

First Sequenced Mitochondrial Genome from the Phylum Acanthocephala (*Leptorhynchoides thecatus*) and Its Phylogenetic Position Within Metazoa

Michelle L. Steinauer,¹ Brent B. Nickol,² Richard Broughton,³ Guillermo Orti²

¹ Department of Biology, University of New Mexico, MSC 03 2020, 1 University of New Mexico, Albuquerque, NM 87131-0001, USA

² School of Biological Sciences, University of Nebraska—Lincoln, Lincoln, NE 68588-0118, USA

³ Oklahoma Biological Survey and Department of Zoology, University of Oklahoma, Norman, OK 73019, USA

Received: 25 May 2004 / Accepted: 2 February 2005 [Reviewing Editor: Dr. Rafael Zardoya]

Abstract. The complete sequence of the mitochondrial genome of *Leptorhynchoides thecatus* (Acanthocephala) was determined, and a phylogenetic analysis was carried out to determine its placement within Metazoa. The genome is circular, 13,888 bp, and contains at least 36 of the 37 genes typically found in animal mitochondrial genomes. The genes for the large and small ribosomal RNA subunits are shorter than those of most metazoans, and the structures of most of the tRNA genes are atypical. There are two significant noncoding regions (377 and 294 bp), which are the best candidates for a control region; however, these regions do not appear similar to any of the control regions of other animals studied to date. The amino acid and nucleotide sequences of the protein coding genes of *L. thecatus* and 25 other metazoan taxa were used in both maximum likelihood and maximum parsimony phylogenetic analyses. Results indicate that among taxa with available mitochondrial genome sequences, Platyhelminthes is the closest relative to *L. thecatus*, which together are the sister taxon of Nematoda; however, long branches and/or base composition bias could be responsible for this result. The monophyly of Ecdysozoa, molting organisms, was not supported by any of the analyses. This study represents the first mitochondrial genome of an acanthocephalan to be sequenced and will allow further studies of systematics, population genetics, and genome evolution.

Key words: Acanthocephala — *Leptorhynchoides thecatus* — Mitogenomics — Metazoa — Phylogeny — Parasite

Introduction

In animals, mitochondrial genomes are typically small (15–20 kb), are circular, and encode 37 genes: 2 ribosomal RNAs, 22 tRNAs, and 13 protein coding genes (Wolstenholme 1992; Boore 1999). There are both a large and a small ribosomal RNA (*rrnL* and *rrnS*), and one tRNA for every amino acid except leucine and serine, which have two genes. The protein coding genes consist of ATP synthase subunits 6 and 8 (*atp6* and *atp8*), cytochrome oxidase subunits 1–3 (*cox1-cox3*), cytochrome B (*cob*), and NADH dehydrogenase subunits 1–6 and 4L (*nad1-6* and *nad4L*). Typically, animal mitochondrial genomes also have a large noncoding region that contains elements that control the replication and transcription of the genome (Wolstenholme 1992). With a few exceptions, the gene content for animal mitochondrial genomes is well conserved (Boore 1999), but gene order is more variable (Boore and Brown 1998; Moret et al. 2001).

Complete sequences of mitochondrial genomes are potentially useful for phylogenetic analysis. Gene order, DNA, and amino acid sequences can be used for phylogenetic reconstruction. Mitochondrial genome sequences are becoming prevalent in GenBank;

The entire genome sequence has been deposited with the GenBank Data Libraries under-accession number AY562383.

Correspondence to: Michelle L. Steinauer; email: mls1@unm.edu

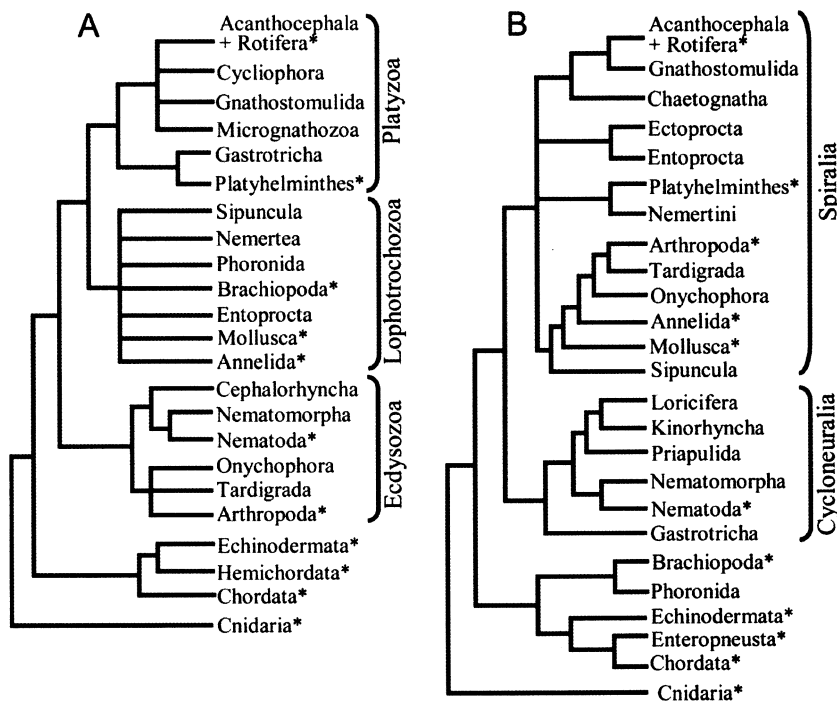


Fig. 1. Phylogenetic hypotheses of metazoan relationships. Asterisks by taxon labels indicate those included in the phylogenetic analysis in this study. **A** Redrawn from Giribet (2002). **B** Redrawn from Nielsen (2001).

however, many groups of organisms are unrepresented, resulting in gaps in taxon sampling for phylogenetic analysis, and gaps in our knowledge of mitochondrial genomes. This presents difficulties in the development of molecular markers for phylogenetic analysis.

Acanthocephala is a group of parasitic organisms with ambiguous phylogenetic status. Phylogenies based on both molecular and morphological data have shown that Acanthocephala is closely affiliated with Rotifera either as sister taxa or as a subtaxon (Winnepenninckx et al. 1995; Abouheif et al. 1998; Zrzavy et al. 1998; Garey et al. 1998; Herlyn et al. 2003). The placement of rotifers and acanthocephalans together in Protostomia differs among hypotheses (Fig. 1). These taxa have been placed in Platyzoa, which includes Cycliophora, Gnathostomulida, Gastrotricha, and Platyhelminthes, and is a sister taxon to Lophotrochozoa, which includes Mollusca, Annelida, Brachiopoda, Phoronida, Entoprocta, Nemertea, and Sipuncula (Giribet et al. 2000; Giribet 2002) (Fig. 1A). A second hypothesis based on morphological data groups Acanthocephala, Rotifera, Gnathostomulida, and Chaetognatha and places them in Spiralia, a large group that includes such taxa as Platyhelminthes, Mollusca, Annelida, Arthropoda, and Entoprocta (Nielsen 2001) (Fig. 1B).

To date, a mitochondrial genome has not been sequenced for any species of Acanthocephala. Platyhelminths are likely the most closely related organisms for which complete mitochondrial genomes have been sequenced; however, they are distantly related according to most phylogenetic hypotheses (Fig. 1). We determined the sequence and structure of the

mitochondrial genome of an acanthocephalan species, *Leptorhynchoides thecatus*, and performed a phylogenetic analysis of Metazoa using amino acid sequences of the protein coding mitochondrial genes. Although many critical taxa still have not been sequenced, comparisons of mitochondrial genomes among lower Metazoa provide a new perspective on the phylogenetic placement of these rarely studied organisms. This study represents the first mitochondrial genome of an acanthocephalan to be sequenced and will allow further studies of systematics, population genetics, and genome evolution.

Materials and Methods

Specimens of *Leptorhynchoides thecatus* were collected from the intestines of fish from Atkinson Reservoir in Holt County, Nebraska (42°32'36"N, 98°58'22"W). Worms were preserved in 95% ethanol and kept at -20°C prior to extraction of DNA. Total DNA was extracted from multiple worms by proteinase K digestion and phenol/chloroform extraction (Sambrook et al. 1989) and concentrated by ethanol precipitation. Initially, 340 bp of the *cox1* gene was amplified with primers from Bowles et al. (1992) using polymerase chain reaction (PCR). This sequence was used to design primers specifically oriented to amplify the remainder of the genome in one fragment. This fragment was amplified by long PCR with the GeneAmp PCR kit (Applied Biosystems) using the following protocol: 94°C for 1 min, 15 cycles at 94°C for 15 s and 60°C for 10 min, and 15 cycles at 94°C for 15 s and 60°C for 10 min, with an extension of 15 s for each cycle. Primers used were as follows: forward, 5'-TTG GTG GGT TGA CTG GTG TAA TTA TCT CTA A-3'; and reverse, 5'-GAG CTC ACA CTA AAC TAC CTA AAA CCC CAA T-3'. The PCR product was purified using the Qiax II Gel Extraction Kit (Qiagen). Approximately 1 µg of amplified mtDNA was sheared randomly into smaller fragments via sonication. These fragments were made blunt-ended with an end-repair reaction containing 1 unit Klenow DNA polymerase, 1

unit T4 DNAPolymerase, and 2.5 units T4 polynucleotide kinase, in the presence of dNTPs and ATP. The entire reaction mixture was electrophoresed in 1% agarose. Fragments in the range 750 bp to 4 kb were cut out of the gel, and fragments were recovered from the gel block with the MinElute gel purification kit (Qiagen). Purified fragments were ligated into *Sma*I-cut phosphatased pGEM-3Z vector with the LigaFast ligation system (Promega). Recombinant plasmids were transformed into competent *Escherichia coli* JM109 cells, and colonies were screened via blue–white color selection. White (positive) colonies were transferred with sterile toothpicks into individual wells of 96-well blocks containing 1.2 ml LB medium. Two blocks of cultures (= 192 clones) were grown overnight, and plasmids were isolated with the Montage 96 plasmid purification kit (Millipore). Plasmids were sequenced with M13 primers using the BigDye terminator kit (Applied Biosystems) and a BaseStation automated sequencer (MJ Research). Sequences were assembled with Sequencher™ (Gene Codes Corporation). To ensure that proper assembly of the sequences was achieved, the resulting genome sequence was used to design eight sets of primers to be used in PCR. The primer sets covered the entire mitochondrial genome except a 184-bp region encompassing the end of *nad5* and the beginning of *cob*. The primer sets were used for PCR of genomic DNA, and the resulting fragments were visualized on a 0.8% agarose gel and compared to standards. The entire genome sequence has been deposited with the GenBank Data Libraries under accession number AY562383.

Protein coding genes were identified by finding open reading frames and performing BLAST searches on the amino acid sequences (Altschul et al. 1997). To confirm their identity, hydrophobic profile comparisons using MacVector 7.1 (Accelrys Inc.) and amino acid sequence comparisons were performed with organisms of the following species: *Hymenolepis diminuta*, *Ascaris suum*, *Trichinella spiralis*, *Terebratalia transversa*, and *Triops cancriformis* (Table 1). Following Helfenbein et al. (2001), if the 5' end of the protein coding genes could not be determined through comparisons with other species, it was inferred to occur at the first possible start codon with no overlap with the preceding gene with the exception of overlap with the 3' end of an upstream tRNA. Genes were assumed to terminate early with an abbreviated stop codon if they overlapped with a downstream gene. Ribosomal genes were located by performing BLAST searches and the beginning and ends were inferred to be located at the end of the consecutive genes. The tRNAs were identified using the program tRNAscan-SE (Lowe and Eddy 1997) or through recognition by eye of potential secondary structures surrounding possible anticodon consensus sequences (NTXXXRN; XXX = anticodon). If no suitable tRNAs were discovered, then the restrictions on the consensus sequence were relaxed and the search was repeated. Folding of the tRNAs was accomplished by eye or as suggested by tRNAscan-SE. The complete mitochondrial sequence was searched for tandem repeats using Tandem Repeats Finder (Benson 1999; <http://tandem.bu.edu/trf/trf.html>).

Phylogenetic analyses using maximum likelihood (ML) and maximum parsimony (MP) optimality criteria were performed on amino acid sequences for all protein coding genes except *nad4L*, *atp8*, and *atp6* (due to their lack of conservation). A broad representation of metazoan taxa with complete mitochondrial genomes available was chosen, and sequences were obtained from GenBank (Table 1). Sequences were aligned using ClustalX with the following parameters: gap opening = 2, gap extension = 0.5, and Gonnet Series protein weight matrix. The alignments of the genes were concatenated and areas of ambiguous alignment were removed so that the final alignment consisted of 1523 characters. The aligned amino acid sequences were then converted back into their original nucleotide sequences so that phylogenetic analyses could be performed on the DNA sequences as well as the protein sequences. The third codon position was not utilized in any of the analyses due to the high probability of substitution saturation at these sites.

Table 1. GenBank accession numbers of sequences used in this study

Taxon	Group	Accession No.
<i>Ascaris suum</i>	Nematoda	X54253
<i>Necator americanus</i>	Nematoda	AJ417719
<i>Caenorhabditis elegans</i>	Nematoda	X54252
<i>Trichinella spiralis</i>	Nematoda	AF293969
<i>Hymenolepis diminuta</i>	Platyhelminthes	AF314223
<i>Fasciola hepatica</i>	Platyhelminthes	AF216697
<i>Echinococcus multilocularis</i>	Platyhelminthes	AB018440
<i>Schistosoma japonicum</i>	Platyhelminthes	AF215860
<i>Triops cancriformis</i>	Arthropoda	AB084514
<i>Drosophila melanogaster</i>	Arthropoda	U37541
<i>Thyropygus sp.</i>	Arthropoda	AY055728
<i>Terebratalia transversa</i>	Brachiopoda	AF331161
<i>Lumbricus terrestris</i>	Annelida	U24570
<i>Platynereis dumerilii</i>	Annelida	AF178678
<i>Loligo bleekeri</i>	Mollusca	AB029616
<i>Pupa strigosa</i>	Mollusca	AB028237
<i>Lampsilis ornata</i>	Mollusca	AY365193
<i>Albinaria coerulea</i>	Mollusca	X83390
<i>Katharina tunicata</i>	Mollusca	U09810
<i>Asterina pectinifera</i>	Echinodermata	D16387
<i>Arbacia lixula</i>	Echinodermata	X80396
<i>Florometra serratissima</i>	Echinodermata	AF049132
<i>Balanoglossus carnosus</i>	Hemichordata	AF051097
<i>Danio rerio</i>	Chordata	AC024175
<i>Acropora tenuis</i>	Cnidaria	AF338425

Maximum parsimony tree searches of the DNA and amino acid sequences were performed heuristically using PAUP* 4.0b10 (Swofford 2001) with tree bisection–reconnection (TBR) branch swapping and random stepwise addition of taxa replicated 1000 times. Analysis of the DNA sequences was unweighted and the amino acid analysis employed the PROTPARS substitution matrix for mitochondrial DNA given in the PAUP* example files. This matrix gives the minimum number of amino acid replacement substitutions needed to convert one amino acid to another, using the mammalian mitochondrial code. Support for nodes was assessed using bootstrap analysis (Felsenstein 1985) using a heuristic search with TBR branch swapping and random stepwise addition of taxa with 100 pseudoreplicates.

Phylogenetic analysis using ML optimality criteria on nucleotide sequences was performed using the GTR + Γ model, which was determined as the best fit model for the data by the likelihood ratio test implemented in MODELTEST 3.0 (Posada and Crandall 1998). ML analysis was carried out heuristically with PAUP* 4.0b10 (Swafford 2001) with 100 replications of a random stepwise addition of taxa and TBR branch swapping. Bootstrap analysis, executed as described earlier, was used to assess node support. Maximum likelihood tree searches based on the amino acid data were performed using Tree-Puzzle 5.0 (Strimmer and von Haeseler 1996) selecting parameter settings corresponding to the mtREV24 (Adachi and Hasegawa 1996) model of amino acid substitution and using 10,000 puzzling steps. Among-site rate variation was estimated with a gamma distribution, with sites grouped into eight different categories, and the shape parameter α estimated from the data by Tree-Puzzle ($\alpha = 0.66$).

A topology test was used to compare the topology of the ML DNA tree to both hypotheses A and B from Fig. 1. The topologies of hypotheses A and B were obtained by performing constrained ML tree searches to find the most likely tree given the hypotheses and the data. A Shimodaira–Hasegawa (SH) (1999) test was performed on the three trees as implemented in PAUP* 4.0b10

Table 2. Inferred organization of the mitochondrial genome of *Leptorhynchoides thecatus*: Shaded boxes indicate putative genes.

Gene	Position	Strand	bp	Start	Stop	3' spacer
<i>cox1</i>	1–1528	+	1528	GTG	T	0
<i>TrnG</i>	1529–1582	+	54	–	–	41
<i>TrnQ</i>	1566–1632	+	104	–	–	–16
<i>TrnY</i>	1624–1676	+	53	–	–	–8
<i>RrnL</i>	1677–2602	+	925	–	–	0
<i>TrnL_(GAG)</i>	2603–2657	+	55	–	–	0
<i>Nad6</i>	2658–3095	+	438	ATG	TAA	2
<i>TrnD</i>	3098–3151	+	54	–	–	104
<i>atp8</i>	3152–3262	+	111	GTG	TAA	
<i>trnS_(ACU)</i>	3163–3226	–	64	–	–	
<i>atp6</i>	3256–3816	+	561	ATG	TAA	1
<i>nad3</i>	3818–4159	+	342	TTG	TAA	0
<i>trnW</i>	4160–4219	+	60	–	–	0
Noncoding region 1	4213–4596		377	–	–	0
<i>trnK</i>	4251–4303	+	53	–	–	
<i>atp8</i>	4314–4466	+	153	ATA	TAG	
<i>trnV</i>	4597–4652	+	56	–	–	6
<i>trnK</i>	4659–4711	+	53	–	–	0
<i>trnE</i>	4712–4763	+	52	–	–	0
<i>trnT</i>	4764–4820	+	57	–	–	61
<i>nad4L</i>	4881–5132	+	252	ATA	TAG	0
<i>nad4</i>	5133–6396	+	1264	ATG	T	0
<i>trnH</i>	6397–6450	+	54	–	–	0
<i>nad5</i>	6451–8092	+	1642	ATG	T	0
<i>trnL_(UAA)</i>	8093–8147	+	55	–	–	6
<i>trnP</i>	8153–8205	+	53	–	–	0
<i>cob</i>	8206–9322	+	1117	ATA	T	–1
<i>nad1</i>	9320–10235	+	916	ATT	T	0
<i>trnI</i>	10236–10295	+	60	–	–	0
Noncoding region 2	10296–10589		294	–	–	0
<i>trnS_(UCU)</i>	10481–10542	–	62	–	–	
<i>trnM</i>	10590–10647	+	58	–	–	1
<i>trnS_(UGA)</i>	10649–10717	+	69	–	–	0
<i>rrnS</i>	10718–11230	+	513	–	–	0
<i>trnF</i>	11231–11286	+	56	–	–	1
<i>cox2</i>	11288–11935	+	645	GTG	TAA	–1
<i>trnC</i>	11934–11988	+	55	–	–	21
<i>cox3</i>	12010–12741	+	732	ATG	TAA	1
<i>trnA</i>	12743–12799	+	57	–	–	17
<i>TrnR</i>	12816–12879	+	64	–	–	–8
<i>TrnN</i>	12870–12924	+	55	–	–	57
<i>nad2</i>	12982–13887	+	906	ATG	TAG	1

(Swofford 2001) using reestimated log likelihoods (RELL) approximation and 1000 nonparametric bootstrap replicates.

Results and Discussion

General Features

The mitochondrial genome of *L. thecatus* is circular and 13,888 bp in length. This is smaller than that of most vertebrate taxa, but it is similar in size to that of many other invertebrates such as nematodes, cestodes, and brachiopods, which typically range from 13,600 to 14,500 bp (e.g., Okimoto et al. 1992; Noguchi et al. 2000; von Nickisch-Rosenegk et al. 2001; Hu et al. 2002). The circular structure and proper assembly of the genome was confirmed by genome-wide PCR amplifications from genomic DNA. All resulting PCR

products were of the expected length; therefore, the circular structure was confirmed and proper alignment was expected to be achieved.

The genome contains at least 36 of the 37 genes typically found in animal mitochondrial genomes; however, an *atp8* gene could not be positively identified. Two putative *atp8* genes were identified, as were two possible *trnK* and three possible *trnS* genes. All genes except two alternative *trnS* genes occur on the same strand, which has a base composition of 44.5% T, 26.8% A, 19.9% G, and 8.4% C. The total base composition of the positive strand was similar to the base composition of the protein coding genes: 46.6% T, 24.7% A, 20.3% G, and 7.9% C. The genome also contains two significant noncoding regions that might be broken up by the putative *atp8*, *trnK*, and *trnS* genes. The inferred organization of the

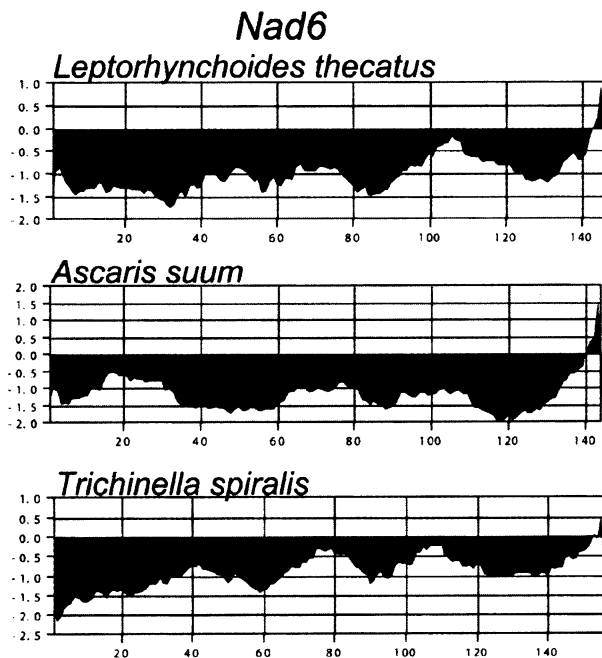


Fig. 2. Comparisons of Hopp/Woods hydrophilicity profiles of *nad6*.

genome is listed in Table 2. Nucleotide positions are numbered starting with the beginning of the *cox1* gene.

Protein Coding Genes

BLAST searches identified all protein coding genes except *nad6* and *atp8*. The hydrophobic profile of open reading frame (ORF) 2658–3095 was very similar to those of the *nad6* gene of two species of nematodes (Fig. 2); therefore, this ORF is likely *nad6* of *L. thecatus*. Two possible open reading frames could code for *atp8*, ORF 3152–3262 and ORF 4314–4466. Both are similar in length to the *atp8* gene of other organisms and ORF 3152–3262 directly precedes *atp6*, a common gene order in arthropod and duetostome mitochondrial genomes (Boore 1999). However, the hydrophobic profiles of both regions are very different from the *atp8* genes of *Terebratalia transversa* and *Mus musculus* and the putative *atp8* of *Trichinella spiralis* (Fig. 3); therefore, it is unlikely that either open reading frame codes for the *atp8* gene. Definite *atp8* genes have not been found in the mitochondrial genomes of either nematodes or platyhelminths (Boore 1999; Le et al. 2000; von Nickisch-Roseneck et al. 2001).

The most common initiation codon is ATG (6 genes), but ATA, GTG, ATT, and TTG are also used (Table 2). Many genes have alternative possible initiation codons, but those that best match the amino acid sequences and the hydrophobic profiles of homologous genes are inferred to be correct. Seven genes and the

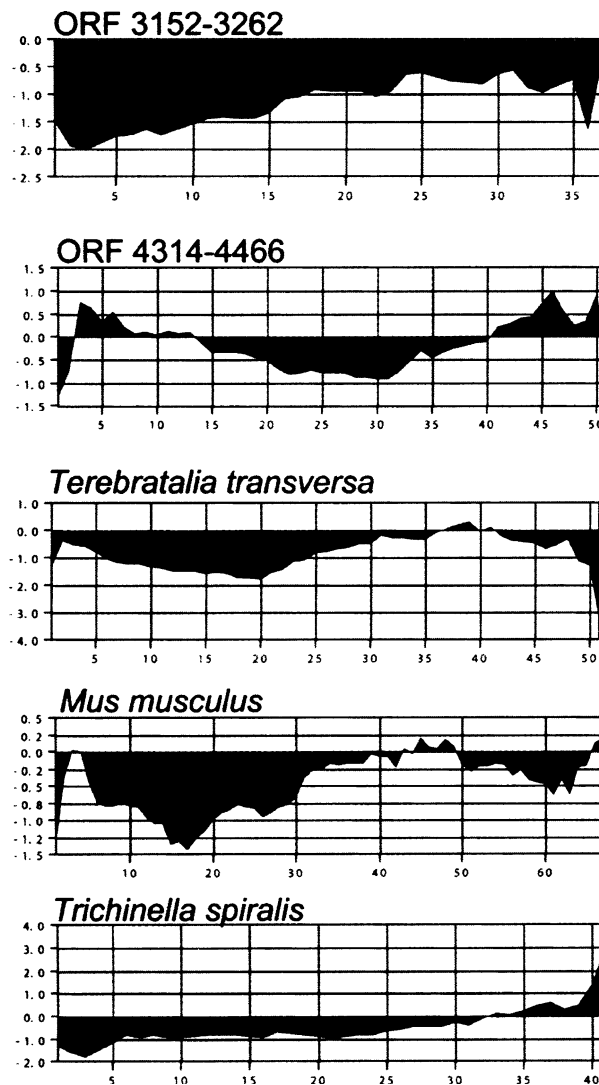


Fig. 3. Hopp/Woods hydrophilicity profiles of two open reading frames (ORF) of *Leptorhynchoides thecatus* compared to the *atp8* profiles of three other species. The identity of the gene for the profile of *Trichinella spiralis* is hypothesized to be *atp8*, but it has not been confirmed.

two possible *atp8* genes terminate with complete stop codons, TAA or TAG. Truncated stop codons of a single T apparently occur for *cox1*, *nad4*, *nad5*, *cob*, and *nad1*. Truncated stop codons are common among bilaterian mitochondrial protein coding genes, and are probably completed through posttranscriptional polyadenylation (Ojala et al. 1981).

Codon usage of all protein coding genes except the putative *atp8* genes is depicted in Table 3. The third nucleotide position is heavily biased toward T (51.5%) and A (29.7%) and corresponds to the overall nucleotide bias of the coding strand.

rRNA Genes

BLAST searches indicated sequence similarity between small portions of the *L. thecatus* mitochondrial

Table 3. Codon usage within 12 mitochondrial protein coding genes of *Leptorhynchoides thecatus*.

Amino acid (anticodon)	Codon group	Usage of codon ending in				Total	%
		A	C	G	T		
Ala (ugc)	GCN	17 (0.5)	3 (0.1)	4 (0.1)	46 (1.3)	70	2.1
Arg (ucg)	CGN	11 (0.3)	0	7 (0.2)	23 (0.7)	41	1.2
Asn (guu)	AAU	—	6 (0.2)	—	88 (2.6)	94	2.8
Asp (guc)	GAY	—	5 (0.1)	—	69 (2.0)	74	2.2
Cys (gca)	TGY	—	6 (0.2)	—	68 (2.0)	74	2.2
Gln (uug)	CAR	15 (0.4)	—	17 (0.5)	—	32	1.0
Glu (uuc)	GAR	34 (1.0)	—	43 (1.2)	—	77	2.3
Gly (ucc)	GGN	46 (1.3)	7 (0.2)	77 (2.2)	163 (4.7)	293	8.7
His (gug)	CAY	—	4 (0.1)	—	43 (1.2)	47	1.4
Ile (gau)	ATY	—	19 (0.6)	—	264 (7.7)	283	8.4
Leu							
uag	CTN	51 (1.5)	3 (0.1)	14 (0.4)	61 (1.8)	129	3.8
uua	TTR	336 (10.6)	—	86 (2.5)	—	422	12.5
Lys (uuu)	AAR	37 (1.1)	—	15 (0.4)	—	52	1.5
Met (cau)	ATR	163 (4.7)	—	77 (2.2)	—	240	7.1
Phe (gaa)	TTY	—	20 (0.6)	—	314 (9.1)	334	9.9
Pro (ugg)	CCN	16 (0.5)	3 (0.1)	5 (0.1)	49 (1.4)	73	2.2
Ser							
uga	TCN	32 (0.9)	4 (0.1)	0	96 (2.8)	132	3.9
ucu	AGN	58 (1.7)	14 (0.4)	47 (1.4)	103 (3.0)	222	6.6
Thr (ugu)	ACN	34 (1.0)	1	8 (0.2)	38 (1.1)	81	2.4
Trp (uca)	TGR	59 (1.7)	—	39 (1.1)	—	98	2.9
Tyr (gua)	TAY	—	19 (0.6)	—	136 (4.0)	155	4.6
Val (uac)	GTN	90 (2.6)	14 (0.4)	64 (1.9)	173 (5.0)	341	10.1
Total		999	128	503	1734	3364	
%		29.7	3.8	15.0	51.5		

genome and the *rrnL* and *rrnS* genes of other organisms; however, the similarity is low for *rrnS*. The boundaries of these genes were assumed to abut adjacent genes. The lengths of *rrnL* and *rrnS* are short, even compared to other invertebrates such as nematodes and flatworms, which have shorter mitochondrial genes than most bilaterians. The *rrnL* gene of *L. thecatus* is 925 bp, compared to 960 bp in *A. suum*, 958 bp in *Necator americanus*, 966 bp in *H. diminuta*, and 1105 bp in *T. transverse*. The length of *rrnS* is very short, 513 bp, compared to 701 bp in *A. suum*, 699 bp in *N. americanus*, 715 bp in *H. diminuta*, and 762 bp in *T. transversa*.

tRNA Genes

The tRNAscan algorithm detected only two tRNA genes (for amino acids R and I) within the complete mitochondrial genome sequence. Additional putative tRNA genes were detected by searching by eye for the anticodon sequences and surrounding sequences for secondary structure. A total of 24 possible tRNA genes were found in the mitochondrial genome of *L. thecatus*. This total includes the 22 tRNA genes typically found in animal mitochondrial genomes plus additional *trnK* and *trnS* genes. Many of the putative tRNA

genes exhibit unusual structures and should be considered as hypotheses of the tRNAs present in the genome. Most of the tRNAs are missing the variable loop and TΨC arm, which are replaced by a TV replacement loop (Fig. 4). This structure is similar to that described for the tRNA genes of many nematode species (Wolstenholme et al. 1987). The tRNAs range from between 52 to 104 nt and typically contain an aminoacyl stem of 6 or 7 ntp, a DHU stem of 3 or 4 ntp, and an anticodon stem of 4 or 5 ntp. Exceptions include *trnS2* (3163-3226), which has a DHU stem of only 1 ntp and *trnS2* (10,481-10,542), which has an anticodon stem of only 3 ntp. The DHU loop ranged from 1 to 11 nt and the anticodon loop of all tRNAs was 7 nt.

Noncoding Regions

There are potentially two significant noncoding regions in the genome that are 377 and 294 bp in length; however, these regions might be broken up by the putative *atp8*, *trnK*, and *trnS* genes (Table 2). These regions are the most likely candidates for a control region, but they do not contain any of the features found in the control regions of other organisms such as tandem repeats, a high A + T

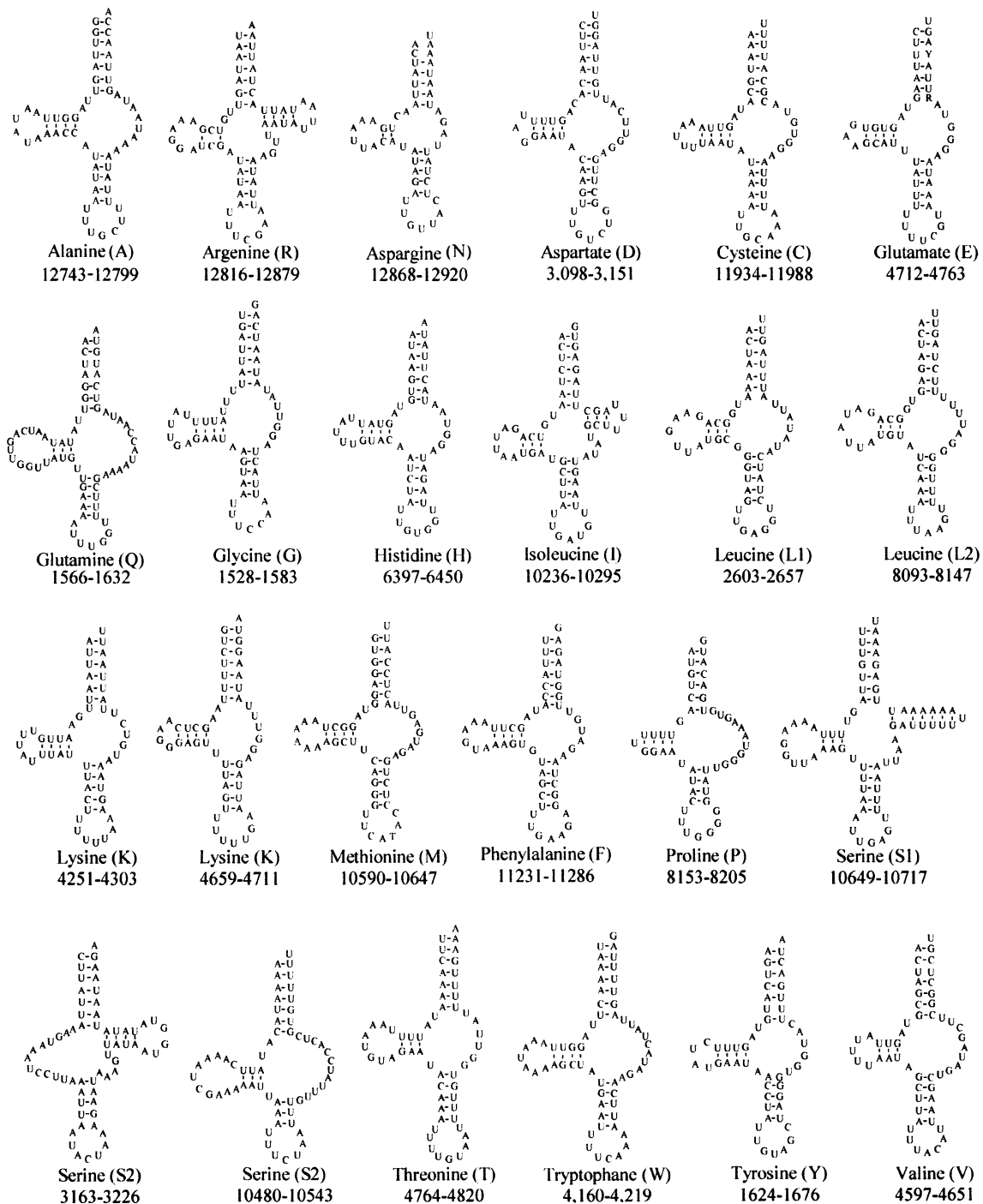


Fig. 4. Inferred secondary structure of tRNAs from the mitochondrial genome of *Leptorhynchoides thecatus*. Numbers indicate the position of each putative tRNA within the mitochondrial genome.

content, a stem-loop structure, or a T-stretch (Wolstenholme 1992; Zhang and Hewitt 1997). The A + T content of the first and second regions are 28.5% T, 32.2% A and 41.7% T, 32.1% A, respectively. These values are lower than those for the protein coding genes and therefore suggest that the non-coding regions are not A-T rich compared to the rest of the genome. A third significant noncod-

ing region of 104 bp might also be present if the putative *atp8* and *trnS* genes do not exist (Table 2). This region is 51% T and 30.7% A but does not contain any of the other aforementioned elements. None of these three regions is a good candidate for a control region, leaving the location of the signals that control replication and transcription unknown.

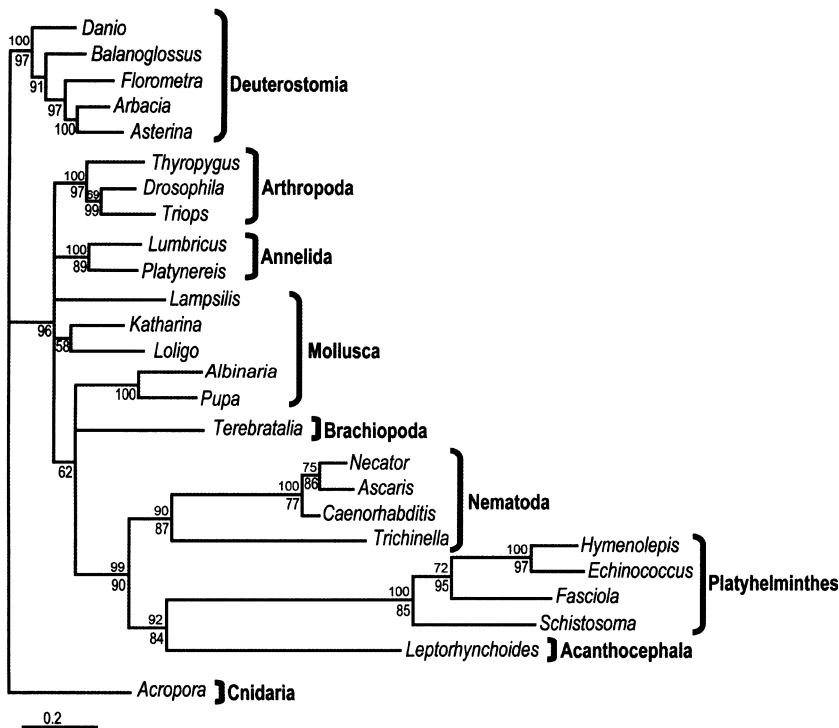


Fig. 5. Maximum likelihood phylogram from analysis of 1523 amino acids from the mitochondrial genome of 26 invertebrate taxa. Numbers below branches indicate quartet puzzling reliability values from the maximum likelihood analysis. Numbers above branches indicate bootstrap values from the parsimony analysis of nucleotide sequences. Only values for the shared nodes are shown.

Phylogenetic Analysis

Parsimony searches of the DNA and amino acid data each resulted in one most parsimonious tree 10,782 and 8,242 steps long, respectively. Tree statistics for the nucleotide derived tree are: consistency index (CI) = 0.406, retention index (RI) = 0.483, rescaled consistency index (RC) = 0.196, and homoplasy index (HI) = 0.594. Statistics for the amino acid derived tree are CI = 0.589, RI = 0.532, RC = 0.313, HI = 0.411. The topologies of the parsimony and maximum likelihood trees are very similar in that the deuterostomes, arthropods, annelids, nematodes, and platyhelminths were each recovered as monophyletic taxa with high bootstrap support (Fig. 5). Additionally, the platyhelminths, nematodes, and acanthocephalans were recovered as a monophyletic taxon in all of the analyses and the bootstrap support of this clade was high (89–99%). *Leptorhynchoides thecatus* is the sister taxon to Platyhelminthes in trees from all analyses except the ML tree derived from DNA, which placed *L. thecatus* as a sister to Nematoda, but with low (58%) bootstrap support. Additionally, an SH test comparing the alternative positions of *L. thecatus* in this topology (as a sister to nematodes or platyhelminths) indicated that there was no difference in the two alternatives ($p = 0.173$). The placement of *L. thecatus* and Platyhelminthes as sister taxa in this analysis supports the hypotheses of Winnepeninckx et al. (1995), Giribet (2000), and Giribet et al. (2002) (Hypothesis A, Fig. 1), who placed these taxa together in Platyzoa. The rela-

tionship does not, however, conflict with the hypothesis of Nielsen (2001) (Hypothesis B, Fig. 1), who left the relationships undetermined as a polytomy. However, the sister relationship of Nematodes to the Acanthocephala and Platyhelminthes as depicted by these analyses conflicts with all of the aforementioned hypotheses. If the nodes with less than 50% bootstrap support are collapsed, the relationships of the taxa are mainly unresolved except for the aforementioned clades.

The topology of our ML DNA tree is significantly better than either hypothesis A or B ($p < 0.019$ and $p < 0.001$, respectively). The Shimodaira and Hasegawa (1999) likelihoods of the trees are 45,767.28 (ML DNA), 45,822.52 (A), and 45,911.57 (B). The main difference between all of the topologies derived from our study and either of the hypotheses presented in Fig. 1 is the placement of the nematodes and their relationships with the arthropods, platyhelminths, and acanthocephalans. Hypothesis A groups the nematodes and arthropods in Ecdysozoa, a clade of molting organisms, which are distantly related to the platyhelminths and the acanthocephalans. Nematodes and arthropods are the only representatives of Ecdysozoa in our analyses, but they show very distant relationships in all topologies. Ecdysozoa has been supported in some analyses of molecular data (Aguinaldo et al. 1997; de Rosa et al. 1999; Giribet et al. 2000; Manuel et al. 2000; Giribet 2002; Copely et al. 2004) but not in others (Mushegian et al. 1998; Wang et al. 1999; Blair et al. 2002). In Hypothesis B, Ecdysozoa are not recognized, but the nematodes are distantly related to the

platyhelminths and acanthocephalans. Additionally, this hypothesis groups the brachiopods with the deuterostomes, which is a relationship not supported by our analyses. Another difference between our analyses and both a priori hypotheses is the status of the mollusks. In our analyses, the mollusks were not monophyletic as they are in both a priori hypotheses. Nonmonophyly of mollusks has also occurred in other analyses using molecular data (Winnepennckenx et al. 1995; Carranza et al. 1997).

Taxon sampling could pose a problem in this analysis because mitochondrial genomes of many taxa have not been sequenced. Missing taxa cause the divergences among the taxa in phylogenetic analyses to be large and can lead to error (Hillis et al. 2003). Additionally, base composition bias could also influence the results of the phylogenetic analyses because most of the nematode, platyhelminth, and acanthocephalan genomes contained high percentages of T's and low percentages of C's, which could influence analyses using DNA sequences and possibly amino acid sequences (Foster et al. 1997). The results from Tree-Puzzle indicated that amino acid composition of 14 of 25 taxa differed significantly from the frequency distribution assumed in the maximum likelihood model, which could also bias the analysis. However, these 14 taxa included not only those from the nematode, platyhelminth, acanthocephalan clade, but also those from other parts of the tree. Long branch attraction could also cause the acanthocephalan, nematode, and platyhelminth clade to group together since all of these taxa have relatively long branches. These organisms may be evolving rapidly or the long branches could be an artifact of missing taxa within the analysis.

Conclusions

The mitochondrial genome of *L. thecatus* is unusual in the small size of the rRNA genes, the structure of the tRNAs, and the absence of a control region with signaling elements similar to other organisms. More acanthocephalan species need to be sequenced to determine whether these are features typical of the group. In *L. thecatus*, mitochondrial genes and their amino acid sequences generally showed low similarity to their homologues in other organisms during BLAST searches and when aligned with sequences from representative nematodes, platyhelminths, brachiopods, and arthropods. The estimated structure of the majority of the tRNAs is similar to that reported from nematodes in that a TV-replacement loop takes the place of the variable loop and TΨC arm.

Phylogenetic analyses of Metazoa using amino acid and nucleotide sequences of conserved regions of protein coding genes support the monophyly of most recognized phyla except Mollusca. However, in the

analysis, Acanthocephala, Nematoda, and Platyhelminthes formed a monophyletic group, a hypothesis that has not previously been proposed. Furthermore, the data did not support the Ecdysozoa, a clade of molting organisms supported by previous studies. These results could be artifacts of missing taxa, codon bias, or long branch attraction. The utility of mitochondrial genome sequences for phylogenetic analysis will improve as sequence data become available from more "minor" phyla such as the Acanthocephala.

Acknowledgment. Partial funding for this project was provided by the University of Nebraska—Lincoln Initiative for Ecological and Evolutionary Analysis. We would also like to thank members of the Ortí lab at the University of Nebraska—Lincoln for their helpful comments on the manuscript.

References

- Abouheif E, Zardoya R, Meyer A (1998) Limitations of metazoan 18S *rRNA* sequence data: implications for reconstructing a phylogeny of the animal kingdom and inferring the reality of Cambrian Explosion. *J Mol Evol* 47:394–405
- Adachi J, Hasegawa M (1996) Model of amino acid substitution in protein encoded by mitochondrial DNA. *J Mol Evol* 42:459–468
- Aguinaldo AMA, Tuberville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA (1997) Evidence for a clade of nematodes, arthropods, and other molting animals. *Nature* 387:489–493
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27:573–580
- Blair JE, Ikeno K, Gojobori T, Hedges SB (2002) The evolutionary position of nematodes. *BMC Evol Biol* 2:1–7
- Boore JL (1999) Animal mitochondrial genomes. *Nucleic Acids Res* 27:1767–1780
- Boore JL, Brown WM (1998) Big trees from little genomes mitochondrial gene order as a phylogenetic tool. *Curr Opin Genet Dev* 8:668–674
- Bowles J, Blair D, McManus DP (1992) Genetic variants within the genus *Echinococcus* identified by mitochondrial DNA sequencing. *Mol Biochem Parasitol* 54:165–174
- Carranza S, Bagañá J, Riutort M (1997) Are the Platyhelminthes a monophyletic primitive group? An assessment using 18S rDNA sequences. *Mol Biol Evol* 14:485–497
- Copley RR, Aloy P, Russell RB, Telford MJ (2004) Systematic searches for molecular synapomorphies in model metazoan genomes give some support for Ecdysozoa after accounting for the idiosyncrasies of *Caenorhabditis elegans*. *Evolution* Dec. 6:164–169
- De Rosa R, Grenier JK, Andreeva T, Cook CE, Adoutte A, Akam M, Carroll SB, Balavoine G (1999) HOX genes in brachiopods and priapulids and protostome evolution. *Nature* 399:772–776
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791
- Foster PG, Jermini LS, Hickey DA (1997) Nucleotide composition bias affects amino acid content in proteins coded by animal mitochondria. *J Mol Evol* 44:282–288
- Garey JR, Schmidt-Rhaesa A, Near TJ, Nadler SA (1998) The evolutionary relationships of rotifers and acanthocephalans. *Hydrobiologia* 387/388:83–91

- Giribet G (2002) Current advances in the phylogenetic reconstruction of metazoan evolution. A new paradigm for the Cambrian explosion. *Mol Phylogenet Evol* 24:345–357
- Giribet G, Distel DL, Polz M, Sterrer W, Wheeler WC (2000) Triploblastic relationships with emphasis on the acoelomates and the position of Gnathostomulida, Cyclophora, Platyhelminthes, and Chaetognatha: a combined approach of 18S rDNA sequences and morphology. *Syst Biol* 49:539–562
- Helfenbein KG, Brown WM, Boore JL (2001) The complete mitochondrial genome of the articulate brachiopod *Terebratalia transversa*. *Mol Biol Evol* 18:1734–1744
- Herlyn H, Piskurek O, Schmitz J, Ehlers U, Zischler H (2003) The syndermatan phylogeny and the evolution of acanthocephalan endoparasitism as inferred from 18S rDNA sequences. *Mol Phylogenet Evol* 26:155–164
- Hillis DM, Pollock DD, McGuire JA, Zwickl DJ (2003) Is sparse taxon sampling a problem for phylogenetic inference? *Syst Biol* 52:124–126
- Hu M, Chilton NB, Gasser RB (2002) The mitochondrial genomes of the human hookworms, *Ancylostoma duodenale* and *Necator americanus* (Nematoda: Secernentea). *Int J Parasitol* 32:145–158
- Le TH, Blair D, Agatsuma T, Humair PF, Campbell NJH, Iwagami M, Littlewood DTJ, Peacock B, Johnston DA, Bartley J, Rollinson D, Herniou EA, Zarlenga DS, McManus DP (2000) Phylogenies inferred from mitochondrial gene orders—A cautionary tale from the parasitic flatworms. *Mol Biol Evol* 17:1123–1125
- Lowe TM, Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964
- Manuel MM, Kruse M, Müller WEG, Le Parco Y (2000) The comparison of β -thymosin homologues among Metazoa supports an arthropod-nematode clade. *J Mol Evol* 51:378–381
- Morret BME, Wang LS, Warnow T, Wyman SK (2001) New approaches for reconstructing phylogenies from gene order data. *Bioinformatics* 17:S165–S173
- Mushegian AR, Garey JR, Martin J, Liu LX (1998) Large-scale taxonomic profiling of eukaryotic model organisms: A comparison of orthologous proteins encoded by the human, fly, nematode, and yeast genomes. *Genome Res* 8:590–598
- Nielsen C (2001) *Animal Evolution: Interrelationships of the living phyla*, 2nd ed. Oxford University Press, Oxford
- Noguchi Y, Endo K, Tajima F, Ueshima R (2000) The mitochondrial genome of the brachiopod *Laqueus rubellus*. *Genetics* 155:245–259
- Ojala D, Montoya J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290:470–474
- Okimoto R, Macfarland JL, Clary DO, Wolstenholme DR (1992) The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. *Genetics* 130:471–498
- Posada D, Crandall KA (1998) Modeltest: Testing the model of DNA substitution. *Bioinformatics* 14:817–818
- Sambrook JE, Fritsch F, Maniatis T (1989) *Molecular Cloning: A laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, New York
- Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* 16:1114–1116
- Strimmer K, von Haeseler A (1996) Quartet puzzling: A quartet maximum likelihood method for reconstructing tree topologies. *Mol Biol Evol* 13:964–969
- Swofford DL (2001) PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sinauer Associates, Sunderland, MA
- Von Nickisch-Roseneck M, Brown WM, Boore JL (2001) Complete sequence of the mitochondrial genome of the tapeworm *Hymenolepis diminuta*: Gene arrangements indicate that Platyhelminths are eutrochozoans. *Mol Biol Evol* 18:721–730
- Wang DY, Kumar S, Hedges SB (1999) Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. *Proc R Soc Lond B Biol Sci* 266:163–171
- Winnepenninckx B, Backeljau T, DeWachter R (1995a) Phylogeny of protostome worms derived from 18S *rrnA* sequences. *Mol Biol Evol* 12:641–649
- Winnepenninckx B, Backeljau T, Mackey LY, Brooks JM, DeWachter R, Kumar S, Garey JR (1995b) 18S rRNA data indicate that Aschelminthes are polyphyletic in origin and consist of at least three distinct clades. *Mol Biol Evol* 12:1132–1137
- Wolstenholme DR (1992) Animal mitochondrial DNA: structure and evolution. *Int Rev Cytol* 141:173–216
- Wolstenholme DR, Macfarlane JL, Okimoto R, Clary DO, J. Wahleithner A (1987) Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. *Proc Natl Acad Sci USA* 84:1324–1328
- Zhang DX, Hewitt GM (1997) Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochem Syst Ecol* 25:99–120
- Zrzavy J, Mihulka S, Kepka P, Bezedek A, Tietz O (1998) Phylogeny of Metazoa based on morphological and 18S ribosomal DNA evidence. *Cladistics* 14:249–285