Molecular Phylogeny of the Animal Kingdom

KATHARINE G. FIELD, GARY J. OLSEN, DAVID J. LANE, STEPHEN J. GIOVANNONI, MICHAEL T. GHISELIN, ELIZABETH C. RAFF, NORMAN R. PACE, RUDOLF A. RAFF*

A rapid sequencing method for ribosomal RNA was applied to the resolution of evolutionary relationships among Metazoa. Representatives of 22 classes in 10 animal phyla were used to infer phylogenetic relationships, based on evolutionary distances determined from pairwise comparisons of the 18S ribosomal RNA sequences. The classical Eumetazoa are divided into two groups. Cnidarians arose from a protist ancestry different from the second group, the Bilateria. Within the Bilateria, an early split gave rise to Platyhelminthes (flatworms) and the coelomate lineage. Coelomates are thus monophyletic, and they radiated rapidly into four groups: chordates, echinoderms, arthropods, and eucoelomate protostomes.

THE RESOLUTION OF DISTANT PHYLOGENETIC RELATIONships among animals is one of the most challenging problems in systematic zoology. There is good agreement on the groupings of animals into particular phyla and classes. Thus, the vertebrates form a natural group that is quite clearly demarcated from such phyla as echinoderms or arthropods. Difficulties emerge, however, in the determination of phylogenetic relationships among more distantly related groups. Data from comparative anatomy, embryology, and paleontology have been used to infer relationships (*1–2a*). However, phylogenetic trees based on these approaches remain highly speculative.

Phyla have quite different body plans; few characters are available to unite them into larger units. Many features are restricted to single phyla, and are therefore of no use for establishing relationships among phyla. Even when similar features (for example, segmentation) are shared by different phyla, homology is often uncertain (*3*). There is no fossil record establishing historical continuity of structure for most characters that might be used to assess relationships among phyla. Embryological data suggesting relationships are hard to interpret because the developmental processes underlying these features are incompletely understood (*4*).

It is therefore not surprising that the relationships among major animal groups are controversial. A commonly presented, but not universally accepted, phylogenetic tree for the Metazoa is shown in Fig. 1 (*5*). This tree suggests that the Eumetazoa, or animals with the exception of Porifera (sponges), were monophyletically derived from protists. In this view, the coelenterates (Cnidaria and perhaps Ctenophora), which are radially symmetric and have fewer cell types than members of other major phyla, are the sister group of the bilateral metazoa. Bilateria share various traits such as bilateral symmetry and well-developed mesoderm. Platyhelminthes (flatworms), which branch from the Bilateria first in this tree, lack a coelom (epithelium-lined cavity in the mesoderm containing body organs). Such phyla as annelids, arthropods, brachiopods, echinoderms, and chordates possess a true coelom and complex body plans. Two possible origins have been proposed for mollusks, as true coelomates or as acoelomates with a separate flatworm ancestry (*2a*). Eucoelomate animals are placed in three superphyletic assemblages: protostomes, deuterostomes, and lophophorates. Most protostomes are united by mode of larval development, mesoderm development, and formation of the coelom by splitting of mesoderm. Arthropods have greatly modified development but share other features with protostomes, particularly annelids. However, there is disagreement as to whether arthropods represent a coherent phylum of monophyletic origin or a polyphyletic group with several distinct annelid ancestors (*6, 7*). Deuterostomes are linked by the mode of formation of the mouth, larval morphology, and formation of the mesoderm and coelom from outpouchings of the gut. Lophophorates (brachiopods, phoronids, and bryozoans) are treated as a distinct group on the basis of their embryological features (*8*).

Phylogenetic Inference from 18S Ribosomal RNA Sequences

To address some of the questions left unresolved by traditionally derived phylogenies, we have used 18S ribosomal RNA (rRNA) to investigate metazoan relationships. The primary structures of macromolecules provide an alternative to traditional data for the inference of genealogical relationships (*9*). The use of these structures solves problems of homology and allows an assessment of relationships independent of morphological, biochemical, and developmental traits (*4*). The advantages of 16S to 18S rRNA for phylogenetic studies have been discussed (*10, 11*). We have directly sequenced 18S rRNAs in bulk cellular RNA (*12, 13*) to determine more than 1000 nucleotides of sequence data from each species examined. The use of cellular RNA for these studies guarantees that sequences will represent commonly transcribed RNA genes, not minor or inactive genes. The method provides sequences in the most conservative portions of the 18S rRNA molecule, which are the most useful for broad phylogenetic comparisons. For distantly related organisms, it is not possible to establish homology between nucleotides in the rapidly evolving portions of the molecule; thus, even if the entire 18S rRNA sequence is known, only some parts of it can be used for phylogenetic inference.

Regions of homologous primary and secondary structure in the 18S rRNA sequences were aligned; where necessary, gaps were

K. G. Field, G. J. Olsen, D. J. Lane, S. J. Giovannoni, E. C. Raff, N. R. Pace, and R. A. Raff are at the Institute for Molecular and Cellular Biology and Department of Biology, Indiana University, Bloomington, IN 47405. M. T. Ghiselin is at the California Academy of Sciences, San Francisco, CA 94118. The present address for K. G. Field and S. J. Giovannoni is Department of Microbiology, Oregon State University, Corvallis, OR 97331 and that for D. J. Lane is Gene-Trak Systems, 31 New York Avenue, Framingham, MA 01701.

*To whom correspondence should be addressed.

inserted to compensate for areas of sequence length variation (*14*). Areas of questionable homology or extensive length variation were not used in our analyses; thus residual systematic and random errors were kept small, and errors in sequence determination and alignment are not expected to contribute significantly to uncertainty in branching order (*12*). Sequences were compared by a distance matrix method (*14–16*). The average number of fixed point mutations per position separating each pair of sequences (evolutionary distance) was estimated from the number of observed nucleotide differences between the pair by introducing a correction for parallel and superimposed mutations (*17*). Evolutionary distance estimates were used to infer phylogenetic trees; branching order and branch lengths were adjusted so that the pairwise evolutionary distance estimates were optimally reproduced by the corresponding paths through the tree (*18*). The length of each branch in the inferred phylogenetic tree represents the amount of sequence change assumed to have occurred along that branch; the evolutionary distance separating two sequences is the sum of the lengths of the branches joining the sequences.

Polyphyletic Origins of Metazoa

Comparison of animal 18*S* rRNA sequences shows that species classically defined as Eumetazoa comprise two groups (Fig. 2). The two classes of Cnidaria (coelenterates) used in this study form a natural group with a protist ancestry different from that of the second group, the Bilateria. The relative branching order of the cnidarians, plants, fungi, and ciliates is not resolved by the data.

The position of the root in Fig. 2 separates the cnidarians from the other animals. Other rRNA-based studies of eukaryote phylogeny (*15, 19*) have placed the root elsewhere, with the ciliate lineage diverging before the splitting of plants, animals, and fungi. The differences are a consequence of the sequence positions used in the analyses; we have been restrictive in our identification of homologous residues. The addition of sequence positions not covered by the reverse transcriptase data, but otherwise satisfying our criterion for comparing homologous residues, does not change the tree shown in Fig. 2. Conversely, we can reproduce different root positions (*15, 19*) by including nucleotides of less certain homology between sequences.

Cnidaria have long been held to represent an early and fundamental divergence from other metazoans (*20*). Cnidarians are treated as animals in part because they have muscle and nervous tissue, features characteristic of more complex metazoans. However, the molecular bases for both of these tissue types are evident in protists as well (*21*): actin-myosin motility is well documented in protists, and the ciliates have excitable membranes and molecular mechanisms of sensory transduction resembling the membrane events in neurons of metazoans.

Rapid Radiation of Four Coelomate Groups

Within the Bilateria, an early split separated Platyhelminthes (flatworms) from coelomate animals (Fig. 2). The close relationship among eucoelomate lineages renders it implausible that the coelom originated more than once (*2a*). Our data suggest a rapid radiation of coelomates, resulting in the divergence of four major groups: (i) Chordata, (ii) Echinodermata, (iii) Arthropoda, and (iv) "eucoelomate protostomes," a group consisting of Annelida, Mollusca, Brachiopoda, Sipuncula, and Pogonophora (Vestimentifera). The data do not resolve the branching order of these four major groups; this is evident in the different relative positions of the four eucoelomate groups in Figs. 2 through 5. There may be other primary eucoelomate lineages, but they are not represented in our database.

The lophophorate lineage, represented in our sample by a brachiopod, is solidly affiliated with the protostome group and does not form a separate superphyletic lineage (Fig. 5). In contrast, the arthropods are not close relatives of the annelids. This fact does not invalidate the strong evidence for relationships among metameric organisms, but it indicates an early divergence of arthropods from other metameric lineages.

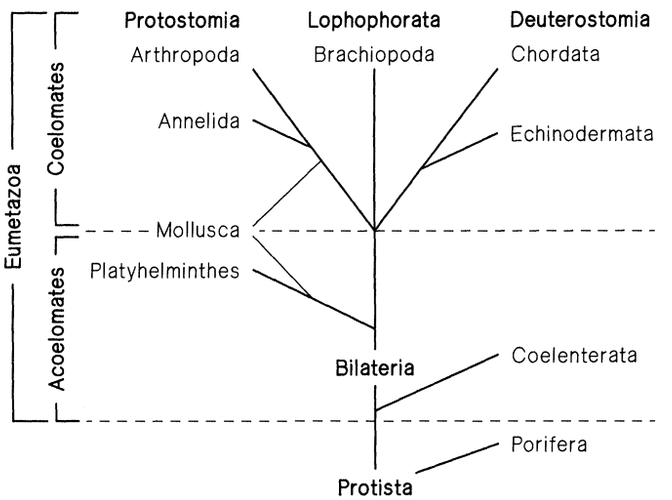

Fig. 1. Phylogenetic tree for the Metazoa, based on the views of Hyman (*5*). This phylogeny is based on morphology of both adults and embryos. Phylum names are shown in lightface lettering.

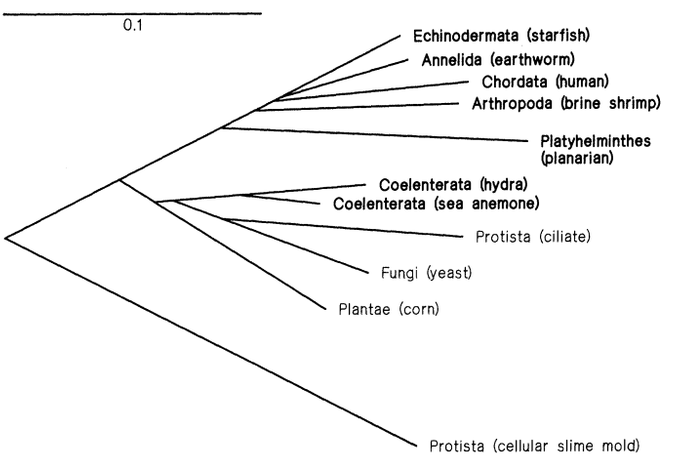

Fig. 2. An evolutionary tree for animals that is based on partial sequences of 18*S* rRNAs. The tree is read from left to right. The root of the tree is provided by the most distantly related organism, the cellular slime mold. The scale bar indicates an evolutionary distance of 0.1 nucleotide substitution per sequence position. Evolutionary distance in the tree is represented by segment length. There is no time scale because constancy of rates of change of the 18*S* rRNA molecule cannot be assumed. Each line terminus represents a living organism for which we have obtained sequence data. The sequences representing Eumetazoa are shown in boldface lettering. Organisms represented: starfish, *Asterias forbesi*; earthworm, *Lumbricus sp.*; human, *Homo sapiens* (*46*); brine shrimp, *Artemia salina* (*47*); planarian, *Dugesia tigrina*; hydra, *Hydra sp.*; sea anemone, *Metridium senile*; ciliate, *Oxytricha nova* (*15*); yeast, *Saccharomyces cerevisiae* (*48*); corn, *Zea mays* (*49*); and cellular slime mold, *Dictyostelium discoideum* (*50*). Sequence positions analyzed: 147 to 190, 209 to 226, 292 to 302, 309 to 321, 330 to 533, 551 to 582, 798 to 833, 841 to 1106, 1126 to 1157, 1443 to 1551, and 1576 to 1658 (numbered according to the human sequence).

Chordates and echinoderms. The branching order of three chordate groups (Fig. 3) agrees with the substantial body of comparative anatomical and embryological data and the extensive fossil record of vertebrates (*22*). The phylogeny of echinoderms and the relationships among them have been a subject of controversy and, in fact, paleontological, embryological, and morphological data have been used to support several alternative trees for the classes (*23*). Our 18*S* rRNA–derived tree (Fig. 3) has allowed us to eliminate most of the possible alternatives.

On embryological grounds, echinoderms and chordates are envisaged to be members of a monophyletic group, the deuterostomes (*24*). Because our data do not resolve closely spaced events in the coelomate radiation, we cannot confirm this hypothesis. However, we can test certain specific hypotheses; one is the suggestion that ancestral vertebrates arose directly from echinoderms during the lower Paleozoic, approximately 500 million years ago (*25*). This possibility is ruled out by our data, which show that echinoderms and chordates have had separate lineages since the eucoelomate radiation. It is known from the fossil record that this radiation occurred before 700 million years ago, well before the lower Paleozoic radiation of living echinoderm classes.

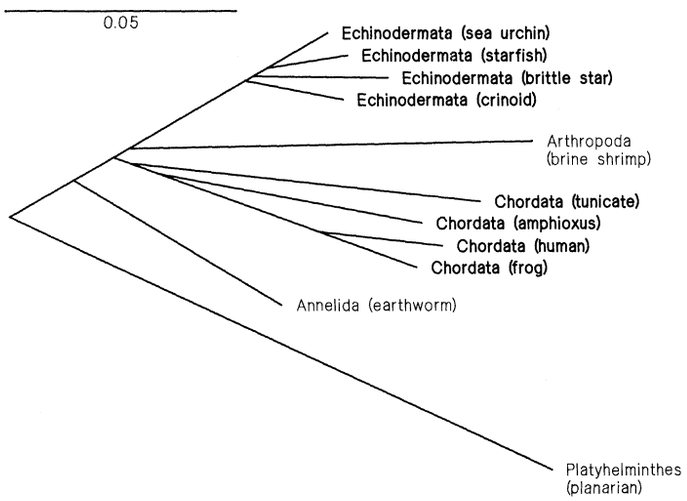

Fig. 3. An expansion of the chordate and echinoderm portions (boldface) of the 18*S* rRNA evolutionary tree for animals. The planarian is included as an outgroup. An arthropod and an annelid represent the other eucoelomate groups. Organisms shown: sea urchin, *Arbacia punctulata*; starfish, *Asterias forbesi*; brittle star, *Ophiocoma wendtii*; crinoid, *Lamprometra palmata*; brine shrimp, *Artemia salina* (*47*); tunicate, *Styela clava*; amphioxus, *Branchiostoma californiense*; human, *Homo sapiens* (*46*); frog, *Xenopus laevis* (*51*); earthworm, *Lumbricus* sp.; planarian, *Dugesia tigrina*. Sequence positions analyzed: 209 to 237, 287 to 302, 309 to 321, 330 to 533, 551 to 582, 798 to 833, 841 to 1157, 1443 to 1551, and 1576 to 1658 (numbered according to the human sequence).

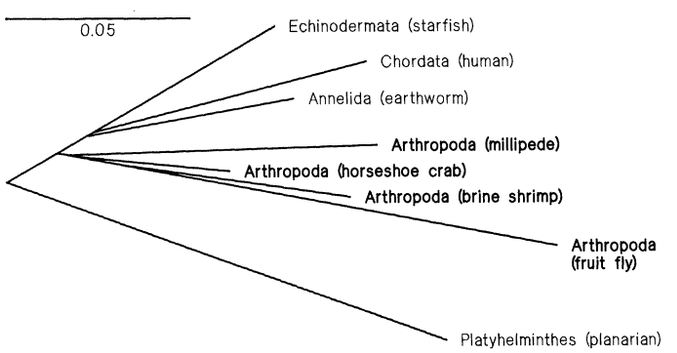

Fig. 4. An expansion of the arthropod portion (boldface) of the 18*S* rRNA tree for animals. The planarian is included as an outgroup. An echinoderm, a chordate, and an annelid represent the other eucoelomate groups. Organisms shown: starfish, *Asterias forbesi*; human, *Homo sapiens* (*46*); earthworm, *Lumbricus* sp.; millipede, *Spirobolus marginatus*; horseshoe crab, *Limulus polyphemus*; brine shrimp, *Artemia salina* (*47*); fruit fly, *Drosophila melanogaster*, and planarian, *Dugesia tigrina*. Sequence positions analyzed: 147 to 192, 207 to 237, 287 to 302, 309 to 321, 330 to 533, 551 to 582, 798 to 833, 841 to 1157, 1443 to 1551, and 1576 to 1658 (numbered according to the human sequence).

Arthropods. The fossil record shows that the radiation of Arthropoda was complex (*26*). The three living subphyla of arthropods are chelicerates (horseshoe crabs, spiders, scorpions, and related groups), crustaceans (crabs, lobsters, and related groups) and unirames (millipedes, insects, and related groups). In the traditional view, Arthropoda are treated as a monophyletic group, with unirames and crustaceans cited as Mandibulata (*7*). A second view is that arthropods are polyphyletic because embryological features indicate that crustaceans may have arisen from polychaete-like annelids, whereas unirames may have evolved from oligochaete-like annelids (*6*).

In our tree for arthropods (Fig. 4), the chelicerates are represented by *Limulus*, the crustaceans by *Artemia*, and the unirames by *Drosophila* and *Spirobolus*. Arthropoda is the largest animal phylum, and the 18*S* rRNA has diversified within the group on a scale comparable to that of the rest of the eucoelomates together. Our data do not indicate polyphyletic origin of Arthropoda from any taxa in our sample. Two of the Mandibulata (*Artemia* and *Drosophila*) are linked, but the millipede occupies an unexpectedly deep position in the tree. Placement of the arthropod lineages is difficult because only a small sample of arthropod diversity is represented, and because three of the four arthropods sampled (*Artemia*, *Drosophila*, and *Spirobolus*) are species that have accumulated large numbers of nucleotide substitutions ("fast clock" species).

Eucoelomate protostomes. The fourth eucoelomate group defined by 18*S* rRNA sequences includes mollusks, annelids, a brachiopod, a sipunculan, and a vestimentiferan (Fig. 5). These phyla diverged within a narrow interval. The close branch points and widely differing amounts of 18*S* rRNA divergence in these animals render the exact branching order uncertain. The 18*S* rRNA phylogeny suggests complex relationships between the annelids and mollusks. These data, and the fact that arthropods, which are presumed to have arisen from annelid-like ancestors, form a more distant branch, suggest that many eucoelomate lineages arose from a metameric ancestor. Sipunculids show no metamerism in their ontogeny, but they are highly modified. There is strong evidence for annelidan features in the ancestry of mollusks (*27*), but it is usually assumed that mollusks diverged from the annelidan line before segment addition occurred by teloblastic growth (*28*). An alternative scenario is that segments of mollusks are homologous to larval segments in primitive arthropods and annelids, and that posterior segments have been lost through progenesis (*29*). This hypothesis would explain why the earliest mollusks were very small. Our results rule out the suggestion that mollusks and other eucoelomate animals arose independently from separate acoelomate lineages (*2a, 30*).

Brachiopods, representing the lophophorates, belong to the eucoelomate protostome group (Fig. 5). This assignment is supported by classical investigations suggesting that the mouth of brachiopods arises near the site of the embryonic blastopore (*31*), although this point has been disputed (*32*). There has been some controversy as to whether *Riftia*, the giant tube worm that inhabits hydrothermal vent sites, should be placed in the phylum Pogonophora, among the annelids, or in a separate phylum, the Vestimentifera (*33*). *Riftia* 18*S* rRNA is solidly affiliated with that of the protostomes (Fig. 5).

Close phylogenetic relationships among the classic protostome groups (except arthropods) have been confirmed. It is also apparent that the relationships among protostomes are more complex than was previously realized, although these 18S rRNA sequence data should not be overinterpreted among groups with such closely spaced and complex radiations. Analyses of additional portions of the 18S rRNA molecule, coupled with data from other molecules, will better establish the branching order.

Animal Radiations

The 18S rRNA relationships distinguish at least three periods of radiation of multicellular "animals." The first, which gave rise to lineages leading to modern Cnidaria and modern Bilateria, was probably part of the extensive diversification that occurred among eukaryotic protists in the Precambrian era. This was not a radiation of animals per se, but of protist lineages, some of which independently gave rise to multicellular animals (*34*). The second radiation occurred in the Bilateria, with the separation of the lineage leading to modern flatworms. Our data for acoelomate animals show only the dim outline of what may have been a more diverse radiation. A third radiation occurred with the splitting of the four eucoelomate groups.

The 18S rRNA comparisons suggest rapid phyletic splitting of major animal groups. These radiations may have followed major evolutionary innovations. The first radiation may have resulted from the creation of diverse new niches for eukaryotes by higher atmospheric oxygen concentrations. The second radiation may correlate with the evolution of the bilaterian body plan, opening up possibilities for more advanced locomotion and sensory systems. The third radiation may have been a consequence of the evolution of the true coelom, which made large body sizes and more powerful burrowing and other locomotory patterns possible (*35*).

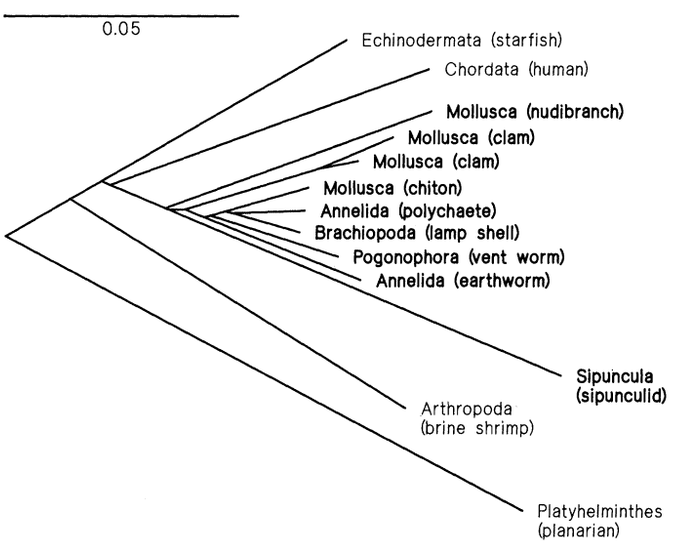

Fig. 5. An expansion of the eucoelomate protostome portion (boldface) of the 18S rRNA evolutionary tree for animals. The planarian is included as an outgroup. A chordate, an echinoderm, and an arthropod represent the other eucoelomate groups. Organisms shown: starfish, *Asterias forbesi*; human, *Homo sapiens* (*46*); nudibranch, *Anisodoris nobilis*; clam, *Mya arenaria*; clam, *Spisula solidissima*; chiton, *Cryptochiton stelleri*; polychaete, *Chaetopteris* sp.; lamp shell, *Lingula reevi*; vent worm, *Riftia pachyptila*; earthworm, *Lumbricus* sp.; sipunculid (peanut worm), *Golfingia gouldii*; brine shrimp, *Artemia salina* (*47*); and planarian, *Dugesia tigrina*. Sequence positions analyzed: 209 to 237, 287 to 302, 309 to 321, 330 to 533, 551 to 582, 798 to 833, 841 to 1157, 1443 to 1551, and 1576 to 1658 (numbered according to the human sequence).

These innovations cannot be dated with molecular sequence data because we cannot assume a constant rate of change in molecules; dates must come from paleontology. The timing of the first two radiations cannot be observed directly because the appropriate fossils have not been discovered. Data from micropaleontology indicate that eukaryotic cells had arisen by approximately 1400 million years ago (*36*). The first unequivocal metazoan fossils, of the late Precambrian Ediacaran fauna, date from about 600 to 700 million years ago (*37*). These animals were all soft-bodied forms, but of diverse body plan, including possible cnidarians and polychaete annelids as well as animals with body plans that are no longer extant (*37*). The radiation of animals with readily fossilizable skeletons occurred over the next 100 million years (*38*), with recognizable representatives of most living phyla and classes of invertebrates appearing at least 450 to 500 million years ago.

The coding sequences of several proteins have been used to estimate the timing of the separation of animals from other eukaryotes and the separation of coelomate animal phyla (*39*). This estimation requires extrapolation of evolutionary rates from these molecules, a procedure of arguable validity (*40*); nevertheless, the time estimates are in reasonable accord with the paleontological data. Thus, fungi and animals are estimated to have diverged 1100 to 1200 million years ago, and coelomate animal phyla between 700 and 900 million years ago (*39*). A date of 550 million years ago for coelomate animal divergence, drawn from 5S rRNA data, is clearly an underestimate (*41*).

Conclusions and Perspectives

For the past century, the use of detailed descriptions of animal adult morphology and embryology has been at the heart of the study of evolutionary relationships among distant groups such as phyla. However, the methodology can have both implicit problems and practical difficulties. Phylogenetic relationships are established by grouping together animals that share derived, divergent features—innovations that should indicate a common ancestry. Often, however, because it is difficult to determine the direction of evolutionary change, it is not known which characters are the derived ones, or whether differences represent an early divergence or a later one. Furthermore, convergent evolution has occurred, obscuring both the establishment of homology and the determination of the direction of evolution. An independent method of inferring phylogeny allows models based on assumptions about homology and primitive and derived traits to be tested.

For example, one model of animal relationships is based on the idea that the gastral pouches of coelenterates are homologous with the gastral pouches (enterocoels) that give rise to the coelom in deuterostomes (*42*). The condition in larval echinoderms and hemichordates, with pairs of coelomic vesicles (and thus mesoderm) arising from gastral pouches (enterocoely), then serves as a kind of ancestral model for the deuterostomes, and for the lophophorates as well, yielding a group sometimes called the Oligomeria because its members have just a few segments. Features of deuterostome development are therefore assumed to be primitive among Bilateria in general. According to this model, conditions in protostomes are derived, perhaps with flatworms having lost the coelom. Hence, the hypothetical ancestor of the bilateral metazoa would be an "archaecoelomate" (*43*).

However, the 18S rRNA data suggest that the Cnidaria are not members of the lineage leading to the rest of the animals. Thus there is no reason for considering enterocoely or indeterminate development to be primitive in the Bilateria. Platyhelminthes becomes the outgroup of all eucoelomates, and the canon of parsimony implies

that the primitive condition was spiral, determinate cleavage. The enterocoely of brachiopods can be treated as a derived condition. The oligomery of both deuterostomes and brachiopods need not be interpreted as homologous; both groups may have arisen from segmented ancestors (*44*).

The inference of phylogenies from molecular sequence data has its own limitations. For example, there are no simple measures of reliability for the position of given branch points. The distance method used here allows the estimation of a statistical error, in standard deviations, for each pairwise calculation of evolutionary distance. However, the conversion of these pairwise errors to an overall error for branch points is complex, even for simple trees. Nevertheless, the validity of the position of a sequence within a given tree compared to other possible trees may be assessed by the "robustness" of its placement in a given branching order. The position of a given sequence is considered to be robust if it is relatively insensitive to the choice of nucleotides used in the analyses and is independent of the organisms represented in the trees.

Using the criterion of robustness, we cannot resolve the relative branching order of the four coelomate groups. In addition, within the chordate cluster, the tunicate *Styela* represents the deepest branch; depending on the subset of organisms included, in some trees the tunicate falls outside the chordate lineage. We interpret this to indicate that the affinity of the tunicate with chordates is distant. Finally, the relationships among arthropods are a problem. One of the arthropod sequences, the horseshoe crab *Limulus*, is "slow clock" (has accumulated relatively few sequence changes since its divergence from a common ancestor), whereas the three other sequences are "fast clock." Including "fast clock" species in a phylogeny introduces systematic error: "fast clock" species often appear to branch more deeply in trees than they should (*45*). *Limulus* and the other three arthropods usually cluster when all are included in the tree. In some phylogenies, however, *Limulus* appears loosely associated with the eucoelomate protostomes, but the three mandibulate arthropod sequences form a cluster closer to the root of the tree. The possibilities remain that the true position of the arthropod cluster is the deepest branch within the protostome cluster, the position taken by *Limulus* when the three "fast clock" arthropods are left out of the tree; or arthropods may be polyphyletic, with chelicerates allied with protostomes.

The categorical rank of a taxon does not necessarily correspond to its time of divergence or evolutionary distance. The evolutionary distances among classes of echinoderms are less than those among chordate subphyla (Fig. 2). These results suggest that the classes of echinoderms represented in our 18*S* rRNA sequence data split from one another considerably after the initial divergence of echinoderms, whereas subphyla of chordates diverged relatively early in the evolution of the phylum. Among eucoelomate protostomes (Fig. 5), phyla and classes exhibit indistinguishable depths of splitting. It is not clear whether the intermixing of phyla in this group is due to polyphyly or simply reflects the limited sampling provided by the current data.

Many questions remain, including the branching order of the eucoelomate lineages, the relationships within the arthropod and eucoelomate protostome groups, and the origins of pseudocoelomate groups. Solving these problems will require more extensive data sets (such as from 28*S* rRNA) and development of techniques to analyze data from regions of rapid sequence evolution.

The 18*S* rRNA trees are internally consistent, and in groups such as the chordates, which have a relatively well understood phylogenetic history, our results are congruent. Several relationships have been established by our study, and these are robust by the criteria discussed earlier. We have provided evidence for the polyphyletic origins of Bilateria and Cnidaria, the early splitting of the flatworm and coelomate lines, and the rapid radiation of the major coelomate groups. Furthermore, we have tested some specific phylogenetic models and have demonstrated that mollusks are eucoelomate animals, that deuterostomy is not likely to have been primitive, and that at least some lophophorates (brachiopods) are protostomes.

Our molecular approach to distant phylogenetic relationships in animals does not displace the study of morphology and embryology in evolution, but it should allow a better understanding of the course of evolutionary change and its underlying mechanisms.

REFERENCES AND NOTES

1. J. M. Valentine, in *Patterns of Evolution, as Illustrated by the Fossil Record*, A. Hallam, Ed. (Elsevier, Amsterdam, 1977), pp. 27–57.
2. M. R. House, Ed., *The Origin of Major Invertebrate Groups* (Academic Press, New York, 1979).
2a. R. B. Clark, in *ibid.*, pp. 55–101.
3. M. T. Ghiselin, in *Evolution, Brain, and Behavior: Persistent Problems*, R. B. Masterton, W. Hodos, H. Jerison, Eds. (Erlbaum, Hillsdale, NJ, 1976), pp. 129–142; V. L. Roth, *Biol. J. Linn. Soc.* **22**, 13 (1984); in *Ontogeny and Systematics*, C. J. Humphries, Ed. (Columbia Univ. Press, New York, in press).
4. R. A. Raff and T. C. Kaufman, *Embryos, Genes, and Evolution* (Macmillan, New York, 1983); P. Alberch, *Syst. Zool.* **34**, 46 (1985).
5. L. H. Hyman, *The Invertebrates: Platyhelminthes and Rhynchocoela* (McGraw-Hill, New York, 1951), vol. 2.
6. D. T. Anderson, *Embryology and Phylogeny in Annelids and Arthropods* (Pergamon, Oxford, 1973); S. M. Manton, *The Arthropoda: Habits, Functional Morphology and Evolution* (Oxford Univ. Press, Oxford, 1977); J. L. Cisne, *Science* **186**, 13 (1974).
7. R. E. Snodgrass, *Smithson. Misc. Collect.* **97**, 1 (1938); A. G. Sharov, *Basic Arthropodan Stock* (Pergamon, New York, 1966); M. Locke and P. Huie, *Nature (London)* **270**, 341 (1977); H. B. Boudreaux, in *Arthropod Phylogeny*, A. P. Gupta, Ed. (Van Nostrand–Reinhold, New York, 1979), pp. 551–586; C. P. Mangum, J. L. Scott, R. E. L. Black, K. I. Miller, K. E. Van Holde, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3721 (1985); P. Weygolt, *Z. Zool. Syst. Evolutionsforsch.* **24**, 19 (1986).
8. M. Kume and K. Dan, *Invertebrate Zoology* (Nolit, Belgrade, Yugoslavia, 1968); O. M. Ivanova-Kazas, *Zool. Zh.* **65**, 757 (1986); *ibid.*, p. 1445.
9. E. Zuckerkandl and L. Pauling, *J. Theor. Biol.* **8**, 357 (1965).
10. S. J. Sogin, M. L. Sogin, C. R. Woese, *J. Mol. Evol.* **1**, 173 (1972); E. Stackebrandt and C. R. Woese, in *Molecular and Cellular Aspects of Microbial Evolution*, M. J. Carlile, J. R. Collins, B. E. B. Mosely, Eds. (Cambridge Univ. Press, Cambridge, 1981), pp. 1–31.
11. C. R. Woese, *Microbiol. Rev.* **51**, 221 (1987).
12. D. J. Lane *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 6955 (1985).
13. RNA was sequenced with reverse transcriptase (D. J. Lane, K. G. Field, G. J. Olsen, N. R. Pace, *Methods Enzymol.*, in press) by a modification of the base-specific dideoxynucleotide chain termination method [F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977)]. Five oligodeoxynucleotide primers (TCAGGCTCCCTCTCCGG, GWATTACCGCGGGKCTG, TTGGCAAATGCTTTCGC, ATTCCTTTRAGTTTC, and ACGGGCGGTGTGTRC; W = A or T, K = G or T, R = A or G) yielded 1000 to 1400 nucleotides of data from each 18*S* rRNA.
14. G. J. Olsen, *Methods Enzymol.*, in press.
15. H. J. Elwood. G. J. Olsen, M. L. Sogin, *Mol. Biol. Evol.* **2**, 399 (1985).
16. W. M. Fitch and E. Margoliash, *Science* **155**, 279 (1967). A distance method is relatively robust to known sources of systematic and random errors. An alternative method based on the parsimony criterion requires that the total amount of sequence change be small [J. Felsenstein, *Annu. Rev. Ecol. Syst.* **14**, 313 (1983)]; if not, a systematic error can result from the accumulation of parallel changes in the lineages that undergo the greatest total amount of change. In our distance method (*14*), the conversion of observed sequence differences to evolutionary distance estimates (*17*) partially corrects for the occurrence of parallel changes, although there can still be difficulties with the inferred branching orders of lineages with disparate amounts of sequence change. Also, in the analysis of amino acid sequence data there is evidence that the parsimony criterion is more sensitive to statistical errors than are distance methods [R. M. Schwartz and M. O. Dayhoff, *Science* **199**, 395 (1978); M. O. Dayhoff, *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **35**, 2132 (1976)]. This difference is less dramatic with nucleotide sequence data, but the trend is the same (G. Olsen, unpublished observation).
17. T. H. Jukes and C. R. Cantor, in *Mammalian Protein Metabolism*, H. N. Monro, Ed. (Academic Press, New York, 1969), pp. 21–132.
18. The inferred phylogenetic tree is the additive tree that "best fits" the table of pairwise distance estimates. The tree error for each pair of organisms is the difference between the estimated pairwise distance and the sum of lengths of the tree branches that connect the specific pair of organisms in the tree. For each pair, the error was squared and divided by the estimated variance of the distance estimate [M. Kimura and T. Ohta, *J. Mol. Evol.* **2**, 87 (1972)]; weighted, squared errors were summed for all pairs of organisms to give the total tree error. Branching order and branch lengths were chosen to minimize the total tree error.
19. M. L. Sogin, H. J. Elwood, J. H. Gunderson, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1383 (1986) [also compare figures 4 and 14 in (*11*)].
20. E. A. Robson, in *The Origins and Relationships of Lower Invertebrates*, S. Conway Morris, J. D. George, R. Gibson, H. M. Platt, Eds. (Clarendon, Oxford, 1985),

pp. 60–77.
21. T. D. Pollard, *J. Cell Biol.* **91**, 156s (1981); G. A. B. Shelton, Ed., *Electrical Conduction and Behavior in "Simple" Invertebrates* (Clarendon, Oxford, 1982); Y. Fukui and S. Yumura, *Cell Motil. Cytoskeleton* **6**, 662 (1986); H. D. Gortz, Ed., *Paramecium* (Springer-Verlag, New York, 1987).
22. A. G. Kluge, *Chordate Structure and Function* (Macmillan, New York, ed. 2, 1977); J. G. Maisey, *Cladistics* **2**, 201 (1986); H. B. Whittington, *The Burgess Shale* (Yale Univ. Press, New Haven, 1985).
23. L. H. Hyman, *The Invertebrates: Echinodermata* (McGraw-Hill, New York, 1955), vol. 4; A. B. Smith, *Palaeontology* **27**, 431 (1984); S. Smiley, in *Echinoderm Phylogeny and Evolutionary Biology*, C. R. C. Paul and A. B. Smith, Eds. (Oxford Univ. Press, Oxford, in press); R. A. Raff *et al.*, *ibid.*
24. T. H. Huxley, *Q. J. Microsc. Sci.* **15** (1875); K. Grobben, *Verh. Zool. Bot. Ges. Wein* **58**, 491 (1908); S. F. Gilbert, *Developmental Biology* (Sinauer, Sunderland, MA, 1985).
25. R. P. S. Jefferies, *Symp. Zool. Soc. London* **36**, 253 (1975); in (*2*), pp. 443–477.
26. H. B. Whittington, in (*2*), pp. 253–268.
27. A. Naef, *Ergeb. Fortschr. Zool.* **6**, 27 (1924).
28. K. G. Wingstrand, *Galathea Rep.* **16**, 7 (1985).
29. P. P. Iwanoff, *Z. Morphol. Oekol. Tiere* **10**, 62 (1928).
30. J. Vagrolgyi, *Syst. Zool.* **16**, 153 (1967); L. v. Salvini-Plawen, *ibid.* **17**, 192 (1968); C. R. Stasek, in *Chemical Zoology*, M. Florkin and B. T. Scheer, Eds. (Academic Press, New York, 1972), pp. 1–44.
31. L. H. Hyman, *The Invertebrates: Smaller Coelomate Groups* (McGraw-Hill, New York, 1959), vol. 5.
32. R. L. Zimmer, in *Living and Fossil Bryozoa*, G. P. Larwood, Ed. (Academic Press, New York, 1973), pp. 593–599.
33. M. Jones, *Bull. Biol. Soc. Wash.* **6**, 117 (1985).
34. J. T. Bonner, *The Evolution of Development* (Cambridge Univ. Press, Cambridge, 1956).
35. B. Runnegar, *J. Geol. Soc. Aust.* **29**, 395 (1982).
36. G. Vidal, *Sci. Am.* **250**, 48 (February 1984); P. Cloud, *Paleobiology* **2**, 351 (1976); H. J. Hofmann and J. Chen, *Can. J. Earth Sci.* **18**, 443 (1981); J. W. Schopf and D. Z. Oehler, *Science* **193**, 47 (1976); J. W. Schopf, *Sci. Am.* **239**, 110 (September 1978); M. R. Walter, J. H. Oehler, D. Z. Oehler, *J. Paleontol.* **50**, 872 (1976).
37. M. F. Glaesner, *The Dawn of Animal Life* (Cambridge Univ. Press, Cambridge, 1984); R. A. Raff and E. C. Raff, *Nature (London)* **228**, 1003 (1970); B. Runnegar, *Alcheringa* **6**, 223 (1982); A. Seilacher, in *Patterns of Change in Earth Evolution*, H. D. Holland and A. F. Trendall, Eds. (Springer-Verlag, Berlin, 1984), pp. 159–168.
38. M. D. Brasier, in (*2*), pp. 103–159.
39. R. E. Dickerson, *J. Mol. Evol.* **1**, 26 (1971); B. Runnegar, *Palaeontology* **29**, 1 (1985); *Lethaia* **15**, 199 (1982); *J. Mol. Evol.* **22**, 141 (1985).
40. R. J. Britten, *Science* **231**, 1393 (1986); M. Goodman, M. L. Weiss, J. Czelusniak, *Syst. Zool.* **31**, 376 (1982); W.-H. Li and M. Tanimura, *Nature (London)* **326**, 93 (1987).
41. T. Ohama, H. Hori, S. Osawa, *Nucleic Acids Res.* **11**, 5181 (1983).
42. A. Sedgwick, *Q. J. Microsc. Sci.* **24**, 43 (1884).
43. R. Siewing, *Zool. Jahrb. Abt. Anat. Ontog. Tiere* **103**, 439 (1980).
44. L. v. Salvini-Plawen, *Zool. Scr.* **11**, 77 (1982).
45. G. J. Olsen, *Cold Spring Harbor Symp. Quant. Biol.*, in press.
46. R. M. Torczynski, M. Fuke, A. P. Bollon, *DNA* **4**, 283 (1985).
47. L. Nelles, B.-L. Fang, G. Volckhaert, A. Vandenberghe, R. De Wachter, *Nucleic Acids Res.* **12**, 8749 (1984).
48. P. M. Rubtsov *et al.*, *ibid.* **8**, 5779 (1980).
49. J. Messing, J. Carlson, G. Hagen, I. Rubenstein, A. Oleson, *DNA* **3**, 31 (1984).
50. R. McCarroll, G. J. Olsen, Y. D. Stahl, C. R. Woese, M. L. Sogin, *Biochemistry* **22**, 5858 (1983).
51. M. Salim and B. E. H. Maden, *Nature (London)* **291**, 205 (1981).
52. Voucher specimens, where available, have been deposited at California Academy of Sciences, Golden Gate Park, San Francisco, CA. Copies of the sequences and alignments are available on written request. We acknowledge gifts of RNA or animals from R. Anderson, W. Jeffery, J. Ruderman, L. Slobodkin, and J. Valois. We thank B. Parr and J. M. Turbeville for comments on the manuscript. Supported by NSF grants BSR 85-16582 (R.A.R., N.R.P., M.T.G., and E.C.R.) and DCB 83-02149 (E.C.R.); NIH grants GM34527 (N.R.P.), HD21337 (R.A.R.), and HD16739 (E.C.R.); Office of Naval Research grant N14-86-K-0268 (G.J.O. and N.R.P.); and a MacArthur Prize Fellowship (M.T.G.).

Research Articles

Glycosyl-Phosphatidylinositol Moiety That Anchors *Trypanosoma brucei* Variant Surface Glycoprotein to the Membrane

MICHAEL A. J. FERGUSON, STEVE W. HOMANS, RAYMOND A. DWEK, THOMAS W. RADEMACHER

Two forms of protein-membrane anchor have been described for the externally disposed glycoproteins of eukaryotic plasma membranes; namely, the hydrophobic transmembrane polypeptide and the complex glycosylphosphatidylinositol (G-PI) moiety. The chemical structures of the major species of G-PI anchors found on a single variant surface glycoprotein (VSG) of the parasitic protozoan *Trypanosoma brucei* were determined by a combination of nuclear magnetic resonance spectroscopy, mass spectrometry, chemical modification, and exoglycosidase digestions. The G-PI anchor was found to be heterogeneous with respect to monosaccharide sequence, and several novel glycosidic linkages were present. The results are pertinent to the mechanism of the biosynthesis of G-PI anchors.

THE PARASITIC PROTOZOAN *Trypanosoma brucei* HAS A CONtinuous cell-surface coat made up of a tightly packed monolayer of variant surface glycoprotein (VSG) molecules. This VSG coat acts as a macromolecular diffusion barrier protecting the parasite from lytic host-serum components. A single trypanosome expresses only one type of VSG (variant) at a time, but has several hundred VSG genes encoding immunologically distinct VSG variants. It is the sequential expression of different VSG coats that allows the parasite to evade the host's immune response by antigenic variation (*1*). All of the different VSG variants analyzed have molecular sizes of about 55 kD and one or more asparagine glycosylation sites. Despite the lack of extensive primary amino acid sequence homology the VSG molecules are thought to share similar tertiary structures (*2*).

The authors are members of the Oxford Oligosaccharide Group, Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, England.