

The Complete Sequence of the Zebrafish (*Danio rerio*) Mitochondrial Genome and Evolutionary Patterns in Vertebrate Mitochondrial DNA

Richard E. Broughton,^{1,3} Jami E. Milam,² and Bruce A. Roe²

¹Oklahoma Biological Survey and Department of Zoology, ²Department of Chemistry and Biochemistry, University of Oklahoma, Norman, Oklahoma 73019, USA

We describe the complete sequence of the 16,596-nucleotide mitochondrial genome of the zebrafish (*Danio rerio*); contained are 13 protein genes, 22 tRNAs, 2 rRNAs, and a noncoding control region. Codon usage in protein genes is generally biased toward the available tRNA species but also reflects strand-specific nucleotide frequencies. For 19 of the 20 amino acids, the most frequently used codon ends in either A or C, with A preferred over C for fourfold degenerate codons (the lone exception was AUG: methionine). We show that rates of sequence evolution vary nearly as much within vertebrate classes as between them, yet nucleotide and amino acid composition show directional evolutionary trends, including marked differences between mammals and all other taxa. Birds showed similar compositional characteristics to the other nonmammalian taxa, indicating that the evolutionary trend in mammals is not solely due to metabolic rate and thermoregulatory factors. Complete mitochondrial genomes provide a large character base for phylogenetic analysis and may provide for robust estimates of phylogeny. Phylogenetic analysis of zebrafish and 35 other taxa based on all protein-coding genes produced trees largely, but not completely, consistent with conventional views of vertebrate evolution. It appears that even with such a large number of nucleotide characters (11,592), limited taxon sampling can lead to problems associated with extensive evolution on long phyletic branches.

Mitochondria provide the primary source of cellular ATP in eukaryotes via the process of oxidative phosphorylation (Saraste 1999). In animals, extranuclear mitochondrial genomes are typically circular, and with few exceptions, code for 13 subunits of the oxidative phosphorylation machinery as well as genes for two rRNA subunits and 22 tRNAs (Boore 1999). Mutations in mitochondrial DNA (mtDNA) have a number of known deleterious effects. At least 50 base substitutions and hundreds of insertion/deletion mutations have been identified in human mtDNA (MITOMAP 2000), with effects ranging from degenerative diseases (Wallace 1999) to aging (Michikawa et al. 1999) to cancer (Polyak et al. 1998; Fliss et al. 2000). In addition to their role as the powerhouse of the cell, mitochondria are also involved in regulating programmed cell death (apoptosis) (Green and Reed 1998; Susin et al. 1999), and mutagenic reactive oxygen species are generated in the process of energy production (Croteau and Bohr 1997). Pathologies can result directly from the loss of ATP production in affected tissues, the build-up of oxygen radicals due to downstream blockage of the oxidative phosphorylation pathway, or unregulated apoptosis (for review, see Wallace 1999).

Hundreds of mitochondria and thousands of mtDNAs are inherited maternally through the cytoplasm of the oocyte (Lightowers et al. 1997). If a zygote receives more than one form of mtDNA (heteroplasmy), different forms can be randomly distributed to daughter cells during cell division and, over many cell generations, can drift to high or low frequen-

cies in various cell lineages (Hauswirth and Laipis 1982). Thus, if one of the mutant forms is deleterious, disease may affect lineages where it reaches sufficiently high frequency (Wallace 1999). Somatic mutations in mtDNA appear to behave similarly and may be a significant source of mitochondrial disease. Given the central role of mitochondria in cell physiology, mutations (either inherited or somatic) are probably responsible for many developmental abnormalities. Although a high frequency of mutant mtDNA molecules is likely to be lethal during embryogenesis, oocytes with moderate to low levels of heteroplasmy occur at detectable levels (Lightowers et al. 1997). Mitochondrial mutations probably affect a number of both general and tissue-specific developmental processes; however, the role of mitochondria in early development has not been well characterized.

Structurally, most animal mitochondrial genomes contain the same 37 genes, and among vertebrates the gene order is highly conserved (Brown 1985; Boore 1999). Vertebrate mitochondrial genomes are typically ~16 kb and are extremely compact with no introns and few, if any, intergenic spacers. The only significant noncoding sequence is the control region, which is involved in regulating transcription and replication (Clayton 1982; Shadel and Clayton 1997) and is usually <5% of the total genome size. Since its endosymbiotic origin around 1.5 billion yr ago, a substantial fraction of original mitochondrial genes have moved to the nucleus (Gray et al. 1999). The products of many of these genes remain essential for oxidative phosphorylation or housekeeping functions and are selectively transported into the mitochondria after translation in the cytoplasm (Schatz 1996). A result of this evolutionary trend is that some mitochondrial abnormalities are due to mutations in genes that now reside in the nucleus and are inherited in a Mendelian fashion, rather than through

³Corresponding author.

E-MAIL rbroughton@ou.edu; FAX (405) 325-7702.

Article published on-line before print: *Genome Res.*, 10.1101/gr.156801.
Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.156801>.

the maternal, haploid inheritance of mtDNA (Wallace 1999). Knowledge of gene location (and thus mode of inheritance) is therefore essential for accurate characterization of the developmental-genetic basis of mitochondrial abnormalities.

Mitochondrial genomes from ~100 species have now been sequenced (Boore 1999; Curole and Kocher 1999). Their small size and relative autonomy from the nucleus makes mitochondrial genomes valuable windows on the process of genome evolution and with respect to cytonuclear interactions (Babcock and Asmussen 1996; Gray et al. 1999). Mitochondrial sequences have also proven to be of great utility in molecular phylogenetic studies, and complete genome sequences have provided valuable insights into deeper-level phylogenetic problems (e.g., Zardoya and Meyer 1997; Boore and Brown 1998; Mindell et al. 1998; Naylor and Brown 1998; Rasmussen and Arnason 1999).

The zebrafish, *Danio rerio*, (Actinopterygii, Cyprinidae) has become a prominent vertebrate genetic and developmental model system (Detrich et al. 1998). Here we describe the zebrafish mitochondrial genome with respect to gene content and organization, codon usage, nucleotide composition, and putative functional motifs. We also use a phylogeny based on the mitochondrial genomes of 36 vertebrate species to characterize the zebrafish relative to other vertebrates. We test evolutionary hypotheses pertaining to nucleotide and amino acid composition, genome-wide patterns of variability, and evolutionary rates among major vertebrate lineages. The wild-type zebrafish mitochondrial genome and its relation to other vertebrates may provide an important baseline for studies of development, disease, and mitochondrial function. It will also complement the sequence of the zebrafish nuclear genome that may be completed in the near future.

RESULTS AND DISCUSSION

The total length of the zebrafish mitochondrial genome was 16,596 bp, within 100 bp of the length of other teleost fishes. It is particularly close to two other cyprinid fishes, only 18 bp longer than the goldfish (*Carassius auratus*) and 21 bp longer than the carp (*Cyprinus carpio*). The zebrafish gene order and content (Table 1) is identical to the common vertebrate form. The genome contains 13 protein-coding genes: seven subunits of the NADH ubiquinone oxidoreductase complex (ND), one subunit of the ubiquinol cytochrome c oxidoreductase complex (*Cytb*), three subunits of the cytochrome c oxidase complex (CO), and two subunits of ATP synthase (ATP). Also contained are small (12S) subunit and large (16S) subunit ribosomal RNA genes and 22 tRNA genes. A noncoding control region contains the origin of heavy strand replication, and the origin of light strand replication is found within a complex of five tRNA genes. Sequence data provided no evidence of length or nucleotide site heteroplasmy in the individual zebrafish studied.

Protein-Coding Genes

Salient features of zebrafish mitochondrial genes are listed in Table 1. All but one of the protein-coding genes begin with the orthodox ATG start codon; the lone exception is *COI*, which begins with GTG. Stop codons include seven TAA and three TAG. The *COII*, *ND4*, and *Cytb* genes do not possess proper stop codons but do show a terminal T or TA. This condition is not uncommon among vertebrate mitochondrial genes, and it appears that TAA stop codons are created via

posttranscriptional polyadenylation (Ojala et al. 1981). Reading frames of two pairs of genes, *ATP8-ATP6* and *ND4L-ND4*, each overlap by seven nucleotides, and *ND5-ND6* overlap by four nucleotides. This is also a common vertebrate feature, although in mammals *ATP8* and *ATP6* overlap by 40–46 bp. A number of other genes share one or two nucleotides in common with adjacent tRNA genes (Table 1).

Nucleotide composition (Table 2) reflects a vertebrate bias against G on the light strand. (For all protein genes except *ND6*, the heavy strand serves as the template for transcription; however, we discuss gene sequences in terms of the light strand, which is the sense strand with respect to mRNA. Only for *ND6* is the heavy strand the sense strand). The bias against G is particularly marked at third codon positions, where only 7% of sites are G. A complementary bias is maintained in *ND6*, in which C is found at only 12% of all sites and only 3% of third positions.

Codon usage patterns in zebrafish are shown in Table 3. Three other vertebrates (another fish, a frog, and human) are shown for comparison. For amino acids with fourfold degenerate third positions, codons ending in A are always the most frequent in zebrafish, followed in frequency by codons ending in T or C. Among twofold degenerate codons, C appears to be used somewhat more than T. Consistent with the overall bias against G, G is the least common third position nucleotide in all categories except for arginine and glycine codons (where G is similar in frequency to C and T but still much less than A). These patterns are generally similar across vertebrate groups, although the frog, *Xenopus*, shows a tendency to use T more frequently than C (Roe et al. 1985).

Because there are only 22 tRNAs in the mitochondria, there is only one specific tRNA species for most amino acids. The exceptions are leucine and serine, which have two tRNAs serving six possible codons each. Thus, of the possible codons for any amino acid, only one (occasionally two) will be perfectly complementary to the available tRNA anticodon, and translation of the others must involve nonspecific (wobble) base pairing. For each amino acid, only codons ending in A or C will be perfectly matched by a complementary tRNA anticodon (Table 3). The matched nucleotide is A for fourfold codons and twofold purine codons, and C for twofold pyrimidine codons (except methionine, AUG, which also serves as a start codon). Thus in zebrafish, the most frequently used codons are those with matching tRNAs except for isoleucine, methionine, and phenylalanine. This indicates that there may be some advantage to matched codons and anticodons in protein translation, although, as seen in *Xenopus*, the phenomenon is not universal.

Codon bias may be associated with strand-specific nucleotide bias in mtDNA. It is noteworthy that, for *ND6*, 16 of the 20 amino acids are preferentially specified by codons that do not match the available tRNA. This is almost exclusively due to the high number of codons ending in T. Strand-specific substitution bias favoring A over T on the light strand and thus T over A on the complementary heavy strand appears to be the explanation. This bias is particularly evident at third codon positions, where substitutions are less likely to alter encoded amino acids and nucleotides are freer to vary. For all protein genes on the light strand (= sense strand), the frequency of A at third positions is 44% and the frequency of T is 27%. Conversely, for *ND6* (sense strand = heavy strand) the frequency of A at third positions is 24% and the frequency of T is 52%. It therefore appears that nucleotide bias is a strand-specific phenomenon rather than being strictly a result

Table 1. Organization of the Zebrafish Mitochondrial Genome

Gene/element	Strand	Position	Size	Start	Stop	3'
Control region	—	1–950	951	—	—	0
<i>tRNA^{phe}</i>	H	951–1019	69	—	—	0
12S <i>rRNA</i>	H	1020–1971	952	—	—	0
<i>tRNA^{val}</i>	H	1972–2042	71	—	—	0
16S <i>rRNA</i>	H	2043–3725	1683	—	—	0
<i>tRNA^{leu} (TAA)</i>	H	3726–3802	77	—	—	0
<i>ND1</i>	H	3803–4777	975	ATG	TAA	3
<i>tRNA^{ile}</i>	H	4781–4854	74	—	—	-3
<i>tRNA^{gln}</i>	L	4852–4922	71	—	—	1
<i>tRNA^{met}</i>	H	4924–4992	69	—	—	0
<i>ND2</i>	H	4993–6039	1047	ATG	TAG	0
<i>tRNA^{trp}</i>	H	6038–6107	70	—	—	1
<i>tRNA^{ala}</i>	L	6109–6176	68	—	—	1
<i>tRNA^{asn}</i>	L	6178–6250	73	—	—	0
<i>O_L</i>	—	6251–6282	32	—	—	0
<i>tRNA^{cys}</i>	L	6283–6354	72	—	—	-2
<i>tRNA^{tyr}</i>	L	6353–6423	71	—	—	1
<i>COI</i>	H	6425–7975	1551	GTG	TAA	0
<i>tRNA^{ser} (TGA)</i>	L	7976–8046	71	—	—	2
<i>tRNA^{asp}</i>	H	8049–8117	69	—	—	12
<i>COII</i>	H	8130–8820	691	ATG	T..	0
<i>tRNA^{lys}</i>	H	8821–8893	73	—	—	0
<i>ATP8</i>	H	8894–9058	165	ATG	TAA	-7
<i>ATP6</i>	H	9052–9735	684	ATG	TAA	-1
<i>COIII</i>	H	9735–10520	786	ATG	TAG	-1
<i>tRNA^{gly}</i>	H	10520–10589	70	—	—	0
<i>ND3</i>	H	10590–10940	351	ATG	TAG	-2
<i>tRNA^{arg}</i>	H	10939–11008	70	—	—	0
<i>ND4L</i>	H	11009–11305	297	ATG	TAA	-7
<i>ND4</i>	H	11299–12680	1382	ATG	TA.	0
<i>tRNA^{his}</i>	H	12681–12750	70	—	—	-1
<i>tRNA^{ser} (GCT)</i>	H	12750–12819	70	—	—	1
<i>tRNA^{leu} (TAG)</i>	H	12823–12895	73	—	—	1
<i>NDS5</i>	H	12897–14717	1821	ATG	TAA	-4
<i>ND6</i>	L	14714–15232	519	ATG	TAA	0
<i>tRNA^{glu}</i>	L	15233–15301	69	—	—	6
<i>Cytb</i>	H	15308–16448	1141	ATG	T..	0
<i>tRNA^{thr}</i>	H	16449–16520	72	—	—	5
<i>tRNA^{pro}</i>	L	16527–16596	70	—	—	0

L and H refer to light and heavy strands, respectively. The 3' column indicates the number of nucleotides downstream to the start of the next gene (this is the 5' end of genes on the L strand). The value is negative for genes with overlapping reading frames.

of codon preference for translation efficiency. The preference for A may also be related to transcription efficiency, as ATP is generally the most common ribonucleotide available in mitochondria (Xia 1996). The substantial bias against G may also be due in part to selection against less stable G nucleotides on the light strand, which is exposed as a single strand for a considerable length of time during the asymmetrical replication of mtDNA (Clayton 1982).

Transfer and Ribosomal RNA Genes

All zebrafish mitochondrial tRNA genes possess anticodons that match the vertebrate mitochondrial genetic code. Each tRNA sequence may be folded into a cloverleaf structure with 7 bp in the aminoacyl stem, 5 bp in the TψC and anticodon stems, and 4 bp in the DHU stem. tRNA stem regions include numerous noncomplementary and T-G base pairings, several of which are shared with carp. Such mutations appear to accumulate in mitochondrial genes, in part because mtDNA is not subject to the process of recombination, which may facilitate elimination of deleterious mutations (Lynch 1997). As in other vertebrates, it appears that CCA nucleotides are

added posttranscriptionally to the 3' ends to form mature, functional species (Roe et al. 1985).

Both ribosomal gene sequences may be folded into secondary structures. Potential secondary structures have free energies of -220.3 kcal/mole for the 12S subunit and -335.6

Table 2. Nucleotide Frequencies and Proportions in Zebrafish Protein-coding Genes

Codon position	Nucleotide			
	A	C	G	T
All	3399	2717	1765	3522
	0.30	0.24	0.15	0.31
1	1053	871	971	906
	0.28	0.23	0.26	0.24
2	713	1020	510	1558
	0.19	0.27	0.13	0.41
3	1633	826	284	1058
	0.43	0.22	0.07	0.28

Table 3. Comparison of Codon Usage (Number of Codons) among Zebrafish and Selected Vertebrates

Amino acid	Codon	Zebrafish	Carp	Frog	Human
Lys	AAA*	81	73	77	85
	AAG	8	4	8	10
Asn	AAC*	65	76	80	131
	AAT	56	45	70	33
Thr	ACA*	140	146	145	133
	ACG	13	8	8	10
	ACC	77	104	71	155
	ACT	62	41	83	51
Stop	AGA	0	0	1	0
	AGG	0	0	0	0
Ser	AGC*	38	46	42	39
	AGT	14	9	15	14
Met	ATA	139	126	161	167
	ATG*	52	49	35	40
Ile	ATC*	107	137	110	196
	ATT	201	157	228	125
Gln	CAA*	84	98	93	81
	CAG	10	1	6	9
His	CAC*	68	79	56	79
	CAT	34	23	40	18
Pro	CCA*	114	105	121	52
	CCG	13	12	11	7
	CCC	40	52	20	119
	CCT	40	33	52	41
Arg	CGA*	53	45	50	29
	CGG	9	10	1	2
	CGC	4	15	5	25
	CGT	9	9	10	7
Leu	CTA*	180	297	141	276
	CTG	29	37	10	45
	CTC	38	97	54	167
	CTT	146	82	159	65
Glu	GAA*	80	93	83	64
	GAG	20	8	14	24
Asp	GAC*	54	64	38	51
	GAT	30	11	36	15
Ala	GCA*	132	129	114	80
	GCG	13	10	6	8
	GCC	96	144	83	124
	GCT	87	55	73	43
Gly	GGA*	116	137	113	67
	GGG	45	35	23	34
	GGC	39	44	40	88
	GGT	36	32	44	24
Val	GTA*	111	118	75	70
	GTG	22	18	17	18
	GTC	21	30	25	49
	GTT	69	43	62	30
Stop	TAA	7	4	3	2
	TAG	4	1	0	2
Tyr	TAC*	57	65	57	89
	TAT	55	50	60	46
Ser	TCA*	108	82	123	83
	TCG	9	9	11	7
	TCC	19	57	44	99
Trp	TCT	57	29	66	32
	TGA*	107	112	101	93
	TGG	11	8	16	11
Cys	TGC*	15	20	18	17
	TGT	14	5	13	5
Leu	TTA*	181	100	217	73
	TTG	26	11	20	16
Phe	TTC*	88	146	103	141
	TTT	148	76	125	77

*Codons for which a tRNA with matching anticodon occurs in mitochondria.

kcal/mole for the 16S subunit. Stem regions appear to be conserved, whereas loop regions are somewhat more variable relative to other vertebrate sequences. The functional requirement for specific base pairing appears to constrain the evolution of stems relative to some portions of loops. This pattern is consistent with phylogenetic studies of a number of vertebrate groups (e.g., Sullivan et al. 1995).

Noncoding Sequences

The major noncoding region (control region) in mtDNA regulates replication and transcription (Clayton 1982,1991; Shadel and Clayton 1997). The primary sequence of much of the control region does not appear to be particularly important for regulatory function, as this region shows extensive variability across taxonomic groups and even among closely related species. The 950-bp zebrafish control region was much less similar to other fishes than were the coding sequences, with numerous nucleotide substitutions and insertions and deletions. However, several important regulatory elements are present. Conserved sequence blocks (CSBs) 1–3, found in the 3' end of the control region, appear to be involved in positioning RNA polymerase both for transcription and for priming replication (Clayton 1991; Shadel and Clayton 1997). All three of these elements are identifiable in zebrafish and they show strong similarity to CSBs identified by authors of other vertebrate sequences (Roe et al. 1985; Foran et al. 1988) (Table 4). Another relatively conserved element is the termination associated sequence (TAS) located at the 5' end of the control region. This sequence appears to act as a signal for termination of D-loop strand (7S DNA) synthesis. A sequence present in zebrafish (tacataaaatgcat) shows limited similarity with TAS sequences identified in other vertebrates. Regulatory elements in the zebrafish also show strong similarity to a North American cyprinid fish, the spotfin shiner (*Cyprinella spiloptera*; Broughton and Dowling 1994). Between tRNA^{asn} and tRNA^{cys} of zebrafish is a 32-bp noncoding sequence similar to the origin of light strand replication (O_L) in other vertebrates. Secondary structures at O_L appear to act as initiation signals for light strand replication (Wong and Clayton 1985). The zebrafish sequence may potentially form such a secondary structure consisting of a perfect 11-bp stem and a 14-bp loop.

Table 4. Sequences of Conserved mtDNA Regulatory Elements

Element	Taxon	Sequence
CSB-1	zebrafish	ATATTAAATATCTAAAGACATAA
	carpGT.....
	shiner	.GG.....G..TAT.....
	frog	TG.AA.TGC.ATG.CT.....T
CSB-2	human	TA.....GC.TGT.G.....
	zebrafish	CAAACCCCTTACCCCC
	carp
	shiner
CSB-3	frog	T.....CCT.....
	humanCCT.....
	zebrafish	TGTCAAACCCCTAAACCA
	carpG.....
	shinerG...G..
	frog	C.....A.....G
	human	..C.....A...A..

Dots indicate the same nucleotide as the top sequence; dashes indicate inferred gaps.

All of these features reflect strong conservation of vertebrate mitochondrial regulatory elements.

Mitochondrial Genome Evolution

Phylogeny

Phylogenetic analysis was used to estimate relationships among the zebrafish and 35 other vertebrates and to assess historical information content of mitochondrial genomes. The majority of complete mitochondrial sequences in GenBank are from mammals; some of these were not included to provide more balanced sampling of vertebrate groups. Maximum parsimony, maximum likelihood, and distance-based phylogenetic methods each have strengths and weaknesses under different evolutionary conditions (Swofford et al. 1996); thus, substantial differences in trees estimated by different methods may indicate violation of assumptions specific to each. Maximum likelihood using the general time reversible model (Lanave et al. 1984) yielded the tree that best matched the generally accepted phylogeny of vertebrates (Fig. 1). The position of zebrafish in all trees was as expected: sister to the other members of the family Cyprinidae (carp and goldfish) within actinopterygian fishes.

Relationships of the coelacanth and reptiles did not match the expected phylogeny in trees produced by any method in which sea lamprey was used as the outgroup, and their positions varied considerably with different methods. The coelacanth was sister to bony fishes rather than the expected sister-group to tetrapods for all methods. The amphibians were basal to all other ingroup taxa (including fishes) in the parsimony analysis and they were the sister-group to fishes in the distance tree using LogDet distances (Steel 1994). However, when lamprey was excluded and no outgroup was specified, the coelacanth and amphibians assumed their expected positions for most methods (data not shown).

Samples for each of the problematic groups were small and the included taxa are at the ends of long phyletic branches. Extreme branch-length variation is known to cause problems for all phylogenetic methods because multiple changes per nucleotide site resulting from extensive evolution on long branches can obscure historical signal and lead to spurious similarity between long-branch taxa. Maximum likelihood appears to be least susceptible to this condition (Hillis et al. 1994; Huelsenbeck 1995). Alternatively, LogDet distances are relatively insensitive to nucleotide frequency heterogeneity and may yield trees different from methods that do not account for nucleotide composition when this attribute varies substantially among taxa. We note that in tests of nucleotide composition homogeneity (not shown), nearly all comparisons of taxa showed significant differences, probably due (in part) to the greater statistical power afforded by large sample sizes (i.e., many nucleotides per sequence). However, because taxa with similar nucleotide frequencies were not always grouped together, the present results appear to be primarily due to problems associated with long branches. We suggest that relationships among basal tetrapods may be better resolved when long branches are broken up as additional mitochondrial genome sequences become available. This issue is particularly relevant to the outgroup. The lamprey was on the longest branch on the tree and the extensive amount of evolution between it and the ingroup taxa has likely obscured much historical signal, resulting in misleading character-state polarities. The more accurate result obtained when lamprey was excluded supports this hypothesis. The position

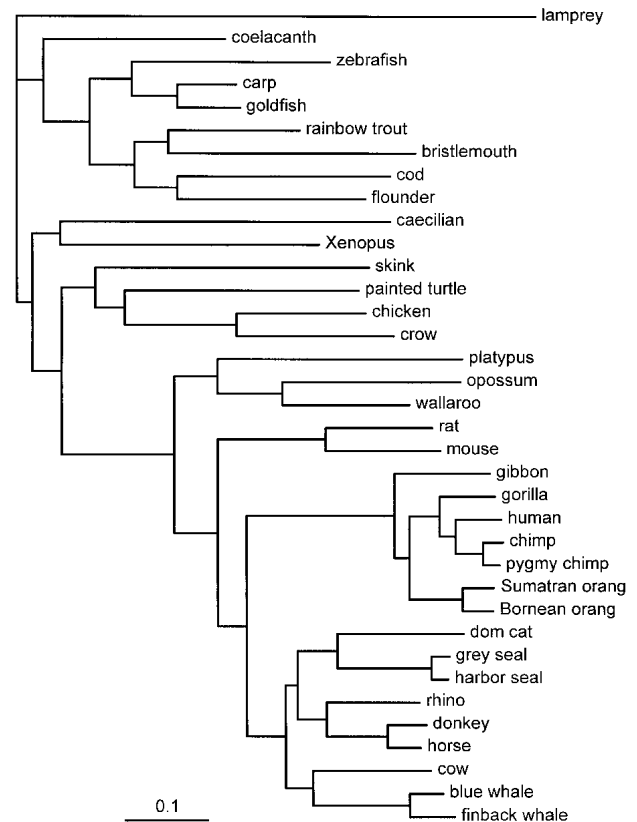


Figure 1 Maximum likelihood estimate of vertebrate phylogeny (–ln likelihood = 239855.997). Likelihood analysis used the GTR + I + Γ model with parameter values estimated from the maximum parsimony tree. This model was significantly better by likelihood ratio test than the Tamura-Nei + I + Γ model ($\chi^2 = 15,792$, d.f. = 3, $P < 0.001$). The tree shown had significantly greater likelihood for the present data than the generally accepted vertebrate tree, the maximum parsimony tree, and the LogDet distance tree indicated by the Shimodaira–Hasegawa (Shimodaira and Hasegawa 1999) test ($P < 0.001$ in each case). Scale bar indicates the probability of change per nucleotide site over the branch length interval shown. Lamprey was the designated outgroup.

of the turtle (rather than the lizard) as sister to birds in all analyses conflicts with most hypotheses of reptilian evolution, although a similar result has been observed previously (Zardoya and Meyer 1998) and may also be due to sparse taxon sampling and long branches. However, the turtle sequence appears to be unusual with respect to nucleotide and amino acid composition (see following). Within mammals, where taxon sampling is denser, relationships are largely as expected (for all methods). The nontraditional arrangement of the nonplacental mammals as a monophyletic group and the sister-group relationship of perissodactyls with the carnivore plus pinnipeds are consistent with recent evidence (Penny et al. 1999; Waddell et al. 1999). Ultimately, the fact that complete protein-coding sequences were not sufficient to correctly resolve all relationships emphasizes the importance of extensive taxon sampling even when the amount of data per taxon is high.

Evolutionary Genomic Patterns

Among vertebrate lineages there is substantial variation in thermal regulation, metabolic rate, and generation time. De-

spite extreme conservation of mitochondrial gene content and order, variation in these factors could influence genome-wide evolutionary rates, nucleotide composition, and/or functional constraints on proteins. Therefore, we might predict differences in genomic features that are independent of selection acting on individual genes, such as a correlation between rates of sequence divergence and body size or thermal habit (endotherms vs. ectotherms), two factors correlated with metabolic rate (Martin and Palumbi 1993).

Rates of Evolution across the Mitochondrial Genome

Nucleotide and amino acid substitution rates vary across mitochondrial genomes and within individual genes. Figure 2 shows a strong correlation between regional nucleotide variation and amino acid variation for all protein-coding genes. The magnitude of variability at third codon positions is consistently higher than at first and second positions, but the spatial pattern of variability is consistent over all positions. Rates of amino acid change appear to be lowest for the *COI* and *Cytb* genes and somewhat higher for others. All of the ND genes as well as *ATPase6* show a clear pattern of alternating regions of high and low variability. For *Cytb* it has been shown that the most variable regions frequently correspond to hydrophobic membrane-spanning domains of the protein (Irwin et al. 1991; Naylor et al. 1995). Across the genome there

is not perfect correspondence between variability and hydrophathy (Engelman et al. 1986); however, peaks of variability appear to coincide with extremes of hydrophathy, either positive or negative. This indicates that in many polar or nonpolar regions the specific identity of amino acid residues may be less important, relative to function, than simply maintaining the relative polarity of the domain.

Rates of Evolution among Lineages

There are two primary methods for evaluating molecular evolutionary rates. Absolute rates may be obtained by relating the amount of time that two taxa have been evolving independently (based on fossil or stratigraphic divergence dates) to the amount of change that has accumulated on particular lineages. Alternatively, relative rates may be evaluated by comparing the amount of change in two lineages since they diverged from a common ancestor, and do not require dates to be known. The latter method involves counting the number of unique nucleotide differences on each lineage relative to a third (the outgroup). The test statistic is asymptotically chi-square distributed and can be used to determine whether the number of changes on two lineages is significantly different (Tajima 1993). In the present case with 36 species, the number of possible combinations of two species plus an outgroup is prohibitively large; we therefore report selected com-

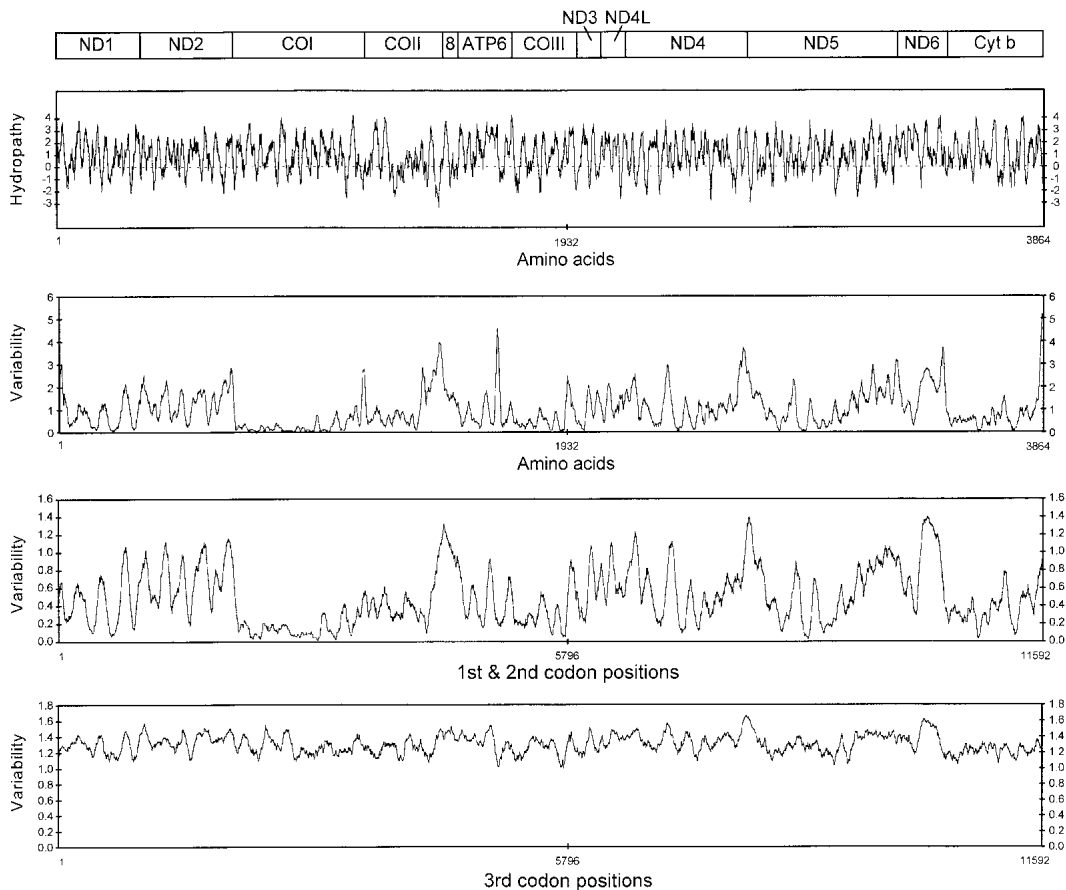


Figure 2 Variability of nucleotides, amino acids, and hydrophathy across the mitochondrial genome. The sequence is of the concatenated protein-coding genes and the bar at top indicates the limits of each gene. The sequence for ND6 is the reverse complement to maintain a consistent reading frame. Plots are the mean values for all 36 vertebrate taxa with a sliding window of 10 for amino acids and third-position nucleotides or 20 for (combined) first- and second-position nucleotides.

parisons with the phylogenetically closest outgroups (Table 5). Absolute rates are more difficult to quantify because few divergence dates are known with confidence, and branch lengths (the amount of divergence) on the phylogenetic tree seem disproportionately short for older taxa. Hence, we report only relative rate comparisons.

Relative rate tests indicate that evolutionary rates vary considerably within all vertebrate groups. Nearly all comparisons among taxonomic classes showed significant differences; however, many intraclass comparisons were significantly different as well (e.g., zebrafish–cod, chicken–crow, platypus–mouse, cat–rhino; Table 5). Rate differences within major groups are frequently as great as differences between groups. For example, the rate for the chicken is not significantly different from the rate for the lizard, yet rates for chicken and crow are significantly different, apparently due to a rate increase in the crow lineage. Rates in birds differ from mammals in some cases (chicken–human, crow–mouse) but not others (chicken–mouse, crow–rat). The fact that rates differ among

Table 5. Relative Rate Tests among Pairs of Vertebrate Lineages

Taxon pairs	Outgroup	Changes on lineage ^a	χ^2	P
Coelacanth	zebrafish	1182	38.15	<0.001
Frog		1502		
Frog	coelacanth	1426	0.90	0.344
Lizard		1477		
Lizard	frog	1283	2.02	0.155
Chicken		1356		
Zebrafish	coelacanth	1191	8.54	0.003
Cod		1338		
Mouse	frog	1408	5.48	0.019
Crow		1535		
Rat	frog	1509	1.34	0.246
Chicken		1446		
Human	frog	1565	28.76	<0.001
Chicken		1279		
Platypus	chicken	1309	8.50	0.004
Mouse		1164		
Chicken	lizard	890	7.10	0.008
Crow		1006		
Horse	gibbon	827	11.38	0.001
Cow		970		
Cat	cow	1050	7.91	0.005
Rhino		925		
Cow	rhino	951	1.28	0.258
Blue whale		1001		
Cat	rhino	972	1.16	0.281
Harbor seal		925		
Rhino	cow	798	0.98	0.323
Horse		759		
Carp	zebrafish	576	0.97	0.324
Goldfish		543		
Mouse	wallaroo	799	0.27	0.602
Rat		820		
Gibbon	rhino	754	0.01	0.938
Human		757		
Horse	rhino	364	1.98	0.159
Donkey		403		
Harbor seal	cat	180	0.13	0.715
Gray seal		187		
Human	gorilla	557	1.49	0.222
Chimp		517		

^aNumber of unique nucleotide changes on each lineage relative to the outgroup.

Table 6. Range of Nucleotide Frequencies (Percent) by Taxonomic Group and Codon Position

Taxon	A	C	G	T	G+C
Combined 1 st and 2 nd positions					
Fishes	22–23	25–28	19–20	30–32	45–48
Coelacanth	25	27	18	30	45
Amphibians	24	25–28	18–19	30–33	46
Reptiles	24–25	26–27	17–18	31–32	43–45
Birds	23–24	29	17–18	30	46–47
Mammals	25–26	24–28	16–17	30–34	40–44
3 rd positions					
Fishes	29–44	22–39	6–11	17–34	29–49
Coelacanth	48	28	7	17	35
Amphibians	36–43	22–35	5–12	17–30	27–47
Reptiles	37–46	29–33	5–11	19–20	36–43
Birds	38–41	35–43	6	14–18	41–48
Mammals	35–46	19–44	3–8	14–32	22–48

some endothermic vertebrates but may be similar between endotherms and ectotherms indicates that there are factors other than metabolic rate (and its associated oxygen radical production) that have important influence on rates of sequence evolution.

Nucleotide and Amino Acid Frequencies

Nucleotide frequencies varied by codon position and by taxonomic group (Table 6). At first and second positions, the relative frequencies were T > C? ≥ A > G and at third positions, relative frequencies were A > C > T > G. Although nucleotide frequencies and GC percent at third positions varied greatly within groups and there were no clear trends among groups, first and second position GC percent varied less and there was a clear difference between mammals (always <44%) and nonmammals (>45%, except for the turtle). This difference corresponds to a marked difference in amino acid composition of translated proteins. Amino acids with G or C in the first or second position of their codons were much more frequent in nonmammals than in mammals (Table 7). In particular, the frequency distribution of arginine (CGN) is completely nonoverlapping between the two groups, alanine (GCN) is also nonoverlapping except for the turtle, and glycine (GGN) is nearly so. For proline (CCN), the nonprimate mammals show frequencies lower than nonmammals; however, the primates appear to reverse this trend, having frequencies comparable to nonmammals. In a complementary way, A- and U-containing codons are more frequent in mammals. For example, isoleucine (AUY) is completely nonoverlapping and asparagine (AAY) is nearly so. Tyrosine (UAY) would be completely nonoverlapping except (again) for the turtle. Thus, whereas third position nucleotide frequencies vary widely from lineage to lineage, first and second position frequencies vary less and they reflect a strong difference in amino acid frequencies among mammalian and nonmammalian taxa. It is curious that the greater frequency of G- and C-containing codons in nonmammals is opposite of the expectation that organisms operating under cooler temperatures should contain more A and T and that G and C should be more frequent under warmer metabolic conditions.

Tests of amino acid frequency homogeneity within vertebrate groups (Table 8) indicate that significant departure

Table 7. Frequency Range of Selected Amino Acids among All Mitochondrial Proteins for Taxonomic Groups

Taxon	Ala	Arg	Asn	Gly	Ile	Pro	Tyr
Fishes	325–353	69–79	112–121	236–249	242–307	206–230	97–114
Coelacanth	285	73	147	239	284	205	116
Amphibians	274–321	68–70	136–148	224–227	273–334	203–220	111–116
Reptiles	243–322	71	132–138	225–227	279–300	204–208	111–125
Birds	289–301	72	126–130	215–221	301–304	221–234	105–108
Mammals	226–263	62–67	143–169	208–224	312–378	187–224	153–189

from homogeneity only occurs when the sample contains both amniotes and nonamniotes or mammals and nonmammals. It appears that there is substantially more variation among the nonmammals or nonamniotes than within the mammals or amniotes. It is also noteworthy that endothermic birds are more similar to ectotherms than to the endothermic mammals in amino acid frequency. These results indicate that common ancestry and long-term historical trends may have more influence on nucleotide and amino acid composition than does thermal habit or metabolic rate.

Conclusions

Despite extensive variation in mitochondrial genome structure among animal phyla, gene order and content are identical among zebrafish and the majority of vertebrates. Of 56 known vertebrate mitochondrial genomes, 44 have this same gene order, including all 30 placental mammals and 12 fishes (all but the sea lamprey) (Boore 1999). Gene order variants exist in some birds (Desjardins and Morais 1990), crocodylians (Janke and Arnason 1997), and snakes (Kumazawa et al. 1998); however, their taxonomic scope tends to be limited. The patterns of strand-specific nucleotide bias and unequal codon usage seen in the zebrafish are also conserved among vertebrates. In contrast, there are substantial differences in evolutionary rates and nucleotide and amino acid composition among vertebrate groups. The most striking finding is that rates of sequence evolution vary widely both within and among major groups, whereas nucleotide composition tends to be conserved within groups but varies substantially between mammals and all other taxa. This trend is also shown at

the amino acid level where lower GC percent in mammals (particularly at first and second positions) results in amino acids with A + T-containing codons being more frequent than those with G + C-containing codons. All other taxa show higher GC percent and corresponding differences in amino acid composition. Whether this is based on differences in nucleotide bias at the genome level or on adaptive significance of proteins with different amino acid composition is not clear. However, these differences appear to be independent of selection on specific genes and metabolic rate or thermal habit.

METHODS

Mature specimens of *Danio rerio*, strain ABC, were obtained from the University of Oregon Zebrafish Facility. Total DNA was isolated from ~25 mg of muscle tissue from a single individual. The preparation used the QIAamp tissue kit (QIAGEN) with DNA resuspended in 10 mM Tris (pH 8.0). Generation of sequencing templates used shotgun cloning of mtDNA. The mitochondrial genome was amplified by long PCR in two overlapping fragments of ~9 kb each. Reactions used Herculase DNA polymerase blend (Stratagene) according to supplier recommendations. One primer set consisted of 16S-AH (5' atgttttggtaaacagcg 3') located in the 16S rRNA gene and Arg-BL (5' caagacccttgattcgctca 3') in the arginine tRNA gene. The other primer set consisted of 16S-BL2 (5' tggcgcagcgcctattaagg 3') in the 16S rRNA gene and NAP-2H (5' tggagcttctactgtrgctt 3') in the ND4 gene. The two fragments overlapped over ND4L and part of ND4. The small gap between the divergent 16S primers was amplified (and sequenced separately) with primers 16S-ArL (5' ctgcgaaca-caagcctcgc 3') and 16S-BrH (5' tyayagatagaactgacctgg 3'). Products of long PCR were purified with QIAquick PCR purification columns (QIAGEN) and randomly fragmented using a nebulizer. Broken ends of DNA strands were repaired with Klenow polymerase and phosphorylated with T4 polynucleotide kinase. Fragments in the range of 600–3000 bp were selectively recovered from a 1.0% agarose gel. These fragments were blunt-end ligated into (*Sma*I cut, dephosphorylated) pUC19 at 4°C overnight and transformed into *Escherichia coli* JM109 cells. White colonies were picked and grown in 96-deep well blocks (1.75 mL Luria Bertani medium plus ampicillin per well).

Plasmids from four blocks (= 384 clones) were purified via alkaline lysis methods using a Hydra microdispenser (Robbins Scientific). Sequencing used universal primers, BigDye terminator chemistry, and ABI 377 instruments. Individual sequences were assembled and analyzed with phred/phrap/CONSED (Ewing et al. 1998; Gordon et al. 1998). A few small gaps were closed via sequencing with specific primers on whole-genome templates generated via long-PCR using primers 16S-AH and 16S-BL2. Sequence coverage was sufficient to achieve a CONSED error rate of 0.01%. The zebrafish sequence was manipulated and aligned with Sequencher (Gene-

Table 8. Tests of Amino Acid Frequency Homogeneity^a for all 13 Genes within Selected Groups of Vertebrate Taxa

Group	χ^2	P
All taxa	1010.03	<0.001
Non-amniotes	211.30	0.128
Amniotes	392.91	0.985
Non-mammals	321.12	0.012
Mammals	237.09	1.000
Endotherms	328.11	0.999
Ectotherms	277.41	0.014
Fishes	61.51	1.000
Amphibians	20.40	0.371
Reptiles (including birds)	108.11	0.638
Reptiles (not including birds)	21.89	0.290
Birds	10.30	0.945

^aThe null hypothesis is that amino acid frequencies are not different among all taxa within a particular group; P values <0.05 indicate significant heterogeneity within the group.

Codes). Genes were identified by alignment and comparison with sequences from other vertebrate taxa, particularly the common carp (*Cyprinus carpio*; Chang et al. 1994) and goldfish (*Carassius auratus*; Murakami et al. 1998). The complete sequence is available under GenBank accession no. AC024175.

Protein-coding genes from 35 vertebrate species obtained from GenBank were concatenated and aligned using Sequencher and corrected by eye to preserve reading frame. Regions of gene overlap and regions of unique insertions (gap in all but one sequence) were excluded, yielding a total alignment of 11,592 nucleotides. Of these, 8,189 were variable and 7,354 were parsimony informative (variants occurring in two or more sequences). Phylogenetic analyses using maximum parsimony, maximum likelihood, and neighbor-joining were performed with the computer program PAUP* ver. 4b8 (Swofford 2001). Parsimony searches used 100 random addition sequence replicates with TBR branch swapping and equal weighting of all characters/character-states. Maximum likelihood analysis used empirical mean nucleotide frequencies and empirical transition:transversion ratios. Other model parameters for likelihood were obtained by optimization on trees obtained either from maximum parsimony or neighbor-joining with LogDet distances. These parameters included the shape parameter (α) for a discrete (four-class) gamma approximation of among site rate variation, the estimated proportion of invariant sites, and values for nucleotide change-rate matrices. The choice of likelihood rate model was based on model performance using the likelihood ratio test. Given the large numbers of taxa and characters, maximum likelihood analyses were computationally intensive and only a few rearrangements (<200) were evaluated for each search. However, all likelihood searches using the GTR model quickly converged on the same tree, regardless of starting parameters. Calculation of codon usage and nucleotide and amino acid frequencies used the computer program DAMBE ver. 4.0.43 (Xia 2000) and relative rate tests were conducted with MEGA ver. 2.0 (Kumar et al. 2001).

ACKNOWLEDGMENTS

We thank W. Trevarrow for providing zebrafish specimens and S. Kenton for assistance with computational analysis and interpretation. Funding for this work was provided by grants from the National Institutes of Health and National Science Foundation EPSCoR program to B.A.R.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Babcock, C. and Asmussen, M.A. 1996. Effects of differential selection in the sexes on cytonuclear polymorphism and disequilibria. *Genetics* **144**: 839–853.
- Boore, J.L. 1999. Animal mitochondrial genomes. *Nucleic Acids Res.* **27**: 1767–1780.
- Boore, J.L. and Brown, W.M. 1998. Big trees from little genomes: Mitochondrial gene order as a phylogenetic tool. *Curr. Opin. Genet. Dev.* **8**: 668–674.
- Broughton, R.E. and Dowling, T.E. 1994. Length variation in mitochondrial DNA of the minnow *Cyprinella spiloptera*. *Genetics* **138**: 179–190.
- Brown, W. M. 1985. The mitochondrial genome of animals. In *Molecular evolutionary genetics* (ed. R.J. MacIntyre), pp. 95–130. Plenum, New York.
- Chang, Y.-S., Huang, F.-L., and Lo, T.-B. 1994. The complete nucleotide sequence and gene organization of carp (*Cyprinus carpio*) mitochondrial genome. *J. Mol. Evol.* **38**: 138–155.
- Clayton, D.A. 1982. Replication of animal mitochondrial DNA. *Cell* **28**: 693–705.
- . 1991. Nuclear gadgets in mitochondrial DNA replication and transcription. *Trends Biochem. Sci.* **16**: 107–111.
- Croteau, D.L. and Bohr, V.A. 1997. Repair of oxidative damage to nuclear and mitochondrial DNA in mammalian cells. *J. Biol. Chem.* **272**: 25409–25412.
- Curole, J.P. and Kocher, T.D. 1999. Mitogenomics: Digging deeper with complete mitochondrial genomes. *Trends Ecol. Evol.* **14**: 394–398.
- Desjardins, P. and Morais, R. 1990. Sequence and gene organization of the chicken mitochondrial genome. *J. Mol. Biol.* **212**: 599–634.
- Detrich, H.W., Westerfield, M., and Zon, L.I. 1998. *The zebrafish: Biology*. Academic Press, New York.
- Engelman, D.M., Steitz, T.A., and Goldman, A. 1986. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu. Rev. Biophys. Biomol. Struct.* **15**: 321–353.
- Ewing, B., Hillier, L., Wendl, M., and Green, P. 1998. Basecalling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**: 175–185.
- Fliss, M.S., Usadel, H., Caballero, O.L., Wu, L., Buta, M.R., Eleff, S.M., Jen, J., and Sidransky, D. 2000. Facile detection of mitochondrial DNA mutations in tumors and bodily fluids. *Science* **287**: 2017–2019.
- Foran, D.R., Hixson, J.E., and Brown, W.M. 1988. Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop synthesis. *Nucleic Acids Res.* **13**: 5841–5861.
- Gordon, D., Abajian, C., and Green, P. 1998. Consed: A graphical tool for sequence finishing. *Genome Res.* **8**: 195–202.
- Gray, M.W., Burger, G., and Lang, B.F. 1999. Mitochondrial evolution. *Science* **283**: 1476–1481.
- Green, D.R. and Reed, J.C. 1998. Mitochondria and apoptosis. *Science* **281**: 1309–1312.
- Hauswirth, W.W. and Laipis, P.J. 1982. Mitochondrial DNA polymorphism in a maternal lineage of Holstein cows. *Proc. Nat. Acad. Sci.* **79**: 4686–4690.
- Hillis, D.M., Huelsenbeck, J.P., and Cunningham, C.W. 1994. Application and accuracy of molecular phylogenies. *Science* **264**: 671–677.
- Huelsenbeck, J.P. 1995. Performance of phylogenetic methods in simulation. *Syst. Biol.* **44**: 17–48.
- Irwin, D.M., Kocher, T.D., and Wilson, A.C. 1991. Evolution of the cytochrome b gene of mammals. *J. Mol. Evol.* **32**: 128–144.
- Janke, A. and Arnason, U. 1997. The complete mitochondrial genome of *Alligator mississippiensis* and the separation between recent Archosauria (birds and crocodiles). *Mol. Biol. Evol.* **14**: 1266–1272.
- Kumar, S., Tamura, K., Jakobsen, I.B., and Nei, M. 2001. MEGA2: Molecular Evolutionary Genetics Analysis software, *Bioinformatics* (In press).
- Kumazawa, Y., Ota, H., Nishida, M., and Ozawa, T. 1998. The complete nucleotide sequence of a snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions. *Genetics* **150**: 313–329.
- Lanave, C., Preparata, G., Saccone, C., and Serio, G. 1984. A new method for calculating evolutionary substitution rates. *J. Mol. Evol.* **20**: 86–93.
- Lightowlers, R.N., Chinnery, P.F., Turnbull, D.M., and Howell, N. 1997. Mammalian mitochondrial genetics: Heredity, heteroplasmy and disease. *Trends Genet.* **13**: 450–455.
- Lynch, M. 1997. Mutation accumulation in nuclear, organelle, and prokaryotic genomes: Transfer RNA genes. *Mol. Biol. Evol.* **14**: 914–925.
- Martin, A.P. and Palumbi, S.R. 1993. Body size, metabolic rate, generation time, and the molecular clock. *Proc. Nat. Acad. Sci.* **90**: 4087–4091.
- Michikawa, Y., Mazzucchelli, F., Bresolin, N., Scarlato, G., and Attardi, G. 1999. Aging-dependent large accumulation of point mutations in the human mtDNA control region for replication. *Science* **286**: 774–779.
- Mindell, D.P., Sorenson, M.D., and Dimcheff, E.E. 1998. Multiple independent origins of mitochondrial gene order in birds. *Proc. Nat. Acad. Sci.* **95**: 10693–10697.
- MITOMAP: A Human Mitochondrial Genome Database. 2000. Center for Molecular Medicine, Emory University, Atlanta, GA. <http://www.gen.emory.edu/mitomap.html>.
- Murakami, M., Yamashita, Y., and Fujitani, H. 1998. The complete sequence of mitochondrial genome from a gynogenetic triploid 'ginbuna' (*Carassius auratus langsdorffi*). *Zool. Sci.* **15**: 335–337.
- Naylor, G.J.P. and Brown, W.M. 1998. Amphioxus mitochondrial DNA, Chordate phylogeny, and the limits of inference based on comparisons of sequences. *Syst. Biol.* **47**: 61–76.

- Naylor, G.J.P., Collins, T.M., and Brown, W.M. 1995. Hydrophobicity and phylogeny. *Nature* **373**: 565–566.
- Ojala, D., Montoya, J., and Attardi, G. 1981. tRNA punctuation model of RNA processing in human mitochondria. *Nature* **290**: 470–474.
- Penny, D., Hasegawa, M., Waddell, P.J., and Hendy, M.D. 1999. Mammalian evolution: Timing and implications from using the LogDeterminant transform for proteins of differing amino acid composition. *Syst. Biol.* **48**: 76–93.
- Polyak, K., Li, Y., Zhu, H., Lengauer, C., Willson, J.K., Markowitz, S.D., Trush, M.A., Kinzler, K.W., and Vogelstein, B. 1998. Somatic mutations of the mitochondrial genome in human colorectal tumours. *Nat. Genet.* **20**: 291–293.
- Rasmussen, A.-S. and Arnason, U. 1999. Phylogenetic studies of complete mitochondrial DNA molecules place cartilaginous fish within the tree of bony fish. *J. Mol. Evol.* **48**: 118–123.
- Roe, B.A., Ma, D.-P., Wilson, R.K. and Wong, J.F.-H. 1985. The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. *J. Biol. Chem.* **260**: 9759–9774.
- Saraste, M. 1999. Oxidative phosphorylation at the *fin de siècle*. *Science* **283**: 1488–1493.
- Schatz, G. 1996. The protein import system of mitochondria. *J. Biol. Chem.* **271**: 31763–31766.
- Shadel, G.S. and Clayton, D.A. 1997. Mitochondrial DNA maintenance in vertebrates. *Annu. Rev. Biochem.* **66**: 409–435.
- Shimodaira, H. and Hasegawa, M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**: 1114–1116.
- Steel, M.A. 1994. Recovering a tree from the leaf colourations it generates under a Markov model. *Appl. Math. Lett.* **7**: 19–23.
- Sullivan, J., Holsinger, K.E., and Simon, C. 1995. Among-site rate variation and phylogenetic analysis of 12S rRNA in sigmodontine rodents. *Mol. Biol. Evol.* **12**: 988–1001.
- Susin, S.A., Lorenzo, H.K., Zamzami, N., Marzo, I., Snow, B.E., Brothers, G.M., Mangion, J., Jacotet, E., Constantini, P., Loeffler, M., et al. 1999. Molecular characterization of mitochondrial apoptosis-inducing factor. *Nature* **397**: 441–446.
- Swofford, D.L. 2001. *PAUP*: Phylogenetic analysis using parsimony and other methods*. Sinauer, Sunderland, MA.
- Swofford, D.L., Olsen, G.J., Waddell, P.J., and Hillis, D.M. 1996. Phylogenetic inference. In *Molecular systematics* (eds. D.M. Hillis, C. Moritz, and B. Mable), pp. 407–514. Sinauer, Sunderland, MA.
- Tajima, F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* **135**: 599–607.
- Waddell, P.J., Cao, Y., Hasegawa, M., and Mindell, D.P. 1999. Assessing the cretaceous superordinal divergence times within birds and placental mammals by using whole mitochondrial protein sequences and an extended statistical framework. *Syst. Biol.* **48**: 119–137.
- Wallace, D.C. 1999. Mitochondrial diseases in man and mouse. *Science* **283**: 1482–1488.
- Wong, T.W. and Clayton, D.A. 1985. In vitro replication of human mitochondrial DNA: Accurate initiation at the origin of light-strand synthesis. *Cell* **42**: 951–958.
- Xia, X. 1996. Maximizing transcription efficiency causes codon usage bias. *Genetics* **144**: 1309–1320.
- . 2000. *Data analysis in molecular biology and evolution*. Kluwer Academic Publishers, Boston.
- Zardoya, R. and Meyer, A. 1997. The complete DNA sequence of the mitochondrial genome of a 'living fossil', the coelacanth (*Latimeria chalumnae*). *Genetics* **146**: 995–1010.
- . 1998. Complete mitochondrial genome suggests diapsid affinities of turtles. *Proc. Nat. Acad. Sci.* **95**: 14226–14231.

Received June 15, 2001; accepted in revised form July 31, 2001.