28S and 18S rDNA Sequences Support the Monophyly of Lampreys and Hagfishes

Jon Mallatt and Jack Sullivan†*

*Department of Zoology, Washington State University; and †Department of Biological Sciences, University of Idaho

Resolving the interrelationships of three major extant lineages of vertebrates (hagfishes, lampreys, and gnathostomes) is a particularly important issue in evolution, because the basal resolution critically influences our understanding of primitive vertebrate characters. A consensus has emerged over the last 20 years that lampreys are the sister group to the gnathostomes and the hagfishes represent an ancient, basal lineage. This hypothesis has essentially displaced the classical hypothesis of monophyly of the cyclostomes (lampreys plus hagfishes). To test these hypotheses, we compared nearly complete ribosomal DNA sequences from each of these major lineages, as well as those from a cephalochordate and a urochordate, which represent a paraphyletic outgroup for assessing the basal vertebrate relationships. For this comparison, 92%–99% complete 28S rDNA sequences were obtained from the lancelet *Branchiostoma floridae*, the hagfish *Eptatretus stouti*, the lamprey *Petromyzon marinus*, and cartilaginous fishes *Hydrolagus colliei* and *Squalus acanthias* and were then analyzed with previously reported 28S and 18S rDNA sequences from other chordates. We conducted conventional (nonparametric) bootstrap analyses, under maximum-likelihood, parsimony, and minimum-evolution (using LogDet distances) criteria, of both 28S and 18S rDNA sequences considered separately and combined. All these analyses provide moderate to very strong support for the monophyly of the cyclostomes. Furthermore, the currently accepted hypothesis of a lamprey-gnathostome clade is moderately rejected by the Kishino-Hasegawa test ($P = 0.099$) and resoundingly rejected by parametric bootstrap tests ($P < 0.01$) in favor of monophyly of living cyclostomes. Another significant finding is that the hagfish *E. stouti* has the longest 28S rDNA gene known in any organism (>5,200 nt).

Introduction

Lampreys and hagfishes comprise the living jawless vertebrates, whereas all other extant vertebrates are jawed (gnathostomes). The question of how these three groups are interrelated is critical to understanding vertebrate evolution because it involves the most basal branches of the vertebrate phylogeny and therefore has great influence on our assessment of primitive vertebrate characters. The first jawless vertebrates are thought to have evolved about 500 MYA, followed by the first gnathostomes about 460 MYA (Sansom, Smith, and Smith 1996). Although traditional classifications united lampreys with hagfishes as cyclostomes, the increasing influence of cladistics in the 1970s led to a reassessment of their relationships (Hardisty 1979, 1982; Janvier 1981). In recent years, a consensus has emerged that lampreys are the sister group to gnathostomes, with the hagfish lineage having diverged as the basal lineage. Agreement on this is remarkably complete among vertebrate morphologists and paleontologists (Nelson 1994; Janvier 1996; Rovainen 1996; Sower 1998; Tree of Life Website: http://phylogeny.arizona.edu/tree/phylogeny.html). Under any scenario, however, it seems likely that all three groups are very distantly related, and that hagfishes have many uniquely derived characters that result in ambiguous assessment of homologies and thereby confuse phylogenetic analyses. Structural characters thought to be synapomorphies uniting lampreys and gnathostomes include the presence of at least rudimentary vertebral arches, Mauthner neurons in the brainstem, renal collecting ducts, a hypoglossal nerve, at least a rudimentary spleen, true taste buds, and loss of a subcutaneous nerve plexus/slime-gland plexus (Hardisty 1982; Braun 1996; H. Wicht, personal communication).

However, three voices have spoken against this new consensus: (1) Mallatt (1997*a*, 1997*b*) argued that hagfishes and lampreys uniquely share many structural synapomorphies, including a large vertically-biting ''tongue'' apparatus (that itself contains many putative synapomorphies; Yalden 1985), a highly elongated oral region, and a respiratory flap called the velum; none of these characters can be either seen or reasonably reconstructed in the fossil jawless fishes (ostracaderms) that would be expected to reveal ancestral vertebrate characters. (2) At the molecular level, Stock and Whitt (1992) provided evidence that 18S ribosomal DNA sequences support a lamprey/hagfish grouping with strong bootstrap values from both likelihood and parsimony analyses; this conclusion was upheld by the reanalysis of Turbeville, Schulz, and Raff (1994). (3) Lanfranchi et al. (1994) conducted a parsimony analysis of hemoglobin amino acid sequences and reached the same conclusion, although their outgroup was not ideal (an echinoderm, rather than nonvertebrate chordates, was used).

By contrast, another recent molecular analysis used vasotocin and united lampreys with gnathostomes (Suzuki et al. 1995). However, this finding might be questioned, because an extremely distant outgroup was used (snail). Outgroup choice is especially critical in resolving basal relationships from molecular data because of long-branch-attraction problems (Felsenstein 1978) and uncertainty in rooting (see below).

Here, we extend Stock and Whitt's (1992) study of 18S rDNA by sequencing and analyzing almost the entire 28S rRNA gene (plus part of the small 5.8S rRNA gene) from a hagfish, a lamprey, a lancelet, and some

Key words: cyclostome, chordate phylogeny, maximum likelihood, gnathostome relationships, 28S rDNA, LSU rDNA.

Address for correspondence and reprints: Jon Mallatt, Box 644236, Department of Zoology, Washington State University, Pullman, Washington 99164-4236. E-mail: jmallatt@mail.wsu.edu.

Mol. Biol. Evol. 15(12):1706–1718. 1998

gnathostome taxa. As in Stock and Whitt's study, our alternative hypotheses of basal-vertebrate relationships are (1) the classical cyclostome view that lampreys and hagfishes are monophyletic versus (2) the currently accepted view that lampreys and gnathostomes are sister taxa. When analyzed suitably, 28S sequences should be appropriate for reconstructing deep phylogenetic branches such as the basal radiation of the vertebrates (Hillis and Dixon 1991), and should yield at least as much phylogenetic information as does the 18S gene (e.g., Kuzoff et al. 1998). In addition to providing a large new data set, this molecular-phylogenetic approach has the following advantages over the morphology-based approach in resolving cyclostome–gnathostome relations: (1) vertebrates are so anatomically dissimilar from their chordate relatives, cephalochordates (lancelets) and urochordates (tunicates), that it is difficult to assess homologies and determine primitive character states for anatomical structures; (2) molecular phylogenetic studies can use powerful maximum-likelihood methods which accommodate various well-characterized features of molecular evolution, including different rates for different types of nucleotide substitution (e.g., Yang 1994*a*), among-site rate heterogeneity (e.g., Yang 1994*b*; Gu, Fu, and Li 1995; Waddell and Penny 1996), and unequal base frequencies (e.g., Felsenstein 1981). These model-based approaches are not available for morphological analyses, which are therefore restricted to the use of parsimony (nor were such improved models available to Stock and Whitt [1992] for their systematic study using 18S rDNA).

Materials and Methods

Specimens and DNA Extraction

Nearly complete 28S and partial 5.8S rDNA sequences were obtained from five chordate species (GenBank accession numbers AF061796–AF061800): three individuals each of (1) the lancelet *Branchiostoma floridae* (4–4.5 cm long, from Tampa Bay, Fla.), (2) the hagfish *Eptatretus stouti* (22–26 cm, collected off La Jolla in 150 m of water at Scripps Institution of Oceanography, Calif.), (3) the lamprey *Petromyzon marinus* (46–54 cm, Cheboygan River, Cheboygan County, Mich.), (4) the chimaeroid *Hydrolagus colliei* (a cartilaginous fish; 40–54 cm, from Puget Sound), and (5) the dogfish shark *Squalus acanthias*, (1 m, from the Seattle Aquarium, originally from Puget Sound). Additionally, one trout *Oncorhynchus mykiss* was used (from Oregon State University Hatchery, Corvallis) to provide a small segment of the 28S gene not already sequenced by Zardoya and Meyer (1996; AF061801; see table 1). For all individuals, 0.5-cm^2 pieces were clipped from the tails and preserved immediately in 95% ethanol; thus, the DNA came mainly from skin and muscle. DNA was extracted with the CTAB method (Winnepenninckx, Backeljau, and DeWachter 1993).

DNA Amplification and Purification

The large-subunit (LSU) rRNA genes were amplified entirely by PCR. For the initial amplification, the reaction mixtures contained: 10.05 μl of millipore-filtered distilled water, 2 μl of 2.5 mM dNTP solution, 5 μl of PCR Buffer B from the Invitrogen PCR Optimizer Kit (pH 8.5, $[Mg^{2+}] = 2.0$ mM), 0.25 μl of *Taq* polymerase (5–8 U/μl), 0.4 μl of Stratagene *Taq* Extender PCR Additive, 1.6 μl of DMSO, 2.5 μl each of two primer solutions (10 pM/μl), and 1 μl of DNA template (3–20 ng/μl) covered with 2 drops of mineral oil. The amplifications were performed in an Amplitron II (Thermolyne) cycler with the following settings: (1) a predwell at 94°C for 4 min; (2) 30 cycles of: 94°C for 1 min, 53–61°C for 1 min, 72°C for 3:15 min; and (3) a postdwell at 72°C for 10 min. To generate sufficient DNA for sequencing, reamplification was performed as follows: the initial PCR product was separated by electrophoresis on a low-melt gel (SeaPlaque agarose, FMC BioProducts; stained with ethidium bromide), excised, and diluted three times in TE (pH 7.5); then, 1–2 μl of this solution was used as the DNA template under the same PCR-reaction conditions as given above, except with fewer than 25 amplification cycles; products were again gel-purified, and then extracted using GeneClean III (Bio 101, Inc.). Template concentrations were measured on a Hoefer DNA fluorometer Model TKO 100 using Hoechst dye, then diluted to between 3.5 and 10 ng/μl for sequencing (see below).

Primers

The following primers (from Oligos Etc., Inc., Wilsonville, Oreg.) were used for amplification and sequencing in all new taxa. They extend from the 3′ end of the 18S rRNA gene through the internal transcribed spacers and the 5.8S gene, to about 55 nt upstream from the 3′ end of the 28S gene, and are separated by an average of 300 nt: universal primers 4, 5, 12, 13, 16–19, 26, 27, 30, and 31 of Van der Auwera, Chapelle, and De Wachter (1994); primers D8 and C2, and their complements, of Philippe et al. (1994); primers 18d, 5.8c, 5.8d, 28dd, ee, ff, hh, ii, w, x, and z of Hillis and Dixon (1991); plus a new 28S primer nearest the 3′ end: ACTTTCAATAGATCGCAG. Two additional sequencing primers were required for hagfishes, from the middle of the long D2 and D8 domains: TGGGTGATCCACCGGGTCCG and CCCTCTTTCGTGGAGGTGCGGG. For all taxa but hagfish, the rRNA gene sequence was amplified in three overlapping fragments as follows: fragment 1, from the 3′ end of the 18S gene (primer 18d above) or the 5.8S gene (primer 5.8d) to about 300 nt into the 28S gene (primer C2); fragment 2, from about 55 nt after the 5′ end (primer 4) to about 300 nt before the 3′ end (primer 31) of the 28S gene; fragment 3, from about 500 nt before (primer 28ii) to about 55 nt before (new primer ACTTTCAATAGATCGCAG) the 3′ end of the 28S gene. In hagfish, the second of these fragments (the bulk of the 28S gene) was so long that it had to be amplified in three roughly equal subfragments.

Sequencing

DNA sequencing was performed on an automated ABI 373 DNA Sequencer using fluorescent dye-termi-

Table 1
Parts of the rRNA Genes Used, from Different Chordates

Taxa for which new LSU rRNA sequences (28S and partial 5.8S) were determined in this study

1. *Hydrolagus colliei* (chimaera: jawed cartilaginous fish)
   Available: all of 18S gene except first ~40 nt (Stock 1992); all of 5.8S; all of 28S except last ~54 nt at 3′ end (present study; GenBank accession number AF061799)
   Used: all of available 18S and 28S, last 50 nt of 5.8S
2. *Squalus acanthias* (shark: jawed cartilaginous fish)
   Available: all of 18S gene except first ~60 nt (Bernardi, Sordino, and Powers 1992; GenBank accession number M91179); no 5.8S; all of 28S except the last ~300 nt (present study; GenBank accession number AF061800)
   Used: all available
3. *Petromyzon marinus* (lamprey)
   Available: all of 18S gene except first ~40 nt (Stock and Whitt 1992; GenBank accession number M97575); all of 5.8S; all of 28S except the last ~54 nt (present study; GenBank accession number AF061798)
   Used: all of available 18S and 28S, last 50 nt of 5.8S
4. *Eptatretus stouti* (hagfish)
   Available: all of 18S except first ~40 nt (Stock and Whitt 1992; GenBank accession number M97572); last 50 nt of 5.8S; all of 28S except last ~54 nt (present study; GenBank accession number AF061797)
   Used: all available
5. *Branchiostoma floridae* (lancelet: invertebrate chordate)
   Available: all of 18S except first ~40 nt (Stock and Whitt 1992; GenBank accession number M97571); last 50 nt of 5.8S; all of 28S except last ~54 nt (present study; GenBank accession number AF061796)
   Used: all available

Taxa for which information was already available

1. *Xenopus laevis* (frog)
   Available: complete 18S, 5.8S, 28S sequences (Ajuh, Heeney, and Maden 1991; GenBank accession number X59734)
   Used: all 18S except first ~40 nt; last 50 nt of 5.8S; all of 28S except last 54 nt
2. *Latimeria chalumnae* (coelacanth: lobe-finned bony fish)
   Available: all 18S except first ~40 nt (Stock et al. 1991; GenBank accession number 111288); no 5.8S; 28S missing first 56 and last ~300 nt (Zardoya and Meyer 1996; GenBank accession number lcu34336)
   Used: all available
3. *Oncorhynchus mykiss* (trout: ray-finned bony fish)
   Available: no 18S; all of 28S except first 56 and last ~54 nt (Zardoya and Meyer 1996; GenBank accession number omu34341; present study; GenBank accession number AF061801)
   Used: all available
4. *Acipenser brevirostrum* (sturgeon: ray-finned bony fish)
   Available: all of 18S except first ~40 nt (Stock 1992); no 5.8S; 28S missing first 56 and last ~300 nt (Zardoya and Meyer 1996; GenBank accession number abu34340)
   Used: all available
5. *Rhinobatos lentiginosus* (ray: cartilaginous fish)
   Available and used: all of 18S except first ~40 nt (Stock and Whitt 1992; GenBank accession number M97576)
6. *Styela plicata* (tunicate: invertebrate chordate)
   Available and used: all 18S except first ~40 nt (Stock and Whitt 1992; GenBank accession number M97577)
7. *Herdmania momus* (tunicate: invertebrate chordate)
   Available: complete 18S, 5.8S, 26S sequences (Degnan et al. 1990; GenBank accession number X53538)
   Used: all of 18S except first ~40 nt; last ~50 nt of 5.8S; all of 26S except last ~51 nt

nator biochemistry. The reaction mixture for cycle sequencing consisted of: 7.3 μl of terminator dye premix (Perkin Elmer/ABI); 35–100 ng of DNA in 10 μl of water; 1.6 μl of sequencing-primer solution (2 pM/μl); 1.0 μl DMSO; and 0.6 μl *Taq* Extender PCR Additive. The cycle-sequencing conditions were: (1) predwell for 3 min at 96°C and (2) 25 cycles of: 96°C for 30 s, 50°C for 2 min, and 60°C for 4 min. The DNA product was cleaned with Centri-Sep columns (Princeton Separations, Inc.) and vacuum-dried before sequencing. Full sequences were obtained from three animals per taxon. For all the basal-chordate taxa, the sequence was determined from both strands for at least 90% of the sequence.

Fragment Assembly and Alignment

Consensus sequences were assembled from contigs using the GCG fragment assembly system (Genetics Computer Group 1994), consisting of the programs GelStart, GelEnter, GelMerge, and GelAssemble. Our newly obtained 28S and partial 5.8S sequences were imported into SeqLab (see Smith et al. 1994); we also imported corresponding 18S rDNA sequences and comparable 28S, 18S, and 5.8S rDNA sequences of other chordates from GenBank and from Stock (1992). Table 1 lists all of the sequences used, from numerous taxa.

The sequences were first divided into stems and loops according to the models of rRNA secondary structure in the electronic database at the University of Colorado (http://pundit.colorado.edu:8080/RNA/23S/23s.html and /16S/16s.html; Schnare et al. 1996; Gutell 1994). Alignments were made by eye using these models as a guide (the LSU models of *Xenopus laevis* and *Herdmania momus* and the SSU models of *X. laevis* and *Strongylocentrotus purpuratus*). The alignments are available on request from the authors and from the EMBL Nucleotide Sequence Database (alignment numbers DS35047–DS35049).

In preparation for phylogenetic analyses, we manually omitted sites that we considered to be poorly

aligned, that were absent from the lamprey or the hagfish, or that were absent from both the invertebrate outgroups, *Branchiostoma* and tunicate. The 28S data were divided into: (1) the conserved core of this molecule (2,344 aligned nucleotides) and (2) the core plus the alignable parts of the divergence domains (2,883 aligned nucleotides). Divergence domains, which are more variable and more rapidly evolving than the conserved core (Schnare et al. 1996), were originally identified in mouse (Hassouna, Michot, and Bachellerie 1984; GenBank accession number X00525) and were identified for the present taxa by location of the corresponding boundary sequences. All but the most readily alignable parts of these domains were omitted in order to minimize the confounding effects of cryptic simplicity, indels, and multiple substitutions on phylogenetic analysis (see, e.g., Kuzoff et al. 1998).

Phylogenetic Analysis

All phylogenetic analyses were performed using PAUP* (test versions 4.0d55–64, by D. L. Swofford). Gaps were treated as missing data. The 28S and 18S rDNA alignments were analyzed separately and also were combined, along with the incomplete (50 aligned nucleotides) 5.8S data. Combining these three different genes was warranted primarily because the 28S and 18S data produced nearly identical phylogenetic trees (see below). Furthermore, the incongruence-length difference test (Farris et al. 1994; Cunningham 1997) suggested that these genes are not significantly more incongruent than random partitions ($P < 0.44$) (but see Sullivan [1996] for a criticism of that test as an arbiter of data combination). For both the 28S and combined data sets, we conducted analyses using both the core regions (with all data from 28S divergence domains omitted; designated ''CORE'') and a data set that included the parts of the 28S divergence domains we judged to be aligned reliably (designated ''CORE + DIV'').

Analyses were conducted primarily under the likelihood criterion (Felsenstein 1981), following the iterative search strategy proposed by Swofford et al. (1996; see Sullivan, Markert, and Kilpatrick [1997], Sullivan and Swofford [1997], and Naylor and Brown [1998] for examples of this strategy). First, initial trees were obtained by exact searches (branch and bound) using equally weighted parsimony; the best parsimony trees consistent with each of the alternative hypotheses of basal-vertebrate relationships were then used to evaluate the relative fits of evolutionary models and to estimate model parameters.

Sixteen evolutionary models (four nucleotide-substitution matrices, each with four ways to accommodate among-site rate variation) were evaluated. Because the most general of these models (GTR + I + Γ; general time reversible substitution matrix, with some proportion of sites invariable and rates at variable sites following an eight-category discrete approximation of the gamma distribution) includes each of the 15 simpler models as special cases, we compared the relative fit of alternative models using the likelihood-ratio test and the χ^2-approximation of the null distribution (e.g., Yang, Goldman, and Friday 1995). The assumptions of these models are presented in Swofford et al. (1996) and Sullivan and Swofford (1997). Exact searches for optimal trees were then conducted under the likelihood optimality criterion with the substitution model fully defined. Strength of nodal support in the likelihood analyses was estimated using the conventional nonparametric bootstrap (Felsenstein 1985; 100 replicates, heuristic search with random input orders, and TBR branch swapping), with model parameters fixed to values estimated from the original data.

Because the 18S data and, as a result, the combined data set with both 18S and 28S sequences may violate the assumption of stationary base frequencies made by all of the likelihood models used above, we also used LogDet/paralinear distances (Lockhart et al. 1994) with 50% invariable sites to construct minimum-evolution trees. In these LogDet analyses, we assessed the strength of nodal support with the conventional bootstrap, based on 500 bootstrap replicates (heuristic searches with random input orders, and TBR branch swapping).

Conventional bootstrapping (Felsenstein 1985) was also performed under maximum-parsimony criteria with equal weighting of characters (500 bootstrap replicates, branch-and-bound searches).

Statistical Tests of Alternative Hypotheses

The two alternative resolutions of basal relationships among vertebrates that have been proposed represent a priori hypotheses and are thus imminently suited to statistical phylogenetic testing (Huelsenbeck and Rannala 1997). Of the several tests that are available, the most powerful are the Kishino-Hasegawa test (KH test; Kishino and Hasegawa 1989) and the parametric bootstrap (Huelsenbeck, Hillis, and Jones 1996). We constrained each of the data sets (18S, $28S_{CORE}$, $28S_{CORE+DIV}$, $combined_{CORE}$ and $combined_{CORE+DIV}$) to fit the two alternative hypotheses and assessed the significance of the test statistic (δ = difference in ln likelihood) in two ways: (1) The KH test uses the distribution of single-site differences in ln likelihoods between the two trees being compared as a null distribution, under the null hypothesis that these trees are equivalent. (2) For the 28S data sets, we used the parametric bootstrap approach and generated the null distributions by simulation. In these parametric tests, we used the best tree that unites lampreys and gnathostomes as the model tree and generated 100 replicate data sets by simulation using the model parameters estimated from the real data (on the model tree). Because of the possible violation of stationarity in the 18S and combined data sets (see above), we did not conduct parametric bootstrap analyses on them.

Results

General

The long 28S rDNA sequences obtained here match short partial sequences that were reported previously (*P. marinus* and *E. stouti*: Le, LeCointre, and Perasso 1993; Philippe et al. 1994), and therefore we are

Table 2
Length and Composition of 28S rDNA Genes in Various Chordates

Taxon	Length (bp)	G+C Content		
		Entire 28S	Core	Divergence Domains
1. *Homo sapiens*	5,025	0.693	0.542	0.829
2. *Mus musculus*	4,712	0.667	0.541	0.797
3. *Xenopus laevis*	4,115	0.658	0.540	0.822
4. *Hydrolagus colliei*[a]	3,856 + 54 est.[b]	0.610	0.538	0.724
5. *Squalus acanthias*[a]	3,465 + 300 est. (incomplete)[c]	(0.60)[c]	(0.53)[c]	(0.73)[c]
6. *Petromyzon marinus*[a]	3,924 + 54 est.	0.620	0.543	0.734
7. *Eptatretus stouti*[a]	5,205 + 53 est.	0.664	0.560	0.750
8. *Branchiostoma floridae*[a]	3,836 + 54 est.	0.606	0.528	0.729
9. *Herdmania momus* (26S) ...	3,566	0.573	0.514	0.689

Note.—These are the most complete sequences available from chordates. Sources of the sequences and GenBank accession numbers are given in table 1, except for *Homo* (Gonzalez et al. 1985; M11167) and *Mus* (Hassouna, Michot, and Bachellerie 1984; J01871, X00525).

[a] Data from the present study.

[b] Estimated: the last 52–54 nt of the 3′ end were not sequenced in the present study. However, 52–54 seems to be a conserved number across chordates based on *Homo, Mus,* and *Herdmania.*

[c] Parentheses indicate these *Squalus* values, less complete than in the other species, are estimates.

confident that they are accurate. Some basic information about the new sequences is presented in tables 2–4, along with comparable rDNA data from previously studied chordates. As seen in table 2, the 28S rDNA genes of *Hydrolagus, Squalus, Petromyzon,* and *Branchiostoma* are similar in length (around 3,825–3,980 nt), but that of *Eptatretus* is larger (>5,200 nt). In fact, this hagfish has the largest 28S rRNA gene of any known organism (Schnare et al. 1996).

As is typical for the 28S rRNA genes of animals, the divergence domains of our taxa are more CG-rich (table 2) and more variable in size (table 3) than the conserved core region. The core of the hagfish gene is unusually CG-rich, and this gene's large size is due to the exceptional lengths of divergence domains D3, D5, D8, and D10 (plus D7b and D2). Although the gene of the lancelet *Branchiostoma* is not particularly large, it has an unusually large D7a domain.

For all of the data sets except the 18S sequences, the best fit under likelihood was provided by the most complex evolutionary model, GTR + I + Γ. For the 18S sequences alone, the GTR + Γ model was indistinguishable from the more parameter-rich GTR + I + Γ. Estimates of the likelihood-model parameters are presented in table 5.

Phylogenetic Analyses

All 28S, 18S, and combined-data trees estimated by all methods supported a monophyletic cyclostome clade containing lampreys and hagfishes, although different methods and data sets provided differing levels of support. The optimal trees are shown in figures 1–3. In all the data sets, the hagfish sequence was by far the most divergent, as indicated by the long hagfish branches in the phylograms.

In the analyses of the 28S data (fig. 1), there is rather strong bootstrap support for lamprey-hagfish monophyly from both the maximum-likelihood (84% for CORE and 75% for CORE+DIV) and LogDet analyses (96% and 93%), but less support from maximum parsimony (74% and 55%). According to the χ^2 test of homogeneity of base frequencies, these 28S sequences show no indication of nonstationarity of base frequencies among taxa: $P = 0.97$ and $P = 0.57$ for CORE and CORE+DIV, respectively (table 5).

For the 18S data (fig. 2), two different groups of taxa were analyzed. One included the taxa used in the 28S analysis (fig. 2*A*). The other (fig. 2*B*) contained the taxa used by Stock and Whitt (1992), whose sequences were reanalyzed using our new alignments based on secondary-structure models and using the new models of nucleotide substitution that have been developed subsequent to their study. For both groups of taxa, the results were the same: the lamprey-hagfish clade received strong bootstrap support from maximum-likelihood analyses (95% for the present taxa; 97% for the taxa of Stock and Whitt 1992), but moderate-to-weak support from LogDet/paralinear (62% and 56%) and maximum parsimony (60% and 62%). However, there was some indication of nonstationary base frequencies in the 18S sequences: $P = 0.15$ and $P = 0.09$, respectively, for our taxa and Stock and Whitt's (1992) taxa (table 5). Although a P value of 0.15 is usually taken to indicate no significant effect, the χ^2-homogeneity test that we used ignores correlation due to phylogeny, and the degrees of freedom are thus inflated. Because of this, the P values obtained with this test should be interpreted with reservations.

Results from the combined data set (28S, 18S, partial 5.8S) are shown in figure 3. Here, the lamprey-hagfish group received strong bootstrap support from maximum-likelihood (96% for CORE, 98% for CORE+DIV) and LogDet analyses (90% and 91%), but weaker support from maximum parsimony (68% and 53%). Nonstationarity is indicated for the combined data set and is especially evident when the divergence domains are included: $P < 0.027$ for CORE and $P < 0.003$

Table 3
Divergence Domains in the 28S rDNA Genes of Various Chordates: Position Ranges and Lengths

Taxon	D1	D2	D3	D4	D5	D6	D7a	D7b	D8	D9	D10	D12
Homo sapiens	122–268 147	427–1291 865	1333–1482 150	1674–1703 30	1784–1812 29	2057–2255 199	2431–2489 59	2526–2569 44	2874–3574 701	3944–4006 63	4047–4129 83	4689–4932 244
Xenopus laevis	122–276 155	435–964 530	1006–1179 174	1371–1381 11	1462–1491 30	1737–1779 43	1954–2012 59	2049–2128 80	2434–2770 336	3140–3166 27	3207–3294 88	3855–4022 168
Hydrolagus colliei[a]	122–275 154	434–956 523	998–1142 145	1334–1342 9	1423–1452 30	1696–1736 41	1911–1969 59	2006–2049 44	2355–2593 239	2963–2970 8	3001–3094 84	3654–3816 162
Squalus acanthias[a]	122–275 154	434–866 433	909–1049 141	1241–1252 12	1333–1362 30	1606–1647 42	1822–1880 59	1917–1960 44	2265–2449 185	2819–2826 8	2867–2950 84	incomplete ?
Petromyzon marinus[a]	122–277 156	436–956 521	998–1188 146	1380–1391 12	1471–1501 30	1745–1786 42	1961–2019 59	2055–2107 53	2412–2659 248	2930–2956 27	2997–3102 106	3662–3829 167
Eptatretus stouti[a]	122–278 157	437–1241 805	1283–1640 358	1833–1845 13	1926–1990 65	2236–2309 73	2484–2545 62	2581–2661 81	2964–3769 806	4150–4207 68	4248–4399 152	4962–5165 204
Branchiostoma floridae[a]	121–270 149	429–859 431	901–1044 144	1235–1243 8	1324–1353 30	1597–1635 39	1810–1989 180	2024–2077 54	2381–2554 174	2924–2933 10	2974–3080 107	3641–3799 158
Herdmania momus	118–268 151	426–797 372	838–972 142	1168–1175 8	1255–1283 29	1525–1575 51	1755–1811 57	1846–1887 42	2181–2272 92	2638–2645 8	2683–2774 92	3335–3474 140

NOTE.—Coordinates were deduced from positions of the divergence domains in mouse (Hassouna, Michot, and Bachellerie 1984; J01871, X00525), after aligning these sequences with the mouse sequence.
[a] Data from the present study.

for CORE+DIV (table 5). This nonstationarity probably was due to a high frequency of A and T in the tunicate versus a high frquency of C and G in the hagfish rRNA genes (see the right half of table 4), because eliminating these two taxa restores stationarity (not shown). For the lamprey, the C and G content is also larger than average (table 4).

Statistical Tests of Alternative Hypotheses

Both the KH test and parametric bootstrap were used to compare likelihood scores of the best lamprey-gnathostome tree versus the optimal likelihood (lamprey-hagfish) tree (table 5). The KH test did not significantly reject the lamprey-gnathostome hypothesis in favor of the lamprey-hagfish hypothesis for the 28S, 18S, or combined$_{CORE+DIV}$ data (P values ranged from 0.63 to 0.12), although the 18S values approached significance at the 0.1 level. For the combined$_{CORE}$ data, the KH test supports cyclostome monophyly at the 0.1 level ($P = 0.099$). Additionally, the parametric bootstrap tests using both the 28S data sets strongly reject the lamprey-gnathostome sister-group hypothesis in favor of cyclostome monophyly (fig. 4; $P < 0.01$ and $P = 0.03$ for CORE and CORE+DIV data, respectively).

Discussion

Overall, the present results support the conclusion of Stock and Whitt (1992), which was based on 18S rDNA sequences alone, that lampreys and hagfishes are a natural group. In fact, our own 18S-based results (using statistically more defensible models) closely matched theirs (fig. 2 and table 5). The addition of 28S sequences, along with recent advances in phylogenetic and statistical techniques, allow a more rigorous evaluation of these relationships than was previously possible. While not all methods used in this study provide equally strong support for cyclostome monophyly, the most rigorous methods provide the strongest support.

Maximum parsimony only weakly supported cyclostome monophyly, but there are a priori reasons for deemphasizing our parsimony-based results. Because it does not correct for multiple substitutions, the parsimony method underestimates branch lengths; this is especially true for rapidly evolving lineages and when there is a great deal of among-site rate variation (e.g., Tateno, Takezaki, and Nei 1994), as seen here. This renders parsimony particularly susceptible to systematic error associated with ''long-branch attraction'' and related phenomena (Felsenstein 1978; Gaut and Lewis 1995). Such error could readily affect the quickly evolving and highly divergent hagfish, which is the main focus of this study. The weakness of support for cyclostome monophyly in the parsimony analyses, with bootstrap values of only 53%–74%, may result from the spurious attraction of the hagfish to the outgroup in some of the bootstrap replicates (not shown). The application of appropriate weighting in parsimony could improve phylogenetic accuracy over that obtained here (e.g., Chippendale and Wiens 1994); however, the efficacy of such a weighting approach is not easily evaluated because dif-

Table 4
Nucleotide Frequencies in the rRNA Genes of Various Chordates

	28S GENE				COMBINED$_{CORE}$[a]			
	A	C	G	T	A	C	G	T
Xenopus laevis	0.184	0.302	0.357	0.158	0.245	0.242	0.300	0.213
Latimeria chalumnae	0.205	0.268	0.341	0.185	0.248	0.230	0.300	0.222
Oncorhynchus mykiss	0.202	0.260	0.336	0.203	—	—	—	—
Acipenser brevirostrum	0.208	0.266	0.332	0.193	0.251	0.232	0.295	0.222
Hydrolagus colliei[b]	0.204	0.273	0.336	0.186	0.250	0.236	0.297	0.218
Squalus acanthias[b]	0.215	0.261	0.338	0.191	0.254	0.232	0.294	0.219
Petromyzon marinus[b]	0.191	0.275	0.345	0.189	0.241	0.238	0.305	0.217
Eptatretus stouti[b]	0.153	0.303	0.361	0.182	0.230	0.255	0.317	0.198
Branchiostoma floridae[b]	0.203	0.275	0.332	0.191	0.247	0.238	0.297	0.218
Herdmania momus	0.225	0.255	0.318	0.202	0.257	0.220	0.291	0.232

[a] Combined$_{CORE}$ consists of the 18S gene, the conserved core of the 28S gene, and 50 bases of the 5.8S rRNA gene; for more precise information on the parts of the genes included, see tables 1 and 3.

[b] Data from the present study.

ferent parsimony weighting schemes cannot be objectively compared (Swofford et al. 1996).

We therefore focused on maximum-likelihood methods because they can account for unobserved substitutions, are consistent across a broader range of conditions than are parsimony methods (e.g., Huelsenbeck 1995), and allow for the comparison of the relative fits of alternative models of nucleotide substitution because the likelihood score is an objective indicator of goodness of fit between model and data that is comparable across models (e.g., Sullivan and Swofford 1997). Thus, the (nonparametric) bootstrap values generated under likelihood models should be better estimates of reliability than those under parsimony (DeBry and Abele 1995; Waddell 1995; Frati et al. 1997; Sullivan, Markert, and Kilpatrick 1997). At 75%–98% (figs. 1–3), these likelihood bootstrap values provide good support for cyclostome monophyly.

The K-H test is often used to examine the significance of an observed difference in likelihood scores between two alternative phylogenetic hypotheses (here, lamprey-hagfish vs. lamprey-gnathostome). With this test, the combined$_{CORE}$ sequences support the lamprey-hagfish relationship at the 0.1 level ($P = 0.099$). Although it was the only data set providing such support, we consider this combined$_{CORE}$ to be the most appropriate set analyzed. That is, it is more complete than the 28S or 18S set alone, while avoiding the potentially confounding effects of faster evolution in the divergence domains of the combined$_{CORE+DIV}$ sequences.

Table 5
Some Results of the Phylogenetic Analysis, Based on 28S, 18S, and Combined (28S, 18S, partial 5.8S) rDNA Genes

	28S Core	28S Core + Divergence Domains	18S	18S (taxa of Stock and Whitt 1992)	Combined: Core	Combined: Core + Divergence Domains
Number of aligned sites	2,344	2,883	1,762	1,762	4,151	4,694
Likelihood-model parameters						
r-matrix						
(AC)	1.0786	1.1887	0.9846	1.1721	0.9907	1.0491
(AG)	2.6398	2.8206	1.9817	2.1281	2.1620	2.2720
(AT)	0.6296	0.7419	0.8019	0.7841	0.7552	0.8160
(CG)	1.1362	1.2810	0.8906	1.0	1.0030	1.1126
(CT)	6.2856	6.9173	4.1655	4.4527	4.9454	5.2175
$p_{invariable}$	0.525	0.442	—	—	0.414	0.374
Alpha	0.776	0.697	0.3389	0.311	0.708	0.696
ln likelihood values						
Best lamprey-hagfish (L-H) tree	−6,894.75	−9,693.28	−6,388.27	−5,761.32	−13,374.19	−15,986.06
Best lamprey-gnathostome (L-G) tree	−6,899.96	−9,695.87	−6,394.49	−5,767.89	−13,387.10	−15,997.58
KH test values[a]	$\delta = 5.21$	$\delta = 2.59$	$\delta = 6.23$	$\delta = 6.57$	$\delta = 12.91$	$\delta = 11.51$
	$\sigma = 5.99$	$\sigma = 5.29$	$\sigma = 4.82$	$\sigma = 4.21$	$\sigma = 7.81$	$\sigma = 7.53$
	$P < 0.38$	$P < 0.63$	$P < 0.20$	$P < 0.12$	$P < 0.099$	$P < 0.13$
Parametric bootstrapping						
(best L-H vs. best L-G tree)	$P < 0.01$	$P < 0.03$	—	—	—	—
Homogeneity of nucleotide frequencies[b]	$P < 0.97$	$P < 0.57$	$P < 0.15$	$P < 0.09$	$P < 0.027$	$P < 0.003$

NOTE.—For the specific taxa and gene regions used in each analysis, see figures 1–4 and table 1. Bootstrap values are given in figures 1–4.

[a] δ = observed difference in ln likelihood, σ = standard deviation of distribution of single-site likelihood differences.

[b] These P values were calculated with the χ^2 test of homogeneity of base frequencies across taxa. This test ignores correlation due to phylogeny, and thus the values are likely biased.

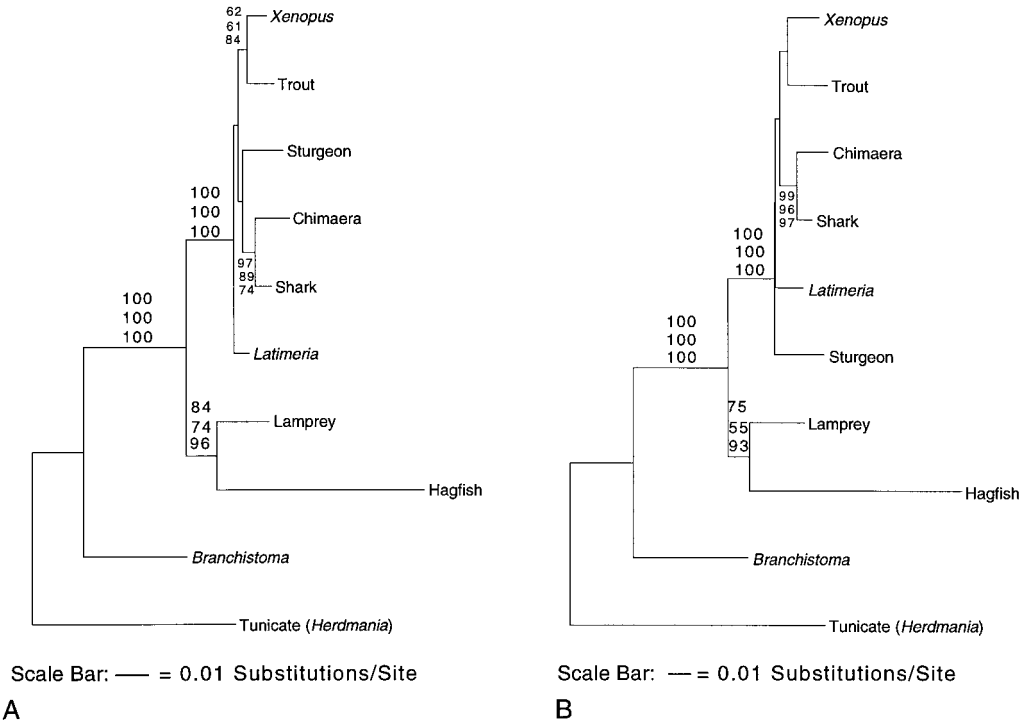

FIG. 1.—Phylogenetic trees estimated from the 28S rDNA sequences under the maximum-likelihood optimality criterion, based on (*A*) core regions (lnL = −6,894.75) and (*B*) core plus divergence domains (lnL = −9,693.28). The percentages of bootstrap replicates supporting the clades are indicated at the branch points, based on (from top to bottom) maximum likelihood, equally-weighted parsimony, and minimum evolution using LogDet/paralinear distances. Bootstrap values <50% are not shown.

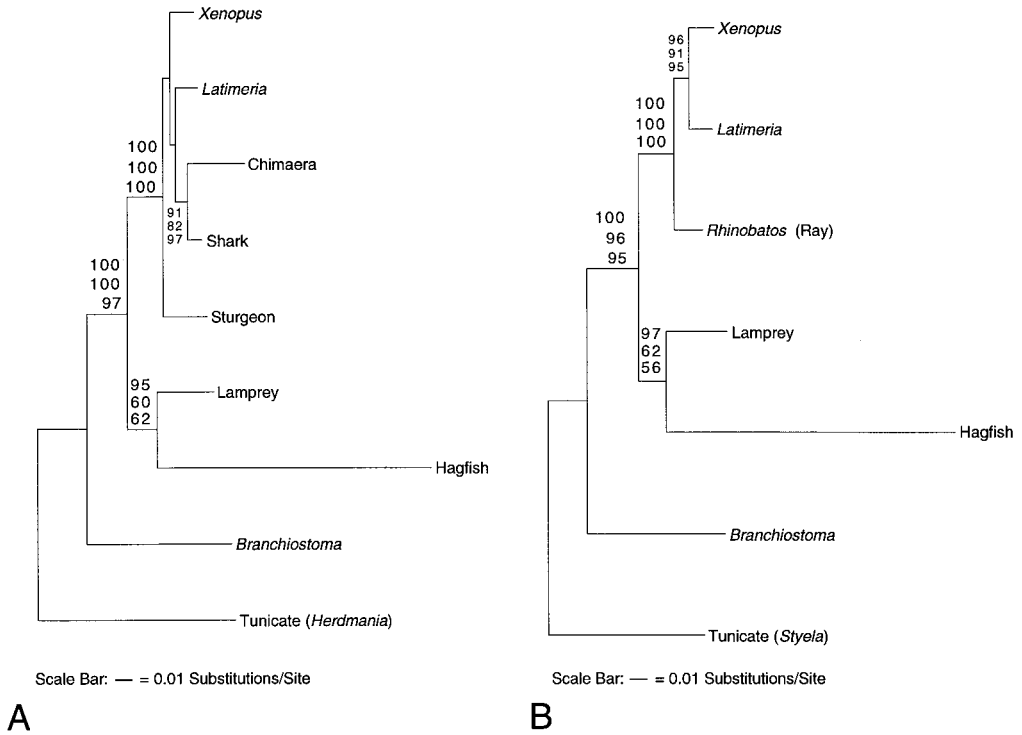

FIG. 2.—Phylogenetic trees estimated from the 18S rDNA sequences under the maximum-likelihood optimality criterion, based on (*A*) the taxa of the present study (lnL = −6,388.27) and (*B*) the somewhat-different taxa used by Stock and Whitt (1992; lnL = −5,761.32; as in part *A*, the hagfish and lamprey sequences are *E. stouti* and *P. marinus*). Percentages of bootstrap replicates supporting the clades are indicated at the branch points, as in figure 1 (likelihood, parsimony, LogDet). Bootstrap values <50% are not shown.

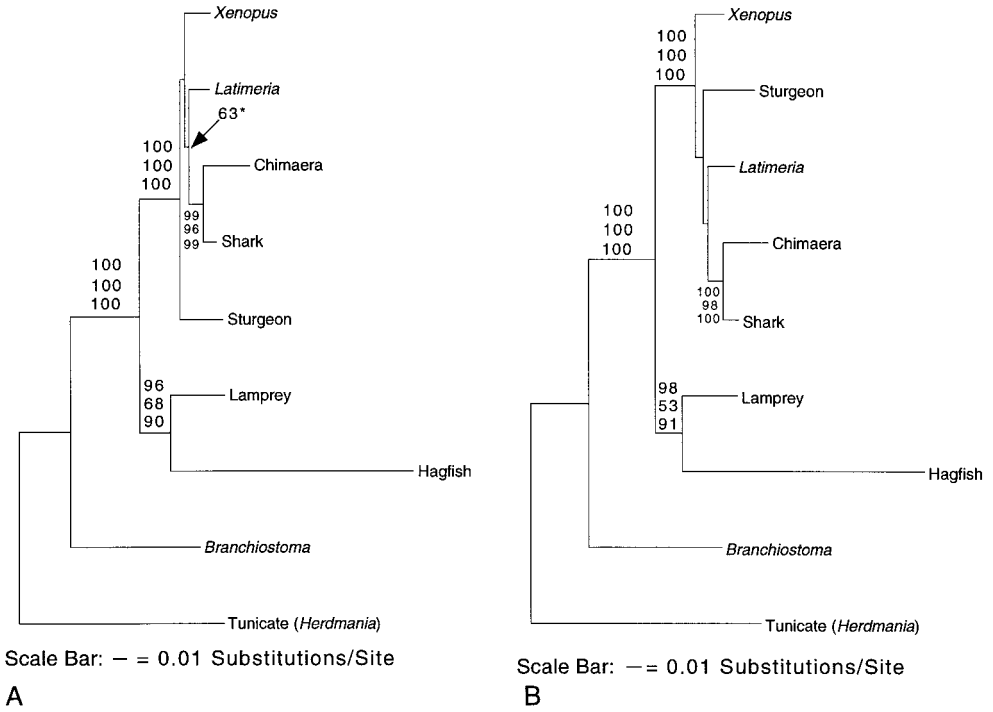

FIG. 3.—Phylogenetic trees estimated from the combined-gene sequences (28S, 18S, partial 5.8S) under the maximum-likelihood optimality criterion, based on (*A*) combined$_{CORE}$, in which the 28S genes contain only the core region (lnL = −13,374.19), and (*B*) combined$_{CORE+DIV}$, in which the 28S genes also contain the divergence domains (lnL = −15,986.06). Percentages of bootstrap replicates supporting the clades are indicated at the branch points, as in figure 1 (likelihood, parsimony, LogDet). Bootstrap values <50% are not shown. In tree *A*, the ''63*'' at the arrow is from parsimony, which is the only method that provided >50% bootstrap support for that node.

The results of the KH test might be questioned on the grounds that the combined-gene sequences indicated nonstationarity of nucleotide substitution (table 5), which could perhaps lead to a spurious grouping of lamprey and hagfish. To determine if nonstationarity was distorting these test values, we performed the LogDet analyses, which are robust to changing base compositions (Lockhart et al. 1994). The LogDet analyses strongly supported cyclostome monophyly, with bootstrap values around 90% (fig. 3), in agreement with the KH test. Thus, nonstationarity does not appear to be unduly influencing the signal for cyclostome monophyly in the combined-gene sequences.

If the results of the KH test are taken at face value, then most of the phylogenetic signal for cyclostome monophyly is in the 18S gene. That is, near-significant KH support was obtained from the 18S sequences alone, and adding the LSU sequences to our 18S sequences to form the combined gene merely improved the *P* value from 0.2 to 0.099 (table 5). However, the high bootstrap values shown in figure 1, as well as the parametric bootstrap (fig. 4), indicate that the 28S sequences also support cyclostome monophyly and strongly reject the lamprey-gnathostome clade.

The parametric bootstrap procedure generates replicate data sets by simulation using model parameters derived from the original data. This is a very powerful method for evaluating competing a priori phylogenetic hypotheses (Huelsenbeck, Hillis, and Jones 1996; Huelsenbeck and Ranalla 1997). It can be used to assess the probability of observing a maximum-likelihood (or maximum-parsimony) tree that is as much better than the model tree predicted by the hypothesis under examination as is the observed tree if the hypothesized tree is the true tree. In the case of the 28S$_{CORE}$ (fig. 4*A*), if the lamprey-gnathostome tree were actually the true tree, we would expect to see an optimal tree 5.232 ln likelihood units better than the true tree far less than 1% of the time ($P < 0.01$) just due to stochastic error. Thus, we can reject the lamprey-gnathostome hypothesis statistically.

Some Challenges to Cyclostome Monophyly

Stock and Whitt's (1992) previous support for cyclostome monophyly has been criticized on the grounds that when these authors used tunicates as their only outgroup, parsimony bootstrap analysis did not unite lampreys with hagfishes, but weakly supported (62%) the lamprey-gnathostome clade (Forey and Janvier 1993; Janvier 1998; Rasmussen, Janke, and Arnason 1998). This criticism is simply devoid of merit. With likelihood analysis, which has advantages over parsimony (as discussed above), Stock and Whitt (1992) actually obtained strong bootstrap support (81%) for a cyclostome clade with the tunicate outgroup. Furthermore, because the issue of the basal resolution of these taxa requires reliable rooting, use of two outgroup taxa is vastly preferable to use of a single outgroup. All of the analyses conducted by Stock and Whitt (1992) that included both the tunicate and lancelet outgroups supported the monophyly of living cyclostomes, and those results were upheld by our reanalyses.

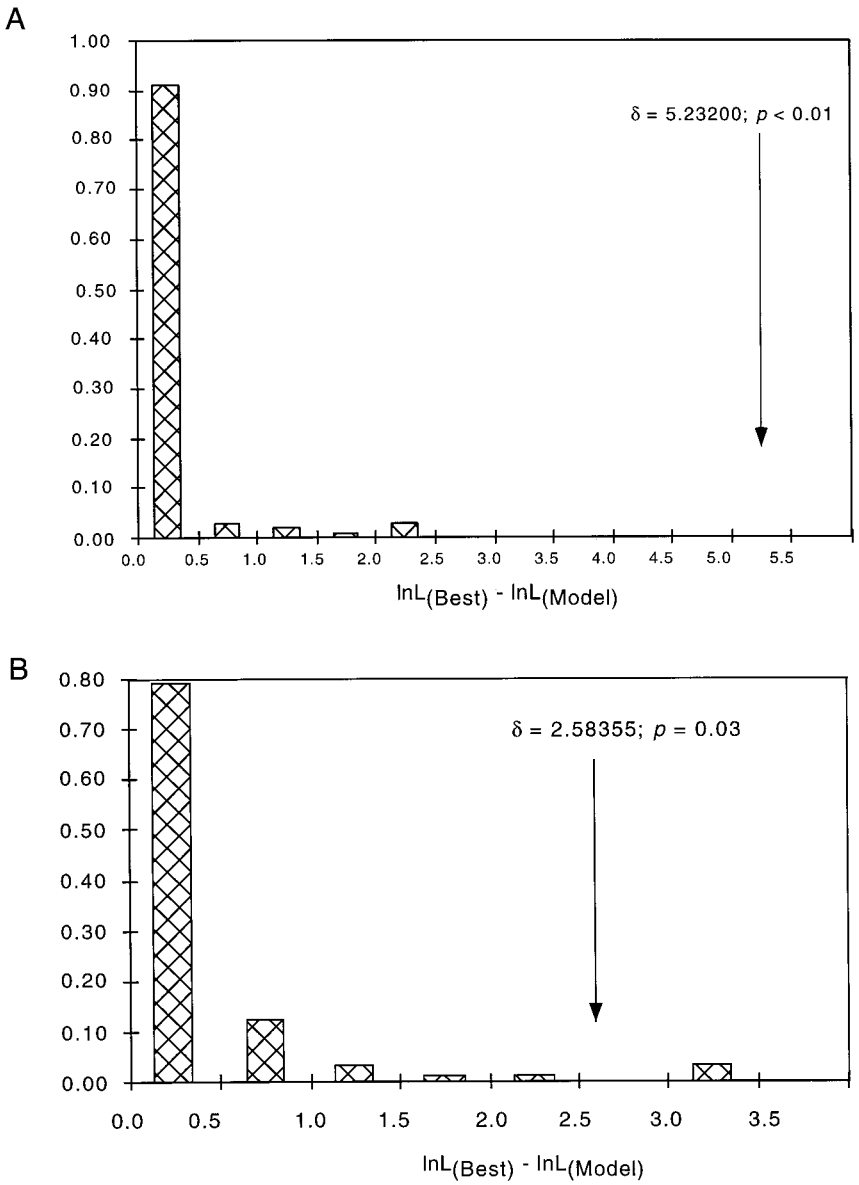

FIG. 4.—Results of parametric bootstrap tests of the null hypothesis that lampreys group with gnathostomes, based on only the 28S sequences. *A*, The test using the core region. *B*, The test using core plus divergence domains. In each test, the optimal lamprey/gnathostome tree was chosen as a model, and then 100 replicate data sets were simulated using optimal likelihood parameters that had been calculated from the original data. Each graph shows the frequency distribution of difference in likelihood scores between the best tree and the model (lamprey/gnathostome) tree. Almost none of the simulated differences in lnL are as large as the observed differences between the best lamprey/hagfish and the model lamprey/gnathostome trees, as indicated by the δ values at right. Thus, the null hypothesis can be rejected; that is, lampreys do not group with gnathostomes.

Recently, the combined sequences of 12 protein-coding mitochondrial genes were compared for hagfishes, lampreys, and various gnathostomes by Rasmussen, Janke, and Arnason (1998), who claimed that their analyses supported a lamprey-gnathostome clade, based on parsimony bootstrap, neighbor-joining bootstrap, and quartet puzzling (an approximate maximum-likelihood method; Strimmer and von Haeseler 1996). However, in addition to evidence that accurate phylogenetic estimates may not be obtainable from the mitochondrial genes of basal chordates (Naylor and Brown 1998; Zardoya et al. 1998), the conclusions of Rasmussen, Janke, and Arnason (1998) have at least three major flaws. First, they did not use nonvertebrate chordates (tunicates and lancelets) as outgroups, but instead used echinoderms. Again, because accurate resolution of relationships among the basal-vertebrate lineages is entirely dependent on reliable rooting, using the most closely related taxa that are undeniable outgroups is far superior to using more distant outgroups; this is axiomatic. Second, these authors claimed that because the same trees were produced by all three methods, their results were robust to any violations of the assumptions of the analyses. However, none of the methods they used accounted for among-site rate variation explicitly, and failure to do this has been shown (Sullivan and Swofford 1997,

p. 84) to induce a great deal of systematic error in similar analyses of mtDNA sequences (D'Erchia et al. 1996); such error exacerbates long-branch attraction artifacts, which could have spuriously united lampreys with gnathostomes by pulling the divergent hagfish sequence toward the echinoderm outgroup (see fig. 1 in Rasmussen, Janke, and Arnason 1998). Third, these authors reported results of statistical (KH) tests between the cyclostome and lamprey/gnathostome hypotheses that could not reject the monophyly of cyclostomes (their table 1). They then cited quartet-puzzling reliability values that supposedly provided moderate support (64%) for a lamprey-gnathostome group. However, quartet-puzzling reliability values have been shown to be biased upward (Cao, Adachi, and Hasegawa 1998). Thus, even based on their own results, the mtDNA data of Rasmussen, Janke, and Arnason (1998) failed to reject the monophyly of living cyclostomes.

As another potential challenge to cyclostome monophyly, none of the published studies based on 18S rDNA sequences has found support for the monophyly of the chordates (Turbeville, Schulz, and Raff 1994; Wada and Satoh 1994), a fundamental taxon that is overwhelmingly supported by other data. This shortcoming could be used to argue that all rDNA genes are useless for estimating basal-chordate relationships, including those of lampreys and hagfishes. The problematic finding of the previous 18S studies was that tunicates grouped with the nonchordate hemichordates rather than with lancelets and vertebrates based on neighbor-joining and parsimony analyses. In contrast, however, when we realigned and reanalyzed these 18S sequences with the likelihood techniques used in the present study, tunicates grouped with the vertebrates, producing a traditional monophyletic chordata (unpublished data).

Some Challenges to Our Analyses

The major findings of the present study could be challenged in several ways. The hagfish branch is very long in all the molecular data sets, as was first pointed out by Stock and Whitt (1992). Hagfish *Eptatretus* has many autapomorphies in its rRNA gene sequences, just as hagfishes have many autapomorphies in their morphologies, their vasotocins, and their mitochondrial gene sequences (see the long branches for hagfish in the phylograms presented by Suzuki et al. [1995] and Rasmussen, Janke, and Arnason [1998]). This implies that an evolutionary rate acceleration occurred sometime in hagfish history, perhaps in association with an ecological shift into their unusual burrowing existence in the cold, deep sea (for more on hagfish ecology, see Hardisty 1979). Rapidly evolving taxa can confound phylogenetic analyses, not only through long-branch attraction and rogue-taxon effects, but also because their long branches may indicate a deviation from stationarity of process and, thus, a violation of the assumptions of the likelihood models commonly used. However, if long-branch attraction artifacts were actually present in our likelihood analyses, we would have expected hagfish to be attracted to the outgroups (the next longest branches) rather than to lamprey. Stock and Whitt (1992) also pointed this out in their analyses of the 18S data.

Second, our study used 28S rDNA sequences from just one hagfish and one lamprey species. Additional taxa of cyclostomes may have enhanced our study and should be sequenced in the future. However, neither living hagfishes nor lampreys are terribly diverse; both contain only a single order. The core of the 28S gene is so conserved that significant interfamily differences among such sequences are not expected. Furthermore, for 18S rDNA sequences, Stock and Whitt (1992) found almost no differences between two different families of hagfishes. This suggests that the use of additional cyclostome species would not effectively break up these long branches.

Third, in our rDNA-based trees, the interrelationships of the major groups of jawed bony fishes and tetrapod vertebrates (Osteichthyes) are neither consistent across analyses nor consistent with accepted relationships. For example, the frog *Xenopus* never groups with the coelocanth *Latimeria* (figs. 1–3), although such an association is expected (e.g., Nelson 1994). Such inconsistency could be because, in emphasizing conserved regions, we eliminated, from the alignments, faster-evolving sites that would resolve osteichthyan relationships. In no analyses do any relationships among the osteichthyan taxa (expected or otherwise) receive strong support (figs. 1–3). These considerations suggest that the osteichthyan relationships may converge toward the accepted scheme with increased data.

Conclusions

The goal of this study was to add both more molecular data and analytical rigor to the phylogenetic study of basal relationships of vertebrates. Our results, based on almost the entire set of ribosomal DNA genes (about 4,700 aligned sites), uphold Stock and Whitt's (1992) conclusion that lampreys and hagfishes form a natural group. Likelihood analyses based on the best fitting of the available evolutionary models all support cyclostome monophyly (with bootstrap values of 75%–98%), as do LogDet analyses. Among the statistical tests, cyclostome monophyly received some support from the KH test and especially strong support from the parametric bootstrap based on 28S rDNA.

While our results support the older view of the monophyly of living hagfishes and lampreys, this does not mean that all the fossil jawless vertebrates also belong in this group, as was traditionally believed (Moy-Thomas and Miles 1971). In fact, two of these extinct jawless groups, thelodonts and osteostracans, appear to be related to gnathostomes based on the presence of denticles in the pharynx and a stomach in the thelodonts, and characters of the bone tissue, tail, occiput, eye, and ear in the osteostracans (Janvier 1996). Lampreys resemble a more-basal group of jawless fossils, the anaspids, so our results suggest that hagfishes are sister to, or nested within, the anaspid-lamprey clade (see fig. 1*A* in Mallatt 1996).

This study has increased the number of nearly complete 28S rDNA sequences available in chordates from 13 (Schnare et al. 1996; Zardoya and Meyer 1996) to 18 (+38%). There may now be enough raw sequences, from a wide enough variety of chordates, to allow better reconstruction of the secondary structure of the LSU rRNA molecule, especially for the more-basal taxa and in the poorly understood divergence domains. Because the 28S gene of the hagfish *Eptatretus* is the largest ever recorded, its unique features should be explored further.

Acknowledgments

Thanks are extended to the following people for providing animals: Chris Braun, Jeff Christianson, David Duggins, Kim Fredericks, Linda and Nicholas Holland, and R. Glenn Northcutt. Others kindly provided advice and help: Karen Adams, Cory Bergey, Joe Brunelli, Susan Johns, Robert Kuzoff, Aparna Palmer, Derek Pouchnik, Dana Rogoff, Christopher Shirkey, Doug Soltis, Steve Thompson, Gary Thorgaard, and William Young. We thank David L. Swofford for permission to publish the results of analyses conducted with test versions of PAUP*. Many of the calculations were performed using the facilities of the VADMS Center, a regional biocomputing resource at Washington State University supported by WSU's Division of Sciences and the Biochemistry/Biophysics Department.

LITERATURE CITED

AJUH, P. M., P. A. HEENEY, and B. E. H. MADEN. 1991. *Xenopus borealis* and *Xenopus laevis* 28S ribosomal DNA and the complete 40S ribosomal precursor RNA coding units of both species. Proc. R. Soc. Lond. B Biol. Sci. **245**:65–71.

BERNARDI, G., P. SORDINO, and D. A. POWERS. 1992. Nucleotide sequence of the 18S rRNA gene from two teleosts and two sharks and their molecular phylogeny. Mol. Mar. Biol. Biotechnol. **1**:187–194.

BRAUN, C. B. 1996. The sensory biology of living jawless fishes: a phylogenetic assessment. Brain Behav. Evol. **48**: 262–276.

CAO, Y., J. ADACHI, and M. HASEGAWA. 1998. Comment on the quartet puzzling method for finding maximum-likelihood tree topologies. Mol. Biol. Evol. **15**:87–89.

CHIPPENDALE, P. T., and J. J. WIENS. 1994. Weighting, partitioning and combining characters in phylogenetic analysis. Syst. Biol. **43**:278–287.

CUNNINGHAM, C. W. 1997. Can three incongruence tests predict when data should be combined? Mol. Biol. Evol. **14**: 733–740.

DEBRY, R. W., and L. G. ABELE. 1995. The relationship between parsimony and maximum-likelihood analyses: tree scores and confidence estimates for three real data sets. Mol. Biol. Evol. **12**:291–297.

DEGNAN, B. M., J. YAN, C. J. HAWKINS, and M. F. LAVIN. 1991. rRNA genes from the lower chordate *Herdmania momus*: structural similarity with higher eukaryotes. Nucleic Acids Res. **18**:7063–7070.

D'ERCHIA, A. M., C. GISSI, G. PESOLE, C. SACCONE, and U. ARNASON. 1996. The guinea-pig is not a rodent. Nature **381**:597–600.

FARRIS, J. S., M. KALLERSJO, A. G. KLUGE, and C. BULT. 1995. Testing significance of incongruence. Cladistics **10**:315–319.

FELSENSTEIN, J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. Syst. Zool. **27**:401–410.

———. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17**:368–376.

———. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**:783–791.

FOREY, P., and P. JANVIER. 1993. Agnathans and the origin of jawed vertebrates. Nature **361**:129–134.

FRATI, F., C. SIMON, J. SULLIVAN, and D. L. SWOFFORD. 1997. Evolution of the mitochondrial COII gene in Collembola. J. Mol. Evol. **44**:145–158.

GAUT, B. S., and P. O. LEWIS. 1995. Success of maximum likelihood phylogeny inference in the four-taxon case. Mol. Biol. Evol. **12**:152–162.

GENETICS COMPUTER GROUP. 1994. Program manual for the GCG package, version 9.1. Madison, Wis.

GONZALEZ, I. L., J. L. GORSKI, T. J. CAMPEN, D. J. DORNEY, J. M. ERICKSON, J. E. SYLVESTER, and R. D. SCHMICKEL. 1985. Variation among human 28S ribosomal RNA genes. Proc. Natl. Acad. Sci. USA **82**:7666–7670.

GU, X., Y.-X FU, and W.-H. LI. 1995. Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. Mol. Biol. Evol. **12**:546–557.

GUTELL, R. R. 1994. Collection of small subunit (16S- and 16S-like) ribosomal RNA structures. Nucleic Acids Res. **22**: 3502–3507.

HARDISTY, M. W. 1979. Biology of the cyclostomes. Chapman and Hall, London.

———. 1982. Lampreys and hagfishes: analysis of cyclostome relationships. Pp. 166–260 *in* M. W. HARDISTY and I. C. POTTER, eds. The biology of lampreys. Vol. 4B. Academic Press, London.

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

HILLIS, D. M., and M. T. DIXON. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. Q. Rev. Biol. **66**:411–453.

HUELSENBECK, J. P. 1995. Performance of phylogenetic methods in simulation. Syst. Biol. **44**:17–48.

HUELSENBECK, J. P., D. M. HILLIS, and R. JONES. 1996. Parametric bootstrapping in molecular phylogenetics: applications and performance. Pp. 19–45 *in* J. D. FERRARIS and S. R. PALUMBI, eds. Molecular zoology. Wiley and Sons, New York.

HUELSENBECK, J. P., and B. RANNALA. 1997. Phylogenetic methods come of age: testing hypotheses in an evolutionary context. Science **276**:227–232.

JANVIER, P. 1981. The phylogeny of the craniata, with particular reference to the significance of fossil ''agnathans''. J. Vertebr. Paleontol. **1**:121–159.

———. 1996. Early vertebrates. Oxford University Press, New York.

———. 1998. A cold look at odd vertebrate phylogenies. J. Mol. Evol. **46**:375–377.

KISHINO, H., and M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order of Hominoidea. J. Mol. Evol. **29**:170–179.

KUZOFF, R. K., J. A. SWEERE, D. E. SOLTIS, P. S. SOLTIS, and E. A. ZIMMER. 1998. The phylogenetic potential of entire 26S rDNA sequences in plants. Mol. Biol. Evol. **15**:251–263.

LANFRANCHI, G., A. PALLAVICINI, P. LAVEDER, and G. VALLE. 1994. Ancestral hemoglobin switching in lampreys. Dev. Biol. **164**:402–408.

LE, H. L. V., G. LECOINTRE, and R. PERASSO. 1993. A 28S rRNA-based phylogeny of the gnathostomes: first steps in the analysis of conflict and congruence with morphologically based cladograms. Mol. Phylogenet. Evol. **2**:31–51.

LOCKHART, P. J., M. A. STEELE, M. D. HENDY, and D. PENNY. 1994. Recovering evolutionary distances under a more realistic model of sequence evolution. Mol. Biol. Evol. **11**: 605–612.

MALLATT, J. 1996. Ventilation and the origin of jawed vertebrates: a new mouth. Zool. J. Linn. Soc. **117**:329–404.

———. 1997*a*. Crossing a major morphological boundary: the origin of jaws in vertebrates. Zoology **100**:128–140.

———. 1997*b*. Hagfish do not resemble ancestral vertebrates. J. Morphol. **232**:293.

MOY-THOMAS, J. A., and R. S. MILES. 1971. Palaeozoic fishes. 2nd edition. Saunders, Philadelphia.

NAYLOR, G. J. P., and W. B. BROWN. 1998. Amphioxus mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences. Syst. Biol. **47**: 61–67.

NELSON, J. S. 1994. Fishes of the world. 3rd edition. Wiley, New York.

PHILIPPE, H., U. SORHANNUS, A. BAROIN, R. PERASSO, F. GASSE, and A. ADOUTTE. 1994. Comparison of molecular and paleontological data in diatoms suggests a major gap in the fossil record. J. Evol. Biol. **7**:247–265.

RASMUSSEN, A.-S., A. JANKE, and U. ARNASON. 1998. The mitochondrial DNA molecule of the hagfish (*Myxine glutinosa*) and vertebrate phylogeny. J. Mol. Evol. **46**:382–388.

ROVAINEN, C. M. 1996. Preface (to the 1995 Karger workshop on agnathan neurobiology). Brain Behav. Evol. **48**:235–236.

SANSOM, I. J., M. M. SMITH, and M. P. SMITH. 1996. Scales of thelodont and shark-like fishes from the Ordovician of Colorado. Nature **379**:628–629.

SCHNARE, M. N., S. H. DAMBERGER, M. W. GRAY, and R. GUTELL. 1996. Comprehensive comparison of structural characteristics in eukaryotic cytoplasmic large subunit (23S-like) ribosomal RNA. J. Mol. Biol. **256**:701–719.

SMITH, S. W., R. OVERBEEK, C. R. WOESE, W. GILBERT, and P. M. GILLEVET. 1994. The genetic data environment an expandable GUI for multiple sequence analysis. CABIOS **10**:671–675.

SOWER, S. 1998. Brain and pituitary hormones of lamprey, recent findings and their evolutionary significance. Am. Zool. **38**:15–38.

STOCK, D. W. 1992. A molecular phylogeny of fishes. Ph.D. thesis, University of Illinois at Urbana-Champaign.

STOCK, D. W., K. D. MOBERG, L. R. MAXSON, and G. S. WHITT. 1991. A phylogenetic analysis of the 18S ribosomal RNA sequence of the coelacanth *Latimeria chalumnae*. Environ. Biol. Fishes **32**:99–117.

STOCK, D. W., and G. S. WHITT. 1992. Evidence from 18S ribosomal RNA sequences that lampreys and hagfishes form a natural group. Science **257**:787–789.

STRIMMER, K., and A. VON HAELESER. 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. **13**:964–969.

SULLIVAN, J. 1996. Combining data with different distributions of among-site rate variation. Syst. Biol. **45**:375–380.

SULLIVAN, J., J. A. MARKERT, and C. W. Kilpatrick. 1997. Phylogeography and molecular systematics of the *Peromyscus aztecus* species group (Rodentia: Muridae) inferred using parsimony and likelihood. Syst. Biol. **46**:426–440.

SULLIVAN, J., and D. L. SWOFFORD. 1997. Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. J. Mamm. Evol. **4**:77–86.

SUZUKI, M., K. KUBOKAWA, H. NAGASAWA, and A. URANO. 1995. Sequence analysis of vasotocin cDNAs of the lamprey *Lampetra japonica*, and the hagfish, *Eptatretus burgeri*: evolution of cyclostome vasotocin precursors. J. Mol. Endocrinol. **14**:67–77.

SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, and D. M. HILLIS. 1996. Phylogenetic inference. Pp. 407–514 *in* D. M. HILLIS, C. MORITZ, and B. K. MABLE, eds. Molecular systematics. 2nd edition. Sinauer, Sunderland, Mass.

TATENO, Y., N. TAKEZAKI, and M. NEI. 1994. Relative efficiencies of the maximum-likelihood, neighbor joining, and maximum-parsimony methods when substitution rate varies with site. Mol. Biol. Evol. **11**:261–277.

TURBEVILLE, J. M., J. R. SCHULZ, and R. A. RAFF. 1994. Deuterostome phylogeny and the sister group of the chordates: evidence from molecules and morphology. Mol. Biol. Evol. **11**:648–655.

VAN DER AUWERA, G., S. CHAPELLE, and R. DE WACHTER. 1994. Structure of the large ribosomal subunit RNA of *Phytophthora megasperma*, and phylogeny of the oomycetes. FEBS Lett. **338**:133–136.

WADA, H., and N. SATOH. 1994. Details of the evolutionary history from invertebrates to vertebrates, as deduced from the sequences of 18S rDNA. Proc. Natl. Acad. Sci. USA **91**:1801–1804.

WADDELL, P. 1995. Statistical methods of phylogenetic analysis, including Hadamard conjugations, LogDet transforms, and maximum likelihood. Ph.D. dissertation, Massey University, Palmerston North, New Zealand.

WADDELL, P., and D. PENNY. 1996. Evolutionary trees of apes and humans from DNA sequences. Pp. 53–73 *in* A. J. LOCK and C. R. PETERS, eds. Handbook of symbolic evolution. Clarendon Press, Oxford.

WINNEPENNINCKX, B., T. BACKELJAU, and R. DEWACHTER. 1993. Extraction of high molecular weight DNA from molluscs. Trends Genet. **9**:407.

YALDEN, D. W. 1985. Feeding mechanisms as evidence for cyclostome monophyly. Zool. J. Linn. Soc. **84**:291–300.

YANG, Z. 1994*a*. Estimating the pattern of nucleotide substitution. J. Mol. Evol. **39**:105–111.

———. 1994*b*. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. **39**:306–314.

YANG, Z., N. GOLDMAN, and A. FRIDAY. 1995. Maximum likelihood trees from DNA sequence data: a peculiar statistical estimation problem. Syst. Biol. **44**:384–399.

ZARDOYA, R., Y. CAO, M. HASEGAWA, and A. MEYER. 1998. Searching for the closest living relative(s) of tetrapods through evolutionary analyses of mitochondrial and nuclear data. Mol. Biol. Evol. **15**:506–517.

ZARDOYA, R., and A. MEYER. 1996. Evolutionary relationships of the coelacanth, lungfishes, and tetrapods based on the 28S ribosomal RNA gene. Proc. Natl. Acad. Sci. USA **93**: 5449–5454.

MASAMI HASEGAWA, reviewing editor

Accepted September 2, 1998