The Phylogenetic Status of Arthropods, as Inferred from 18S rRNA Sequences[1]

J. McClintock Turbeville, David M. Pfeifer,[2] *Katharine G. Field,*[3] *and Rudolf A. Raff*

Institute for Molecular and Cellular Biology and Department of Biology, Indiana University

Partial 18S rRNA sequences of five chelicerate arthropods plus a crustacean, myriapod, insect, chordate, echinoderm, annelid, and platyhelminth were compared. The sequence data were used to infer phylogeny by using a maximum-parsimony method, an evolutionary-distance method, and the evolutionary-parsimony method. The phylogenetic inferences generated by maximum-parsimony and distance methods support both monophyly of the Arthropoda and monophyly of the Chelicerata within the Arthropoda. These results are congruent with phylogenies based on rigorous cladistic analyses of morphological characters. Results support the inclusion of the Arthropoda within a spiralian or protostome coelomate clade that is the sister group of a deuterostome clade, refuting the hypothesis that the arthropods represent the "primitive" sister group of a protostome coelomate clade. Bootstrap analyses and consideration of all trees within 1% of the length of the most parsimonious tree suggest that relationships between the nonchelicerate arthropods and relationships within the chelicerate clade cannot be reliably inferred with the partial 18S rRNA sequence data. With the evolutionary-parsimony method, support for monophyly of the Arthropoda is found in the majority of the combinations analyzed if the coelomates are used as "outgroups." Monophyly of the Chelicerata is supported in most combinations assessed. Our analyses also indicate that the evolutionary-parsimony method, like distance and parsimony, may be biased by taxa with long branches. We suggest that a previous study's inference of the Arthropoda as paraphyletic may be the result of (*a*) having too few arthropod taxa available for analysis and (*b*) including long-branched taxa.

Introduction

Nucleic acid sequences from small-subunit ribosomal RNAs (18S rRNAs) have proved useful for phylogenetic analysis in eukaryotes. Because of their ubiquity and evolutionary conservation, these molecules are useful for inferring distant phylogenetic relationships, providing a means of assessing relationships between organisms which lack any informative homologous morphological or developmental traits (Sogin et al. 1977; Woese 1987; Field et al. 1988).

In 1988, Field and colleagues published a metazoan phylogeny based on comparisons of partial 18S rRNA sequences by using a distance-matrix method of analysis.

1. Key words: arthropods, chelicerates, phylogeny, rRNA, inference methods.

2. Present address: United States Environmental Protection Agency, Region 5, Water Division, 230 South Dearborn, Chicago, Illinois 60604.

3. Present address: Department of Microbiology, Oregon State University, Corvallis, Oregon 97331.

Address for correspondence and reprints: R. A. Raff and J. M. Turbeville, Institute for Molecular and Cellular Biology and Department of Biology, Indiana University, Bloomington, Indiana 47405.

Mol. Biol. Evol. 8(5):669–686. 1991.
© 1991 by The University of Chicago. All rights reserved.
0737-4038/91/0805-0008$02.00

Downloaded from https://academic.oup.com/mbe/article/8/5/669/1044485 by guest on 21 August 2022

An attempt was made to determine the phylogeny of the arthropods, with particular emphasis on resolving the question of arthropod monophyly or polyphyly, by using representatives of four arthropod groups: *Limulus polyphemus* (horseshoe crab, Chelicerata), *Drosophila melanogaster* (fruit fly, Insecta), *Spirobolus marginatus* (millipede, Myriapoda), and *Artemia salina* (brine shrimp, Crustacea). Also included in the analysis were sequences representing Echinodermata, Chordata, Annelida, and Platyhelminthes, as well as several other phyla. The resulting phylogenetic trees showed that the arthropods were more closely related to each other than they were to any of the other organisms represented, supporting the hypothesis of arthropod monophyly (In order to minimize confusion, we have not restricted the terms "monophyly" and "monophyletic" to cladistic analyses). However, the authors expressed doubt concerning confidence in the placement of the arthropods relative to the other taxa included in the analyses. Very little could be determined about the relationships within the Arthropoda. This was attributed to the following two factors: (1) the small sample of arthropods in the study and (2) the fact that, with the exception of *L. polyphemus*, all the arthropod 18S rRNA sequences had accumulated more changes than had the sequences from other organisms in the study. It has been shown that long branches may be artifactually linked in both distance and parsimony trees, potentially obscuring the pattern of relationships between lineages (Olsen 1988*a*; Swofford and Olsen 1990). We suspect this was the case for the long-branch arthropods in the distance analyses of Field et al. (1988). Maximum-parsimony analysis of the same set of sequences supports arthropod monophyly (Patterson 1989). Results of a reanalysis of the sequences of Field et al. (1988) and Abele et al. (1989) by a rate-invariant method termed "evolutionary parsimony" (Lake 1990) support a paraphyletic Arthropoda, although Lake (1990) questioned the robustness of this result.

The question of arthropod monophyly versus polyphyly has long been debated. Proponents of the hypothesis of arthropod polyphyly suggest that features of development, functional anatomy, and morphology of living arthropods are so distinct that arthropods must lack a common arthropod ancestor and that these groups have instead acquired such arthropod characters as a chitinous exoskeleton and jointed appendages independently (Tiegs and Manton 1958; Manton 1972, 1973, 1977; Anderson 1973, 1979; Manton and Anderson 1979). Proponents of arthropod monophyly correctly argue that advocates of arthropod polyphyly have not shown that any of the arthropod taxa are more closely related to any known taxonomic group than they are to each other (Weygoldt 1979, 1986). Cladistic analyses of morphological and developmental characters support monophyly of the Arthropoda (fig. 1; Paulus 1979; Weygoldt 1979, 1986; Ax 1984; Briggs and Fortey 1989). Hypothesized synapomorphies of the Arthropoda include, among others, a unique cuticle containing chitin and protein, transitory coelomic cavities (mixocoel), dorsal blood vessel with paired ostia, pericardial sinus with septum, and a uniquely organized brain. Molecular characters have the potential to test arthropod monophyly independently of morphological and developmental characters.

This study presents 18S rRNA nucleic-acid-sequence characters bearing on the monophyly/polyphyly controversy. We focus on chelicerate arthropods here because our previous study (Field et al. 1988) showed that, unlike the single chelicerate (*L. polyphemus*) studied, mandibulate arthropod 18S rRNA sequences exhibited a large number of substitutions per sequence position, which may cause artifactual association of taxa (Swofford and Olsen 1990). *Limulus polyphemus* is morphologically primitive and represents only a small part of chelicerate diversity. The addition of several other

Downloaded from https://academic.oup.com/mbe/article/8/5/669/1044485 by guest on 21 August 2022

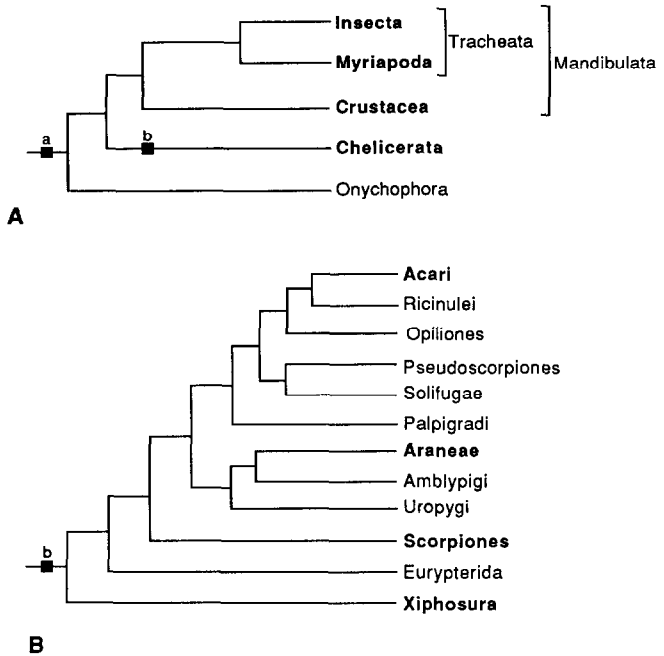

FIG. 1.—Cladograms illustrating relationships among Arthropoda (A) and Chelicerata (B), inferred from analysis of morphological and developmental features. Branch lengths are *not* proportional to number of character-state changes. In cladogram A members of the taxa shown in boldface were included in the present study. Synapomorphies of the Arthropoda indicated by black square "a" include a unique chitinous cuticle with protein, transitory coelomic cavities, dorsal blood vessel with paired ostia, pericardial sinus with septum, and a uniquely organized brain (Ax 1984; Weygoldt 1986). Synapomorphies of the Chelicerata indicated by black square "b" include, among others, chelicerae, prosoma and opisthosoma (=body divisions), and lack of antennae (Weygoldt and Paulus 1979*b*; Weygoldt 1986; also Shultz 1990). Synapomorphies for other taxa may be found in the given references. Of the extant groups represented in cladogram B, those shown in boldface were sampled in the present study and are as follows: Xiphosura, *Limulus polyphemus*; Scorpiones, *Urodacus hoplurus*; Aranae, *Rhechostica chalcodes* and *Eurypelma californica*; and Acari, *Amblyomma americanum*. The Eurypterida are extinct. Cladogram A is redrawn from Ax (1984) and Weygoldt (1986); cladogram B is redrawn from Weygoldt and Paulus (1979*b*).

chelicerates adds diversity to the arthropod data base. In addition, we wanted to determine whether other chelicerates share with *L. polyphemus* a low number of substitutions in 18S rRNA, relative to other arthropods. The present paper confirms the monophyly of the arthropods and of the chelicerates and provides a partial test of arthropod and chelicerate phylogenies that is independent of morphological and developmental characters. Finally, we compare inferences resulting from analyses of ribosomal sequence data by using a distance method, a maximum-parsimony method, and an invariants method.

Material and Methods

Animals

Organisms were chosen to provide a wide sampling of the Chelicerata. In addition, published sequences were utilized. Table 1 lists the animals used for phylogenetic comparisons and gives the source of the sequences. The three new sequences have been submitted to EMBL (see fig. 2).

Downloaded from https://academic.oup.com/mbe/article/8/5/669/1044485 by guest on 21 August 2022

Table 1
Organisms Used for Sequence Comparisons

Classification	Organism	Source
Arthropoda		
Chelicerata		
Arachnida		
Acari	*Amblyomma americanum*	Present study
Scorpiones	*Urodacus hoplurus*	Present study
Araneae	*Rhechostica chalcodes*	Present study
	Eurypelma californica	Hendriks et al. 1988*b*
Xiphosura	*Limulus polyphemus*	Field et al. 1988
Mandibulata		
Crustacea	*Procambarus leonensis*	Kim and Abele 1990
Tracheata		
Myriapoda	*Spirobolus marginatus*	Field et al. 1988
Insecta	*Tenebrio molitor*	Hendriks et al. 1988*a*
Platyhelminthes	*Dugesia tigrina*	Field et al. 1988
Annelida	*Chaetopterus variopedatus*	Field et al. 1988
Echinodermata	*Asterias forbesii*	Field et al. 1988
Chordata	*Homo sapiens*	Torczynski et al. 1985

RNA Extraction

The 18S rRNA sequences of the scorpion *Urodacus* and the spider *Rhechostica chalcodes* were determined by direct sequencing of total cellular RNA. RNA from *U. hoplurus* was extracted using guanidine hydrochloride (Cox 1968; Strohman et al. 1977; Paterson and Roberts 1981). Tissues (excluding the gut) of freshly killed animals were ground in liquid nitrogen with a mortar and pestle. The powder was homogenized in 10 vol 8.0 M guanidine hydrochloride, 25 mM sodium acetate pH 5.0, 0.1 M 2-mercaptoethanol. The supernatant of a 10-min centrifugation at 10,000 *g* at 0°C was saved, and 0.5 vol cold ethanol was added. The sample was stored overnight at −20°C. The preparation was then centrifuged at 10,000 *g* for 10 min at 0°C. The pellet was resuspended with an 18-gauge needle in ~5 vol 5.7 M guanidine hydrochloride containing 20 mM ethylenediaminetetraacetic acid (EDTA) pH 7–8. To this suspension 0.05 vol 2 M potassium acetate pH 5.0 and 0.5 vol ethanol were added next, and the preparation was held at 20°C for 2–6 h. After centrifugation at 10,000 *g* for 20 min at 0°C, the pellet was dissolved in ~1 ml 20 mM EDTA pH 8.0 and was extracted, until the interface was clear, with phenol:chloroform:isoamyl alcohol (25:24:1), and the RNA was precipitated with ethanol and stored at −20°C.

Rhechostica chalcodes RNA was extracted with guanidine thiocyanate (Chirgwin et al. 1979; Turpen and Griffith 1986). After being ground as described above, the powder was homogenized in 10 vol 4.0 M guanidine isothiocyanate, 25 mM sodium citrate, 0.5% sodium lauryl sarkosinate, 0.1% Sigma anti-foam A. RNA was separated through a cesium chloride gradient by centrifugation (14 h at 150,000 *g* average). The pellet was resuspended in 20 mM EDTA and was extracted as above with phenol:chloroform:isoamyl alcohol, and the RNA was precipitated with ethanol.

DNA Extraction

As RNA extraction from the tick *Amblyomma americanum* did not prove feasible, we amplified cloned, and sequenced its 18S rRNA gene. DNA was prepared by a

Downloaded from https://academic.oup.com/mbe/article/8/5/669/1044485 by guest on 21 August 2022

modification of the protocol of Strauss (1987). Twenty-five whole adult unfed ticks were ground as above. The powder was suspended in digestion buffer (100 mM NaCl, 10 mM Tris-HCl pH 8.0, 25 mM EDTA pH 8.0, 0.5% SDS). Proteinase K was added to a concentration of 0.1 mg/ml. After overnight incubation at 50°C, the digested mixture was extracted as above with phenol:chloroform:isoamyl alcohol, followed by extraction twice with chloroform:isoamyl alcohol (24:1). DNA was recovered by ethanol precipitation.

rDNA Amplification and Cloning

The 18S gene of *A. americanum* was amplified using PCR (Sakai et al. 1985, 1988; Scharf et al. 1986; Mullis and Faloona 1987) with eukaryotic-specific 18S rDNA primers and a Cetus thermal cycler (2–22 forward, 5′-CCGTCGACGAGCTCAACCTGGTTGATCCTGCCAGT-3′, 1864–1842, reverse, 5′-CCCGGGTACCAAGCTTGATCCWTCTGCAGGTTCACCTAC-3′; adapted from Medlin et al. 1989). The forward and reverse primers possess, respectively, 14-base and 16-base 5′ extensions that include restriction sites. The amplified product was extracted twice with phenol:chloroform:isoamyl alcohol as above. The product was electrophoresed on a 1% agarose gel, the desired band was excised, and DNA was recovered by electroelution. DNA was digested with *Eco*RI and *Xba*I. Cut fragments were inserted into a Bluescript phagemid vector (Stratagene) which was used to transform *Escherichia coli* cells (strain NM 522). The transformed cells were used to produce single-stranded DNA for sequencing as described in the Sequenase protocol booklet (U.S. Biochemical).

RNA and DNA Sequencing

The RNA was used as a template for dideoxy-terminated sequencing (Sanger et al. 1977) using reverse transcriptase, by following the protocols described by Lane et al. (1985, 1988). DNA sequencing was performed with a Sequenase kit (U.S. Biochemical). Sequencing utilized the five universal reverse primers used by Field et al. (1988), as well as three additional primers (1574–1557, 5′-AGTCCTWTCCGGACAGGG-3′, 860–838, 5′-TTCCATGCACCATTATTCAGGC-3′, 243–222, 5′-AGTCCGATTGGTCTTGG-3′). Sequencing reactions were electrophoresed on 8% acrylamide, 8 M urea gels for 2 h and 4 h (RNA sequencing) or 2 h and 7 h (DNA sequencing).

Data Analysis

Sequence data were aligned by hand, and the alignment was checked against available secondary-structure models (Hendriks et al. 1988*a*, 1988*b*). Highly divergent regions that could not be reliably aligned were excluded from analyses. The data were subjected to three different analysis programs—PAUP 3.0d, the maximum-parsimony program developed by David Swofford (Illinois State Natural History Survey); the distance-matrix program developed by Gary Olsen (1988*a*, 1988*b*) which is based on the method of Fitch and Margoliash (1967); and James Lake's (1987) method of phylogenetic invariants or evolutionary parsimony. The version of Lake's method compiled by D. L. Swofford for the PAUP 3.0d package was utilized. For maximum-parsimony analysis, all data were entered unordered, and gaps and "Ns" were treated as missing data. The BRANCH AND BOUND algorithm of PAUP was used in all cases, except for bootstrapping. Because of the excessive amount of computer time (>5 h) estimated for each bootstrap replication with the BRANCH AND BOUND algorithm, the HEURISTIC search was employed using CLOSEST stepwise addition

Downloaded from https://academic.oup.com/mbe/article/8/5/669/1044485 by guest on 23 August 2022

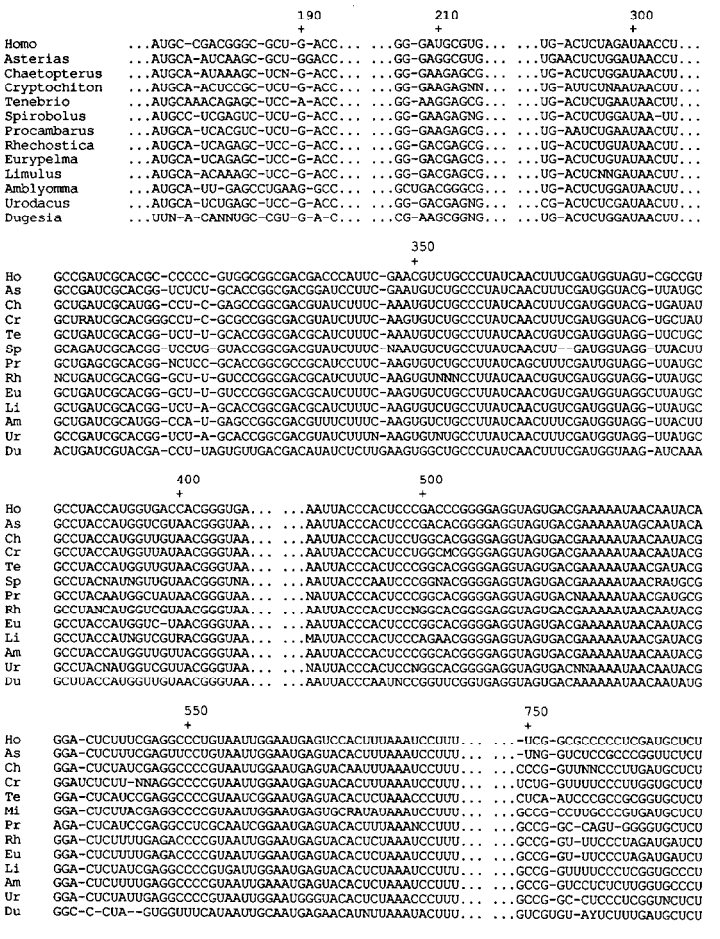

FIG. 2.—Alignment of 18S rRNA sequence data for taxa used. Only positions included in analyses are shown. Numbering of positions is based on alignment to human sequence (Torczynski 1985). Ho = *Homo sapiens;* Eu = *Eurypelma californica;* Li = *Limulus polyphemus;* Am = *Amblyomma americanum;* Rh = *Rhechostica chalcodes;* Ur = *Urodacus hoplurus;* Pr = *Procambarus leonensis;* Sp = *Spirobolus marginatus;* Cr = *Cryptochiton stelleri;* Ch = *Chaetopterus variopedatus;* and Du = *Dugesia tigrina.* Partial sequences of *Amblyomma, Rhechostica,* and *Urodacus* have been deposited in EMBL under accession numbers M60487, M60488, and M60489.

sequence and holding 20 trees at each step. This accelerated the analysis considerably. A comparison between 10 replicates with BRANCH AND BOUND and the same 10 replicates with the described combination of options with the HEURISTIC search revealed identical solutions.

Evolutionary-parsimony analyses were carried out by dividing taxa into four groups. The three possible trees for all quartets composed of a single sequence from each of the four groups were evaluated. The χ^2 values for combined trees were calculated by following the method given by Lake (1987, appendix). The method is designed to take into account any correlation between taxa. In some cases, negative correlation values [ρ in Lake's (1987, appendix) eq. (N4)] were obtained, and in these cases the estimates of correlation were set to zero before the χ^2 values were calculated.

Downloaded from https://academic.oup.com/mbe/article/8/5/669/1044485 by guest on 21 August 2022

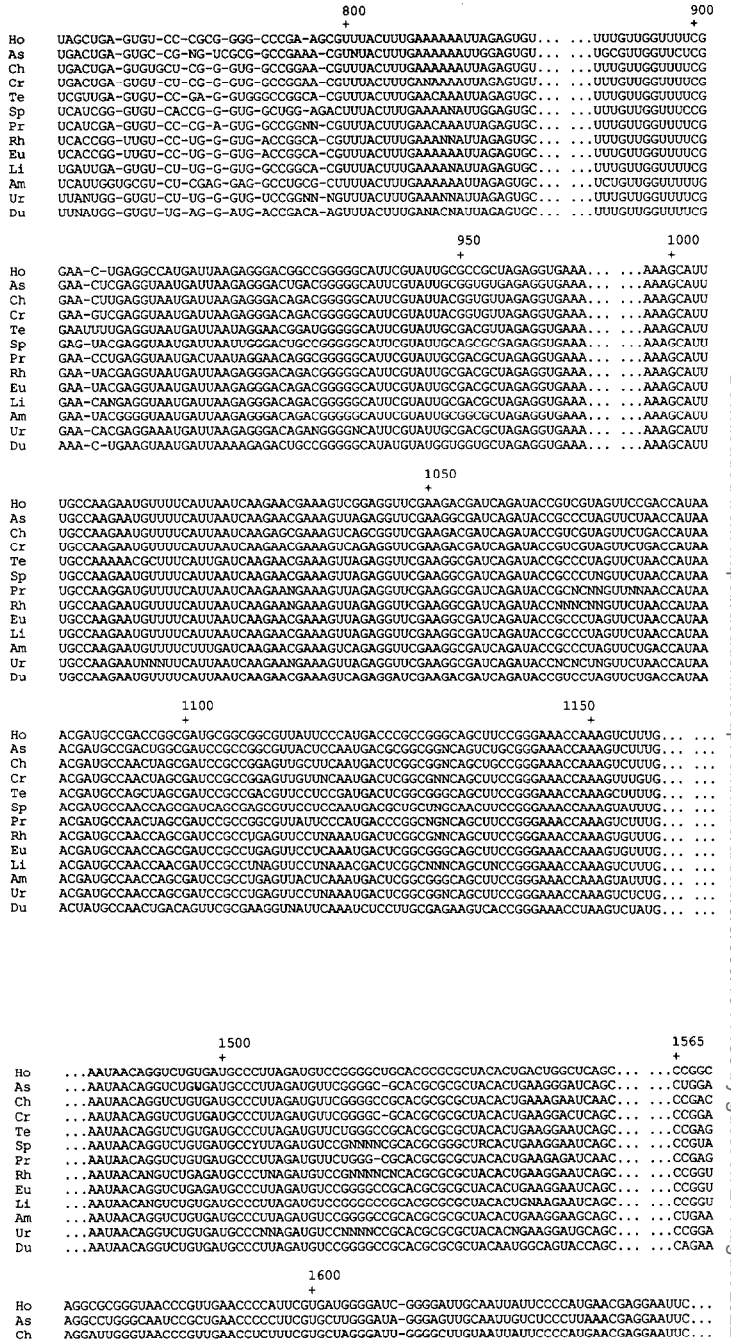

FIG. 2 (*Continued*)

Downloaded from https://academic.oup.com/mbe/article/8/5/669/1044485 by guest on 21 August 2022

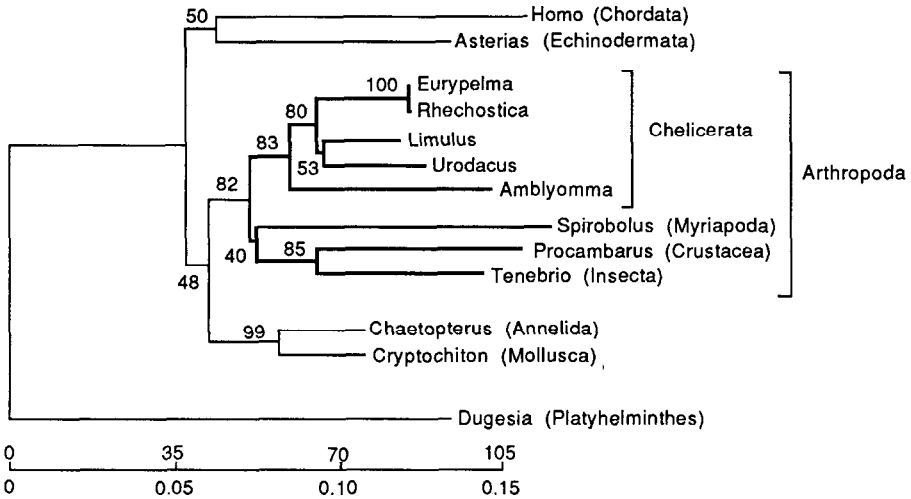

FIG. 3.—Phylogenetic inference derived from distance analysis of sequence data. *Dugesia tigrina* (Platyhelminthes) is included as an outgroup. Average number of substitutions per sequence position is indicated by numbers at the bottom of the scale bar. Numbers at the top of the scale bar indicate average number of substitutions. Numbers at nodes are percentiles and represent bootstrap estimates.

Results

The alignments of the chelicerate and human 18S rRNA sequences are shown in figure 2. The portions chosen represent several relatively well-conserved parts of the 18S rRNA. Sequence was sampled from throughout the molecule to minimize the effects of local areas of rapid change. The total number of nucleotides included in the analysis was ~700. There are 147 cladistically informative sites within the sequences analyzed. The numbering is based on the human sequence.

Very few differences exist between the two spider sequences (99% similar), although the sequence of *Rhechostica chalcodes* was obtained by direct RNA sequencing and that of *Eurypelma californica* by DNA sequencing of a genomic library. This observation confirms that the data obtained by direct RNA sequencing with reverse transcriptase are accurate and can be analyzed with data obtained by DNA sequencing of chelicerate 18S rRNA genes.

Apart from the different placement of the millipede *Spirobolus*, distance analysis and maximum parsimony give essentially congruent results (figs. 3 and 4). The tree generated using the distance method is shown in figure 3. The distance matrix used to generate this tree is provided in table 2. The distance tree indicates that the chelicerates and the other arthropods included in the analysis (*Tenebrio molitor, Spirobolus marginatus,* and *Procambarus leonensis*) form a lineage distinct from the other organisms included in the study. In addition, the distance analysis supports monophyly of the Chelicerata and Mandibulata but not monophyly of the Arachnida or Tracheata.

The exact relationships between members of the Chelicerata are not solidly established with this analysis; in particular, the positions of the tick *Amblyomma americanum* and of the horseshoe crab *Limulus polyphemus* within the chelicerate clade is unexpected on the basis of phylogenies inferred from morphological characters. Although all the chelicerates are relatively conservative in the rates at which the sequences of their 18S rRNA molecules are evolving, the 18S rRNA of *A. americanum* appears to have accumulated more substitutions than have those of the other chelicer-

Downloaded from https://academic.oup.com/mbe/article/8/5/669/1044485 by guest on 21 August 2022

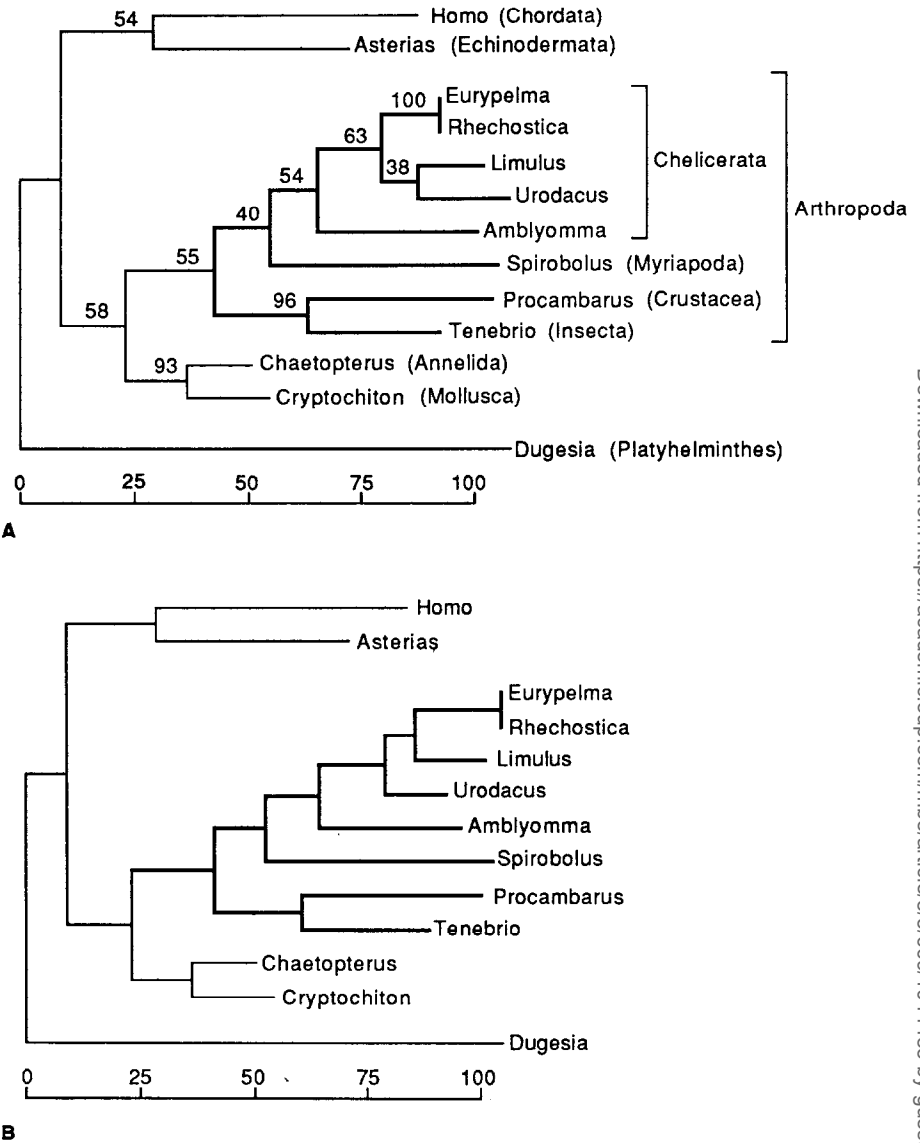

FIG. 4.—Phylogenetic inferences derived from maximum parsimony analysis using PAUP. Branch lengths are proportional to the number of nucleotide substitutions. A, Minimal length tree. The tree length is 576 nucleotide substitutions long, with an overall consistency index of 0.688 and with a consistency index, when uninformative characters are excluded, of 0.560. Numbers at nodes represent the frequency with which clades descending from that node are found by bootstrapping. B, Next most parsimonious tree. Tree length was 577 nucleotide substitutions. The overall consistency index is 0.686, and the consistency index when uninformative characters are excluded is 0.559. Scale bars indicate number of nucleotide substitutions.

ates, particularly in regions where other chelicerate sequences are conservative. This may have the effect of placing the tick deeper into the tree than might otherwise be expected. Bootstrap analysis (Felsenstein 1985) with the distance-matrix program indicates moderate support for arthropod monophyly (82%) and chelicerate monophyly (83%).

Downloaded from https://academic.oup.com/mbe/article/8/5/669/1044485 by guest on 24 August 2022

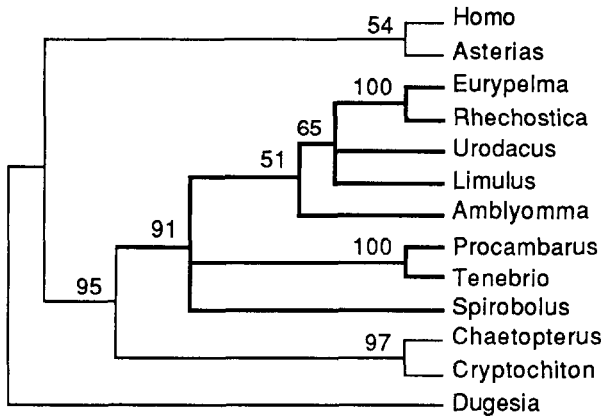

FIG. 5.—Fifty-percent majority-rule consensus tree of 97 trees that lie within ~1% of the length (582) of the shortest tree (576). Values at nodes indicate the number of times the clades descending from nodes were found among the 97 trees saved. The tree was generated by the CONSENSUS function of PAUP.

The shortest tree generated using the maximum-parsimony program PAUP is shown in figure 4. *Dugesia tigrina* (Platyhelminthes) was used as the outgroup. The BRANCH AND BOUND algorithm of PAUP finds a single, most parsimonious tree of 576 nucleotide substitutions, with an overall consistency index of 0.688. The parsimony analysis supports monophyly of the arthropods. The myriapod (*Spirobolous*) is identified as the sister group of a chelicerate clade. This monophyletic group (Myriapoda+Chelicerata) is identified as the sister group of the other arthropods. Monophyly of the Arachnida, of the Mandibulata, and of the Tracheata is not supported. In the next shortest tree *Urodacus* is placed between the tick (*Amblyomma*) and a clade consisting of *Limulus* plus the two spiders. The topology is otherwise similar to that of the shortest tree. The length of the next most parsimonious tree is 577 nucleotide substitutions. This tree has an overall consistency index of 0.686.

A bootstrap analysis was performed to check the robustness of the most parsimonious solution. The results for maximum parsimony are shown in figure 4A. Monophyly of the Arthropoda and the Chelicerata is supported, although not strongly. Relationships among the Chelicerata, exclusive of the two spiders, are weakly to moderately supported (fig. 4A).

In addition to bootstrapping, all trees within 1% (six steps) of the length of the most parsimonious solution were saved, and a 50% majority rule consensus tree was computed in order to further assess the reliability of the phylogeny (fig. 5). Of 97 trees saved, 88 (91%) indicate monophyly of the Arthropoda, 49 (51%) monophyly of the Chelicerata, and 25 (26%) monophyly of the Mandibulata. In only four (~4%) of 97 trees is the Arachnida monophyletic (fig. 5). No tree identifies a monophyletic Tracheata (i.e., Myriapoda+Insecta). This analysis and the bootstrap analyses suggest that the available sequence data are insufficient to unambiguously infer relationships within either the Arthropoda or the Chelicerata.

Evolutionary parsimony indicates support for arthropod and chelicerate monophyly in most combinations evaluated, depending on the "outgroup taxa" included. Inferences of relationships by evolutionary parsimony are shown in figures 6–8. These results are a subset of all possible combinations. Only the favored topologies are illustrated. The *Eurypelma* sequence was omitted from these analyses because it was

Downloaded from https://academic.oup.com/mbe/article/8/5/669/1044485 by guest on 21 August 2022

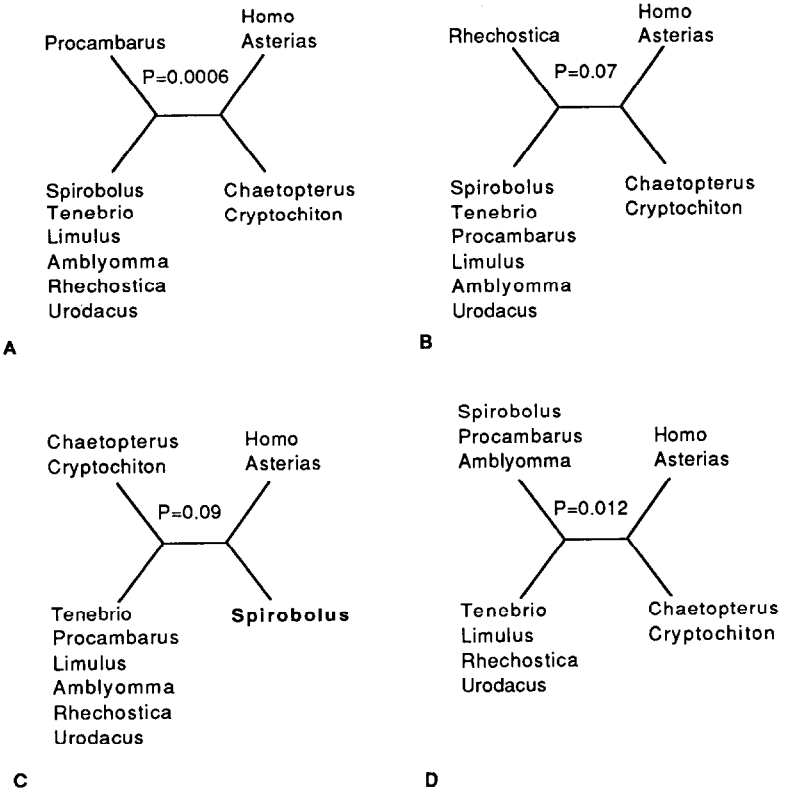

FIG. 6.—Summary of results obtained with evolutionary parsimony for arthropod monophyly. The topologies shown represent the favored topologies from the combination of all trees found when a single taxon from each of the four groups was evaluated in turn. For example, for tree A, six combinations (one arthropod × six arthropods × two spiralian coelomates × two deuterostomes) were combined. *P* values were obtained by first calculating the χ^2 value for correlated data as described by Lake (1987). A, Favored tree when crustacean *Procambarus* is compared with all other arthropods. B, Favored tree when crustacean is replaced by chelicerate *Rhechostica*. Note that support for arthropod monophyly drops. C, Favored tree when myriapod *Spirobolus* is compared with all arthropods. Note that the favored tree links the myriapod with the deuterostomes (*Homo* and *Asterias*) rather than with the rest of the arthropods. However, support is not significant. The *P* value for the expected tree (i.e., a monophyletic arthropoda) is 0.22. D, Favored tree when arthropod taxa are randomly divided into two groups.

99% similar to that of *Rhechostica*. We have not attempted to construct a multitaxon tree from quartets, as no explicit method has been published.

Significant support (i.e., $P \leq 0.05$) for arthropod monophyly is dependent on the taxa included in the analyses. For example, significant support for monophyly of *Procambarus* plus all other arthropods is indicated when the annelid *Chaetopterus* and mollusk *Cryptochiton* and the two deuterostomes (*Homo* and *Asterias*) are included as outgroups and combined as shown in figure 6. Significant support for arthropod monophyly was also indicated when the crustacean was replaced by the insect *Tenebrio* and the chelicerate *Amblyomma*. However, support for arthropod monophyly drops when the remaining taxa are used. When *Rhechostica* is compared with all other arthropods, support decreases ($P = 0.07$), but arthropod monophyly is still favored over the other alternatives. When the millipede *Spirobolus* is compared with all other arthropods, the millipede is linked with the two deuterostomes rather than with the

Downloaded from https://academic.oup.com/mbe/article/6/6/678/1048836 by guest on 21 August 2022

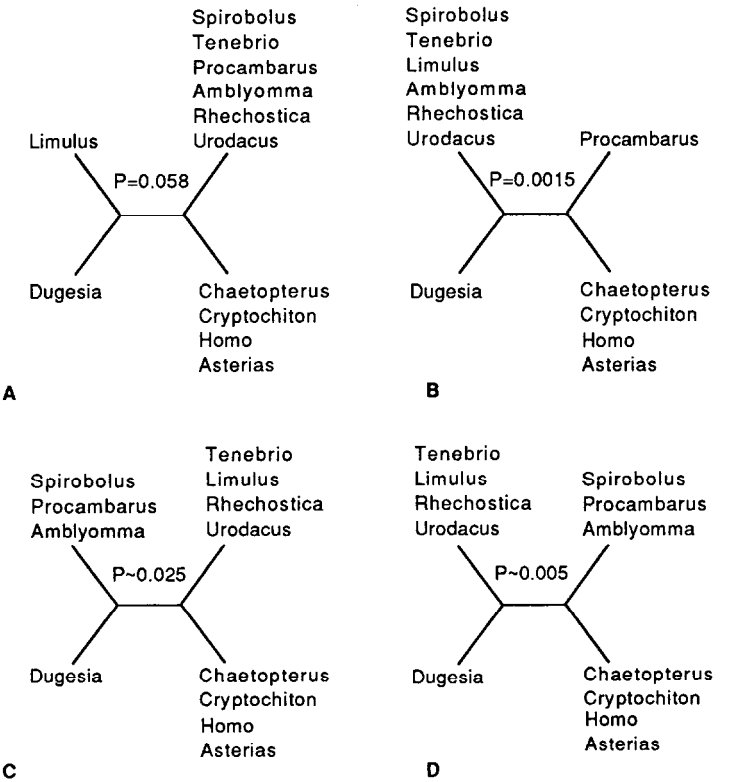

FIG. 7.—Summary of results for arthropod monophyly when long-branched taxon *Dugesia* (Platyhelminthes) is used as "outgroup." A, Favored tree links horseshoe crab *Limulus* with flatworm *Dugesia* rather than with rest of arthropods. B, Favored tree linking crustacean *Procambarus* with deuterostomes and spiralian coelomates rather than with rest of arthropods. C, One of two supported trees when arthropod taxa are divided into two groups. Note that arthropods are not linked. D, Alternative topology from analyses of same combinations as in C. The *P* value of the expected topology (linking the arthropods) is 0.8. Analyses and calculations were performed as described in the text and in fig. 6.

arthropods, although support for this topology is not significant ($P = 0.097$; fig. 6C). The *P* value for the expected tree based on morphological data, linking the millipede with the other arthropods is 0.22. When the arthropod taxa are randomly divided into two groups, results indicate *significant* support for arthropod monophyly. A test of the effects of a long-branch taxon is shown in figure 7. In an analysis of quartets including the long-branch taxon *Dugesia* as an outgroup, in all combinations evaluated the expected tree—i.e., the tree linking the arthropods—is never favored (fig. 7A–D). Note that when the flatworm *Dugesia* is used, the crustacean *Procambarus* is no longer linked with the arthropods (compare fig. 6A with fig. 7B). Results of tests of chelicerate monophyly are shown in figure 8. Chelicerate monophyly is favored in all combinations analyzed, although support is not always significant.

In some cases, we detected a problem with the technique for combining trees. For example, significant support was sometimes found for two mutually conflicting trees (fig. 7C and D). In another instance, the χ^2 value indicated significant support for two topologies, even though the number of informative transversion counts (parsimony-like counts minus background counts) for one topology was four times higher than the number of counts for the other significantly supported topology (not shown).

Downloaded from https://academic.oup.com/mbe/article/8/5/669/1044485 by guest on 21 August 2022

Table 2
Distance Matrix Used to Generate Distance Tree

	As	Ch	Cr	Sp	Li	Am	Ur	Eu	Rh	Te	Pr	Du
Ho	0.16614	0.15266	0.14844	0.21266	0.17866	0.19572	0.17589	0.17538	0.18118	0.19604	0.19368	0.28134
As		0.14337	0.13908	0.18201	0.14182	0.17357	0.14752	0.14663	0.14905	0.15772	0.17661	0.27939
Ch			0.05207	0.15521	0.09282	0.12363	0.11410	0.10982	0.10733	0.13405	0.14639	0.23863
Cr				0.16136	0.10454	0.12780	0.11260	0.10534	0.10562	0.12973	0.13823	0.24603
Sp					0.13860	0.16037	0.14675	0.13517	0.13761	0.16974	0.16253	0.29504
Li						0.10313	0.05495	0.05391	0.05406	0.11003	0.13020	0.24244
Am							0.09873	0.09352	0.09568	0.15122	0.16809	0.26232
Ur								0.06132	0.06083	0.12690	0.13976	0.25360
Eu									0.00149	0.11338	0.12853	0.25624
Rh										0.11639	0.13256	0.25715
Te											0.11360	0.28805
Pr												0.30597

NOTE.—Numbers represent average number of accepted point mutations per sequence position as described by Olsen (1988*a*, 1988*b*). Abbreviations are as follows: Ho = *Homo;* As = *Asterias;* Ch = *Chaetopterus;* Cr = *Cryptochiton;* Sp = *Spirobolus;* Li = *Limulus;* Am = *Amblyomma;* Ur = *Urodacus;* Eu = *Eurypelma;* Rh = *Rhechostica;* Te = *Tenebrio;* Pr = *Procambarus;* Du = *Dugesia.*

Downloaded from https://academic.oup.com/

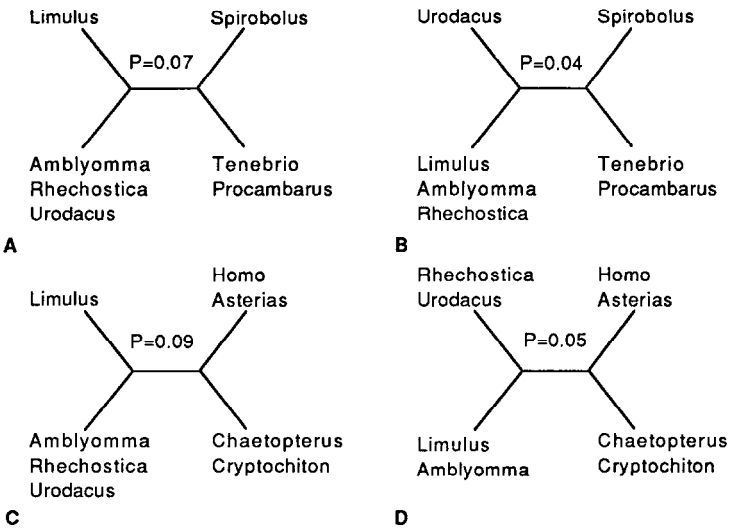

FIG. 8.—Summary of tests for chelicerate monophyly with evolutionary parsimony. Only favored topologies are illustrated. None of the two alternative topologies was supported. Note that the level of significance depends on the taxa compared. A, Test comparing *Limulus* with all other chelicerates by using nonchelicerate arthropods as "outgroups." B, As in A but substituting *Urodacus* for *Limulus*. C, Test using nonarthropods as "outgroups" and comparing *Limulus* to other chelicerates. D, Favored tree when chelicerate taxa are divided into two groups and nonarthropod "outgroups" are used. Analyses and calculations were carried out as explained in the text and in fig. 6.

Discussion

Monophyly of the Arthropoda and the Chelicerata is supported by several morphological synapomorphies (fig. 1; Ax 1984; Weygoldt 1986; Shultz 1990), and analyses of partial 18S rRNA sequences by both maximum-parsimony and distance methods are congruent with these hypotheses. Inferences generated by the evolutionary-parsimony method are congruent with these hypotheses in most cases.

A recent paper (Lake 1990) which reanalyzed published sequences (Field et al. 1988; Abele et al. 1989) by the method of evolutionary parsimony suggested that the Arthropoda is paraphyletic. Results of our analyses with evolutionary parsimony, although not exhaustive, reveal support for arthropod monophyly in most cases when the coelomate taxa are included as "outgroups" but suggest that the method is somewhat sensitive to the taxa chosen and thus to the number of taxa available for evaluation (fig. 6). Our analyses also suggest that, as with maximum parsimony and distance analysis, the evolutionary-parsimony method may be "biased" by long-branch taxa—i.e., taxa which have accumulated a large number of substitutions. This may explain, in part, the uncertain placement of *Spirobolus,* which is the longest-branched arthropod (fig. 6C). Support for arthropod monophyly is never found when the long-branched taxon *Dugesia* is included as an "outgroup." When *Dugesia* is included in the analyses, support is indicated for hypotheses that are untenable on the basis of analyses of morphological characters as well as on the basis of parsimony and distance analysis of the sequence data (fig. 7). This observation of what may be long-branch bias confirms Lake's (1990) own supposition that, despite the fact that it is intended to do so, the evolutionary-parsimony method apparently does not sufficiently correct for the effects of homoplasy in all cases. In agreement with Lake's (1990) qualifications, our results

Downloaded from https://academic.oup.com/mbe/article/5/6/1044/1048180 by guest on 21 August 2022

suggest that the inference of the Arthropoda as paraphyletic (Lake 1990) may be the result of (*a*) having too few arthropod taxa for analysis in combination with (*b*) the potential bias introduced by long-branched taxa. It is also important to note that partial 18S rRNA sequences provide relatively few informative transversion positions [parsimony-like term (P) minus background term (B), which equals 0 to 8 per quartet of sequences in our analyses] for evolutionary-parsimony analysis, and it will be important to determine whether inferences obtained with this method will be more robust, as seems likely, when a larger number of informative positions and additional taxa are available for analysis.

Assessment of the relationship within the Arthropoda is difficult by 18S rRNA sequence analysis. It is clear, however, that the Chelicerata is a monophyletic group within the Arthropoda. Neither the distance nor the parsimony analysis places any of the chelicerates within any other group; they are clearly more closely related to each other than they are to any other group. In addition, arthropod and chelicerate monophyly is supported when the long-branched arthropods *Drosophila melanogaster* (Insecta; Field et al. 1988) and *Artemia salina* (Crustacea; Nelles et al. 1984) are added to the data set (data not shown).

The reason that inference of relationships among the arthropods is not robust may be, in part, because the regions of the molecule sampled in the present study are highly conserved. Although this makes them very useful for comparing distantly related organisms, it limits the ability of the method to discern between more closely related organisms, because the number of sites where the sequences differ is simply not enough to reliably determine close phylogenetic relationships. Even *Amblyomma americanum* and *Limulus polyphemus,* the two most distantly related chelicerates, have 18S rRNA sequences which are ~90% similar. Better resolution should come by using the rest of the 18S rRNA molecule, thereby potentially doubling the number of informative sites being compared, and by including data from other genes. It is significant to note in this regard that although the specific association of echinoderm and chordate 18S rRNA is only indicated by 54% of the bootstrapping outcomes in maximum-parsimony analysis (fig. 5A) and by only 50% of the outcomes in distance analysis, molecular data obtained from the structural analysis of a different gene indicate a close relationship between these phyla (Delgadillo-Reynoso et al. 1989).

The unexpected placement of taxa such as the chelicerate *Amblyomma* and the myriapod *Spirobolus* may be attributable to their having relatively long branches. The potential increase in homoplasy may obscure both the true position of *Amblyomma* within the Chelicerata and the position of *Spirobolus* within the Arthropoda. Neither maximum-parsimony nor distance analysis gives the answer consistent with the morphological and developmental data for these taxa (fig. 1; Weygoldt and Paulus 1979*a,* 1979*b;* Weygoldt 1986; Shultz 1990). Given both the possibility that long-branch taxa may be placed artifactually and the limitations of using conservative portions of the molecule (see above), rigorous treatment of conflicts between the 18S rRNA phylogeny and morphological phylogenies should be deferred until additional molecular data are available and more taxa have been sampled.

With an increased number of arthropods studied at the molecular level, there is moderate support for the existence of the Arthropoda as a monophyletic group comprising the Chelicerata, Crustacea, Myriapoda and Insecta. This finding is consistent with cladistic analyses of morphological characters (see Ax 1984; Weygoldt 1986). Although the exact relationships between Crustacea, Myriapoda, and Insecta remain uncertain on the basis of the 18S rRNA data, it is clear that the Chelicerata form a

Downloaded from https://academic.oup.com/mbe/article/8/5/669/1044485 by guest on 21 August 2022

distinct clade within the Arthropoda. We are presently unable, with the existing 18S rRNA sequence data set, to reliably infer the relationships within the Chelicerata. Distance and maximum-parsimony analyses provide roughly equivalent trees, differing in the placement of the myriapod within the arthropod clade. Results with evolutionary parsimony indicate support for arthropod and chelicerate monophyly in most combinations analyzed, when the coelomate taxa were used as "outgroups."

Inferences with maximum-parsimony and distance methods support the hypothesis that the Arthropoda belongs within a spiralian coelomate clade that is the sister group of a deuterostome clade. This is consistent, in part, with hypotheses based on morphological characters, further supporting the interpretation that segmentation of annelids and arthropods is homologous (see Weygoldt 1986). The results concerning the arthropods are not congruent with the analyses of Field et al. (1988), which were unable to reliably infer the relationships of the arthropods. Nor are these results congruent with those of Lake (1990), which suggest that the arthropods arose before the rest of the protostome coelomates.

Downloaded from https://academic.oup.com/mbe/article/8/5/669/1044435 by guest on 01 August 2022

Acknowledgments

We thank Drs. N. Pace, B. A. Parr, and T. Schmidt for technical advice and Dr. C. Marshall for informative discussions about inference methods and a critical review of the manuscript. Dr. Marshall also provided assistance with statistical methods. We are grateful to G. Olsen, J. Coddington, and three anonymous reviewers for providing helpful comments. Thanks are also due to Noel Tait (Macquarie University), Steve Rissing (Arizona State University) and James Tucker (Oklahoma State University) for assistance in obtaining animals. This study was supported by National Science Foundation grant BSR-8818044 to R.A.R. and K.G.F. and by an Alfred P. Sloan Postdoctoral Fellowship to J.M.T.

LITERATURE CITED

ABELE, L. G., W. KIM, and B. E. FELGENHAUER. 1989. Molecular evidence for inclusion of the phylum Pentastomida in the Crustacea. Mol. Biol. Evol. **6**:685–691.

ANDERSON, D. T. 1973. Embryology and phylogeny in annelids and arthropods. Pergamon, New York and London.

———. 1979. Embryos, fate maps, and the phylogeny of arthropods. Pp. 59–105 *in* A. P. GUPTA, ed. Arthropod phylogeny. Van Nostrand Reinhold, New York.

AX, P. 1984. Das phylogenetische System. Gustav Fischer, Stuttgart.

BRIGGS, D. E. G., and R. A. FORTEY. 1989. The early radiation and relationships of the major arthropod groups. Science **246**:241–243.

CHIRGWIN, J. M., A. E. PRZYBYLA, R. J. MACDONALD, and W. J. RUTTER. 1979. Isolation of biologically active ribonucleic acids from sources enriched in ribonuclease. Biochemistry **18**: 5294–5299.

COX, R. A. 1968. The use of guanidinium chloride in the isolation of nucleic acids. Methods Enzymol. **12B**:120–129.

DELGADILLO-REYNOSO, M. G., D. R. ROLLO, D. A. HURSH, and R. A. RAFF. 1989. Structural analysis of the uEGF gene in the sea urchin *Strongylocentrotus purpuratus* reveals more similarity to vertebrate than invertebrate genes with EGF-like repeats. J. Mol. Evol. **29**:314–327.

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

FIELD, K. G., G. J. OLSEN, D. J. LANE, S. J. GIOVANNONI, M. T. GHISELEN, E. C. RAFF,

N. R. PACE, and R. A. RAFF. 1988. Molecular phylogeny of the animal kingdom. Science **239**:748–753.

FITCH, W. M., and E. MARGOLIASH. 1967. Construction of phylogenetic trees. Science **155**: 279–284.

HENDRIKS, L. R. DE BAERE, C. VAN BROECKHOVEN, and R. DE WACHTER. 1988*a*. Primary and secondary structure of the 18S ribosomal RNA of the insect species *Tenebrio molitor*. FEBS Lett. **232**:115–120.

HENDRIKS, L., C. VAN BROEKHOVEN, A. VANDENBERGHE, Y. VAN DE PEER, and R. DE WACHTER. 1988*b*. Primary and secondary structure of the 18S ribosomal RNA of the bird spider *Eurypelma californica* and evolutionary relationships among eukaryotic phyla. Eur. J. Biochem. **177**:15–20.

KIM, W., and L. G. ABELE. 1990. Molecular phylogeny of selected decapod crustaceans based on 18S ribosomal RNA nucleotide sequences. J. Crustacean Biol. **10**:1–13.

LAKE, J. A. 1987. A rate-independent technique for analysis of nucleic acid sequences: evolutionary parsimony. Mol. Biol. Evol. **4**:167–191.

———. 1990. Origin of the metazoa. Proc. Natl. Acad. Sci. USA **87**:763–766.

LANE D. J., K. G. FIELD, G. J. OLSEN, and N. R. PACE. 1988. Reverse transcriptase sequencing of ribosomal RNA for phylogenetic analysis. Methods Enzymol. **167**:138–144.

LANE, D. J., B. PACE, G. J. OLSEN, D. A. STAHL, M. L. SOGIN, and N. R. PACE. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc. Natl. Acad. Sci. USA **82**:6955–6959.

MANTON, S. M. 1972. The evolution of arthropod locomotory mechanisms, part 10. J. Linnaean Soc. Zool. **51**:203–400.

———. 1973. Arthropod phylogeny—a modern synthesis. J. Zool. Lond. **171**:111–130.

———. 1977. The Arthropoda: habits, functional morphology, and evolution. Oxford University Press, Oxford.

MANTON, S. M., and D. T. ANDERSON. 1979. Polyphyly and the evolution of the arthropods. Pp. 269–321 *in* M. R. HOUSE, ed. The origin of major invertebrate groups. Systematics Association special volume **12**. Academic Press, New York and London.

MEDLIN, L., H. J. ELWOOD, S. STICKEL, and M. L. SOGIN. 1989. The characterization of enzymatically-amplified eukaryotic 16S-like rRNA coding regions. Gene **71**:491–499.

MULLIS, K. B., and F. A. FALOONA. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol. **155**:335–350.

NELLES, L. B., G. VOLCKAERT, A. VANDENBERGHE, and R. DE WACHTER. 1984. Nucleotide sequence of a crustacean 18S ribosomal RNA gene and secondary structure of eukaryotic small subunit ribosomal RNAs. Nucleic Acids Res. **12**:8749–8768.

OLSEN, G. J. 1988*a*. The earliest phylogenetic branchings: comparing rRNA-based evolutionary trees inferred with various techniques. Cold Spring Harbor Symp. Quant. Biol. **52**:825–837.

———. 1988*b*. Phylogenetic analysis using ribosomal RNA. Methods Enzymol. **164**:793–812.

PATERSON, B. M., and B. E. ROBERTS. 1981. Structural gene identification utilizing eukaryotic cell-free translational systems. Pp. 417–437 *in* J. G. CHIRIKJIAN and T. S. PAPAS, eds. Gene amplification and analysis, vol. **2**: Structural analysis of nucleic acids. Elsevier–North Holland, Amsterdam.

PATTERSON, C. 1989. Phylogenetic relationships of major groups: conclusions and prospects. Pp. 471–488 *in* B. FERNHOLM, K. BREMER, and H. JÖRNVALL, eds. The hierarchy of life. Elsevier, Amsterdam.

PAULUS, H. F. 1979. Eye structure and the monophyly of Arthropoda. Pp. 299–383 *in* A. P. GUPTA, ed. Arthropod phylogeny. Van Nostrand Reinhold, New York.

SAKAI, R., D. H. GELFAND, S. STOFFEL, S. J. SCHARF, R. HIGUCHI, G. T. HORN, K. B. MULLIS, and H. A. ERLICH. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239**:487–489.

SAKAI, R., S. SCHARF, R. FALOONA, K. B. MULLIS, G. T. HORN, H. A. ERLICH, and N. A.

Downloaded from https://academic.oup.com/mbe/article/8/5/669/1044486 by guest on 21 August 2022

ARNHEIM. 1985. Enzymatic amplification of B-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science **230**:1350–1354.

SANGER, F., S. NICKLEN, and A. R. COULSEN. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5468.

SCHARF, S. J., G. T. HORN, and H. A. ERLICH. 1986. Direct cloning and sequence analysis of enzymatically amplified genomic sequences. Science **233**:1076–1078.

SHULTZ, J. W. 1990. Evolutionary morphology and phylogeny of Arachnida. Cladistics **6**:1–38.

SOGIN, S. J., M. L. SOGIN, and C. R. WOESE. 1977. Phylogenetic measurement in procaryotes by primary structural characterization. J. Mol. Evol. **1**:173–184.

STRAUSS, W. M. 1987. Current protocols, unit 2.2: Preparation of genomic DNA from mammalian tissue. John Wiley, New York.

STROHMAN, R. C., P. S. MOSS, J. MICOUD-EASWOOD, D. SPECTOR, A. E. PRZYBYLA, and B. M. PATERSON. 1977. Messenger RNA for myosin polypeptides: isolation from single myosin cell cultures. Cell **10**:265–273.

SWOFFORD, D. L., and G. J. OLSEN. 1990. Phylogeny reconstruction. Pp. 411–501 *in* D. M. HILLIS and C. MORITZ, eds. Molecular systematics. Sinauer, Sunderland, Mass.

TIEGS, O. W., and S. M. MANTON. 1958. The evolution of the Arthropoda. Biol. Rev. **33**:255–337.

TORCZYNSKI, R. M., M. FUKE, and A. P. BOLLON. 1985. Cloning and sequencing of a human 18S ribosomal RNA gene. DNA **4**:283–291.

TURPEN, T. H., and O. M. GRIFFITH. 1986. Rapid isolation of RNA by a guanidinium thiocyanate/cesium chloride gradient method. BioTechniques **4**:11–15.

WEYGOLDT, P. 1979. Significance of later embryonic stages and head development in arthropod phylogeny. Pp. 107–136 *in* A. P. GUPTA, ed. Arthropod phylogeny. Van Nostrand Reinhold, New York.

———. 1986. Arthropod interrelationships—the phylogenetic systematic approach. Z. Zool. Syst. Evolutionsforsch. **24**:19–35.

WEYGOLDT, P., and H. F. PAULUS. 1979*a*. Untersuchungen zur Morphologie, Taxonomie und Phylogenie der Chelicerata. I. Morphologische Untersuchungen. Z. Zool. Syst. Evolutionsforsch. **17**:85–116.

———. 1979*b*. Untersuchungen zur Morphologie, Taxonomie, und Phylogenie der Chelicerata. II. Cladogramme und die Entfaltung der Chelicerata. Z. Zool. Syst. Evolutionsforsch. **17**:117–200.

WOESE, C. R. 1987. Bacterial evolution. Microbiol. Rev. **51**:221–271.

WESLEY M. BROWN, reviewing editor

Received July 18, 1990; revision received March 8, 1991

Accepted March 11, 1991

Downloaded from https://academic.oup.com/mbe/article/8/5/669/1044485 by guest on 21 August 2022