Bilaterian Phylogeny Based on Analyses of a Region of the Sodium–Potassium ATPase α-Subunit Gene

Frank E. Anderson,[1,2] Alonso J. Córdoba,[2] Mikael Thollesson[1,3]

[1] Laboratory of Molecular Systematics, Smithsonian Institution, Museum Support Center, 4210 Silver Hill Road, Suitland, MD 20746, USA
[2] Department of Zoology and Center for Systematic Biology, Southern Illinois University, Carbondale, IL 62901, USA
[3] Department of Molecular Evolution, Evolutionary Biology Centre, and Linnaeus Centre for Bioinformatics, Uppsala University, Norbyvägen 18C, SE-752 36, Uppsala, Sweden

Abstract. Molecular investigations of deep-level relationships within and among the animal phyla have been hampered by a lack of slowly evolving genes that are amenable to study by molecular systematists. To provide new data for use in deep-level metazoan phylogenetic studies, primers were developed to amplify a 1.3-kb region of the α subunit of the nuclear-encoded sodium–potassium ATPase gene from 31 bilaterians representing several phyla. Maximum parsimony, maximum likelihood, and Bayesian analyses of these sequences (combined with ATPase sequences for 23 taxa downloaded from GenBank) yield congruent trees that corroborate recent findings based on analyses of other data sets (e.g., the 18S ribosomal RNA gene). The ATPase-based trees support monophyly for several clades (including Lophotrochozoa, a form of Ecdysozoa, Vertebrata, Mollusca, Bivalvia, Gastropoda, Arachnida, Hexapoda, Coleoptera, and Diptera) but do not support monophyly for Deuterostomia, Arthropoda, or Nemertea. Parametric bootstrapping tests reject monophyly for Arthropoda and Nemertea but are unable to reject deuterostome monophyly. Overall, the sodium–potassium ATPase α-subunit gene appears to be useful for deep-level studies of metazoan phylogeny.

Key words: Sodium–potassiumn ATPase α subunit — Metazoan phylogeny — Maximum likelihood — Bayesian inference — Molecular systematics

Introduction

Recent investigations of the relationships within and among the metazoan (multicellular animal) phyla have employed morphological data (e.g., Eernisse et al. 1992; Nielsen et al. 1996), molecular data (e.g., Eeckhaut et al. 2000; Field et al. 1988; Giribet and Ribera 1998; Ruiz-Trillo et al. 1999), and combined morphological and molecular data sets (Giribet et al, 2000; Giribet et al. 2001; Zrzavy et al. 1998). Most molecular investigations of metazoan phylogeny to date have utilized the nuclear small-subunit (18S) ribosomal RNA gene. Analyses of 18S data have provided a number of important insights into metazoan relationships at several levels, including evidence supporting a clade of molting animals (e.g., nematodes and arthropods) dubbed Ecdysozoa (Aguinaldo et al. 1997). However, some aspects of metazoan phylogeny remain unclear, most notably relationships within Lophotrochozoa, a clade that includes Annelida, Mollusca, and several other phyla (Halanych 1998; Halanych et al. 1995).

Correspondence to: Frank E. Anderson, Department of Zoology, Southern Illinois University, Carbondale, IL 62901, USA; *email:* feander@siu.edu

To address these remaining ambiguities, some researchers have studied large multilocus data sets from representatives of a few taxa (those for which complete genomic sequences or large EST databases are available) (e.g., Blair et al. 2002; Mushegian et al. 1998; Wang et al. 1999), while others have attempted to gather sequences for one or a few genes from representatives of many taxa (e.g., Giribet et al. 2001). Both approaches have merit, but at this time large sequence databases exist for only a few metazoans, and most molecular systematists simply do not have the resources to generate complete genomic sequences (or EST profiles) for representatives of several metazoan taxa. Therefore, PCR-based studies of relatively small numbers of genes will continue to be the predominant approach employed by most molecular systematists, because it allows several taxa of interest to be studied in a cost-effective manner.

Other approaches to investigating phylogeny beyond comparisons of nucleotide substitutions within one or more genes are also being explored. Rare genomic changes (RGCs) such as intron indels, SINEs and LINEs, mitochondrial and nuclear genetic code variants, and gene duplications may also play an important role in resolving phylogenetic problems at various levels, including relationships among and within the animal phyla (Rokas and Holland 2000). Thus far, three types of RGC have been used to investigate aspects of metazoan phylogeny: mtDNA code variants (Castresana et al. 1998), signature Hox gene motifs (de Rosa et al. 1999), and mtDNA gene order (Boore 1999). Of these, mtDNA gene order has received the most attention.

Although studies of mitochondrial gene order and complete mitochondrial genome sequences seem to corroborate findings based on analyses of 18S data alone (Boore and Brown 2000; Boore et al. 1998; von Nickisch-Rosenegk et al. 2001), this line of research seems to have produced few novel insights to date (but see Boore and Staton 2002). This is due in part to the difficulty and expense of sequencing complete mitochondrial genomes. Mitochondrial gene order rearrangements initially seemed to be an ideal source of information for studying ancient divergences, but the presumed rarity of mitochondrial gene rearrangement events has been called into question recently (Dowtin and Austin 1999; Flook et al. 1995; Hickerson and Cunningham 2000; Le et al. 2000; Mindell et al. 1998; Rawlings et al. 2001; Shao et al. 2001a,b), and the development of appropriate methods for analyzing gene order data has been difficult (Blanchette et al. 1999). Complete mitochondrial sequence data seem promising, but even $14+$ kb of mitochondrial protein-coding sequence is insufficient for resolving relationships among basal vertebrates (Cao et al. 1998), suggesting that the ability of these data to resolve interphylum relationships also may be

limited. These observations suggest that mitochondrial data alone may be insufficient for the resolution of persistent problems in deep-level animal phylogeny—additional data from the nuclear genome, in the form of either sequences of additional nuclear genes from representatives of several phyla or new RGCs (or both) will be necessary.

A few studies have focused on nuclear genes other than 18S, most notably the large-subunit (28S) ribosomal rRNA gene (Christen et al. 1991; Mallatt and Winchell 2002) and the gene that encodes elongation factor 1-α (EF-1α) (Kobayashi et al. 1996; Kojima 1998; Kojima et al. 1993; McHugh 1997; Regier and Shultz 1997, 1998). Unfortunately, the value of the EF-1α gene for elucidating relationships among phyla may be somewhat limited (Littlewood et al. 2001), although it has recovered some interesting patterns, including paraphyly of Annelida with respect to Echiura and Pogonophora (McHugh 1997) and paraphyly of Crustacea with respect to Hexapoda (Regier and Shultz 1997). Comparative studies of a few other nuclear genes have also been published, including studies of aldolase and triose phosphate isomerase (Nikoh et al. 1997), intermediate filament proteins (Erber et al. 1998), lamins (Erber et al. 1999), heat shock proteins (Borchiellini et al. 1998), protein kinases (Kruse et al. 1998), elongation factor-2 (Regier and Shultz 2001), and myosin heavy chain type II (Ruiz-Trillo et al. 2002). Thus far, however, these data sets have been very limited in taxonomic scope, likely reflecting the difficulty of designing ''universal'' PCR primers that can amplify a particular gene region from representatives of most (or all) metazoan phyla.

Studies of metazoan relationships have clearly been hampered by a relative lack of slowly evolving genes that can be studied using standard laboratory techniques available to most molecular systematists (e.g., PCR). Unfortunately, little exploratory work has been done to develop additional nuclear protein-coding gene markers for studies of deep intraphylum and interphylum relationships within Metazoa. Friedlander et al. (1992) described 14 promising candidate genes for phylogenetic research within Metazoa and listed EF-1α, elongation factor-2, Na^+/K^+ ATPase, enolase, and glucose-6-phosphate dehydrogenase as genes that may be suitable for use in deep-level phylogenetic studies of metazoans. In an effort to develop new nuclear-protein coding markers to address remaining questions in inter- and intraphylum metazoan relationships, we developed three pairs of oligonucleotide primers for amplification of a 1.3-kb region of coding sequence from the gene encoding the α subunit of Na^+/K^+ ATPase (the ''sodium pump'' enzyme). We used these primers to amplify this gene region from mRNA extracts from representatives of a diverse group of thirty-one

Table 1. Oligonucleotide sequences and expected PCR fragment sizes for all primer

Name of oligonucleotide	Oligonucleotide sequence (5′ ⇒ 3′)	Expected nested PCR fragment size (bp)
fATPa	ATGACNGTNGCNCATATGTGGT	712
rATPa	ATNGGGTGGTCNCCNGTNACCAT	
fATPb	GTNATGAAGGGNGCNCCNGA	691
rATPb	CCCATNGCNACNCCNATGTCNGCTTT	
fATPc	ATGGTNACNGGNGATCATCCNAT	
rATPc	ATNGCNGGNACCATGTCNGTNCC	619

Note. For nested PCR reamplifications, primers with M13-21 and M13 rev sequences attached to their 5′ ends were used to allow direct sequencing with M13 primers. M13 linkers attached to the 5′ ends of each of these primers are as follows: 5′-CAGGAAA-CAGCTATGACC-3′ (*f* primers) and 5′-TGTAAAACGACGGCCAGT-3′ (*r* primers).

bilaterian (bilaterally symmetrical) metazoans. These sequences were used to reconstruct relationships among and within several important metazoan taxa and to evaluate the potential value of this gene region for future studies of metazoan phylogeny.

Materials and Methods

Primer Design

An alignment of several metazoan ATPase α-subunit amino acid sequences downloaded from GenBank was visually scanned for short conserved motifs of seven or more amino acids that could serve as PCR primer sites. Oligonucleotides matching five conserved regions were synthesized for use in RT-PCR and PCR (Table 1). The primers have several degenerate sites and are thus unsuitable for use in cycle sequencing reactions. To allow PCR fragments to be directly sequenced, all primers were synthesized with M13 linker sequences attached to their 5′ ends, following Cho et al. (1995) and Regier and Schulz (1997). The 5′-most and 3′-most primers (fATPa and rATPc) were also synthesized without M13 linkers; this primer pair was used in initial RT-PCR amplifications.

RT-PCR and Nested PCR Amplification

Total RNA was extracted from 50–100 mg of tissue using TRI reagent (Molecular Research Center) following the manufacturer's protocols. The taxa sampled, along with collection locality and GenBank accession numbers, are listed in Table 2. RT-PCR amplification was performed using Ready-To-Go RT-PCR Beads (Amersham Pharmacia Biotech) and the following cycling parameters: 42°C for 20 min (1 cycle), 95°C for 5 min (1 cycle), 95°C for 30 s—38°C for 30 s—72°C for 1 min (32 cycles), 72°C for 7 min (1 cycle). RT-PCR reagent concentrations followed the manufacturer's suggestions, and a poly (T) primer was used to enhance mRNA amplification. To produce smaller overlapping fragments suitable for direct sequencing, nested PCR reamplifications of the RT-PCR product were performed using the primer pairs fATPa–rATPa, fATPb–rATPb, and fATPc–rATPc (Table 1). Reamplifications were performed under a variety of cycling parameters and reagent concentrations. A regime consisting of 94°C for 90 s (1 cycle), 94°C for 30 s—50°C for 30 s—72°C for 1 min (10 cycles), 94°C for 30 s—44°C for 30 s—72°C for 1 min (25 cycles), followed by a single terminal extension step (72°C for 7 min) on a Perkin–Elmer Gene Amp 9700 was usually satisfactory. Each PCR product was gel-purified using Centri-Sep spin columns (Princeton Separations) and the purified fragment was directly sequenced in both directions using dye terminator chemistry (BigDye Terminators; Applied Biosystems) with M13 sequencing primers. Sequencing reactions were run out on an ABI 373XL automated sequencer. The resulting

sequences were assembled and edited in Sequencher (GeneCodes). Edited sequences were aligned using ClustalX (Thompson et al. 1997) with modification by eye in Se-Al (Rambaut 1995) and MacClade 4.03 (Maddison and Maddison 2001).

MP, ME, ML, and Bayesian Analyses

The nucleotide data alignment was analyzed under maximum parsimony (MP; equal weights and implicit weights), minimum evolution (ME; LogDet + invariant distances, using an ML estimate of the proportion of invariant sites), and maximum likelihood (ML) criteria in PAUP* 4.0 (Swofford 2002) and using a Bayesian inference approach in MrBayes 2.0 (Huelsenbeck and Ronquist 2001). Several ATPase α-subunit coding sequences were included with the taxa sampled in this study to increase the diversity of taxa represented (Table 2). Two taxon sets were analyzed—a large taxon set including 24 vertebrate ATPase α-subunit sequences (representing several isoforms) and a smaller taxon set including only 13 vertebrate isoform 1 sequences, each with all available nonvertebrate metazoan ATPase sequences. Two regions of the data set (corresponding to amino acid positions 423–429 and 509–519 in the *Drosophila melanogaster* sequence) were difficult to align with confidence; these regions were excluded from all analyses. Third codon position sites were also excluded from all nucleotide analyses due to lack of base frequency homogeneity across all sequences (see below). The nucleotide data were also translated to produce an amino acid data matrix, which was analyzed using MP and ML. All MP and ME heuristic searches were performed using 200 random addition sequence replicates with tree bisection–reconnection (TBR) branch-swapping, holding 10 trees at each addition step.

The MP trees were used to evaluate nucleotide substitution models for ML analysis. Likelihood scores were calculated for one MP tree for each taxon set under 56 substitution models, and ModelTest 3.0 (Posada and Crandall 1998) was used to select an appropriate model via a series of hierarchical likelihood ratio tests. The general time-reversible (GTR) DNA substitution model with among-site rate heterogeneity modeled using a mixed-model γ + invariant sites distribution (Γ + I) was chosen by ModelTest as the most adequate model for both taxon sets. Investigation of the likelihood score output revealed that a GTR + Γ + I submodel not evaluated by ModelTest was not significantly worse than a full GTR + Γ + I model (likelihood-ratio test; $p = 0.414427$ for the small taxon set, $p = 0.223130$ for the large taxon set). This GTR submodel employed four (rather than six) relative rate parameters: one for A–C transversions and A–G transitions, one for A–T and C–G transversions, one for C–T transitions, and one for G–T transversions (the PAUP* option used was "RCLASS = (a a b b c d)"). To reduce computational time and sampling error, this simpler model was used for all ML searches. To further reduce analysis time, a "successive approximations" approach (e.g., Anderson 2000; Frati et al. 1997) was used. All substitution model parameters

were estimated from the data on five starting trees—two MP trees (if more than two MP trees were produced, two very different MP trees were chosen to maximize topological diversity in the starting tree pool), a successive weighting parsimony tree, an implicit weights parsimony tree ($k = 2$) and a LogDet + I tree. In cases where only one MP tree was found in initial searches, two random trees were also used as starting points for ML searches. Each of these trees (with model parameter estimates fixed) was then used as a starting tree for a round of nearest-neighbor interchange (NNI) branch swapping under ML, followed by reestimation of model parameters and a subsequent round of TBR branch swapping. For each taxon set, all starting points converged on the same ML topology; this tree was accepted as the ML tree. The WAGf + Γ + I substitution model (Whelan and Goldman 2001) was selected as the best-fitting model for the amino acid data matrix following a series of likelihood-ratio tests. TREE-PUZZLE (Schmidt et al. 2002; Strimmer and von Haeseler 1996) was used to perform a maximum likelihood quartet puzzling analysis of the amino acid matrix under this model, with amino acid frequencies, the Γ shape parameter (α), and the proportion of invariant sites (I) estimated from the data.

Nodal support was estimated using equal-weights parsimony bootstrap analysis (100 pseudoreplicates, each consisting of a heuristic search using 100 random addition sequence replicates), maximum likelihood bootstrap analysis using the GTR + Γ + I submodel described above (100 pseudoreplicates generated with SEQBOOT in PHYLIP and analyzed using a successive approximations approach in PAUP*) (Felsenstein 1985, 1995), Bayesian analysis (1 million generations, with the first 200,000 generations discarded as burn-in) (Huelsenbeck and Ronquist 2001), and quartet puzzling (for the amino acid data). The ML submodel used in PAUP* for analysis of the nucleotide data is not available in MrBayes; a standard GTR + Γ + I model was used instead.

Hypothesis Testing

A parametric bootstrapping approach (Hillis et al. 1996; Huelsenbeck et al. 1996) was used to test the level of support in the small taxon set for monophyly of three higher taxa—Deuterostomia, Arthropoda, and Nemertea. Model trees for each hypothesis (deuterostome monophyly, arthropod monophyly, and nemertean monophyly) were constructed by performing ML analyses in PAUP* using the GTR + Γ + I submodel described above and topological constraints to enforce monophyly for each group of interest. One hundred data sets were simulated on each model tree with Seq-Gen 1.25 (Rambaut 1995) using branch lengths and model parameters resulting from the constrained ML analysis. Each simulated data set was then analyzed under ML twice: once under topological constraints to enforce monophyly of the clade of interest and once with no topological constraints. The distribution for the test statistic was generated by calculating the difference in ln$L$ scores between the best unconstrained tree and the best constrained tree for each simulated data set. The ln$L$ difference between the best constrained and unconstrained trees for the original data set was then directly compared to the appropriate distribution to estimate p values.

Results

Sequence Comparisons and Practical Considerations

Evaluation of sequences collected in this study supports the view that they represent coding region sequences of the ATPase α-subunit gene. Initial comparisons of these sequences with ATPase sequences in GenBank using the BLAST search protocol returned all ATPase α-subunit coding sequences as top matches. There is evidence of among-sequence base composition heterogeneity when all sequences and all codon positions are evaluated (mean nucleotide frequencies: A = 0.25113, C = 0.24410, G = 0.25153, and T = 0.25324 [$p \ll 0.01$, χ^2 test of homogeneity of base frequencies across taxa]), but the heterogeneity appears to be confined to third codon positions. With third positions excluded, the mean nucleotide frequencies are A = 0.29011, C = 0.20925, G = 0.25809, and T = 0.24255, and there is no significant heterogeneity ($p = 0.94098848$).

Unfortunately, the primers developed here are not universal metazoan primers—despite several attempts, the primers failed to amplify the target fragment from RNA extracts from two bryozoans (*Bugula nerita* and *Pectinatella magnifica*), two nematodes (*Camallanus oxycephalus* and an unidentified mermithid nematode), a nematomorph (*Gordius* sp.), two sipunculans (*Golfingia pugettensis* and *Phascolosoma agassizii*), an echiuran (*Urechiscaupo*), two echinoderms (the holothuroid *Eupentactaquinquesemita* and the ophiuroid *Ophiothrix spiculata*), an apterygote insect (*Thermobia domestica*), and four annelids (the polychaetes *Nereis virens* and *Phragmatopoma californica*, the hirudinid *Placobdella parasitica*, and an unidentified lumbricid earthworm). For two of the three gastropods studied (*Tegula brunnea* and *Onchidella borealis*), only about 1000 bp could be successfully amplified and sequenced. Entire sequences were obtained for the pseudoscorpion *Garypus californica* and the brown recluse spider *Loxosceles reclusa*, but some regions had several ambiguous base calls. The practical value of this gene region for studies of bilaterian phylogeny is admittedly somewhat limited if universal PCR primers cannot be developed. Development of consensus-degenerate hybrid oligonucleotide (a.k.a. CODEHOP) primers (Rose et al. 1998), which employ a degenerate "core" region at the 3′ end of the primer and a nondegenerate consensus "clamp" region at the 5′ end, may allow amplification of this gene fragment from additional taxa; we are currently exploring this possibility.

Phylogenetic Analyses

Nucleotide Data

The nucleotide and amino acid data matrices, as well as several trees, have been submitted to TreeBASE (www.treebase.org; study accession number = S904, matrix accession numbers = M1485–M1487). MP analysis of the small nucleotide alignment (including only isoform 1 sequences from vertebrates) yielded 32 equally parsimonious trees (TL = 3819; trees not shown). Successive weighting and

Table 2. List of all taxa sampled for this study

Higher taxon	Subclade 1	Subclade 2	Scientific name	Common name	Collection locality	GenBank No.
Arthropoda	Chelicerata	Pseudoscorpiones	*Garypus californica*	Pseudoscorpion	Santa Cruz, CA	AY303367
		Araneae	*Loxosceles reclusa*	Brown recluse spider	Jackson County, IL	AY303368
		Opiliones	Opilionid*	Harvestman	Suitland, MD	AY303366
	Crustacea	Amphipoda	*Caprella* sp.	Skeleton shrimp	Santa Cruz, CA	AY303369
		Isopoda	Oniscidean*	Pill bug	Arlington, VA	AY303370
		Brachiura	*Artemia franciscana*^	Brine shrimp	—	X56650
		Decapoda	*Callinectes sapidus*^	Blue crab	—	AF327439
			Carcinus maenas^	Green crab	—	AY035550
	Hexapoda	Hemiptera	*Acrosternum hilare*	Green stink bug	Jackson County, IL	AY303376
		Coleoptera	*Harmonia axyridis*	Ladybird beetle	Jackson County, IL	AY303371
			Tenebrio molitor	Mealworm (larva)	Pet store	AY303372
		Dermaptera	*Doru lineare*	Earwig	Jackson County, IL	AY303373
		Diptera	Brachyceran fly	Brachiopod	Jackson County, IL	AY303374
			Nematoceran fly*	Mosquito	Jackson County, IL	AY303375
			Drosophila melanogaster^	Fruit fly	—	X14476
		Siphonoptera	*Ctenocephalides felis*^	Cat flea	—	S66043
	Myriapoda	Chilopoda	*Scolopocryptops sexspinosa*	Centipede	Jackson County, IL	AY303365
		Diplopoda	Diplopodan 1*	Millipede	Arlington, VA	AY303364
			Diplopodan 2*	Millipede	Jackson County, IL	AY303363
Brachiopoda			*Terebratalia transversa*	Brachiopod	Friday Harbor, WA	AY303380
Bryozoa			*Watersipora subtorquata*	Bryozoan	Bodega Bay, CA	AY303378
			Conopeum sp.	Bryozoan	Chesapeake Bay, MD	AY303377
Chordata	Vertebrata	Actinopterygii	*Anguilla anguilla*^	European eel	—	X76108
			Catostomus commersoni^	White sucker	—	X58629
			Danio rerio^	Zebrafish	—	NM_131683, NM_131684, NM_131686
			Electrophorus electricus^	Electric eel	—	AF356351
			Oreochromis mossambicus^	Tilapia	—	U82549, AFI09409
		Anura	*Bufo marinus*^	Cane toad	—	Z11798
			Xenopus laevis^	African clawed frog	—	U10108
		Aves	*Gallus gallus*^	Chicken	—	J03230, M59959, M59960
		Chondrichthyes	*Torpedo californica*^	Electric ray	—	X02810
		Mammalia	*Canis familiaris*^	Domestic dog	—	L42173
		Mammalia	*Homo sapiens*^	Human	—	NM_000701, NM_000702, NM_000703
			Ovis aries^	Domestic sheep	—	X02813
			Rattus norvegicus^	Norway rat	—	NM_012504, NM_012505, NM_012506, NM_022848
			Sus scrofa^	Domestic pig	—	X03938
Cnidaria	Hydrozoa		*Hydra vulgaris*^	Hydra	—	M75140
Echinodermata	Asteroidea		*Asterina miniata*	Bat star	Palos Verdes Estates, CA	AY303361
	Echinoidea		*Strongylocentrotus purpuratus*	Purple sea urchin	Palos Verdes Estates, CA	AY303362

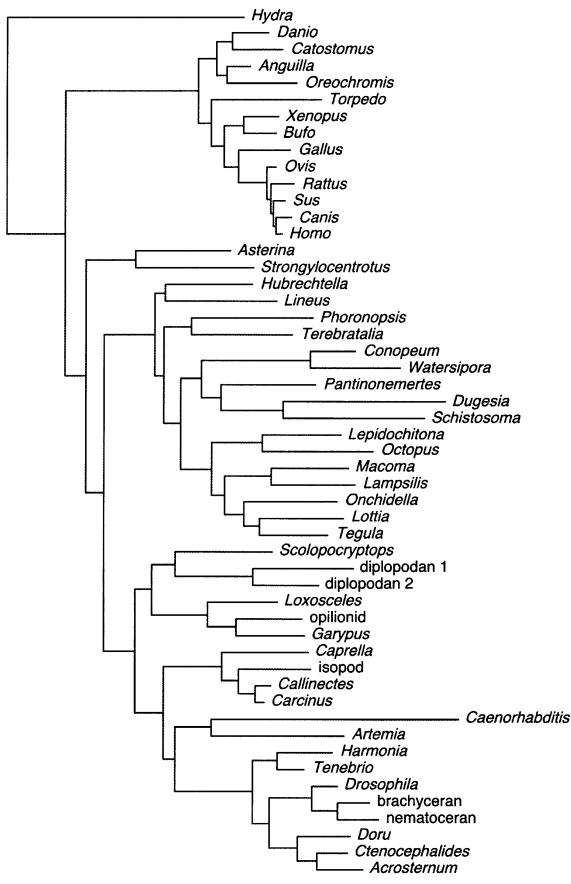|						
Mollusca	Bivalvia	Unionidae	*Lampsilis cardium*	Plain pocketbook	Kentucky	AY303383
	Bivalvia		*Macoma nasuta*	Bent-nosed clam	Friday Harbor, WA	AY303382
	Cephalopoda	Octopoda	*Octopus rubescens*	Red octopus	Pigeon Point, CA	AY303384
	Gastropoda	Acmaeidae	*Lottia austrodigitalis*	Limpet	Palos Verdes Estates, CA	AY303385
		Onchidiidae	*Onchidella borealis*	Leather limpet	Friday Harbor, WA	AY303386
		Trochidae	*Tegula brunnea*	Brown turban snail	Palos Verdes Estates, CA	AY303387
	Polyplacophora		*Lepidochitona hartwegi*	Chiton	Palos Verdes Estates, CA	AY303381
Nematoda			*Caenorhabditis elegans*^	Nematode	—	NM_073868
Nemertea	Heteronemertea		*Lineus ruber*	Nemertean	Nahant, MA	AJ563804
	Hoplonemertea		*Pantinonemertes* sp. nov.	Nemertean	Fort Pierce, FL	AJ563802
	Pilidiophora		*Hubrechtella dubia*	Nemertean	Fort Pierce, FL	AJ563803
Phoronida			*Phoronopsis viridis*	Phoronid	Bodega Bay, CA	AY303379
Platyhelminthes	Tricladida		*Dugesia japonica*^	Planarian	—	AB012391
	Trematoda		*Schistosoma mansoni*^	Human blood fluke	—	AF303222

Note. *Voucher destroyed prior to species-level identification. ^Sequence downloaded from GenBank.

implicit weight MP analyses, as well as ME analysis of LogDet + I distances, each yielded one tree (not shown). ML analysis yielded a single tree (Figs. 1 and 2; $\ln L = 19{,}806.34902162$). Several clades are recovered with high bootstrap support and Bayesian clade posterior probability. A large clade comprising all mollusk, nemertean, flatworm, and bryozoan sequences, along with the phoronid and brachiopod sequences, was recovered with moderate support (ML bootstrap support value [LBS] = 77, MP bootstrap support value [PBS] = 62, Bayesian posterior probability [BPP] = 1.0). This clade, known as Lophotrochozoa (Halanych et al. 1995) or Eutrochozoa (Eernisse et al. 1992; Ghiselin 1989), has been recovered in one form or another in several recent studies (e.g., Peterson and Eernisse 2001). Within this group, only a bryozoan clade (LBS and PBS = 100, BPP = 1.0), a mollusk clade (LBS = 65, PBS = 63, BPP = 0.986), and a pairing of the two bivalve sequences (LBS = 84, PBS = 89, BPP = 1.0) are supported. It should be noted, however, that taxon sampling is extremely limited and that future studies of this gene that include more taxa may reach different conclusions.

Another large clade, consisting of the lone nematode sequence and all arthropod sequences, was also recovered with moderate support (LBS = 61, PBS = 49, BPP = 1.0), although the position of the nematode sequence (as sister to *Artemia*, a brine shrimp; LBS = 72, PBS = 49, BPP = 1.0) is unexpected. The LogDet + Invariant ME tree has the *Caenorhabditis* sequence in a different position—as sister to a Crustacea + Hexapoda clade—but Arthropoda itself is still not monophyletic (Table 3). Within Arthropoda, arachnid (LBS = 97, PBS = 97, BPP = 1.0), hexapod (LBS/PBS = 100, BPP = 1.0), and myriapod (LBS = 81, PBS = 63, BPP = 0.991) monophyly are well supported. An opilionid + pseudoscorpion grouping (LBS = 97, PBS = 89, BPP = 1.0) and monophyly for a group of malacostracan crustaceans consisting of an isopod, an amphipod, and two decapods (LBS/PBS = 100, BPP = 1.0) are also strongly supported. Hexapods are found to be nested within Crustacea (including the *Artemia* + *Caenorhabditis* clade). Within Hexapoda, dipteran monophyly (LBS = 96, PBS = 89, BPP = 1.0), coleopteran monophyly (LBS = 82, PBS = 64, BPP = 0.999) and monophyly for a dermapteran + hemipteran + siphonopteran clade (LBS = 96, PBS = 83, BPP = 1.0) are all moderately supported.

A monophyletic Deuterostomia is not recovered in these analyses (i.e., the echinoderms do not form a clade with the vertebrates; Fig. 1). There is, however, appreciable support for a pairing of the two echinoderms (LBS = 77, PBS = 71, BPP = 1.0). Vertebrate monophyly is also strongly supported (LBS/PBS = 100, BPP = 1.0), as is monophyly of Tetra-

Hydra

Danio
Catostomus
98/93
1.0
Anguilla
Oreochromis
61/35
0.915
88/87
1.0

Torpedo
Xenopus
Bufo
100
1.0
Gallus
Ovis
Rattus
100
1.0
Sus
Canis
Homo
42/51
0.819
47/25
0.756
65/69
0.896
95/90
1.0
91/91
1.0
69/16
0.981
100
1.0

Vertebrata

Strongylocentrotus
Asterina **A**
77/71
1.0

Hubrechtella
Lineus
35/31
0.786

Phoronopsis
Terebratalia
49/34
0.980

Conopeum **B**
Watersipora
100
1.0

Pantinonemertes
Dugesia **C**
Schistosoma
99/95
1.0
44/28
0.999
17/7
0.635
15/11
0.817

Lepidochitona
Octopus
79/65
1.0
Macoma **D**
Lampsilis
84/89
1.0
*Onchidella** **E**
*Tegula**
Lottia
57/25
0.803
42/59
0.834
49/63
0.804
65/63
0.986
26/8
0.991
77/62
1.0

Mollusca
Lophotrochozoa

Scolopocryptops
diplopodan 1
diplopodan 2 **F**
100
1.0
81/63
0.991
Loxosceles **G**
opilionid
Garypus
97/89
1.0
97/97
1.0
65/41
0.926

Caprella
isopod
Callinectes
Carcinus
75/93
0.999
58/74
0.875
100
1.0

Caenorhabditis
Artemia
72/49
1.0

Harmonia **H**
Tenebrio
82/64
0.999

Drosophila **I**
brachyceran
nematoceran
92/88
1.0
96/89
1.0
Doru
Ctenocephalides
Acrosternum **J**
73/67
0.988
96/83
1.0
96/64
0.918
100
1.0
48/8
0.951
66/11
1.0
61/49
1.0

Hexapoda
"Ecdysozoa"

64/48
0.892

46/33
0.922

100
1.0

Fig. 2. Maximum likelihood phylogram for the small taxon set. The scale bar is equivalent to 0.05 substitution per site.

poda (LBS = 91, PBS = 91, BPP = 1.0), Amniota (LBS = 95, PBS = 90, BPP = 1.0), and Mammalia (LBS/PBS = 100, BPP = 1.0), although relationships within Mammalia are poorly supported.

All MP and ME trees, as well as the majority-rule consensus tree from Bayesian analysis, are highly congruent with the ML tree (Table 3). The major differences between trees resulting from different analyses are in regions of low bootstrap support and Bayesian posterior probability, particularly within Lophotrochozoa and Arthropoda. Deuterostome monophyly is not recovered in any analysis of the small nucleotide alignment, and some analyses do not

Fig. 1. Maximum likelihood tree for the small taxon set (including only vertebrate isoform I sequences), with groups of interest highlighted. Support values are as follows: above branches (left)—equal-weights parsimony bootstrap support values; above branches (right)—maximum likelihood bootstrap support values; below branches—Bayesian clade posterior probabilities. Branches marked with a single boldface italicized "100" had 100% MP and ML bootstrap support. An asterisk (*) indicates a partial sequence (approximately 1000 bp). Letters denote the following clades of interest: A—Echinodermata, B—Bryozoa (Cheilostomata), C—Platyhelminthes, D—Bivalvia, E—Gastropoda, F—Myriapoda, G—Arachnida, H—Coleoptera, I—Diptera and J—"Dermaptera–Hemiptera–Siphonoptera" (DHS) clade.

recover a Myriapoda + Arachnida clade or a monophyletic Myriapoda (Table 3). Monophyly for Gastropoda, Coleoptera, Osteichthyes, Echinodermata, Protostomia, the *Artemia* + *Caenorhabditis* clade, and the "Diptera + DHS" clade are also recovered in most, but not all, analyses.

MP analysis of the large nucleotide alignment (including isoform 1, 2, 3, and 4 sequences from several vertebrates) yielded eight equally parsimonious trees (TL = 4679; trees not shown). As above, successive weighting and implicit weighting MP analyses produced one tree, as did ME analysis of LogDet + I distances (not shown). ML analysis yielded a single tree (Fig. 3; InL = 23871.83723707) that is congruent with the ML tree for the small taxon set (Fig. 1) in all but one respect. ML analysis of the large taxon set recovered a monophyletic Osteichthyes (all vertebrates but *Torpedo californica*, a cartilaginous fish) in the isoform 1 clade with low support (LBS = 52, PBS = 45, BPP = 0.642). A monophyletic Osteichthyes was not recovered in ML analysis of the small taxon set (Fig. 1).

Amino Acid Data

Equal-weights MP analysis of the amino acid alignment resulted in six trees (TL = 2743; strict consensus shown in Fig. 4). Implicit weights parsimony and successive weighting analyses each produced one tree (not shown). The amino acid parsimony trees are highly congruent with trees recovered in analyses of the nucleotide data (Table 3). The only major difference between the amino acid trees and the nucleotide trees is the lack of support for a Myriapoda + Arachnida clade and variable recovery of a monophyletic Myriapoda in the amino acid trees (Table 3). The quartet puzzling ML tree for the amino acid data (not shown) is less resolved than—but largely congruent with—the amino acid MP trees and nucleotide trees, with a notable exception; *Pantinonemertes* is not found with the flatworms (as for DNA data) but as sister group to *Lineus*.

Parametric Bootstrapping Analyses

The results of the parametric bootstrapping tests of deuterostome, arthropod, and nemertean monophyly are shown in Fig. 5. Deuterostome monophyly cannot be rejected by these data (1.0 > p > 0.05), while arthropod and nemertean monophyly are rejected ($p < 0.01$ in both cases).

Discussion

Phylogenetic Hypotheses

In general, phylogenetic analyses of this fragment of the sodium–potassium ATPase α-subunit result in

Table 3. Cross-analysis evaluation of several clades for the small taxon set

Clade	Nucloeitde data						Amino acid data			
	ML	EWP	SWP	IW	LDI	B	EWP	SWP	IW	OP
Echinodermata	✓	✓	✓	N	✓	✓	✓	✓	✓	✓
Vertebrata	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Osteichthyes	N	✓	✓	✓	✓	N	✓	✓	N	✓
Anura	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Amniota	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Mammalia	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Protostomia	✓	N	✓	N	✓	✓	✓	✓	✓	✓
Ecdysozoa	✓	✓	✓	✓	✓	✓	✓	✓	✓	N
Myriapoda + Arachnida	✓	N	N	✓	✓	✓	N	N	N	N
Myriapoda	✓	N	N	✓	✓	✓	N	N	✓	N
Arachnida	✓	✓	✓	✓	✓	✓	✓	✓	✓	N
Opilionid + *Garypus*	✓	✓	✓	✓	✓	✓	✓	✓	✓	N
Malacostraca	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Artemia + Caenorhabditis	✓	✓	✓	✓	N	✓	✓	✓	✓	✓
Hexapoda	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Coleoptera	✓	N	N	✓	✓	✓	✓	✓	✓	✓
Diptera + DHS	✓	✓	✓	✓	✓	✓	✓	✓	N	N
Diptera	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
DHS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Lophotrochozoa	✓	✓	✓	✓	✓	✓	✓	✓	✓	N
Phoronida + Brachiopoda	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Platyhelminthes	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Mollusca	✓	✓	✓	✓	✓	✓	✓	✓	✓	N
Bivalvia	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gastropoda	✓	N	N	✓	N	✓	✓	✓	✓	✓
Polyplacophora + Cephalopoda	✓	✓	✓	✓	✓	✓	✓	✓	N	✓

Note. Checks denote that the clade was recovered in all optimal trees for the given analysis; N denotes clades that were not recovered (often due to lack of resolution). ML—maximum likelihood; EWP—equal-weights parsimony; SWP—successive-weights parsimony; IW—implicit weights parsimony; LDI—LogDet + invariant distances; B—Bayesian; QP—quartet puzzling; DHS—Dermaptera + Hemiptera + Siphonoptera.

trees that are concordant with results of previous studies of metazoan phylogeny based on morphological and/or 18S sequence data. Monophyly of most phyla is supported by the ATPase analyses, as is monophyly of a few major superphylum-level groups (Protostomia, Ecdysozoa, and Lophotrochozoa).

Relationships Within Protostomia

Recovery of a variant of Ecdysozoa—a clade consisting of all molting animals sampled for this study—is consistent with previous findings based on both 18S data (Aguinaldo et al. 1997) and some analyses of morphological data (Eernisse et al. 1992). Although a monophyletic Ecdysozoa was recovered in all analyses, the position of the *C. elegans* sequence—usually as sister to *Artemia franciscana*—was unexpected (this finding is evaluated further below). Overall, most other inferred relationships within Arthropoda match previous findings rather closely. The presence of a Pancrustacea clade (disregarding the enigmatic position of the *C. elegans* sequence) in the ATPase phylogeny serves as independent support of this finding in studies of other sources of data, including 18S (Giribet and Ribera 2000), EF-1α (Regier

and Shultz 1997), combined morphology and 18S (Zrzavy et al. 1998), a combined data set consisting of several genes and morphology (Giribet et al. 2001), and mitochondrial gene order (Boore et al. 1998). The positions of Myriapoda and Chelicerata remain unclear, with some previous studies proposing a Myriapoda + Chelicerata clade (Friedrich and Tautz 1995; Giribet et al. 1996; Hwang et al. 2001; Kusche and Burmester 2001; Turbeville et al. 1991), while others support a basal position for Myriapoda (Regier and Shultz 2001) or Chelicerata (Regier and Shultz 1997; Shultz and Regier 2000; Wheeler et al. 1993) within Arthropoda. A Myriapoda + Chelicerata clade is supported by a putative molecular synapomorphy—a deletion of the second α-helix of hemocyanin domain 1 (Kusche and Burmester 2001). Support for any particular resolution of myriapod/chelicerate relationships is, however, generally low in most studies. The ATPase nucleotide data favor a Myriapoda + Chelicerata (or, more accurately, Arachnida) pairing, although bootstrap support and posterior probability are not high (Fig. 1). Analyses of the amino acid data set, however, do not support a Myriapoda + Chelicerata clade (Fig. 4). Within Arachnida, the ATPase data generally support a close
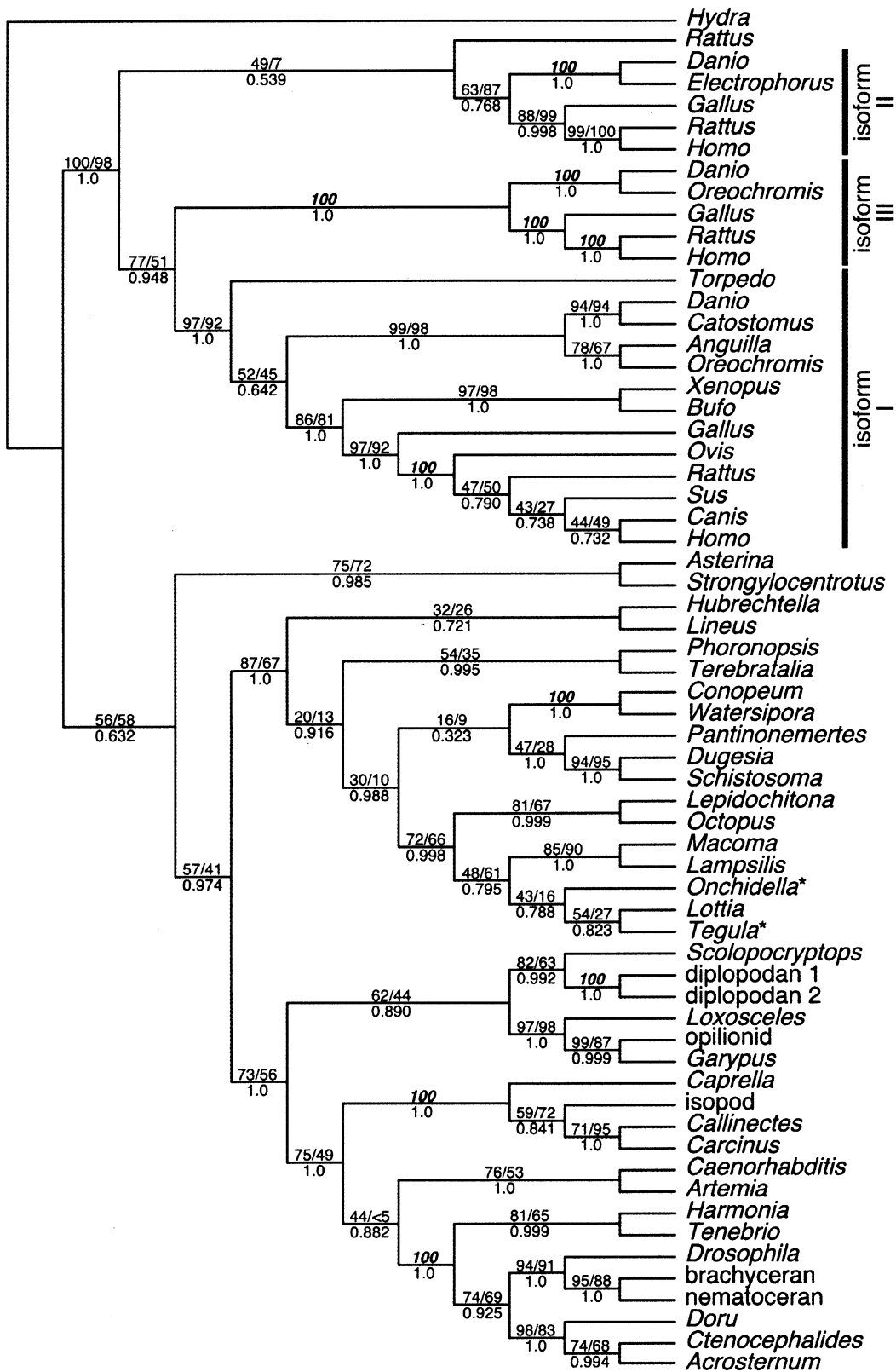
Fig. 3. Maximum likelihood tree for the large taxon set with different isoform groups highlighted. Support values are as described for Fig. 1.

relationship between an opilionid (harvestman) and *Garypus californica* (a pseudoscorpion), with *Loxosceles reclusa* (a spider) basal to this pair. Although relationships within Arachnida remain unclear, this

finding is also not unprecedented (see, e.g., Figs. 3–5 of Giribet and Ribera 2000).

Within Pancrustacea, a pairing of two brachyuran crabs (*Carcinus* and *Callinectes*) and monophyly for
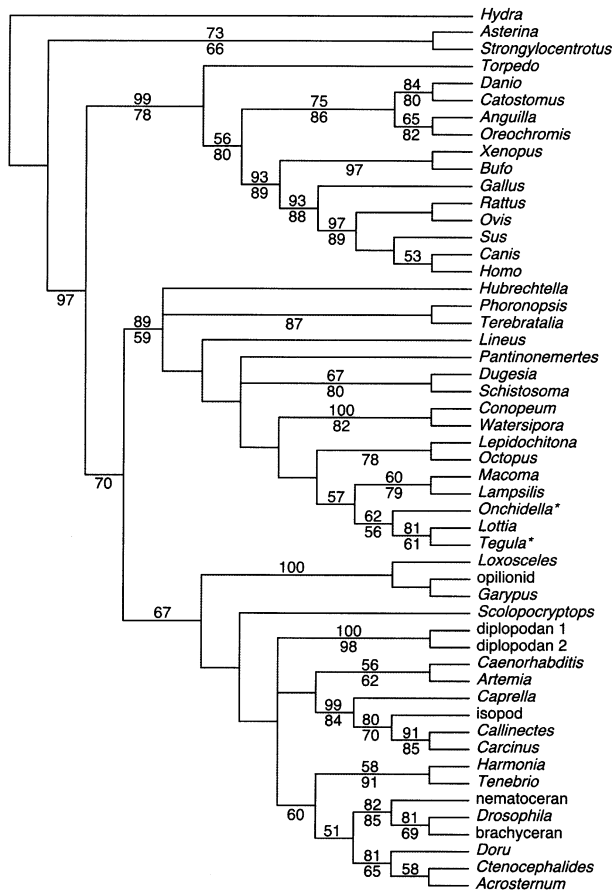
Fig. 4. The tree resulting from equal-weights maximum parsimony analysis of the amino acid data matrix. Values above the branches are maximum parsimony bootstrap support values, and values below the branches are maximum likelihood quartet puzzling support values (WAGf + Γ + I model); only values above 50 are shown.

Malacostraca (represented here by a caprellid amphipod, a terrestrial isopod, and the two brachyurans) are unsurprising. The ATPase data suggest, however, that Crustacea is paraphyletic not only with respect to Hexapoda, but also with respect to Nematoda, a rather unlikely proposition. The only non-malacostracan crustacean sequence in the analysis (*Artemia franciscana*, a branchiopod) lies closer to hexapods than to the malacostracan sequences. This finding is also incongruent with some recent publications based on 18S sequences (Giribet and Wheeler 1999) and complete mitochondrial genome sequences (Wilson et al. 2000), which propose a close relationship between malacostracans and hexapods, with branchiopods in a relatively basal position. Within Hexapoda, a monophyletic Coleoptera and Diptera were expected. However, the "DHS" (Dermaptera–Hemiptera–Siphonoptera) clade, and the hemipteran–flea pairing found within this clade, is radically different from previous hypotheses of hexapod relationships (Wheeler 2001 and citations therein).

With respect to the hexapod taxa included here, Wheeler et al, (2001) proposed a relatively close relationship between Siphonoptera and Diptera, with Hemiptera and Dermaptera outside this pairing in successively basal positions, a hypothesis clearly different from the ATPase tree. Taxon sampling is unquestionably an issue here—Wheeler and coworkers' study included 128 terminal taxa representing 33 hexapod orders, while only 8 sequences from 5 hexapod orders are included here.

Within Lophotrochozoa, apart from some phylum-level groupings (e.g., Bryozoa and Mollusca) and a phoronid–brachiopod pair, there is very little support for any particular set of interphylum relationships. Unfortunately, this has been the case in most molecular phylogenetic studies of relationships within this clade (e.g., Mallatt and Winchell 2002; Winnepenninckx et al. 1995). The majority of clades within Lophotrochozoa are supported in fewer than 50% of ML and MP bootstrap replicates, and Bayesian clade posterior probabilities are also relatively low (below 0.9) for several groups. It is becoming increasingly clear that single molecules (18S, EF-1α and ATPase) are insufficient to resolve interphylum relationships within Lophotrochozoa. Recently, several researchers have turned to analyses of combined data (both morphological and molecular) to resolve these issues (Mallatt and Winchell 2002; Zrzavy et al. 1998), thus far with limited success. It appears, however, that only combined analyses of multiple data sets will resolve relationships within this problematic group.

Two clades that are recovered with relatively high support on the ATPase ML tree are Mollusca and, within Mollusca, Bivalvia. In contrast, most studies of 18S data alone have failed to resolve a monophyletic Mollusca or, for that matter, a monophyletic Bivalvia (Adamkewicz et al. 1997; Canapa et al. 1999; Steiner and Muller 1996; Winnepenninckx et al. 1996). Taxon sampling within Mollusca in this study is very limited, however, and with additional molluscan exemplars, ATPase may also reveal disconcerting patterns of molluscan para- or polyphyly. This is particularly true for Bivalvia, as the two bivalve species sampled here—*Lampsilis cardium* (a freshwater unionid) and *Macoma nasuta* (a marine tellinid)—are fairly closely related (Brusca and Brusca 1990). More extensive taxon sampling is clearly necessary before strong conclusions can be made regarding mollusk (or bivalve) monophyly and the utility of this gene for resolving relationships within Mollusca or Lophotrochozoa.

Relationships Within Deuterostomia

Apart from the peculiar rooting of the tree between the echinoderms and the vertebrates (discussed fur-
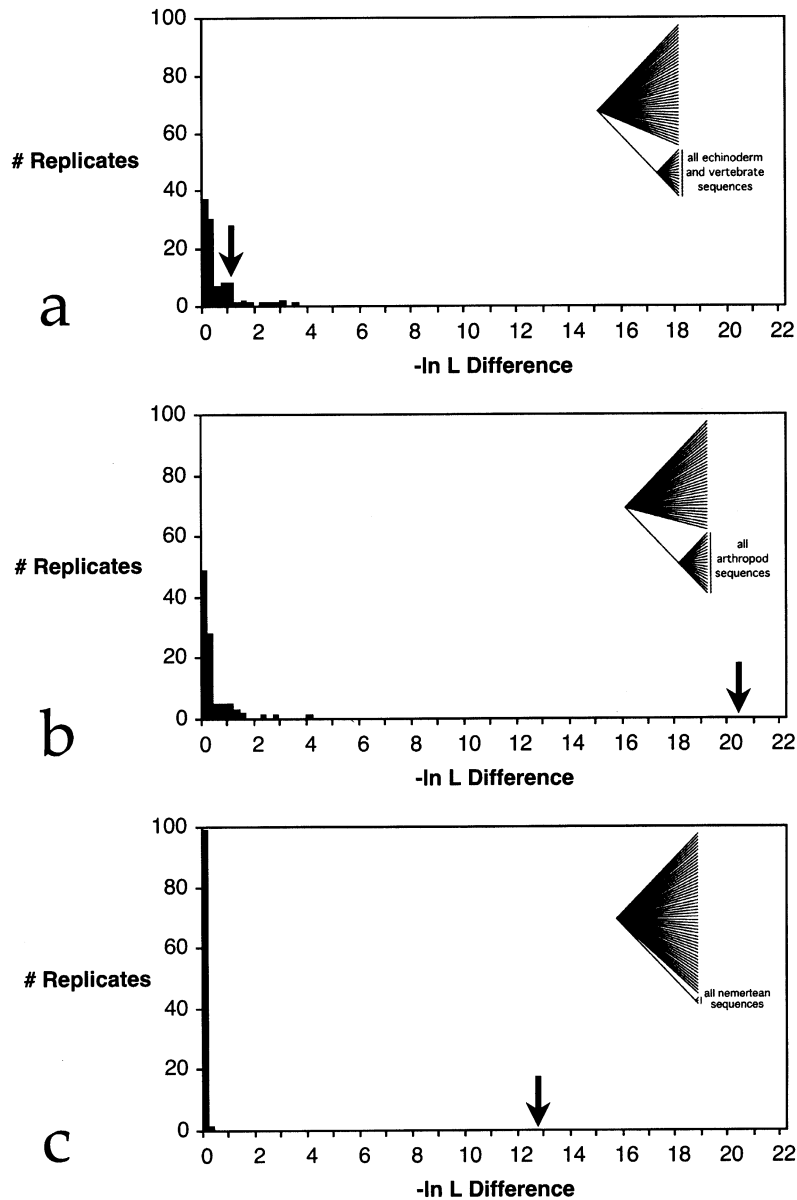
Fig. 5. Results of the parametric bootstrapping analyses. The −lnL difference between the best constrained and the best unconstrained trees for each test is denoted by an arrow. **a** Test of monophyly for Deuterostomia (including 2 echinoderm and 13 vertebrate sequences). **b** Test of monophyly for Arthropoda (19 arthropod sequences). **c** Test of monophyly for Nemertea (three nemertean sequences). The difference in likelihood scores for the original data were 1.10542415 (Deuterostomia monophyly; $1.0 > p > 0.05$), 20.44706 (Arthropoda monophyly; $p < 0.01$, and 12.90524706 (Nemertea monophyly; $p < 0.01$). Please see the text for further details.

ther below), this gene fragment successfully recovered nearly all postulated relationships within Deuterostomia, generally with high clade posterior probabilities and MP and ML bootstrap support. Mammalia (represented here by *Rattus norvegicus*, *Ovis aries*, *Sus scrofa*, *Canis familiaris*, and *Homo sapiens*), Amniota (*Gallus gallus* + Mammalia), Anura (*Bufo marinus* + *Xenopus laevis*), Tetrapoda (Anura + Amniota), and Actinopterygii (*Anguilla anguilla*, *Catostomus commersoni*, *Danio rerio*, and *Oreochromis mossambicus*) are all strongly supported, and Osteichthyes (Actinopterygii + Sarcopterygii, represented here only by various tetrapods) is also recovered in ML analyses of the large data set (Fig. 3). However, neither tree is completely concordant with traditional views of relationships within Actinopterygii or Mammalia. Within Actinopterygii, *Anguilla* (a relatively basal actinopterygian) is expected

to be sister to the other three actinopterygian species (Lauder and Liem 1983), although support for a *C. commersoni*–*D. rerio* clade was expected (as both species are cypriniforms). Support for branches within this region of the tree is not high, so this may simply be a region of the tree that ATPase sequences are unable to resolve accurately. Support for any particular hypothesis of relationships within Mammalia is also low, as has been seen in many other studies of mammal ordinal relationships (see Waddell et al. 1999b and citations therein).

Parametric Bootstrapping Tests

Despite the general congruence between the ATPase phylogeny and previous findings, some aspects of the trees presented here were surprising. Three of the

most peculiar findings—deuterostome, arthropod, and nemertean para/polyphyly—were investigated using parametric bootstrapping tests. There is substantial molecular and morphological support for monophyly of these groups (particularly Arthropoda and Deuterostomia), yet nearly all analyses of the ATPase fragment fail to recover these groups. The parametric bootstrapping tests of deuterostome, arthropod, and nemertean monophyly produced mixed results—deuterstome monophyly could not be rejected, but nemertean and arthropod monophyly was rejected. If the position of the outgroup *Hydra* sequence is ignored, the ATPase data do support deuterostome monophyly. The difference in likelihood scores between a "deuterostome monophyly" tree and the best ML topology is insignificant ($p > 0.08$; Fig. 5). The branch leading to the *Hydra* sequence is connected to the longest internal branch on the ML tree (the branch between the basal vertebrate node and the rest of the tree; Fig. 2), suggesting that long-branch attraction—even weak attraction, given the lack of signal supporting Deuterostomia—may be playing a role. Removal of the *Hydra* sequence prior to ML analysis produces a nearly identical topology, but on this unrooted phylogeny the two echinoderm sequences are no longer sisters—yielding the topology ((((Vertebrata) *Strongylocentrotus*) *Asterina*) all other sequences)—and the *Phoronopsis + Terebratalia* clade is sister to the *Hubrechtella + Lineus* clade (tree not shown). This loss of echinoderm monophyly when the *Hydra* sequence is excluded suggests that inclusion of even this single outgroup sequence may aid resolution in parts of the phylogeny. Additional outgroup sequences (from ctenophores, other cnidarians, or sponges) or sequences from other deuterostomes (e.g., hemichordates, urochordates, or cephalochordates) may allow a monophyletic Deuterostomia to be recovered. Some analyses of the alignment including vertebrate ATPase isoforms 2, 3, and 4 do recover deuterostome monophyly (trees not shown), suggesting that more thorough taxon sampling may alleviate this problem.

Parametric bootstrapping analysis rejects arthropod monophyly ($p < 0.01$; Fig. 5), but the rejection is solely due to the nematode (*C. elegans*) sequence. The nematode branch is one of the longest branches on the ML tree, and in this case, it is joined to the second-longest terminal branch (leading to *Artemia*) (Fig. 2). The parametric bootstrapping analysis strongly rejects arthropod monophyly, however, and sequential exclusion of the *Caenorhabditis* and *Artemia* sequences from the data set produces ML trees that are identical to those produced when both sequences are included (trees not shown). Sodium–potassium ATPase is known to play a critical role in the extraordinary osmoregulatory abilities of brine shrimp. The Na^+/K^+ α-subunit gene in *Artemia* has been studied extensively by Saez et al. (2000), who found higher levels of polymorphism in Na^+/K^+ ATPase α-subunit isoform 1 sequences than expected and suggested that this locus is evolving under positive selection. This could explain the high relative rate of change of this gene in *Artemia* and the seemingly spurious grouping of the *Artemia* sequence with the *Caenorhabditis* sequence.

Nemertean monophyly, based on the three sequences included here, is rejected by parametric bootstrapping analysis ($p < 0.01$; Fig. 5) and is also not recovered in any analysis of either the nucleotide or the amino acid data. This result can be examined in light of a recent phylogenetic study of nemertean relationships. Based on analyses of 16S, 28S, histone H3, and COI gene regions, Thollesson and Norenburg (2003) have proposed that Nemertea comprises two major clades—Hoplonemertea and Pilidiophora (a group comprising Heteronemertea and *Hubrechtella*)—and a paraphyletic "paleonemertean" assemblage. Two of the nemertean sequences included here are members of Pilidiophora (*Lineus* and *Hubrechtella*), while the other is a member of Hoplonemertea (*Pantinonemertes* sp. nov.). A *Lineus + Hubrechtella* pair is recovered in nearly all analyses based on DNA sequences, but the *Pantinonemertes* sequence typically forms a clade with the two platyhelminth sequences (from *Dugesia* and *Schistosoma*). In the quartet puzzling ML analysis of amino acid sequences, however, *Pantinonemertes* and *Lineus* form a clade to the exclusion of *Hubrechtella*, and there is no support for a nonmonophyletic Nemertea (the position of *Hubrechtella* is unresolved). It is thus likely that the odd result is an artifact in the analytical methods rather than a case of paralogy or sample contamination. As with other genes that have been used to evaluate metazoan relationships (e.g., 18S), there is low support for most clades within Lophotrochozoa on the ATPase trees (Figs. 1 and 4), but the peculiar position of *Pantinonemertes* (and the rejection of nemertean monophyly in the parametric bootstrapping test) merits further exploration.

The Phylogenetic Utility of the Sodium–Potassium ATPase α-Subunit Gene

Will the sodium–potassium ATPase α-subunit gene region studied here be useful for future deep-level studies of animal phylogeny? Practical as well as empirical considerations are required to address this question fully. To be of practical use for molecular phylogenetic studies, the fragment must be amenable to PCR amplification across several phyla (or at least within phyla of interest). The gene should also be single copy, so that erroneous comparisons among paralogous gene copies can be avoided (Martin and

Burg 2002). Alternatively, if there are multiple copies of the gene, sequences in the PCR priming region should allow PCR to amplify only orthologs. Empirical considerations might include the ability of a novel gene fragment to recover clades that are well supported based on studies of morphological data or sequence data from other genes. In addition, the ability of a new phylogenetic marker to resolve relationships that other data sets have failed to resolve (or to aid in the resolution of such relationships in combination with other sources of data) may also suggest that a new marker can provide an important contribution.

Is the ATPase fragment studied here practically useful for metazoan phylogenetic studies? We believe that it can be useful for some questions. The gene was successfully amplified from representatives of seven phyla (Arthropoda, Brachiopoda, Bryozoa, Echinodermata, Mollusca, Phoronida, and Nemertea) and is particularly easy to amplify from some subgroups (e.g., Hexapoda). Future studies of relationships within groups where amplification was least problematic may benefit from inclusion of data from this gene. Unfortunately, the failure of PCR amplification for representatives of several phyla (most notably annelids, sipunculans, cnidarians, and nematodes) limits the usefulness of these primer pairs in studies of relationships among the metazoan phyla. Future work on this gene will involve further attempts to amplify the gene from additional phyla and evaluation of additional primer pairs designed using the CODEHOP approach described above.

There is some evidence from previous studies (Baxter-Lowe et al. 1989; Friedlander et al. 1992) that this gene is multicopy in some taxa. In this study, several ambiguities were found in two of the three arachnids sequenced (*Garypus* and *Loxosceles*), suggesting that there may be two (or more) functional copies of this gene in at least some arachnids. The recovery of a monophyletic Arachnida suggests that this putative duplication may only be an issue within this group. Any future phylogenetic studies using this gene will require a more thorough examination of gene copy number in the taxon (or taxa) of interest.

Finally, are there empirical benefits to the use of this ATPase α-subunit gene region in metazoan phylogenetic studies? The divergences between taxa studied here span a broad range of time. For example, minimum estimated time depths based on other data for splits among the major vertebrate groups represented here range from 65/92 million years ago (mya) for the split between carnivores and primates all the way to 439/528 mya (the split between Chondrichthyes and Osteicthyes). Between these extremes lie the synapsid–diapsid split (310 mya), the Lissamphibia–Amniota split (323/360 mya), and the Actinopterygii–Sarcopterygia split (424/450 mya) (all

pairs of dates are crude fossil-based estimates derived from several sources [Alroy 1999; Benton 1997; Janvier 1996; Zug et al. 2001] and molecular clock-based estimates [Kumar and Hedges 1998; Waddell et al. 1999a], respectively). The finding that the Na^+/K^+ ATPase α-subunit accurately recovers these splits with high support suggests that splits of similar time depth within other metazoan phyla might be investigated fruitfully using this gene. In addition, the Lophotrochozoa/Ecdysozoa split occurred at least 560 mya, given fossil considerations and a recent Bayesian estimate of the minimum divergence time for these two clades (Aris-Brosou and Yang 2002). The recovery of these two clades in all analyses, along with generally reasonable patterns of relationships within Vertebrata, suggests that this gene fragment may be useful for both intra- and interphylum relationships within Metazoa, particularly in combination with data from other genes.

References

Adamkewicz SL, Harasewych MG, Blake J, Saudek D, Bult CJ (1997) A molecular phylogeny of the bivalve mollusks. Mol Biol Evol 14:619–629

Aguinaldo AMA, Turbeville JM, Lindford LS, Rivera MC, Garey JR, Raff RA, Lake JA (1997) Evidence for a clade of nematodes, arthropods and other moulting animals. Nature 387:389–393

Alroy J (1999) The fossil record of North American mammals: Evidence for a Paleocene evolutionary radiation. Syst Biol 48:107–118

Anderson FE (2000) Phylogeny and historical biogeography of the loliginid squids (Mollusca: Cephalopoda) based on mitochondrial DNA sequence data. Mol Phylogenet Evol 15:191–214

Aris-Brosou S, Yang ZH (2002) Effects of models of rate evolution on estimation of divergence dates with special reference to the metazoan 18S ribosomal RNA phylogeny. Syst Biol 51:703–714

Baxter-Lowe LA, Guo JZ, Bergstrom EE, Hokin LE (1989) Molecular cloning of the Na,K-ATPase alpha-subunit in developing brine shrimp and sequence comparison with higher organisms. FEBS Lett 257:181–187

Benton MJ (1997) Vertebrate palaeontology. Chapman & Hall, London

Blair JE, Ikeo K, Gojobori T, Hedges SB (2002) The evolutionary position of nematodes. BMC Evol Biol 2:7

Blanchette M, Kunisawa T, Sankoff D (1999) Gene order breakpoint evidence and animal mitochondrial phylogeny. J Mol Evol 49:193–203

Boore JL (1999) Animal mitochondrial genomes. Nucleic Acids Res 27:1767–1780

Boore JL, Brown WM (2000) Mitochondrial genomes of Galathealinum, Helobdella, and Platynereis: Sequence and gene arrangement comparisons indicate that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. Mol Biol Evol 17:87–106

Boore JL, Staton JL (2002) The mitochondrial genome of the sipunculid Phascolopsis gouldii supports its association with Annelida rather than Mollusca. Mol Biol Evol 19:127–137

Boore JL, Lavrov DV, Brown WM (1998) Gene translocation links insects and crustaceans. Nature 392:667–668

Borchiellini C, Boury-Esnault N, Vacelet J, Le Parco Y (1998) Phylogenetic analysis of the Hsp70 sequences reveals the monophyly of Metazoa and specific phylogenetic relationships between animals and fungi. Mol Biol Evol 15:647–655

Brusca RC, Brusca GJ (1990) Invertebrates. Sinauer Associates, Sunderland, MA

Canapa A, Marota I, Rollo F, Olmo E (1999) The small-subunit rRNA gene sequences of venerids and the phylogeny of Bivalvia. J Mol Evol 48:463–468

Cao Y, Waddell PJ, Okada N, Hasegawa M (1998) The complete mitochondrial DNA sequence of the shark Mustelus manazo: Evaluating rooting contradictions to living bony vertebrates. Mol Biol Evol 15:1637–1646

Castresana J, Feldmaier-Fuchs G, Yokobori S, Satoh N, Paabo S (1998) The mitochondrial genome of the hemichordate Balanoglossus carnosus and the evolution of deuterostome mitochondria. Genetics 150:1115–1123

Cho S, Mitchell A, Regier JC, Mitter C, Poole RW, Friedlander TP, Zhao S (1995) A highly conserved nuclear gene for low-level phylogenetics: Elongation factor-1 alpha recovers morphology-based tree for heliothine moths. Mol Biol Evol 12:650–656

Christen R, Ratto A, Baroin A, Perasso R, Grell KG, Adoutte A (1991) An analysis of the origin of metazoans, using comparisons of partial sequences of the 28S RNA, reveals an early emergence of triploblasts. EMBO J 10:499–503

de Rosa R, Grenier JK, Andreeva T, Cook CE, Adoutte A, Akam M, Carroll SB, Balavoine G (1999) Hox genes in brachiopods and priapulids and protostome evolution. Nature 399:772–776

Dowtin M, Austin AD (1999) Evolutionary dynamics of a mitochondrial rearrangement "hot spot" in the Hymenoptera. Mol Biol Evol 16:298–309

Eeckhaut I, McHugh D, Mardulyn P, Tiedemann R, Monteyne D, Jangoux M, Milinkovitch MC (2000) Myzostomida: A link between trochozoans and flatworms? Proc R Soc Lond B 267:1383–1392

Eernisse DJ, Albert JS, Anderson FE (1992) Annelida and Arthropoda are not sister taxa: A phylogenetic analysis of spiralian metazoan morphology. Syst Biol 41:305–330

Erber A, Riemer D, Bovenschulte M, Weber K (1998) Molecular phylogeny of metazoan intermediate filament proteins. J Mol Evol 47:751–762

Erber A, Riemer D, Hofemeister H, Bovenshulte M, Stick R, Panopoulou G, Lebrach H, Weber K (1999) Characterization of the Hydra lamin and its gene: A molecular phylogeny of metazoan lamins. J Mol Evol 49:260–271

Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–791

Felsenstein J (1995) PHYLIP (phylogeny inference package). Distributed by the author, Department of Genetics, University of Washington, Seattle

Field KG, Olsen GJ, Lane DJ, Giovannoni SJ, Ghiselin MT, Raff EC, Pace NR, Raff RA (1988) Molecular phylogeny of the animal kingdom. Science 239:748–753

Flook P, Rowell H, Gellissen G (1995) Homoplastic rearrangements of insect mitochondrial transfer-RNA genes. Naturwissenschaften 82:336–337

Frati F, Simon C, Sullivan J, Swofford DL (1997) Evolution of the mitochondrial cytochrome oxidase II gene in Collembola. J Mol Evol 44:145–158

Friedlander TP, Regier JC, Mitter C (1992) Nuclear gene sequences for higher level phylogenetic analysis: 14 promising candidates. Syst Biol 41:483–489

Friedrich M, Tautz D (1995) Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution of myriapods. Nature 376:165–167

Ghiselin MT (1989) Summary of our present knowledge of metazoan phylogeny. In: Fernholm B, Bremer K, Jörnvall H (eds) The hierarchy of life. Elsevier, Amsterdam, pp 261–272

Giribet G, Ribera C (1998) The position of arthropods in the animal kingdom: A search for a reliable outgroup for internal arthropod phylogeny. Mol Phylogenet Evol 9:481–488

Giribet G, Ribera C (2000) A review of arthropod phylogeny: New data based on ribosomal DNA sequences and direct character optimization. Cladistics 16:204–231

Giribet G, Wheeler WC (1999) The position of arthropods in the animal kingdom: Ecdysozoa, islands, trees and the "parsimony ratchet." Mol Phylogenet Evol 13:619–623

Giribet G, Carranza J, Baguñà J, Riutort M, Ribera C (1996) First molecular evidence for the existence of a Tardigrada + Arthropoda clade. Mol Biol Evol 13:76–84

Giribet G, Distel DL, Polz M, Sterrer W, Wheeler WC (2000) Triploblastic relationships with emphasis on the acoelomates and the position of Gnathostomulida, Cycliophora, Plathelminthes, and Chaetognatha: A combined approach of 18S rDNA sequences and morphology. Syst Biol 49:539–562

Giribet G, Edgecombe GD, Wheeler WC (2001) Arthropod phylogeny based on eight molecular loci and morphology. Nature 413:157–161

Halanych KM (1998) Considerations for reconstructing metazoan history: Signal, resolution, and hypothesis testing. Am Zool 38:929–941

Halanych KM, Bacheller JD, Aguinaldo AM, Liva SM, Hillis DM, Lake JA (1995) Evidence from 18S ribosomal DNA that the lophophorates are protostome animals. Science 267:1641–1643

Hickerson MJ, Cunningham CW (2000) Dramatic mitochondrial gene rearrangements in the hermit crab Pagurus longicarpus (Crustacea, Anomura). Mol Biol Evol 17:639–644

Hillis DM, Mable B, Moritz C (1996) Applications of molecular systematics and the future of the field. In: Hillis DM, Mable B, Moritz C (eds) Molecular systematics. Sinauer Associates, Sunderland, MA, pp 515–543

Huelsenbeck JP, Hillis DM, Jones R (1996) Parametric bootstrapping in molecular phylogenetics: Applications and performance. In: Ferraris JD, Palumbi SR (eds) Molecular zoology: Advances, strategies and protocols. Wiley–Liss, New York, pp 19–46

Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogeny. http://morphbank.ebc.uu.se/mrbayes/

Hwang UW, Friedrich M, Tautz D, Park CJ, Kim W (2001) Mitochondrial protein phylogeny joins myriapods with chelicerates. Nature 413:154–157

Janvier P (1996) Early vertebrates. Clarendon Press, New York

Kobayashi M, Wada H, Satoh N (1996) Early evolution of the Metazoa and phylogenetic status of the diploblasts as inferred from amino acid sequence of elongation factor-1α. Mol Phylogenet Evol 5:414–422

Kojima S (1998) Paraphyletic status of Polychaeta suggested by phylogenetic analysis based on the amino acid sequences of elongation factor-1α. Mol Phylogenet Evol 9:255–261

Kojima S, Hashimoto T, Hasegawa M, Murata S, Ohta S, Seki H, Okada N (1993) Close phylogenetic relationship between Vestimentifera (tube worms) and Annelida revealed by the amino acid sequence of elongation factor-1α. J Mol Evol 37:66–70

Kruse M, Leys SP, Muller IM, Muller WEG (1998) Phylogenetic position of the Hexactinellida within the phylum Porifera based on the amino acid sequence of the protein kinase C from *Rhabdocalyptus dawsoni*. J Mol Evol 46:721–728

Kumar S, Hedges SB (1998) A molecular timescale for vertebrate evolution. Nature 392:917–920

Kusche K, Burmester T (2001) Diplopod hemocyanin sequence and the phylogenetic position of the Myriapoda. Mol Biol Evol 18:1566–1573

Lauder GV, Liem KF (1983) The evolution and interrelationships of the actinopterygian fishes. Bull Mus Comp Zool 150:95–197

Le TH, Blair D, Agatsuma T, Humair P-F, Campbell NJH, Iwagami M, Littlewood TJ, Peacock B, Johnston DA, Bartley J, Rollinson D, Herniou EA, Zarlenga DS, McManus DP (2000) Phylogenies inferred from mitochondrial gene orders—A cautionary tale from the parasitic flatworms. Mol Biol Evol 17:1123–1125

Littlewood DT, Olson PD, Telford MJ, Herniou EA, Riutort M (2001) Elongation factor 1-alpha sequences alone do not assist in resolving the position of the acoela within the metazoa. Mol Biol Evol 18:437–442

Maddison DR, Maddison WP (2001) MacClade 4: Analysis of phylogeny and character evolution. Sinauer Associates, Sunderland, MA

Mallatt J, Winchell CJ (2002) Testing the new animal phylogeny: First use of combined large-subunit and small-subunit rRNA gene sequences to classify the protostomes. Mol Biol Evol 19:289–301

Martin AP, Burg TM (2002) Perils of paralogy: Using HSP70 genes for inferring organismal phylogenies. Syst Biol 51:570–587

McHugh D (1997) Molecular evidence that echiurans and pogonophorans are derived annelids. Proc Natl Acad Sci USA 94:8006–8009

Mindell DP, Sorenson MD, Dimcheff DE (1998) Multiple independent origins of mitochondrial gene order in birds. Proc Natl Acad Sci USA 95:10693–10697

Mushegian AR, Garey JR, Martin J, Liu L-X (1998) Large-scale taxonomic profiling of eukaryotic model organisms: A comparison of orthologous proteins encoded by the human, fly, nematode and yeast genomes. Genome Res 8:590–598

Nielsen C, Scharff N, Eibye-Jacobsen D (1996) Cladistic analyses of the animal kingdom. Biol J Linn Soc Lond 57:385–410

Nikoh N, Iwabe N, Kuma K-i, Ohno M, Sugiyama T, Watanabe Y, Yasui K, Shi-cui Z, Hori K, Shimura Y, Miyata T (1997) An estimate of divergence time for Parazoa and Eumetazoa and that of Cephalochordata and Vertebrata by aldolase and triose phosphate isomerase clocks. J Mol Evol 45:97–106

Peterson KJ, Eernisse DJ (2001) Animal phylogeny and the ancestry of bilaterians: Inferences from morphology and 18S rDNA gene sequences. Evol Dev 3:170–205

Posada D, Crandall KA (1998) Modeltest: Testing the model of DNA substitution. Bioinformatics 14:817–818

Rambaut A (1995) Se-Al: Sequence Alignment Program. Distributed by the author

Rawlings TA, Collins TM, Bieler R (2001) A major mitochondrial gene rearrangement among closely related species. Mol Biol Evol 18:1604–1609

Regier JC, Shultz JW (1997) Molecular phylogeny of the major arthropod groups indicates polyphyly of crustaceans and a new hypothesis for the origin of hexapods. Mol Biol Evol 14:902–913

Regier JC, Shultz JW (1998) Molecular phylogeny of arthropods and the significance of the Cambrian "explosion" for molecular systematics. Am Zool 38:918–928

Regier JC, Shultz JW (2001) Elongation factor-2: a useful gene for arthropod phylogenetics. Mol Phylogenet Evol 20:136–148

Rokas A, Holland PW (2000) Rare genomic changes as a tool for phylogenetics. Trends Ecol Evol 15:454–459

Rose TM, Schultz ER, Henikoff JG, Pietrokovski S, McCallum CM, Henikoff S (1998) Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. Nucleic Acids Res 26:1628–1635

Ruiz-Trillo I, Riutort M, Littlewood DT, Herniou EA, Baguna J (1999) Acoel flatworms: Earliest extant bilaterian metazoans, not members of Platyhelminthes. Science 283:1919–1923

Ruiz-Trillo I, Paps J, Loukota M, Ribera C, Jondelius U, Baguna J, Riutort M (2002) A phylogenetic analysis of myosin heavy chain type II sequences corroborates that Acoela and Nemertodermatida are basal bilaterians. Proc Natl Acad Sci USA 99:11246–11251

Saez AG, Escalante R, Sastre L (2000) High DNA sequence variability at the alpha 1 Na/K-ATPase locus of *Artemia franciscana* (brine shrimp): Polymorphism in a gene for salt-resistance in a salt-resistant organism. Mol Biol Evol 17:235–250

Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18:502–504

Shao R, Campbell NJ, Barker SC (2001a) Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthiraptera). Mol Biol Evol 18:858–865

Shao R, Campbell NJ, Schmidt ER, Barker SC (2001b) Increased rate of gene rearrangement in the mitochondrial genomes of three orders of hemipteroid insects. Mol Biol Evol 18:1828–1832

Shultz JW, Regier JC (2000) Phylogenetic analysis of arthropods using two nuclear protein-encoding genes supports a crustacean + hexapod clade. Proc R Soc Lond B 267:1011–1019

Steiner G, Muller M (1996) What can 18S rDNA do for bivalve phylogeny? J Mol Evol 43:58–70

Strimmer K, von Haeseler A (1996) Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies. Mol Biol Evol 13:964–969

Swofford DL (2002) PAUP*. Phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, MA

Thollesson M, Norenburg JL (2003) Ribbon worm relationships—A phylogeny of the phylum Nemertea. Proc R Soc Lond B 270:407–415

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882

Turbeville JM, Pfeifer DM, Field KG, Raff RA (1991) The phylogenetic status of arthropods, as inferred from 18S rRNA sequences. Mol Biol Evol 8:669–686

von Nickisch-Rosenegk M, Brown WM, Boore JL (2001) Complete sequence of the mitochondrial genome of the tapeworm *Hymenolepis diminuta*: Gene arrangements indicate that platyhelminths are eutrochozoans. Mol Biol Evol 18:721–730

Waddell PJ, Cao Y, Hasegawa M, Mindell DP (1999a) Assessing the Cretaceous superordinal divergence times within birds and placental mammals by using whole mitochondrial protein sequences and an extended statistical framework. Syst Biol 48:119–137

Waddell PJ, Okada N, Hasegawa M (1999b) Towards resolving the interordinal relationships of placental mammals. Syst Biol 48:1–5

Wang DY-C, Kumar S, Hedges SB (1999) Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. Proc R Soc Lond B 266:163–171

Wheeler WC, Cartwright P, Hayashi CY (1993) Arthropod phylogeny—A combined approach. Cladistics 9:1–39

Wheeler WC, Whiting M, Wheeler QD, Carpenter JM (2001) The phylogeny of the extant hexapod orders. Cladistics 17:113–169

Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18:691–699

Wilson K, Cahill V, Ballment E, Benzie J (2000) The complete sequence of the mitochondrial genome of the crustacean *Penaeus monodon*: Are malacostracan crustaceans more closely related to insects than to branchiopods? Mol Biol Evol 17:863–874

Winnepenninckx B, Backeljau T, De Wachter R (1995) Phylogeny of protostome worms derived from 18S rRNA sequences. Mol Biol Evol 12:641–649

Winnepenninckx B, Backeljau T, DeWachter R (1996) Investigation of molluscan phylogeny on the basis of 18S rRNA sequences. Mol Biol Evol 13:1306–1317

Zrzavy J, Mihulka S, Kepka P, Bezdek A, Tietz D (1998) Phylogeny of the Metazoa based on morphological and 18S ribosomal DNA evidence. Cladistics 14:249–285

Zug GR, Vitt LJ, Caldwell JP (2001) Herpetology: An introductory biology of amphibians and reptiles. Academic Press, San Diego