© 1989 Elsevier Science Publishers B.V. (Biomedical Division)
The Hierarchy of Life
B. Fernholm, K. Bremer and H. Jörnvall, editors

CHAPTER 26

Vertebrate phylogeny: evidence from 28S ribosomal DNA sequences

DAVID M. HILLIS and MICHAEL T. DIXON

Department of Zoology, The University of Texas at Austin, Austin, TX 78712, U.S.A.

Introduction

The higher relationships of vertebrates have been more extensively studied than have the relationships of any other comparable group of organisms. Vertebrate phylogeny has been the bastion of comparative anatomy; in fact, many people equate the field of comparative anatomy with *vertebrate* (or at least chordate) comparative anatomy. Although several molecular data sets have been collected on the higher relationships of vertebrates [e.g., 3,8,11-14], these studies have almost uniformly treated the morphological status quo as a yardstick against which the value of the molecular data set under study could be measured. Given the minuscule size of the total molecular data set for vertebrates (collected over the past few decades for proteins and over the past few years for nucleic acids) when compared to the breadth of the morphological data set (collected over the past few centuries), it is clear that the phylogeny of vertebrates as deduced from morphology will continue to be the standard to which molecular studies aspire, at least for the near future. However, this does not mean that molecular studies cannot make immediate contributions to our knowledge of higher vertebrate relationships. A number of problems of vertebrate relationships have proven recalcitrant to morphological systematists, despite extensive study. For instance, the relationship of the coelacanths to the rest of the vertebrates is extensively debated by morphologists [13,22,29,30,33,39,41], and the relationship of the birds to the other amniotes has recently become controversial [9, 10]. It is likely that molecular data sets can immediately add new insights to these debates.

The long-term outlook for molecular resolutions of vertebrate relationships is even brighter. Although there will always be limitations on the molecular data set that do not apply to the morphological data set (the greatest being the inaccessibility of fossils to the vast majority of molecular methods), molecular techniques also have clear advantages [16]. The most important advantage of the molecular data set is simply one

of size: all heritable information is potentially accessible to molecular techniques, whereas only a small fraction of this information can be garnered from the study of morphology. In this sense, morphological information is a subset of molecular information. However, we do not suggest that molecular techniques give ultimate answers. If one considers the taxonomic breadth of the data sets, the taxa that can be studied by molecular methods are a subset of the taxa that can be studied by morphology. The importance of this latter point is demonstrated by the differences that are obtained if one excludes fossils from objective morphological studies of vertebrate phylogeny [10]: the inferred relationships of the extant taxa can shift markedly. Both morphological and molecular techniques have a number of other advantages and disadvantages [16,25], so that a combined approach of molecules plus morphology is not only desirable but necessary for a comprehensive view of phylogeny.

The preceding discussion suggests that molecular studies of phylogeny are in their infancy and cannot yet challenge most of the framework of higher vertebrate phylogeny. However, progress from molecular studies on reconstructing the relationships of more closely related groups has been remarkable. This is undoubtedly a function of the high variability of most of the molecules studied to date (particularly proteins and mitochondrial DNA). Thus, for many groups of morphologically similar organisms, our only knowledge of phylogenetic relationships comes from molecular data. Even large groups of morphologically cryptic species often yield their phylogenetic history to molecular studies [20,24]. But until recently, evolutionarily conserved genes have received far less attention than have evolutionarily variable regions of the genome. Although studies of amino acid sequences of several relatively conserved proteins have been applied to problems in higher phylogeny for more than two decades [e.g., 4,8,11–14], only within the last decade has it become technically practical to sequence the primary structure of highly conserved genes shared by all living organisms.

Ribosomal DNA

Many studies of higher phylogeny have recently been conducted on the ribosomal RNA genes and spacer regions (rDNA) [7,17,18,21,34,42]. There are several reasons for this emphasis. Perhaps the most important reason is the great diversity of rates of evolution within the rDNA arrays. The three rRNA genes in the array (in vertebrates, the 18S, 5.8S, and 28S genes) contain the most highly conserved sequences that occur in living organisms, and portions of these genes have been used to reconstruct the phylogeny back to the origin of life nearly four billion years ago [43]. The genes contain regions that show greater variation as well; these regions (called divergent domains or expansion segments) often show both length and sequene differences among relatively closely related taxa that have diverged for only a few hundred million years [5,15,17,18]. The three rRNA genes are separated by transcribed spacers (Fig. 1) that evolve at a considerably more rapid rate than the genes [2,17]. Finally, the adjacent arrays are separated by non-transcribed spacers (Fig. 1) that evolve even more rapidly [2,17]. The transcribed and non-transcribed spacer sequences have been

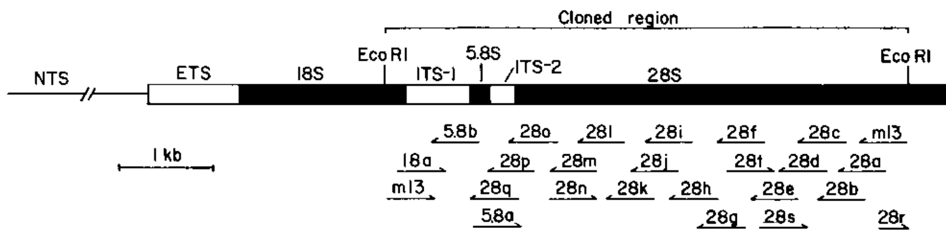

Fig. 1. The ribosomal repeat unit of vertebrates and the cloned *Eco*RI–*Eco*RI fragment sequenced in this study. Primers are indicated below the cloned fragment; arrows indicate direction of sequencing. Abbreviations: ETS: external transcribed spacer; ITS: internal transcribed spacer; NTS: non-transcribed spacer.

used to reconstruct the phylogeny of taxa that have not shared a common ancestor in 1–50 million years [17,35]. Therefore, the various regions of the rDNA arrays can be used to reconstruct phylogeny over virtually the entire history of life.

Several other reasons make rDNA arrays amenable and desirable to study. The rDNA array consists of dozens to hundreds of copies in most eukaryotes, so probing by Southern blotting, amplification, and cloning procedures are all facilitated, and the highly conserved nature of the ribosomal genes makes possible the use of cloned probes from relatively distantly related species [17]. The mature rRNAs account for a large percentage of cellular RNA, and are present in high enough copy number that they can be sequenced directly [21]. The numerous copies of rDNA evolve in a concerted fashion [6], so intraspecific variation in sequence is low; this enables the use of small sample sizes in phylogenetic studies [17,19]. These reasons combine to produce a molecular probe of evolution that can be applied relatively easily to a large diversity of problems.

Direct sequencing versus cloning

Much of the sequence information that has been collected from rDNA arrays has been produced by directly sequencing rRNA using reverse transcriptase [21]. With the development of new methods in DNA amplification (especially the use of a thermostable DNA polymerase in the polymerase chain reaction, or PCR, technique [31]), many laboratories are beginning to amplify and sequence rDNA directly. Because we have deliberately chosen to clone and sequence rDNA rather than use the more common direct sequencing methods, some comment on this point is necessary. The direct methods have the distinct advantage of very rapid collection of data, which is why they are widely used. However, there are also some disadvantages. Compared to cloning or PCR amplification, direct RNA sequencing has the disadvantages of providing access only to the transcribed portions of the array, of being limited to the single transcribed strand (so two-strand verification is not possible), and of being limited to relatively high quality tissue samples. PCR amplification eliminates these problems, but requires highly conserved flanking regions so that large variable regions are inac-

cessible. Because our interests concern all parts of the rDNA array, cloning provides the only sure means of obtaining the relevant information. Direct RNA sequencing involves higher error rates than does sequencing of cloned DNA (as a function of both the error rate of reverse transcriptase and of single-strand sequencing). Finally, the primary reason we chose to clone and sequence was repeatability and extension of the work. Our clones and gene libraries are available in essentially inexhaustable supply for other laboratories to use. Therefore, our rDNA clones can be used as probes for further studies or resequenced to verify our results, and our libraries can be screened for any other genes of interest. We have prepared gene libraries of most major groups of vertebrates, and plan to prepare such libraries from at least one species in each family of vertebrates that becomes accessible to us. As this gene library collection progresses, it will be possible to screen for virtually any gene from a species in the same family as any vertebrate of interest.

Materials and methods

DNA from *Latimeria chalumnae* (Actinistia), *Rhineura floridana* (Squamata), and *Cacatua alba* (Aves) was extracted and restricted with *Eco*RI according to standard methods [23]. Cleaved DNA was used to construct *Eco*RI partial libraries in bacteriophage lambda ZAP II (Stratagene Cloning Systems). Approximately 50,000–250,000 plaques were screened by hybridizing nylon filter lifts with the cloned 28S rDNA gene of *Rana catesbeiana* (pE2528) described by Hillis and Davis [18]. Positive plaques were isolated and purified, and the cloned inserts were subcloned into pBluescript (Stratagene). In addition to the species listed above, an orthologous clone of *Notropis lutrensis* (Actinopterygii) was provided by John Gold and Scott Davis (in pUC19). The resulting plasmids, containing an approximately 5–6 kb insert with a portion of the 18S gene, the internal transcribed spacers, all of the 5.8S gene, and most of the 28S gene (Fig. 1), were verified by restriction endonuclease digestion. Plasmid DNA was isolated by cesium chloride centrifugation [23] and sequenced by the base-specific dideoxynucleotide chain termination method [32] using the modified bacteriophage T7 DNA polymerase described by Tabor and Richardson [36] and 55 cm field gradient gels [1] of 4–6% acrylamide. Primers are given in Table 1 and sequencing strategy is shown in Fig. 1.

Sequences were aligned using the alignment subroutines of the IBI/Pustell sequence analysis software described by Pustell and Kafatos [26–28], with adjustments made manually to increase similarity. Two previously sequenced vertebrate 28S genes (*Mus musculus*, Mammalia [15], and *Xenopus laevis*, Amphibia [40]) were also aligned with the new sequences, as was the 28S rDNA sequence of *Drosophila melanogaster* (Insecta [37]), which was used as the outgroup. Regions with unambiguous alignments were analyzed for phylogenetic information using the software package PAUP (D. Swofford, University of Illinois). Analyses were performed on both unweighted and

TABLE 1

Primers used to sequence the 18S–28S rDNA *Eco*RI-*Eco*RI cloned region (see Fig. 1). Primers marked with an asterisk were used in this study. Positions of the primers in the *Mus* sequence correspond to the positions of the RNA nucleotides in the mature transcripts [15]. 'S' stands for the strand synonymous to RNA; 'C' stands for the complementary strand.

Primer	*Mus* position	Strand	Sequence
5.8a	3–17	C	5′-CTCTTAGCGGTGGAT-3′
5.8b	50–64	S	5′-AATTCTCGCAGCTAG-3′
18a	1844–1858	C	5′-CGTAGGTGAACCTGC-3′
28a*	3913–3927	S	5′-CCTTCTGCTCCACGG-3′
28b*	3695–3709	S	5′-AGAGTAGTGGTATTT-3′
28c*	3481–3495	S	5′-ACAGTGGGAATCTCG-3′
28d*	3284–3298	S	5′-TTAAACAGTCGGATT-3′
28e*	2976–2990	S	5′-GTCCAGAGTCGCCGC-3′
28f	2617–2631	S	5′-TCCCGAAGTTACGGA-3′
28g*	2386–2400	S	5′-CTGCCCTTCACAAAG-3′
28h*	2101–2115	S	5′-CTACCACCAAGATCT-3′
28i*	1840–1854	S	5′-GCGCCATCCATTTTC-3′
28j*	1665–1679	S	5′-CCAGTTCTGCTTACC-3′
28k*	1402–1416	S	5′-CGATTTGCACGTCAG-3′
28l	1131–1145	S	5′-GGTCCGTGTTTCAAG-3′
28m	824–838	S	5′-CGGCGAGTGCTGCTG-3′
28n	381–395	C	5′-TGAAAAGAACTTTGA-3′
28o	426–440	S	5′-ACCCGTTTACCTCTT-3′
28p	197–211	S	5′-CGATCAGAAGGACTT-3′
28q	4–18	S	5′-GTCTGATCTGAGGTC-3′
28r	3828–3842	C	5′-CAGGTGGGGAGTTTG-3′
28s	2559–2573	C	5′-AGGTGAACAGCCTCT-3′
28t	2338–2352	C	5′-ACCGATCCCGGAGAA-3′
m13 (forward)*	NA	S/C	5′-GTTTTCCCAGTCACGAC-3′
m13 (reverse)*	NA	S/C	5′-CAGGAAACAGCTATGAC-3′

weighted character sets. For the latter, transversions were weighted as two steps, whereas transitions, insertions, and deletions were weighted as one step. Analyses were also conducted will all insertion/deletion characters removed for both weighted and unweighted character sets. All possible trees were examined using the ALLTREES command of PAUP, and strict and majority consensus trees were constructed from the shortest cladograms. Previously proposed hypotheses were tested against the most parsimonious cladograms using the Wilcoxon matched-pairs signed-ranks test described by Templeton [38], under the assumption that transversions are less likely to occur than are transitions.

```
            1111111111 1111111111 1111111111 1111111111 1111111111 1111111111 1111111111
            1111111111 1111111111 1111122222 2222222222 2222222333 3333333333 3333333333
            5555666677 7777778888 8999903344 4444666667 8999999111 1222222255 5566677889
            6789013601 2367893457 9023460245 6788012347 7234799237 9034578913 6723829380
                                                +             +       +

Mus         TGCGCAGCGT CCGAGCCTTC GCGCACGTCT CCG+CGGCTC CGGCGCOTGG ATGOATCCCA CCCGGTACCA
Rhineura    TGTGCGGCGT CTGAGCCTTC GOGCACGTCT CCG+CGCCTC CGGOACOCAG GCAGOTCCCG CCCGGTACCA
Cacatua     CGTGCAGCGT CCGOGCCTTC OOGCGCATCC CCG+CGGGGC CACGGG+CAG AGAGOTCCCG CCCG--ACCA
Xenopus     TGCGCGGCGC CTCTGACTTG OOACACATCG GCG+GGCCGC CGGGACOCAG GCGTATCCCG CCCGGTACCA
Latimeria   TGCGCAGCGT CTCGTCCTTC GOGTATATGT GCG+CGCCCT TGTCACOTAG ACATOTCCCG CCCGGTACCA
Notropis    TCGGTGGCGT CTCAOCTGCC GGGTGCGOGA GGCOCTCCCC TCAGGCOOOT GCGTOCCACG CCCGGTACCO
Drosophila  TGTATAATAT AACTAT---- ---------- ---------- ---------- -----CACTA TAAAACGTTT


  1111111111 1111111111 1111111111 1111111111 1111111111 1111111111 1111111111 1111111111
  3333444444 4444445555 5555555555 5555666666 6666666666 6666666666 6666677777 7777777777
  9999011222 2233470000 2222334556 6779000001 1111122222 3334444445 5899900000 0111112222
  1468179356 7803852356 1257138593 8154567890 1256912347 1226667990 9501612348 9123670123
          +                        ++                      + ++  +

M GACCGCGGOG GACTAGGCAG G+CAAGAACG OOGTTTCGGG CCGCCGCACC CCOAOOCGOA TGACCGGCCT TCGCGCCGGC
R GACCGOGGOG GACTAGGCAG G+CAAAAACG OOGTTTGGCG CCGCCGCACC CCOGOGCGOA TGACCGGCCT TCGCGCCGGC
C TACCGCGGOG GGCGAGGTAG C+CAGGAACG OOGGCCCCGG CCGCCGCACC GCGAOOCAOG TGACCGGCCT TCGCGCCGGC
X GACCGCGGOG GACTCGGCA- G+CAAGAACG ATGTTTCGGG CCGCCGAACC CCCGTGCGTA TGACCGGCCT TCGCGCCGGC
L -ACCGCGGTG GACTAGACAG G+CAAAGACG OOGTTTCGGG CCGCCGAACC CCOAOOCGOA TGACCGGCCC ACGCGCCGGC
N GGCCGCGGOG GACTAGGCAG G+CAAGGGCG OOGTTTCGGG CCGCCGAACC CCOTOCGGOG CGACCGGCCT TCGCGCCGGC
D GTTTATATTO AATTGAACTA OOTGAAAGTA OOATTTGGAA TTAAATATAO OOOOOOOGOG TACATATTAT GACGTTATTG


  1111111111 1111111111 1111111111 1111111111 1122222222 2222222222 2222222222 2222222222
  7777777777 7777777788 8888888888 8888888888 8801111122 2222222222 2223333333 3333333333
  2233344455 6677889233 3345566666 6677777888 8890333900 0111233455 7990111222 2223333334
  5616712879 0278180856 7806700345 6934567123 4552167525 9569425946 4288015125 6781256790
    +                         +                   + ++      +             +

M GCOGGCTATT ACCGAGCACG GGCCCGOGCC CGGGGCCAGC GGTCOTOOCT CCAOGGCAAC GACAC+OCGC ACGATCGCGG
R GCOGGCTATT ACCGAACACG GGCCCGOGCC CGCGGCCAGC GGTAOTOOTT TCGOGGCAAC GACGC++AGC ACGATCGCGG
C GCOCGCTATG ACCGAGCACG GGCCCGOGCC CGCGGCCAGC GG-----OTG TCGOCGCAAT GACGG++GGC ACGGCCGTCG
X GCOGGCTATT ACCGAACACG GGCCCGOGCC CGGGGCCAGC GGCCOTOOTT CCGOCGCGAC GGCGC++AGC ACGGCCGCGG
L GCOGGCTATT ACCGAGCACG GGCCCGOGCC CGGGGCCAGC GCCCOTOOTT CTGOGGCGAT AATCC++AGC ACGGTCGTGO
N GCOGGCTATT ACCGAGCACG GGCCCGOGCC CGGGGCCAGC TTCCCCCTTO CGGTCOGGGC GGCGOO+GGC ACGATCGTTG
D TOCATACGGA TTTATGTCGC CCAATAATAT OAGTAATGAT TT-------- ---------- -----O+GAA GGAGTTCCTG


  2222222233 3333333333 3333333333 3333333333 3333333333 3333333333 333333333
  3333339900 0000000000 0022222222 2233333333 3333444455 5555555555 555556666
  4444459900 0111233333 3355556677 7711122222 3489007900 1444456889 999990001
  3567852407 8069501256 7867890212 8946925679 9190078014 0234898490 123684697
                               +                   +          +

M AGGCCGGTCG GACACCCCAG CGGGCTGOAC GTGGTGCGCG CCACTOCGGT GGCCGOCCCG TGCCTTCTT
R AGGCCGGTCG AACGCGCAGG GCGACTGOAC GTGGTGCGCA CCACTOCGGT AGTCGCCCCG GGACTTCTT
C A--------- ---------- --GGCCGOAC GTGGTGCGCG CCACTGCGGT GGCCGOCCCG CGACTTCTT
X AGGCCGGGCG AACGCCGCGA GAGACTGOAC -TGGTGCTCG CCACTOCGGT AGCCGOCCGT TGACTTCTC
L AGGCCGGTTG AACACCGCGC CCGACTGOAC GTGGTGCGCG CCACTO---- ---------- -----OCTC
N TGACCAATCA GGACTAGCAG CCGGCCGOAC GTGGTGCGCG CCACTOCGGC AGTCGOCCCG TGACTTCCT
D GTATTG---- ---------- --TAGTTGGT AAAAACATOG TTCAGOAAAT ATTAAOATCG TCATCTTTT
```

Fig. 2. Data matrix of variable 28S rDNA sites in vertebrates and *Drosophila* (outgroup). The numbers refer to the nucleotide positions in the mature RNA transcript of *Mus* [15]. A '+' indicates an insertion of one or more specific bases after the corresponding *Mus* positions; an 'O' indicates that the insertion does not occur in the species. Dashes indicate missing data for vertebrates or non-aligned areas of the *Drosophila* sequence.

Results

Sequences were obtained for the 28S rDNA genes of each of the study species from the position equivalent to *Mus* 1117 to the *Eco*RI site at *Mus* 3628 [15], except for a 400 bp region in divergent domain 8 [18]. Sequences were not aligned in divergent domains 4–10, because of considerable sequence differences among the species. It was also not possible to align the *Drosophila* sequence to the vertebrate sequences over the region from *Mus* 1183–1325, 2056–2310, and 2970–3255, so *Drosophila* was scored as missing data for these regions. We located 299 substitutions, deletions, and insertions among the aligned sequences (Fig. 2). Identical insertions of several adjacent base pairs were treated as a single character, and the insertion was scored as present or absent.

The distribution of lengths of all possible cladograms from the unweighted data set is shown in Fig. 3. Three shortest trees were obtained; the strict consensus of these three trees is shown in Fig. 4A. The three trees all placed *Latimeria* in a group with the tetrapods, clustered the amniotes in a monophyletic group, and grouped the bird (*Cacatua*) with the mammal (*Mus*) rather than with the squamate (*Rhineura*). One of these trees and the number of supporting character changes along each branch is shown in Fig. 5.

The weighted character set produced similar results. The distribution of lengths of

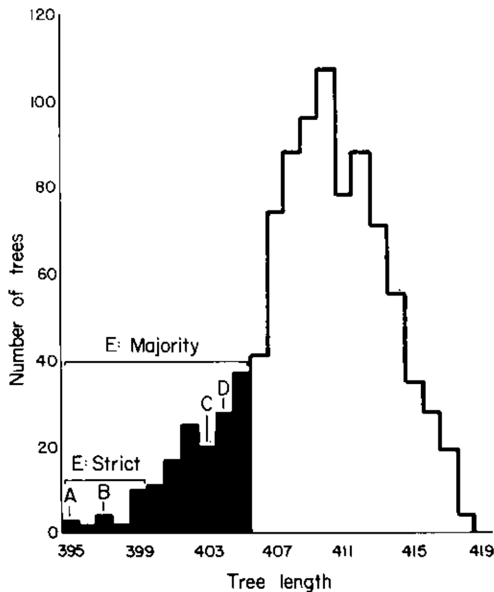

Fig. 3. Distribution of tree lengths of all possible trees in the unweighted analysis (under the constraint of outgroup rooting). Letters refer to the length class of particular trees shown in Fig. 4.

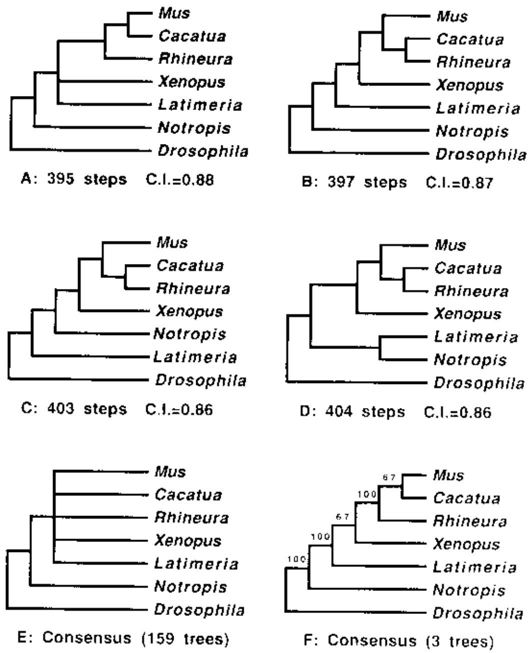

Fig. 4. Topologies, lengths, and consistency indices (C.I.) of various hypotheses of vertebrate relationships. Lengths and consistency indices refer to the unweighted analysis. (A) Consensus tree of three shortest cladograms in the unweighted analysis. (B) Hypothesis of Romer [29]. (C) Hypothesis of Løvtrup [22], von Wahlert [39], and Wiley [41]. (D) Alternate hypothesis of non-sister group relationship between *Latimeria* and tetrapods. (E) Consensus tree of unweighted and weighted analyses. This is the strict consensus tree of the shortest 21 unweighted trees and the shortest 18 weighted trees, as well as the majority consensus tree of the shortest 159 trees in both analyses. (F) Majority consensus tree of the shortest three trees in the weighted analysis. Numbers above branches indicate the percent of trees that agree with the topology.

all possible bifurcating trees for the weighted data set is shown in Fig. 6. The majority consensus tree for the three shortest trees is shown in Fig. 4F. As with the unweighted data set, *Latimeria* is placed with the tetrapods. Two of the three trees also place *Mus* with *Cacatua*; the remaining resolution places *Mus* with *Rhineura*.

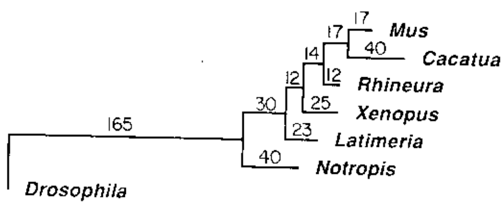

Fig. 5. One of the three shortest cladograms from the unweighted analysis. Numbers above the branches indicate the number of character changes.

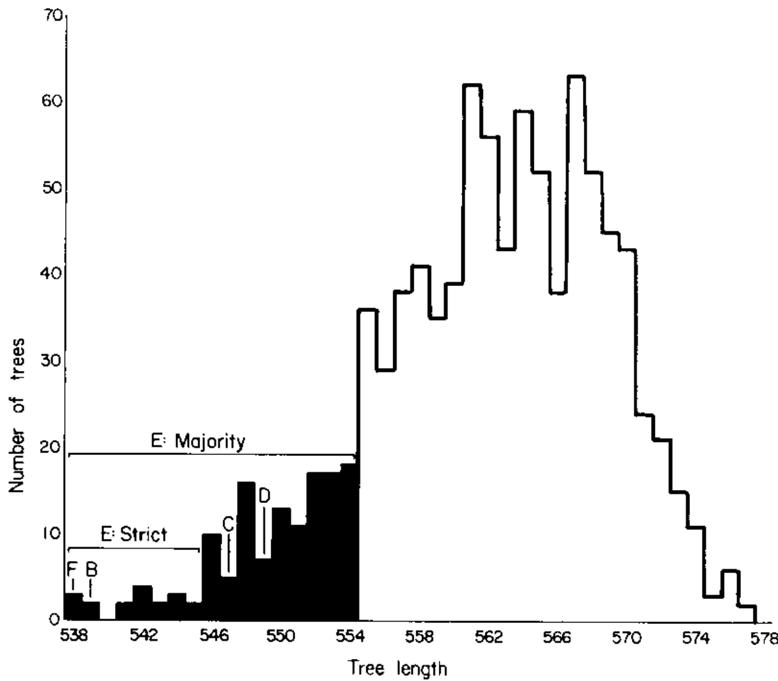

Fig. 6. Distribution of tree lengths of all possible trees in the weighted analysis (under the constraint of outgroup rooting). Letters refer to the length class of particular trees shown in Fig. 4.

The strict consensus tree of the 21 shortest trees in the unweighted analysis and of the 18 shortest trees in the weighted analysis places *Latimeria* in a group with the tetrapods to the exclusion of *Notropis* (see Figs. 3, 4E, 6). This is also the majority consensus tree if the shortest 159 trees in either analysis are included.

Removal of the insertion/deletion events from the character matrix had little effect on the results. In the unweighted analysis, removal of insertion/deletion characters resulted in the same three shortest trees as were found when these characters were included (length 319 steps, no other trees at 320 steps; consensus tree is Fig. 4A). In the weighted analysis, only two shortest trees result if the insertion/deletion events are excluded; these two trees are the two resolutions of Fig. 4A that do not place *Latimeria* and *Xenopus* as sister taxa (length 453 steps, no other trees at 454 steps).

Discussion

The relationship of the coelacanths to the rest of the vertebrates has been the subject of considerable debate among morphologists [13,22,29,30,33,39,41]. The traditional view [29] is that the coelacanths are closely related to the tetrapods (Fig. 4B). Al-

though some debate exists as to whether *Latimeria* or lungfishes are the closest living relatives of tetrapods, the close relationship of coelacanths to tetrapods is still held by many authors [30, 33]. However, many authors place the Actinistia as the sister group to tetrapods plus actinopterygian fishes, either including [41] or excluding [39] the lungfishes from the tetrapod line (Fig. 4C). This suggested relationship has been the basis for two formal classifications of the vertebrates [13,41]. Other authors completely exclude *Latimeria* from Osteichthyes (*sensu* Wiley [41]), and place the Actinistia with Chondrichthyes [22].

Until now, few molecular phylogenetic data have been collected for the only living coelacanth, *Latimeria chalumnae* (primarily because of the difficulties of obtaining tissues from this species), so the debate over the relationships of coelacanths has been limited to morphologists. However, the 28S rDNA data provide a strong statement about *Latimeria* relationships. The most robust finding of our analysis places *Latimeria* with the tetrapods (Figs. 4, 5). Not only do the most parsimonious trees of the unweighted data set suggest this relationship, all 21 trees within four steps of the shortest trees and a majority of the 159 trees within 10 steps of the shortest trees also place *Latimeria* with the tetrapods (Figs. 3, 4). With the weighted data set, all 18 trees within four steps of the shortest trees and a majority of the 159 trees within 10 steps of the most parsimonious cladograms group tetrapods and *Latimeria* together. The suggestion of Løvtrup [22], von Wahlert [39], and Wiley [41] that *Latimeria* belongs outside of Tetrapoda plus Actinopterygii requires eight additional steps in the unweighted analysis and 11 additional steps in the weighted analysis (Figs. 3, 4C, 6). The other possible placement of *Latimeria* as a non-sister group to the tetrapods, namely as a sister group to actinopterygian fishes (Fig. 4D), requires even more steps in both the unweighted (Fig. 3) and weighted (Fig. 6) analyses. A comparison of trees that differ only in the position of *Latimeria* (i.e., Fig. 4B versus 4C or 4D) using a Wilcoxon matched-pairs signed-ranks test [38] shows that the trees with *Latimeria* as the sister group to the tetrapods are significantly better descriptions of the data than the other possibilities (4B versus 4C: $n=12$, $t=16.5$, $P<0.05$; 4B versus 4D: $n=10$, $t=4.5$, $P<0.01$).

Another major controversy of the phylogeny of vertebrates concerns the relationships of birds to the rest of the amniotes. The traditional view is that in order for the Reptilia to be monophyletic, it must include the birds [29]. Recently, however, Gardiner [9] has suggested that the neontological evidence places Aves as the sister group to Mammalia. Gauthier et al. [10] expanded and reevaluated the available data set, and found that if fossil taxa were not included, birds plus crocodiles did in fact cluster with mammals. However, they then showed that the inclusion of fossil taxa resulted in the traditional view of a monophyletic Reptilia (including birds).

At present, the 28S rDNA data set provides only a very weak resolution of the amniote relationships. Although all three of the most parsimonious unweighted cladograms and two of the three most parsimonious weighted cladograms place the bird (*Cacatua*) with the mammal (*Mus*) rather than with the squamate reptile (*Rhineura*),

the traditional view (Fig. 4B) is only two steps longer in the unweighted analysis and a single step longer in the weighted analysis. These two trees (Fig. 4F versus 4B) are not significantly different in a Wilcoxon matched-pairs signed-ranks test ($n=10$, $t=22$, $P>0.05$). Therefore, no strong statement can be made about amniote relationships from these data as can be made for the relationship of *Latimeria*.

A greater resolution of the amniote relationships may be obtained from rDNA sequences in several ways. First, we only aligned the more conserved portions of the gene, so that we could examine relatively ancient events. Although sequences within the divergent domains are too dissimilar between the amniotes and anamniotes to align, portions of these regions appear to be informative about amniote relationships, and we are currently pursuing this line of investigation as well as expanding the analysis into the transcribed spacer regions (Fig. 1). Second, Gauthier et al. [10] have shown that the exclusion of taxa can have a dramatic effect on phylogenetic relationships. Although we cannot sequence the genes of fossil taxa, inclusion of additional amniote lineages in the analysis may help clarify the analysis. Finally, the most closely related outgroup sequence available for this analysis (*Drosophila*) is highly divergent from the vertebrate sequences; inclusion of invertebrate chordates as outgroup taxa may also result in a clearer picture of the relationships of vertebrates as recorded in the ribosomal genes.

Conclusions

The sequences of rRNA genes are useful for resolving some controversies of phylogenetic relationships of the major vertebrate groups. The rDNA data clearly suggest that *Latimeria* is related to the tetrapod lineage, rather than a sister group to the tetrapods plus actinopterygian fishes as suggested by several previous authors. However, over 2000 base pairs of sequence information was insufficient to unambiguously infer the relationships of several major amniote lineages. Although the rDNA data weakly support a relationship between birds and mammals as suggested by Gardiner [9], the results are not sufficiently robust to reject the traditional relationships between birds and the other reptiles. Inclusion of additional taxa, additional sequence data, and alignment within the divergent domains may extend these findings.

Acknowledgements

We thank John Gold and Scott Davis for the 28S rDNA clone of *Notropis lutrensis*. Loren Ammerman, Joseph Johnson, and William Ramos provided laboratory assistance. James Bull and Michael Ryan provided helpful input on the manuscript. Susan Clubb and E.O. Wiley provided tissues, and Ann M. and Ereckson B. Hillis assisted in data input. This work was supported by National Science Foundation (U.S.A.) Grants BSR 8657640, BSR 8614622, and BSR 8796293 to D.M.H.

References

1 Ansorge W, Labeit S. Field gradients improve resolution on DNA sequencing gels. J Biochem Biophys Methods 1984; 10: 237–243.

2 Appels R, Dvorák J. Relative rates of divergence of spacer and gene sequences within the rDNA region of species in the Triticeae: implications for the maintenance of homogeneity of a repeated gene family. Theoret Appl Genet 1982; 63: 361–365.

3 Bishop MJ, Friday AE. Tetrapod relationships: the molecular evidence. In: Patterson C (Ed.) Molecules and Morphology in Evolution: Conflict or Compromise? Cambridge: Cambridge University Press, 1987; 123–139.

4 Boulter D, Ramshaw JAM, Thompson EW, Richardson M, Brown RH. A phylogeny of higher plants based on the amino acid sequences of cytochrome *c* and its biological implications. Proc R Soc London B 1972; 181: 441–455.

5 Clark CG, Tague BW, Ware VC, Gerbi SA. *Xenopus laevis* 28S ribosomal RNA: a secondary model and its evolutionary and functional implications. Nucl Acids Res 1984; 12: 6197–6220.

6 Dover GA, Coen E. Springcleaning ribosomal DNA: a model for multigene evolution? Nature 1981; 290: 731–732.

7 Field KG, Olsen GJ, Lane DJ, et al. Molecular phylogeny of the animal kingdom. Science 1988; 239: 748–753.

8 Fitch WM, Margoliash E. Construction of phylogenetic trees. Science 1967; 155: 279–284.

9 Gardiner BG. Tetrapod classification. Zool J Linnean Soc 1982; 74: 207–232.

10 Gauthier J, Kluge AG, Rowe T. Amniote phylogeny and the importance of fossils. Cladistics 1988; 4: 105–209.

11 Goodman M. Decoding the pattern of protein evolution. Prog Biophys Mol Biol 1981; 38: 105–164.

12 Goodman M, Czelusniak J, Moore GW, Romero-Herrera AE, Matsuda G. Fitting the gene lineage into its species lineage: a parsimony strategy illustrated by cladograms constructed from globin sequences. Syst Zool 1979; 28: 132–163.

13 Goodman M, Miyamoto MM, Czelusniak J. Pattern and process in vertebrate phylogeny revealed by coevolution of molecules and morphologies. In: Patterson C (Ed.) Molecules and Morphology in Evolution: Conflict or Compromise? Cambridge: Cambridge University Press, 1987; 141–176.

14 Goodman M, Moore GW, Matsuda G. Darwinian evolution in the genealogy of haemoglobin. Nature 1975; 253: 603–608.

15 Hassouna N, Michot B, Bachellerie J-P. The complete nucleotide sequence of mouse 28S rRNA gene: implications for the process of size increase of the large subunit rRNA in higher eukaryotes. Nucl Acids Res 1984; 12: 3563–3583.

16 Hillis DM. Molecular versus morphological approaches to systematics. Annu Rev Ecol Syst 1987; 18: 23–42.

17 Hillis DM, Davis SK. Evolution of ribosomal DNA: fifty million years of recorded history in the frog genus Rana. Evolution 1986; 40: 1275–1288.

18 Hillis DM, Davis SK. Evolution of the 28S ribosomal RNA gene in anurans: regions of variability and their phylogenetic implications. Mol Biol Evol 1987; 4: 117–125.

19 Hillis DM, Davis SK. Ribosomal DNA: intraspecific polymorphism, concerted evolution, and phylogeny reconstruction. Syst Zool 1988; 37: 63–66.

20 Hillis DM, Frost JS, Wright DA. Phylogeny and biogeography of the *Rana pipiens* complex: a biochemical evaluation. Syst Zool 1983; 32: 132–143.

21 Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analysis. Proc Natl Acad Sci USA 1985; 82: 6955–6959.

22 Lovtrup S. The Phylogeny of Vertebrata. London: John Wiley and Sons, 1977.

23 Maniatis T, Fritsch EF, Sambrook J. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor: Cold Spring Harbor Laboratory, 1982.

24 Moritz C, Dowling TE, Brown WM. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu Rev Ecol Syst 1987; 18: 269–292.
25 Patterson C (Ed.) Molecules and Morphology in Evolution: Conflict or Compromise? Cambridge: Cambridge University Press, 1987.
26 Pustell J, Kafatos FC. A convenient and adaptable package of DNA sequence analysis programs. Nucl Acids Res 1982; 10: 51–59.
27 Pustell J, Kafatos FC. A convenient and adaptable package of computer programs for DNA and protein sequence management, analysis, and homology determination. Nucl Acids Res 1984; 12: 643–655.
28 Pustell J, Kafatos FC. A convenient and adaptable microcomputer environment for DNA and protein manipulation and analysis. Nucl Acids Res 1986; 14: 479–488.
29 Romer AS. Vertebrate Paleontology, 3rd edn. Chicago: University of Chicago Press, 1966.
30 Rosen DE, Forey PL, Gardiner BG, Patterson C. Lungfishes, tetrapods, paleontology, and plesiomorphy. Bull Am Mus Nat Hist 1981; 167: 159–276.
31 Saiki RK, Gelfand DH, Stoffel S, et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 1988; 239: 487–491.
32 Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 1977; 74: 5463–5467.
33 Schultze H-P. Hennig und der Ursprung der Tetrapoda. Paläont Z 1981; 55: 71–86.
34 Sogin ML, Elwood HJ, Gunderson JH. Evolutionary diversity of eukaryotic small subunit rRNA genes. Proc Natl Acad Sci USA 1986; 83: 1383–1387.
35 Sytsma KJ, Schaal BA. Phylogenetics of the *Lisianthius skinneri* (Gentianaceae) species complex in Panama utilizing DNA restriction fragment analysis. Evolution 1985; 39: 594–608.
36 Tabor S, Richardson CC. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc Natl Acad Sci USA 1987; 84: 4767–4771.
37 Tautz D, Hancock JM, Webb DA, Tautz C, Dover GA. Complete sequences of the rRNA genes of *Drosophila melanogaster*. Mol Biol Evol 1988; 5: 366–376.
38 Templeton AR. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution 1983; 37: 221–244.
39 von Wahlert G. Latimeria und die Geschichte der Wirbeltiere: eine evolutionsbiologische Untersuchung. Stuttgart: Gustav Fischer Verlag, 1968.
40 Ware VC, Tague BW, Clark CG, Gourse RL, Brand RC, Gerbi SA. Sequence analysis of 28S ribosomal DNA from the amphibian *Xenopus laevis*. Nucl Acids Res 1983; 11: 7795–7817.
41 Wiley EO. Ventral gill arch muscles and the interrelationships of gnathostomes, with a new classification of the Vertebrata. Zool J Linnean Soc 1979; 67: 149–179.
42 Woese CR. Bacterial evolution. Microbiol Rev 1987; 51: 221–271.
43 Woese CR. Macroevolution in the microscopic world. In: Patterson C (Ed.) Molecules and Morphology in Evolution: Conflict or Compromise? Cambridge: Cambridge University Press, 1987; 177–202.