6

MOLECULAR SYSTEMATICS OF THE MAJOR LINEAGES OF THE GASTROPODA

Gastropod mollusks are among the oldest and most evolutionarily successful groups of animals to inhabit the earth. With origins in the Cambrian (Erwin and Signor 1991; Tracey et al. 1993), gastropods have come to inhabit nearly all marine, freshwater, and terrestrial habitats and have achieved a diversity at all taxonomic levels that is exceeded only by the Insecta. Because of their long and extensive fossil record, Gastropoda have figured prominently in paleontological studies spanning the entire Phanerozoic. By virtue of their abundance and diversity in the Recent fauna, gastropods have been used in studies spanning numerous biological disciplines, including ecology, physiology, developmental biology, population genetics, biodiversity, biomechanics, biogeography, and molecular evolution. Gastropods are also of considerable economic importance as sources of food, ornament, and pharmacological compounds, and as agricultural and mariculture pests and vectors of disease (see Cheng 1967; Faust et al. 1968; Abbott 1972; Olivera 1997, and references therein).

The higher classification of gastropods had become entrenched for much of the twentieth century, being based on an arrangement advanced by Thiele (1929–31) and only slightly modified by subsequent authors (Wenz 1938–44; Moore 1960; Boss 1982; Brusca and Brusca 1990). Thiele (1929–31) adopted Milne-Edwards's (1848) division of Gastropoda into Prosobranchia, Opisthobranchia, and Pulmonata, but further subdivided the Prosobranchia into the anagenic series Archaeogastropoda, Mesogastropoda, and Stenoglossa (renamed Neogastropoda by Wenz 1938–1944). Thorough, historical reviews of gastropod classification are provided in Bieler (1992) and Ponder and Lindberg (1997).

The dramatically divergent classification of prosobranch gastropods proposed by Golikov and Starobogatov (1975) catalyzed a renewed interest in gastropod phylogeny. This interest was fueled by the discoveries of numerous new, higher taxa, especially at hydrothermal vents and sulphide seeps; the rapid increase in the amount of new data provided by new morphological and molecular techniques; and, perhaps most importantly, by the application of increasingly rigorous methodologies for analyzing the data and generating phylogenetic hypotheses. Although discrepancies remain among phylogenies based on morphological (e.g., Salvini-Plawen 1980; Haszprunar 1988; Ponder and Lindberg 1996, 1997), DNA–RNA sequence (e.g., Tillier et al. 1992, 1994; Rosenberg et al. 1994, 1997; Winnepenninckx et al. 1994, 1998; Harasewych et al. 1997a, b, 1998; McArthur and Koop 1999; Colgan et al. 2000; Harasewych and McArthur 2000), and paleontological (e.g., Wagner 1995; Bandel 1997) datasets, these recent studies converge on a broad outline of gastropod evolution that differs significantly from those advocated by Thiele and Wenz. Virtually all modern classifications divide the Gastropoda into the following monophyletic groups: Patellogastropoda, Cocculinoidea, Lepetelloidea, Neritopsina, Neomphalina, Vetigastropoda (generally, but not always including the Pleurotomarioidea), Caenogastropoda, and Heterobranchia. However, differences arise regarding the rank and relationships among these clades (see Figure 6.1).

Some features, such as the basal position of the patellogastropod limpets within Gastropoda, which had been recognized in earlier classifications (e.g., Lister 1678; Troschel 1861; Pelseneer 1906), have been reaffirmed by morphological (e.g., Golikov and Starobogatov 1975; Haszprunar 1988; Ponder and Lindberg 1997; Sasaki 1998) and molecular (e.g., Harasewych et al. 1997a; Harasewych and McArthur 2000) studies. Similarly, the monophyly of the Apogastropoda and the sister-group relationship of Caenogastropoda and Heterobranchia are strongly confirmed by multiple datasets. However, the relationships of a number of basal taxa, among them the cocculiniform superfamilies Cocculinoidea and Lepetelloidea, the vetigastropod superfamily Pleurotomarioidea, and the order Neritopsina, are poorly resolved. Their positions have been strongly influenced by such factors as data type (morphological versus molecular), taxon sampling, and outgroup selection.

Over the last decade, many studies have been published in which sequence data have been used to investigate phylogenetic relationships within Gastropoda at varying hierarchical levels (e.g., Tillier et al. 1992, 1994; Rosenberg et al. 1994, 1997; Winnepenninckx et al. 1994, 1996, 1998; Harasewych et al. 1997a, 1997b, 1998; Colgan et al. 2000; Harasewych and McArthur 2000). In general, these studies have readily identified the major gastropod lineages (outlined
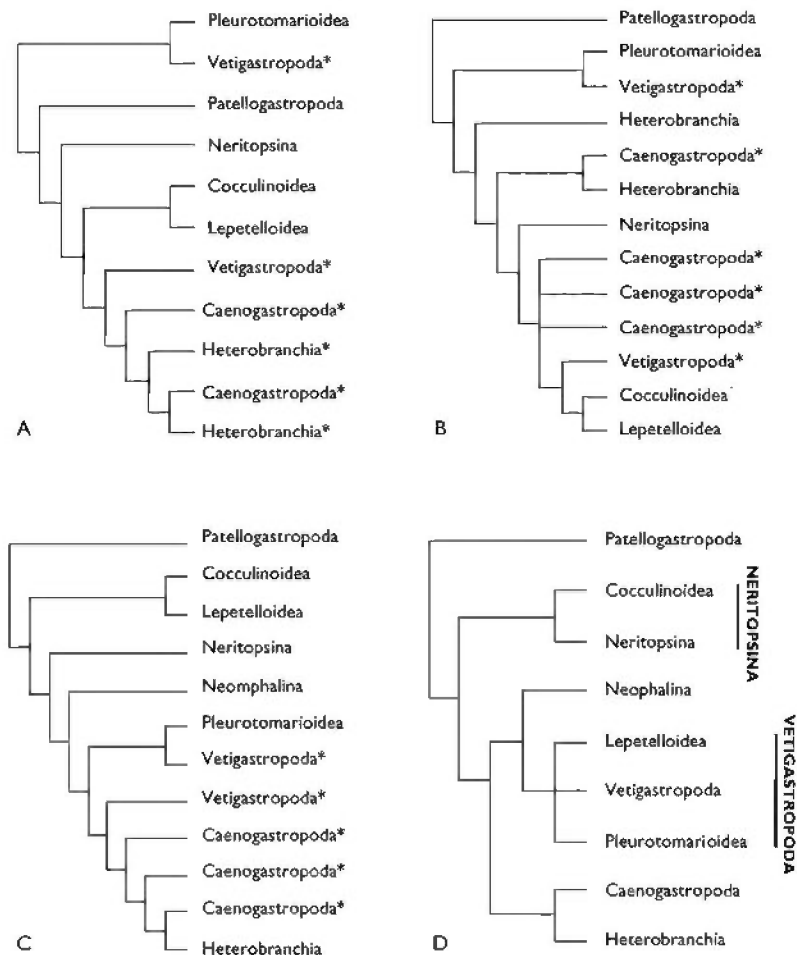
Figure 6.1. Morphology-based classifications of Gastropoda, with trees drawn to represent relationships among the nine clades (see text). Taxon names and ranks have been modified for the sake of uniformity and clarity. Hypothesis of nonmonophyly for the modern major groupings are marked with an asterisk. (A) Thiele (1929–31), represented as a phylogenetic tree (modified from Ponder and Lindberg 1996). (B) Golikov and Starobogatov (1975). (C). Haszprunar (1988). (D). Ponder and Lindberg (1997).

above), but have had much difficulty in resolving basal relationships. In this chapter, we examine the growing consensus concerning the major lineages of the Gastropoda and use the largest alignment of 18S (small subunit) ribosomal RNA sequences yet assembled as an independent assessment of confidence.

EXAMINING CONSENSUS USING SUPERTREE ANALYSIS

With so many independent studies, each using a different dataset, it is difficult to obtain an overall view of our current understanding of gastropod phylogeny. In particular, it is often hard to compare the results of molecular phylogenetic investigations (based on sequences from individual specimens) and parsimony analysis of anatomical characters (often based on summaries of character states for higher-order taxonomic groups). Even among studies using the same kind of data, differences in taxonomic sampling, the exact characters considered, and how the analysis was performed can make comparison difficult. Table 6.1 shows the data used by a metaanalysis of 11 studies of gastropod phylogeny, ranging from the precladistic hypotheses of Thiele (1929–31) and Golikov and Starobogatov (1975) to the recent molecular phylogenetic investigation of Colgan et al. (2000). We used the supertree method of Baum (1992) and Ragan (1992) (for a review, see Sanderson et al. 1998), in which the phylogenetic trees of each study are recoded as new metadata by the matrix representation using parsimony (MRP) method. In essence, the new metadata matrix reflects the presence or absence (or lack of data) of all of the possible clades of gastropod groups found in these studies. The resulting matrix was subjected to a cladistic analysis under parsimony to provide a consensus supertree in which shared results among original studies are recovered and differences in branching order among the original studies are resolved using the most parsimonious explanation. The advantage of this method is that the studies do not have to be compatible—each can use a different selection of taxa. For example, the detailed neontological study of Ponder and Lindberg (1997) includes the architaenioglossan families Cyclophoridae and Ampullariidae, whereas the molecular study of McArthur and Koop (1999) includes representatives of Ampullariidae and Viviparidae.

The metadata were compiled by recoding each of the original phylogenetic hypotheses to the taxonomic names used by Ponder and Lindberg (1997), with the addition of a few taxa present in one or more of the other studies and some changes to more inclusive higher-order taxa. Because some molecular studies had multiple representatives of a single taxonomic group (e.g., multiple representatives of the Trochidae), we trimmed these phylogenetic trees to single representatives, albeit with some difficult choices, when molecular trees did not support monophyly. Some taxonomic groups were not included because they were not informative about overall gastropod phylogeny. The resulting dataset included 33 taxa and 151 parsimony-informative characters. Because not all of the original studies included bootstrapping or other measures of internal con-

Table 6.1

Sources of data used for the supertree metaanalyses

Source	Type of Data	Type of Analysis	Number of MRP Characters
Thiele (1929–31)	Morphology	Nonphylogenetic	16
Golikov and Starobogatov (1975)	Morphology	Nonphylogenetic	19
Haszprunar (1988)	Anatomy	Cladistic	17
Ponder and Lindberg (1997)	Anatomy and Morphology	Phylogenetic	30
Tillier et al. (1994)	28S rDNA	Phylogenetic	9
Harasewych et al. (1997a)	18S rDNA	Phylogenetic	12
Harasewych et al. (1998)	18S rDNA	Phylogenetic	11
Winnepenninckx et al. (1998)	18S rDNA	Phylogenetic	6
McArthur and Koop (1999)	28S rDNA	Phylogenetic	14
Colgan et al. (2000)	28S rDNA and Histone H3	Phylogenetic	16
Harasewych and McArthur (2000)	18S rDNA	Phylogenetic	11

Note: Trees were recoded as binary characters by the matrix representation using parsimony (MRP) method (Baum 1992; Ragan 1992).

fidence, each of the new metadata characters was given equal weight in the final analysis.

We subjected the metadata to analysis under the parsimony optimality criteria using the computer program PAUP* (Swofford 2001). We used 100 random taxa addition replicates with tree bisection and reconnection (TBR) branch swapping to find the set of most parsimonious trees. We assessed confidence of the results by performing bootstrap analysis on 100 random bootstrap replicate datasets, with 10 random taxa addition replicates per bootstrap dataset and a limit of 100 trees in memory for TBR swapping. Eighty equally parsimonious reconstructions of the metadata were found; the majority-rule consensus tree is presented in Figure 6.2. The resulting phylogenetic hypothesis was quite conventional, with Ponder and Lindberg's (1996) hypothesized basal position of the Patellogastropoda and monophyly of all of the major groups, although this was assumed for the Patellogastropoda (based on Harasewych and McArthur 2000), Neritopsina (consistent with all of the studies), Neomphalina (based on McArthur and Koop 1999), and Pulmonata. As with recent studies, bootstrapping revealed negligible support for the basal branching order of the Gastropoda. Monophyly of groups such as the Cocculiniformia (Cocculinoidea + Lepetelloidea), Caenogastropoda, Vetigastropoda, and Heterobranchia was not statistically supported. However, there was strong support for monophyly of the Apogastropoda (Caenogastropoda + Heterobranchia) and a clade containing the Architectonicoidea, Pulmonata, and a monophyletic Opisthobranchia. Resolu-
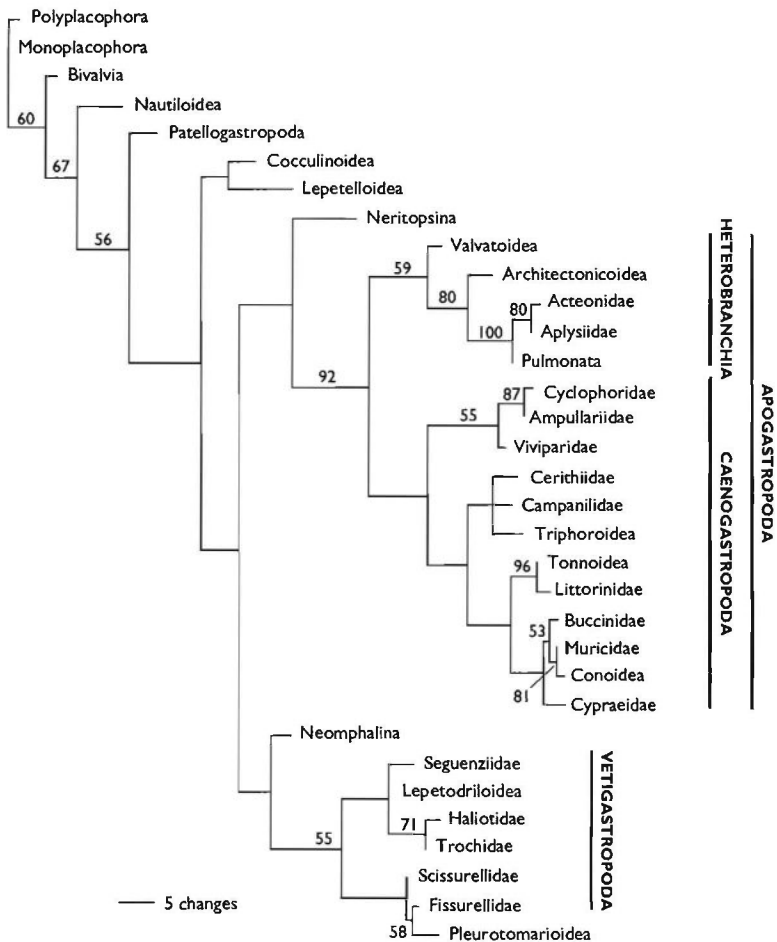
Figure 6.2. Majority-rule consensus tree of the 80 most parsimonious supertrees found using all of the studies listed in Table 6.1. Bootstrap proportions where higher than 50% are shown.

tion of evolutionary relationships within the Prosobranchia did not improve when the studies of Thiele (1929–31), Golikov and Starobogatov (1975), and Haszprunar (1988) were excluded and the data reanalyzed (Figure 6.3). This second result reflected common themes found in molecular studies: unexpected placement of the root of the Gastropoda (Neritopsina as the basal clade?), lack of resolution of the prosobranch grade, yet good support for monophyly of the Apogastropoda, Caenogastropoda, and Heterobranchia. As did the molecular studies, this method suggested shared ancestry of the Cocculinoidea, Lepetelloidea, and Patellogastropoda.
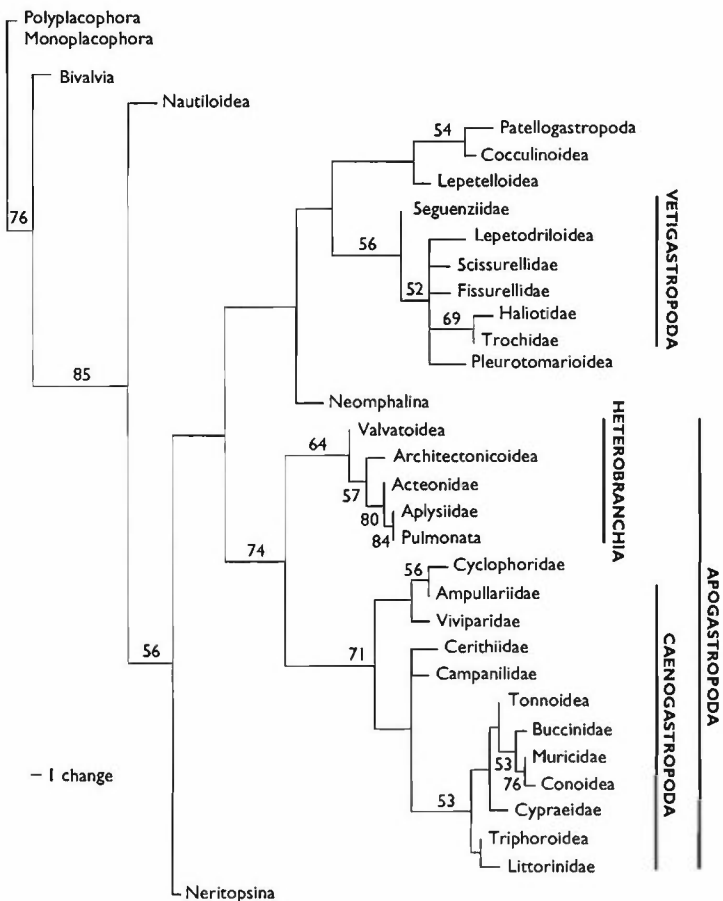
Figure 6.3. Majority-rule consensus tree of the 20 most parsimonious supertrees found using all the studies listed in Table 6.1, except those of Thiele (1929–31), Golikov and Starobogatov (1975), and Haszprunar (1988). Bootstrap proportions where higher than 50% are shown.

A PRELIMINARY BAYESIAN PERSPECTIVE

Molecular investigations of gastropod phylogeny have mainly used the DNA–RNA sequences of the small (18S) and large (28S) subunits of ribosomal RNA genes. Every gene examined in molecular systematics has its limits. The consistently poor resolution of early gastropod phylogeny suggests that ribosomal sequences have a difficult time "reaching back" to the early Paleozoic and Cambrian. This is contrary to expectation, because these gene sequences have been used successfully to examine the Tree of Life and have helped elu-

cidate the existence of the three domains of life—the Bacteria, Archaea, and Eukaryota (Woese et al. 1990; Embley et al. 1994). Our lack of significant progress in the Gastropoda is more likely a reflection of sampling effort. Outside of the Euthyneura, most gastropod rRNA sequences are partial. Statistical resolution improves with the number of characters sequenced. Similarly, representative taxon sampling can be very important in resolving major phylogenetic patterns because crucial taxa can split long internal branches, reducing artifacts and providing more phylogenetic signal about early evolutionary events. We have attempted to address these concerns by building as large an 18S rDNA dataset as possible. To that end, an 18S rDNA alignment of 163 gastropod taxa has been constructed (available on request, including source information). This alignment includes all partial and full-length 18S rRNA sequences available in GenBank, aligned according to the secondary structure model of the Ribosomal Database Project (Olsen et al. 1992) using an iterative application of ClustalW (Thompson et al. 1994) and subsequent manual editing using MacClade (Maddison and Maddison 2000). The alignment also includes 21 new sequences (Table 6.2) determined using the methods outlined in Harasewych and McArthur (2000). These new sequences attempt to improve the taxon sampling throughout the Prosobranchia, although most were partial sequences because preserved specimens were used.

To incorporate maximum likelihood, phylogenetic uncertainty, and easy examination of competing hypotheses, we analyzed this large alignment using a Bayesian statistical procedure, as implemented by the computer program MrBayes (Huelsenbeck and Ronquist 2001). MrBayes performs a Metropolis-coupled Markov chain Monte Carlo (MCMCMC) estimation of posterior probabilities (see Shoemaker et al. 1999; Huelsenbeck et al. 2001; Lewis 2001b). Bayesian methods have the advantage of incorporating maximum likelihood, complex substitution models, and fast analyses. Their goal is not to find the best tree, but to instead sample the "cloud of best trees." Bayesian methods recognize that there is uncertainty in phylogenetic analyses owing to the finite amount of data sampled. Instead of finding the single best tree (a very difficult procedure), Bayesian methods instead use their sampling of "best trees" to estimate posterior probabilities of relationships between taxa. Any possible clade can be assigned an easy-to-interpret posterior probability. For example, Bayesian methods make it easy to ask, "What is the probability that the sister taxon to the Caenogastropoda is the Heterobranchia versus the probability that it is the Neritopsina?" The consensus of all trees sampled from the "cloud of best trees" provides an overview of the most probable phylogeny. In practice, this is often the same tree found using maximum likelihood heuristic searches. As our dataset was too large for maximum likelihood heuristic searching, we performed MCMCMC estimation of posterior probabilities using noninformative prior

Table 6.2

New 18S rDNA sequences included in this study, localities, preservation means, vouchers, sequence lengths, and GenBank accession numbers

Taxon	Specimen Details	GenBank Accession No.
Caenogastropoda		
Viviparus georgianus	Lake Talquin, Tallahassee, Florida (Frozen, USNM 1003901)	AY090794 (1797 bp)
Cocculinoidea		
Coccopigya hispida	On wood, off Cape Palliser, 41°45.2′S 175°26.8′E, 1,039–1,077 m (EtOH, Marshall 87048)	AY090795 (537 bp)
Cocculina messingi	On deployed wood, Bahamas, 26°37.30′N 78°58.55′W, 1,372 ft. (Frozen, USNM 888655)	AY090796 (1752 bp)
Heterobranchia		
Rissoella caribaea	Fiesta Key, Florida (Frozen, USNM 881221)	AY090797 (2123 bp)
Lepetelloidea		
Caymanabyssia fosteri	On deployed wood, 11°51.00′N 103°50.00′W, East Pacific Rise, 2,700 m (EtOH, USNM 784765)	AY090798 (145 bp)
Copulabyssia gradata	On wood, off Cape Egmont 38°58.5′S 172°10.2′E, 1,045–1,055 m (EtOH, USNM 888730)	AY090799 (530 bp)
Mesopelex zelandica	On kelp holdfast, Chatham Rise, 42°53′S 176°04′E, 370–420 m (EtOH, Marshall 118916)	AY090800 (547 bp)
Pyropelta musaica	Hydrothermal vents, 45°57.00′N 130°01.00′W, Juan de Fuca Ridge, 1,546 m (EtOH, USNM 858229)	AY090801 (149 bp)
Tentaoculus haptricola	On algal holdfast, off Chatham Island, 42°50′S 176°30′W, 945 m (EtOH, USNM 888731)	AY090802 (141 bp)
Neomphalina		
Cyathermia naticoides	Hydrothermal vents, 20°49.9′N 109°06.0′W, East Pacific Rise, 2,615 m (Frozen, Lutz A2232)	AY090803 (526 bp)
Depressigyra globulus	Hydrothermal vents, 44°59.43′N, 130°12.08′W, Juan de Fuca Ridge, 2,249 m (Frozen, Tunn. HYS202)	AY090804 (561 bp)
Melanodrymia aurantiaca	Hydrothermal vents, 20°47.0′N 109°08.9′W, East Pacific Rise, 2,577 m (Frozen, Lutz A2233)	AY090805 (490 bp)
Neomphalus fretterae	Hydrothermal vents, Oyster Bed and Garden of Eden, Galapagos Rift (EtOH, USNM 784638)	AY090806 (576 bp)
Peltospira operculata	Hydrothermal vents, 00°48′N 86°13′W, Galapagos Rift, 2,462 m (Frozen, Lutz A2010)	AY090807 (538 bp)
Symmetromphalus regularis	Hydrothermal vents, 18°12.36′N 144°42.24′E, Mariana back-arc basin, 3,640 m (EtOH, USNM 784763)	AY090808 (538 bp)
Pleurotomarioidea		
Entemnotrochus adansonianus	Guadeloupe (Frozen, USNM 888647)	AY090809 (1993 bp)

Table 6.2 continued

Taxon	Specimen Details	GenBank Accession No.
Vetigastropoda		
Bathymargarites symplector	Hydrothermal vents, 20°49.9'N 109°06.0'W, East Pacific Rise, 2,615 m (Frozen, Lutz A2232)	AY090810 (534 bp)
Cittarium pica	Jamaica (Frozen, USNM 888661)	AY090811 (403 bp)
Lepetodrilus fucensis	Hydrothermal vents, 40°58'N 129°05.5'W, Juan de Fuca Ridge, 2,125 m (Frozen, Tunn. F20-A2413)	AY090812 (529 bp)
Sinezona confusa	Long Key, Florida (Frozen, USNM 888716)	AY090813 (530 bp)
Temnocinclis euripes	Hydrothermal vents, 44°56'N 130°15'W, Juan de Fuca Ridge, 2282 m. (EtOH, Tunn. A2078-1452)	AY090814 (541 bp)

Note: Sequences for *Cocculina messingi*, *Rissoella caribaea*, and *Entemnotrochus adansonianus* are extensioos of previously published sequences.

Preservation method: EtOH = fixation io formalin, followed by storage io ethanol.

Source codes: Lutz = collection of R. Lutz, Rutgers University, U.S.A.; Marshall = collection of B. Marshall, Museum of New Zealand, New Zealand; Tunn. = collection of V. Tunnicliffe, University of Victoria, Canada; USNM = Mollusk collectioo, Natiooal Museum of Natural History, Smithsonian Institution.

probabilities, the TrN+I+Γ substitution model with inclusion of unequal nucleotide frequencies (see Swofford et al. 1996), and four incrementally heated Markov chains. The TrN+I+Γ substitution model was selected using the Likelihood Ratio Test as implemented in the computer program ModelTest (Posada and Crandall 1998). The Markov chains were run for 150,000 generations, with sampling of topologies every 100 generations. Posterior probabilities of topologies, clades, and parameters were estimated from the sampled topologies after removal of MCMCMC burn-in. We included all taxa in the alignment, except that we used representative taxon sampling for the Caenogastropoda and Heterobranchia, because both of these groups included a heavy sampling of closely related genera. Regions of poor or uncertain alignment were excluded from analyses and the final dataset included 81 taxa and 1,431 characters (899 of which were constant). Based on our experience with these data, we did not include a nongastropod outgroup because the preponderance of partial sequences at the base of the Gastropoda make rooting extremely unreliable. Instead, we used the Patellogastropoda as a visual root when presenting trees. As such, all discussion of monophyletic and sister relationships from our results should be treated as conditional on the final placement of the root of the Gastropoda.

A consensus of posterior probabilities as determined by Bayesian analysis
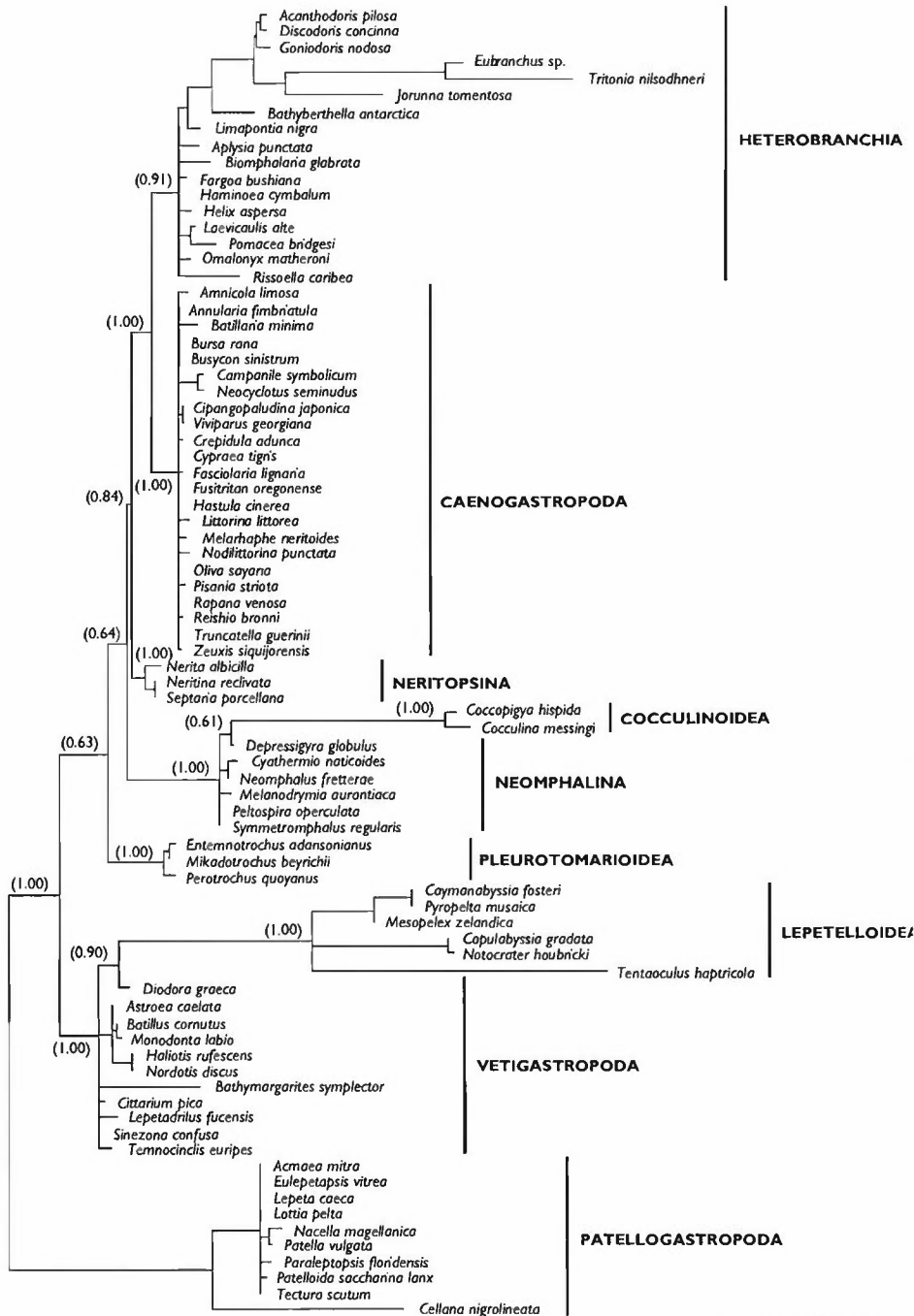
Figure 6.4. Majority-rule consensus tree of trees sampled from the "cloud of best trees" by MCMCMC (after removal of burn-in). Such a sampling is representative of posterior probabilities, which are given in parentheses for the major clades and a few internal nodes related to the placement of the Cocculinoidea and Lepetelloidea.

of the 81 taxa alignment is shown in Figure 6.4. Monophyly of the Heterobranchia, Caenogastropoda, Apogastropoda, Neritopsina, and Patellogastropoda was supported. In contrast, the Cocculiniformia was divided into two groups — a monophyletic Cocculinoidea appeared as a member of the hydrothermal vent Neomphalina and a monophyletic Lepetelloidea appeared as a member of the Vetigastropoda. Both of these results had strong support. As found previously, the Pleurotomarioidea branched separately from the Vetigastropoda. Unfortunately, the sampling used to determine these posterior probabilities was very small. The MCMCMC was only able to run for 150,000 generations before MrBayes ran out of memory, and the first 100,000 generations had to be discarded as burn-in. Only 500 trees were sampled. As in heuristic searching, in which the number of random addition replicates needed to find the best tree is unknown, the number of generations needed by MCMCMC to obtain an unbiased estimate of posterior probabilities is also unknown. However, because our sampling ran only half as long as the burn-in, the posterior probabilities presented in Figure 6.4 are almost certainly inaccurate estimates. As such, we must view Figure 6.4 as a preliminary Bayesian estimate of gastropod phylogeny.

REPRESENTATIVE TAXON SAMPLING

Although our goal in using Bayesian methods was to obtain an estimate of gastropod phylogeny using the largest possible sampling of taxa, we ran aground of computational limits. As such, we decided to take a smaller, representative sampling of gastropod phylogeny from our 163 taxon alignment and examine gastropod phylogeny under maximum parsimony, minimum evolution, maximum likelihood, and Bayesian perspectives. Based on our broader Bayesian analysis (Figure 6.4), we sampled representative sequences from the Patellogastropoda, Vetigastropoda, and Apogastropoda and included all sequences available for the Lepetelloidea, Pleurotomarioidea, Neomphalina, Cocculinoidea, and Neritopsina. In all, 33 taxa and 1,431 characters (1,070 constant) were included. We searched for the best tree under the maximum likelihood optimality criterion using 10 random taxon addition replicates, TBR branch swapping, and the TrN+I+Γ substitution model with inclusion of unequal nucleotide frequencies (determined by ModelTest). Bootstrap measures of internal confidence were determined using 100 bootstrap replicates with 10 (maximum parsimony / minimum evolution) or 2 (maximum likelihood) random addition

Figure 6.4 continued
Branch lengths represent the amount of evolutionary change. The posterior probabilities shown are probably inaccurate estimates caused by the overly short Markov chain lengths (see text).

replicates each. All searches and bootstrapping were restricted to 20 trees in memory during branch swapping. In addition, we performed Bayesian estimation of posterior probabilities using noninformative prior probabilities, the TrN+I+Γ substitution model with inclusion of unequal nucleotide frequencies, and four incrementally heated Markov chains. The Markov chains were run for a million generations, with sampling of topologies every 100 generations. Posterior probabilities of topologies, clades, and parameters were estimated from the sampled topologies after removal of MCMCMC burn-in.

The best tree found under maximum likelihood (Figure 6.5) agreed in general with the tree found with the broader Bayesian analysis. There was strong bootstrap and posterior probability support for the major clades, with the exception of the Vetigastropoda/Lepetelloidea grouping. All analyses found strong support for the Apogastropoda, but disagreed on the Neritopsina being its sister taxa (the maximum likelihood bootstrap value was low). There was strong support under maximum likelihood and by posterior probabilities that the Cocculinoidea shared common ancestry with the hydrothermal vent Neomphalina. The placement of the Lepetelloidea within the Vetigastropoda is merely an unsupported suggestion because bootstrap values and posterior probabilities were negligible for the placement of vetigastropod taxa, although support was strong for independent origins of the Pleurotomarioidea.

The combination of new discoveries such as the novel Neomphalina from hydrothermal vents, new tools such as electron microscopy and DNA sequencing, and new perspectives such as parsimony and maximum likelihood have all combined to revolutionize our understanding of gastropod phylogeny. As shown in our metaanalyses, neontological and molecular data do not conflict for much of gastropod phylogeny. We have clear evidence that traditional groupings such as the Prosobranchia, Archaeogastropoda, Mesogastropoda, and Streptoneura are phylogenetically meaningless or are grades of organization. We now understand that a great deal of gastropod diversity, including marine, terrestrial, and aquatic forms with a variety of lifestyles is the product of a single phylogenetic lineage—the Apogastropoda. This clade, with its internal sister relationship between the Heterobranchia and Caenogastropoda, finds strong support in all neontological and molecular studies. However, neontological and molecular studies are not in agreement on the remainder of gastropod phylogeny. The most notable difference is that neontological studies find resolved (i.e., bootstrap support) branching patterns for basal gastropod phylogeny (e.g., Ponder and Lindberg 1997), whereas molecular studies often resolve the major clades, but cannot resolve the relationships among them. The differences in basal branching pattern and consistent poor basal bootstrap support in our two supertrees (Figures 6.2 and 6.3) are due in part to introduction of basal phylogenetic noise by molecular studies. This noise, within the molecular studies
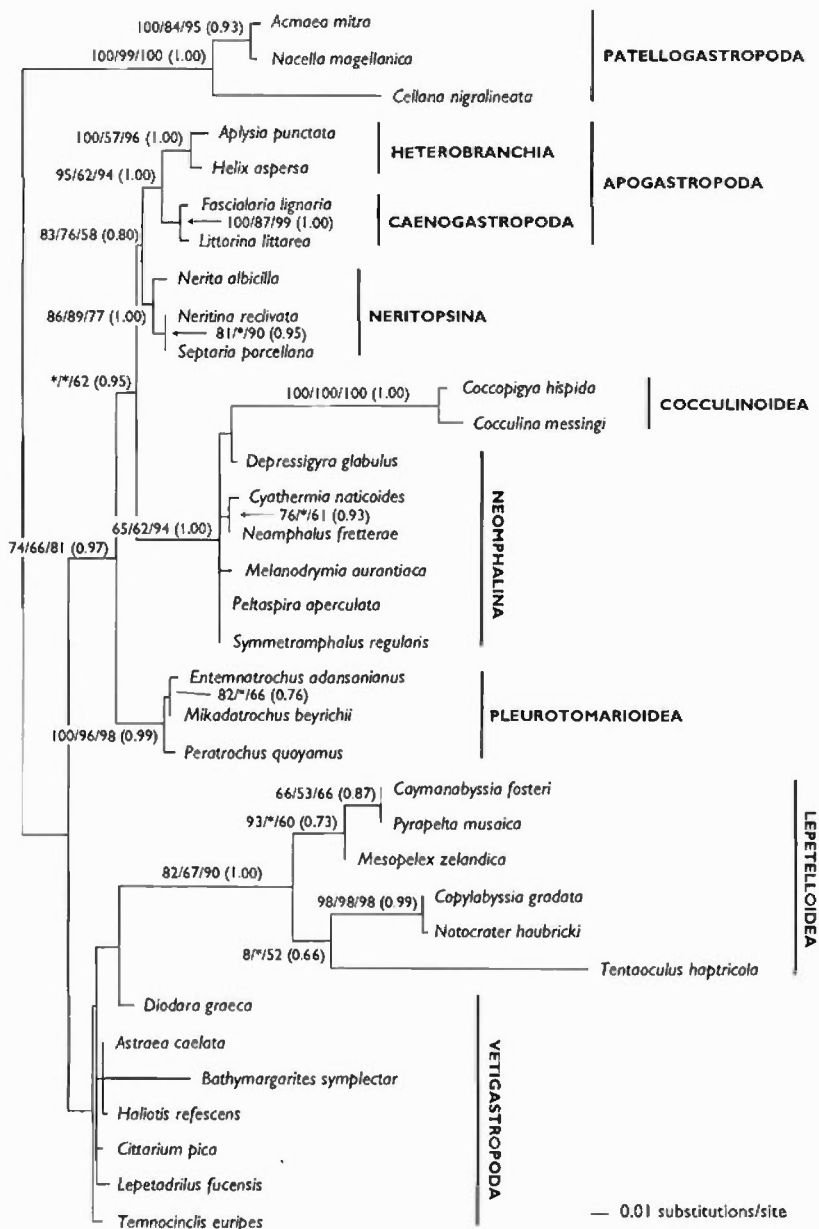
Figure 6.5. Best maximum likelihood tree found for the representative sampling of the Gastropoda. Bootstrap values are shown where higher than 50% (maximum parsimony / minimum evolution / maximum likelihood). Bayesian posterior probabilities are given in parentheses where higher than 50%. * represent bootstrap values lower than 50%. Branch lengths represent amount of evolutionary change.

themselves or in our supertree analyses, produces phylogenetic artifacts such as the basal placement of the Neritopsina (Figure 6.3). Most notably, basal position of the Patellogastropoda finds considerable, very believable support in studies of gastropod anatomy and ultrastructure (e.g., Ponder and Lindberg 1996, 1997), but long branch problems essentially make the Patellogastropoda a rogue taxon in ribosomal sequence investigations (McArthur and Koop 1999; Harasewych and McArthur 2000). Thus, molecular investigations of gastropod phylogeny need to pay careful attention to sources of variation, noise, and bias.

Molecular investigations of gastropod phylogeny provide a very powerful tool for independent assessment of neontological hypotheses. Our preliminary Bayesian investigation and thorough maximum likelihood analysis confirm the monophyly of the Patellogastropoda, Neritopsina, Cocculinoidea, Lepetelloidea, Pleurotomarioidea, Caenogastropoda, Heterobranchia, and Apogastropoda. Use of 18S rDNA sequences confirms the earlier finding of McArthur and Koop (1999) that the hydrothermal vent endemic Neomphalina represents a major gastropod lineage, but additionally supports common ancestry with the Cocculinoidea, another group of deep-sea sulphophiles known from hydrothermal vents, hydrocarbon seeps, whale falls, and sunken wood. We also find considerable evidence for origins of the Lepetelloidea independent of the Cocculinoidea, with a statistically unsupported association of the Lepetelloidea with the Vetigastropoda. The lack of support for monophyly of the Vetigastropoda (less the Pleurotomarioidea) contrasts with previous molecular investigations and illustrates some concerns with these data. Long internal branches separate the Patellogastropoda, Cocculinoidea, and Lepetelloidea from their closest relatives. Long branches can obscure phylogenetic history through introduction of localized noise, although maximum likelihood handles this artifact quite well (Swofford et al. 1996). Although the Cocculinoidea does not appear to be acting as a rogue taxon (i.e., high support for monophyly of Neomphalina + Cocculinoidea), the combination of very long branches in the Lepetelloidea and drastic loss of support for the Vetigastropoda makes the association of these two taxa suspect. The separation of the Pleurotomarioidea from the Vetigastropoda has been found in previous molecular investigations (e.g., Harasewych and McArthur 2000). The Pleurotomarioidea has novel 18S rDNA inserts of considerable size associated with elevated sequence variation in conservative regions of the alignment (Harasewych et al. 1997a; Harasewych and McArthur 2000). On a smaller scale, the same is true for the Patellogastropoda and Lepetelloidea (Harasewych and McArthur 2002; alignment used in this study). Models of nucleotide substitution used by maximum likelihood analyses assume that the rules of evolution are homogeneous for all lineages within the sampled phylogenetic tree (Swofford et al. 1996). The opposite-elevated

or differing patterns of variation in some of the sampled lineages (covarion pattern of variation) can confuse tree reconstruction considerably (Lockhart et al. 1998). Thus, although the Pleurotomarioidea does not exhibit long branches, examinations of the rDNA alignments suggest it may be contributing a covarion structure to the data (as may be the Patellogastropoda and Lepetelloidea in association with long branch problems). Although Ponder and Lindberg's (1997) detailed investigation of gastropod anatomy supported a monophyletic Vetigastropoda that included the Lepetelloidea and Pleurotomarioidea, a covarion structure in the rDNA data may be degrading our ability to accurately examine the Vetigastropoda and its relatives.

Consistent with other molecular investigations, we do not find a clear resolution of basal gastropod phylogeny. Although Bayesian methods hold promise for future resolution of basal gastropod phylogeny, we were unable to run sufficiently long Markov chains in our dataset with a large taxon sampling (Figure 6.4). We suspect that the lack of support for the Vetigastropoda + Lepetelloidea in the maximum likelihood analysis (Figure 6.5) could also in part be the product of restricted taxon sampling. Overall, there was a general agreement between maximum likelihood bootstrap values and Bayesian posterior probabilities in the smaller taxon sampling (Figure 6.5), with the exception of two of the deeper nodes in the tree (position of the Neritopsina and placement of the Neomphalina + Cocculinoidea). Posterior probabilities support a sister relationship between the Neritopsina and the Apogastropoda, with this grouping in a sister relationship with the Neomphalina + Cocculinoidea clade. There is no current understanding of the relationship between bootstrap values and posterior probabilities. That the two are in conflict for these two hypothetical sister relationships means these relationships should be taken as preliminary hypotheses testable by additional data and future advances in our understanding of the bootstrap–posterior probability relationship.

Instead of focusing on the search for the best tree, molecular systematic investigations can often be more fruitful if they are considered opportunities to examine important hypotheses. One of the advantages of Bayesian methods is that posterior probabilities can be determined for any number of phylogenetic hypotheses. A series of important hypotheses and their associated posterior probabilities are presented in Table 6.3. These data do not support monophyly of the Cocculiniformia (Cocculinoidea + Lepetelloidea; $p = 0.00$), nor their joint association with another major lineage, such as the Patellogastropoda ($p = 0.00$) or Neomphalina ($p = 0.00$). Although the Cocculinoidea appear to share common ancestry with the Neomphalina ($p = 1.00$), the data are equivocal on whether this could be a sister relationship ($p = 0.38$). These data also reject common ancestry of the Neomphalina and Vetigastropoda ($p = 0.00$). In fact, any

Table 6.3

A Bayesian look at several alternative phylogenetic hypotheses

Hypothesis	Posterior Probability [a,b]
Monophyly of Cocculinoidea + Lepetelloidea	0.00
Monophyly of Cocculinoidea + Neomphalina (exists in Figure 6.5)	1.00
Cocculinoidea as sister taxon to the Neomphalina only	0.38
Monophyly of Lepetelloidea + Vetigastropoda (exists in Figure 6.5)	0.30
Lepetelloidea as sister taxon to the Vetigastropoda only	0.17
Monophyly of the Vetigastropoda (regardless of Lepetelloidea)	0.17
Monophyly of Vetigastropoda + Pleurotomarioidea	0.00
Monophyly of Patellogastropoda + Cocculinoidea + Lepetelloidea	0.00
Monophyly of Neomphalina + Vetigastropoda	0.00
Monophyly of Cocculinoidea + Neritopsina	0.00
Monophyly of Neomphalina + Cocculinoidea + Lepetelloidea	0.00

[a]Posterior probabilities are from the Bayesian analysis shown in Figure 6.5.
[b]Rounded to two decimal places.

hypothesis that does not place the Lepetelloidea, Vetigastropoda (not including Pleurotomarioidea), and Patellogastropoda basal to all other gastropod lineages is strongly rejected ($p = 0.97$, maximum likelihood bootstrap = 81). However, because Bayesian approaches are a recent introduction to molecular systematics and many aspects of their strengths, weaknesses, and biases have yet to be investigated, we should exercise some caution in interpreting these results.

It should be noted that the long branches associated with the Lepetelloidea in our analyses are in part due to use of some very short sequences (Table 6.2). Short sequences increase error in branch length estimates. Resolution of the position of the Lepetelloidea, plus improved resolution of basal relationships, could improve with use of longer sequences. Overall, our large 18S rDNA alignment contained little variation for the taxa studied (532 variable sites for Figure 6.4, 361 variable sites for Figure 6.5). Given the dramatically decreased cost of DNA sequencing and development of long PCR techniques, future studies should attempt to amplify and sequence the entire ribosomal operon when examining overall gastropod phylogeny. Combined analysis of both large and small subunit ribosomal sequences is proving quite powerful in resolving deep animal relationships (Medina et al. 2001; Mallatt and Winchell 2002), so amplifying the entire operon would be the most efficient and cost effective way to obtain these data. Statistical and probabilistic approaches to gastropod phylogeny are proving to be powerful and, as our emphasis shifts from finding the very best tree to recognizing phylogenetic uncertainty, we are certain to expand

our use of morphological and molecular data for phylogenetic hypothesis testing. For example, Lewis (2001a) recently described a maximum likelihood method for discrete morphological characters that could be applied to our rich knowledge of gastropod morphology, anatomy, and ultrastructure. Because statistical and model-based methods have associated error, it will be important to minimize random error by using the longest sequences and largest datasets possible. Given the expanding use of diverse approaches such as electron microscopy, examination of development, DNA sequencing, comparative genomics, and broad taxon sampling, the vigorous malacological research community is sure to rise to the challenge.

ACKNOWLEDGMENTS

REFERENCES

Abbott, R. T. 1972. Kingdom of the seashell. Crown Publishers, New York.

Bandel, K. 1997. Higher classification and pattern of evolution of the Gastropoda. Courier Forschungsinstitut Senckenberg 201:57–81.

Baum, B. R. 1992. Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. Taxon 41:3–10.

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

Boss, K. 1982. Mollusca. Pp. 945–1161 in Synopsis and Classification of Living Organisms (S. P. Parker, ed.). McGraw Hill, New York.

Brusca, R. C., and G. J. Brusca. 1990. Invertebrates, Sinauer Associates, Sunderland, MA.

Cheng, T. C. 1967. Marine Molluscs as Hosts for Symbiosis: With a Review of Known Parasites of Commercially Important Species. Academic Press, New York.

Colgan, D. J., W. F. Ponder, and P. E. Eggler. 2000. Gastropod evolutionary rates and phylogenetic relationships assessed using partial 28S rDNA and histone H3 sequences. Zoologica Scripta 29:29–63.

Embley, T. M., R. P. Hirt, and D. M. Williams. 1994. Biodiversity at the molecular level: The domains, kingdoms and phyla of life. Philosophical Transactions of the Royal Society of London, B 345:21–33.

Erwin, D. H., and P. W. Signor. 1991. Extinction in an extinction–resistant clade: The evolutionary history of the Gastropoda. Pp. 152–160 in The Unity of Evolutionary Biology. Proceedings of the 4th International Congress of Systematic and Evolutionary Biology, 1 (E. C. Dudley, ed.). Dioscorides Press, Portland, OR.

Faust, E. C., P. C. Beaver, and R. C. Jung. 1968. Animal Agents and Vectors of Human Disease, 3rd ed. Lea and Febiger, Philadelphia.

Golikov, A. N., and Y. I. Starobogatov. 1975. Systematics of prosobranch gastropods. Malacologia 15:185–232.

Harasewych, M. G., S. L. Adamkewicz, J. A. Blake, D. Saudek, T. Spriggs, and C. J. Bult. 1997a. Phylogeny and relationships of pleurotomarid gastropods (Mollusca: Gastropoda): an assessment based on partial 18S rDNA and cytochrome oxidase I sequences. Molecular Marine Biology and Biotechnology 6:1–20.

Harasewych, M. G., S. L. Adamkewicz, J. A. Blake, D. Saudek, T. Spriggs, and C. J. Bult. 1997b. Neogastropod phylogeny: A molecular perspective. Journal of Molluscan Studies 63:327–351.

Harasewych, M. G., S. L. Adamkewicz, M. Plassmeyer, and P. M. Gillevet. 1998. Phylogenetic relationships of the lower Caenogastropoda (Mollusca, Gastropoda, Architaenioglossa, Campaniloidea, Cerithioidea) as determined by partial 18S rDNA sequences. Zoologica Scripta 27:361–372.

Harasewych, M. G., and A. G. McArthur. 2000. A molecular phylogeny of the Patellogastropoda (Mollusca: Gastropoda). Marine Biology 137:183–194.

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

Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755.

Huelsenbeck, J. P., F. Ronquist, R. Nielsen, and J. P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294:2310–2314.

Lewis, P. O. 2001a. A likelihood approach to estimating phylogeny from discrete morphological character data. Systematic Biology 50:913–925.

Lewis, P. O. 2001b. Phylogenetic systematics turns over a new leaf. Trends in Ecology and Evolution 16:30–37.

Lister, M. 1678. Historiae Animalium Angliae Trest Tractatus. III. Cochleis Marinis, Royal Society, London.

Lockhart, P. J., M. A. Steel, A. C. Barbrook, D. H. Huson, M. A. Charleston, and C. J. Howe. 1998. A covariotide model explains apparent phylogenetic structure of oxygenic photosynthetic lineages. Molecular Biology and Evolution 15:1183–1188.

Maddison, W. P., and D. R. Maddison. 2000. MacClade, version 4.0b10. Sinauer Associates, Sunderland, MA.

Mallatt, J., and C. J. Winchell. 2002. Testing the new animal phylogeny: First use of combined large–subunit and small–subunit rRNA gene sequences to classify the protostomes. Molecular Biology and Evolution 19:289–301.

McArthur, A. G., and B. F. Koop. 1999. Partial 28S rDNA sequences and the antiquity of hydrothermal vent endemic gastropods. Molecular Phylogenetics and Evolution 13:255–274.

Medina, M., A. G. Collins, J. D. Silberman, and M. L. Sogin. 2001. Evaluating hypotheses of basal animal phylogeny using complete sequences of large and mall subunit rRNA. Proceedings of the National Academy of Sciences U.S.A. 98:9707–9712.

Milne-Edwards, H. 1848. Note sur la classification naturelle chez Mollusques Gasteropodes. Annales des Sciences Naturalles 9:102–112.

Moore, R. C. 1960. Summary of classification. Pp. I157–I169 in Treatise on Invertebrate Paleontology, Part I (R. C. Moore, ed.). Geological Society of America and University of Kansas Press, Lawrence, KS.

Olsen, G. J., R. Overbeek, N. Larsen, T. L. Marsh, M. J. McCaughey, M. A. Maciukenas, W.-M. Kuan, T. J. Macke, Y. Xing, and C. R. Woese. 1992. The ribosomal database project. Nucleic Acids Research Supplement 2199–2200.

Olivera, B. 1997. *Conus* venom peptides, receptor and ion channel targets, and drug design: 50 million years of neuropharmacology. Molecular Biology and Cell 8:2101–2109.

Pelseneer, P. 1906. A Treatise on Zoology. V. Mollusca, A and C Black, London.

Ponder, W. F., and D. R. Lindberg. 1996. Gastropod phylogeny. Challenges for the 90s. Pp. 135–154 in Origin and Evolutionary Radiation of the Mollusca (J. D. Taylor, ed.). Oxford University Press, Oxford.

Ponder, W. F., and D. R. Lindberg. 1997. Towards a phylogeny of gastropod molluscs: An analysis using morphological characters. Zoological Journal of the Linnean Society 119:83–265.

Posada, D., and K. Crandall. 1998. Modeltest: Testing the model of DNA substitution. Bioinformatics 14:817–818.

Ragan, M. A. 1992. Phylogenetic inference based on matrix representation of trees. Molecular Phylogenetics and Evolution 1:53–58.

Rosenberg, G., G. S. Kuncio, G. M. Davis, and M. G. Harasewych. 1994. Preliminary ribosomal RNA phylogeny of gastropod and unioidean bivalve mollusks. The Nautilus, Supplement 2:111–121.

Rosenberg, G., S. Tillier, A. Tillier, G. Kuncio, R. T. Hanlon, M. Masselot, and C. J. Williams. 1997. Ribosomal RNA phylogeny of selected major clades in the Mollusca. Journal Molluscan Studies 63:301–309.

Salvini-Plawen, L. von. 1980. A reconsideration of systematics in the Mollusca (phylogeny and higher classification). Malacologia 19:249–278.

Sanderson, M. J., A. Purvis, and C. Henze. 1998. Phylogenetic supertrees: Assembling the trees of life. Trends in Ecology and Evolution 13:105–109.

Sasaki, T. 1998. Comparative anatomy and phylogeny of the Recent Archaeogastropoda (Mollusca: Gastropoda). Bulletin, The University Museum, The University of Tokyo 38:1–224.

Shoemaker, J. S., I. S. Painter, and B. S. Weir. 1999. Bayesian statistics in genetics. Trends in Genetics 15:354–357.

Swofford, D. L. 2001. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods), v. 4.0b8. Sinauer Associates, Sunderland, MA.

Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. Phylogenetic infer-
ence. Pp. 407–514 in Molecular Systematics (D. M. Hillis, C. Moritz, and B. K.
Mable, eds.). Sinauer Associates, Sunderland, MA.

Thiele, J. 1929–31. Handbuch de systematischen weichtierkunde, Gustav Fischer Ver-
lag, Jena.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: Improving the
sensitivity of progressive multiple sequence alignment through sequence weighting,
position-specific gap penalties and weight matrix choice. Nucleic Acids Research
22:4673–4680.

Tillier, S., M. Maselot, J. Guerdox, and A. Tillier. 1994. Monophyly of major gastro-
pod taxa tested from partial 28S rRNA sequences, with emphasis on Euthyneura and
hot–vent limpets Peltospiroidea. The Nautilus, Supplement 2:122–140.

Tillier, S., M. Masselot, P. Herve, and A. Tillier. 1992. Phylogénie moléculaire des Gas-
tropoda (Mollusca) fondée sur le séquençage partiel de l'ARN ribosomique 28S.
Comptes Rendus Academie de Sciences (Paris) 134:79–85.

Tracey, S., J. A. Todd, and D. H. Erwin. 1993. Mollusca: Gastropoda. Pp. 137–167 in
The Fossil Record 2 (M. J. Benton, ed.). Chapman and Hall, London.

Troschel, F. H. 1861. Das Gebiss der Schnecken, zur Begründung einer natürlichen Clas-
sification, Nicolaische Verlagsbuchhandlung, Berlin.

Wagner, P. J. 1995. Diversity patterns among early gastropods: Contrasting taxonomic
and phylogenetic descriptions. Paleobiology 21:410–439.

Wenz, W. 1938–44. Gastropoda, Teil 1: Allgemeiner Teil und Prosobranchia. in Hand-
buch der Paläozoologie (O. H. Schindewolf, ed.). Gebrüder Borntraeger, Berlin.

Winnepenninckx, B., T. Backeljau, and R. De Wachter. 1994. Small ribosomal subunit
RNA and the phylogeny of Mollusca. The Nautilus, Supplement 2:98–110.

Winnepenninckx, B., T. Backeljau, and R. De Wachter. 1996. Investigation of mollus-
can phylogeny on the basis of 18S rRNA sequences. Molecular Biology and Evolu-
tion 13:1306–1317.

Winnepenninckx, B., G. Steiner, T. Backeljau, and R. De Wachter. 1998. Details of gas-
tropod phylogeny inferred from 18S rRNA sequences. Molecular Phylogenetics and
Evolution 9:55–63.

Woese, C. R., O. Kandler, and M. L. Wheelis. 1990. Towards a natural system of or-
ganisms: Proposal for the domains Archaea, Bacteria and Eucarya. Proceedings of the
National Academy of Sciences U.S.A. 87:4576–4579.