Nucleic Acids Research, Vol. 20, No. 18 4853–4858

The complete nucleotide sequence of the *Crossostoma lacustre* mitochondrial genome: conservation and variations among vertebrates

Chyng-Shyan Tzeng[1,2], Cho-Fat Hui[1,*], Shih-Chieh Shen[2] and P.C.Huang[1,3]
[1]Institute of Molecular Biology, Academia Sinica, Nankang, Taipei 11529, [2]Institute of Zoology, National Taiwan University, Taipei 10764, Taiwan, Republic of China and [3]Department of Biochemistry, School of Hygiene and Public Health, the Johns Hopkins University, 615 N. Wolf Street, Baltimore, MD 21205, USA

Received June 8, 1992; Revised and Accepted August 24, 1992 GenBank accession no. M91245

ABSTRACT

The complete mitochondrial (mt) genome of *Crossostoma lacustre*, a freshwater loach from mountain stream of Taiwan, has been cloned and sequenced. This fish mt genome, consisting of 16558 base-pairs, encodes genes for 13 proteins, two rRNAs, and 22 tRNAs, in addition to a regulatory sequence for replication and transcription (D-loop), is similar to those of the other vertebrates in both the order and orientation of these genes. The protein-coding and ribosomal RNA genes are highly homologous both in size and composition, to their counterparts in mammals, birds, amphibians, and invertebrates, and using essentially the same set of codons, including both the initiation and termination signals, and the tRNAs. Differences do exist, however, in the lengths and sequences of the D-loop regions, and in space between genes, which account for the variations in total lengths of the genomes. Our observations provide evidence for the first time for the conservation of genetic information in the fish mitochondrial genome, especially among the vertebrates.

INTRODUCTION

The conservation in amino acid sequences of most of the animal mitochondrial proteins with those from bacteria suggests a prokaryotic origin of mitochondria, and the dependence of mitochondrial proteins for cytosolic subunits encoded by nuclear genes indicates a symbiotic ancestry. However, there remains an enigma when one contemplates the variations exist among mitochondrial genomes. On the one hand, there is the extraordinary range in size, in gene organization, and in mode of expression. On the other hand, certain phyla, notably the vertebrates are highly parsimonious, confining their genetic information within a narrow range, 14 to 18 kb of DNA, in contrast to 250–2500 kb in the plants, 30–70 kb in fungi, and 18 to 78 kb in protozoa (1).

The complete nucleotide sequences have been determined in only a few of the metazoa mitochondrial genomes. Among them, mammalia [human *Homo sapiens* (2); domestic cow *Bos taurus* (3); house mouse *Mus domesticus* (4); rat *Rattus norvegicus* (5); fin whale *Balaenoptera physalus* (6)], aves [chicken *Gallus gallus domesticus* (7)], amphibian [*Xenopus laevis* (8)], sea urchins [*Strongylocentrotus purpuratus* (9); *Paracentrotus lividus* (10)], nematode worm [*Ascaris suum* and *Caenorhabditis elegans* (11)], and insect [fruit fly *Drosophila yakuba* (12)]. However, mtDNA sequences from two other major vertebrate groups, reptilia, and pisces are still lacking.

Fishes, the most primitive vertebrate, constitute the largest portion of vertebrates, yet only partial sequences have been obtained from some of their species such as Atlantic cod (13), carp (14,15,35), trout (16), salmon (17), and sturgeon (18,19).

In this study we have examined the whole mitochondrial genome of an indigenous Oriental stream loach, *Crossostoma lacustre* Steindachner (Homalopteridae, Pisces), with an attempt to extend our knowledge concerning the phylogeny of mtDNA.

C.lacustre is one of the four endemic species of Homalopterid fishes (20), small benthic, loach-like, living in the swift mountainous stream of Taiwan (21). This family of fishes is believed to have originated in the southern east of China, and from there dispersed throughout the southeast Asia (22,23). In the process, at least 100 species have evolved, and they have scattered widely extending from Peninsular India in the west to Taiwan in the northeast and from the islands of the Malay Archipelago in the south to Yangtze River basin in the northwest. Most of the homalopterid fishes are endemic to ecological niches within a given river system in which they are distributed. Due to their highly adaptative morphological features, a continuous evolutionary trend among this group may be expected which renders these fishes one of the most desirable for the study of evolutionary ichthyology (23).

In this paper, we report the complete nucleotide sequence of *C.lacustre* mt genome, 16558 bp in size, contrasting it with those of others for their genetic information and organization.

* To whom correspondence should be addressed

MATERIALS AND METHODS

Materials

Restriction endonucleases were purchased from New England Biolabs, T4 DNA ligase was from Boehringer Mannheim, and radioisotopes were from Amersham and Du Pont-New England Nuclear. Bluescript plasmid, exonuclease III, and mung bean nuclease were purchased from Stratagene. Sequenase kit (version 2.0), and T7 DNA polymerase sequencing system were purchased from United States Biochemical Corp. Oligonucleotides for sequencing primers and PCR primers were synthesized on a Pharmacia LKB Biotechnology Gene Assembler at the Institute of Molecular Biology, Academia Sinica, Nankang, Taipei.

Mitochondrial DNA isolation, cloning and subcloning

MtDNA was extracted from mature oocytes of a single individual of *C.lacustre* which was obtained from Dahu River, northern west of Taiwan, by using the rapid methods described previously (24,25). The crude mtDNA was further purified by twice subjecting it to isopynic ultracentrifugation of CsCl/ethidium bromide gradient (26). The supercoiled fish mitochondrial genome was digested into nine fragments with *Hind*III, each of which was cloned individually into Bluescript KS^+ plasmid vector and amplified using *E.coli* JM101 or JM109 as hosts. Each of the nine inserted fragments was characterized by restriction endonuclease analysis as reported previously, from which a physical map was constructed (25). Clones containing nested set of deletions were constructed from these inserts using the exonuclease III/mung bean nuclease technique (27,28). DNA sequencing of these clones was carried out as described below.

DNA sequencing

Nucleotide sequences were determined directly from double-stranded plasmid DNA (29,30) using the dideoxynucleotide chain-termination-method (31) with Sequenase in the sequencing reaction, and [α^{35}S]dATP for labeling. Eight pairs of PCR primers, each located two to three hundred bases apart at the junction of two neighbouring *Hind*III fragments, were designed. The PCR fragments generated were sequenced using that particular primer pair after a second asymmetric PCR (32,33).

Computer Analysis

The sequences were analyzed on VAX II using the software GCG (Genetics Computer Group, Version 6.1; University of Wisconsin) (34). Coding genes were identified by homology with known mitochondrial nucleotide sequences retrieved from the GenBank and EMBL collections by using GCG softwares of Strings and Fetch or the search program Fasta. The protein sequences were deduced by the GCG codon programs of mammalian mitochondrial DNA (2).

RESULTS AND DISCUSSION

Cloning and sequencing of *Crossostoma* mtDNA

A physical map of mtDNA of *Crossostoma* was constructed by using 14 restriction endonucleases (25). The nine *Hind*III mtDNA fragments, 3.1, 2.85, 2.32, 2.1, 1.8, 1.62, 1.44, 0.96 and 0.67 kb in length (Fig. 1), needed for sequencing were subcloned into the *Hind*III cloning site of Bluescript KS^+ plasmid vector, then sequenced. In order to confirm that no other small *Hind*III fragments had been missed, pairs of primers that located typically two to three hundred bases from each other at the junction of neighboring fragments were designed. By using PCR amplification (33), the fragments that covered the boundaries of neighboring fragments were generated and then sequenced directly from the products generated by a second asymmetric PCR (32). The sites of the primers are shown in figure 1. The results confirmed that no other sequences present between any neighbouring *Hind*III fragments.

The complete nucleotide sequence of the *C.lacustre* mtDNA (H-strand) has been submitted to the GenBank and EMBL databases with the accession number M91245.

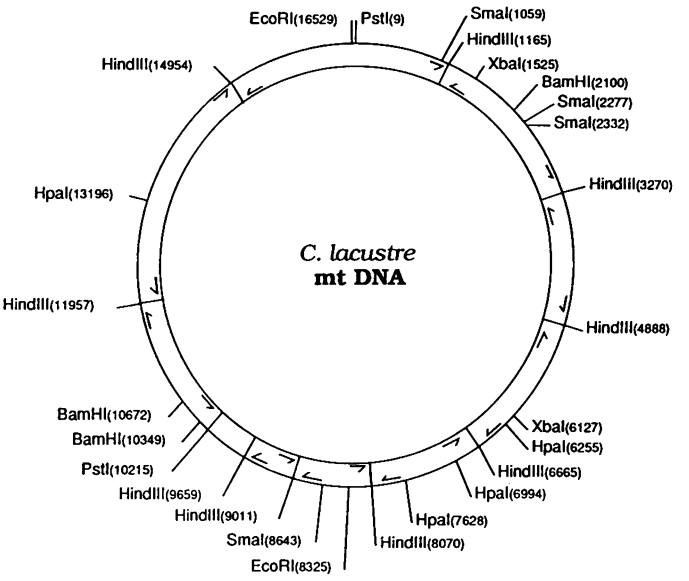

Figure 1. The physical map of mtDNA of *C.lacustre*. Restriction enzymes used are identical to those in our previous work (25). The arrows inside of the circle indicate the directions and the sites of the primers that were designed for PCR.

Table 1. Percentages of base composition of the H-strand of animal mitochondrial genomes.

	A	T	C	G	C/G
Human	30.9	24.7	31.2	13.1	2.38
Cow	33.4	27.2	25.9	13.5	1.92
Whale	32.7	26.7	27.3	13.3	2.05
Mouse	34.5	28.7	24.4	12.3	1.98
Rat	34.1	27.2	26.2	12.5	2.10
Chicken	30.5	23.8	32.5	13.5	2.41
Xenopus	33.1	30.0	23.5	13.5	1.64
Carp	31.7	25.0	27.5	15.8	1.74
Loach	29.4	25.1	28.6	16.9	1.69
Sea urchin	28.7	30.2	22.7	18.4	1.23
Drosophila	39.5	39.1	12.2	9.2	1.33

The percentages calculated here are based on the complete genome of each species. All of the genomes are obtained from the available sequences in the GenBank and EMBL databases (2–8,9,11,35), and sea urchin is represented by *S.purpuratus* (9).

Base composition

In *C.lacustre*, just as in the cases of carp (35) and fin whale (6), the most represented base in the H-strand is A, followed by C > T > G. In chicken and human, the most represented base is C instead of A, then followed by A > T > G (Table 1). In other

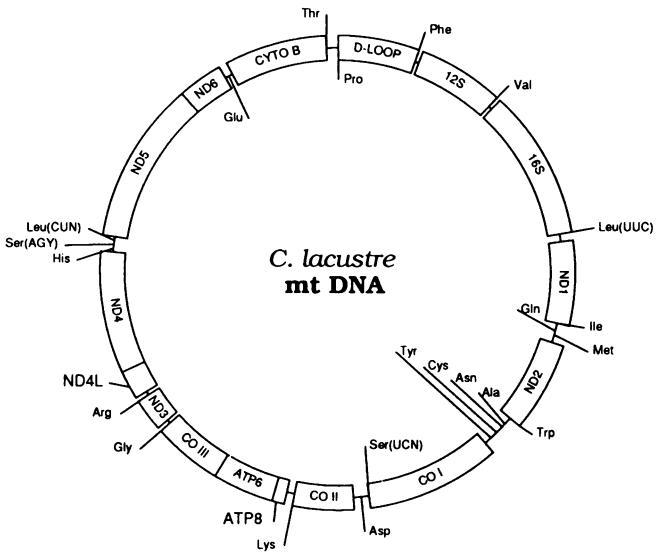

Figure 2. The genetic map of mt genome of *C.lacustre*. The protein-coding genes are shown as boxes. The tRNA genes are indicated according to their related positions. The tRNA genes encoded on the H-strand are shown on the outside of the circle, and the tRNA genes indicated in the inner side of the circle are encoded on the L-strand. The gene locations and direction of transcription are presented in more detail in Table 2. ND 1−6 and 4L: NADH dehydrogenase subunits 1−6 and 4L; CO I, CO II, and CO III: cytochrome c oxidase subunits I, II, and III; ATP 6, and ATP 8: ATPase subunits 6 and 8; CYTO B: cytochrome b; Phe, Val, Leu(UUC), Ile, Gln, Met, Trp, Ala, Asn, Cys, Tyr, Ser(UCN), Asp, Lys, Gly, Arg, His, Ser(AGY), Leu(CUN), Glu, Thr and Pro: tRNAs of those particular amino acids.

species such as cow, rat, mouse, *Xenopus, Drosophila,* and Sea urchin, the order is A>T>C>G (2−12). The mitochondrial genome of *C.lacustre* has a G-C content of 45.5%, and it falls within the range of G-C content of all other vertebrates with that of mouse (36%) being lowest, and that of Chicken (46%) being highest (Table 1). It is interesting to note that the C/G ratio of the H-strand of *C.lacustre* (1.69) is similar to poikilothermic *Xenopus* (1.64) and carp (1.74), lower than homeothermic vertebrates (1.92 in cow, 2.38 in human), and significantly higher than invertebrates (1.23 in sea urchin, 1.33 in *Drosophila*).

Genome organization

The total 16558 bp of the fish mtDNA contains the coding sequences for two rRNA genes (12S, and 16S), thirteen protein genes [seven respiratory chain NADH dehydrogenase components 1−6 and 4L (ND 1−6, and 4L); three cytochrome c oxidase subunits I, II, and III (CO I, CO II, and CO III); two ATPase subunits 6 and 8 (ATP 6, and ATP 8); and cytochrome b (Cyto b)], and 22 tRNAs genes for the mitochondrion's protein synthesizing system, as found in other animal mt genomes (2−12) (Fig. 2, Table 2). In invertebrates, a whole gene such as ATP 8 might have been lost in nematodes (11), or an entire mt genome was replaced by certain microbodies called hydrogenosomes (36).

The 12 protein genes and 15 tRNA genes in the *C.lacustre* mt genome are encoded on the H-strand, while only the ND 6 gene and the tRNA genes of Gln, Ala, Asn, Cys, Tyr, Ser(UCN), Glu and Pro are encoded on the L-strand (Fig. 2, Table 2). These gene arrangements are identical to those of vertebrates (2−8).

Various extents of overlapping between contiguous genes have been observed in vertebrate mitochondrial genomes: for instance

Table 2. The gene locations of the complete mt genome of *C.lacustre*.

Name of Gene	Location	Remark
D-loop	(1..896)	
tRNA Phe	(898..967)	H
12S rRNA	(968..1918)	H
tRNA Val	(1919..1990)	H
16S rRNA	(1991..36700)	H
tRNA Leu(UUC)	(3671..3744)	H
ND 1	(3746..4720)	H
tRNA Ile	(4728..4801)	H
tRNA Gln	(4799..4869) complement	L
tRNA Met	(4871..4939)	H
ND 2	(4940..5986)	H
tRNA Trp	(5985..6055)	H
tRNA Ala	(6058..6126) complement	L
tRNA Asn	(6128..6200) complement	L
tRNA Cys	(6231..6298) complement	L
tRNA Tyr	(6298..6365) complement	L
CO I	(6367..7917)	H
tRNA Ser(UCN)	(7920..7988) complement	L
tRNA Asp	(7992..8064)	H
CO II	(8078..8768)	H
tRNA Lys	(8769..8844)	H
ATP 8	(8846..9013)	H
ATP 6	(9004..9687)	H
CO III	(9687..10472)	H
tRNA Gly	(10472..10543)	H
ND 3	(10544..10894)	H
tRNA Arg	(10893..10962)	H
ND 4L	(10963..11259)	H
ND 4	(11253..12635)	H
tRNA His	(12634..12704)	H
tRNA Ser(AGY)	(12704..12773)	H
tRNA Leu(CUN)	(12774..12846)	H
ND 5	(12847..14683)	H
ND 6	(14682..15203) complement	L
tRNA Glu	(15205..15271) complement	L
Cyto b	(15278..16418)	H
tRNA Thr	(16419..16490)	H
tRNA Pro	(16489..16558) complement	L
BASE COUNT	4876 A 4743 C 2791 G	4148 T

The number 1 in gene location represents the first base of the 5′ end of the D-loop region in the H-strand. Abbreviations of protein names are identical to those used in Fig. 2. H or L on the remark column signifies that the indicated gene is transcripted from the H-strand or the L-strand respectively.

as many as nine nucleotides between CO I and tRNA-Ser (UCN) in chicken (7). In *C.lacustre*, intergenic spacers of variable lengths of 1 to 13 nucleotides are also found. A 'butt-joined' configuration is adopted between tRNA-Met and ND 2, tRNA-Asn and tRNA-Cys, tRNA-Gly and ND 3, tRNA-Arg and ND 4L, tRNA-Thr genes, tRNA-Ser, tRNA-Leu, and ND 5, and Cyto b and tRNA-Thr genes (Fig. 2, Table 2). In the cases of ND 2, ND 3, and ND 4, the last two bases of their stop codon TAG may serve as the initiation sequences of their downstream genes, tRNAs-Trp, Arg, and His (Fig. 2, Table 2). If so, these would be overlapping contiguous genes. However, if the first T of the TAG is a one base stop codon (37), then these genes could remain in a butt-joined configuration. Gene-overlap can be observed between the contiguous genes ATP 8 and ATP 6, and ND 4L and ND 4; they overlap by ten and seven bases, respectively (Fig. 2, Table 2).

Other gene overlaps in *C.lacustre* may include tRNA-Ile and tRNA-Gln, CO I and tRNA-Ser (UCN), ND 5 and ND 6, tRNA-Thr and tRNA-Pro, and they overlap by three, four, two and two bases, respectively. However, these overlapping genes are encoded on the opposite strands (Fig. 2, Table 2).

Table 3. The length variations in base pairs of mitochondrial genomes, and genes of all available mt genomes of vertebrates, sea urchin, and *Drosophila* (2–9,12,35).

Genes	Human	Bovine	Whale	Rat	Mouse	Chicken	Frog	Carp	Loach	*Drosophila*	Sea Urchin
D-LOOP	1043	910	929	898	879	1227	2134	927	896	1077	121
12S rRNA	954	955	957	955	975	819	951	951	789	867	976
16S rRNA	1559	1571	1574	1559	1582	1621	1631	1681	1680	1326	1530
ND 1	956	956	957	957	957	975	970	975	975	975	969
ND 2	1042	1042	1044	1038	1036	1038	1039	1047	1047	1025	1059
CO I	1541	1545	1551	1545	1545	1548	1549	1551	1551	1536	1554
CO II	684	684	684	684	684	684	688	691	691	685	690
ATP 8	207	201	192	204	204	165	168	165	168	162	168
ATP 6	679	681	681	681	681	681	679	684	684	674	690
CO III	784	784	786	783	784	783	781	786	768	789	783
ND 3	346	347	345	348	345	348	342	351	351	354	351
ND 4L	297	297	297	297	294	294	297	297	297	290	294
ND 4	1378	1378	1377	1380	1378	1377	1384	1383	1383	1339	1380
ND 5	1811	1821	1821	1836	1824	1818	1815	1824	1837	1720	1914
ND 6	528	528	528	519	519	519	513	519	522	525	495
CYTO B	1141	1137	1140	1143	1144	1140	1140	1141	1141	1137	1157
TOTAL	16569	16338	16398	16298	16295	16775	17553	16575	16558	16019	15650

Abbreviations of genes are identical to those used in Figure 2.

The order of arrangement of the two rRNA genes, 13 protein genes and 22 tRNA genes in *C.lacustre* is identical to those of *Xenopus* and Mammalians (Fig. 2). However, it differs from those of the galliform birds, in whose mitochondrial genomes the ND 6 and tRNA-Glu genes that are upstream of the Cyto b gene in the cases of *Crossostoma, Xenopus*, and Mammalians, are sandwiched between the tRNA-Pro gene and the D-loop (7). It also differs from those of the bull frog, *Rana catesbeiana* (38), and the tree frog, Rhacophoids (our unpublished data), whose mitochondrial tRNA genes of Thr, and Pro, are located between the D-loop and the tRNA gene of Leu. Our observation supports the notion that the mitochondrial gene arrangement among vertebrates is highly conserved, while, contrarily, the mt genomes of invertebrates seem to be subjected to high frequency of gene transposition (12,39,40). The majority of the protein-coding genes and both rRNA genes are punctuated by at least one tRNA gene which has been thought to act as a recognition signal(s) for mitochondrial RNA processing (7,41). However, from data just discussed above, it seems possible that the high frequency of gene transposition could be related to the major speciation events during the process of evolution.

Length of genes

The mt genomes of most vertebrates vary between 16.3–17.6 kb in length, and the *C.lacustre* mt DNA, which is 16558 bp in length, is quite similar with those of the mammalians (rat, 16298 bp; fin whale, 16398 bp; cow, 16338 bp; human, 16569 bp) and bird (chicken, 16775 bp), but shorter than that of *Xenopus* (17553 bp) by about 1 kb (2–8) (Table 3). However, in invertebrates, the mt genome lengths varied extensively (15–78 kb) (1), and yet in one species of *Drosophila*, two species of sea urchins, and two species of nematodes, their mt genomes are actually shorter than those of the vertebrates (9–12). In the cases of two species of nematodes, the shortened mt genomes are due to the absence of the ATP 8 gene (11), and in the cases of the two species of sea urchins, they are due to the shortened noncoding D-loop region (9,10), and in the case of *Drosophila yakuba*, it is due to the shortened 16S rRNA and ND 5 genes (12) (Table 3).

All of the 13 protein-coding genes, the two rRNA genes, and the D-loop sequence of the *C.lacustre* genome have been identified by their homologies to the similar regions in other mitochondrial genomes that are available in the GenBank and EMBL databases (1–12). Only small variations in the length of these gene sequences are observed (Table 3). Among genes, the longest length variation found was in the 16S rRNA of *Drosophila* (12), which was shorter than all other vertebrates and sea urchin by at least 210 bp (Table 3). The broadest length variations among all sequences examined occurred in the non-coding region, the D-loop, where it could vary from 121 bp in sea urchin *Strongylocentrotus purpuratus* (9) to 2134 bp in *Xenopus* (8) (Table 3). When gene length variations did occur in the protein coding sequences or the rRNA genes, they were found mostly in either ends of the genes. When insertion or deletion was found in the middle of a gene, such as the 27 bases insertion in the ND 6 gene of the human mt genome (aa118-aa126) (2), the same insertion in that gene seemed to occur to all members of that particular Class (2–6) (Table 3). The lengths of all other mt genes vary by less than ten bases among vertebrates (2–8) (Table 3). Finally, the absence of intron in the *C.lacustre* mtDNA sequences is consistent with all the available mtDNA sequences of all metazons (2–12) (Fig. 2, Table 2).

Translational initiation codons and termination codons

The first in-phase ATG codon is the translational initiation codon of all but one of the mt protein-coding genes of *C.lacustre*, just as in most of known mitochondrial genomes (2–12), while only GTG seems to be the initiation codon of the CO I gene (Table 4). Such a usage is identical to several identified mt genes of other species of fishes (13,35), as well as to chicken (7). Among the available sequence data of complete mt genomes from the databases, ATG seems to be the most common initiation codon for the mt-protein genes, exceptions are noted in which GTT, GTG, ATA, ATC, and ATT have also been proposed to be used as initiation codons (Table 4). In the cases of nematode mt protein genes, TTG and ATT seem to be more commonly use for initiations (11). In mammalians, ND 2, ND 3, ND 5, and several other mt-genes among invertebrates, ATN is used instead of ATG as initiation codon (9–12) (Table 4). Other more rarely used initiation codons include GTG in ND 1 of rodents (4,5), ND 5 of *Drosophila* (12), ATP 8 of sea urchins (9,10), CO I of chicken (7) and fishes (13,35), and ATAA of the *Drosophila* CO I gene

Table 4. The initiation codon usage of mitochondrial protein genes of all metazoa.

	ND1	ND2	ND3	ND4	ND5	ND6	ATP8	ATP6	COI	COII	COIII	CYTOB
Hum	---	ATT	ATA	---	ATA	---	---	---	---	---	---	---
Bov	---	ATA	ATA	---	ATA	---	---	---	---	---	---	---
Bpt	---	ATA	ATA	---	ATA	---	---	---	---	---	---	---
Rat	GTG	ATA	ATT	---	ATA	---	---	---	---	---	---	---
Mus	GTG	ATA	ATC	---	ATC	---	---	---	---	---	---	---
Ave	---	---	---	---	---	---	---	---	GTG	---	---	---
Xen	---	---	---	---	---	---	---	---	---	---	---	---
Fis	---	---	---	---	---	---	---	---	GTG	---	---	---
Sus	---	---	---	---	---	---	GTG	ATA	---	---	---	---
Dro	ATA	ATT	ATT	---	TTT	ATT	ATT	---	ATAA	---	---	---
Nem1	ATA	TTG	ATT	TTG	ATT	ATA	---	ATT	ATT	ATT	ATA	TTG
Nem2	TTG	TTG	TTG	TTG	ATT	TTG	---	ATT	ATT	TTG	GTT	ATT

--- represents ATG, which is the most common translational initiation codon of genes. Abbreviations of proteins are identical to those used in the text. Hum:Human, Bov:Bovine, Bpt:fin whale, Rat:rat, Mus:mouse, Ave:bird, Xen:*Xenopus*, Fis:*Crossostoma* and carp, Sus:sea urchin, Dro:Fruit fly, Nem1:Nematode *Caenorhabditis elegans*, Nem2:Nematode *Ascaris suum* (2–12,35).

Table 5. The genetic codes and codon usage in *C.lacustre* mitochondrial DNA.

AA	Cod	No	/1000	AA	Cod	No	/1000	AA	Cod	No	/1000	AA	Cod	No	/1000
Phe	TTT	114	29.91	Ser	TCT	37	9.71	Tyr	TAT	53	13.91	Cys	TGT	6	1.57
Phe	TTC	111	29.13	Ser	TCC	72	18.89	Tyr	TAC	63	16.53	Cys	TGC	20	5.25
Leu	TTA	81	21.25	Ser	TCA	69	18.11	End	TAA	8	2.10	Trp	TGA	98	25.72
Leu	TTG	29	7.61	Ser	TCG	8	2.10	End	TAG	3	0.79	Trp	TGG	22	5.77
Leu	CTT	115	30.18	Pro	CCT	20	5.25	His	CAT	19	4.99	Arg	CGT	6	1.57
Leu	CTC	107	28.08	Pro	CCC	104	27.29	His	CAC	84	22.04	Arg	CGC	25	6.56
Leu	CTA	237	62.19	Pro	CCA	74	19.42	Gln	CAA	91	23.88	Arg	CGA	39	10.23
Leu	CTG	59	15.48	Pro	CCG	15	3.94	Gln	CAG	10	2.62	Arg	CGG	10	2.62
Ile	ATT	159	41.72	Thr	ACT	49	12.86	Asn	AAT	35	9.18	Ser	AGT	7	1.84
Ile	ATC	126	33.06	Thr	ACC	127	33.32	Asn	AAC	84	22.04	Ser	AGC	39	10.23
Met	ATA	112	29.39	Thr	ACA	114	29.91	Lys	AAA	65	17.06	End	AGA	0	0.00
Met	ATG	55	14.43	Thr	ACG	12	3.15	Lys	AAG	13	3.41	End	AGG	0	0.00
Val	GTT	54	14.17	Ala	GCT	47	12.33	Asp	GAT	15	3.94	Gly	GGT	27	7.08
Val	GTC	61	16.01	Ala	GCC	177	46.44	Asp	GAC	64	16.79	Gly	GGC	55	14.43
Val	GTA	80	20.99	Ala	GCA	113	29.65	Glu	GAA	73	19.16	Gly	GGA	110	28.86
Val	GTG	29	7.61	Ala	GCG	16	4.20	Glu	GAG	25	6.56	Gly	GGG	59	15.48

The codon usage was constructed by using the codon frequency program with the mammalian mitochondrial GCG translation table (2). Frequencies of codon usage in thousand is calculated with all of the 13 mt-protein genes of *C.lacustre* mtDNA. Total number (No) of each type of codon found among the 13 mt-protein genes of *C.lacustre* mtDNA are indicated.

(12). *Xenopus* is the only species where all the mt-protein genes start with ATG (8). Apparently, frogs, fishes, and birds are identical in their mitochondrial translational initiation codon usage (7,8,13,35), except in the CO I gene, however, high variation in initiation codon usage is common among lower invertebrates (9–12) (Table 4).

While most of the *C.lacustre* mt protein-coding genes ended with the termination codon TAA, three of them, ND 2, ND 3, and ND 4 ended with TAG. It is interesting to note that in *C.lacustre*, the genes for CO II and Cyto b may end with a single base stop codon T. If this is the case, it would be similar to that of CO II in *Xenopus*, and Cyto b in human (8,2). The stop codons AGA and AGG are not used in *C.lacustre* (Table 5).

Genetic codes and codon usage

The genetic codes and codon usage in *C.lacustre* 13 mt-protein genes were identified by using the Codonfrequency program (GCG) with the translate codons of mammalians (2). Not only the genetic codes are highly conserved, the codon usage is similar to that of all of the vertebrate mtDNAs (Table 5). The only major variation was detected on the ending codons. In fish, codons ending in C are used most frequently (36.0%), then A (33.2%), T (20.8%), and G are rarely used (10.0%) (Table 5). However, codons ending in C and A are more frequently used in chicken (80%) and mammalians (72–77%), than in *C.lacustre* (69%) and *Xenopus* (65%) (2–8).

CONCLUSIONS

We report here in this paper the complete mitochondrial nucleotide sequence of *Crossostoma lacustre*. The genes it encodes, its gene organization, and codon usage are identical to those of other vertebrates. Although partial sequences of mt genome of several other fishes have been reported, this is the first instance in which the whole genome is studied. This information should fill a gap in our knowledge on the evolution of mt DNA.

As mt genome is considerably more variable in size and gene organization than has generally been recognized (42–44), differences among the closely related species usually can be detected in the non-coding sequences, particularly in the D-loop. This is demonstrated also in *C.lacustre*. Since variations in genome size and heteroplasmy that may have arisen from differences in copy-number of short, tandemly repeated

sequences, or by duplication of longer sequences of the entire control region or structural genes have hampered interpretation of results obtained solely by restriction endonuclease and RFLP analyses alone (45), sequence information on the entire genome would be tantamount to the appreciation of the complexity involved in mt DNA evolution. The availability of powerful techniques such as PCR used here has allowed the approach feasible.

Our study has extended the established list of twelve animal species showing that extreme economy in the coding of genetic information is characteristic to vertebrate mt genomes (37), including that of fish, we also have applied this knowledge to monitor alterations in the genome of *C.lacustre*, and other marine species under selection pressure, which may be a prelude to speciation. We have also carried out detailed phylogenetic analysis of the 13 proteins, two rRNAs, and 22 tRNAs encoded in the mt genome of higher eukaryotes. The results of these studies will be presented elsewhere.

ACKNOWLEDGEMENTS

We thank Dr. Igor Dawid for a plasmid clone of *Xenopus laevis* mtDNA used for homology assays in the early screening for clones used in this study. We are grateful to Dr. L.-Y.Lin, Dr. J.-R.Wu, Mrs. Y.-J.Chu and Dr. H.-C.Chi for their helpful discussions. We thank Dr. M.-T.Tam for his assistance in oligonucleotide synthesis, Mr. Cicero Chen for computer analyses. This work was supported by grants from Academia Sinica, and National Science Council (Grant No. NSC80-0209-B-002-05), Taiwan, Republic of China.

This paper is in partial fulfillment of the requirement for a PhD degree for C.S.Tzeng.

REFERENCES

1. Gray,M.W. (1989) Annu. Rev. Cell Biol., 5, 25–50.
2. Anderson,S., Bankier,A.T., Barrell,B.G., DeBruijn,M.H.L., Coulson,A.R., Drouin,J., Eperon,I.C., Nierlich,D.P., Roe,B.A., Sanger,F., Schreier,P. H., Smith,A.J.H., Staden,R. and Young,I.G. (1981) Nature, 290, 457–465.
3. Anderson,S., DeBruijn,M.H.L., Coulson,A.R., Eperon,I.C., Sanger,F. and Young,I.G. (1982) J. Mol. Biol., 156, 683–717.
4. Bibb,M.J., Van Etten,R.A., Wright,C.T., Walberg,M.W. and Clayton,D.A. (1981) Cell, 26, 167–180.
5. Gadaleta,G.,Pepe,G., Decandia,G.,Quagliariello,C., Sbisa,E. and Saccone, C. (1989) J. Mol. Evol., 28, 497–516.
6. Arnason,U., Gullberg,A. and Widegren,B. (1991) J. Mol. Evol., 33, 556–568.
7. Desjardins,P. and Morais,R. (1990) J. Mol. Biol., 212, 599–634.
8. Roe,B.A., Ma,D-P., Wilson,R.K. and Wong,J.F-H. (1985) J. Biol. Chem., 260, 9759–9774.
9. Jacobs,H.T., Elliott,D.J., Math,V.B. and Farquharson,A. (1988) J. Mol. Biol., 202, 185–217.
10. Cantatore,P., Roberti,M., Rainaldi,G., Gadaleta,M.N. and Saccone,C. (1989) J. Biol. Chem., 264(19), 10965–10975.
11. Okimoto,R., Macfarlane,J.L., Clay,D.O. and Wolstenholme,D.R. (1992) Genetics, 130, 471–498.
12. Clary,D.O. and Wolstenholme,D.R. (1985) J. Mol. Evol., 22, 252–271.
13. Johansen,S., Guddal,P.H. and Johansen,T. (1990) Nucleic Acids Res., 18(3), 411–419.
14. Araya,A., Amthauer,R., Leon,G. and Krauskopf,M. (1984) Mol. Gen. Genet., 196, 43–52.
15. Huang,C.J., Huang,F.L., Chang,Y.S., Tsai,Y.J. And Lo,T.B. (1990) Nucleic acids Res., 18, 1056.
16. Normark,B.B., McCune,A.R. and Harrison,R.G. (1991) Mol. Biol. Evol., 196, 43–52.
17. Davidson,W.S., Bartlett,S.E., Birt,T.P. and Green,J.M. (1988) Curr. Genet., 14, 483–486.
18. Gilbert,T.L., Brown,J.R., O'Hara,P.J., Buroker,N.E., Beckenbach,A.T. and Smith,M.J. (1988) Nucleic Acids Res., 16, 11825.
19. Brown,J.R., Gilbert,T.L., Kowbel,D.J., O'Hara,P.J., Buroker,N.E., Beckenbach,A.T. and Smith,M.J. (1989) Nucleic Acids Res., 17, 4389.
20. Tzeng,C.S. and Shen,S.C. (1982) Bull. Inst. Zool. Acad. Sinica, 21(2), 161–169.
21. Tzeng,C.S. (1986) Jour. Taiwan Mus., 39(2), 127–146.
22. Chen,Yi-Yu, (1980) Acta Zool. Sinica, 5(2), 200–211.
23. Sawada,Y., (1982) Mem. Fac. Fish. Hok. Univ., 28(2), 65–223.
24. Chapman,R.W. and Powers,D. (1984) A method for the rapid isolation of mitochondrial DNA from fishes. Tec. Rep. Maryl. Sea Grand Program. UM-SG-TS-84–05, pp. 1–11.
25. Tzeng,C.S., Shen,S.C. and Huang,P.C. (1990) Bull. Inst. Zool. Acad. Sinica, 29(1), 11–19.
26. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) Molecuar Cloning: A Laboratory manul. (2nd edition) Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
27. Henikoff,S. (1984) Gene, 28, 351–359.
28. Lawler,J. and Hynes,R.D. (1986) J. Cell Biol., 103, 1635–1648.
29. Chen,E.Y. and Seeburg,P.H. (1985) DNA, 4, 165–170
30. Chi,H.-C., Hsieh,J.-C., Hui,C.-F. and Tam,M.F. (1988) Nucleic Acids Res., 16, 10382.
31. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. USA, 74, 5463–5467.
32. Gyllensten,U.B. and Ehrlich,H.A. (1988) Proc. Natl. Acad. Sci. USA, 85, 7652–7655.
33. Saiki,R.K., Scharf,S., Faloona,F., Mullis,K.B., Horn,G.T., Erlich,H.A. and Arnheim,N. (1985) Science, 230, 1350–1354.
34. Devereux,J., Haeberli,P. and Smithies,O. (1984) Nucleic Acids Res., 12, 387–395.
35. Chang,Y.S. and Huang,F.L. (unpublished, personal communicatiion). X61010.
36. Sleigh,M.A. (1989) Protozoa and other protists. Edward Arnold, London.
37. Attardi,G. (1985) Int. Rev. Cytol., 93, 93–145.
38. Yoneyama,Y. (1987) Nippon Ika Daigaku Zasshi, 54, 429–440.
39. Wolstenholme,D.R., Macfarlane,J.L., Okimoto,R., Clary,D.O. and Wahleithner,J.A. (1987) Proc. Natl. Acad. Sci. USA 84, 1324–1328.
40. Cantatore,P. and Saccone,C. (1987) Int. Rev. Cytol., 108, 149–208.
41. Ojala,D., Montoya,J. and Attardi,G. (1981) Natur, 290, 470–474.
42. Avis,J.C. and Lansman,R.A. (1983) Polymorphism of mitochondrial DNA in populations of higher animals. In Nei,M. and Koehn,R.K. (ed), Evolotion of genes and proteins. Sunderland, Sinauer,pp. 147–164.
43. Bermingham,E., Lam,T. and Avise,J.C. (1986) J. Hered., 77, 249–252.
44. Moritz,C. and Brown,W.M. (1987) Proc. Natl. Acad. Sci. USA, 84, 7183–7187.
45. Moritz,C., Dowling,T.E. and Brown,W.M. (1987) Ann. Rev. Evol. Syst., 18, 269–292.