# Preliminary Ribosomal RNA Phylogeny of Gastropod and Unionoidean Bivalve Mollusks

**Gary Rosenberg**
Academy of Natural Sciences
1900 Benjamin Franklin Parkway
Philadelphia, PA 19103 USA

**Gerald S. Kuncio**
Department of Medicine
University of Pennsylvania
Philadelphia, PA 19104 USA

**George M. Davis**
Academy of Natural Sciences
1900 Benjamin Franklin Parkway
Philadelphia, PA 19103 USA

**M. G. Harasewych**
Department of Invertebrates
National Museum of Natural History
Smithsonian Institution
Washington, D.C. 20560 USA

## ABSTRACT

Sequences of about 150 nucleotides in the D6 region of the large (28S) ribosomal molecule were obtained from 20 unionoidean bivalves and 13 gastropods, including 9 truncatellids, 1 muricid, 1 cancellariid, 1 melongenid, and 1 pleurotomariid. These were analyzed along with sequences from Emberton *et al.* (1990) for 8 pulmonates, a helicinid and a pomatiopsid. Rates of divergence varied by a factor of two, with unionoidean and the helicinid sequences differing by 15–20% relative to mouse, pulmonates by 22–24%, neogastropods by 24–27%, and rissooideans by 27–32%. Length variation among sequences occurred mainly in the D6 loop, and complementary mutations were seen in the D6 stem. Cladistic analysis found 24 equally parsimonious trees; the strict consensus supports monophyly of Unionoidea, Rissooidea, Pulmonata and Stylommatophora. Monophyly of Neogastropoda is not contradicted.

Some groupings are anomalous when compared to morphology-based phylogenies. Helicinidae groups with Unionoidea, Pleurotomariidae with Neogastropoda, and *Geomelania* (Truncatellidae) with Pomatiopsidae. In each case, addition of taxa that intersect long branches (e.g. Chitonidae, Patellogastropoda) might show that characters interpreted as synapomorphic are plesiomorphic or convergent. The observed grouping of Muricidae and Cancellariidae is well-supported, indicating that cancellariids are a highly derived group within the Stenoglossa. Pleurotomariidae are more closely related to the other gastropods in the analysis than are Helicinidae, supporting Haszprunar's (1988) anatomy-based conclusion.

To date, sequence studies of mollusks have not overturned phylogenies based on morphology, but rather have helped in choosing among competing morphology-based hypotheses. Like morphological data, sequence data are subject to problems of convergence, unequal rates of evolution, and choice of taxa. Classifications must be based on all available data to maximize the potential for detecting convergences and correctly resolving phylogenetic relationships.

*Key words:* Gastropoda, Bivalvia, 28S ribosomal RNA, phylogeny, cladistics, rates of divergence.

## INTRODUCTION

Ribosomal RNA (rRNA) and rDNA sequences have proven to be valuable and versatile sources of data for phylogenetic inferences. The great variation in rates of evolution of different parts of rDNA allows evolutionary investigations from the level of population and species through kingdom, by study of appropriately variable regions. This variation has caused debate as to the reliability of some types of rDNA sequence data in phylogenetic analysis, but it has become clear that, when analyzed with care, all parts of the sequence are potentially informative (Wheeler & Honeycutt, 1988; Swofford & Olsen, 1990; Hillis & Dixon, 1991; Dixon & Hillis, 1993). Ribosomal sequences have most often been used in determining relationships among bacteria (e.g., Woese & Olsen, 1986) and vertebrates (e.g., Hedges *et al.*, 1990), but can be used with any organism (e.g., Sogin *et al.*, 1986; Field *et al.*, 1988).

Of the more than 150 phylogenetic studies of rDNA sequences published to date (Hillis & Dixon, 1991), only a few have been devoted to mollusks. Ghiselin (1988) looked at molluscan origins using 18S rRNA, Emberton *et al.* (1990) at pulmonate relationships using D6 28S rRNA and Tillier *et al.* (1992) at gastropod phylogeny using D1 28S rRNA. These studies have demonstrated the potential for rDNA sequences to sharpen and resolve ideas of molluscan phylogeny, particularly at the ordinal level and above.

We have supplemented the 10 gastropod sequences obtained by Emberton *et al.* (1990) with data from 33 more molluscan species. All 43 sequences were used in this study with the aims of 1) analyzing aspects of caenogastropod, archaeogastropod, and unionoidean relationships, 2) surveying variability in 28S rRNA sequences in mollusks, and 3) examining how choice of taxa affects

**Table 1.** Localities and catalogue numbers of voucher specimens for this study. Depository is the Academy of Natural Sciences of Philadelphia (ANSP) unless otherwise noted; USNM = United States National Museum. Condition: f = frozen; l = lyophilized; w = whole live animal. Voucher lot of *Anodonta imbecillis* was collected in 1974; tissue sample was from same population in 1975. See Emberton *et al.* (1990) for vouchers of *Helicina orbiculata, Oncomelania hupensis, Biomphalaria glabrata, Mesodon inflectus, Mesodon normalis, Neohelix albolabris, Triodopsis hopetonensis, Haplotrema concavum, Mesomphix latior, Ventridens cerinoideus.*

| Species | Condition | Locality | Catalogue no. |
|---|---|---|---|
| *Anodonta cataracta* | l | NE of Swedesboro, Gloucester Co., New Jersey | 333526, 341937 |
| | l | Swartswood, Sussex Co., New Jersey | 334429, 341946 |
| *A. grandis* | l | Ramah Borrow Canal, Iberville Parish, Louisiana | 341888 |
| *A. imbecillis* | l | Magnolia Springs, Jenkins Co., Georgia | 333563 |
| *Amblema plicata* | f | Bogue Chitto Creek, Dallas Co., Alabama | 373820, A12742 |
| *Elliptio complanata* | l | Swartswood, Sussex Co., New Jersey | 334428 |
| | l | Lake Lacawac, Lake Ariel, Wayne Co., Pennsylvania | 339430 |
| | l | Deep Creek, Nanticoke River, Sussex Co., Delaware | 339340 |
| *Fusconaia cerina* | f | Bogue Chitto Creek, Dallas Co., Alabama | 397248 |
| *Gonidea angulata* | l | Pit River, SW of Canby, Modoc Co., California | 339965 |
| *Lampsilis teres* | f | Bogue Chitto Creek, Dallas Co., Alabama | 373821, A12744 |
| *L. claibornensis* | f | Bogue Chitto Creek, Dallas Co., Alabama | 397249 |
| *Megalonaias boykiniana* | l | Ochlockonee River, Leon Co., Florida | 346111 |
| *Obliquaria reflexa* | f | Bogue Chitto Creek, Dallas Co., Alabama | 397247, A12722 |
| *Quadrula cylindrica* | l | Kyles Ford, Clinch River, Hancock Co., Tennessee | 335041 |
| *Q. quadrula* | f | Bogue Chitto Creek, Dallas Co., Alabama | 397246, A12728 |
| *Plectomerus dombeyanus* | l | Ramah Borrow Canal, Iberville Parish, Louisiana | vouchers not kept |
| *Pleurobema cordatum* | l | Ouachita River, Arkadelphia, Clark Co., Arkansas | 340629 |
| *Unio pictorum* | f | Shropshire Canal, north of Chester, near Mollington Grange, England | 350622 |
| *Uniomerus tetralasmus* | l | Magnolia Springs, Jenkins Co., Georgia | 353133 |
| *Cumberlandia monodonta* | l | Kyles Ford, Clinch River, Hancock Co., Tennessee | 341956 |
| *Margaritifera falcata* | l | Siletz River, Lincoln Co., Oregon | 339339 |
| *M. margaritifera* | l | Locust Creek, Schuylkill River, Pennsylvania | 334867 |
| *Perotrochus maureri* | f | 90 miles east of Charleston, South Carolina | USNM 875218 |
| *Truncatella* sp. | w | Harrison Point Lighthouse, Barbados | 397286 |
| *T. caribaeensis* | w | Bay side, Mile 57, Grassy Key, Florida Keys | 397275 |
| *T. clathrus* | w | Bay side, Mile 57, Grassy Key, Florida Keys | 397273 |
| *T. pulchella* | w | Bay side, Mile 57, Grassy Key, Florida Keys | 397274 |
| | w | Falmouth, Trelawny Parish, Jamaica | 397264 |
| *T. reclusa* | w | Cumaca, Northern Range, Trinidad | 397285 |
| *T. scalaris* | w | Falmouth, Trelawny Parish, Jamaica | 397263 |
| *T. subcylindrica* | w | The Fleet, Dorset, England | 397280 |
| *Geomelania* sp. | w | North of Quickstep, Trelawny Parish, Jamaica | 397283 |
| *G. typica* | w | Wallingford, St. Elizabeth Parish, Jamaica | 397284 |
| *Busycon carica* | f | Cape Henlopen, Sussex Co., Delaware | USNM 847010 |
| *Mancinella deltoidea* | f | South Beach, Miami, Dade Co., Florida | USNM 870850 |
| *Progabbia cooperi* | f | Off La Jolla, San Diego Co., California | USNM 846054 |

phylogenetic inference. The results presented here must be considered preliminary until data for longer sequences and additional ordinal level taxa are available.

## MATERIALS AND METHODS

We obtained sequences from 20 unionoidean bivalves and 13 gastropods, including 9 truncatellids, 1 muricid, 1 cancellariid, 1 melongenid, and 1 pleurotomariid. Species names, localities, voucher information, and higher classification are given in Tables 1 and 2. Truncatellid RNA was obtained by homogenizing live animals, other gastropod RNA from frozen tissue, and unionoidean RNA from lyophilized or frozen tissue. Methods for sequencing followed Emberton *et al.* (1990) and were done in the same laboratory, using the same primer for the D6 region, complementary to nucleotides 2099 through 2118 for mouse as published by Hassouna *et al.* (1984). Each species was sequenced at least twice, or more often as necessary to resolve ambiguities in nucleotide identity.

**Sequence alignment:** Gross alignment of the sequences was easily achieved because of large conservative stretches in the D6 flanks. The MALIGN program of Wheeler and Gladstein (version 1.73, 1993), was used to refine the manual alignment. The following weights (costs), with options alignaddswap and treeaddswap, yielded alignments that matched overall the manual alignments of the D6 flanks, while providing improvement in details:

transitions 1, transversions 3; internal gaps 10; leading gaps 5; trailing gaps 5.

**Phylogenetic analysis:** Informative and variable nucleotide positions, indicated by "i" or "v" in Figure 1, were analyzed using Hennig86. The sequence of commands "mhennig; bb; ie*;" was used, which guarantees finding all of the most parsimonious trees. All characters were equally weighted and unordered (command "cc-."). Those gaps marked with a hyphen (-) in Figure 1 were scored as characters, except for the deletion from positions 74 to 79 in *Truncatella clathrus*. Species that showed no differences in sequence were combined for the purpose of the phylogenetic analysis.

## RESULTS

Aligned sequences are shown in Figure 1. The 5′ flanking region (positions 1–46), and the 3′ flank (positions 99–161) are conservative, and only a few gaps were inserted to align the sequences. The D6 loop shows considerable variation in length, and were too variable in the nonpulmonate gastropods to be reliably aligned. A number of complementary changes can be seen in the stem region, positions 47–55 and 90–98( Figure 2).

Among the 43 species, 26 different sequences were found. All species with identical sequences were confamilial. As reported by Emberton *et al.* (1990), sequences were invariant among the four polygyrids. Among 20 unionoideans, only 6 different sequences were found. *Cumberlandia*, *Gonidea* and the two *Margaritifera* all had distinct sequences. The other 16 unionids differed from each other by at most one nucleotide, falling into two groups, referred to here as the *Anodonta* and *Amblema* groups. In contrast, sequences differed strongly among truncatellids: each of the nine species had a unique sequence. Within the genus *Truncatella*, all species differed from each other by at least five nucleotides. Differences were concentrated in the D6 loop, and involved significant variation in length, in addition to nucleotide substitution. Only a partial sequence was obtained for *Perotrochus*.

Molluscan sequences differed from those of mouse at 15 to 32 percent of the sites (Table 3). Sequences from the unionoideans and *Helicina* were the most conservative, differing from mouse at 15 to 20 percent of sites. Gastropods, excluding *Helicina*, differed from mouse at 22 to 32 percent of sites, with pulmonates showing less sequence divergence (22 to 24 percent) than neogastropods (24 to 27 percent) and rissooideans (27 to 32 percent).

Use of published sequences (Gutell & Fox, 1988) from *Mus*, *Rattus*, *Xenopus* or *Homo* as outgroup did not affect polarization of characters. Sequences for *Caenorhabditis*, *Physarum* and *Saccharomyces* were more divergent from molluscan sequences than were vertebrate sequences, and in some regions could not be aligned satisfactorily with them. They were therefore judged less appropriate as outgroups. The sequence from *Drosophila* (Tautz *et al.*, 1988), shown in Figure 1, could be aligned, but differs from the molluscan sequences at twice as many sites as does the mouse sequence. Of 148 nucleotide positions scored in bivalves, 35 (24%) are variable relative to mouse, whereas 72 (49%) are variable relative to *Drosophila*. This degree of divergence made *Drosophila* unreliable as an outgroup.

Out of 153 alignable sites, 73 (48%) are variable in mollusks relative to mouse and 55 are potentially informative for cladistic analysis. With mouse as the outgroup, cladistic analysis yielded 24 equally parsimonious trees, the strict consensus tree of which is shown in Figure 3.

**Table 2.** Higher classification of genera for which sequence data were analyzed. Classification follows Davis and Fuller (1981) for Unionoidea, Haszprunar (1988b) for Gastropoda, Rosenberg (1989) for Rissooidea, Kantor and Harasewych (1992) for Neogastropoda and Emberton *et al.* (1990) for Pulmonata.

- Bivalvia
  - Paleoheterodonta
    - Unionoidea
      - Unionidae
        - Unioninae
          - *Unio*
        - Ambleminae
          - *Amblema*
          - *Megalonaias*
          - *Plectomerus*
          - *Quadrula*
        - Pleurobemini
          - *Elliptio*
          - *Fusconaia*
          - *Pleurobema*
          - *Uniomerus*
        - Gonideini
          - *Gonidea*
        - Lampsilini
          - *Lampsilis*
          - *Obliquaria*
        - Anodontinae
          - *Anodonta*
        - Margaritiferinae
          - *Margaritifera*
          - *Cumberlandia*
- Gastropoda
  - Neritopsina
    - Neritoidea
      - Helicinidae
        - *Helicina*
  - Vetigastropoda
    - Pleurotomarioidea
      - Pleurotomariidae
        - *Perotrochus*
  - Caenogastropoda
    - Neotaenioglossa
      - Rissooidea
        - Pomatiopsidae
          - *Oncomelania*
        - Truncatellidae
          - Truncatellinae
            - *Truncatella*
          - Geomelaniinae
            - *Geomelania*
    - Neogastropoda
      - Stenoglossa
        - Muricoidea
          - Muricidae
            - *Mancinella*
          - Melongenidae
            - *Busycon*
        - Cancellarioidea
          - *Progabbia*
- Pulmonata
  - Basommatophora
    - Planorboidea
      - Planorbidae
        - *Biomphalaria*
  - Stylommatophora
    - Holopoda
      - Polygyroidea
        - Polygyridae
          - *Mesodon*
          - *Neohelix*
          - *Triodopsis*
      - Holopodopes
        - Rhytidoidea
          - Haplotrematidae
            - *Haplotrema*
      - Aulacopoda
        - Zonitoidea
          - Zonitidae
            - *Mesomphix*
            - *Ventridens*

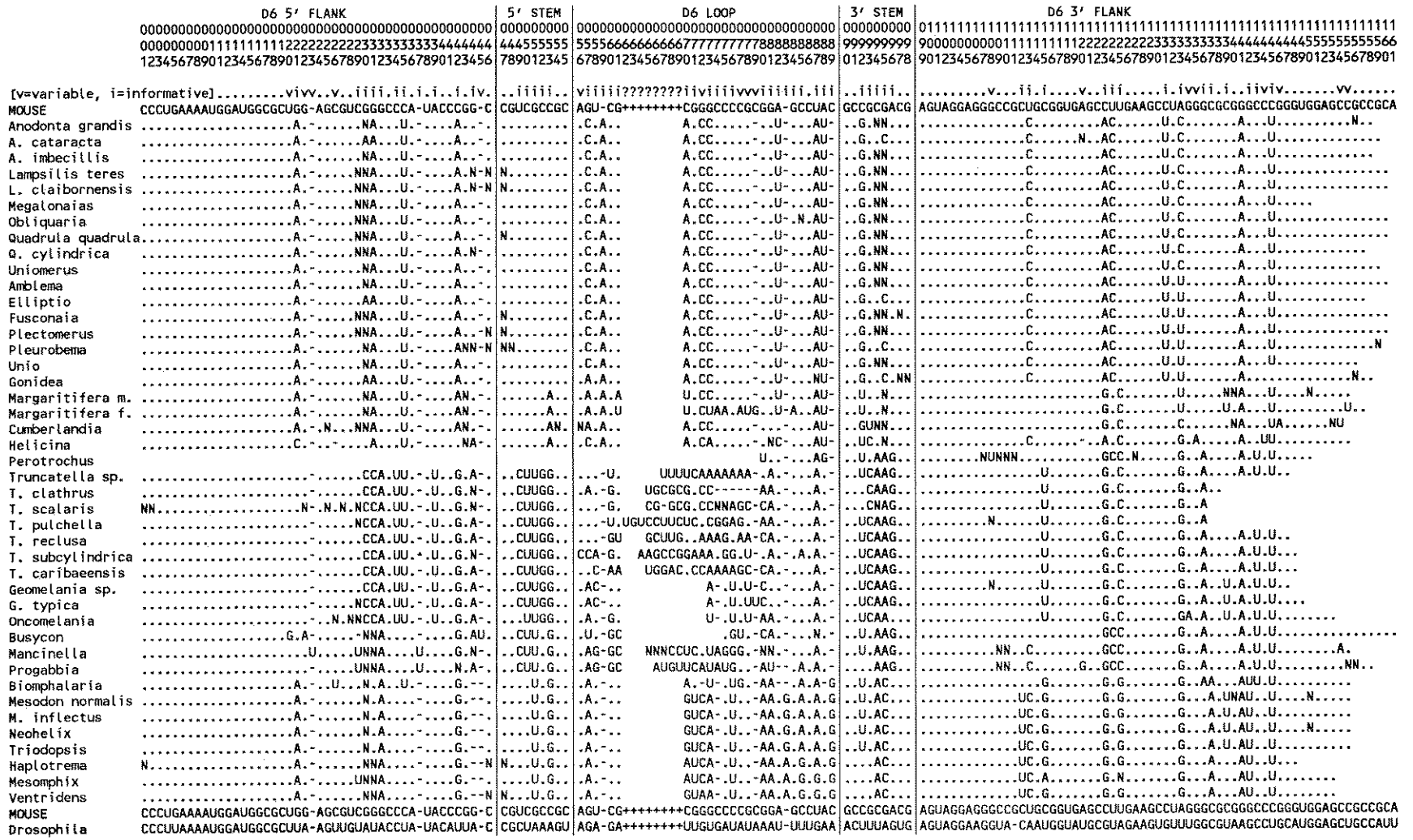


**Figure 1.** Aligned sequences. Numbers across the top in vertical triplets refer to the nucleotide position. The 5′ end of the sequences corresponds to mouse position 1836 of Hassouna *et al.* (1984). Nucleotides unchanged from mouse are represented by a dot (.); gaps by a dash (-). Nucleotides that could not be scored are marked "N". Pluses (+) in the D6 loop of mouse and *Drosophila* (at bottom) are an ellipsis where those sequences are longer than molluscan ones. Above the mouse sequence, positions variable in mollusks are marked "v"; those potentially informative for cladistic analysis are marked "i". The vertical lines delimit the 5′ stem, where the ribosomal molecule folds back and base pairs with itself (see Figure 2).

## DISCUSSION

Analyses of rRNA data must take into account several complicating factors: the reliability of the alignments, bias in nucleotide composition, the ratio of transitions to transversions, and the affect of complimentary mutations in base-paired regions. The latter two are often handled by weighting the data in various ways. The significance of results, in terms of the reliability of nodes in a phylogram, is often assessed by bootstrapping.

### Reliability of Alignments

Hillis and Dixon (1991) noted that alignments are often ambiguous when sequences differ by more than 30 percent in a given region. Except in the D6 loop, the molluscan sequences exhibited less than 30% divergence, and alignment was not problematic. In the phylogenetic analysis (Figure 3), all variable and informative nucleotide positions in the D6 loop were included, except numbers 62 to 69 (Figure 1). The alignment of these and of nucleotides 70 to 78 was uncertain in the rissooideans and neogastropods because of length polymorphism and sequence variation; most of these sequences are shown flush right with no gaps inserted. When characters 70 to 78 for these taxa were excluded from the analysis, the same consensus tree as shown in Figure 3 was obtained, except that all rissooideans formed a unresolved polychotomy, save the two *Geomelania*, which grouped together.

### Bias in Nucleotide Composition

The sequences analyzed in this study exhibited a significant bias in favor of guanine and cytosine: the GC:AU ratio of mouse was 3:1, and all taxa had a ratio greater than 1.3:1 (Table 3). If mouse and mollusks had evolved this bias independently, it might affect the reliability of the mouse sequence as an outgroup for polarizing characters, because convergences could be mistaken for plesiomorphies. Because the sequences of mouse and mollusks have diverged only 15 to 30%, as discussed above, it is unlikely that the large GC biases evolved independently in these taxa. Nevertheless, as a control, the analysis was rerun with bivalves as outgroup, and with bivalves plus *Helicina* as outgroup. Topology of the ingroup was not affected by the change in outgroup.

### Weighting of Character Data

Differential weighting schemes for character data are often used when analysis of a data set gives poorly resolved or unexpected phylogenetic inferences. If subsets of the data can be reasonably regarded as being less likely to suffer from homoplasy, they are given more weight. In rRNA sequences, transversions can be up to five times less frequent than transitions and are sometimes accorded greater weight (Crother & Presch, 1992). Mishler *et al.* (1988) suggested weighting according to the observed ratio of transitions to transversions in a given data set, but cautioned that this ratio is not likely uniform among

| Taxon | 0000000000 4445555555 7890123456 | 0000000000 9999999999 0123456789 | paired sites |
|---|---|---|---|
| Mouse | CGUCGCCGC | GCCGCGACG | 8 |
| *Anodonta* group | CGUCGCCGC | GC**G**GCCACG | 8 |
| *Amblema* group | CGUCGCCGC | GC**G**GCCACG | 8 |
| *Gonidea* | CGUCGCCGC | GC**G**GCCANN | 8 |
| *M. margaritifera* | CGUCGC**A**GC | GCUGCNACG | 8 |
| *M. falcata* | CGUCGC**A**GC | GCUGCNACG | 8 |
| *Cumberlandia* | CGUCGCANC | GCGUNNACG | 7? |
| *Helicina* | CGUCGC**A**GC | GCUCCNACG | 7? |
| *Perotrochus* | ?????????? | GCUG**AA**GCG | ? |
| *Truncatella sp.* | CG**CUUG**GGC | GCU**CAA**GCG | 8.5 |
| *T. clathrus* | CG**CUUG**GGC | GC**CCAA**GCG | 9 |
| *T. scalaris* | CG**CUUG**GGC | GC**CCNA**GCG | 9 |
| *T. pulchella* | CG**CUUG**GGC | GCU**CAA**GCG | 8.5 |
| *T. reclusa* | CG**CUUG**GGC | GCU**CAA**GCG | 8.5 |
| *T. subcylindrica* | CG**CUUG**GGC | GCU**CAA**GCG | 8.5 |
| *T. caribaeensis* | CG**CUUG**GGC | GCU**CAA**GCG | 8.5 |
| *Geomelania sp.* | CG**CUUG**GGC | GCU**CAA**GCG | 8.5 |
| *G. typica* | CG**CUUG**GGC | GCU**CAA**GCG | 8.5 |
| *Oncomelania* | CG**UUUG**GGC | GCU**CAAA**CG | 8.5 |
| *Busycon* | CG**CUU**CGGC | GCUG**AA**GCG | 8.5 |
| *Mancinella* | CG**CUU**CGGC | GCUG**AA**GCG | 8.5 |
| *Progabbia* | CG**CUU**CGGC | GC**C**G**AA**GCG | 9 |
| *Biomphalaria* | CGUCUCGGC | GCUGACACG | 7.5 |
| Polygyridae | CGUCUCGGC | GCUGACACG | 7.5 |
| *Haplotrema* | NGUCUCGGC | GC**C**GACACG | 8 |
| *Mesomphix* | CGUCUCGGC | GC**C**GACACG | 8 |
| *Ventridens* | NGUCUCGGC | GC**C**GACACG | 8 |

**Figure 2.** Detail of the D6 stem regions. Numbers across the top correspond to nucleotide positions in Figure 1. These regions fold back on each other (position 47 corresponding to 98, 48 to 97, etc.); base-pairs form where cytosine (C) matches guanine (G) and adenine (A) matches uracil (U). Mutations can disrupt the pairing, and complementary mutations sometimes restore the pairing. Inferred complementary mutations are in boldface. The number of paired sites is shown in the left column for each taxon. (The weaker pairing of uracil with guanine is counted as 0.5). Note that only a few species have all nine stem positions paired; complementary mutations do not always occur.

taxa or in different regions within a sequence. Wheeler (1990) suggested a combinatorial weighting method based on observed nucleotide distributions among taxa, but this method assumes that there are no hidden intermediate states, and has been criticized on the grounds that it was applied *a posteriori* to trees based on equally weighted data (Albert & Mishler, 1992).

The observation of compensatory mutations in base-paired stem regions of rRNA molecules has led to the suggestion that such regions should be excluded from analyses, or weighted one-half because of non-independence of characters (Steele *et al.*, 1988; Wheeler & Honeycutt, 1988). In contrast, Smith (1989) found that results based on paired nucleotide positions were more reliable than those based on unpaired positions. Hillis and Dixon (1991) and Dixon and Hillis (1993) found that paired regions can contain most of the informative sites in some analyses, and suggested that paired regions be down-weighted by no more than twenty percent, because compensatory mutations do not always occur. This is true in the mollusk data, where perfect complementarity is not maintained (Figure 2).

We have opted to run our data with equal weights for

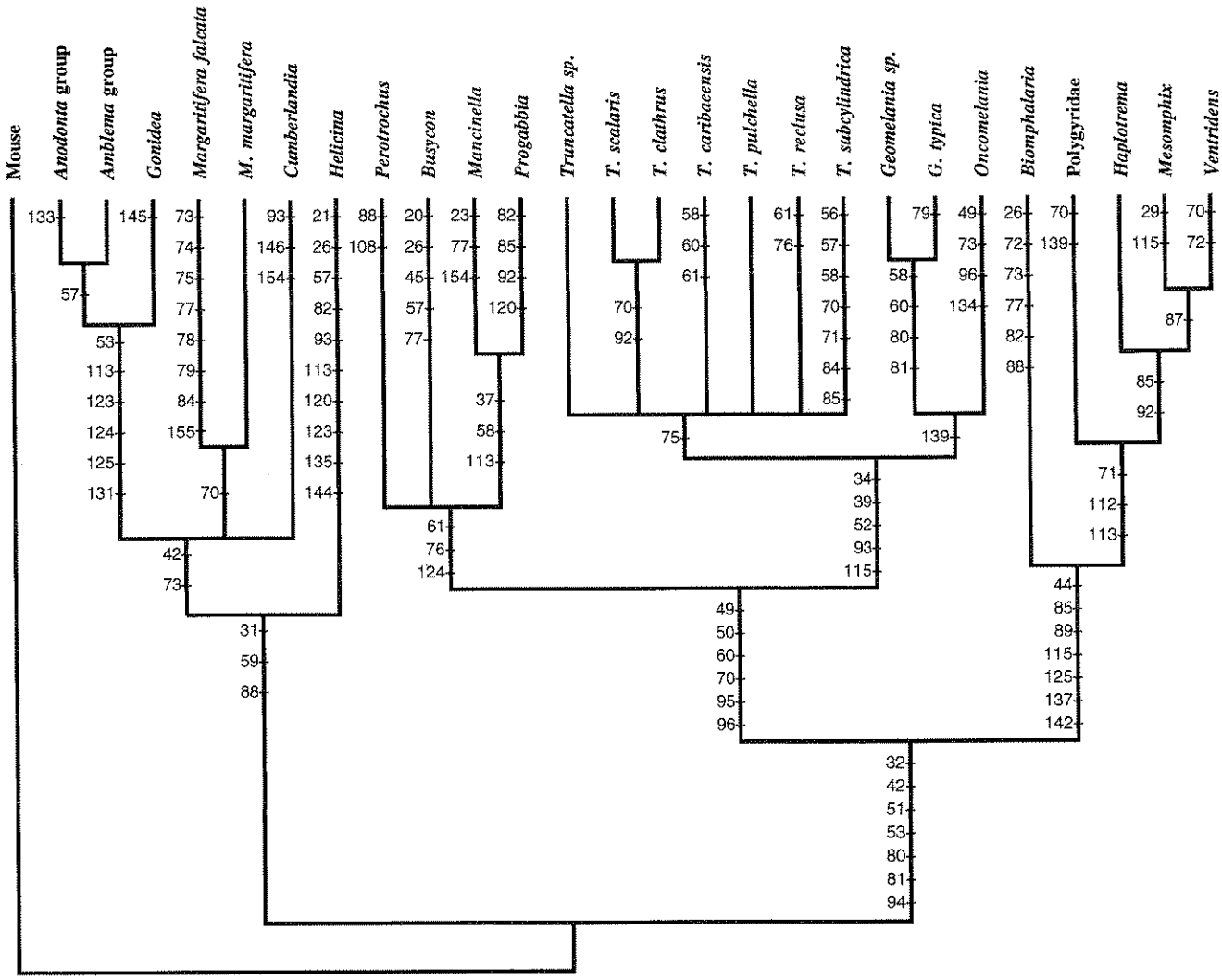


**Figure 3.** Strict consensus of 24 equally parsimonious trees of length 205 resulting from cladistic analysis of the sequence data. The consistency index of each of the most parsimonious trees is 0.64, the retention index is 0.81, and the rescaled consistency index is 0.52. Characters having states that define each node are indicated by nucleotide number. These characters support the nodes in each of the 24 most parsimonious trees making up the consensus tree, as shown using the character tracing feature of MacClade 3.01 (Maddison & Maddison 1992). Characters whose assignment to a node is ambiguous are not shown.

all characters for two reasons. First, because weighting schemes for transitions versus transversions and for compensatory mutations are still controversial, we prefer the assumptions of equal weighting to those that these schemes require. Second, the purpose of weighting is to compensate for the misleading effects of undetected homoplasy. (Homoplasy that has been detected is not misleading.) Undetected homoplasy exists when there are hidden intermediate character states between taxa that appear identical for a particular character. Given a sufficient number of undetected homoplasies, taxa that are convergent can be mistakenly inferred to be closely related. This is the long branch length problem first noted by Felsenstein (1978). One cannot know *a priori* what weights (if any) will overcome the long branch problem for a particular data set. Also, weighting cannot determine unknown intermediate states, so weighting cannot completely overcome the problem. Often, however, taxa can be added to the analysis that will intersect long branches, revealing the intermediate states. We therefore view convergence in sequence data primarily as a problem of taxon sampling rather than of weighting. In the following discussions, we attempt to identify areas where the addition of taxa is likely to change the relationships shown in Figure 3.

## Reliability of Nodes

Hillis and Bull (1993) have recently shown that bootstrap proportions are conservative estimators of the probability of a clade's being real, if certain conditions are met. Hedges (1992) has shown that the number of bootstrap

**Table 3.** Differences in sequence and nucleotide composition between mollusks and mouse. If a position was not scored in a species, it was considered to match the consensus sequence for the higher taxon to which the species belongs. For example, at position 30 (figure 1), all unionoideans were considered to have "A," even though the nucleotide was not scorable in some species. Similarly, at the 3′ end, unionoidean sequences were all considered to extend to position 160, rissooideans to 153, neogastropods to 161, and pulmonates to 155. Positions 62 to 69 for mouse shown by pluses (+) in figure 1 are CGCGGCGU, extending from the 5′ side.

| Sequence | Nucleotide differences | Total compared | Percentage difference | GC/AU |
|---|---|---|---|---|
| Mouse | — | — | — | 3.1 |
| *Anodonta* group | 24 | 147 | 0.16 | 2.3 |
| *Amblema* group | 24 | 147 | 0.16 | 2.2 |
| *Gonidea* | 22 | 147 | 0.15 | 2.2 |
| *Margaritifera falcata* | 30 | 147 | 0.20 | 1.6 |
| *M. margaritifera* | 24 | 147 | 0.16 | 1.9 |
| *Cumberlandia* | 25 | 147 | 0.17 | 1.9 |
| *Helicina* | 27 | 139 | 0.19 | 2.1 |
| *Perotrochus* | 16 | 70 | 0.23 | 2.2 |
| *Truncatella* sp. | 42 | 143 | 0.29 | 1.4 |
| *T. clathrus* | 42 | 145 | 0.29 | 1.8 |
| *T. scalaris* | 41 | 145 | 0.28 | 1.9 |
| *T. pulchella* | 46 | 148 | 0.31 | 1.6 |
| *T. reclusa* | 43 | 145 | 0.30 | 1.6 |
| *T. subcylindrica* | 46 | 146 | 0.32 | 1.7 |
| *T. caribaeensis* | 46 | 145 | 0.32 | 1.7 |
| *Geomelania* sp. | 38 | 140 | 0.27 | 1.7 |
| *G. typica* | 39 | 140 | 0.28 | 1.6 |
| *Oncomelania* | 38 | 140 | 0.27 | 1.4 |
| *Busycon* | 36 | 149 | 0.24 | 1.9 |
| *Mancinella* | 40 | 152 | 0.26 | 1.9 |
| *Progabbia* | 41 | 153 | 0.27 | 1.8 |
| *Biomphalaria* | 34 | 142 | 0.24 | 1.6 |
| Polygyridae | 34 | 143 | 0.24 | 1.7 |
| *Haplotrema* | 33 | 143 | 0.23 | 1.8 |
| *Mesophix* | 32 | 143 | 0.22 | 1.6 |
| *Ventridens* | 33 | 143 | 0.23 | 1.8 |

replicates necessary for bootstrap proportions to be reliable is much higher than typically used in phylogenetic studies. We attempted to bootstrap our data using the general heuristic algorithm in PAUP 3.1.1 (Swofford, 1993). Unfortunately, some bootstrap replicates generated more than 5,000 equally parsimonious trees, and with available computing power, it was not possible to complete the several hundred replicates needed to obtain reliable bootstrap values. Felsenstein (1985), however, has shown that when all characters are perfectly compatible, bootstrapping will show significant support for a group if it is defined by at least three characters. Therefore, groups defined by fewer than three characters in Figure 3 cannot be considered to have significant support. Groups defined by three or more characters might not be significant if some of the characters are homoplastic, or incorrectly polarized.

Because an explicit, morphology-based cladistic analysis of gastropod phylogeny has not yet been published (see Bieler, 1990), we could not perform a combined analysis of molecular and morphological data (see Bull *et al.*, 1993; Eernisse & Kluge, 1993). We have instead attempted to evaluate on a case-by-case basis the evidence supporting the major clades shown in Figure 3. Particular characters are referred to in the form "21>A," meaning position 21, character state A.

**Archaeogastropoda**

Two taxa traditionally classified as archaeogastropods were included in the analysis, Neritopsina (*Helicina*) and Pleurotomariidae (*Perotrochus*). The Neritopsina (= Neritimorpha, Neritoida) traditionally have been regarded as having strong affinities with the Caenogastropoda (Bieler, 1992). Recent classifications, however, have reversed this, with Neritopsina being placed basally to the Archaeo- and Caenogastropoda (Haszprunar, 1988a, b; Healy, 1988; Hickman, 1988) on the basis of lack of skeletal rods in the ctenidium, and on sperm morphology. Haszprunar (1988a, b) regarded similarities of Neritopsina and Caenogastropoda as convergences due to the specialized reproductive biology of Neritopsina.

The sequence data show Pleurotomariidae rather than Neritopsina to be more closely related to the other gastropods, supporting Haszprunar's morphology-based classification. Trees with branching order (Pleurotomariidae (Neritopsina (Caenogastropod Pulmonata))) are three steps longer than those with the order (Neritopsina (Pleurotomariidae (Caenogastropod (Pulmonata))). The latter tree is supported by three characters: 94>A, 136>A and 143>U. Only a partial sequence was obtained for Pleurotomariidae, and several positions (32, 42, 51, 53), may also support its derived position relative to Neri-

topsina. The position of Pleurotomariidae is discussed further under Caenogastropoda.

Neritopsina groups with the Bivalvia based on three characters (31>A, 59>A and 88>U) whereas only a single character supports grouping it with the Gastropoda (133>G). All other nucleotide positions are uninformative as to its relationships. The three characters supporting affinity with bivalves may prove plesiomorphic for mollusks when sequences from other molluscan classes are added. The sequence of Neritopsina has diverged relatively little from that of bivalves, and several places where it has changed are uniquely autapomorphic (positions 21, 26, 82, 120, 135, 144). Sequences from additional taxa such as Patellogastropoda, Cocculiniformia, and Neomphalidae might show that some of the autapomorphies of Neritopsina are actually basal synapomorphies that will unite the Gastropoda.

**Caenogastropoda**

Neotaenioglossa, Neogastropoda and Pleurotomariidae group together in our analysis, with Pleurotomariidae basal to the Neogastropoda. This is consistent with Ponder's (1973) hypothesis of archaeogastropod origins of the Neogastropoda, and contradicts the monophyly of Caenogastropoda. However, osphradial and spermatozoic characters have been found recently that indicate that the Neogastropoda probably evolved from the higher mesogastropods (Haszprunar, 1988a, b; Healy, 1988; Taylor & Morris, 1988). In Figure 3, sixteen characters define branches leading to the *Perotrochus* lineage. In *Perotrochus*, no data are available for ten of these because of the incomplete sequence, and two positions differ (80>U and 81>G). The single character that unites Pleurotomariidae with Neogastropoda in our analysis (124>C) is insufficient to refute morphological characters defining the Neogastropoda and is likely convergent. Addition of taxa such as Littorinoidea or Cerithioidea that would intersect the branch between Pleurotomariidae and Neogastropoda might reveal this convergence, as might completion of the partial sequence for Pleurotomariidae. Because the position of *Perotrochus* in Figure 3 is weak, it cannot be taken as contradicting the monophyly of Caenogastropoda or Neogastropoda.

**Neogastropoda**

Taylor and Morris (1988) concluded on the basis on morphological characters that Neogastropoda is monophyletic. If *Perotrochus* is excluded, the nucleotide sequences supports this monophyly, character 61>C being a unique synapomorphy for the Neogastropoda. Another character also supports the node (76>G) but with homoplasy elsewhere in the tree. The relationships of the three neogastropod taxa used in our study differ from those postulated by Taylor and Morris (1988) who regarded Cancellarioidea to be a possible sister group of Rachiglossa + Conoidea (i.e., Stenoglossa + Toxoglossa). In our analysis, the Muricidae and Cancellariidae group together, with Melongenidae as a sister group. Three characters support the monophyly of Muricidae and Cancellariidae: 37>U (an insertion), 58>G and 113>C, and none contradict it.

The relationship of Cancellariidae to other neogastropod taxa has been uncertain, with the group having been included by various authors in the Toxoglossa, Stenoglossa, and its own order, the Nematoglossa (see Petit & Harasewych, 1990). Most subsequent authors have followed Ponder (1973) in dividing the Neogastropoda into the corresponding superfamilies Conoidea, Muricoidea and Cancellarioidea. More recently, Kantor and Harasewych (1992) noted anatomical similarities between Cancellariidae and the stenoglossan family Volutomitridae, and suggested that a reassessment of the taxonomic rank and systematic position of these taxa was warranted. The present data support the hypothesis that the Cancellariidae comprise a highly derived group within the Stenoglossa, as reflected in such earlier classifications as those of Thiele (1929) and Wenz (1943).

**Neotaenioglossa**

All of the neotaenioglossans studied are rissooideans. Five characters support the monophyly of Rissooidea: 34>U, 39>U, 52>G, 93>C, 115>U. All of these are uniquely derived, except 93>C, which is convergent with *Helicina*. Within Rissooidea, *Geomelania* grouped with the Pomatiopsidae rather that the Truncatellidae, but a number of morphological characters argue that it is a truncatellid. These include truncation of the apical whorls of the shell, reduction in the number of anterior rachidian cusps, the looping mode of locomotion, and shortening of the pleuro-supraesophageal connective (Davis, 1979; Rosenberg, 1989). The single nucleotide character that appears synapomorphic for *Geomelania* and Pomatiopsidae (139>U) may prove to be plesiomorphic for rissooideans when sequences from more taxa, such Assimineidae and Hydrobiidae are added to the analysis. Within the Truncatellinae, relationships were unresolved, except that *Truncatella scalaris* and *Truncatella clathrus* grouped together as sister species. This grouping is confirmed by allozyme and morphological data (Rosenberg, 1989).

**Pulmonata**

Our alignment for species treated by Emberton *et al.* (1990) differs in minor details from their published alignment, and polarities of some characters have changed with added data, but these differences do not affect inferred relationships among the pulmonates. The monophyly of Pulmonata is strongly supported by seven characters and monophyly of Stylommatophora is supported by three characters. No other nodes within the pulmonates are supported by more than two characters.

**Bivalvia**

The sequence data support the distinction between Ambleminae and Margaritiferinae advocated by Davis and Fuller (1981), but do not reflect the distinctiveness of Anodontinae, which nests within Ambleminae (represented by *Amblema* and *Gonidea* in Figure 3). Six characters give support for grouping Anodontinae and Ambleminae apart from the Margaritiferinae, but a few of these would be plesiomorphic if *Helicina* were rerooted. Only character 133>C supports the grouping of *Anodonta* with some of the amblemines. Unpublished sequences from the 5′ terminus obtained during this study were also uninformative, containing only a single variable site, corresponding to position 23 of Emberton *et al.* (1990). Given the large body of immunological, electrophoretic, and anatomical evidence showing the distinctiveness of the Anodontinae from the Ambleminae (Davis & Fuller, 1981, Davis *et al.*, 1981), we maintain the tripartite subfamilial classification of Unionidae, with Margaritiferinae as sister group to the clade containing Ambleminae and Anodontinae. The sequences for *Unio* and *Amblema* are identical, indicating that Ambleminae may be a synonym of Unioninae, however, we refrain from changing the taxonomy until more data sets for *Unio* are available.

Sequence Variability in Mollusks

Emberton *et al.* (1990) found that 13% of sites in the D6 divergent domain of 28S rRNA were informative for stylommatophoran phylogeny, but only 1% in the D6 flanking regions were informative. They concluded that divergent domains of LrRNA would be of "some value in resolving stylommatophoran phylogeny." We have found that the D6 region is more variable in mollusks than anticipated from the results of the first study, with 23% of sites in the D6 flank and more than 70% in the divergent domain being informative at some level for molluscan phylogeny.

Variability in the bivalves and pulmonates is almost entirely in the form of nucleotide substitutions; there are only a few insertions and deletions. In the caenogastropods, the D6 loop region is subject to considerable variation in length, in addition to nucleotide substitutions.

The selection of taxa for phylogenetic analysis is extremely important because taxa display varying rates of DNA sequence evolution (slower in unionoideans, faster in rissooideans); differing rates in different regions (slower in the D6 flanks, faster in the D6 loop); and different proportions of substitutions versus insertion and deletions in various taxa. A single species or genus often is not representative of its higher taxon, as three examples show.

1.) *Helicina* has diverged from ancestral sequences more slowly than the other gastropods, and groups with the slowly evolving bivalves, perhaps because of retained plesiomorphies, as discussed above.

2.) *Margaritifera falcata* has eight autapomorphies, whereas *M. margaritifera* has none, and *Cumberlandia* has three (Figure 3). No other unionoidean had more than one autapomorphy. There were no convergent autapomorphies between *M. falcata* and *Cumberlandia*, but there easily could be between derived unionoidean species. Such convergences can be mistaken for synapomorphies, as we interpreted happened between *Anodonta* and some amblemines.

3.) *Truncatella* resembles *Mancinella* and *Progabbia* in having long D6 loop sequences; *Geomelania* and *Oncomelania* resemble *Busycon* in having short D6 loop sequences. From positions 65 to 74, *Truncatella sp.* shares six of eight nucleotides with *Progabbia*. The convergence is due in part to compensatory mutations in the D6 loop. *Truncatella sp.* has nine adenines between positions 72 and 82 some of which base-pair with six uracils between positions 58 and 70. *Progabbia* has a corresponding A/U rich region. The convergence is revealed as such by comparison to the sequences of close relatives. Thus, *Truncatella clathrus* shares no nucleotides with *T. sp.* in the region where *Progabbia* shares six, but it is identical in the D6 flanking regions.

Because species may not be representative of their higher taxa, and because long branches attract, a mixture of closely and distantly related species must be incorporated into phylogenetic analyses to minimize the chance of convergences remaining undetected.

If convergence is undetected, it is often revealed by comparison to relationships inferred from other data sets. In our analysis, comparisons to morphological phylogenies have identified several areas where the sequence data by themselves seem to be misleading. Problematic areas in Figure 3, the grouping of *Anodonta* with *Amblema*, *Helicina* with the bivalves, *Perotrochus* with the neogastropods, and *Geomelania* with *Oncomelania*, are not strongly supported by the sequence data, with the alternative trees being in each case only one or two steps longer. In the cases of *Helicina* and *Geomelania*, the addition of taxa is likely to change the polarity of those characters supporting the doubtful groupings. That is, the problem is not necessarily convergence, but that plesiomorphy has been mistaken for apomorphy. With *Perotrochus*, addition of taxa such as patellogastropods might also show that characters scored as autapomorphic are synapomorphic with basal taxa. This would reveal the convergence that groups *Perotrochus* with neogastropods. In the case of *Anodonta*, adding taxa would not help, but more variable sequences from another region might.

To date, sequence studies of mollusks have not overturned phylogenies based on morphology, but rather have helped in choosing among competing morphology-based hypotheses. Like morphological data, sequence data are subject to problems of convergence, unequal rates of evolution, and choice of taxa. Phylogenies based on sequence data alone can be misleading. Molecular and morphological data are often complementary, serving to define different nodes within a tree. Analyses and classifications must be based on all available data to maximize

the potential for detecting convergences and correctly resolving phylogenetic relationships.

## ACKNOWLEDGMENTS

Supported in part by NIH grant AI 11373-TMP to George M. Davis. We thank Edward Theriot and two anonymous reviewers for helpful comments on the manuscript, Caryl Hesterman and Kathleen M. Monderewicz for technical assistance, and S. Michael Phillips for providing laboratory space at the University of Pennsylvania.

## LITERATURE CITED

Albert, V. A. and B. D. Mishler. 1992. On the rationale and utility of weighting nucleotide sequence data. Cladistics 8:73–83.

Bieler, R. 1990. Haszprunar's "clado-evolutionary" classification of the Gastropoda–a critique. Malacologia 31:371–380.

Bieler, R. 1992. Gastropod phylogeny and systematics. Annual Review of Ecology and Systematics 23:311–338.

Bull, J. J., J. P. Huelsenbeck, C. W. Cunningham, D. L. Swofford, and P. J. Waddell. 1993. Partitioning and combining data in phylogenetic analysis. Systematic Biology 42: 384–397.

Crother, B. I. and W. F. Presch. 1992. The phylogeny of xantusiid lizards: The concern for analysis in the search for a best estimate of phylogeny. Molecular Phylogenetics and Evolution 1:289–294.

Davis, G. M. 1979. The origin and evolution of the gastropod family Pomatiopsidae, with emphasis on the Mekong River Triculinae. Academy of Natural Sciences of Philadelphia, Monograph 20: vii + 120 pp.

Davis, G. M. and S. L. H. Fuller. 1981. Genetic relationships among Recent Unionacea (Bivalvia) of North America. Malacologia 20:217–253.

Davis, G. M., W. H. Heard, S. L. H. Fuller and C. Hesterman. 1981. Molecular genetics and speciation in *Elliptio* and its relationships to other taxa of North American Unionidae (Bivalvia). Biological Journal of the Linnean Society 15: 131–150.

Dixon, M. T. and D. M. Hillis. 1993. Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis. Molecular Biology and Evolution 10:257–267.

Eernisse, D. J. and A. G. Kluge. 1993. Taxonomic congruence versus total evidence, and amniote phylogeny inferred from fossils, molecules and morphology. Molecular Biology and Evolution 10:1170–1195.

Emberton, K. C., G. S. Kuncio, G. M. Davis, S. M. Phillips, K. M. Monderewicz, Y. H. Guo. 1990. Comparison of Recent classifications of stylommatophoran land-snail families, and evaluation of large-ribosomal-RNA sequencing for their phylogenetics. Malacologia 31:327–352.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Systematic Zoology 27:401–410.

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

Field, K. C., G. J. Olsen, D. J. Lane, S. J. Giovannoni, M. T. Ghiselin, E. C. Raff, N. R. Pace, and R. A Raff. 1988. Molecular phylogeny of the animal kingdom. Science 239: 748–753.

Ghiselin, M. T. 1988. The origin of molluscs in light of molecular evidence. Oxford Surveys of Evolutionary Biology 5:66–95.

Gutell, R. R. and G. E. Fox. 1988. A compilation of large subunit RNA sequences presented in a structural format. Nucleic Acids Research Sequences Supplement 16:175–269.

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

Haszprunar, G. 1988a. A preliminary phylogenetic analysis of the streptoneurous gastropods. Malacological Review, Supplement 4:7–16.

Haszprunar, G. 1988b. On the origin and evolution of the major gastropod groups, with special reference to the Streptoneura. Journal of Molluscan Studies 54:367–441.

Healy, J. M. 1988. Sperm morphology and its systematic importance in the Gastropoda. Malacological Review, Supplement 4:251–266.

Hedges, S. B. 1992. The number of replications needed for accurate estimation of the bootstrap *P* value in phylogenetic studies. Molecular Biology and Evolution 9:366–369.

Hedges, S. B., K. D. Moberg and L. R. Maxson. 1990. Tetrapod phylogeny inferred from 18S and 28S ribosomal RNA sequences and a review of the evidence for amniote relationships. Molecular Biology and Evolution 7:607–633.

Hickman, C. S. 1988. Archaeogastropod evolution, phylogeny and systematics: a re-evaluation. Malacological Review, Supplement 4:17–34.

Hillis, D. M. and J. J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Systematic Biology 42:182–192.

Hillis, D. M. and M. T. Dixon. 1991. Ribosomal DNA: Molecular evolution and phylogenetic inference. Quarterly Review of Biology 66:411–453.

Kantor, Y. I. and M. G. Harasewych. 1992. Morphology of the digestive system of *Volutomitra alaskana* Dall, 1902 (Gastropoda, Pectinibranchia, Volutomitridae), with notes on the possible mechanism of feeding. Ruthenica 2:45–53.

Maddison, W. P. and D. R. Maddison. 1992. MacClade Version 3. Sinaeur Associates: Sunderland, Massachusetts. xi + 398 pp.

Mishler, B. D., K. Bremer, C. J. Humphries and S. P. Churchill. 1988. The use of nucleic acid sequence data in phylogenetic reconstruction. Taxon 37:391–395.

Petit, R. E. and M. G. Harasewych. 1990. Catalogue of the superfamily Cancellarioidea Forbes and Hanley, 1851 (Gastropoda: Prosobranchia). The Nautilus, Supplement 1, 69 pp.

Ponder, W. F. 1973. The origin and evolution of the Neogastropoda. Malacologia 12:295–338.

Ponder, W. F. and A. Warén. 1988. Classification of the Caenogastropoda and Heterostropha—A list of family-group names and higher taxa. Malacological Review, Supplement 4:288–328.

Rosenberg, G. 1989. Phylogeny and evolution of terrestriality of the Atlantic Truncatellidae (Prosobranchia: Gastropoda: Mollusca). Ph.D. thesis, Harvard University. x + 296 pp.

Smith, A. B. 1989. RNA sequence data in phylogenetic reconstruction: testing the limits of its resolution. Cladistics 5:321–344.

Sogin, M. L., H. J. Elwood, and J. H. Gunderson. 1986. Evolutionary diversity of eukaryotic small-subunit rRNA genes. Proceedings of the National Academy of Sciences, USA 83:1383–1387.

Steele, K. P., K. E. Holsinger, R. K. Jansen, and D. W. Taylor. 1988. Phylogenetic relationships in green plants—A comment on the use of 5S ribosomal RNA sequences by Bremer et al. Taxon 37:135–138.

Swofford, D. L. 1993. PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1.1. Computer program distributed by the Illinois Natural History Survey.

Swofford, D. L. and G. J. Olsen. 1990. Phylogenetic reconstruction. Pp. 411–501 *in* D. M. Hillis and C. Moritz (eds.), Molecular Systematics. Sinauer Associates: Sunderland, Massachusetts.

Tautz, D., J. M. Hancock, D. A. Webb, C. Tautz and G. A. Dover. 1988. Complete sequences of the rRNA genes of *Drosophila melanogaster*. Molecular Biology and Evolution 5:366–376.

Taylor, J. D. and N. J. Morris. 1988. Relationships of neogastropods. Malacological Review, Supplement 4:167–179.

Thiele, J. 1929. Handbuch der Systematischen Weichtierkunde. 1(1). Loricata./Gastropoda. I:Prosobranchia (Vorderkiemer). Jena: 376 pp.

Tillier, S., M. Masselot, H. Philippe and A. Tillier. 1992. Phylogénie moléculaire des Gastropoda (Mollusca) fondeé sur le séquençage partiel de l'ARN ribosomique 28 S. Comptes Rendus de l'Académie des Sciences 314:79–85.

Wenz, W. 1943. Gastropoda. Teil I: Allgemeiner Teil und Prosobranchia. Pp. 1201–1506 (1943) *in* O. H. Schindewolf, ed., Handbuch der Paläozoologie, vol. 6. Berlin.

Wheeler, W. C. 1990. Combinatorial weights in phylogenetic analysis: a statistical parsimony procedure. Cladistics 6:269–275.

Wheeler, W. and D. Gladstein. 1993. MALIGN. Version 1.73 1/4/93. Published by the authors. 28 pp.

Wheeler, W. C. and R. L. Honeycutt. 1988. Paired sequence difference in ribosomal RNA's: evolutionary and phylogenetic implications. Molecular Biology and Evolution 5:90–96.

Woese, C. R. and G. J. Olsen. 1986. Archaebacterial phylogeny: perspectives on the urkingdoms. Systematic and Applied Microbiology 7:161–177.