

Rate Acceleration and Long-branch Attraction in a Conserved Gene of Cryptic Daphniid (Crustacea) Species

Angela R. Omilian and Derek J. Taylor

Department of Biological Sciences, State University of New York at Buffalo, Buffalo

The nuclear large subunit (LSU) rRNA gene is a rich source of phylogenetic characters because of its large size, mosaic of slowly and rapidly evolving regions, and complex secondary structure variation. Nevertheless, many studies have indicated that inconsistency, bias, and gene-specific error (e.g., within-individual gene family variation, cryptic sequence simplicity, and sequence coevolution) can complicate animal phylogenies based on LSU rDNA sequences. However, most of these studies sampled small gene fragments from expansion segments—among animals only five nonchordate complete LSU sequences are published. In this study, we sequenced near-complete nuclear LSU genes from 11 representative daphniids (Crustacea). The daphniid expansion segment V6 was larger and showed more length variation (90–351 bp) than is found in all other reported LSU V6 sequences. Daphniid LSU (without the V6 region) phylogenies generally agreed with the existing phylogenies based on morphology and mtDNA sequences. Nevertheless, a major disagreement between the LSU and the expected trees involved a positively misleading association between the two taxa with the longest branches, *Daphnia laevis* and *D. occidentalis*. Both maximum parsimony (MP) and maximum likelihood (ML) optimality criteria recovered this association, but parametric simulations indicated that MP was markedly more sensitive to this bias than ML. Examination of data partitions indicated that the inconsistency was caused by increased nucleotide substitution rates in the branches leading to *D. laevis* and *D. occidentalis* rather than among-taxon differences in base composition or distribution of sites that are free to vary. These results suggest that lineage-specific rate acceleration can lead to long-branch attraction even in the conserved genes of animal species that are almost morphologically indistinguishable.

Introduction

Phylogenetic robustness is often inferred when trees found using different optimality criteria are congruent, but with some complex modes of molecular evolution, phylogenetic methods often agree on the incorrect phylogeny. The most insidious case is an incorrect phylogeny that is strengthened as more data are added. Felsenstein (1978) described this statistical inconsistency as positively misleading, and demonstrated a particular scenario (long-branch attraction [LBA]) whereby phylogenetic methods are misled by pronounced branch length differences (Hendy and Penny 1989). Simulations and experimental viral phylogenies have shown that branch length differences can create positively misleading trees for the maximum parsimony (MP), minimum evolution, and maximum likelihood (ML) methods (Hillis et al. 1992; Gaut and Lewis 1995; Huelsenbeck 1995; Cunningham, Zhu, and Hillis 1998). Additionally, the problem may be exacerbated by incorrect assumptions about the model of evolution (Chang 1996) and by parameters that vary across taxa (nonstationarity), such as among-site rate heterogeneity, among-taxon base compositional bias (Cunningham, Zhu, and Hillis 1998), and among-taxon differences in the distribution of nucleotide sites that are free to vary (Lockhart et al. 1998). So, distantly related taxa may be misleadingly recovered as

sister groups solely because of parallel mutations among rapidly evolving lineages, incorrect models of evolution, shared base composition, or shared distributions of variable sites (also known as covariotide evolution).

Although LBA is commonly invoked as a source of inconsistency, it remains unclear how common the theoretical conditions that lead to LBA are in nature (Huelsenbeck 1998; Sanderson et al. 2000; Weins and Hollingsworth 2000). Part of the problem is the lack of a consistent reconstruction method for real data that might remedy the bias. The initial assumption that ML is a more consistent method than MP has been challenged in several studies (Sanderson and Kim 2000). Even when potential biases are clearly identified in real data, as in the well-studied Strepsiptera-Diptera problem, it is difficult to determine if the inferred phylogeny results from a bias or from shared evolutionary history (Huelsenbeck 1997; Whiting 1998; Steel, Huson, and Lockhart 2000). Independent phylogenetic evidence is necessary to rule out shared evolutionary history as a source of a given phylogenetic relationship (Weins and Hollingsworth 2000). Also, in many cases multiple biases interact and are difficult to tease apart (Steel, Huson, and Lockhart 2000).

One nuclear gene that exhibits complex molecular evolution and a demonstrated potential for biased phylogenetic results is the nuclear large subunit (LSU) rRNA gene. Among-taxon base composition and substitution rate differences are commonly observed in this gene. For example, insects exhibit a nonstationarity pattern because dipterans possess LSU stem regions with 10% greater adenine-thymine content than most nondipteran relatives (Friedrich and Tautz 1997). Also, the dipteran stem lineage possesses a 20-fold greater substitution rate than other holometabolous insects. The overall mosaic of slowly and rapidly evolving regions in LSU

Abbreviations: CI, consistency index; HSP, heat shock protein; LBA, long-branch attraction; LSU, large subunit; ML, maximum likelihood; MP, maximum parsimony; PCR, polymerase chain reaction; RI, retention index; SSU, small subunit.

Key words: *Daphnia*, long-branch attraction, nuclear large subunit, rDNA, rate acceleration, cryptic species.

Address for correspondence and reprints: Angela R. Omilian, Department of Biology, Jordan Hall, 1001 E. Third Street, Indiana University, Bloomington, Indiana 47405. E-mail: aomilian@indiana.edu.

Mol. Biol. Evol. 18(12):2201–2212, 2001

© 2001 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

rDNA, with variable domains evolving faster (6–10 times faster in plants) than conserved domains, creates strong among-site rate heterogeneity (Hillis and Dixon 1991; Kuzoff et al. 1998). This mosaic pattern also predisposes the gene to differing distributions of variable sites among taxa that possess different overall rates of substitution.

In addition to potential bias, some initial studies of the LSU reported several unique sources of phylogenetic error. These include sequence coevolution, within-individual gene family variation, frequent indel mutations, and cryptically simple repeats (Tautz, Trick, and Dover 1986; Hancock and Dover 1988; Bult, Sweere, and Zimmer 1995; Nunn et al. 1996). Nevertheless, detailed studies of the entire LSU gene have shown that these sources of error are negligible in some eukaryotes and that this gene yields strong phylogenetic information. For example, Kuzoff et al. (1998) examined complete LSU sequences from fifteen plant taxa and found significant phylogenetic concordance with 18S rDNA and *rbcL* gene phylogenies, greater phylogenetic information than other genes, and few apparent sources of phylogenetic error. Likewise, Mallat and Sullivan (1998) used the entire LSU sequence of 10 chordates to test the hypothesis of cyclostome monophyly. The results indicated strong phylogenetic signal for this question and were concordant with phylogenies based on several nuclear protein-coding genes (Kuraku et al. 1999) as well as whole mtDNA sequence phylogenies (Delarbre et al. 2000). Finally, Mugridge et al. (2000) found that complete sequences recovered the expected topology of sarcocystid protozoans, but the use of shorter LSU segments compromised the phylogeny. Thus, accurate phylogenetic information is present in the few existing studies of the entire gene despite multiple potential sources of error.

Here, we present the first exploration of phylogenetic utility, bias, and inconsistency in nearly complete nuclear LSU ribosomal RNA gene sequences from non-chordate animals. Our data set consists of 12 new sequences from daphniid crustaceans for which there are several existing robust associations based on independent mtDNA sequence, heat shock protein (HSP) 90, and morphological information (Lehman et al. 1995; Colbourne and Hebert 1996; Taylor, Hebert, and Colbourne 1996; Taylor, Finston, and Hebert 1998; unpublished data). Although these groups (fig. 1) are approximations, we explore a disagreement that involves the breaking up of a strongly supported cryptic species clade, *Daphnia laevis* and *D. dentifera*. These species are so similar in morphology that they have often been incorrectly synonymized as *D. longispina* (see Brooks 1957).

Materials and Methods

Specimen Collections

We used specimens from eight species of the genus *Daphnia* and from two species of the genus *Daphniopsis* (Ds.) for sequencing and phylogenetic analysis (table 1). These ingroup species were chosen to represent each of

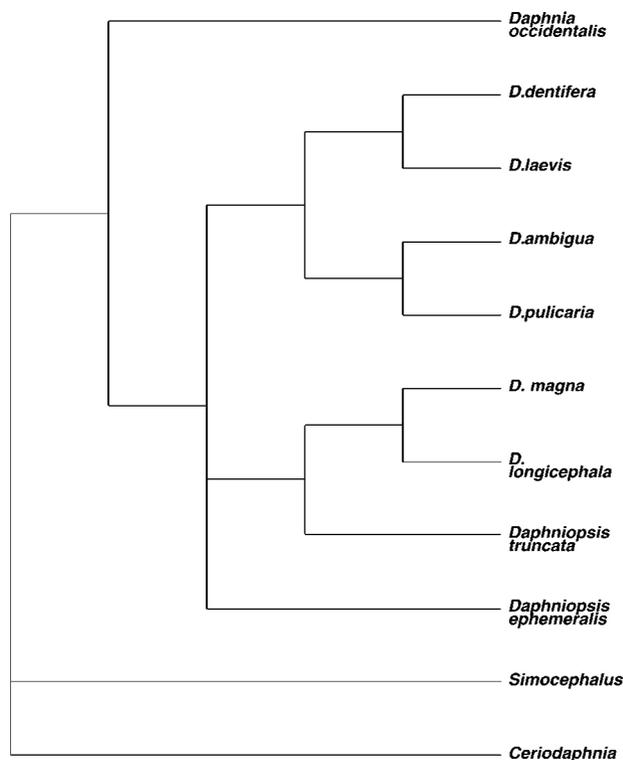


FIG. 1.—Concordance tree hypothesis for the representative daphniid taxa used in this study. Evidence is based on mtDNA, HSP90, and morphology (see text).

the proposed subgenera and genera from previous phylogenetic studies (Colbourne and Hebert 1996). A partial *D. dubia* sequence (3,955 bp of the 4,611-bp alignment) was used in only one analysis to verify the authenticity of the *D. laevis* sequence, and as a taxon addition to break up long branches. *Daphnia dubia* proved to be very closely related to *D. laevis*. We used the daphniid genera *Simocephalus* and *Ceriodaphnia* as outgroup taxa because they are noncontroversially closely related to, but not part of, *Daphnia* or *Daphniopsis* (Olesen 1998; Taylor, Crease, and Brown 1999).

Laboratory Protocols

DNA was extracted from previously frozen individuals using the 2× cetyl-trimethylammonium bromide buffer protocol (Doyle and Doyle 1987). Each polymerase chain reaction (PCR) consisted of 40 μl irradiated H₂O, 5 μl 10× buffer, 1 μl deoxynucleotide triphosphates, 1.5 μl of each primer, 1 μl *Taq* DNA polymerase, and 1 μl of DNA template. Initially, primers were designed from conserved regions within the nuclear LSU rDNA gene in *Drosophila*, and later from daphniids (table 2). Because of the nuclear LSU's large size, we partitioned the gene into three overlapping segments for amplification. The PCR conditions for amplification of all three segments consisted of 40 cycles of 60 s at 94°C, 60 s at 55°C, 90 s at 72°C, followed by one cycle of 6 min at 72°C. PCR was conducted on a Stratagene RoboCycler.

Table 1
Species Names, Collection Sites, and GenBank Accession Numbers for Daphniid Specimens Sequenced in this Study. *Daphnia* is Divided into Subgenera (in parentheses)

Species	Collection Location	GenBank Accession Codes
<i>Daphnia (Ctenodaphnia)</i>		
<i>D. longicephala</i>	Fleurieu Peninsula, Australia	AF346516
<i>D. magna</i>	Crescent Lake, Neb.	AF346515
<i>Daphnia (Daphnia)</i>		
<i>D. ambigua</i>	Round Pond, Mendon Park, N.Y.	AF346513
<i>D. dubia</i>	Corner Brook, Newfoundland, Canada	AF348426, AF404853
<i>D. laevis</i>	Truro, Cape Cod, Mass.	AF346512
<i>D. occidentalis</i>	Northcliffe, Western Australia	AF346510
<i>D. pulicaria</i>	Buffalo, N.Y.	AF346514
<i>D. dentifera</i>	Deep Lake, Mendon Park, N.Y.	AF346511
<i>Daphniopsis</i>		
<i>Ds. ephemeralis</i>	Buffalo, N.Y.	AF346518
<i>Ds. truncata</i>	Moora, Western Australia	AF346517
<i>Ceriodaphnia rotunda</i>	Greeleyville, S.C.	AF346519
<i>Simocephalus serrulatus</i>	First Sister Lake, Ann Arbor, Mich.	AF346520

PCR products were purified using either the QIAEX II agarose gel-extraction protocol or the Amicon kit for DNA extraction. We sequenced the nuclear LSU in both directions. Some templates proved difficult to sequence through the hypervariable expansion segments and were subsequently cloned with the Invitrogen TOPO TA cloning kit for sequencing (Version B). More than three clones were sequenced per fragment. The ABI PRISM BigDye terminator cycle sequencing ready reaction kit or the Amersham Pharmacia Biotech Thermo Sequenase fluorescent-labeled primer cycle sequencing kit with 7-deaza-dGTP and the ABI 377 or a LI-COR 4200 automated DNA sequencer were used for sequencing.

Sequence Assembly and Alignment

Sequences were assembled and edited with Sequencher 4.0 (Gene Codes Corporation, Ann Arbor, Mich.) and then aligned with Clustal X using the default parameters (Thompson, Higgins, and Gibson 1994;

Thompson et al. 1997). The alignment was manually adjusted with BioEdit 4.7.1 (Hall 1999) according to core region rRNA secondary structure (De Rijk et al. 2000). The alignment length was 4,661 base pairs, but 1,038 sites could not be aligned unambiguously and were excluded from the phylogenetic analysis (the alignment is available at <http://www.herbaria.harvard.edu/treebase>, study accession number S657, or from A.R.O. upon request). Variable domain boundaries were based on the proposal of De Rijk et al. (2000). Sequences were deposited into GenBank and accession numbers are listed in table 1.

Phylogenetic Analyses

All phylogenetic analyses were conducted in PAUP* 4.0 (Swofford 2000). Base compositions were calculated for entire LSU sequences, conserved cores, variable domains, and parsimony-informative sites. In order to assess phylogenetic signal in the sequences, the g_1 skewness statistic (Hillis and Huelsenbeck 1992) was

Table 2
Primers Used for PCR Amplification and Sequencing of Nuclear LSU rDNA. Primers Used for Amplification Were D1f and D6br for the First Segment of the Nuclear LSU rDNA, 28ee and D8r for the Second Segment, and 28v' and D12r for the Third Segment. The Second Segment of *D. occidentalis* Was Amplified by Primers 4bf and D7r, and D7f and D11r

Primer Name	Direction	5'-3' Sequence	Primer Origin
D1f.....	Forward	GGGACTACCCCTGAATTTAAGCAT	Modified from Littlewood et al. (1994)
4bf.....	Forward	ACACGGACCAAGGAGTCTAACA	Taylor, Crease, and Brown (1999)
28ee.....	Forward	ATCCGCTAAGGAGTGTGTAACAACCTACC	Hillis and Dixon (1991)
D7f.....	Forward	CCGCAGCTGGTCTCCAAGGT	This study
28v'.....	Forward	CTTAAGGTAGCCAAATGCCTC	This study
D10f.....	Forward	GATCCTTCGATGTCGGCTCTTC	This study
D2r.....	Reverse	AGCCTTAGAAGGAGTTTACCT	This study
D6r.....	Reverse	CCAGCTATCCTGAGGAAACTTCG	This study
D6br.....	Reverse	CACACGAAACCCTTCTCCAC	This study
D7r.....	Reverse	GAGCCAATCCTTTTCCCGAAGTTACGG	This study
D8r.....	Reverse	GAGTCAAGCTCAACAGGGTCTTCTTTCCC	This study
D10r.....	Reverse	GGCCGCCATAAGCCAGTTATCCCTATGGTA	This study
D11r.....	Reverse	GCCTTCTCCAGCATCGATTCTAC	This study
D12r.....	Reverse	CTGCTCTGCCGTGTACAACAAC	This study

Table 3
Results of Hierarchical Likelihood Ratio Tests Comparing Statistical Differences Between Increasingly Complex Models of DNA Sequence Evolution for Daphniid Nuclear LSU rDNA Sequences. The Degrees of Freedom are the Difference in the Number of Free Parameters Between Models

Models of DNA Sequence Evolution	$2(\ln L_1 - \ln L_0)$	Degrees of Freedom	<i>P</i> -value
JC versus F81	62.2471	3	<0.000001
F81 versus HKY	151.6777	1	<0.000001
HKY versus TrN	31.1025	1	<0.000001
TrN versus TIM	1.6162	1	0.203621
TrN versus TrN + Γ	275.1660	1	<0.000001
TrN + Γ versus TrN + I + Γ	63.1152	1	<0.000001

NOTE.—JC: Jukes-Cantor model (Jukes and Cantor 1969); F81: Felsenstein model (Felsenstein 1981); HKY: Hasegawa-Kishino-Yano model (Hasegawa et al. 1985); TrN: Tamura-Nei model (Tamura and Nei 1993); TIM: Transitional model (Rodriguez et al. 1990).

calculated from 10,000 random tree length distributions. One taxon of each strongly supported clade was then removed to determine if phylogenetic signal was present in the deeper branches.

Maximum parsimony analysis was used with a branch and bound search algorithm, all characters weighted equally, and gaps treated as characters and as missing data. Nonparametric bootstrapping was performed with 1,000 bootstrap replicates and the MP search settings just listed. Nonparametric bootstrap resampling with MP was carried out with increasing numbers of resampled bases.

Fifty-six ML models were assessed by a series of likelihood ratio tests with the program Modeltest 3.0 (Posada and Crandall 1998). Hierarchical model fitting indicated that the Tamura-Nei model (TrN) with invariable sites and the gamma parameter (Tamura and Nei 1993) had the best fit to the data (TrN + I + Γ ; table 3). This is a special case of the general time-reversible model with among-site rate variation and the following parameters being estimated from the data: three types of base substitutions, the proportion of invariable sites, and the gamma estimate of among-site rate variation with four rate categories. In order to find the best tree under the ML criterion, we used this model with a heuristic search, tree bisection-reconnection branch swapping, and 10 random sequence taxon additions. To test the hypothesis that the observed tree was more likely than the expected tree, we used parametric bootstrapping (Swofford-Olsen-Waddell-Hillis-test; partial optimization under H_A ; direct estimation of *P*-value) (Huelsenbeck and Crandall 1997; Shimodaira and Hasegawa 1999; Goldman, Anderson, and Rodrigo 2000).

In order to assess the potential of long branches to attract because of branch length disparity alone, we performed parametric simulations (Hillis, Moritz, and Mable 1996; Huelsenbeck 1998). We used the best-fit TrN + I + Γ model with branch length parameters estimated from the best ML tree found to contain the expected *D. laevis/D. dentifera* clade. Then, 100 data sets were simulated by Seq-Gen 1.21 (Rambaut and Grassly 1997),

using these parameters, and analyzed using MP and ML (TrN + I + Γ model) optimality criteria. Topologies found from each of the 100 simulated data sets were tallied in four categories: expected clade (*D. laevis/D. dentifera*), LBA clade (*D. laevis/D. occidentalis*), mixed LBA and other clades, and other clades. For some of the analyses, there was more than one best tree. Search results were assigned to expected or LBA only if the strict consensus of trees contained this clade. As a control, a second parametric bootstrap analysis was carried out as described earlier, but with the branch length parameters based on the inferred ML (*D. laevis/D. occidentalis*) tree.

The effect of differing distributions of invariant sites on tree structure was examined by phylogenetic analysis of covariotide data partitions (Lockhart et al. 1998). The data were partitioned into five categories for two groups—long-branched taxa (*D. laevis* and *D. occidentalis*) and other taxa. The categories were: type 1, sites that are invariant across all taxa; type 2, sites that are invariant within *D. laevis/D. occidentalis*, and within the remaining taxa, but different between these groups (e.g., a site with the bases AA/GGGGGGGG, where the first two taxa are *D. laevis* and *D. occidentalis*); type 3, sites that are invariant across all taxa but vary in *D. laevis/D. occidentalis*; type 4, sites that are invariant in *D. laevis/D. occidentalis* but vary in the other taxa; and type 5, sites that vary in both groups. Type 3 and 4 sites are generally considered to be the covariotide sites (Lockhart et al. 1998). Nevertheless, if the putative LBA group has just two taxa, and a positively misleading shared character occurs at a site that is invariable in other taxa, then a type 2 site will result. Removing these type 2 sites from the analysis should then eliminate or markedly reduce the covariotide bias.

Results

Nuclear LSU rDNA features

Nearly complete nuclear LSU rDNA in the Daphniidae ranged in length from 4,038 to 4,376 base pairs. Length variation was considerable among species with an approximate 340-bp difference between the shortest (*D. ephemeralis*) and longest (*D. pulicaria*) daphniid nuclear LSU genes. Much of this size variation (90–351 bp) occurred in the putative hidden break or V6 area (table 4). After alignment, there were 280 variable sites and 106 parsimony informative sites (128 informative sites with gapped sites included). Among taxa, there was no evidence for base composition heterogeneity in the total data (table 5; $\chi^2 = 2.8782$, *df* = 30, *P* > 0.99; mean content: A = 0.2273, C = 0.2509, G = 0.3184, T = 0.2033) or in the parsimony informative sites ($\chi^2 = 18.368$, *df* = 30, *P* = 0.95). The mean GC content was 0.5844 for the variable domains and 0.5364 for the conserved cores.

Nuclear LSU rDNA Phylogeny

The g_1 value ($g_1 = -1.4127$, *P* = 0.01; -1.7 with gaps) for 11 taxa suggested that significant phylogenetic signal exists for the data. When one of each closely

Table 4
Size (bp) of Expansion Regions (V1–V11) in the Nuclear LSU rDNA Genes of Daphniid Crustaceans

Species	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
<i>D. ambigua</i>	175	381	194	51	55	236	282	269	24	247	15
<i>D. dentifera</i>	175	373	182	51	55	178	282	281	24	252	15
<i>D. laevis</i>	177	374	192	58	55	168	289	231	32	243	15
<i>D. longicephala</i>	175	364	171	53	55	118	282	221	24	237	15
<i>D. magna</i>	175	392	171	56	55	199	282	221	25	236	15
<i>D. occidentalis</i>	175	415	189	54	55	108	287	283	32	307	15
<i>D. pulicaria</i>	175	385	181	50	55	351	282	270	24	236	15
<i>Ds. ephemeralis</i>	175	361	169	51	55	90	282	235	20	237	15
<i>Ds. truncata</i>	175	394	173	53	55	335	282	226	24	238	15
<i>C. rotunda</i>	175	341	160	56	53	142	279	229	26	257	15
<i>S. serrulatus</i>	176	404	163	61	53	230	279	268	24	249	15

NOTE.—Expansion region boundaries are those proposed by De Rijk et al. (2000). Sizes for V12 are not given because the 3' boundary is difficult to establish from alignment with existing database sequences.

related taxon pairs was removed from the analysis, the g_1 value remained significant (e.g., $g_1 = -1.4457$, -1.3517 , -1.6305 , -0.6126 ; $P = 0.01$), suggesting that the phylogenetic signal exists beyond the sister taxa.

The MP trees generally shared the topology of the expected concordance tree (tree length = 388, consistency index [CI] = 0.789, retention index [RI] = 0.539; with gapped sites, two best trees were found of 465 steps; CI = 0.794, RI = 0.549; fig. 2). The only differences in the MP bootstrap consensus tree and the reference tree (fig. 1) were the disruption of the *D. laevis/D. dentifera* clade in favor of a *D. laevis/D. occidentalis* clade and the movement of *D. ephemeralis* to a basal position in the ingroup. This position for *D. ephemeralis* had weak support, but the branch leading to the *D. laevis/D. occidentalis* clade had strong bootstrap support when gapped sites were included as a fifth character (87%) and moderate support (75%) when gapped sites were scored as missing. The observed tree was eight steps shorter (465 steps) than the best tree found from an analysis constrained to have the expected *D. laevis/D. dentifera* clade (473 steps). When the number of sites available for resampling was varied (and gapped sites were excluded), the bootstrap value for the *D. laevis/D. occidentalis* clade increased with the number of sites

used, reaching a maximum value of 77% at 3,623 sites (fig. 3).

The ML tree had the same topology as the bootstrap MP tree, but the support for the *D. laevis/D. occidentalis* group was weak at 48% (fig. 2). Also, the parametric bootstrap analysis failed to reject the hypothesis that this observed *D. laevis/D. occidentalis* tree was more likely than the expected *D. laevis/D. dentifera* tree ($\delta = L_1 - L_2 = 1.9843$; $P = 0.07$).

LBA and Covariotide Analyses

The unexpected *D. laevis/D. occidentalis* grouping involved the two branches that are the longest in the ingroup. Indeed, at an ML length of 0.042 (fig. 2), the branch leading to *D. laevis* is the longest branch among all taxa and greater than five times longer than the other ingroup branches. Simulations designed to determine if the branches are long enough to be misleading were carried out and supported an LBA scenario that involved the long branches leading to *D. laevis* and *D. occidentalis*. Even when the simulated data were parameterized with a *D. laevis/D. dentifera* clade, parsimony recovered the long-branched *D. laevis/D. occidentalis* clade in a majority of replicates (62%; fig. 4A) and strict consensus

Table 5
Length and Nucleotide Composition of the Nuclear LSU rDNA Gene in Various Daphniid Crustaceans

TAXON	LENGTH (bp)	NUCLEOTIDE FREQUENCIES OF COMPLETE NUCLEAR LSU				G + C CONTENT	
		A	C	G	T	Conserved Cores	Variable Domains
<i>D. ambigua</i>	4282	0.220	0.243	0.309	0.226	0.537	0.567
<i>D. dentifera</i>	4220	0.215	0.245	0.312	0.226	0.536	0.578
<i>D. laevis</i>	4226	0.216	0.246	0.310	0.226	0.529	0.583
<i>D. longicephala</i>	4064	0.219	0.250	0.318	0.211	0.537	0.601
<i>D. magna</i>	4173	0.219	0.247	0.316	0.216	0.537	0.590
<i>D. occidentalis</i>	4276	0.210	0.250	0.314	0.224	0.537	0.590
<i>D. pulicaria</i>	4376	0.222	0.240	0.308	0.227	0.534	0.562
<i>Ds. ephemeralis</i>	4038	0.220	0.249	0.317	0.212	0.537	0.596
<i>Ds. truncata</i>	4320	0.218	0.245	0.315	0.219	0.535	0.584
<i>C. rotunda</i>	4063	0.231	0.243	0.310	0.213	0.538	0.572
<i>S. serrulatus</i>	4252	0.216	0.252	0.319	0.211	0.539	0.601

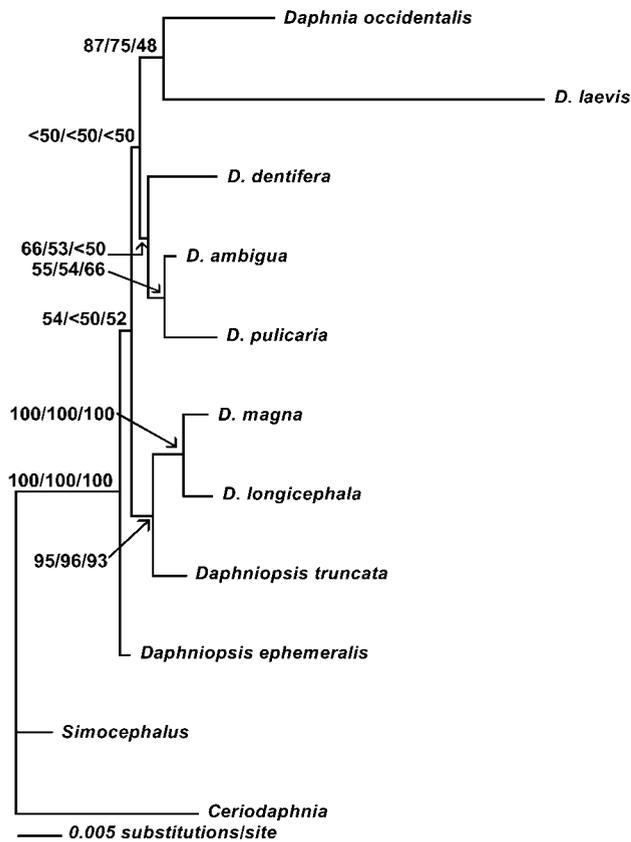


FIG. 2.—Phylogram of daphniid relationships based on ML analysis of the nuclear LSU rDNA (tree score = 7,350.3874). The branch lengths reflect the amount of evolution. Numbers indicate nonparametric bootstrap support for MP with gaps as characters, MP with gaps as missing characters, and ML.

trees of replicates (52%). The true tree was recovered in only 32% of the replicates. The opposite pattern was recovered by ML analysis as the correct *D. laevis*/*D. dentifera* clade was recovered in 65% of the replicates, whereas the long-branched clade was recovered in only 14% of replicates. When the true tree was changed to the inferred tree (which contains a *D. laevis*/*D. occidentalis* clade), both MP and ML showed strong recovery of the correct tree (90%–96%; fig. 4B).

In order to explore the sensitivity of the putatively inconsistent placement of *D. laevis* to taxon addition and removal, different subsets of the data were analyzed. First, when the potential long-branch attractor, *D. occidentalis*, was removed, the best MP tree (fig. 5A) placed *D. laevis* back in the expected grouping with *D. dentifera* (one best tree of 451 and 342 steps, with and without gapped sites, respectively). Second, the removal of the other long-branch taxon, *D. laevis*, resulted in the expected placement of *D. occidentalis* in a basal position to *Daphnia*/*Daphniopsis* (fig. 5B). The addition of another taxon, *D. dubia*, a species in the *D. laevis*/*D. dentifera* group (Taylor, Finston, and Hebert 1998), failed to free *D. laevis* of its putative long-branch attraction with *D. occidentalis* in the MP analysis (bootstrap = 90; fig. 5C). The long branch leading to the *D. laevis*/*D. dubia* clade strongly indicates that a substitu-

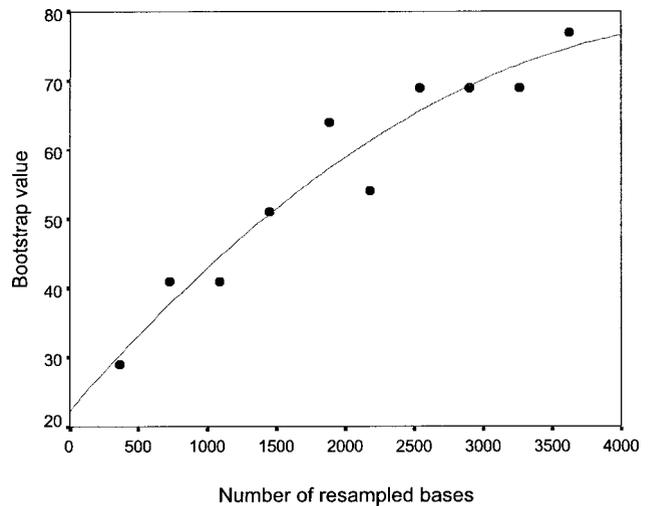


FIG. 3.—Bootstrap support for the *D. laevis*/*D. occidentalis* clade using parsimony analyses of nuclear LSU rDNA sequences. Percentage values are given as a function of the number of resampled nucleotides.

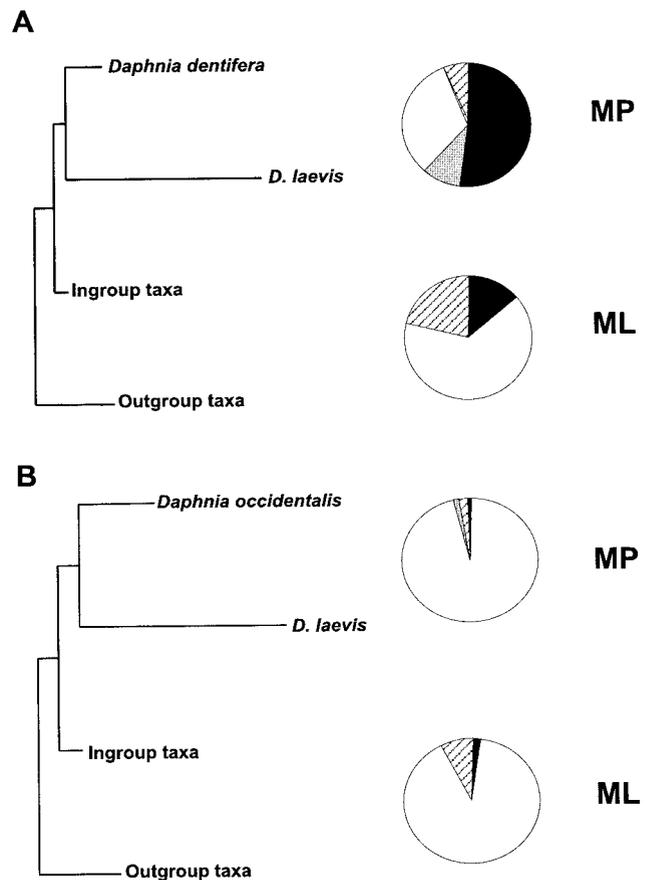


FIG. 4.—Tally of the parametric simulation replicates involving long-branched taxa from daphniid nuclear LSU rDNA sequences. A cartoon summary of the correct tree for each simulation is shown on the left. Each pie chart indicates the proportion of 100 replicates that were recovered by either MP or ML analyses: (A) the correct *D. laevis*/*D. dentifera* clade (white), the putative long-branch attraction clade of *D. laevis*/*D. occidentalis* (black), a mix of correct and LBA clades (shading), or other topologies (hatched); and (B) the correct long-branch *D. laevis*/*D. occidentalis* clade (white), the *D. laevis*/*D. dentifera* clade (black), a mix of correct and *D. laevis*/*D. dentifera* clades (shading), or other topologies (hatched).

tion rate increase occurred before the *D. dubia* and *D. laevis* lineages diverged. A final taxon set excluded both putative long-branch taxa, and the alignment and exclusion sets were adjusted to allow for the inclusion of an increased number of unambiguous sites. Realignment of the dataset provided for 3,759 nucleotide sites and 43 additional parsimony-informative characters. One best MP tree was obtained (tree length = 422, CI = 0.810, RI = 0.673) that agreed with the concordance tree, and bootstrap support was high for the concordance clades (see fig. 6). For the ML analysis, the TrN + I + Γ model showed the best fit to the realigned dataset (fig. 6; tree score = 7,704.2484, model tests not shown), and the optimal ML tree had the same topology as the MP tree.

Partitioning the data according to Lockhart et al. (1998) gave the following distribution of characters: type 1 = 3,285, type 2 = 6, type 3 = 114, type 4 = 161, and type 5 = 57. It is clear that *D. laevis* and *D. occidentalis* do have an increased number of sites that are free to vary (manifested in the number of type 3 characters). Nevertheless, MP analysis of the partitions revealed that most of the support for the *D. laevis/D. occidentalis* clade occurred at sites that are free to vary in the rest of the taxa (type 4 in this analysis; one best tree of 168 steps without gaps and 205 steps with gaps). If the inconsistency for *D. laevis* and *D. occidentalis* was caused by shared bases at sites that were invariant in the other taxa, then the removal of type 2 sites should remove the inconsistency. But type 2 sites are few in number, and their removal failed to prevent a strongly supported *D. laevis/D. occidentalis* clade.

Discussion

Since the theoretical paper of Felsenstein (1978) on the extreme conditions that create positively misleading phylogenies, the existence of such scenarios in real data has been controversial (Huelsenbeck 1997). Empirical studies have shown that many biases do occur in real data, but in order to demonstrate that a tree is misled by a bias, shared evolutionary history should reasonably be ruled out (Weins and Hollingsworth 2000). The topology of a concordance tree based on independent evidence (mtDNA, HSP90, and morphology) for daphniids suggests that the placement of *D. laevis* on the LSU tree is misleading. The finding that *D. laevis* groups with *D. dentifera* (i.e., agrees with the reference tree) when *D. occidentalis* is removed from the LSU data set is also suggestive of a taxon-specific misleading effect. Finally, statistical inconsistency is evident as support for the incorrect *D. laevis/D. occidentalis* clade increases as more characters are added.

What is the source of the bias? Studies of bias have generally involved deeper phylogenies, often above the ordinal level, where the source of inconsistency is often blurred by the joint action of multiple biases or by taxon-sampling artifacts. The Diptera-Strepsiptera and eukaryotic phyla studies are examples where base compositional and covariotide biases may be acting (Steel, Huson, and Lockhart 2000). In contrast, the LSU data in daphniids involves an inconsistency at the species

level that seems to lack these biases. Among-taxon base compositional bias, for instance, is not significant for the entire dataset, and the covariotide analysis indicates that although *D. occidentalis* and *D. laevis* have differing distributions of sites that are free to vary from the rest of the taxa, support for the inconsistency comes from sites that are free to vary in the non-long-branch taxa. Therefore, base compositional and covariotide biases fail to explain the inconsistency. Also, because the simulated data sets in the parametric bootstrapping lacked compositional and covariotide bias, but still recovered the same inconsistent topology as the observed data, additional biases must be acting.

The analyses are consistent with an LBA bias that results from an accelerated rate of evolution in the variable sites of the *D. laevis/D. dubia* group. The inconsistency clearly involves the longest branch in the tree and the second longest branch in the ingroup (*D. occidentalis*), a prediction of an LBA scenario. The correct grouping of *D. laevis* when no other long-branch attractor is present (i.e., when *D. occidentalis* is removed) is also consistent with LBA. Finally, even when the simulations contain a *D. laevis/D. dentifera* tree parameter, MP recovered the incorrect *D. laevis/D. occidentalis* clade in the majority of replicates. The evidence for long-branch repulsion is weak as both ML and MP recovered the correct long-branched clade in a separate simulation. These findings strongly suggest that an accelerated rate of evolution in the LSU of the lineage leading to *D. laevis* contributes to an LBA bias.

Long-branch attraction is only one of several explanations for a positively misleading association between *D. laevis* and *D. occidentalis*. Other plausible explanations include alignment artifacts, paralogous gene comparisons, cryptic sequence simplicity, erroneous concordance trees, and suboptimal models of evolution. The alignment that we used was conservative with most of the rapidly evolving expansion segments and length-variable regions removed from the analysis. Adding ambiguously aligned sites to the analysis still results in the *D. laevis/D. occidentalis* clade (not shown). Therefore, alignment is unlikely to be the sole cause of inconsistency. It is possible that a rogue paralogous rDNA copy was sequenced in *D. laevis*. However, this seems unlikely for two reasons. First, within-individual variation was undetected in the sequenced PCR product of *D. laevis* in the sites used for phylogenetic analysis. Secondly, many of the unique substitutions in *D. laevis* were also observed in the sister species *D. dubia* (fig. 6A), suggesting that the changes in this lineage are shared and derived rather than artifacts.

Could it be that the LSU tree is correct and the concordance tree incorrect? More genetic characters are needed for a definitive answer to this question. Both morphology and gene sequence (particularly 16S rDNA and HSP90) support this clade. *Daphnia laevis* and *D. dentifera* are nearly morphologically indistinguishable, and for many decades they were considered to be the same species. Furthermore, the independent reference trees lack the long branches (i.e., they do not violate the molecular clock assumption) that apparently mislead the

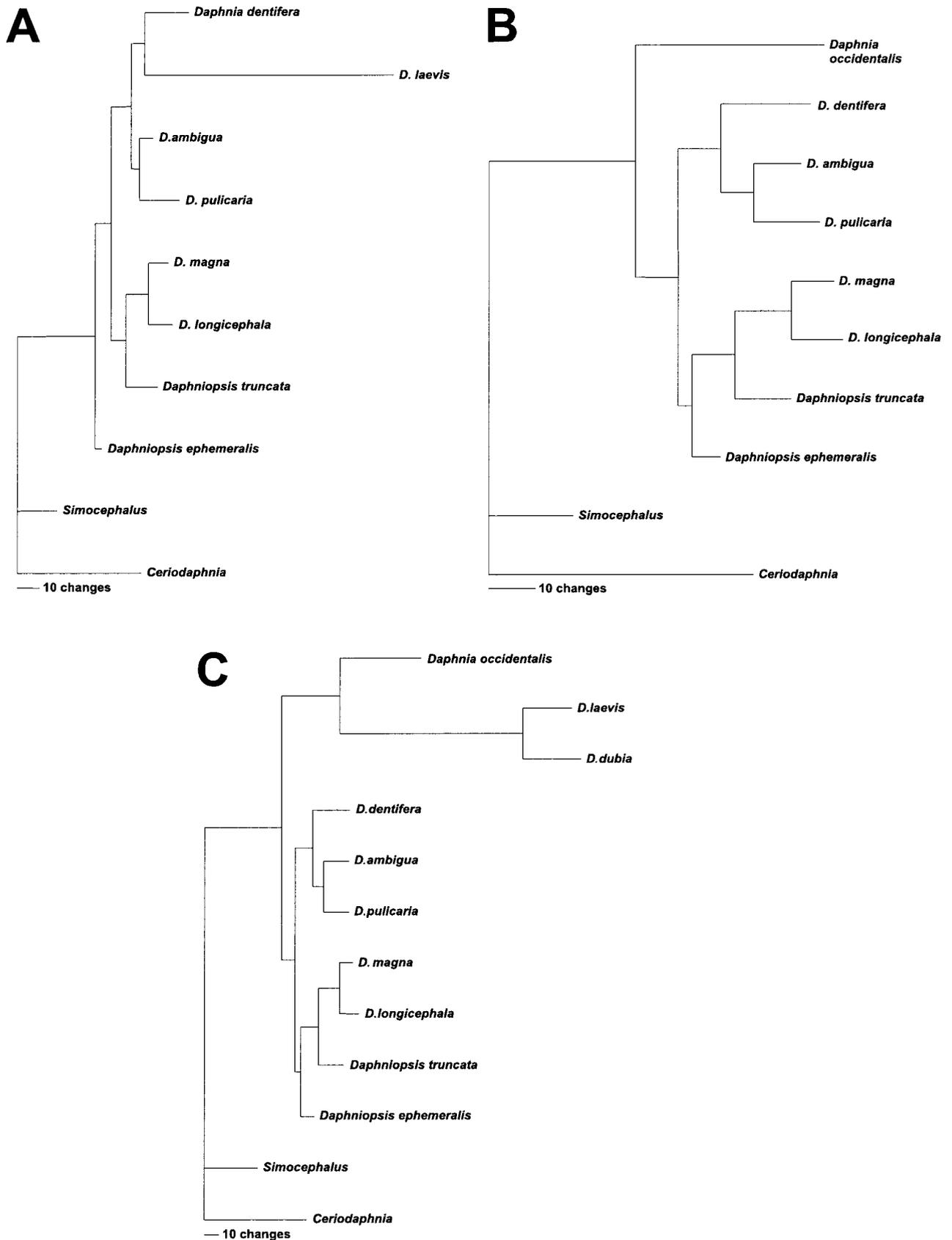


FIG. 5.—Best trees resulting from the removal or addition of taxa. A, MP phylogram (one best tree of 341 steps with gaps coded as missing) found after removing long-branched *D. occidentalis* from the analysis. Note that in contrast to the inferred tree with all taxa (fig. 2), *D. laevis* is now placed with the expected sister taxon, *D. dentifera*. B, MP phylogram (one best tree of 275 steps with gaps coded as missing) found

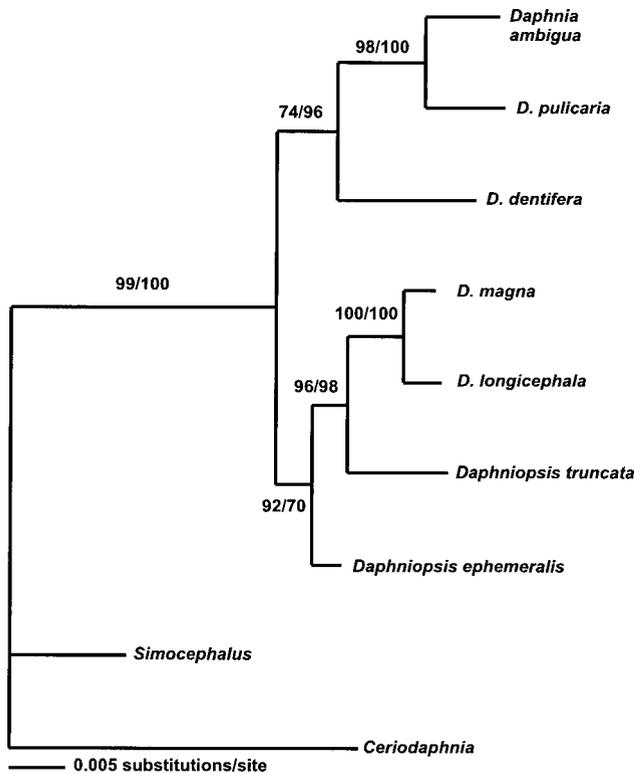


FIG. 6.—Phylogram of daphniid sequences (with branch lengths drawn to reflect the amount of evolution) based on the likelihood tree. Bootstrap values are given for ML and MP optimality criteria. Putative long-branch attractors (*D. occidentalis* and *D. laevis*) are excluded from this analysis.

nuclear LSU analysis (e.g., Taylor, Finston, and Hebert 1998). The available independent data indicate that the *D. laevis*/*D. dentifera* clade is robust.

Complex ML models have been proposed as a remedy for LBA, but this approach recovered the same LBA clade as that recovered by the MP analysis for daphniid LSU data. Nevertheless, parametric bootstrapping of the data revealed that ML did recover the correct tree for the majority of replicates, whereas MP recovered the incorrect LBA clade for the majority of replicates. This finding supports the notion that ML is much less sensitive to LBA than MP (Huelsenbeck 1997). However, the differing results of ML under parametric and nonparametric bootstrapping suggest that the model used in the nonparametric analysis is inadequate or that too few characters have been sampled. Because the nuclear LSU has a very complex mode of evolution, the TrN + I + Γ model may be inappropriate even though model testing determined that it had the best fit to the data. For example, the known aspects of rRNA secondary structure coevolution are unaccounted for in the TrN + I + Γ model. The Akaike information criterion (Hasegawa

1990), an alternative way of comparing different models of DNA sequence evolution, indicated that a more complex GTR + I + Γ model was optimal, but the tree reconstructed with the use of this model also recovered the *D. laevis*/*D. occidentalis* clade (not shown).

The use of a conserved gene and the addition of taxa also failed in this case to stave off LBA. It seems that the acceleration in LSU rates occurred before the radiation of the group that can act as a source of additional taxa (such as *D. dubia*). We know of no other extant species whose addition might break up the long *D. laevis*/*D. dubia* or *D. laevis*/*D. occidentalis* branch. Inspection of the phylogram (fig. 5C) reveals that further taxon addition will most likely fail to break up the *D. laevis* branch. The long branch leading to the *D. laevis* clade is so long that other *Daphnia* species are more distant from *D. laevis* than from the outgroup genera *Simocephalus* and *Ceriodaphnia*.

Often a rogue lineage is associated with some aspect of unusual habitat, life history, or molecular evolution that might increase the substitution rate (Nunn et al. 1996; Stiller and Hall 1999). Nunn et al. (1996), for example, found that the large size of the LSU V3 domain in isopod crustaceans is associated with terrestrial habitats. They proposed that the rapid evolution and large size in terrestrial isopods might be a response to the increased thermal stress of terrestrial life. At present, we can identify nothing exceptional about the biology of the *D. laevis* lineage compared with that of the other daphniid taxa examined. The habitat of *D. laevis* is temperate freshwater ponds and lakes—less extreme than the snowmelt ponds of the coldwater stenotherm *D. ephemeralis*, or the Australian saline habitat of *D. truncata*. There is nothing exceptional about the life history of *D. laevis* compared with that of *D. pulex* (Banta et al. 1939). Also, there are no signatures of base compositional bias that might affect substitution rates as in the dipterans. A positive association of gene size and rate of evolution in rDNA genes that appears to occur in protists (Stiller and Hall 1999) does not exist in daphniids—*D. laevis* has an unexceptional gene size (table 5). The factors that lead to substitutional rate acceleration in the *D. laevis* group are elusive.

Although *D. laevis* has an unexceptional LSU gene size for daphniids, it is clear that daphniids themselves possess unusually large nuclear LSU rDNA genes. At approximately 4.5 kb (about 84 bp are missing from our reported sequence), the *D. pulicaria* LSU sequence is smaller than the largest-reported LSU gene, 5.2 kb in the hagfish, but larger than the LSU of most animals and plants (Kuzoff et al. 1998; Mallat and Sullivan 1998). Crease and Taylor (1998) found that the size of the V2 (also known as D2) expansion segment from the LSU and the small subunit (SSU) are highly correlated

after removing long-branched *D. laevis* from the analysis. *Daphnia occidentalis* is now placed in the expected position basal to the other *Daphnia*/*Daphniopsis*. C, MP phylogram found after adding *D. dubia* to the analysis. *Daphnia dubia* is known from independent evidence (Taylor, Finston, and Hebert 1998) to be closely related to one of the two taxa involved in the putative long-branched attraction, *D. laevis*. One of the two best trees is shown of 612 steps with gaps coded as a fifth base. One tree of 471 was found with gaps coded as missing. Note that this taxon addition failed to break up the putative long-branch association of *D. occidentalis* with *D. laevis*.

in branchiopod crustaceans. When combined with the SSU, intergenic, and internal transcribed spacer data (Crease 1993; Crease and Taylor 1998; unpublished data), our LSU results make it clear that each gene region in the daphniid nuclear rDNA gene family is exceptionally large in length when compared with that of other animals. As daphniids have the smallest genomes of all the crustaceans examined (Lecher, Defaye, and Noel 1995), our results bolster the hypothesis that rDNA variable regions are uncoupled from the factors that regulate genome size evolution (Crease and Taylor 1998).

Slippage and gene family-wide base composition biases are consistent with variable region coevolution but fail to explain the existence of rogue variable regions that are markedly expanded compared with neighboring variable regions in an array. Our results have identified the putative hidden break region or V6 (table 4) as a rogue expansion segment in daphniids. In an extensive comparison of this region with 29 taxa from 12 phyla, Chenuil, Solignac, and Bernard (1997) found very little size variation in the helix that is expanded in daphniids. In their study, *Drosophila* had the longest stem at 44 bp. A recent study of vertebrate LSUs reported one chordate taxon with an expanded V6, *Branchiostoma floridae* (lancelet), which had a size of 180 bp (Mallat and Sullivan 1998). So, with a broad range of 90–351 bp, daphniids possess the largest-reported V6 regions and show more size variation in this region than is found throughout the rest of the reported metazoan phyla. In some animals, the V6 region, which contains the recognition site for the L25 protein (Chenuil, Solignac, and Bernard 1997), is partially or entirely deleted by processing (Ware, Renkawitz, and Gerbi 1985). There is no evidence that the largest V6 regions (from *D. pulex*) undergo any size reduction during processing (Taylor, Omilian, and Swain, unpublished data). Because daphniid V6 regions show a broad range of size from just above average size to the largest yet recorded, their study could provide insights into the evolution of rogue rDNA expansion segments. Such studies are important in the light of proposals to code rDNA size variation and other correlated expansion segment features for phylogenetic studies (Billoud et al. 2000).

Despite several sources of error and potential bias, the nuclear LSU rDNA is useful for reconstructing evolutionary relationships among daphniids that lack strong among-lineage rate heterogeneity. In the present case, rapidly evolving taxa misled the analysis in two ways: LBA involving *D. laevis* and *D. occidentalis*, and a reduction in alignable informative sites that affected the position of *D. ephemeralis*. With these rapidly evolving taxa removed, daphniid phylogenies constructed from nuclear LSU sequence data are well resolved, well supported, and concordant with phylogenies constructed from independent data. Of particular note is a strongly supported and paraphyletic genus *Daphnia*. Also, there is strong support for the following clades: (*D. pulex*/*D. ambigua*), (*Daphniopsis*/*Ctenodaphnia*), (*D. magna*/*D. longicephala*), and the traditional subgenus *Daphnia*. Although many of the expansion segments proved unreliable in the alignment, they may contain further

phylogenetically informative secondary structural information.

Our first comparison of nearly complete LSU sequences from nonchordate animals has revealed a misleading association involving morphologically cryptic species and a conserved gene. Each optimality criterion is misled by this pattern of evolution, but ML is markedly less sensitive to the observed LBA bias than MP. Unlike most existing case studies, this inconsistency seems caused by substitutional rate acceleration at variable sites rather than base compositional, covariotide, or taxonomic sampling biases. This case may provide a simple empirical case to study LBA remedies because there has been insufficient time for multiple interacting biases to evolve. Alternatively, further taxonomic sampling will reveal additional rogue lineages that possess differing or multiple biases in daphniid LSU sequences. The results suggest that current models of evolution are inadequate for some rDNA genes and that rogue taxa are difficult to predict on the basis of morphological divergence, genome type, and gene sequence conservation. Given this situation, the best guard against inconsistency seems to be the comparison of numerous independent genes.

Acknowledgments

We sincerely thank Angela Holliss for technical assistance, and Paul Hebert and Peter Lockhart for helpful advice. Chris Wilson and Paul Hebert provided specimens from Australia. Jim Stamos provided assistance with the tables and figures. This research was supported by NSF Grant OPP9984901 awarded to D.J.T.

LITERATURE CITED

- BANTA A. M., T. R. WOOD, L. A. BROWN, and L. INGLE. 1939. Studies on the physiology, genetics, and evolution of some Cladocera. Carnegie Institute of Washington, Washington, D.C.
- BILLOU, B., M. A. GUERRUCCI, M. MASSELOT, and J. S. DEUTSCH. 2000. Cirripede phylogeny using a novel approach: molecular morphometrics. *Mol. Biol. Evol.* **17**: 1435–1445.
- BROOKS, J. L. 1957. The systematics of North American *Daphnia*. *Mem. Conn. Acad. Arts Sci.* **13**:1–180.
- BULT, C. J., J. A. SWEERE, and E. A. ZIMMER. 1995. Cryptic sequence simplicity, nucleotide composition bias, and molecular coevolution in the large subunit of ribosomal DNA in plants: implications for phylogenetic analyses. *Ann. Mo. Bot. Gard.* **82**:235–246.
- CHANG, J. T. 1996. Inconsistency of evolutionary tree topology reconstruction methods when substitution rates vary across characters. *Math. Biosci.* **134**:189–215.
- CHENUIL, A., M. SOLIGNAC, and M. BERNARD. 1997. Evolution of the large-subunit ribosomal RNA binding site for protein L23/25. *Mol. Biol. Evol.* **14**:578–588.
- COLBOURNE, J. K., and P. D. N. HEBERT. 1996. The systematics of North American *Daphnia* (Crustacea: Anomopoda): a molecular phylogenetic approach. *Philos. Trans. R. Soc. Lond., Ser. B Biol. Sci.* **351**:349–360.
- CREASE, T. J. 1993. Sequence of the intergenic space between the 28S and 18S rRNA-encoding genes of the crustacean, *Daphnia pulex*. *Gene* **134**:245–249.

- CREASE, T. J., and D. J. TAYLOR. 1998. The origin and evolution of variable-region helices in V4 and V7 of the small-subunit ribosomal RNA of branchiopod crustaceans. *Mol. Biol. Evol.* **15**:1430–1446.
- CUNNINGHAM, C. W., H. ZHU, and D. M. HILLIS. 1998. Best-fit maximum-likelihood models for phylogenetic inference: empirical tests with known phylogenies. *Evolution* **52**:978–987.
- DE RIJK, P., J. WUYTS, Y. VAN DE PEER, T. WINKELMANS, and R. DE WACHTER. 2000. The European large subunit ribosomal RNA database. *Nucleic Acids Res.* **28**:177–178.
- DELARBRE, C., H. ESCRIVA, C. GALLUT, V. BARRIEL, P. KOURILSKY, P. JANVIER, V. LAUDET, and G. GACHELIN. 2000. The complete nucleotide sequence of the mitochondrial DNA of the agnathan *Lampetra fluviatilis*: bearings on the phylogeny of cyclostomes. *Mol. Biol. Evol.* **17**:519–529.
- DOYLE, J. J., and J. L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19**:11–15.
- FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27**:401–410.
- FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**:368–376.
- FRIEDRICH, M., and D. TAUTZ. 1997. An episodic change of rDNA nucleotide substitution rate has occurred during the emergence of the insect order Diptera. *Mol. Biol. Evol.* **14**:644–653.
- GAUT, B. S., and P. O. LEWIS. 1995. Success of maximum-likelihood phylogeny inference in the 4-taxon case. *Mol. Biol. Evol.* **12**:152–162.
- GOLDMAN, N., J. P. ANDERSON, and A. G. RODRIGO. 2000. Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.* **49**:652–670.
- HALL, T. 1999. Bioedit sequence alignment editor. Department of Microbiology, North Carolina State University.
- HANCOCK, J. M., and G. A. DOVER. 1988. Molecular coevolution among cryptically simple expansion segments of eukaryotic 26S/28S rRNAs. *Mol. Biol. Evol.* **5**:377–391.
- HASEGAWA, M. 1990. Phylogeny and molecular evolution in primates. *Jpn. J. Genet.* **65**:243–265.
- HASEGAWA, M., H. KISHINO, and T. YANO. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **32**:443–445.
- HENDY, M. D., and D. PENNY. 1989. A framework for the quantitative study of evolutionary trees. *Syst. Zool.* **38**:297–309.
- HILLIS, D. M., J. J. BULL, M. E. WHITE, M. R. BADGETT, and I. J. MOLINEUX. 1992. Experimental phylogenetics: generation of a known phylogeny. *Science* **255**:589–592.
- HILLIS, D. M., and M. T. DIXON. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. *Q. Rev. Biol.* **66**:411–453.
- HILLIS, D. M., and J. P. HUELSENBECK. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. *J. Hered.* **83**:189–195.
- HILLIS, D. M., C. MORITZ, and B. K. MABLE. 1996. *Molecular systematics*. 2nd edition. Sinauer Associates, Sunderland, Mass.
- HUELSENBECK, J. P. 1995. Performance of phylogenetic methods in simulation. *Syst. Biol.* **44**:17–48.
- . 1997. Is the Felsenstein zone a fly trap? *Syst. Biol.* **46**:69–74.
- . 1998. Systematic bias in phylogenetic analysis: is the Strepsiptera problem solved? *Syst. Biol.* **47**:519–537.
- HUELSENBECK, J. P., and K. A. CRANDALL. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. *Annu. Rev. Ecol. Syst.* **28**:437–466.
- JUKES, T. H., and C. R. CANTOR. 1969. *Mammalian protein metabolism*. Academic Press, New York.
- KURAKU, S., D. HOSHIYAMA, K. KATOH, H. SUGA, and T. MIYATA. 1999. Monophyly of lampreys and hagfishes supported by nuclear DNA-coded genes. *J. Mol. Evol.* **49**:729–735.
- 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.
- LECHER, P., D. DEFAYE, and P. NOEL. 1995. Chromosomes and nuclear-Dna of Crustacea. *Invertebr. Reprod. Dev.* **27**:85–114.
- LEHMAN, N., M. E. PFRENDER, P. A. MORIN, T. J. CREASE, and M. LYNCH. 1995. A hierarchical molecular phylogeny within the genus *Daphnia*. *Mol. Phylogenet. Evol.* **4**:395–407.
- LITTLEWOOD, D. T. 1994. Molecular phylogenetics of cupped oysters based on partial 28S rRNA gene sequences. *Mol. Phylogenet. Evol.* **3**:221–229.
- LOCKHART, P. J., M. A. STEEL, A. C. BARBROOK, D. H. HUSON, M. A. CHARLESTON, and C. J. HOWE. 1998. A covariate model explains apparent phylogenetic structure of oxygenic photosynthetic lineages. *Mol. Biol. Evol.* **15**:1183–1188.
- MALLAT, J., and J. SULLIVAN. 1998. 28S and 18S rDNA sequences support the monophyly of lampreys and hagfishes. *Mol. Biol. Evol.* **15**:1706–1718.
- MUGRIDGE, N. B., D. A. MORRISON, T. JAKEL, A. R. HECKEROTH, A. M. TENTER, and A. M. JOHNSON. 2000. Effects of sequence alignment and structural domains of ribosomal DNA on phylogeny reconstruction for the protozoan family Sarcocystidae. *Mol. Biol. Evol.* **17**:1842–1853.
- NUNN, G. B., B. F. THEISEN, B. CHRISTENSEN, and P. ARCTANDER. 1996. Simplicity-correlated size growth of the nuclear 28S ribosomal RNA D3 expansion segment in the crustacean order Isopoda. *J. Mol. Evol.* **42**:211–223.
- OLESEN, J. 1998. A phylogenetic analysis of the Conchostraca and Cladocera (Crustacea, Branchiopoda, Diplostraca). *Zool. J. Linn. Soc.* **122**:491–536.
- POSADA, D., and K. A. CRANDALL. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**:817–818.
- RAMBAUT, A., and N. C. GRASSLY. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput. Appl. Biosci.* **13**:235–238.
- RODRIGUEZ, F., J. L. OLIVER, A. MARÍN, and J. R. MEDINA. 1990. The general stochastic model of nucleotide substitution. *J. Theor. Biol.* **142**:485–501.
- SANDERSON, M. J., and J. KIM. 2000. Parametric phylogenetics? *Syst. Biol.* **49**:817–829.
- SANDERSON, M. J., M. F. WOJCIECHOWSKI, J. M. HU, T. S. KHAN, and S. G. BRADY. 2000. Error, bias, and long-branch attraction in data for two chloroplast photosystem genes in seed plants. *Mol. Biol. Evol.* **17**:782–797.
- SHIMODAIRA, H., and M. HASEGAWA. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**:1,114–1,116.
- STEEL, M., D. HUSON, and P. J. LOCKHART. 2000. Invariable sites models and their use in phylogeny reconstruction. *Syst. Biol.* **49**:225–232.
- STILLER, J. W., and B. D. HALL. 1999. Long-branch attraction and the rDNA model of early eukaryotic evolution. *Mol. Biol. Evol.* **16**:1270–1279.

- SWOFFORD, D. L. 2000. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0b8. Sinauer Associates, Sunderland, Mass.
- TAMURA, K., and M. NEI. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**:512–526.
- TAUTZ, D., M. TRICK, and G. A. DOVER. 1986. Cryptic sequence simplicity in DNA is a major source of genetic variation. *Nature* **322**:652–656.
- TAYLOR, D. J., T. J. CREASE, and W. M. BROWN. 1999. Phylogenetic evidence for a single long-lived clade of crustacean cyclic parthenogens and its implications for the evolution of sex. *Proc. R. Soc. Lond., Ser. B Biol. Sci.* **266**:791–797.
- TAYLOR, D. J., T. L. FINSTON, and P. D. N. HEBERT. 1998. Biogeography of a widespread freshwater crustacean: Pseudocongruence and cryptic endemism in the North American *Daphnia laevis* complex. *Evolution* **52**:1648–1670.
- TAYLOR, D. J., P. D. N. HEBERT, and J. K. COLBOURNE. 1996. Phylogenetics and evolution of the *Daphnia longispina* group (Crustacea) based on 12S rDNA sequence and allozyme variation. *Mol. Phylogenet. Evol.* **5**:495–510.
- THOMPSON, J. D., T. J. GIBSON, F. PLEWNIK, F. JEANMOUGIN, and D. G. HIGGINS. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**:4876–4882.
- THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. Clustal-W—improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
- WARE, V. C., R. RENKAWITZ, and S. A. GERBI. 1985. rRNA processing: removal of only nineteen bases at the gap between 28S alpha and 38S beta rRNA in *Sciara coprophila*. *Nucleic Acids Res.* **13**:3581–3597.
- WEINS, J. J., and B. D. HOLLINGSWORTH. 2000. War of the iguanas: conflicting molecular and morphological phylogenies and long-branch attraction in Iguanid lizards. *Syst. Biol.* **49**:143–159.
- WHITING, M. F. 1998. Phylogenetic position of the Strepsiptera: review of molecular and morphological evidence. *Int. J. Insect Morphol. Embryol.* **27**:53–60.

ROSS CROZIER, reviewing editor

Accepted August 13, 2001