Evidence for a clade of nematodes, arthropods and other moulting animals

Anna Marie A. Aguinaldo*, James M. Turbeville†, Lawrence S. Linford*, Maria C. Rivera*, James R. Garey‡, Rudolf A. Raff§ & James A. Lake*

* *Molecular Biology Institute and MCD Biology, University of California, Los Angeles, California 90095, USA*

† *Department of Biological Sciences, University of Arkansas, Fayetteville, Arkansas 72701, USA*

‡ *Department of Biological Sciences, Duquesne University, Pittsburgh, Pennsylvania 15282, and Department of Biology, University of South Florida, Tampa, Florida 33620, USA*

§ *Department of Biology and Indiana Molecular Biology Institute, Indiana University, Bloomington, Indiana 47405, USA*

The arthropods constitute the most diverse animal group, but, despite their rich fossil record and a century of study, their phylogenetic relationships remain unclear[1]. Taxa previously proposed to be sister groups to the arthropods include Annelida, Onychophora, Tardigrada and others, but hypotheses of phylogenetic relationships have been conflicting[2,3]. For example, onychophorans, like arthropods, moult periodically, have an arthropod arrangement of haemocoel[1,4], and have been related to arthropods in morphological and mitochondrial DNA sequence analyses[4,5]. Like annelids, they possess segmental nephridia and muscles that are a combination of smooth and obliquely striated fibres[6]. Our phylogenetic analysis of 18S ribosomal DNA sequences indicates a close relationship between arthropods, nematodes and all other moulting phyla. The results suggest that ecdysis (moulting) arose once and support the idea of a new clade, Ecdysozoa, containing moulting animals: arthropods, tardigrades, onychophorans, nematodes, nematomorphs, kinorhynchs and priapulids. No support is found for a clade of segmented animals, the Articulata, uniting annelids with arthropods. The hypothesis that nematodes are related to arthropods has important implications for developmental genetic studies using as model systems the nematode *Caenorhabditis elegans* and the arthropod *Drosophila melanogaster*, which are generally held to be phylogenetically distant from each other.

We have analysed relationships of arthropods to other taxa by sequencing complete 18S rDNAs from representative taxa, aligning them with existing 18S sequences from other metazoan taxa, and analysing them by using standard phylogenetic techniques[7]. This study confirms the suspected relationships between arthropods and other taxa, such as tardigrades and onychophorans. But by careful consideration of rates of evolution, we find the surprising result that nematodes are also closely related to arthropods.

An outstanding problem with the molecular phylogeny of nematodes is that their 18S sequences evolve too rapidly to be useful for phylogenetic reconstruction. Previously published sequences of nematodes have a substitution rate 2–3 times greater than those of most other Metazoa. Hence special efforts were made to include only the slowest evolving sequences from representative taxa, because errors due to unequal rate effects and alignment artefacts are compounded by including rapidly evolving sequences[8]. To obtain three slowly evolving nematode sequences, 10–20 nematode 18S genes were sequenced (J.R.G., unpublished results). Marked differences are observed, depending upon whether rapidly or slowly evolving sequences are present (Fig. 1). When both rapidly and slowly evolving nematode sequences (bold type) are included, all nematodes branch from the base of the bilateral animals (Fig. 1a), whereas, when only the slowest nematode sequence is included (Fig. 1b), the nematode branches high within the protostomes as the sister taxon of the arthropods. Furthermore, analysis of the slowly evolving protein-synthesis elongation factor EF-1α also place nematodes within the protostomes (J.R.G., A.M.A.A. and J.A.L., unpublished results), suggesting that other evolutionary processes[9] are not responsible. These results are consistent with unequal rates artefactually placing rapidly evolving, long-branched nematode sequences adjacent to the long branch that joins the outgroup to the tree. Molecular sequence analysis, using the available fast-evolving 18S rRNA nematode sequences or faster evolving molecules, has demonstrated a similarly deep placement[10], whereas some morphological studies have predicted a placement similar to that found using only the slowly evolving nematode sequence[11].

To exclude rapidly evolving taxa, all complete 18S rDNA sequences relevant to this study were systematically surveyed. An alignment of about 50 of the most useful complete sequences was constructed and the distances from each taxon to the last common ancestor of protostomes was calculated using the paralinear/LogDet method[12,13] (Table 1). Guided by these distances, the slowest evolving protostome and outgroup taxa were selected (shown in bold). These included the slowest evolving sequences from the following taxa: a cnidarian as an outgroup to triploblastic animals[7,14], a deuterostome as an outgroup to the protostome animals, a polychaete, an oligochaete, a brachiopod, a mollusc, a non-moulting aschelminth, representatives of the six phyla of non-arthropod moulting animals, and four major arthropod groups (a chelicerate, a crustacean, a myriapod and an insect).

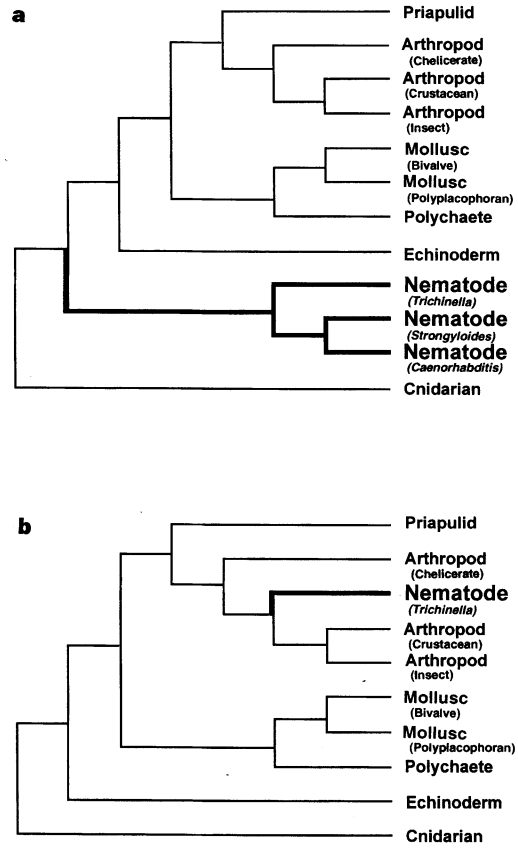

Figure 1 Phylogenetic analysis of 18S rDNA sequence data illustrating the effects of unequal rate biases on nematode placement. In **a**, both rapidly and slowly evolving nematode sequences (*Caenorhabditis* and *Strongyloides*, and *Trichinella*) are included in the analysis; the nematodes branch from the bottom of the tree, even before the deuterostome–protostome divergence. In **b**, only the slowly evolving *Trichinella* sequence is included and this nematode now branches from within the protostome clade, as the sister taxon to the arthropods.

The majority-rule consensus tree derived from phylogenetic reconstructions is shown in Fig. 2. Four reconstruction methods were used, including paralinear (LogDet) distances[12,13], maximum parsimony[9], Kimura two-parameter distances[28], and Jukes–Cantor distances[28]. Paralinear (LogDet) distances[12,13] were emphasized because of their generality (most distance methods are special cases of paralinear distances). As preliminary calculations indicated an excess of constant sites (see Methods), all distance methods were corrected for site-to-site variation. Bootstrap values for these four methods, respectively, are shown adjacent to the interior nodes.

In all of the reconstructions, the protostome taxa are clustered into two monophyletic groups. One clade, containing all the moulting animals (kinorhynch, priapulid, nematomorph, onychophoran, nematode, tardigrade, crustacean, insect, myriapod and chelicerate) is present in 95, 78, 85 and 79% of the trees derived through paralinear distances, maximum parsimony, Kimura two-parameter and Jukes–Cantor distances, respectively. The other protostome clade, containing the articulate brachiopod, mollusc, oligochaete, polychaete and rotifer, is present in 98, 80, 99 and 100% of the bootstrap replicates. A monophyletic protostome clade is also supported in 97, 83, 96 and 95% of the bootstrap replicates. Interpreted using the empirical results of Hillis and Bull[15] as a guideline, these data provide significant support ($P \leq 0.05$) for a clade of arthropod-related moulting animals within the protostomes. This conclusion is further supported by topology-dependent cladistic permutation tail probability tests confirming the significance of the arthropod-related clade ($P \leq 0.01$).

We initially found that flatworm sequences, like rapidly evolving nematode sequences, branched below the base of the bilateral animals. Hence multiple flatworm taxa were sequenced in order to obtain slowly evolving 18S sequences. In experiments similar to those shown in Fig. 1 (with flatworm sequences substituted for nematode sequences), flatworms were shown to branch artefactually deep. Given the importance of the phylogenetic position of the platyhelminthes to theories of the evolution of bilateral animals[16,17], a tree containing slowly evolving lophotrochozoal taxa and the most slowly evolving flatworm, *Stenostomum*, was reconstructed (Fig. 3). Bootstrap support for the clade consisting of the flatworm and other lophotrochozoans is high (91, 88, 83 and 80%, for paralinear distances, maximum parsimony, Kimura two-parameter and Jukes–Cantor distances, respectively), consistent with the placement of the flatworms within the Lophotrochozoa[18].

Divisions within the protostomes have long been a major point of contention among zoologists. Conventional wisdom supports the

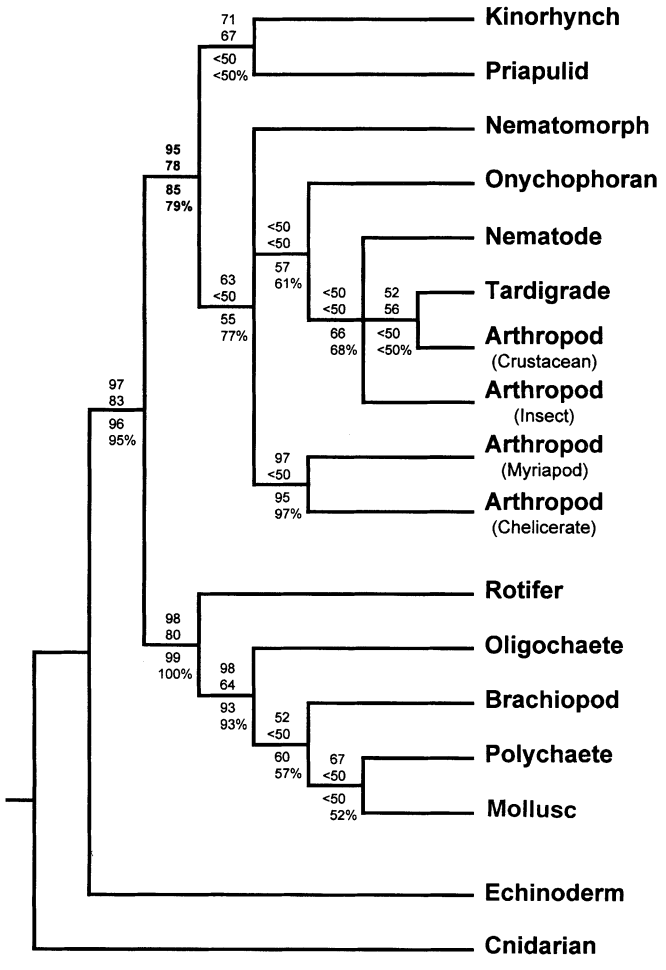

Figure 2 Phylogenetic analysis of 18S rDNA sequence data to determine relationships among the moulting metazoans. The moulting animals are present as the top ten taxa, the Lophotrochozoa are shown in the middle, and outgroups are shown at the bottom. The topology shown here is a majority-rule consensus combining the results from four individual majority-rule consensus trees derived using the following methods: paralinear/LogDet distances, maximum parsimony, Kimura two-parameter distances and Jukes-Cantor distances. All distance methods are corrected for site-to-site variation. The numbers next to the central branches represent the percentage of bootstrap replicates supporting the clades for these methods, respectively (from top to bottom).

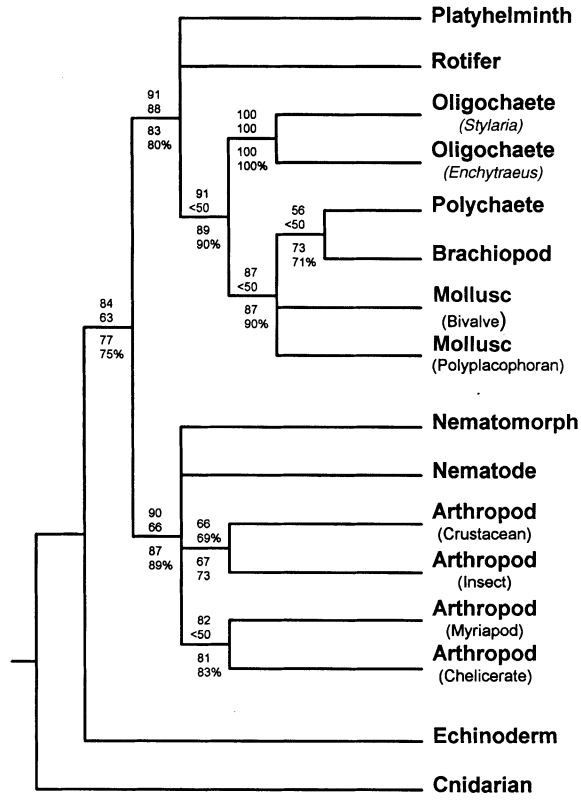

Figure 3 Phylogenetic analysis of 18S rDNA sequence data to illustrate relationships of the flatworm to other protostome animals. The Lophotrochozoa are present as the top eight taxa, the Ecdysozoa are shown in the middle, and outgroups are shown at the bottom. The topology shown here is a majority-rule consensus combining the results from four individual majority-rule consensus trees derived using the following methods: paralinear/LogDet distances, maximum parsimony, Kumura two-parameter distances and Jukes-Cantor distances. All distance methods are corrected for site-to-site variation. The numbers next to the central branches represent the percentage of bootstrap replicates supporting the clades for these methods, respectively (from top to bottom).

existence of a clade, the Articulata, that includes the segmented animals, chiefly the arthropods and the annelids. This concept has a long tradition, but has been called into question by analysis of morphological and palaeontological data[3,19] and of 18S rRNA sequence data[7,20,21]. Eernisse *et al.* characterized two clades within the protostomes, the arthropods and the Eutrochozoa (annelids, molluscs and other protostomes developing from a trochophore larva) with morphological data[3]. A number of studies using 18S data[7,14,20,21,30] identified two clades within the protostomes, the arthropods and the coelomate protostomes of Field *et al.*, now called Lophotrochozoa[7]. The lophotrochozoans include the annelids, molluscs, rotifers, phoronids, brachiopods, bryozoans, platyhelminthes and related phyla. Our data indicate that the sister clade to the lophotrochozoans contains the remaining protostomes, which all develop by moulting. Segmentation does not seem to be a synapomorphy uniting annelids and arthropods. Our analyses, which include four aschelminths (pseudocoelomates), do not support aschelminth monophyly, in agreement with molecular studies[2,10,30]. Our studies are consistent with the clade cephalorhyncha[16].

Our interpretation of these results is shown in Fig. 4. The most obvious feature of this phylogeny is that it separates the protostomes into two groups, an arthropod-related clade exclusively composed of animals that moult, and a lophotrochozoal clade exclusively containing non-moulting animals. All members of the arthropod-related clade undergo ecdysis[22]. In addition, all members lack locomotory cilia, although other groups (for example, chaetognaths and acanthocephalans) also lack them[16]. Given the observed tree topology and these common structural features, this raises the possibility that ecdysis and the cellular modifications associated with it may have been derived only once within this clade.

Because these 18S rDNA analyses support the hypothesis that all moulting animals (arthropods, tardigrades, onychophorans, nematodes, nematomorphs, kinorhynchs and priapulids) share a common ancestor to the exclusion of deuterostomes and the lophotrochozoans, we have chosen the node-based name[23] Ecdysozoa. This group is defined as these taxa plus their last common ancestor and all of its descendants. The name reflects the property that all members of this group, and only members of this group, undergo ecdysis during at least part of their life cycles.

It was unexpected to find nematodes contained within the Ecdysozoa because in previous molecular studies they diverged deep in the protostome tree, even before the deuterostome–protostome bifurcation[10]. Boore *et al.*[24], in their pioneering study using mitochondrial gene order, assumed that nematodes were an outgroup to the protostomes. We realized the results of previous molecular studies could be unequal rate artefacts caused by the extremely rapid nucleotide-substitution rates found in previously published rhabditid nematode sequences, and therefore sequenced numerous nematode species to identify slowly evolving representatives. Unequal rate effects are well documented in theory[15]

Table 1 Substitution rates of 18S rDNA sequences

Phylum	Genus	Substitutions per site
	Lophotrochozoa	
Chaetognatha	*Sagitta*	0.143 ± 0.111
Sipuncula	*Phascolosoma*	0.079 ± 0.007
Pogonophora	*Siboglinum*	0.070 ± 0.008
Platyhelminthes	*Bdelloura*	0.147 ± 0.012
	Fasciolopsis	0.083 ± 0.009
	Stenostomum	**0.063 ± 0.063**
Nemertea	*Lineus*	0.061 ± 0.007
Echiura	*Ochetostoma*	0.058 ± 0.007
Vestimentifera	*Ridgeia*	0.055 ± 0.007
Mollusca	*Lymnaea*	0.060 ± 0.006
	***Placopecten* (bivalve)**	**0.042 ± 0.007**
	***Acanthopleura* (polyplacophoran)**	**0.040 ± 0.006**
Aschelminthes:		
Acanthocephala	*Moniliformis*	0.111 ± 0.009
Gastrotricha	*Lepidodermella*	0.070 ± 0.007
Rotifera	***Brachionus***	**0.058 ± 0.007**
Lophophorates:		
Phoronida	*Phoronis*	0.053 ± 0.007
Ectoprocta	*Plumatella*	0.049 ± 0.006
Brachiopoda	*Glottidia*	0.044 ± 0.006
	Terebratalia	**0.044 ± 0.006**
Annelida	*Eisenia*	0.057 ± 0.007
	Lanice	0.056 ± 0.006
	***Enchytreus* (oligochaete)**	**0.052 ± 0.006**
	***Stylaria* (oligochaete)**	**0.042 ± 0.006**
	***Glycera* (polychaete)**	**0.033 ± 0.005**
	Arthropods and relatives	
Nematoda	*Strongyloides*	0.192 ± 0.014
	Caenorhabditis	0.187 ± 0.013
	Trichuris	0.141 ± 0.012
	Trichinella	**0.110 ± 0.010**
Onychophora	***Euperipatoides***	**0.090 ± 0.009**
Tardigrada	*Milnesium*	0.079 ± 0.008
	Macrobiotus	**0.079 ± 0.009**
Kinorhyncha	***Pycnophyes***	**0.075 ± 0.007**
Nematomorpha	***Gordius***	**0.068 ± 0.007**
Arthropoda	*Artemia*	0.068 ± 0.007
	***Panulirus* (crustacean)**	**0.065 ± 0.008**
	Drosophila	0.121 ± 0.011
	Crossodonthina	0.056 ± 0.007
	***Tenebrio* (insect)**	**0.048 ± 0.006**
	***Scolopendra* (myriapod)**	**0.043 ± 0.006**
	Androctonus	0.046 ± 0.006
	***Eurypelma* (chelicerate)**	**0.038 ± 0.005**
Priapula	***Priapulus***	**0.040 ± 0.005**
	Outgroups	
Chordata	*Lampetra*	0.065 ± 0.007
	Branchiostoma	0.059 ± 0.006
Echinodermata	*Strongylocentrotus*	0.043 ± 0.006
	Antedon	**0.040 ± 0.005**
Ctenophora	*Mnemiopsis*	0.130 ± 0.111
Cnidaria	*Anemonia*	0.101 ± 0.009
	Tripedalia	**0.100 ± 0.009**

Distances are calculated by paralinear/LogDet distances and the ± s.d. estimated from bootstrap replicates. The number of substitutions per position from the last common ancestor of protostomes was calculated with respect to three slowly evolving reference taxa. Distances to protostome taxa were calculated using *Tripedalia* and *Antedon* as outgroup taxa and either *Glycera* or *Priapulus*, depending upon which ingroup taxon was being examined. Distances to outgroup taxa were calculated using *Glycera*, *Priapulus* and *Acanthopleura* as reference taxa.

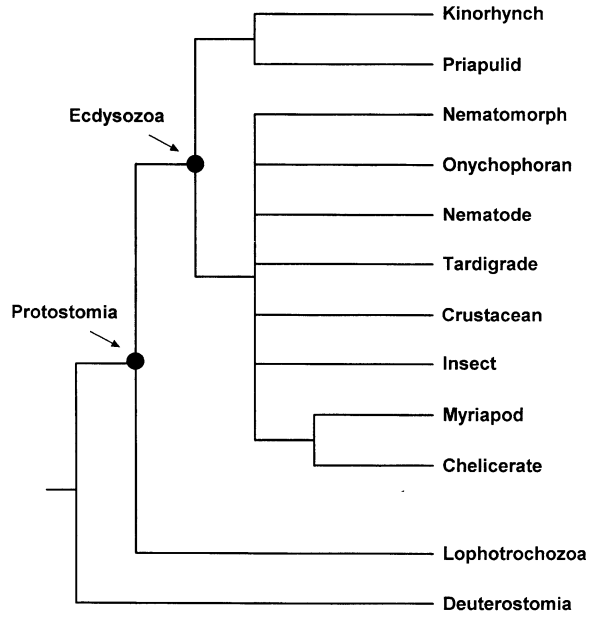

Figure 4 As inferred from 18S rDNA, the Protostomia consists of two major groups. The Lophotrochozoa includes the lophophorates, molluscs, annelids, rotifers and other groups[7]. The Ecdysozoa includes the arthropods, tardigrades, onychophorans, nematomorphs, nematodes, kinorhynchs, priapulids and probably the loriciferans. (So far, no living specimens and fewer than 200 preserved loriciferans (which moult) have been collected. Morphological evidence, however, suggests a close relationship to kinorhynchs and priapulids[16,29].) The common ancestors of these clades are indicated.

but are usually ignored. Morphological studies also support the inclusion of nematodes with many ecdysozoans, although not with arthropods[11,16]. One thoughtful analysis groups nematodes, nematomorphs, priapulids, kinorhynchs and loricifera (but not arthropods, onychophorans and tardigrades) using the synapomorphies, "loss of locomotory cilia, cuticle moulted, introvert with spines, teeth or scalids"[16]. (These first two synapomorphies also serve to unite the ecdysozoa.) Another recent cladistic analysis of morphological characters supports a clade of moulting animals excluding the priapulids[3], although nematomorphs were not included in that analysis.

Given the tremendous interest in the nematode *Caenorhabditis elegans* and the arthropod *Drosophila melanogaster* as model systems, the hypothesis that both are closely related has important implications for developmental and genomic studies. For example, it has been assumed that developmental mechanisms common to *Caenorhabditis* and to *Drosophila* originated before the protostome–deuterostome divergence and hence should also be found in *Homo sapiens*. Our results imply that mechanisms found in both nematodes and fruitflies will not necessarily be found in humans.

The inclusion of the priapulids within an arthropod-containing clade was not anticipated because most morphological studies had not indicated a close priapulid, arthropod phylogenetic relationship[2,3]. Both arthropods and priapulids are numerically prominent members of the Burgess shale faunas[25], indicating the early success (and successful preservation) of ecdysozoans in the Cambrian radiation.

These studies provide evidence that the nematodes are not primitive metazoans but are protostomes related to arthropods. They also support a monophyletic protostome clade. Considering the greatly differing morphologies, embryological features and life histories of the moulting animals, it was initially surprising that the ribosomal RNA tree should group them together. However, given that all moulting taxa sampled are in this clade, and given the significant anatomical modifications associated with moulting, such as the lack of locomotory cilia, ecdysis appears to be a defining synapomorphy for this group, although additional molecular data from other molecules are necessary to test further or confirm the monophyly of the moulting animals. □

Methods

DNA isolation. Total genomic DNA was isolated by standard techniques and amplified by the polymerase chain reaction (PCR). PCR fragments or complete sequences were then cloned into a plasmid vector before sequencing. Replicates of the PCR amplification were sequenced in both directions. A list of the PCR and sequencing oligonucleotides and PCR reaction conditions is available from J.A.L. or J.R.G. (garey@chuma.cas.usf.edu).

Sequences. The following sequences are available in GenBank: *Brachionus plicatilis* (Rotifer; accession number, U49911), *Enchytraeus sp.* (Oligochaete; accession number, U95948), *Euperipatoides leukartii* (Onychophoran; accession number, U49910), *Gordius sp.* (Nematomorph; accession number, U51005), *Macrobiotus sp.* (Tardigrade; accession number, U49912), *Milnesium tardigradum* (Tardigrade; accession number, U49909), *Stenostomum sp.* (Platyhelminth; accession number, U95947), *Stylaria sp.* (Oligochaete; accession number, U95946), and *Trichinella spiralis* (Nematode, accession number, U60231).

Sequence alignments. An alignment of 49 complete sequences was constructed using the star alignment procedure to reduce biases[8], with the slowly evolving *Glycera americana* sequence used as the reference, and then proofread by hand. Pairwise alignments of nucleotide sequences were performed with the ALIGN program, using a break penalty of 6; nucleotide identities, transversions and transitions were scored as +3, +1 and 0, respectively, based on preliminary experiments with EF-1α and 18S rDNA. Regions were excluded from the analysis if extreme length variation existed among sequences, or if many of the sequences contained gaps that could be easily moved with little or no change in alignment score. The alignments are available from J.A.L.

Phylogenetic reconstruction. The 17-taxon phylogenetic trees shown in Figs 2 and 3 were obtained using PAUP version 3.1.1 for maximum parsimony analyses and Bootstrappers gambit[26] for distance analyses. For both methods, 200 bootstrap trees were calculated to determine the 50% majority-rule consensus tree; each search was initiated with 100 replicates of random taxon addition, and positions with gaps were excluded. For parsimony, the following heuristic search options were used: starting trees were obtained by stepwise addition (starting seed was 1) with one tree held at each step; and tree-bisection-reconnection branch-swapping was performed with the MULPARS, but not the steepest descent, option. For paralinear/LogDet, Kimura two-parameter and Jukes–Cantor distances, four-point metrics were used to assess quartet values; the quartet consistency value[26] (53.46%) was selected to ensure that the probability of finding the best solution was >99.9%. A cnidarian and an echinoderm were used as the outgroups, except in Fig. 3 where the two slowest echinoderms were used for parsimony to further reduce unequal rate effects. As site-to-site variation was judged to be significant, distances were corrected for this artefact by estimating nine site categories from the data, calculating distances from the eight non-categories, and estimating trees from the sums of the distances[12].

Site-to-site variation. Site-to-site variation was considered significant when estimated using a diagnostic statistical test for the number of constant sites[27]. Maximum-likelihood trees were calculated using the DNAML program (version 3.4) in PHYLIP. Parameters necessary for the test were calculated for a variety of substitution models using both single and double rate categories determined by the hidden Markov model[28]. An excess of observed constant sites (overpredicted sites) was found for all models, indicating that even two-site categories could not fully explain the data. (Using empirical base frequencies and a transition/transversion ration of 2.0, the best single-site model (rate ratio, 2 : 1; probability of each rate, 0.5, 0.5) predicted 787 ± 39 site versus 1,081 observed sites, and the best two-category model (rate model, 10 : 1; probability of each rate, 0.8, 0.2) predicted 963 ± 38 sites versus 1,081 observed sites. All choices of parameters reconstructed trees with monophyletic ecdysozoal and lophotrochozoal clades, although long computation times prevented bootstrap analysis.)

Bootstrap interpretations. Based on empirical studies of bootstrap analyses, they represent highly conservative estimates of phylogenetic accuracy. Typically for maximum parsimony, bootstrap proportions of ⩾70% correspond to a probability of ⩾95% that the respective clade is a historical lineage. For Gambit, the probabilities are slightly less conservative.

T-PTP test. The topology-dependent cladistic permutation tail probability (T-PTP) test determines whether the difference in length between the shortest tree supporting the monophyly of this clade and the shortest tree not supporting monophyly (5 steps difference) is significantly different from the difference in length expected from randomized data. If the difference in length between the monophyly and non-monophyly trees is outside 95% of the distribution based on randomized data, it can be concluded that the data significantly support monophyly of the clade. We used 200 randomized data sets that were analysed by maximum parsimony.

Received 7 November 1996; accepted 3 April 1997.

1. Boudreaux, H. B. *Arthropod Phylogeny with Special Reference to Insects* (Wiley, New York, 1979).
2. Raff, R. A., Marshall, C. R. & Turbeville, J. M. Using DNA sequences to unravel the Cambrian radiation of the animal phyla. *Annu. Rev. Ecol. Syst.* **25,** 351–375 (1994).
3. Eernisse, D. J., Albert, J. S. & Anderson, F. E. Annelida and Arthropoda are not sister taxa: a phylogenetic analysis of spiralian metazoan morphology. *Syst. Biol.* **41,** 305–330 (1992).
4. Ballard, J. W. *et al.* Evidence from 12S ribosomal RNA sequences that onychophorans are modified arthropods. *Science* **258,** 1345–1348 (1992).
5. Weygoldt, P. Arthropod relationships: the phylogenetic-systematic approach. *Z. Zool. Syst. Evolut.-forsch.* **24,** 19–35 (1986).
6. Brusca, R. C. & Brusca, G. J. *Invertebrates* (Sinauer, Sunderland, MA, 1990).
7. Halanych, K. M. *et al.* Evidence from 18S ribosomal DNA that the lophophorates are protostome animals. *Science* **267,** 1641–1632 (1995).
8. Lake, J. A. The order of sequence alignment can bias the selection of tree topology. *Biol. Evol.* **8,** 378–385 (1991).
9. Stewart, C.-B. The powers and pitfalls of parsimony. *Nature* **361,** 603–607 (1993).
10. Winnepenninckx, B. *et al.* 18S rRNA data indicate that aschelminthes are polyphyletic in origin and consist of at least three distinct clades. *Mol. Biol. Evol.* **12,** 1132–1137 (1995).
11. Wright, K. in *Microscopic Anatomy of Invertebrates* Vol. 4 (eds Harrison, F. W. & Ruppert, E. E.) 111–195 (Wiley-Liss, New York, 1991).
12. Lake, J. A. Reconstructing evolutionary trees from DNA and protein sequences: Paralinear distances. *Proc. Natl Acad. Sci. USA* **91,** 1455–1459 (1994).
13. Lockhart, P. J., Steel, M. A., Hendy, M. D. & Penny, D. Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.* **11,** 605–612 (1994).
14. Field, K. G. *et al.* Molecular phylogeny of the animal kingdom. *Science* **239,** 748–753 (1988).

15. Hillis, D. M. & Bull, J. J. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* **42,** 182–192 (1993).
16. Nielsen, C. *Animal Evolution: Interrelationships of the Living Phyla* (Oxford University Press, 1995).
17. Gee, H. *Before the Backbone: Views on the Origin of the Vertebrates* (Chapman & Hall, London, 1996).
18. Balavoine, G. The early emergence of plaryhelminths is contradicted by the agreement between 18S rRNA and Hox data. *C. R. Acad. Sci. Paris Life Sci.* **320,** 83–94 (1997).
19. Valentine, J. W. Bilaterians of the Precambrian–Cambrian transition and the annelid–arthropod relationship. *Proc. Natl Acad. Sci. USA* **86,** 2272–2275 (1989).
20. Patterson, C. *The Hierarchy of Life* (eds Fernholm, B., Bremer, K. & Jornvall, H.) 471–488 (Elsevier, Amsterdam, 1989).
21. Lake, J. A. Origin of the Metazoa. *Proc. Natl Acad. Sci. USA* **87,** 763–766 (1990).
22. Ruppert, E. E. in *Microscopic Anatomy of Invertebrates* Vol. 4, (eds Harrison, R. W. & Ruppert, E. E.) 1–17 (Wiley-Liss, New York, 1991).
23. de Queiroz, K. & Gauthier, J. Phylogeny as a central principle in taxonomy: Phylogenetic definitions of taxon names. *Syst. Zool.* **39,** 307–322 (1990).
24. Boore, J. L., Collins, T. M., Stanton, S., Daehler, L. L. & Brown, W. M. Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. *Nature* **376** 163–165 (1995).
25. Conway Morris, S. The fossil record and the early evolution of the Metazoa. *Nature* **361,** 219–225 (1993).
26. Lake, J. A. Calculating the probability of multitaxon evolutionary trees: Bootstrappers Gambit *Proc. Natl Acad. Sci. USA* **92,** 9662–9666 (1995).
27. Goldman, N. Simple diagnostic statistical tests of models for DNA substitution. *J. Mol. Evol.* **37,** 650–661 (1993).
28. Felsenstein, J. & Churchill, G. A. A hidden Markov model approach to variation among sites in rate of evolution. *Mol. Biol. Evol.* **13,** 93–104 (1996).
29. Wallace, R. L., Ricci, C. & Malone, G. A cladistic analysis of pseudoceolomate (aschelminth) morphology. *Invert. Biol.* **115,** 104–112 (1996).
30. Garey, J. R., Krotec, M., Nelson, D. R. & Brooks, J. Molecular analysis supports a Tardigrade–Arthropod association. *Invert. Biol.* **115,** 79–88 (1996).

Acknowledgements. We thank C. Marshall for insight and comments, M. Blaxter for nematode DNA; L. Mackey for technical assistance; D. R. Nelson for identifying *Macrobiotus sp.*; and J. Felsenstein for suggestions on how to implement the Goldman Statistical tests. One of us (J.M.T.) did part of this work as a visiting research investigator in the laboratories of R.A.R. and of W. Brown, University of Michigan, Ann Arbor. This research was supported by USDA grants to J.R.G., NSF grants to J.A.L., and NIH grants to R.A.R.

Correspondence and requests for materials should be addressed to J.A.L. (e-mail: lake@mbi.ucla.edu).

An ancestral mitochondrial DNA resembling a eubacterial genome in miniature

B. Franz Lang*, Gertraud Burger*, Charles J. O'Kelly†, Robert Cedergren*, G. Brian Golding‡, Claude Lemieux§, David Sankoff¶, Monique Turmel§ & Michael W. Gray‖

* *Département de biochimie et ¶ Centre de recherches mathématiques, Université de Montréal, Montréal, Québec H3C 3J7, Canada*
† *Bigelow Laboratory for Ocean Sciences, P.O. Box 475, McKown Point, West Boothbay Harbor, Maine 04575, USA*
‡ *Department of Biololgy, McMaster University, Hamilton, Ontario L8S 4K1, Canada*
§ *Département de biochimie, Université Laval, Québec, Québec G1K 7P4, Canada*
‖ *Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada*

Mitochondria, organelles specialized in energy conservation reactions in eukaryotic cells, have evolved from eubacteria-like endosymbionts[1–3] whose closest known relatives are the rickettsial group of α-proteobacteria[4,5]. Because characterized mitochondrial genomes vary markedly in structure[3], it has been impossible to infer from them the initial form of the proto-mitochondrial genome. This would require the identification of minimally derived mitochondrial DNAs that better reflect the ancestral state. Here we describe such a primitive mitochondrial genome, in the freshwater protozoon *Reclinomonas americana*[6]. This protist displays ultrastructural characteristics that ally it with the retortamonads[7,8], a protozoan group that lacks mitochondria[8,9]. *R. americana* mtDNA (69,034 base pairs) contains the largest collection of genes (97) so far identified in any mtDNA, including genes for 5S ribosomal RNA, the RNA component of RNase P, and at least 18 proteins not previously known to be encoded in mitochondria. Most surprising are four genes specifying a multisubunit, eubacterial-type RNA polymerase. Features of gene content together with eubacterial characteristics of genome organization and expression not found before in mitochondrial genomes indicate that *R. americana* mtDNA more closely resembles the ancestral proto-mitochondrial genome than any other mtDNA investigated to date.

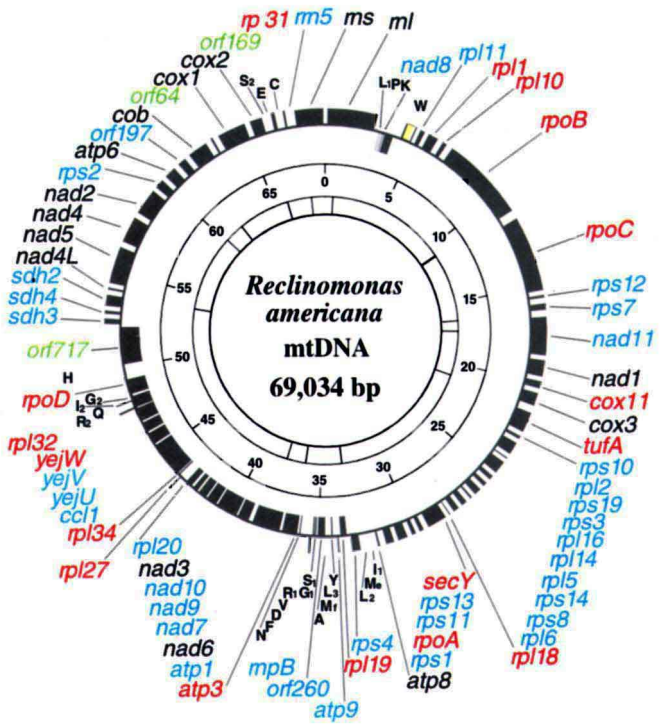

Figure 1 Gene map of the *Reclinomonas americana* mitochondrial genome, with the innermost circle showing the location of *Hin*dIII restriction sites. Identified protein-coding genes are listed in Table 1. The open reading frames (ORFs) *orf197* and *orf260* are homologous to *orf25* (*ymf39*) and *orf244* (*ymf16*), respectively, in liverwort (*Marchantia polymorpha*) mtDNA. Three other ORFs (*orf64*, *orf169* and *orf717*) are unique to *Reclinomonas* mtDNA. Other genes are *rns*, small subunit (SSU) rRNA; *rnl*, large subunit (LSU) rRNA; *rrn5*, 5S rRNA; *rnpB*, RNase P RNA. Transfer RNA genes are indicated by the one-letter amino-acid code, with subscripts denoting different genes specific for the same amino acid. Genes (represented by filled rectangles) shown on the outside of the outermost circle are transcribed in a clockwise direction, whereas those on the inside of the circle are transcribed anti-clockwise. Red, protein-coding genes unique to *R. americana* mtDNA; blue, protein-coding genes absent from vertebrate mtDNAs but generally or occasionally present in plant and protist mitochondrial genomes; green, unique ORFs. A single group II intron (yellow rectangle) is located in the *trnW* gene.

Currently, the inferred set of 'proto-mitochondrial genes' comprises 44 protein-coding genes that specify 23 components of complexes I–V of the electron transport chain, 18 mitoribosomal proteins, and 3 proteins involved in cytochrome c_1 biogenesis (Table 1). In addition, mtDNA encodes up to 3 ribosomal RNAs, up to 27 different transfer RNAs, and (rarely) the RNA subunit of mitochondrial RNase P. At present, therefore, a limited set of about 75 genes of assignable function can be traced directly to the proto-mitochondrial genome, by virtue of their presence in at least several, if not most, contemporary mtDNAs.

In order to provide a more comprehensive picture of mitochondrial genome organization and evolution within the unicellular eukaryotes, which make up the bulk of the biological diversity within the eukaryotic lineage, the Organelle Genome Megasequencing Program (OGMP) is systematically determining the complete mtDNA sequences of selected protists. One of the organisms chosen for this analysis is *Reclinomonas americana* (ATCC 50394), a recently described[6] heterotrophic flagellate. The 'jakobid' assemblage to which *R. americana* has been assigned shares specific