Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements

Jeffrey L. Boore*, Timothy M. Collins, David Stanton*, L. Lynne Daehler & Wesley M. Brown

Department of Biology, University of Michigan, 830 N. University Avenue, Ann Arbor, Michigan 48109-1048, USA

THE origins of arthropods and the phylogenetic relationships among their three major living groups (atelocerates, crustaceans and chelicerates) are vigorously contended. To help resolve this, we determined mitochondrial gene arrangements for a chelicerate, a myriapod, two crustaceans, an onychophoran, a mollusc and an annelid, and compared them with published gene orders of other species. The result strongly supports the monophyly of Arthropoda and of Mandibulata (atelocerates plus crustaceans) and refutes the Uniramia (atelocerates plus onychophorans). Gene arrangement comparisons are emerging as a powerful new tool for resolving ancient phylogenetic relationships.

Three groups of living arthropods can be recognized with confidence: Atelocerata (insects, myriapods), Crustacea (shrimp, lobsters, barnacles, crabs) and Chelicerata (horseshoe crabs, arachnids). Debates rage, sometimes acrimoniously, on whether these taxa constitute a monophyletic group and on how they are related. Cladistic analyses of morphological characters have generally supported arthropod monophyly[1], but some studies of functional morphology conclude that 'arthropodization' occurred independently in various lineages as each evolved a protective chitinous exoskeleton[2]. Some functional morphologists place onychophorans with atelocerates to form the Uniramia[2], although evidence from fossil insects suggests that this is not a monophyletic group[3]. The most closely related arthropod subgroups may be the atelocerates and crustaceans (the Mandibulata); both have mandibles on the fourth head segment and share features of the eye, brain and appendages[1,4]. However, studies of other aspects of morphology[5,6] or of fossil evidence and functional considerations[7,8] unite crustaceans and chelicerates. Comparisons of arthropod ribosomal RNA sequences have yielded conflicting results: some claim arthropod polyphyly[9]; some recognize a monophyletic Arthropoda and Mandibulata, but exclude myriapods from a clade containing the other arthropods[10,11]; and one unites chelicerates with atelocerates[12]. Reasons for conflict include alignment ambiguities, artefactual associations of rapidly evolving taxa, confounding influences of base compositional or substitutional biases, and multiple substitutions at many nucleotide positions[13-18].

We are examining a set of molecular characters, the arrangement of genes in mitochondrial DNA, that promises to be especially useful for resolving ancient relationships. Metazoan mtDNAs typically encode 36 or 37 genes: 2 for rRNAs, 22 for transfer RNAs, and 12 or 13 for electron transport proteins[19]. In some mtDNAs, all genes are transcribed from the same strand; in others, both strands encode genes. The large number of possible gene arrangements makes it improbable that the same order would arise independently. Thus shared-derived arrangements are likely to indicate common ancestry[13,20-22].

Complete mitochondrial gene arrangements have been published for 14 invertebrates (3 echinoderms, 4 insects, 1 crustacean, 3 nematodes, 2 molluscs and 1 cnidarian)[13,19,20,23-25]. The multiplicity of arrangements found among these groups suggests that particular arrangements are probably not selectively constrained, but rather that rearrangements that do not lethally disrupt mitochondrial genome functioning are rare.

We have determined the relative arrangements of numerous mitochondrial genes for a chelicerate (*Limulus*), two crustaceans (*Homarus* and *Daphnia*), a myriapod (*Thyrophygus*), an onychophoran (*Euperipatoides*), a mollusc (*Plicopurpura*) and an annelid (*Lumbricus*). We compared these and other published mitochondrial gene arrangements cladistically, joining taxa only on the basis of shared-derived characters. As can be seen in Fig. 1, the mtDNAs of *Lumbricus*, the mollusc *Katharina* and the nematodes *Ascaris* and *Caenorhabditis* have in common the direct abutment of two gene pairs; $tRNA^{C}/tRNA^{M}$ and $tRNA^{S(AGN)}/ND2$ (*ND* genes code for NADH dehydrogenase subunits). *Katharina* mtDNA also shares the arrangement $tRNA^{W}/tRNA^{Y}$ with another nematode, *Meloidogyne*. In contrast, several arthropods, representing Chelicerata, Crustacea and Atelocerata, share alternative arrangements of these genes: $tRNA^{C}$ is between $tRNA^{W}$ and $tRNA^{Y}$, $tRNA^{Y}$ is inverted relative to $tRNA^{W}$, $tRNA^{M}$ is between $tRNA^{Q}$ and *ND2*, and $tRNA^{S(AGN)}$ is between $tRNA^{N}$ and $tRNA^{E}$ (Fig. 1).

If the arthropod clade shown in Fig. 1 included *Katharina* and/or *Lumbricus*, either the *Katharina* and *Lumbricus* arrangements reverted to those of nematodes, or identical rearrangements occurred independently in two or more arthropod lineages. Neither is likely. The most parsimonious explanation is that the arrangement shared among *Katharina*, *Lumbricus* and nematodes is unchanged from an ancestral state, and the alternative arrangement was derived from it early in the common arthropod lineage. These shared-derived gene arrangements support arthropod monophyly. *Artemia* has apparently acquired the translocation of $tRNA^{I}/tRNA^{Q}$ independently; this arrangement is not shared by any organism yet studied.

Similarly, a tRNA gene rearrangement reveals arthropod subgroup relationships. Metazoan mtDNAs contain two tRNA genes specifying leucine, designated $tRNA^{L(CUN)}$ and $tRNA^{L(UUR)}$, according to the codons these tRNAs recognize. In the mtDNAs of *Limulus* and *Euperipatoides* these genes abut directly in the arrangement $l\text{-}rRNA/tRNA^{L(CUN)}/tRNA^{L(UUR)}/ND1$. This is identical to their arrangement in *Katharina* and *Plicopurpura*, and similar to that in *Mytilus* (in which *l-rRNA* is at a different position) and *Lumbricus* (in which there have been independent translocations of $tRNA^{A}$ and $tRNA^{S(UCN)}$). In contrast, the mtDNAs of the crustaceans *Artemia*, *Daphnia* and *Homarus* share with *Drosophila* the arrangements $COI/tRNA^{L(UUR)}/COII$ (*COI* and *COII* code for cytochrome oxidase subunits I and II) and $l\text{-}rRNA/tRNA^{L(CUN)}/ND1$, and also share the latter arrangement with *Thyrophygus* (Fig. 1).

It is unlikely that $tRNA^{L(UUR)}$ has translocated to a position between *COI* and *COII* independently in the lineages leading to Atelocerata and Crustacea. The more parsimonious explanation is that *Euperipatoides* and *Limulus* retain the ancestral arthropod arrangement, and that the translocation of $tRNA^{L(UUR)}$ to the position between *COI* and *COII* occurred in the lineage common to Atelocerata and Crustacea after its separation from that leading to chelicerates and/or onychophorans. If *Thyrophygus* is correctly inferred as an atelocerate, $tRNA^{L(UUR)}$ must have translocated from the *COI/COII* junction to another position in the mtDNA after myriapods diverged from insects. However, our data do not eliminate the possibility that Atelocerata is not monophyletic, with myriapods diverging earlier than shown in Fig. 1, and $tRNA^{L(UUR)}$ having translocated from the *l-rRNA/ND1* region convergently with the other mandibulates.

Available data do not allow us to place *Euperipatoides* unambiguously on this phylogenetic tree, but are clearly inconsistent with the Uniramia hypothesis[2]. To place *Euperipatoides* in a clade with the atelocerates would require either independent, identical rearrangements in atelocerates and crustaceans or a

* Present addresses: Department of Cell Biology and Neuroanatomy, University of Minnesota, 4-135 Jackson Hall, 321 Church St. SE, Minneapolis, Minnesota 55455, USA (J.L.B.); and Department of Biology, Eastern Michigan University, Ypsilanti, Michigan 48197, USA (D.S.).

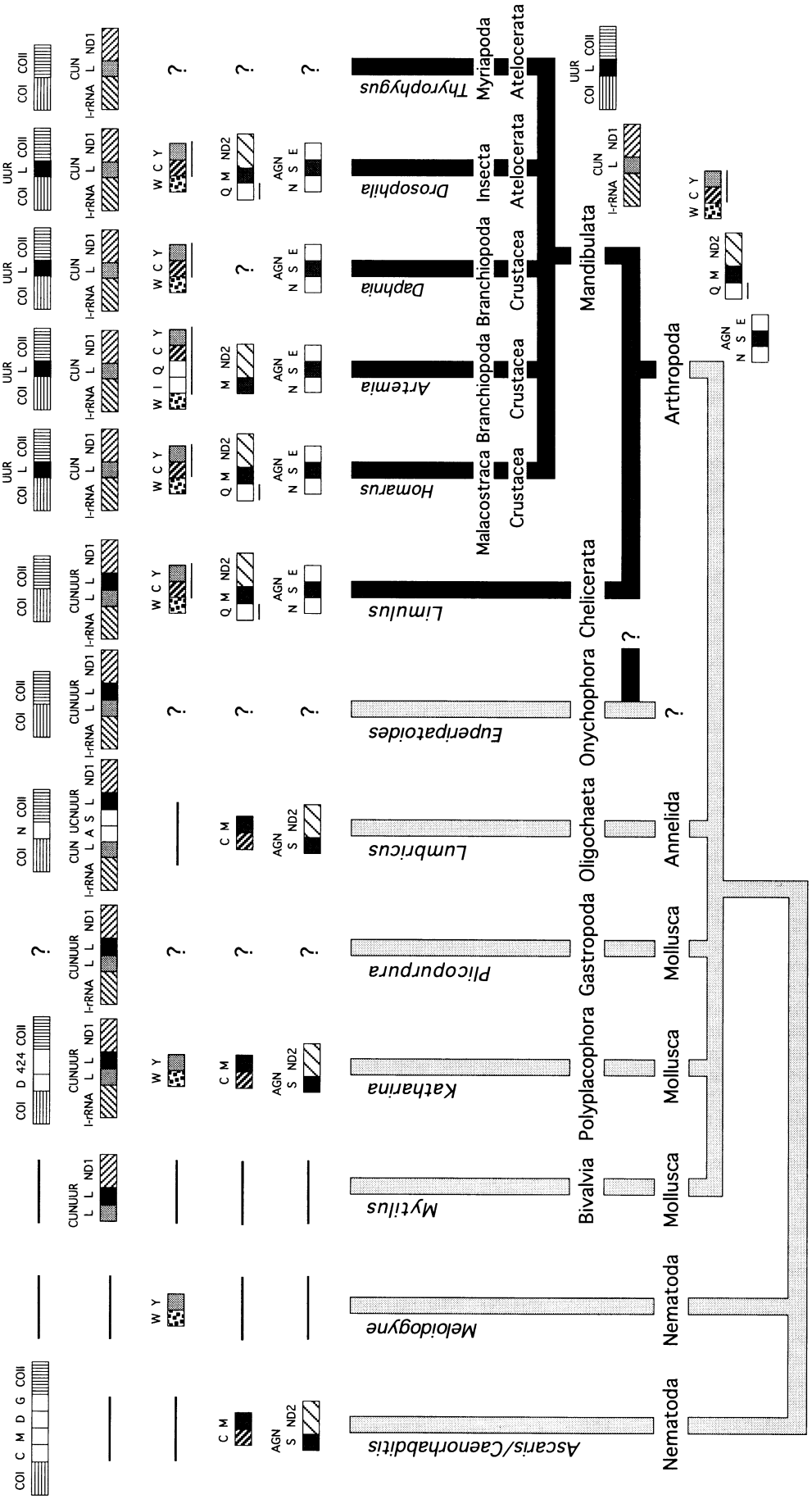

FIG. 1 Mitochondrial gene arrangements informative for arthropod phylogeny. No gene boundaries are shared in a pattern supporting any alternative tree. Shared-derived gene arrangements for Arthropoda and for Mandibulata are shown at the lower right. A matrix was constructed, scoring 74 characters as 'upstream of' and 'downstream of' (according to transcriptional orientation) each of the 37 genes. For each character, there are 74 possible states, each being the 5′ or 3′ end of an adjacent gene. Shared gene boundaries become identical character states; rearrangements generate differing states. Based on analyses of morphological characters[1], nematodes were designated the outgroup taxon, that is, they were chosen as being excluded from a clade that contains the remaining taxa. This matrix was analysed by eye, using features of the computer program MacClade[30], for all patterns that indicate shared-derived character states on which to unite taxa. The relationship among Mollusca, Annelida and Arthropoda is considered here as an unresolved polytomy, as is that among the mandibulates. With available data, the onychophoran cannot be placed any more exactly than its exclusion from the Mandibulata. Complete mitochondrial genome arrangements for *Ascaris suum*, *Caenorhabditis elegans*, *Meloidogyne javonica*, *Mytilus edulis*, *Katharina tunicata*, *Artemia franciscana* and *Drosophila yakuba* have been published[13,19,20,25]. MtDNA preparation, cloning into bacteriophage, subcloning into pBluescript plasmids and DNA sequence determination and analysis for *Plicopurpura columellaris*, *Lumbricus terrestris*, *Limulus polyphemus*, *Homarus americana* and *Daphnia pulex* were essentially as described[20]. Portions of the mtDNAs of *Euperipatoides leuckarti* and *Thyrophygus* sp. were amplified by PCR using oligonucleotides matching well-conserved regions of flanking genes. Horizontal bars indicate unique, phylogenetically uninformative gene arrangements; question marks indicate unknown arrangements. Gene designations: COI/COII, cytochrome oxidase subunits I/II; l-rRNA, large subunit rRNA; ND1/ND2, NADH dehydrogenase subunits 1 and 2; tRNA genes are indicated by the corresponding one-letter amino-acid code. The two leucine and two serine tRNA genes are differentiated by the codon recognized (UUR or CUN for leucine, AGN or UCN for serine). Transcription is from left to right except for underlined genes. '424' refers to a non-coding region of 424 nucleotides[20].

reversion of the onychophoran mtDNA to the more primitive arrangement of *l-rRNA/tRNA*$^{L(CUN)}$/*tRNA*$^{L(UUR)}$/*ND1*. Our inference of Onychophora as either the sister group to Chelicerata or to Arthropoda is consistent with the results of comparing mitochondrial rRNA sequences[11] and morphological features[1].

The positions of the leucine tRNA genes in other insect mtDNAs are similar or identical to those in *Drosophila*. They are identical in another dipteran, the mosquito *Anopheles*[24], and in a hymenopteran, the honeybee *Apis*[23]. In an orthopteran, the grasshopper *Locusta*, *tRNA*$^{L(UUR)}$ is upstream of *COII* and *tRNA*$^{L(CUN)}$ is upstream of *ND1*[26], although the other flanking genes are currently unknown. A lepidopteran, the moth *Spodoptera*[27], and another dipteran, the fly *Simulium*[28], have the arrangement *l-rRNA/tRNA*$^{L(CUN)}$/*ND1*, but the position of *tRNA*$^{L(UUR)}$ is unknown. Finally, polymerase chain reaction amplification of *COII* from representatives of 10 different insect orders has been achieved using a primer that anneals to *tRNA*$^{L(UUR)}$ (ref. 29). Thus *tRNA*$^{L(UUR)}$ must be immediately upstream of *COII* in each of these insects.

Obviously, not all divergences will coincide with mtDNA rearrangements, and independent changes in one or more lineages may subsequently erase traces of relatedness. However, when shared rearrangements are preserved, relationships can be reliably inferred. The great number of arrangements theoretically possible makes it very unlikely that taxa will acquire the same arrangement by chance, so even if some relationships cannot be resolved by such data, those that can are likely to be correct. This is affirmed in this study, where all gene order characters considered are either autapomorphic or support the phylogeny presented here. Although relatively few characters are available for comparison, they are not subject to many of the shortcomings attributed to morphological studies (for example, convergence resulting from common selective pressures, or ambiguity in determining homologous structures), functional arguments (for example, lack of falsifiability), or primary sequence comparisons (for example, problems of alignment, base compositional effects, or excessive homoplasy). The relationships of taxa thought to share a close evolutionary history with arthropods, such as tardigrades and pentastomids, can be addressed with similar data. We anticipate that, as mitochondrial gene arrangements are determined for more taxa, additional evolutionary relationships will be clarified. □

Received 11 October 1994; accepted 18 May 1995.

1. Brusca, R. C. & Brusca, G. L. *Invertebrates* (Sinauer, Sunderland, Massachusetts, 1990).
2. Manton, S. M. & Anderson, D. T. in *The Origin of Major Invertebrate Groups* (ed. House, M. R.) 269–322 (Systematics Association, London, 1979).
3. Kukalová-Peck, J. *Can. J. Zool.* **70,** 236–255 (1992).
4. Boudreaux, H. B. *Arthropod Phylogeny with Special Reference to Insects* (Wiley, New York, 1979).
5. Tiegs, O. W. *Q. J. microsc. Sci.* **82,** 165–268 (1947).
6. Cisne, J. L. *Science* **186,** 13–18 (1974).
7. Hessler, R. R. & Newman, W. A. *Fossils Strata* **4,** 437–459 (1975).
8. Valentine, J. W. *Proc. natn. Acad. Sci. U.S.A.* **86,** 2272–2275 (1989).
9. Lake, J. A. *Proc. natn. Acad. Sci. U.S.A.* **87,** 763–766 (1990).
10. Field, K. G. *et al. Science* **239,** 748–753 (1988).
11. Ballard, J. W. O. *et al. Science* **258,** 1345–1348 (1992).
12. Hendriks, L., van Broeckhoven, C., Vandenberghe, A., van de Peer, Y. & DeWachter, R. *Eur. J. Biochem.* **177,** 15–20 (1988).
13. Boore, J. L. & Brown, W. M. *Nautilus* **108** (Suppl. 2), 61–78 (1994).
14. Wheeler, W. C., Cartwright, P. & Hayashi, C. Y. *Cladistics* **9,** 1–39 (1993).
15. Felsenstein, J. *Syst. Zool.* **27,** 401–410 (1978).
16. Vawter, L. & Brown, W. M. *Genetics* **134**(2), 597–608 (1993).
17. Patterson, C., Williams, D. M. & Humphries, C. J. *A. Rev. Ecol. Syst.* **24,** 153–188 (1993).
18. Adoutte, A. & Philippe, H. *Experientia (Suppl.)* **63,** 1–30 (1993).
19. Wolstenholme, D. R. *Int. Rev. Cytol.* **141,** 173–216 (1992).
20. Boore, J. L. & Brown, W. M. *Genetics* **138,** 423–443 (1994).
21. Sankoff, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **89,** 6575–6579 (1992).
22. Smith, M. J., Arndt, A., Gorski, S. & Fajber, E. *J. molec. Evol.* **36,** 545–554 (1993).
23. Crozier, R. H. & Crozier, Y. C. *Genetics* **133,** 97–117 (1993).
24. Mitchell, S. E., Cockburn, A. F. & Seawright, J. A. *Genome* **36,** 1058–1073 (1993).
25. Valverde, J. R., Batuecas, B., Moratilla, C., Marco, R. & Garesse, R. *J. molec. Evol.* **39,** 400–408 (1994).
26. Haucke, H.-R. & Gellissen, G. *Curr. Genet.* **14,** 471–476 (1988).
27. Pashley, D. P. & Ke, L. D. *Molec. Biol. Evol.* **9**(6), 1061–1075 (1992).
28. Pruess, K., Zhu, X. & Powers, T. *J. med. Ent.* **29**(4), 644–651 (1992).
29. Liu, H. & Beckenbach, A. T. *Molec. Phylog. Evol.* **1,** 41–52 (1992).
30. Maddison, W. P. & Maddison, D. R. *Analysis of Phylogeny and Character Evolution* (Sinauer, Sunderland, Massachusetts, 1992).

ACKNOWLEDGEMENTS. This work was supported by grants from the NIH, NSF and the Rackham Graduate School of the University of Michigan. We thank E. Weinstein and B. Salisbury for technical assistance; J. M. Turbeville for the gift of *Euperipatoides* mtDNA and PCR primers; and S. I. Fuerstenberg, J. M. Turbeville, F. Kraus and G. Naylor for comments. PCR primer sequences and the matrix used for comparing gene arrangements are available from the authors. All DNA sequences except those for *Homarus* and *Daphnia* have been placed in GenBank under accession numbers U24570 and U29703–U29719; *Homarus* and *Daphnia* sequences are available on request.

Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution of myriapods

Markus Friedrich & Diethard Tautz

Zoologisches Institut der Universität München, Luisenstrasse 14, 80333 München, Germany

THE evolutionary relationships among arthropods are of particular interest because the best-studied model system for ontogenetic pattern formation, the insect *Drosophila*, is a member of this phylum. Evolutionary inferences about the developmental mechanisms that have led to the various designs of the arthropod body plan depend on a knowledge of the phylogenetic framework of arthropod evolution. Based on morphological evidence[1–3], but also on palaeontological considerations[4], the sister group of the insects is believed to be found among the myriapods. Using nuclear ribosomal gene sequences for constructing a molecular phylogeny, we provide strong evidence that the crustaceans and not the myriapods should be considered to be the sister group of the insects. Moreover, the degree of sequence divergence suggests that the diversification of the myriapods occurred during the Cambrian. Our findings have general implications for the course of land colonization by the different arthropod groups, as well as for the interpretation of primitive and derived features of arthropod morphology.

To study the phylogenetic relationships of arthropods, we have obtained extensive sequence information from the nuclear ribosomal genes of taxa that cover most of the evolutionary divergence of each of the four major extant arthropod subgroups (Fig. 1). The data were analysed with respect to parameters that are known to be crucial for the accuracy of molecular phylogenetic methods such as homogeneity of nucleotide composition and similarity of evolutionary rates (Fig. 1). To estimate phylogenetic trees, we applied neighbour-joining with distances corrected for multiple hits and gamma-distributed rates across sites[5], weighted maximum parsimony[6] and maximum likelihood[7]. The bootstrap method was used to quantify the support for the tree topology[8].

We find that all three procedures agree with very high support on three major nodes (nodes *a*, *b* and *c* in Fig. 1). Node *a* supports the monophyletic origin of the arthropod taxa included, node *b* concerns the monophyly of the Chelicerata and node *c* unites the Crustacea and the Hexapoda into a monophyletic group. These results have also been partly suggested previously on the basis of less extensive comparisons of both nuclear and mitochondrial ribosomal genes of some relevant taxa (reviewed in refs 3, 9).

The major groups reconstructed in our tree correspond in principle to the concept of the four extant arthropod classes: Myriapoda, Chelicerata, Crustacea and Hexapoda. The support for the monophyly of the Myriapoda is also notable, given the numerous suggestions for their being paraphyletic with respect to the Hexapoda[3,10]. But the support for the monophyletic status of the Crustacea is weak. Our data suggest that these might also be paraphyletic with respect to the Insecta. Whereas the shortest