae. Except for *Sterictiphora* + *Perga* + *Phylacteophaga*, and Pamphiliinae, these are all in agreement with the morphological tree. However, Argidae and *Pamphilius* + Cephalciinae are only weakly supported by morphology. The monophyly of Cimbicidae, well supported by morphology, is supported by five of the molecular trees.

A striking feature of the molecular trees is that Hymenoptera is monophyletic in only one of the nine trees (1:1:1). If the trees are rooted on Raphidioptera + Coleoptera (the most distant outgroups), the trichopteran (OTU Amphiesmenoptera) appears within Hymenoptera in five of the trees; in two of these, it groups with *Megalodontes*, and in the three other cases, it is the sister group to *Megalyra*. In three other trees, those with a gap:change cost ratio of 4:1, *Panorpa*, instead of the trichopteran, appears within Hymenoptera. The trichopteran sequences often differ strongly from those of the other 38 taxa and many of their fragments are shorter and therefore align ambiguously. This is an indication that the placement of the trichopteran within Hymenoptera might be due to "random outgroup attraction" (Wheeler, 1990). This phenomenon occurs if the DNA sequences of a taxon are so different from those of the others that the similarity is about random (25% with equal base frequencies), so that the taxon is placed with that taxon which happens to have the highest (random) similarity.

Among the higher taxa present in the molecular stability tree (Fig. 4) are Tenthredinoidea s.l., Tenthredinoidea s.str., and the Argidae + Pergidae clade, all of which are well supported by morphology as well as the combined evidence. The stablest higher taxon is Tenthredinoidea s.str., which is monophyletic in eight molecular trees.

The monophyly of the siricoidean exemplars (= Anaxyelidae + Siricidae, see above) is supported by five of the molecular trees. This group is paraphyletic in the morphological tree, but the morphological support for a monophyletic Siricidae + Xiphydriidae + Vespina is weak (Fig. 1), as discussed above and by Vilhelmsen (2001).

The strongest disagreement between the molecular and the morphological trees is found in the two clades Cimbicidae + Diprionidae + Tenthredinidae and *Xiphydria* + Cephidae. Both are present in seven molecular trees. The alternative positions of both Cimbicidae and Xiphydriidae are well supported in the morphological tree, see Fig. 1. However, the position of Cimbicidae as the sister taxon to Tenthredinidae + Diprionidae is in agreement with the hypothesis of Rasnitsyn (Rasnitsyn, 1988; and see Ronquist et al., 1999). The placement of *Xiphydria* as a sister taxon to the cephids in seven of the nine molecular trees, which stands in stark contrast to its placement in the morphological and simultaneous trees, can be explained to a large extent by a prominent insertion in the 16S gene (Fig. 8) shared by these taxa. In the molecular analyses placing *Xiphydria* with cephids, POY interpreted these insertions as homologous and therefore as synapomorphies for *Xiphydria* and Cephidae. The assumption that the insertions are the main basis for this sister-group relationship is supported by the fact that the *Xiphydria* + Cephidae clade appears in

only one of the four individual gene trees (Fig. 2), viz., the 16S tree, and by the fact that in an analysis in which we excluded this fragment (and some others), the clade did not appear in any of the molecular trees (results not shown). A sequence motif, which is identical in *Calameuta* and *Xiphydria*, speaks for a homologous origin of these insertions. However, in the simultaneous trees, in which there is obviously a strong opposition to this clade (coming from the morphological partition), these insertions were interpreted as nonhomologous by POY. The effect of the 16S insertion on the molecular trees is enhanced by the treatment of indel events in our analysis. If the entire insertion was counted as a single character or if we had used affine gap costs in our analysis (which were not yet implemented in POY), the additional cost of assuming a convergent origin of this insertion would be much smaller. But because our analysis treats each position independently, there are at least 11 insertion events as synapomorphies for the *Xiphydria* + Cephidae clade.

Combined evidence

Intrafamiliar relationships

To make conclusions about the phylogeny within families, at least two members from each subfamily should be included to test the monophyly of the subfamilies. The taxon sampling of Vilhelmsen's (1999a, 2001) and our studies therefore does not allow conclusions about the relationships within the families, except perhaps in Pamphiliidae. The branches within the families have therefore been collapsed in the final hypothesis (Fig. 7).

In the morphological tree, *Pamphilius* appears more closely related to Cephalciinae than to the other exemplar of Pamphiliinae, *Neurotoma*. The pamphiliine exemplars used for DNA sequencing are *Neurotoma* and *Onycholyda*, which are sister taxa in all nine molecular trees (Fig. 4). This might indicate that Pamphiliini as well as Pamphiliinae could be paraphyletic, with *Onycholyda* being more closely related to *Neurotoma*, while *Pamphilius* might be more closely related to Cephalciinae. This contradiction persists in the simultaneous trees (in which the data from *Pamphilius* and *Onycholyda* were combined in the OTU Pamphiliini): in four of the parameter sets, the exemplars of Pamphiliinae are monophyletic, in five they are paraphyletic (Fig. 5). To resolve this ambiguity, a future study should cover all three known genera of Pamphiliinae as separate OTUs.

Monophyly of families

In the exemplar approach, which was employed in Vilhelmsen's (2001) as well as our studies, a reasonably stringent test of the monophyly of a family can be performed only if representatives from both monophyletic

```
Xyela          CAC--TAATAAGTGAGTTTTT--------------------------------GATCCT
Macroxyela     CAT--TAATGTATGAATTTTT--------------------------------GATCCT

Runaria        CAT--TAATATATGATTTTATT-------------------------------GATCCT
Tenthredo      CAT--TTATTAATGATTTTTT--------------------------------GATCCT
Dolerus        CAT--TAATTTATGAATTTTT--------------------------------GATCTT
Athalia        CAT--TGATTTATGAATTTTT--------------------------------GATCTT
Nematus        CAT--TTATTAATGATTTATT--------------------------------GATCCT
Gilpinia       CAT--TTATTAATGATTTTAT--------------------------------GATCCT
Cimbex         CAT--TGATAAATGAAATTTT--------------------------------GATCCT
Corynis        CAT--TTATTTATGAATTTTT--------------------------------GATCCT
Arge           CAT--TAATTTATGAATTTTT--------------------------------GATCCT
Sterictiph.    CAT--TGATGAATGATTTTTT--------------------------------GATCCT
Perga          CAA--AGATTATTGAAATGTT--------------------------------GATCCT
Phylacteoph.   CAA--TGATGTTTGAATTTTT--------------------------------GATCCT

Megalodontes   CAT--AGATTTATGTTTAAATTT------------------------------TATCTT
Cephalcia      CAT--AAATTAATGATTTTTTTTTT----------------------------GATCTT
Acantholyda    CAT--TAATTAATGATTTTATTAATT---------------------------GATCTT
Pamphiliini    CAT--AAATTAATGATTTATTAATA----------------------------GATCCT
Neurotoma      CAT--TAATTAATGATTTATAAATA----------------------------GATCCT

Hartigia       CAT--TAATTTATGAATTTTTAATAATTAATTAAT---TATTAATTATT------GACCCT
Cephus         CAT--TAATTTATGAAATTTAATATATAAATAATTTATATTTATTATTAATTTTTGATCCT
Calameuta      CAT--TAATTTATGAATTTTTAATATAT------TTTATATTAATTATT------GATCCT

Syntexis       CAT--TAATTTATGAAAATTATT--------------------------------GATTTT
Sirex          CAA--AAATTATTGAATTTTT----------------------------------GAGTAT
Urocerus       CAA--AAGTAATTGAATTTTT----------------------------------GAGTGT
Tremex         CAA--TAATATTTGATTAAAT----------------------------------GAAAAT

Xiphydria      CAT--TAATTTATGAAATTAAAAAAAAATAAAA----ATATTAATTA------GATGATCTT

Orussus        CAA--TAATTTTTGATAAAATT---------------------------------GATCCT

Schletter.     CAATTTAATAATTGAATTGAATTTATATA--------------------GTTGATCTT
Megalyra       CAT--TAATTTATGAATAATT----------------------------------GATCCA
Gasterupt.     CAT--TAATTAATGAAATTAA----------------------------------GTAATA
Labidogon.     AAT--ATATAAATTTAAATTAAAA-------------------------------GATCCA
Ibalia         CAT--TTATAAATGAATTAAAATATAAAA--------------------------GAGTGA
```

Fig. 8. Fragment of the 16S gene showing especially long insertions in *Xiphydria* and the cephid taxa *Cephus*, *Hartigia*, and *Calameuta*. The characters have been arranged (not aligned) by hand for illustrative purposes. They correspond to positions 13,619?–13,585 in the *A. mellifera* genome (Crozier and Crozier, 1993). The taxa are arranged in monophyletic groups, as in the final hypothesis presented in this paper (Fig. 9). The insertions and certain motifs (which are identical in different families) are highlighted.

subclades are included. This requires that the relationships within that family are known to some extent and that the monophyly of the subclades is well supported. Unfortunately, this is not the case for most families of basal hymenopterans because phylogenetic analyses are missing. We can only compromise by assuming that the subfamilies constitute monophyletic groups and that a sampling of all subfamilies would therefore suffice to assess monophyly of the family. As long as these representatives group together, we have at least no indication of paraphyly. With that in mind, we can look at the 14 sawfly families. The information about the subfami-

lies of each family was taken from Abe and Smith (1991).

Xyelidae. The representatives of the two xyelid subfamilies are sister groups in all of the nine simultaneous trees (Fig. 6B), which confirms the findings of Vilhelmsen (2001) and Ronquist et al. (1999). But assessing the monophyly of Xyelidae is a problem because Xyelinae and Macroxyelinae are probably the most basal symphytans and the presumed closest living relatives of Hymenoptera are all highly derived (Ronquist, 1999). Moreover, the monophyly of Xyelidae is supported only by homoplastic morphological characters, no unique

morphological synapomorphies have been found (Vilhelmsen, 2001).

Blasticotomidae. This is an extremely rare family and we were not able to sequence DNA from more than one species. However, Vilhelmsen (2001) included representatives of three genera, which might be all extant genera of Blasticotomidae, a fourth being doubtful (Shinohara, 1983). The three taxa grouped together in his analysis and their monophyly is well supported by seven unreversed autapomorphies, three of which are unique (Vilhelmsen, 2001).

Tenthredinidae. This is the largest of all "symphytan" families. Abe and Smith (1991) list eight subfamilies; their monophyly and relationships have yet to be tested. Representatives of four subfamilies came out as paraphyletic in Vilhelmsen's (2001) analyses, but monophyletic in our reduced morphological analysis (Fig. 1). Nonetheless, the tenthredinid exemplars are paraphyletic in all our simultaneous analyses (Fig. 5). However, the four tenthredinids together with the diprionid *Gilpinia* come out as monophyletic in seven simultaneous trees, among them the tree that minimizes total character incongruence. But because this covers only half of the subfamilies of Tenthredinidae, this finding constitutes only a very weak indicator for the monophyly of Tenthredinidae sensu Rasnitsyn.

Diprionidae. We included only one species in our analysis, but the exemplars of the two subfamilies, Monocteninae and Diprioninae, group together in Vilhelmsen's study (2001), and the apomorphic form of the male antennae present in all diprionid taxa is a good synapomorphy for the whole family. (This form of the antennae also occurs in some species of Pergidae, but this is apparently due to convergence, because the monophyly of Pergidae and Argidae + Pergidae is well supported.)

Cimbicidae. Representatives of all three subfamilies were included in the study. They are monophyletic in all simultaneous trees (Fig. 6B).

Argidae. The two argid exemplars are monophyletic in two-thirds of the simultaneous trees, including the preferred hypothesis. They come out as monophyletic whenever indels and morphological character transformations are weighted higher than transversions and transitions (2:1 or 4:1).

Pergidae = Pterygophoridae. Abe and Smith (1991) list 14 subfamilies. The monophyly of these subfamilies and the relationships among them have never been examined. Only two of them are included in Vilhelmsen's (2001) and our analyses, but these always group together. In a molecular analysis of a larger taxon sample (at 2:1:1), representatives of four pergid subfamilies (Perginae, Phylacteophaginae, Pterygophorinae, and Acorduleceri-nae) also came out as monophyletic (unpublished results). But this still does not say much about the monophyly of Pergidae, and a phylogenetic analysis covering many

more pergid subfamilies will be necessary to allow any conclusion about the monophyly of Pergidae.

Pamphiliidae. Two representatives from each of the two extant subfamilies are included in the analyses. They come out as monophyletic in all simultaneous analyses; the monophyly of Pamphiliidae hence seems to be well established. For the status of the subfamilies, see above.

Megalodontesidae. Only one species of this family was included, so its monophyly could not be tested.

Cephidae. There are two subfamilies, one of which, Athetocephinae, consists of only one genus (Abe and Smith, 1991). Cephinae is divided into the tribes Hartigiini, Cephini, and Pachycephini. The taxon sample of Vilhelmsen's (2001) and our studies include three cephids that belong to the first two tribes of Cephinae. Therefore, the monophyly of neither Cephidae nor Cephinae could be tested, but at least we can say that the three exemplars of Cephinae always group together. A couple of unique and identical insertions in the 28S gene (Fig. 9C) strongly support the monophyly of *Cephus* + *Calameuta* + *Hartigia*. Moreover, a number of morphological synapomorphies (lack of cenchri, slight constriction between first and second abdominal segments, lack of harpes in the male genitalia) present in all known genera of Cephidae indicate that the monophyly of Cephidae is quite likely.

Anaxyelidae. There is only one extant species, *Syntexis libocedrii*.

Siricidae. There are only two extant subfamilies, Tremecinae and Siricinae. Gauld and Mound (1982) found both to be monophyletic, with one synapomorphy for Siricinae and three for Tremecinae. Three representatives from the two subfamilies group together in all simultaneous analyses (Fig. 6B). In addition to a number of morphological synapomorphies found for these three genera (see Appendix G), there are some unique insertions that are present in all of the four siricid species sequenced in the present study (Figs. 9A and B). The sequence of the species *Xeris spectrum* (L.) shown in Fig. 9 was not included in the cladistic analysis presented here; however, in an unpublished molecular analysis of a larger taxon sample (2:1:1), the four siricids *Sirex*, *Xeris*, *Urocerus*, and *Tremex* came out as monophyletic. Therefore, little doubt is left about the monophyly of the Siricidae.

Xiphydriidae. There is only one genus included in Vilhelmsen's (1999a, 2001) and our analyses because the other genera are extremely rare. Therefore, monophyly could not be tested.

Orussidae. This subfamily exhibits the same problem as that which occurs in Xiphydriidae.

Relationships among families

Fig. 7 shows the relationships of the basal hymenopterans as obtained by the simultaneous analysis that minimizes the total character incongruence (gap:trans-

A.

```
Xyela        GACCGAGGGAGGATGGGC--------------------CGCG--------------------------------TCA------------------------CGAT----------------------------------------GCG
Macroxyela   GACCGAGGGAGGATGGGC--------------------CGCG--------------------------------TCA------------------------CGAT----------------------------------------GCG
Runaria      GACCGAGGGAGGATGGGC--------------------CGCG--------------------------------TCA------------------------CGAT----------------------------------------GCG
Tenthredo    GACCGAGGGAGGATGGGC--------------------CGCG--------------------------------TCA------------------------CGAT----------------------------------------GCG
Nematus      GACCGAGGGAGGATGGGC--------------------CGCG--------------------------------TCA------------------------CGAT----------------------------------------GCG
Cimbex       GACCGAGGGAGGATGGGC--------------------CGCG--------------------------------TCA------------------------CGAT----------------------------------------GCG
Zaraea       TGCCGAGGGAGGATGGGC--------------------CGCG--------------------------------TCA------------------------CGAT----------------------------------------GCG
Corynis      GACCGAGGGAGGATGGGC--------------------CGCG--------------------------------TCA------------------------CGAT----------------------------------------GCG
Sterictiph.  GACCGAGGGAGGATGGGC--------------------CGCG--------------------------------TCA------------------------CGAT----------------------------------------GCG
Perga        GACAGAGGGAGGATGGAT--------------------CGCG--------------------------------TCA------------------------CGAT----------------------------------------GCG
Onycholyda   GACCGAGGGAGGATGGGC--------------------CGCG--------------------------------TCA------------------------CGAT----------------------------------------GCG
Neurotoma    GACCGAGGGAGGATGGGC--------------------CGCG--------------------------------TCA------------------------CGAT----------------------------------------GCG
Cephalcia    TACCGAGGGAGGATGGGC--------------------CGCG--------------------------------TCA------------------------CGAT----------------------------------------GCG
Acantholyda  GACCGAGGGAGGATGGGC--------------------CGCG--------------------------------TCA------------------------CGAT----------------------------------------GCG
Cephus       GACCGAGGGAGGATGGGC--------------------CGCG--------------------------------TCA------------------------CGAT----------------------------------------GCG
Calameuta    GACCGAGGGAGGATGGGC--------------------CGCG--------------------------------TCA------------------------CGAT----------------------------------------GCG
Hartigia     GACCGAGGGAGGATGGGT--------------------CGCG--------------------------------TCA------------------------CGAT----------------------------------------GCG
Syntexis     GACCGAGGGAGGATGGGT--------------------CGCG--------------------------------AAA------------------------CGAT----------------------------------------GCG
Xeris        GGCCGAGGGAGGATGGGTCTC--GC--------------GGGGCGCG--CTCG---CGCG----------------TTC------------------------CGTC-----------------------------------------GGG
Sirex        GGCCGAGGGAGGATGGGTCTCCCGGGATCCTCCCTCTCGGGGCCGGCGA--CTCGGTGCGCGCGCGTCGCCGGGGAGCGGGCTCGTCCCGTTTTTCCTTGCTCGCGCGTGTCTCTCCGTTTCGCCGGTCAACCCCGGGGGGGGTTCCCAGAAATGGGGGCTCGGGCG
Urocerus     GGCCGAGGGAGGATGGGTTTT--GG--------------GGCGCCG--CTCG---------------------TCCGGT----------------------CGCT------------------------------------------CCG
Tremex       GGCCGAGGGAGGATGGGTTCT--GG--------------TCGCGAGGGTTCTCCC--CGAG----------------CGTCC-----------------------CG-T------------------------------------------
Xiphydria    GACCGAGGGAGGATGGGC--------------------CGCG--------------------------------TCA------------------------CGAT----------------------------------------GCG
Orussus      GACCGAGGGAGGATGGAC--------------------GGCG--------------------------------TCA------------------------CGAT----------------------------------------GCC
Schletterer. GACCGAGGGAGGATGGGC--------------------CGCG--------------------------------TCA------------------------CGAT----------------------------------------GCG
Trigonalidae GACCGAGGGAGGATGGAC--------------------CGCG--------------------------------TTT------------------------CGAT----------------------------------------GCG
```

Fig. 9. Fragments of the 28S gene. The bases have been arranged by hand for illustrative purposes. The sequence of *Xeris spectrum* was not included in the cladistic analysis. (A) The fragment corresponds to positions 122–179 in the 28S alignment of Whiting et al. (1997) (their alignment is missing the last 2 bp of this fragment). The beginning has no homolog in *Drosophila*; the end corresponds to position 4173 in the *Drosophila* sequence of Tautz et al. (1988). Prominent features are outlined. (B) The fragment corresponds to positions 4433–4466 in the *Drosophila* genome (Tautz et al., 1988). (C) Positions 4551–4593.

B.
```
Xyela      GGCACTCGCGTGC---GA-----AACGTACA----------------------------------------------------------CGAGTCTCATC
Macroxyela GGCACTCGCGTGC---GA-----AACGTACA----------------------------------------------------------CGAGTCTCATC
Runaria    GGCACTCGCGTGC---GA-----AACGTACA----------------------------------------------------------CGAGTCTCATC
Athalia    GGCACTCGCGTAC---AA-----AATGTACA----------------------------------------------------------CGAGTCTCATC
Gilpinia   GGCACTCGCGTAC---AA-----AACGTACA----------------------------------------------------------CGAGTCTCATC
Cimbex     GGCACTCGCGTAC---AA-----AACGTACA----------------------------------------------------------CGAGTCTCATC
Strictiph. GGCACTCGCGTGC---AA-----CTTGTACA----------------------------------------------------------CGAGTCTCATC
Perga      GGCACTCGCGTGC--GA-----AACGTACA----------------------------------------------------------CGAGTCTCATT
Onycholyda GGCACTCGCGTAC---TC-----AACGTACA----------------------------------------------------------CGAGTCTCATC
Neurotoma  GGCACTCGCGTAC---TC-----AAAGTACA----------------------------------------------------------CGAGTCTCATC
Cephalcia  GGCACTCGCGTGC---TC-----AAAGTACA----------------------------------------------------------CGAGTCTCATC
Acantholyda GGCACTCGCGTGC---TC-----AAAGTACA---------------------------------------------------------CGAGTCTCATC
Megalodon. GGCACTCGGGATT-CTTATTTTGAACG----------------------------------------------------------------CGAGTCTCATC
Cephus     GGCACTCG--TAC-GTTCTTATGAACTGAAG----------------------------------------------------------CGAGTCTCATC
Calameuta  GGCACTCG--TAC-GTTCTTATGAACTGAAG----------------------------------------------------------CGAGTCTCATC
Hartigia   GGCACTCG--TAC-GTTCTTATGAACTGAAG----------------------------------------------------------CGAGTCTCATC
Syntexis   GGCACTCGCGT---GTTCTCA-GAACTTGCG----------------------------------------------------------CGAGTCTCATC
Tremex     GGCGCTCG--TACCGCTC-----GACGCGGC--------G--CGCGAAAG------CG-CGC-CG---CG---CGG----TATCGAGTCTCATC
Urocerus   GGCGCTCGY-----GCCC-----AACGTGCC--------G--CACCTACC------CGGTGCGCG-CACG--T-G-----TATCGAGTCTCATC
Sirex      GGCGCTCGC-------T------AAC--GCC----GACCGACCACCGCCGGTCTTCCGGTGGGCGGTGCG-GTCGGT---TGTCGAGTCTCATC
Xeris      GGCGCTCGC-----GCCTT---------GGC-----AGGGGGTGGGCGAGTGTTT-CGGCACTCGTTACGCCTCCCTGCGTATCGAGTCTCATC
Xiphydria  GGCACTCGC---GAGTTCTCA-GAACTTAGG----------------------------------------------------------CGAGTCTCATC
Orussus    GGCACTCGC---GTGTTCCCAGGAACCTGTG----------------------------------------------------------CGAGTCTCATC
Megalyra   GGCACTCGC---CTGTTCTCATGAACTTGCG----------------------------------------------------------CGAGTCTCATC
```

C.
```
Megalodontes CTTGAACG-------------TTGAAGCCATGAGAT-----------TT---------------CGGATCAGAGTGCCAAGTGGG
Onycholyda   CTCGAACG-------------TTGAAGCCACGAGAT-----------TT---------------CGGATCAGAGTGCCAAGTGGG
Neurotoma    CTCGAACG-------------TTGAAGCCACGAGAT-----------TT---------------CGGATCAGAGTGCCAAGTGGG
Cephalcia    CTCGAACG-------------TTGAAGCCACGAGAT-----------TT---------------CGGATCAGAGTGCCAAGTGGG
Acantholyda  CTTGAACT-------------ATGAAGCCACGAGAT-----------TT---------------CGGATCAGAGTGCCAAGTGGG
Cephus       CTTGAACGAAAAATTTAAAAATTGAAGCCACGAGATA----------TTAT------TGAAATACGGATCAGAGTGCCAAGTGGG
Calameuta    CTTGAACGAAAAATTTAAAAATTGAAGCCACGAGATA----------TTAT------TGAAATACGGATCAGAGTGCCAAGTGGG
Hartigia     CTCGAACGAAAAATTTAAAAATTGAAGCCACGAGATACATATATCATTTATGATACATGAAATACGGATCAGAGTGCCAAGTGGG
Syntexis     CTTGAACG-------------TGAAGCCACGAGAT-----------TT---------------CGGATCAGAGTGCCAAGTGGG
Sirex        CTCGAACG-------------TTGAAGCCACGAGAT-----------T---------------CGGATCAGAGTGCCAAGTGGG
Xeris        CTCGAACG-------------TTGAAGCCACGAGAT-----------T---------------CGGATCAGAGTGCCAAGTGGG
Urocerus     CTCGAACG-------------TTGAAGCCACGAGAT-----------T---------------CGGATCAGAGTGCCAAGTGGG
Tremex       CTCGAACG-------------TTGAAGCCACGAGAT-----------T---------------CGGATCAGAGTGCCAAGTGGG
Xiphydria    CTTGAACG-------------TGAAGCCACGAGAT-----------TT---------------CGGATCAGAGTGCCAAGTGGG
Orussus      CTTGAACT-------------ATGAAGCCACGAGAT-----------CA--------------TGGATCAGAGTGCCAAGTGGG
Ibalia       CTTGAACG-------------ATGAAGCCACGAGAT-----------TT--------------CGGATCAGAGTGCCAAGTGGG
```

Fig. 9. (*continued*)

version:transition cost ratio = 4:1:1, Fig. 5, bottom left). The relationships within the sawfly families and within Apocrita have been omitted because the taxon sampling in this study is not sufficient to permit conclusions within these groups. The unambiguously optimized morphological changes on the cladogram of Fig. 7 are listed in Appendix G.

Hymenoptera come out as monophyletic, with a monophyletic Xyelidae as the sister group to all other hymenopterans. This is supported by all simultaneous trees, cf. the strict consensus in Fig. 6B.

The remaining Hymenoptera are divided into two monophyletic groups (Fig. 7): Tenthredinoidea s.l.+ Pamphilioidea and Unicalcarida. The monophyly of the first of these clades is, however, quite sensitive to a variation of the analysis parameters; it appears only in those analyses in which indel events and morphological transformations were weighted four times transversions (Fig. 5). This clade not only appears in some of our simultaneous analyses (and some molecular analyses), but also if Vilhelmsen's data matrices (1999a, 2001) are analyzed nonadditively in the computer program PeeWee with a concavity constant $k = 1$. (However, at k-values of 2 to 6, Pamphilioidea comes out as sister group to Unicalcarida.) According to Vilhelmsen (2001, p. 433), the cladogram with Tenthredinoidea s.l. and Pamphilioidea as sister groups derived from his new matrix by parsimony analysis, with some characters treated as additive, is only one step longer than the shortest tree. Hence it must be concluded that both the molecular and morphological data contain some support for this clade. But the morphological synapomorphies for the clade are weak, all of the characters (see Appendix G) are homoplastic, and four of them have been coded as unknown or inapplicable for many taxa.

In five simultaneous analyses, Pamphilioidea is instead the sister group of Unicalcarida. And in the cladograms resulting from the 1:1:1 simultaneous analysis, Pamphilioidea is even paraphyletic, *Megalodontes* then being the sister taxon to Unicalcarida. The sequences in Fig. 9B show insertions as potential synapomorphies of *Megalodontes* + Unicalcarida. At present, the hypothesis that places Pamphilioidea as sister to Tenthredinoidea s.l. is the one that best explains the data included in our analysis. (But due to the weak support for and lack of stability of this clade, we will refrain from assigning it superfamily status at present.)

The monophyly of Tenthredinoidea s.l. and that of Tenthredinoidea s.str. are stably and strongly supported by the combined evidence (Fig. 6B). Within Tenthredinoidea s.str., the Argidae + Pergidae clade shows up in all simultaneous analyses and exhibits high Bremer support values. It is hence well established. The sister-group relationship of Cimbicidae to this clade is present in eight of the nine simultaneous trees. In contrast, the tenthredinid exemplars are monophyletic in none of the simultaneous

trees. But Tenthredinidae sensu Rasnitsyn (i.e., including Diprionidae) comes out as monophyletic in seven of the nine simultaneous analyses, including the most congruent. It hence appears in our final hypothesis (Fig. 7). However, it must be kept in mind that our analysis included only four of the eight subfamilies of Tenthredinidae; Tenthredinidae sensu Rasnitsyn might still come out as paraphyletic in an analysis that includes members from all its subfamilies. The monophyly of the group must therefore still be viewed with much caution. A better taxon sampling within Tenthredinoidea s.str. is strongly desired.

Unicalcarida is monophyletic in all simultaneous trees; no reasonable alternative is suggested by the molecular data. It is supported by 11 unambiguously optimized morphological changes (see Appendix G). The posterior apical protibial spur (character 71) is missing or very small in all members of Unicalcarida—which is the case otherwise only in *Phylacteophaga*. *Raphidia* and *Chrysopa* are missing both apical spurs, so that even though they have also been coded as "1", their condition is not really equivalent to that found in Unicalcarida and *Phylacteophaga*. The distal epipharyngeal wall (character 6) is sclerotized and continuous with the labrum in all included Unicalcarida (unknown in two Apocrita), which is unique within Hymenoptera. The lateral attachment points of the mesopostnotum with the mesepimera (character 89) are invaginated or reduced in all Unicalcarida but in no other taxon included in this analysis. A mesofurcal bridge (character 99) is also a unique and unreversed synapomorphy of Unicalcarida. Thus, the monophyly of Unicalcarida is well supported.

Cephidae is the sister taxon to all other groups of Unicalcarida, which is supported by all simultaneous analyses. The next branch off the main stem leading to Apocrita is Siricoidea, which (here) comprises Anaxyelidae and Siricidae and comes out as monophyletic in all our simultaneous analyses. This arrangement is not in agreement with the hypotheses of Rasnitsyn (1988), Ronquist et al. (1999), and Vilhelmsen (2001), according to which Siricidae is more closely related to Xiphydriidae + Vespina. In the nonadditive analysis of Vilhelmsen (2001, pp. 427, 435), the trees with either taxon as the sister group to Xiphydriidae + Vespina are equally long, but a sister-group relationship between Siricidae and Anaxyelidae was not suggested; in the additive analysis, the tree with a sister-group relationship between these taxa is three steps longer than the shortest tree (Vilhelmsen, 2001, p. 435).

Although Siricoidea (= *Syntexis* + Siricidae) is monophyletic in all our simultaneous trees, the evidence supporting this is not particularly convincing; very few unique morphological and molecular synapomorphies could be found. It is supported mostly by homoplastic characters. Although there are several unique insertions

shared by the siricid species (Figs. 9A and B), none of them is present in *Syntexis*. Morphological synapomorphies for Siricoidea are the reduction of the metapleural apodemes back to the condition found in Xyelidae (also found in Pamphilioidea) (character 142) and the line of fusion of the cordate apodemes being situated in a depression, which is also the case in *Orussus* (character 194). There is another unambiguously optimized change supporting the branch, but this character (151) is quite homoplastic and coded as unknown for many taxa, including two of the three Siricidae. The morphological evidence placing Siricidae closer to *Xiphydria* + Vespina, on the other hand, consists of three homoplastic characters (see Vilhelmsen, 2001). Hence it must be concluded that there is only weak evidence for the placement of *Syntexis*. The scarcity of morphological evidence might in part be due to the fact that *Syntexis libocedrii* is highly derived, as reflected, for example, in the male genitalia (a drawing of which is given in Middlekauff, 1964). Another possible cause could be that the branch between *Syntexis* + sister taxon and the rest of the hymenopterans might be rather short. An even larger character sampling is required in order to find more, particularly nonhomoplastic, characters supporting the placement of *Syntexis*. However, even though the synapomorphies of Siricoidea may not be particularly convincing, the combined evidence included in the present analysis groups *Syntexis* with Siricidae (Fig. 6B) not only in the simultaneous analysis that maximizes congruence, but in all other simultaneous analyses as well.

The cladogram also shows Xiphydrioidea (= Xiphydriidae) as the sister group to a monophyletic Vespina. These relationships are stable across all nine simultaneous analyses (cf. Fig. 6B). Since the cephids and *Xiphydria* are always separated by Siricoidea (Fig. 6B) in the simultaneous trees, the insertions in the 16S gene present in these two taxa (Fig. 8) must be accounted for by a convergent origin. A monophyletic Vespina (= Orussidae + Apocrita) is present in all our simultaneous trees (Fig. 6B) and is supported by many unambiguously optimized morphological changes (Appendix G). Apocrita is also monophyletic (Fig. 7). Their monophyly is, however, not robust to a variation of the analysis parameters: they appear in only four of the nine simultaneous trees (Fig. 5), because *Orussus* and the stephanid *Schlettererius* come out as sister taxa in the remaining five. Orussidae and Stephanidae do share some morphological features, e.g., a "crown" of cuticular teeth around the ocelli, but these are outnumbered by the morphological apomorphies shared by Stephanidae and other apocritans (see Vilhelmsen, 2001). Both Orussidae and Stephanidae are highly derived taxa that parasitize wood-boring larvae. A better taxon sampling within Vespina, especially for Orussidae and Stephanidae, is definitely needed to provide more stabler support for one of these two alternative hypotheses.

Taxon sampling

It is desirable to have two representatives for each of the groups whose relationships are to be examined (Baverstock and Moritz, 1996, p. 26). This would mean that two species of each family should have been sampled in this study, preferably from different subfamilies. Unfortunately, there are practical limits to taxon sampling. In Anaxyelidae, there is only one recent species, *S. libocedrii*. Members of Blasticotomidae, Megalodontesidae, Xiphydriidae, and Orussidae are very rare and we were glad to obtain even one genus for each of these three families (thanks to the help of colleagues who are mentioned in the Acknowledgments). But it should be possible to include more diprionid genera in future analyses. We have already sequenced a second stephanid (*Neostephanus*), but refrained from including it in our analyses, because it was not present in the morphological matrix (Vilhelmsen, 2001).

Appendix A

List of taxa used in the present analysis

The first column names the family (for sawflies) or superfamily (Apocrita) or order (outgroups) to which the taxa belong. The second column lists the species that Vilhelmsen (2001) used to generate the morphological data matrix; for the complete species names refer to his publication. The last column names the least inclusive supraspecific taxon in which both the molecular and the morphological exemplars are contained and which was therefore used as the operational taxonomic unit in the analysis.

Taxon	Exemplar used for morphology	Exemplar used for DNA	OTU
Raphidioptera	*Raphidia xanthostigma*	*Agulla* sp. and Raphidioptera sp.	Raphidioptera
Mecoptera	*Panorpa communis*	*Panorpa* sp.	Panorpa
Amphiesmenoptera	*Micropterix calthella*	*Hydropsyche sparna* and Trichoptera sp.	Amphiesmeno.
Coleoptera	*Priacma serrata*	*Xyloryctes faunus* and Cantharidae sp.	Coleoptera
Xyelidae	*Xyela julii*	*Xyela* sp.	Xyela
	Macroxyela ferruginea	*Macroxyela ferruginea* (Say, 1824)	Macroxyela
Blasticotomidae	*Runaria reducta*	*Runaria reducta* (Malaise, 1931)	Runaria
Tenthredinidae	*Tenthredo arcuata* & sp.	*Tenthredo mesomela* (Linnaeus, 1758)	Tenthredo
	Dolerus niger & sp.	*Dolerus* sp.	Dolerus
	Athalia sp.	*Athalia* sp.	Athalia
	Nematus sp.	*Nematus* sp.	Nematus
Diprionidae	*Gilpinia* sp.	*Gilpinia* sp.	Gilpinia
Cimbicidae	*Cimbex* sp.	*Cimbex americana* (Leach, 1817)	Cimbex
	Zaraea fasciata & sp.	*Zaraea* sp.	Zaraea
	Corynis sp.	*Corynis crassicornis* (Rossi, 1790)	Corynis
Argidae	*Arge nigr., gracil., pull.*	*Arge cyanocrocea* (Forster, 1771)	Arge
	Sterictiphora furcata	*Sterictiphora furcata* (Villers, 1789)	Sterictiphora
Pergidae	*Perga condei*	*Perga condei* (Benson, 1939)	Perga
	Phylacteophaga froggatti	*Phylacteophaga froggatti* (Riek, 1955)	Phylacteophaga
Pamphiliidae	*Pamphilius sylvat.* & sp.	*Onycholyda amplecta* (Fabricius, 1804)	Pamphiliini
	Neurotoma nemoralis	*Neurotoma fasciata* (Norton, 1862)	Neurotoma
	Cephalcia arvensis	*Cephalcia* sp.	Cephalcia
	Acantholyda erythro. & sp.	*Acantholyda posticalis* (Matsumura, 1912)	Acantholyda
Megalodontesidae	*Megalodontes cephalotes*	*Megalodontes cephalotes* (Fabricius, 1781)	Megalodontes
Cephidae	*Cephus cultratus* & *nigrinus*	*Cephus pygmeus* (Linnaeus, 1767)	Cephus
	Calameuta pallipes & *filif.*	*Calameuta filiformis* (Eversmann, 1847)	Calameuta
	Hartigia linearis & *xantho.*	*Hartigia trimaculata* (Say, 1824)	Hartigia
Anaxyelidae	*Syntexis libocedrii*	*Syntexis libocedrii* (Rohwer, 1915)	Syntexis
Siricidae	*Sirex juvencus*	*Sirex noctilio* (Fabricius, 1793)	Sirex
	Urocerus gigas	*Urocerus gigas* (Linnaeus, 1758)	Urocerus
	Tremex columba	*Tremex columba* (Linnaeus, 1763)	Tremex
Xiphydriidae	*Xiphydria camelus*	*Xiphydria prolongata* (Geoffroy, 1785)	Xiphydria
Orussidae	*Orussus abiet.* & *occident.*	*Orussus abietinus* (Scopoli, 1763)	Orussus
Stephanoidea	*Schlettererius cinctipes*	*Schlettererius cinctipes* (Cresson, 1880)	Schlettererius
Megalyroidea	*Megalyra fasciipennis*	*Megalyra* sp.	Megalyra
Evanioidea	*Aulacus striatus*	*Evania appendigaster* (L. 1758) and *Gasteruption* sp.	Evanioidea
Cynipoidea	*Ibalia rufipes*	*Ibalia anceps* (Say, 1824)	Ibalia
Trigonalyoidea	*Orthogonalys pulchella*	*Labidogonalos* sp.	Trigonalidae
Vespoidea	*Vespula rufa*	*Vespula maculifrons* (Buysson, 1905)	
		Polistes fuscatus (Fabricius, 1793)	Vespidae

Appendix B

Primer sequences (5′ to 3′) used for generating the rDNA sequences

Positions of primers for mitochondrial genes (16S, CO1) are based on the sequence of *A. mellifera* (Crozier and Crozier, 1993). Positions of primers for 18S and 28S are based on the sequence of *D. melanogaster* (Tautz et al., 1988). Reverse primers are marked by an asterisk. The primer called "18S 9R" is actually situated in the ITS1 gene behind the 18S gene.

Primer name	Primer sequence	Position
16S A Hym	5′-TRA CTG TRC AAA GGT AGC-3′	13859–13842 (*Apis*)
16S B Hym*	5′-TTA ATT CAA CAT CGA GGT C-3′	13473–13491 (*Apis*)
18S 1F	5′-TAC CTG GTT GAT CCT GCC AGT AG-3′	1–23 (*Dros.*)
18S 3F	5′-GTT CGA TTC CGG AGA GGG A-3′	378–396 (*Dros.*)
18S 4F	5′-CCA GCA GCC GCG CTA ATT C-3′	573–591 (*Dros.*)
18S a2.0	5′-ATG GTT GCA AAG CTG AAA C-3′	1204–1222 (*Dros.*)
18S 5R*	5′-CTT GGC AAA TGC TTT CGC-3′	1040–1023 (*Dros.*)
18S bi*	5′-GAG TCT CGT TCG TTA TCG GA-3′	1421–1402 (*Dros.*)
18S 7R*	5′-GCA TCA CAG ACC TGT TAT TGC-3′	1631–1611 (*Dros.*)
18S 9R*	5′-GAT CCT TCC GCA GGT TCA CCT AC-3′	1991–1969 (*Dros.*)
28S A	5′-GAC CCG TCT TGA AAC ACG GA-3′	4046–4065 (*Dros.*)
28S B*	5′-TCG GAA GGA ACC AGC TAC TA-3′	4413–4394 (*Dros.*)
28S Bout*	5′-CCC ACA GCG CCA GTT CTG CTT ACC-3′	4625–4602 (*Dros.*)
CO1 lco hym	5′-CAA ATC ATA AAG ATA TTG G-3′	1816–1834 (*Apis*)
CO1 hco extA	5′-GAA GTT TAT ATT TTA ATT TTA CCT GG-3′	2511–2536 (*Apis*)
CO1 hco*	5′-TAA ACT TCA GGG TGA CCA AAA AAT CA-3′	2518–2493 (*Apis*)
CO1 hco out*	5′-CCA GGT AAA ATT AAA ATA TAA ACT TC-3′	2536–2511 (*Apis*)
CO1 hco outout*	5′-GTA AAT ATA TGR TGD GCT C-3′	2668–2650 (*Apis*)
CO1 hco extB*	5′-CCT ATT GAW ARA ACA TAR TGA AAA TG-3′	2938–2913 (*Apis*)

Appendix C

Origins of the sequences

Here, present study; Carp., Carpenter and Wheeler (1999); Dow., Dowton and Austin (1994, 1995); Liu, Liu (in preparation); Whit., Whiting et al. (1997). The fragments of CO1 sequenced by Dowton and Austin correspond to the second CO1 fragment in the present study (see Appendix D).

Exemplar	16S	18S	28S	CO1
Agulla sp.	—	Whit.	—	—
Raphidioptera sp.	Here	—	Here	Here
Panorpa sp.	Here	Here	Here	Here
Hydropsyche sparna	—	Whit.	Whit.	—
Trichoptera sp.	Here	—	—	Here
Xyloryctes faunus	—	Whit.	—	—
Cantharidae sp.	Here	—	Here	Here
Xyela sp.	Here	Here	Here	Here
Macroxyela ferruginea	Here	Here	Here	Here + Dow.
Runaria reducta	Here	Here	Here	Here
Tenthredo mesomela	Here	Here	Here	Here
Dolerus sp.	Here	Here	Here	Here
Athalia sp.	Here	Here	Here	Here
Nematus sp.	Here	Here	Here	Here
Gilpinia sp.	Here	Here	Here	Here

Exemplar	16S	18S	28S	CO1
Cimbex americana	Here	Here	Here	Here
Zaraea sp.	Here	Here	Here	Here
Corynis crassicornis	Here	Here	Here	Here
Arge cyanocrocea	Here	Here	Here	Here
Sterictiphora furcata.	Here	Here	Here	Here
Perga condei	Dow.	Here	Here	Here + Dow.
Phylacteophaga frog.	Dow.	Here	Here	Here + Dow.
Onycholyda amplecta	Here	Here	Here	Here
Neurotoma fasciata	Here	Here	Here	Here
Cephalcia sp.	Here	Here	Carp.	Here
Acantholyda posticalis	Here	Here	Here	Here
Megalodontes cephalotes	Here	Here	Here	Here
Cephus pygmeus	Here	Here	Here	Here
Calameuta filiformis	Here	Here	Here	Here
Hartigia trimaculata	Dow.	Here	Here	Here + Dow.
Sirex noctilio	Here	Here	Here	Here + Dow.
Urocerus gigas	Here	Here	Here	Here
Tremex columba	Here	Here	Here	Here
Syntexis libocedrii	Here	Here	Here	Here
Xiphydria prolongata	Here	Here	Here	Here
Orussus abietinus	Here	Here	Here	Here
Schlettererius cinctipes	Dow.	Here	Here	Here + Dow.
Megalyra sp.	Here	Here	Here	Here
Evania appendigaster	—	Whit.	—	—
Gasteruption sp.	Here	—	Carp.	Carp.
Ibalia anceps	Liu	Liu	Liu	Liu
Labidogonalos sp.	Here	Here	Carp.	Carp.
Vespula maculifrons	Here	—	Carp.	Carp.
Polistes fuscatus	—	Whit.	—	—

Appendix D

Fragments of genes used for the analysis

The first column gives the number and name of the fragment. The next three columns give the position of the beginning and end of the fragments based on the sequence of *A. mellifera* (Crozier and Crozier, 1993), the sequence of *D. melanogaster* (Tautz et al., 1988), and the alignment of Whiting et al. (1997). A "?" means that the homology to the sequence position of *Drosophila* could not be determined because the sequences of *Drosophila* and Hymenoptera are too different in these regions. The basepairs at the end of fragment 11 and the beginning of fragment 12 are missing in the alignment of Whiting et al. (1997).

Fragment	*Apis*	*Drosophila*	Whiting et al. (1997)	Number of unaligned basepairs
01: 16S-1	13850–13707			144–148
02: 16S-2	13706–13652			32–60
03: 16S-3	13653–13480			161–204
04: 18S-1		600–?	205-333	116–120
05: 18S-2		?–824	334-399	72–79
06: 18S-3		825–1165	400–758	340–343
07: 18S-4		1166–1364	759–959	195–199
08: 18S-5		1365–1416	960–1111	52
09: 18S-6		1417–1582	1112–?	67–107

Fragment	Apis	Drosophila	Whiting et al. (1997)	Number of unaligned basepairs
10: 28S-1		4067–?	14–140	100–105
11: 28S-2		?–?	141–miss.	12–150
12: 28S-3		?–?	miss.–	242–244
13: 28S-4		?–?	–	15–52
14: 28S-5		?–?	–	104
15: 28S-6		?–?	–	15–57
16: 28S-7		?–4607	–	34–36
17: CO1-1	1861–2629			763–769
18: CO1-2	2630–2904			275

Appendix E

Fragments that were included in the analysis

The numbers of the fragments correspond to those in Appendix D. A "+" means that the fragment is present for that species; "%" means that more than 25 bp of the fragment are missing for that species.

	16S			18S						28S							CO1	
	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18
Raphidioptera	+	+	+	+		+	+	+	+	+	+	+	+	+	+	+	+	+
Panorpa	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Amphiesmenoptera	+	+	+	+	+	+	+	+		+	+	%					+	
Coleoptera	+	+	+	+	+	+	+	+		+	+	+		+	+	+	+	
Xyela	+	+	+	%	+	+	+			+	+	+	+	+	+	+	+	
Macroxyela	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Runaria	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Tenthredo	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Dolerus	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Athalia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Nematus	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Gilpinia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Cimbex	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Zaraea	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Corynis	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Arge	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Sterictiphora	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Perga	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Phylacteophaga	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Pamphiliini	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Neurotoma	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	%	
Cephalcia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Acantholyda	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Megalodontes	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Cephus	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Calameuta	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Hartigia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Syntexis	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Sirex	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Urocerus	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Tremex	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	

	16S			18S						28S							CO1	
	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18
Xiphydria	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Orussus	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Schlettererius	+	+	+	+	+	+	+	+	+	+	+	%					+	+
Megalyra	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Evanioidea	+	+	+	+	+	+				+	+	%					+	+
Ibalia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Trigonalidae	+	+	+	+	+	+	+			+	+	%					%	+
Vespidae	+	+	+	+	+	+	+	+		+	+	+	+	+	+	+	+	+

Appendix F

Lengths of trees and incongruence values for the nine parameter sets

The first two columns give the indel-to-transversion ratio and the transversion-to-transition ratio. These ratios are summarized in column 3 as the indel-to-transversion-to-transition ratio. Morphological character transformations are always weighted equal to indel (insertion and deletion) events. Columns 4–7 show the lengths of the trees calculated from each gene separately. Column 8 gives the lengths of the trees calculated from all the molecular evidence together (see Fig. 3). Column 9 gives the length of the morphological tree multiplied by the respective weight. Column 10 gives the lengths of the simultaneous trees (see Fig. 5). Column 11 gives the length of all the data on a bush (as a sum of the values estimated by POY for the separate data sets), which is needed to calculate the RILD values. Column 12 gives the RILD values that compare the length of the simultaneous tree with those of the five separate trees, column 13 gives the RILD values that compare the length of the simultaneous tree with those of the molecular tree and the morphological tree, and column 14 gives the RILD values that compare the length of the molecular tree with those of the four single gene trees.

1	2	3	4	5	6	7	8	9	10	11	12	13	14
Id:Tv	Tv:Ti	Id:Tv:Ti	16S	18S	28S	CO1	Molec.	Morph.	Simult.	Bush	Total RILD	MolMor RILD	Molec.RILD
1:1	1:1	1:1:1	1729	578	637	4219	7371	835	8284	11639	0.0785	0.0227	*0.0927*
1:1	2:1	2:2:1	3060	839	1002	6926	12219	1670	14,037	20,137	0.0813	0.0237	0.1019
1:1	4:1	4:4:1	5673	1334	1712	12,187	21,617	3340	25,265	36,878	0.0807	0.0258	0.1009
2:1	1:1	2:1:1	1907	640	636	4279	7728	1670	9518	14,374	0.0736	0.0241	0.1086
2:1	2:1	4:2:1	3437	963	1047	7010	12,998	3340	16,582	25,837	0.0782	0.0257	0.1214
2:1	4:1	8:4:1	6395	1505	1756	12,346	23,115	6680	30,293	48,118	0.0829	0.0272	0.1346
4:1	1:1	4:1:1	2089	721	718	4309	8176	3340	11,682	19,446	*0.0611*	*0.0209*	0.1263
4:1	2:1	8:2:1	3735	1079	1114	7062	13,678	6680	20,762	35,934	0.0671	0.0259	0.1350
4:1	4:1	16:4:1	6969	1770	1868	12,455	24,525	13,360	38,722	68,586	0.0715	0.0273	0.1489

Appendix G

Unambiguously optimized morphological changes on the cladogram shown in Fig. 7

The numbers of the characters correspond to those of Vilhelmsen (2001); for a description of the characters, refer to his publication.

Hymenoptera: 3: 0 → 1; 35: 1 → 0; 46: 0 → 1; 48: 0 → 1; 54: 01; 62: 1 → 2; 70: 1 → 0; 72: 0 → 1; 82: 0 → 1; 86: 0 → 1; 98: 0 → 1; 102: 2 → 0; 104: 0 → 2; 105: 0 → 1; 111: 0 → 1; 116: 0 → 1; 132: 0 → 1; 147: 2 → 0; 148: 0 → 1; 153: 0 → 1; 155: 0 → 1; 156: 1 → 0; 171: 0 → 1; 173: 1 → 0; 176: 1 → 0; 184: 1 → 0; 217: 0 → 1; 220: 0 → 1; 222: 0 → 1; 224: 4 → 3;

Xyelidae: 8: 0 → 1; 16: 1 → 0; 23: 01; 26: 0 → 1; 29: 0 → 1; 90: 3 → 0; 96: 1 → 0; 142: 1 → 0; 145: 1 → 0; 172: 1 → 0; 219: 0 → 1; 224: 3 → 1;

Hymenoptera – Xyelidae: 46: 1 → 2; 48: 1 → 2; 92: 0 → 1; 141: 0 → 1; 151: 0 → 1; 160: 0 → 1; 164: 0 → 1; 178: 0 → 1; 199: 0 → 1;

Tenthredinoidea s.l. + Pamphilioidea: 42: 0 → 1; 91: 0 → 2; 97: 0 → 1; 108: 0 → 1; 131: 0 → 1;

Tenthredinoidea s.l.: 16: 1 → 0; 39: 0 → 1; 41: 0 → 1; 54: 1 → 0; 57: 1 → 0; 59: 1 → 2; 65: 0 → 1; 94: 01; 95: 0→1; 109: 0 → 1; 113: 0 → 1; 115: 0 → 1; 119: 0 → 1; 167: 0 → 1;

Tenthredinoidea s.str.: 50: 0 → 1; 65: 1 → 2; 69: 0 → 1; 76: 1 → 0; 120: 0 → 1; 150: 0 → 1; 156: 0 → 1; 161: 0 → 1; 197: 0 → 1; 201: 0 → 1; 216: 0 → 1;

Tenthredinidae sensu Rasnitsyn: 90: 3 → 1; 91: 2→ 0; 131: 1 → 0;

Cimbicidae + Argidae + Pergidae: 103: 0 → 1; 117: 0→ 1; 118: 0 → 1; 124: 0 → 1; 129: 01; 137: 0 → 1; 139: 0 → 3; 143: 0 → 1; 152: 0 → 1; 176: 0 → 1; 202: 0 → 1; 224: 3 → 5;

Argidae + Pergidae: 13: 1 → 0; 47: 0 → 1; 56: 0 → 1; 70: 0 → 1; 166: 0 → 1; 204: 0 → 1; 210: 0 → 1;

Pergidae: 35: 0 → 1; 52: 1 → 0; 57: 0 → 1; 58: 0 → 1; 60: 1 → 0; 82: 1 → 0; 85: 0 → 1; 132: 1 → 2; 150: 1 → 0; 171: 1 → 3; 174: 0 → 1; 181: 01;

Argidae: 23: 0 → 2; 133: 0 → 1;

Cimbicidae: 66: 0 → 1; 119: 1 → 0; 126: 0 → 1; 127: 0 → 1; 232: 0 → 1;

Runaria (Blasticotomidae): 9: 0 → 1; 12: 0 → 1; 13: 1 → 0; 23: 0 → 2; 43: 01; 48: 2 → 1; 94: 1 → 2; 135: 1 → 0; 139: 0 → 2; 153: 1 → 0; 160: 1 → 0; 189: 0 → 1;

Pamphilioidea: 14: 0 → 1; 26: 0 → 2; 27: 1 → 2; 45: 0 → 1; 47: 0 → 1; 196: 0 → 1; 198: 0 → 1; 207: 0 → 1; 210: 0 → 1; 219: 0 → 1; 223: 0 → 1; 224: 30; 233: 0 → 1; 235: 1 → 0;

Pamphiliidae: 38: 0 → 1; 142: 1 → 0; 145: 1 → 0; 159: 1 → 0; 172: 1 → 0; 178: 1 → 0; 187: 0 → 1;

Pamphiliini + Cephalciinae: 160: 1 → 0;

Megalodontes (Megalodontesidae): 5: 0 → 1; 21: 0→ 1; 22: 0 → 2; 50: 0 → 1; 69: 01; 76: 1 → 2; 90: 3 → 1; 96: 1→ 0; 117: 0 → 1; 124: 0 → 1; 129: 0 → 1; 152: 0 → 1; 174: 0 → 1; 176: 0 → 1;

Unicalcarida: 6: 0 → 1; 12: 0 → 1; 71: 0 → 1; 88: 0 → 1; 89: 0 → 1; 99: 0 → 1; 124: 0 → 1; 128: 0 → 1; 174: 0 → 1; 212: 0 → 1; 234: 0 → 1;

Cephidae: 17: 0 → 1; 21: 0 → 1; 45: 0 → 1; 74: 0 → 1; 96: 1 → 0; 101: 0 → 1; 113: 0 → 1; 114: 0 → 1; 116: 1 → 0; 125: 0 → 1; 132: 1 → 2; 139: 01; 165: 0 → 1; 185: 0 → 1; 218: 0 → 1;

Siricoidea + Xiphydria + Vespina: 5: 0 → 1; 35: 0→ 1; 47: 0 → 1; 99: 1 → 2; 193: 12; 200: 0 → 1; 221: 0 → 1; 224: 3 → 4; 226: 0 → 1; 227: 0 → 1;

Siricoidea: 142: 1 → 0; 151: 1 → 0; 194: 0 → 1;

Siricidae: 18: 0 → 1; 34: 1 → 3; 38: 0 → 1; 40: 0 → 1; 52: 0 → 1; 59: 1 → 2; 77: 0 → 1; 86: 1 → 2; 102: 0 → 1; 126: 1 → 0; 136: 0 → 1; 147: 0 → 2; 159: 1 → 0; 162: 0 → 1; 174: 1 → 0;

Syntexis (Anaxyelidae): 17: 0 → 1; 25: 0 → 1; 33: 0 → 1; 44: 0 → 1; 68: 10; 87: 0 → 1; 88: 1 → 0; 96: 1 → 0; 113: 0 → 1; 124: 1 → 0; 125: 0 → 1; 131: 0 → 1; 165: 0 → 1; 180: 0 → 1; 181: 0 → 1;

Xiphydria + Vespina: 15: 0 → 1; 41: 0 → 1; 42: 0 → 1; 64: 0 → 1; 78: 0 → 1; 87: 0 → 2; 89: 1 → 2; 94: 0 → 1; 100: 0 → 1; 108: 0 → 1; 123: 0→ 1; 143: 0 → 1; 148: 1 → 2; 157: 0 → 1;

Vespina: 18: 0 → 1; 20: 0 → 2; 21: 0→ 1; 25: 0 → 1; 33: 0 → 1; 34: 1 → 2; 36: 2 → 0; 37: 01; 41: 1→ 2; 45: 0 → 1; 53: 0 → 1; 74: 0 → 1; 78: 1 → 2; 92: 1 → 2; 94: 1 → 2; 106: 0 → 1; 109: 0 → 1; 110: 0 → 1; 120: 0 → 1; 128: 1 → 0; 129: 0 → 1; 132: 1 → 2; 134: 0 → 1; 144: 0 → 1; 147: 0 → 1; 165: 0 → 2; 169: 0 → 1; 170: 0 → 1; 171: 1 → 2; 177: 0 → 1; 179: 0 → 1; 180: 0 → 1; 181: 0 → 1; 190: 0 → 1; 191: 0 → 1; 210: 0 → 1; 218: 0 → 1; 224: 4 → 6; 228: 2 → 3; 234: 1 → 0;

Apocrita: 30: 0 → 1; 44: 0 → 2; 60: 0 → 1; 65: 0 → 2; 76: 2 → 3; 91: 0 → 2; 103: 0 → 1; 116: 1 → 0; 130: 0 → 1; 134: 1 → 2; 139: 0 → 3; 171: 2 → 3; 183: 0 → 1; 184: 0 → 1; 185: 0 → 2;

References

Abe, M., Smith, D.R., 1991. The genus-group names of Symphyta (Hymenoptera) and their type species. Esakia 31, 1–115.

Basibuyuk, H.H., Quicke, D.L.J., 1995. Morphology of the antenna cleaner in the Hymenoptera with particular reference to non-aculeate families (Insecta). Zool. Scr. 24, 157–177.

Basibuyuk, H.H., Quicke, D.L.J., 1997. Hamuli in the Hymenoptera (Insecta) and their phylogenetic implications. J. Nat. Hist. 31, 1563–1585.

Basibuyuk, H.H., Quicke, D.L.J., 1999. Gross morphology of multiporous plate sensilla in the Hymenoptera (Insecta). Zool. Scr. 28, 51–67.

Baverstock, P.R., Moritz, C., 1996. Project design. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), Molecular Systematics, second ed. Sinauer, Sunderland, MA, pp. 17–27.

Börner, C., 1919. Stammesgeschichte der Hautflügler. Biol. Zentralblatt 39, 145–186.

Brothers, D.J., 1975. Phylogeny and classification of the aculeate Hymenoptera, with special reference to Mutillidae. Univ. Kansas Sci. Bull. 50, 483–648.

Brower, A.V.Z., DeSalle, R., Vogler, A., 1996. Gene trees, species trees, and systematics: a cladistic perspective. Annu. Rev. Ecol. Syst. 27, 423–450.

Bull, J.J., Huelsenbeck, J.P., Cunningham, C.W., Swofford, D.L., Waddell, P.J., 1993. Partitioning and combining data in phylogenetic analysis. Syst. Biol. 42, 384–397.

Carpenter, J.M., Wheeler, W.C., 1999. Towards simultaneous analysis of morphological and molecular data in Hymenoptera. Zool. Scr. 28, 251–260.

Chippindale, P.T., Wiens, J.J., 1994. Weighting, partitioning, and combining characters in phylogenetic analysis. Syst. Biol. 43, 278–287.

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

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

de Queiroz, A., Donoghue, M.J., Kim, J., 1995. Separate versus combined analysis of phylogenetic evidence. Annu. Rev. Ecol. Syst. 26, 657–681.

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

Derr, J.N., Davis, S.K., Woolley, J.B., Wharton, R.A., 1992b. Reassessment of the 16S rRNA nucleotide sequence from members of the parasitic Hymenoptera. Mol. Phyl. Evol. 1, 338–341.

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

Dowton, M., Austin, A.D., 1995. Increased genetic diversity in mitochondrial genes is correlated with the evolution of parasitism in the Hymenoptera. J. Mol. Evol. 41, 958–965.

Dowton, M., Austin, A.D., 1997. 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.

Farris, J.S., 1983. The logical basis of phylogenetic systematics. In: Platnick, N.I., Funk, V.A. (Eds.), Advances in Cladistics. Columbia University of Press, New York, pp. 7–36.

Farris, J.S., Källersjö, M., Kluge, A.G., Bult, C., 1995. Testing significance of incongruence. Cladistics 10, 315–319.

Gatesy, J., O'Grady, P., Baker, R.H., 1999. Corroboration among data sets in simultaneous analysis: hidden support for phylogenetic

relationships among higher level artiodactyl taxa. Cladistics 15, 271–313.

Gauld, I.D., Mound, L.A., 1982. Homoplasy and the delineation of holophyletic genera in some insect groups. Syst. Entomol. 7, 73–86.

Gibson, G.A.P., 1985. Some pro- and mesothoracic structures important for phylogenetic analysis of Hymenoptera, with a review of terms used for the structures. Can. Entomol. 117, 1395–1443.

Gibson, G.A.P., 1986. Evidence for monophyly and relationships of Chalcidoidea, Mymaridae, and Mymarommatidae. Can. Entomol. 118, 205–240.

Gibson, G.A.P., 1993. Groundplan structure and homology of the pleuron in Hymenoptera based on a comparison of the skeletomusculature of Xyelidae (Hymenoptera) and Raphidiidae (Neuroptera). Mem. Ent. Soc. Can. 165, 165–187.

Gibson, G.A.P., 1999. Sister-group relationships of the Platygastroidea and Chalcidoidea (Hymenoptera)—an alternate hypothesis to Rasnitsyn (1988). Zool. Scr. 28, 125–138.

Giribet, G., Ribera, C., 2000. A review of arthropod phylogeny: new data based on ribosomal DNA sequences and direct character optimization. Cladistics 16, 204–231.

Gladstein, D.S., Wheeler, W.C., 1997. POY: The Optimization of Alignment Characters. Program and documentation. New York, NY. Available at: ftp.amnh.org/pub/molecular. Documentation by D. Janies and W. Wheeler. ("POY: Phylogeny Reconstruction via Direct Optimization of DNA Data" native code version 2.0 by David Gladstein and Ward Wheeler.).

Goulet, H., 1993. Superfamilies Cephoidea, Megalodontoidea, Orussoidea, Siricoidea, Tenthredinoidea, and Xyeloidea. in: Goulet, H., Huber, J.T. (Eds.), Hymenoptera of the World: An Identification Guide to Families. Centre for Land and Biological Resources Research, Research Branch Agriculture Canada, Publication 1894/E, Ottawa, Ontario, pp. 101–115 (Chapter 6).

Hennig, W., 1950. Grundzüge einer Theorie der phylogenetischen Systematik. Deutscher Zentralverlag, Berlin.

Hennig, W., 1966. Phylogenetic Systematics. University of Illinois Press, Urbana.

Heraty, J.M., Woolley, J.B., Darling, D.C., 1994. Phylogenetic implications of the mesofurca and mesopostnotum in Hymenoptera. J. Hym. Res. 3, 241–277.

Johnson, N.F., 1988. Midcoxal articulations and the phylogeny of the order Hymenoptera. Ann. Ent. Soc. Am. 81, 870–881.

Kenneth, J.H., 1960. A Dictionary of Scientific Terms (founded by Henderson and Henderson), seventh ed. Oliver and Boyd, Edinburgh.

Kluge, A.G., 1989. A concern for evidence and a phylogenetic hypothesis for relationships among *Epicrates*. Syst. Zool. 38, 7–25.

Kluge, A.G., Wolf, A.J., 1993. Cladistics: what's in a word? Cladistics 9, 183–199.

Königsmann, E., 1976. Das phylogenetische System der Hymenoptera. Teil 1: Einführung, Grundplanmerkmale, Schwestergruppe und Fossilfunde. Dtsch. Ent. Z. N.F. 23, 253–279.

Königsmann, E., 1977. Das phylogenetische System der Hymenoptera Teil 2: "Symphyta". Dtsch. Ent. Z. N.F. 24, 1–40.

Königsmann, E., 1978a. Das phylogenetische System der Hymenoptera Teil 3: "Terebrantes" (Unterordnung Apocrita). Dtsch. Ent. Z. N.F. 25, 1–55.

Königsmann, E., 1978b. Das phylogenetische System der Hymenoptera. Teil 4: Aculeata (Unterordnung Apocrita). Dtsch. Ent. Z. N.F. 25, 365–435.

Kristensen, N.P., 1991. Phylogeny of extant hexapods. In: Naumann, K. et al. (Eds.), The Insects of Australia, second ed. Melbourne University Press, Melbourne, pp. 125–140.

Larson, A., 1998. The comparison of morphological and molecular data in phylogenetic analysis. In: DeSalle, R., Schierwater, B. (Eds.), Molecular Approaches to Ecology and Evolution, second ed. Birkhäuser Verlag, Basel, Switzerland, pp. 275–296.

Liu, Z. in preparation. Higher-level phylogeny of Cynipoidea (Hyme-noptera, Insecta) based on multiple DNA sequences (provisional title).

Mickevich, M.F., Farris, J.S., 1981. The implications of congruence in *Menidia*. Syst. Zool. 30, 351–370.

Mickevich, M.F., Johnson, M.S., 1976. Congruence between morpho-logical and allozyme data in evolutionary inference and character evolution. Syst. Zool. 25, 260–270.

Mickoleit, G., 1966. Zur Kenntnis einer neuen Spezialhomologie (Synapomorphie) der Panorpoidea. Zool. Jahrbücher Abteilung Anat. Ont. Tiere 83, 483–496.

Mickoleit, G., 1969. Vergleichend-anatomische Untersuchungen an der pterothorakalen Pleurotergalmuskulatur der Neuropteria und Me-copteria (Insecta, Holometabola). Z. Morphol. Tiere 64, 151–178.

Middlekauff, W.W., 1964. Notes and description of the previously unknown male of *Syntexis libocedrii*. Pan-Pacific Entomol. 40, 255–258.

Miyamoto, M.M., Fitch, W.M., 1995. Testing species phylogenies and phylogenetic methods with congruence. Syst. Biol. 44, 64–76.

Nixon, K.C., 2001. Winclada, version 0.99.9. Bailey Hortorium, Ithaca, NY.

Nixon, K.C., Carpenter, J.M., 1996. On simultaneous analysis. Cladistics 12, 221–241.

Oeser, R., 1961. Vergleichend-morphologische Untersuchungen über den Ovipositor der Hymenopteren. Mitt. Zool. Mus. Berlin 37, 3–119.

Paulus, H.F., 1979. Eye structure and the monophyly of Arthropoda. In: Gupta, A.P. (Ed.), Arthropod Phylogeny. Van Nostrand Reinhold, New York, pp. 299–383.

Quicke, D.L.J., Fitton, M.G., Tunstead, J.R., Ingram, S.N., Gaitens, P.V., 1994. Ovipositor structure and relationships within the Hymenoptera, with special reference to the Ichneumonoidea. J. Nat. Hist. 28, 635–682.

Rasnitsyn, A.P., 1969. Origin and evolution of the lower Hymenop-tera. Trudy paleontologicheskii Instituta (Akademiya Nauk SSSR) 123, 1–196 (In Russian; translation by Amerind, New Delhi, 1979).

Rasnitsyn, A.P., 1988. An outline of evolution of the hymenopterous insects. Orient. Insects 22, 115–145.

Rohwer, S.A., Cushman, R.A., 1917. Idiogastra, a new suborder of Hymenoptera with notes on the immature stages of Oryssus. Proc. Ent. Soc. Wash. 19, 89–99.

Ronquist, F., 1999. Phylogeny of the Hymenoptera (Insecta): The state of the art. Zool. Scr. 28, 3–11.

Ronquist, F., Rasnitsyn, A.P., Roy, A., Eriksson, K., Lindgren, M., 1999. Phylogeny of the Hymenoptera: a cladistic reanalysis of Rasnitsyn's (1988) data. Zool. Scr. 28, 13–50.

Ross, H.H., 1937. A generic classification of the nearctic sawflies. Illinois Biol. Monogr. 15 (2), 1–173.

Sankoff, D., Cedergren, R.J., 1983. Simultaneous comparison of three or more sequences related by a tree. In: Sankoff, D., Kruskal, (Eds.), Time Warps, String Edits, and Macromolecules: The Theory and Practise of Sequence Comparison. Addison-Wesley, Reading, MA, pp. 253–263.

Sankoff, D., Rousseau, P., 1975. Locating the vertices of a Steiner tree in an arbitrary metric space. Math. Prog. 9, 240–246.

Shinohara, A., 1983. Discovery of the families Xyelidae, Pamphiliidae, Blasticotomidae, and Orussidae from Taiwan, with descriptions of four new species (Hymenoptera: Symphyta). Proc. Entomol. Soc. Wash. 85, 309–320.

Shipley, A.E., 1904. The orders of insects. Zool. Anz. 27, 259–262.

Tautz, D., Hancock, J.M., Webb, D.A., Tautz, C., Dover, G.A., 1988. Complete sequences of the rDNA genes of *Drosophila melanogas-ter*. Mol. Biol. Evol. 5, 366–376.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24, 4876–4882.

Vilhelmsen, L., 1996. The preoral cavity of lower Hymenoptera (Insecta): comparative morphology and phylogenetic significance. Zool. Scr. 25, 143–170.

Vilhelmsen, L., 1997a. The phylogeny of lower Hymenoptera, with a summary of the early evolutionary history of the order. J. Zool. Syst. Evol. Res. 35, 49–70.

Vilhelmsen, L., 1997b. Head capsule concavities accomodating the antennal bases in Hymenoptera pupating in wood: possible emergence-facilitating adaptations. Int. J. Insect Morphol. Em-bryol. 26, 129–138.

Vilhelmsen, L., 1999a. The comparative anatomy of the basal hyme-nopteran lineages, and its phylogenetic implications. Unpublished Ph.D. thesis. Zoological Museum, University of Copenhagen.

Vilhelmsen, L., 1999b. The occipital region in the basal Hymenoptera (Insecta): a reappraisal. Zool. Scr. 28, 75–85.

Vilhelmsen, L., 2000a. Before the wasp–waist: comparative anatomy and phylogenetic implications of the skeleto-musculature of the thoraco-abdominal boundary region in basal Hymenoptera (In-secta). Zoomorphology 119, 185–221.

Vilhelmsen, L., 2000b. Cervical and prothoracic skeleto-musculature in the basal Hymenoptera (Insecta): comparative anatomy and phylogenetic implications. Zool. Anz. 239, 105–138.

Vilhelmsen, L., 2000c. The ovipositor apparatus of basal Hymenoptera (Insecta): phylogenetic implications and functional morphology. Zool. Scr. 29, 319–345.

Vilhelmsen, L., 2001. Phylogeny and classification of the extant basal lineages of the Hymenoptera. Zool. J. Linnean Soc. 131, 393–442.

von Kéler, S., 1963. Entomologisches Wörterbuch (mit besonderer Berücksichtigung der morphologischen Terminologie, third ed. Akademie Verlag, Berlin.

Wei, M., Nie, H., 1997. Preliminary phylogeny of Symphyta (Hyme-noptera). J. Central South Forestry Univ. 17, 36–41 (In Chinese, with English abstract).

Wheeler, W.C., 1990. Nucleic acid sequence phylogeny and random outgroups. Cladistics 6, 363–367.

Wheeler, W.C., 1995. Sequence alignment, parameter sensitivity and the phylogenetic analysis of molecular data. Syst. Biol. 44, 321–331.

Wheeler, W.C., 1996. Optimization alignment: the end of multiple sequence alignment in phylogenetics? Cladistics 12, 1–9.

Wheeler, W.C., 2000. Heuristic reconstruction of hypothetical-ances-tral DNA sequences: Sequence alignment versus direct optimiza-tion. In: Scotland, R., Pennington, R.T. (Eds.), Homology and Systematics: coding characters for phylogenetic analysis. System-atics Association Special Volume 58. Taylor & Francis, London, pp. 106–113.

Wheeler, W.C., Gladstein, D.S. , 1991–1998. MALIGN. Program and documentation. New York, NY (Documentation by Daniel Janies and Ward Wheeler.).

Wheeler, W.C., Gladstein, D.S., 1994. MALIGN: a multiple sequence alignment program. J. Hered. 85, 417–418.

Wheeler, W.C., Hayashi, C.Y., 1998. The phylogeny of the extant chelicerate orders. Cladistics 14, 173–192.

Whitfield, J.B., Johnson, N.F., Hamerski, M.R., 1989. Identity and phylogenetic significance of the metapostnotum in nonaculeate Hymenoptera. Ann. Ent. Soc. Am. 82, 663–673.

Whiting, M.F., Carpenter, J.M., Wheeler, Q.D., Wheeler, W.C., 1997. The Strepsiptera problem: phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. Syst. Biol. 46, 1–68.

Yuasa, H., 1922. A classification of the larvae of the Tenthredinoidea. Illinois Biol. Monogr. 7 (4), 1–172.