Journal of Biological Research-Thessaloniki **15**: 25 – 35, 2011
J. Biol. Res.-Thessalon. is available online at http://www.jbr.gr
Indexed in: WoS (Web of Science, ISI Thomson), SCOPUS, CAS (Chemical Abstracts Service) and DOAJ (Directory of Open Access Journals)

The mitochondrial genome of the European catfish *Silurus glanis* (Siluriformes, Siluridae)

Spiros VITTAS, Elena DROSOPOULOU, Ilias KAPPAS, Chrysoula N. PANTZARTZI and Zacharias G. SCOURAS*

Department of Genetics, Development & Molecular Biology, School of Biology, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

Received: 9 April 2010 Accepted after revision: 21 May 2010

The catfish *Silurus glanis* is the second largest freshwater fish in Europe. Despite its great commercial value and importance in historical biogeography, data concerning its nuclear and mitochondrial genome are very limited. The complete sequence of the mitochondrial genome of *S. glanis*, the first from the family Siluridae, is presented here. It has a length of 16526 base pairs (bp), containing 37 genes (13 protein-coding, 22 tRNAs, 2 rRNAs) and a control region, all organized in the same order and direction with the majority of vertebrate mtDNAs. Bayesian, parsimony, and maximum likelihood analyses, using all seven available siluriform mitogenomes, provided solid evidence for the specific position of *S. glanis* close to the early diversification of Siluriformes. The sequence of *S. glanis* mtDNA will likely offer new DNA markers useful for phylogenetic investigations concerning the complex evolutionary relationships within Siluriformes.

Key words: mitochondrial DNA, wels catfish, historical biogeography, protein-coding genes, complete sequence.

INTRODUCTION

The order Siluriformes, known as catfishes, constitutes an exceptionally diverse and speciose natural group of primarily freshwater ray-finned fishes (Nelson, 1994). Currently, 36 families and over 3000 species are recognized (Ferraris, 2007) rendering catfishes among the most diverse vertebrate orders (approximately 1 in 10 actinopterygians or 1 in 20 vertebrates is a catfish). Morphological (Fink & Fink, 1981) as well as molecular data (Saitoh *et al.*, 2003; Sullivan *et al.*, 2006) clearly support the monophyly of Siluriformes and place the main burst of their diversification at the late Cretaceous-early Tertiary boundary (Hardman, 2005). Despite the diversity of catfishes, interfamilial relationships still remain controversial and not fully resolved. Phylogenetic investigations have been stymied by the relative lack of dense sampling for certain mitochondrial and nuclear sequence data. For example, complete mitochondrial genomes have been determined for only six siluriform species, namely *Pseudobagrus tokiensis* (Saitoh *et al.*, 2003), *Corydoras rabauti* (Saitoh *et al.*, 2003), *Ictalurus punctatus* (Waldbieser *et al.*, 2003), *Cranoglanis bouderius* (Peng *et al.*, 2006), *Pangasianodon gigas* (Jondeung *et al.*, 2007), and *Liobagrus obesus* (Kartavtsev *et al.*, 2007).

The European catfish *Silurus glanis* (Siluridae) is the second largest freshwater fish in Europe after the European sturgeon (Kottelat & Freyhof, 2007). Its natural distribution extends from the Aral Sea basin to the Danube and Vistula River basins, and southward to Greece and western Anatolia (Banarescu, 1989). The wels catfish (*S. glanis*) is of considerable commercial importance, particularly for central and eastern European countries, due to several characteristics that make it desirable for profitable aquaculture (Legendre *et al.*, 1996; Proteau *et al.*, 1996; Smitherman *et al.*, 1996; Brzuska & Adamek, 1999). Few studies exist on the genetic variability and phylogeography of *S. glanis* populations (Krieg *et al.*, 1999; Triantafyllidis *et al.*, 1999a, b, 2002). Throughout the

* Corresponding author: tel.: +30 2310 998324, fax: +30 2310 998298, e-mail: scouras@bio.auth.gr

native range of the species, the high levels of interpopulation differentiation reported do not seem to be geographically patterned. Results based on mitochondrial DNA data have identified the Ponto-Caspian region as the most genetically diverse. This has been interpreted as an indication that this region was the single *S. glanis* late Pleistocene refuge and the main source of postglacial colonization of European lineages (Krieg *et al.*, 1999; Triantafyllidis *et al.*, 2002). Considering the commercial importance of *S. glanis* and the presence of several transplanted strains in northern and western Europe, additional data are imperatively needed. The development of novel molecular markers is expected to provide new insights into the genealogical history of *S. glanis* and the suspected ecological impacts of introductions as well as to contribute to genetic management and conservation efforts. In this vein, the mitochondrial DNA genome is an indispensable source of information.

Animal mitochondrial DNA (mtDNA) is a double-stranded circular DNA molecule, ranging in length from 16 to 20 kb (Avise *et al.*, 1987). It generally consists of intronless genes, few, if any, intergenic spacers, and one non-coding region of approximately 1 kb, the control region (Displacement loop: D-loop) which contains a number of important features regulating mtDNA replication and transcription (Clayton, 1982, 1991; Fernandez-Silva *et al.*, 2003). All the mtDNA encoded polypeptides are functionally essential since they are components of the mitochondrial energy-generating pathway involved in oxidative phosphorylation. Furthermore, the almost exclusive maternal inheritance, the lack of recombination in general, and the high substitution rate make this molecule very attractive for phylogenetic studies (Nei, 1987; Pissios & Scouras, 1993; Nikolaidis & Scouras, 1996; Peng *et al.*, 2006).

Up to date, higher-level phylogenetic analyses in Siluriformes have used either single mtDNA genes or concatenated protein-coding gene sequences, with variable taxonomic coverage (Saitoh *et al.*, 2003; Waldbieser *et al.*, 2003; Hardman, 2005; Peng *et al.*, 2006; Jondeung *et al.*, 2007; Kartavtsev *et al.*, 2007). The most extended study so far has been that of Hardman (2005) who used cytochrome *b* sequence data to investigate phylogenetic relationships among 170 siluriform species from 29 extant families and also explore the puzzling genealogical connections between North American ictalurids and Southeast Asian relatives.

In the present study, we report the complete mitochondrial genome of *S. glanis*, the first of the family Siluridae. Using all available siluriform complete mitogenomes, we additionally explore the evolutionary relationships of *S. glanis* to its relatives and identify certain lineage relationships critical in the global diversification of catfishes.

MATERIALS AND METHODS

Genomic and mitochondrial DNA extraction

The nomenclature in this paper follows that of FishBase (www.fishbase.org). One *S. glanis* individual was collected from Kastoria Lake in northern Greece and was used for both mitochondrial and genomic DNA extraction. Mitochondrial DNA was extracted from liver tissue using the extraction protocol described in Arnason *et al.* (1991), while genomic DNA was extracted from alcohol-fixed muscle tissue using the NucleoSpin Tissue kit protocol (Macherey-Nagel).

Direct and PCR cloning and sequencing of mtDNA

In order to obtain the complete mtDNA sequence of *S. glanis*, direct and PCR cloning were performed. A 2385-bp *Hin*cII mtDNA fragment was cloned into a pGEM T-Easy Vector following the manufacturer's recommended protocol (Promega). The rest of the molecule was PCR-amplified using homologous and heterologous (based on closely related species) sets of primers (Table 1, Fig. 1) and cloned in either a pGEM T-Easy Vector (Promega) or a pDrive vector system (QIAGEN) following conditions recommended by each manufacturer. PCR amplifications were generally performed using cycling conditions reported by Kartavtsev *et al.* (2007). Purification of all PCR products was performed either with the QIAquick PCR purification kit or the QIAquick Gel Extraction kit (QIAGEN). Plasmid DNA isolation was performed using the NucleoSpin‟ Plasmid kit (Macherey-Nagel). Double-stranded sequencing was performed by an automatic sequencer (Lark Technologies, Essex or Macrogen, Seoul).

The complete mitochondrial genome of *S. glanis* was deposited in GenBank (accession number: AM398435).

Sequence analysis

All obtained sequences were initially subjected to BLAST (Altschul *et al.*, 1990) searches for verification of homology. Protein-coding genes, rRNAs, and tRNAs were identified by comparison with the corresponding known sequences of six other catfish species

TABLE 1. Sequences of primers used in the present study. Primers marked with * were designed using the sequences of six Siluriformes species (GenBank acc. nos: AB054127, AB054128, AF482987, AY898626, DQ321752, AY762971). Primers marked with § were designed using *S. glanis* sequences (AJ969127, AM113592). All other primers were designed using *S. glanis* sequences retrieved from this study

Primer name	**Sequence 5′-3′**
ATP6F (1F)	CTG CCA TAT ACC TTC ACA CCT ACA ACA CAG C
Cytb92R§ (1R)	ATT ATG AGA CAT AGT AGG AGG AGG GAG CCA AAG
DloopF§ (2F)	AGC GAG TCC TAA TTA ATC CTG TCA AAC CCC
COXI473R (2R)	GTT TGT GGT AGT TAG TTC AAC GGA GAG CAC TTC
Cytb442F§ (3F)	TAT TAT CCG CCG TTC CCT ACA TAG GAG ATG
16SR* (3R)	GCA TTA CAG ATA GAA ACT GAC CTG GAT TGC
16SF* (4F)	AAT GAA GAC CTG TAT GAA TGG TGG AAC GAG G
COXI203R (4R)	TAA AAA GGG GGA ATC AGT GAA CGA AAG CTC C
COXIIINF (5F)	CTG ATA GAA GGG GAA CGC AAA C
ND4R* (5R)	CCT CAT GGG GTT TGA ATT AAG ATT CC
ND4F (6F)	TAC TCG CTG GGG CAA TCA AGC
CytGluR (6R)	GGG TGG GTT TTT CGG GTT AC

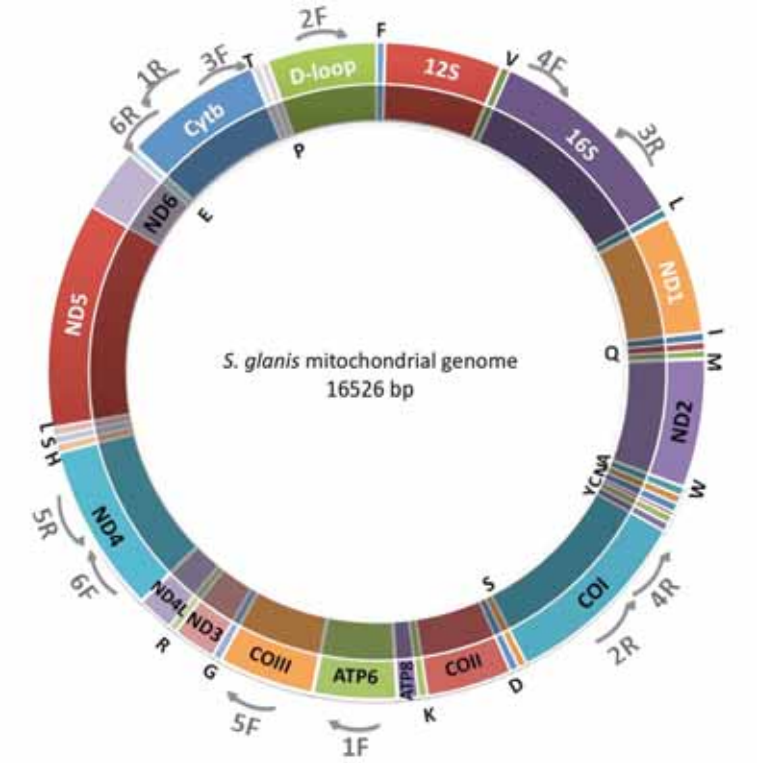

FIG. 1. *Silurus glanis* mitochondrial genome. Genes for proteins and rRNAs are shown with standard abbreviations. Genes for tRNAs are designated by a single letter for the corresponding amino acid. Genes shown at the outer circle are encoded by the H-strand while those at the inner circle are encoded by the L-strand. The arrows represent the primers used for mtDNA amplification (see also Tables 1 and 3).

TABLE 2. Species information of complete mtDNA sequences used in phylogenetic analysis

Species	mtDNA length (bp)	Family	GenBank accession number	Reference
Pseudobagrus tokiensis	16529	Bagridae	AB054127	Saitoh *et al.* (2003)
Corydoras rabauti	16831	Callichthyidae	AB054128	Saitoh *et al.* (2003)
Ictalurus punctatus	16497	Ictaluridae	AF482987	Waldbieser *et al.* (2003)
Cranoglanis bouderius	16539	Cranoglanididae	AY898626	Peng *et al.* (2006)
Pangasianodon gigas	16533	Pangasiidae	AY762971	Jondeung *et al.* (2007)
Liobagrus obesus	16531	Amblycipitidae	DQ321752	Kartavtsev *et al.* (2007)
Silurus glanis	16526	Siluridae	AM398435	this study
		Outgroups		
Salmo salar	16665	Salmonidae	U12143	Hurst *et al.* (1999)
Clupea pallasii	16700	Clupeidae	AP009134	Lavoué *et al.* (2007)

(*Liobagrus obesus*, *Ictalurus punctatus*, *Pangasianodon gigas*, *Corydoras rabauti*, *Pseudobagrus tokiensis*, *Cranoglanis bouderius*) (Table 2) through alignments with ClustalX 1.8 (Thompson *et al.*, 1997). Furthermore, the 13 mitochondrial protein-coding genes were defined by the presence of initiation and stop codons whereas the 22 tRNAs were identified by their tendency to fold into cloverleaf secondary structures (Zuker *et al.*, 1999) and by the presence of specific anticodons. Codon usage was determined using the Sequence Manipulation Suite (SMS2) (www.bioinformatics.org/sms2/) while nucleotide frequency at each codon position was determined using the MEGA 4.0 package (Kumar *et al.*, 2008).

Phylogenetic analysis

The complete mtDNA dataset consisted of all six available mitochondrial genomes of Siluriformes, the mitochondrial genome of *S. glanis* reported in this study, and two outgroup taxa (*Salmo salar*, *Clupea pallasii*) (see Table 2). The METAMiGA database (METAzoan Mitochondrial Genomes Accessible database, http://amiga.cbmeg.unicamp.br/) was used to assemble the mtDNA dataset. The ND6 gene was excluded from the analysis due to its consistent translocation on the L-strand. Nucleotide sequences of individual protein-coding genes were retrieved and grouped in 12 separate files. Sequences were aligned using the RevTrans 1.4 server from DTUCBS (Wernersson & Pedersen, 2003) and subsequently concatenated to produce a final alignment of 10932 nucleotide and 3631 amino acid positions. Phylogenetic analysis was performed on both nucleotide and amino acid alignments under Bayesian inference (BI) in MrBayes 3.1 (Huelsenbeck & Ronquist, 2001) as well as under parsimony (MP) and likelihood (ML) methods in PAUP* 4.0b10 (Swofford, 1998).

For BI, the best-fit nucleotide substitution model employed was the general time reversible model GTR+I+Γ (Tavare, 1986) as determined by Modeltest 3.7 (Posada & Crandall, 1998). Two independent, simultaneous analyses were run for 2×10^6 generations, each starting from different random trees with four chains (one cold and three incrementally heated) and sampling every 200 generations. We used the AWTY software (Nylander *et al.*, 2008) to graphically check convergence of the Markov chain Monte Carlo by monitoring cumulative posterior split probabilities and between-run variability of split frequencies. Stationarity was confirmed after 2500 sampled generations which were discarded as "burn-in". A majority-rule consensus topology was created with the remaining samples, pooled together from the independent runs. The frequencies of each node of the consensus tree were represented as posterior probabilities.

For MP, trees were generated using heuristic searches with TBR (tree-bisection-reconnection) branch swapping and 500 random taxon additions. Nodal support was assessed with 1000 bootstrap replicates. For ML, the best-fit substitution model, determined previously, was employed in a heuristic search with TBR branch swapping and 100 bootstrap replicates. Phylogenetic analysis was also performed on the corresponding amino acid dataset using MrBayes 3.1 (Huelsenbeck & Ronquist, 2001). We used ProtTest 2.4 (Abascal *et al.*, 2005) and selected mtREV (Ada-

chi & Hasegawa, 1996) as the best-fit amino acid substitution model. Parameter settings during Bayesian inference were as previously described.

Prior to phylogeny reconstruction, we tested for homogeneity of base frequencies using the χ^2 heterogeneity test implemented in PAUP* 4.0b10 (Swofford, 1998). We also inspected for substitution saturation by plotting transitions/transversions to corrected sequence divergence using DAMBE 5.0 (Xia & Xie, 2001).

RESULTS AND DISCUSSION

The complete nucleotide sequence of the European catfish *S. glanis* mtDNA was obtained and deposited in GenBank (acc. no. AM398435). The total length of the L-strand is 16526 bases with an overall base composition of A: 30.2%, C: 28.8%, G: 16.1%, and T: 24.9%. The *S. glanis* mitochondrial gene order is identical to that of other fishes (Miya *et al.*, 2003) and higher vertebrates (Boore, 1999). The mtDNA molecule consists of two rRNAs (12S and 16S rRNA), 22 tRNAs, 13 protein-coding genes, and a major non-coding sequence, the control region (Fig. 1, Table 3). Sequence homology between *S. glanis* and the six available siluriform mtDNAs (*I. punctatus*, *C. rabauti*, *P. tokiensis*, *P. gigas*, *C. bouderius*, *and L. obesus*) appears to be relatively high, ranging from 72 to 89% for the protein-coding genes, from 84 to 90% for the rRNA genes, and from 63 to 71% for the control region.

Ribosomal and transfer RNA genes

The *S. glanis* 12S and 16S rRNA genes consist of 954 (positions: 71-1024) and 1679 (positions: 1097-2775) nucleotides, respectively (Table 3). As in other vertebrates (Inoue *et al.*, 2000; Broughton *et al.*, 2001), these genes are located between those for $tRNA^{Phe}$ and $tRNA^{Leu}$ (UUR) and are separated by the $tRNA^{Val}$ gene. The *S. glanis* mitochondrial genome contains 22 tRNAs (14 encoded by the H-strand and 8 by the L-strand) which are dispersed along the genome, range in size from 66 to 75 nucleotides (Table 3), and are predicted to fold into the expected cloverleaf secondary structure (data not shown). The $tRNA^{Ile}$ and $tRNA^{Gln}$ genes as well as the $tRNA^{Gln}$ and $tRNA^{Met}$ genes overlap by one nucleotide each. This is quite common in other fish mitochondrial genomes, differing, however, in the number of the overlapping nucleotides (Lee *et al.*, 2001).

Protein-coding genes

All protein-coding genes generally found in other vertebrate mtDNAs are also present in *S. glanis* mitochondrial genome with the same arrangement. They are all encoded by the H-strand, with the exception of the ND6 gene encoded by the L-strand (Table 3). Restricted overlapping is observed among some of the 13 protein-coding genes (Table 3). In particular, the ATP8 and ATP6 genes share ten nucleotides, as in birds and other fishes, the ATP6 and COIII genes overlap by one nucleotide, as in other teleosts, the ND4L and ND4 genes overlap by seven nucleotides, as in all other chordates, and the ND5 and ND6 genes have an overlap of four nucleotides, as in most vertebrates (4-17 nucleotides) (Broughton *et al.*, 2001; Peng *et al.*, 2006).

All protein-coding genes use ATG as a start codon except COI that uses GTG. Six protein-coding genes end with TAA and three with TAG. The COII, COIII, ND4, and Cytb genes do not possess proper stop codons but do show a terminal T (Table 3) which likely is completed by post-transcriptional polyadenylation (Ojala *et al.*, 1981). This is quite common and typical among mtDNA genes of other fishes (Lee *et al.*, 2001; Kim *et al.*, 2004; Nagase *et al.*, 2005).

The nucleotide frequencies for every codon position (Table 4) in all 13 protein-coding genes of *S. glanis* mtDNA reveal no bias at the 1st codon position. Bias against A and G and towards T is detected at the 2nd position. Similarly, a strong bias towards A and C and against G at the 3rd codon position, typical in vertebrates (Broughton *et al.*, 2001), is also evident in *S. glanis*. The strong bias towards A and C at the 3rd codon position is probably due to the fact that the 22 tRNAs encoded by the mtDNA are specific for codons which have A or C at the 3rd position (Table 3) (Broughton *et al.*, 2001).

Regarding codon usage (Table 5), a strong bias towards A and C at the 3rd position is observed in all codons except for isoleucine, something similar to the mtDNA of the channel catfish *I. punctatus* (Waldbieser *et al.*, 2003). Regarding amino acids with fourfold degenerate third positions, the A-ending codons are not the most frequent in all cases, in contrast to the general pattern observed in other vertebrates (Broughton *et al.*, 2001). In addition, codons for which the anticodons are present in the mitochondrial tRNA genes are generally preferred (Table 3).

TABLE 3. Organization of the *S. glanis* mitochondrial genome

Gene/Element	Abbreviation	Strand	Position	Size (bp)	Start codon	Stop codon	Codon recognized	Amino acids
tRNA^{Phe}	F	H	1-70	70			TTC	
12S ribosomal RNA	12S	H	71-1024	954				
tRNA^{Val}	V	H	1025-1096	72			GTA	
16S ribosomal RNA	16S	H	1097-2775	1679				
tRNA^{Leu} (UUR)	L	H	2776-2850	75			TTA	
NADH dehydrogenase subunit 1	ND1	H	2851-3825	975	ATG	TAA		324
tRNA^{Ile}	I	H	3827-3898	72			ATC	
tRNA^{Gln}	Q	L	3898-3968	71			CAA	
tRNA^{Met}	M	H	3968-4036	69			ATG	
NADH dehydrogenase subunit 2	ND2	H	4037-5083	1047	ATG	TAG		348
tRNA^{Trp}	W	H	5082-5152	71			TGA	
tRNA^{Ala}	A	L	5155-5223	69			GCA	
tRNA^{Asn}	N	L	5225-5297	73			AAC	
tRNA^{Cys}	C	L	5332-5397	66			TGC	
tRNA^{Tyr}	Y	L	5401-5470	70			TAC	
Cytochrome c oxidase subunit 1	COI	H	5472-7022	1551	GTG	TAA		516
tRNA^{Ser} (UCN)	S	L	7023-7093	71			TCA	
tRNA^{Asp}	D	H	7098-7169	72			GAC	
Cytochrome c oxidase subunit 2	COII	H	7184-7874	691	ATG	T-*		230
tRNA^{Lys}	K	H	7875-7948	74			AAA	
ATP synthase F0 subunit 8	ATP8	H	7950-8117	168	ATG	TAA		55
ATP synthase F0 subunit 6	ATP6	H	8108-8791	684	ATG	TAA		227
Cytochrome c oxidase subunit 3	COIII	H	8791-9574	784	ATG	T-*		261
tRNA^{Gly}	G	H	9575-9647	73			GGA	
NADH dehydrogenase subunit 3	ND3	H	9648-9998	351	ATG	TAG		116
tRNA^{Arg}	R	H	9997-10067	71			CGA	
NADH dehydrogenase subunit 4L	ND4L	H	10068-10364	297	ATG	TAA		98
NADH dehydrogenase subunit 4	ND4	H	10358-11738	1381	ATG	T-*		460
tRNA^{His}	H	H	11739-11808	70			CAC	
tRNA^{Ser} (AGY)	S	H	11809-11874	66			AGC	
tRNA^{Leu} (CUN)	L	H	11878-11950	73			CTA	
NADH dehydrogenase subunit 5	ND5	H	11951-13777	1827	ATG	TAA		608
NADH dehydrogenase subunit 6	ND6	L	13774-14292	519	ATG	TAG		172
tRNA^{Glu}	E	L	14293-14361	69			GAA	
Cytochrome b	Cytb	H	14363-15500	1138	ATG	T-*		379
tRNA^{Thr}	T	H	15501-15571	71			ACA	
tRNA^{Pro}	P	L	15570-15639	70			CCA	
Displacement loop (control region)	D-loop	–	15640-16526	887				

* TAA stop codon is completed by the addition of 3′A residues to mRNA

TABLE 4. Nucleotide frequencies (%) by codon position over all mitochondrial protein-coding genes of *S. glanis*

	T	C	A	G
1st position	21.4	26.2	26.6	25.8
2nd position	40.5	27.3	18.8	13.4
3rd position	18.5	34.6	38.0	8.9
Total	26.8	29.3	27.8	16.1

TABLE 5. Codon usage in *S. glanis* mitochondrial protein-coding genes

Amino acid	Codon	Number	Frequency	Codon usage (%)
Ala	GCG	15	0.004	4.5
	GCA	116	0.030	34
	GCU	52	0.014	15.5
	GCC	154	0.040	46
Cys	UGU	10	0.003	31
	UGC	22	0.006	69
Asp	GAU	20	0.005	26
	GAC	57	0.015	74
Glu	GAG	18	0.005	18
	GAA	82	0.022	82
Phe	UUU	90	0.024	41
	UUC	130	0.034	59
Gly	GGG	50	0.013	21
	GGA	87	0.023	36
	GGU	18	0.005	7
	GGC	86	0.023	36
His	CAU	27	0.007	24
	CAC	84	0.022	76
Ile	AUU	160	0.042	57
	AUC	121	0.032	43
Lys	AAG	6	0.002	7
	AAA	75	0.020	93
Leu (UUR)	UUG	27	0.007	19
	UUA	116	0.030	81
Leu (CUN)	CUG	55	0.014	11
	CUA	252	0.066	51
	CUU	80	0.021	16
	CUC	109	0.029	22
Met	AUG	62	0.016	36
	AUA	112	0.029	64
Asn	AAU	39	0.010	32
	AAC	84	0.022	68
Pro	CCG	10	0.003	5
	CCA	80	0.021	37
	CCU	28	0.007	13
	CCC	97	0.025	45
Gln	CAG	16	0.004	16
	CAA	85	0.022	84
Arg	CGG	10	0.003	14
	CGA	39	0.010	53
	CGU	10	0.003	14
	CGC	14	0.004	19
Ser (AGY)	AGU	9	0.002	19
	AGC	38	0.010	81
Ser (UCN)	UCG	5	0.001	3
	UCA	70	0.018	38.5
	UCU	29	0.008	16
	UCC	77	0.020	42.5
Thr	ACG	12	0.003	4
	ACA	137	0.036	45
	ACU	45	0.012	14.5
	ACC	112	0.029	36.5
Val	GUG	29	0.008	13
	GUA	91	0.024	40
	GUU	49	0.013	21.5
	GUC	58	0.015	25.5
Trp	UGG	22	0.006	19
	UGA	95	0.025	81
Tyr	UAU	38	0.010	34
	UAC	73	0.019	66
Stop	AGG	0	0.000	0
	AGA	0	0.000	0
	UAG	3	0.001	23
	UAA	10	0.003	77

Non-coding regions and D-loop

The non-coding regions of vertebrate mtDNAs are restricted to few spacers between tRNA genes, the origin of light strand replication (O_L), and the control region that regulates replication and transcription (Clayton, 1982, 1991; Shadel & Clayton, 1993, 1997). In the *S. glanis* mitochondrial genome several short spacers (1-4 nucleotides long) between tRNA genes were identified, along with few spacers (1-14 nucleotides long) between tRNA and protein-coding genes (Table 3). The origin of light strand replication (O_L) was identified as a 37-bp sequence motif (5′-CTT TCC CCG CCT GTT CGT CGA ATA AAG GCG GGG AAA G-3′) inside a cluster of five tRNA genes (WANCY region) located between the $tRNA^{Asn}$ and $tRNA^{Cys}$ genes (sharing a 3-bp overlap with the latter, positions: 5298-5331, Table 3). This region has the potential to form a stable stem-loop structure, consisting of 24 bp in the stem region and 13 bp in the loop region which is common and crucial for the light strand replication of vertebrate mtDNA (Shadel & Clayton, 1997).

The length of the *S. glanis* control region is 887 bp (positions: 15640-16526, Table 3), located between the $tRNA^{Pro}$ and $tRNA^{Phe}$ genes. Its size is similar to that of most Siluriformes (Waldbieser *et al.*, 2003; Peng *et al.*, 2006; Jondeung *et al.*, 2007) and comparable to the control regions of other teleosts (Lee *et al.*, 1995). Several conserved sequence blocks present in most vertebrates (Walberg & Clayton, 1981; Sbisa *et al.*, 1997) were also identified in *S. glanis* control region and are presented in Table 6.

TABLE 6. Control region conserved boxes (Siluriformes) and their positions in *S. glanis* mitochondrial DNA

Conserved box	Position in *S. glanis* control region
TAS-1	15683-15698
TAS-2	15717-15732
TAS-3	15736-15750
ETAS	15683-15750
CSB-F	15936-15956
CSB-E	15997-16015
CSB-D	16030-16057
CSB-C	16089-16113
CSB-B	16162-16184
CSB-1	16267-16285
CSB-2	16387-16404
CSB-3	16430-16449
O_H-like	16127-16139

Phylogenetic inference

The mtDNA dataset (12 protein-coding genes) consisted of 10932 nucleotide positions of which 3512 were parsimony informative and 5103 were variable. Significant heterogeneity in base frequencies was found ($\chi^2 = 250.99$, df = 24, $p < 0.01$) while substitution saturation was only evident at third-codon positions. The reconstructed topologies applying BI, MP, and ML methods were overall congruent and robust. All groupings were faithfully retained except for the position of *I. punctatus*. The inferred BI phylogeny is shown in Figure 2. The lineage leading to *S. glanis* (Siluridae) branches off the clade consisting of *L. obesus* (Amblycipitidae), *P. tokiensis* (Bagridae), *I. punctatus* (Ictaluridae), *C. bouderius* (Cranoglanididae), and *P. gigas* (Pangasiidae). The specific position of *S. glanis* is particularly stable and is also recovered during MP and ML analyses (data not shown). Given the limited taxon coverage of the mitogenome dataset, our analysis provides further support for the likely position of the family Siluridae at the early stages of diversification of the order (see also Peng *et al.*, 2006; Jondeung *et al.*, 2007; Kartavtsev *et al.*, 2007). Interestingly, the branching order of *I. punctatus* seems to be somewhat variable across analyses. In BI and ML, *I. punctatus* comes out as equally related to *C. bouderius* and the Mekong giant catfish *P. gigas* while in MP it is recovered as a sister species to *C. bouderius*. In a previous investigation based on parsimony and Bayesian analyses of cytochrome *b* sequences (Hardman, 2005), *C. bouderius* was recovered as a sister taxon to Ictaluridae. On the other hand, Ictaluridae and Pangasiidae were inferred as closest relatives during an analysis of 60 catfish species based on 16S rRNA data (Kartavtsev *et al.*, 2007). However, Cranoglanididae had not been sampled in that study. Considering the enigmatic relationship of North American ictalurid catfishes to other catfish families, the current mitogenome analysis upholds previous results showing that North American freshwaters were invaded by the ancestor of Ictaluridae in the late Cretaceous through a Northeastern Asia-Northwestern North America connection (Hardman, 2005). Furthermore, our results strongly highlight *C. bouderius* and *P. gigas* as the most likely extant close relatives of the Ictaluridae. The corresponding amino acid tree (data not shown) was also identical and similarly robust to the Bayesian topology, providing further evidence that the ictalurid sister taxa may have been safely demarcated. However, more mitogenome se-

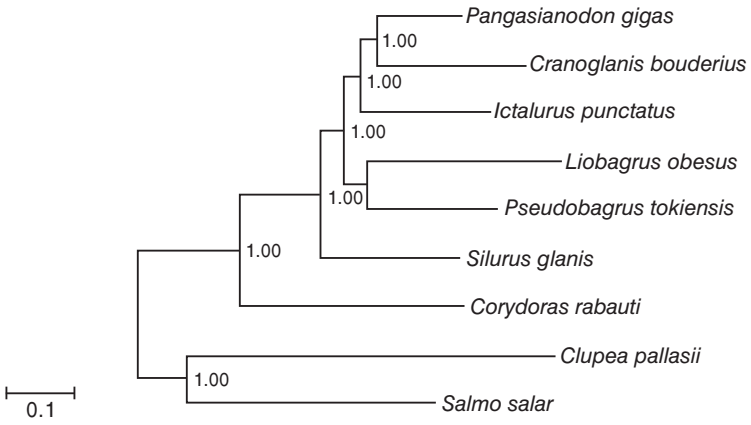

FIG. 2. Bayesian inference topology of siluriform mitogenome sequences (12 protein-coding genes). Accession numbers of the sequences used are given in Table 2. Sequence evolution is based on the GTR+I+Γ model (I = 0.42, gamma distribution shape parameter $\alpha = 0.80$). In two simultaneous runs, four Markov chains (three heated and one cold) were started from random trees and run for 2×10^6 generations with sampling frequency every 200 generations. First 2500 trees were discarded as "burn-in". Branch lengths are means estimated by MrBayes. Posterior probability values higher than 0.90 are shown. The topology is rooted with *Clupea pallasii* and *Salmo salar*. Scale bar: substitutions/site.

quences are needed in order to confidently ascertain this conclusion.

CONCLUSIONS

The mitochondrial genome of *S. glanis* is reported, the first of the family Siluridae. It has a length of 16526 bp, bearing no introns and few intergenic spacers. It codes for 13 polypeptides (subunits of the respiratory chain enzyme complexes), 22 tRNAs, and 2 rRNAs, as in most vertebrates. The *S. glanis* major non-coding region (control region) is found to be 887 bp long, a size comparable to that of other teleosts. Phylogenetic analysis of all available siluriform mitogenomes, and in conjunction with other data, reveals an early split of *S. glanis* and identifies candidate lineages in the intensive search for the extant ictalurid sister taxon. Future addition of more mitochondrial genomes from this order will assist to a better understanding of the phylogenetic relationships and biogeographic history of catfish families which still remain controversial.

ACKNOWLEDGEMENTS

The results presented here are part of a Doctoral Thesis to be submitted at the School of Biology, Aristotle University of Thessaloniki by SV. SV is a recipient of a fellowship from the State Scholarships Foundation (Athens, Greece) and a fellowship from the A.G. Leventis Foundation (Paris, France).

REFERENCES

Abascal F, Zardoya R, Posada D, 2005. ProTest: selection of best-fit models of protein evolution. *Bioinformatics*, 21: 2104-2105.

Adachi J, Hasegawa M, 1996. Model of amino acid substitution in proteins encoded by mitochondrial DNA. *Journal of Molecular Evolution*, 42: 459-468.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, 1990. Basic local alignment search tool. *Journal of Molecular Biology*, 215: 403-410.

Arnason U, Gullberg A, Widegren B, 1991. The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*. *Journal of Molecular Evolution*, 33: 556-568.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC, 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, 18: 489-522.

Banarescu PM, 1989. Zoogeography and history of the freshwater fish fauna of Europe. In: Holcik J, ed. *The freshwater fishes of Europe*. AULA Verlag, Wiesbaden: 88-107.

Boore JL, 1999. Animal mitochondrial genomes. *Nucleic Acids Research*, 27: 1767-1780.

Broughton RE, Milam JE, Roe BA, 2001. The complete sequence of the zebrafish (*Danio rerio*) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA. *Genome Research*, 11: 1958-1967.

Brzuska E, Adamek J, 1999. Artificial spawning of European catfish, *Silurus glanis* L.: stimulation of ovulation using LHRH-a, ovaprim and carp pituitary extract. *Aquaculture Research*, 30: 59-64.

Clayton DA, 1982. Replication of animal mitochondrial DNA. *Cell*, 28: 693-705.

Clayton DA, 1991. Nuclear gadgets in mitochondrial DNA replication and transcription. *Trends in Biochemical Science*, 16: 107-111.

Fernandez-Silva P, Enriquez JA, Montoya J, 2003. Replication and transcription of mammalian mitochondrial

DNA. *Experimental Physiology*, 88: 41-56.

Ferraris CJ, 2007. Checklist of catfishes, recent and fossil (Osteichthyes: Siluriformes), and catalogue of siluriform primary types. *Zootaxa*, 1418: 1-628.

Fink SV, Fink WL, 1981. Interrelationships of the ostariophysan fishes (Teleostei). *Zoological Journal of the Linnean Society*, 72: 297-353.

Hardman M, 2005. The phylogenetic relationships among non-diplomystid catfishes as inferred from mitochondrial cytochrome b sequences; the search for the ictalurid sister taxon (Otophysi: Siluriformes). *Molecular Phylogenetics and Evolution*, 37: 700-720.

Huelsenbeck JP, Ronquist F, 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics*, 17: 754-755.

Hurst CD, Bartlett SE, Davidson WS, Bruce IJ, 1999. The complete mitochondrial DNA sequence of the Atlantic salmon, *Salmo salar*. *Gene*, 239: 237-242.

Inoue JG, Miya M, Tsukamoto K, Nishida M, 2000. Complete mitochondrial DNA sequence of the Japanese sardine, *Sarinops melanostictus*. *Fisheries Science*, 66: 924-932.

Jondeung A, Sangthong P, Zardoya R, 2007. The complete mitochondrial DNA sequence of the Mekong giant catfish (*Pangasianodon gigas*), and the phylogenetic relationships among Siluriformes. *Gene*, 387: 49-57.

Kartavtsev YP, Jung S-O, Lee Y-M, Byeon H-K, Lee J-S, 2007. Complete mitochondrial genome of the bullhead torrent catfish, *Liobagrus obesus* (Siluriformes: Amblycipitidae): Genome description and phylogenetic considerations inferred from the Cytb and 16S rRNA genes. *Gene*, 396: 13-27.

Kim YJ, Kweon HS, Kim IC, Lee YM, Kim JM, Lee JS, 2004. The complete mitochondrial genome of the floating goby, *Gymnogobius petschiliensis* (Periformes, Gobiidae). *Molecular Cell*, 17: 446-453.

Kottelat M, Freyhof J, 2007. *Handbook of European freshwater fishes*. Publications Kottelat, Cornol, Switzerland.

Krieg F, Estoup A, Triantafyllidis A, Guyomard R, 1999. Isolation of microsatellite loci in European catfish, *Silurus glanis*. *Molecular Ecology*, 8: 1964-1966.

Kumar S, Dudley J, Nei M, Tamura K, 2008. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics*, 9: 299-306.

Lavoué S, Miya M, Saitoh K, Ishiguro NB, Nishida M, 2007. Phylogenetic relationships among anchovies, sardines, herrings and their relatives (Clupeiformes), inferred from whole mitogenome sequences. *Molecular Phylogenetics and Evolution*, 43: 1096-1105.

Lee JS, Miya M, Lee YS, Kim CG, Park EH, Aoki Y, Nishida M, 2001. The complete mitochondrial genome of the self-fertilizing fish *Rivulus marmoratus* (Cyprinodontiformes, Rivulidae) and the first description of a duplication of a control region in fish. *Gene*, 280: 1-7.

Lee WJ, Conroy J, Howell WH, Kocher TD, 1995. Structure and evolution of teleost mitochondrial control regions. *Journal of Molecular Evolution*, 41: 54-66.

Legendre M, Linhart O, Billard R, 1996. Spawning and management of gametes, fertilized eggs and embryos in Siluroidei. *Aquatic Living Resources*, 9: 59-80.

Miya M, Takeshima H, Endo H, Ishiguro NB, Inoue JG, Mukai T, Satoh TP, Yamaguchi M, Kawaguchi A, Mabuchi K, Shirai SM, Nishida M, 2003. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, 26: 121-138.

Nagase M, Aimi T, Suginaka K, Kitamoto Y, Morinaga T, 2005. Complete mitochondrial DNA sequence of the Japanese flying fish *Cypselurus hiraii*. *Fisheries Science*, 71: 914-923.

Nei M, 1987. *Molecular evolutionary genetics*. Columbia University Press, New York.

Nelson JS, 1994. *Fishes of the world*. Third edition. John Wiley & Sons, New York.

Nikolaidis N, Scouras ZG, 1996. The *Drosophila montium* subgroup species. Phylogenetic relationships based on mitochondrial DNA analysis. *Genome*, 39: 874-883.

Nylander JAA, Wilgenbusch JC, Warren DL, Swofford DL, 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics*, 24: 581-583.

Ojala D, Montoya J, Attardi G, 1981. tRNA punctuation model of RNA processing in human mitochondria. *Nature*, 290: 470-474.

Peng Z, Wang J, He S, 2006. The complete mitochondrial genome of the helmet catfish *Cranoglanis bouderius* (Siluriformes: Cranoglanididae) and the phylogeny of otophysan fishes. *Gene*, 376: 290-297.

Pissios P, Scouras ZG, 1993. Mitochondrial DNA evolution in the Montium-species subgroup of *Drosophila*. *Molecular Biology and Evolution*, 10: 375-382.

Posada D, Crandall KA, 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics*, 14: 817-818.

Proteau JP, Hilge V, Linhart O, 1996. Etat actuel et perspectives de la production aquacole des poissons-chats (Siluroidei) en Europe. *Aquatic Living Resources*, 9: 229-235.

Saitoh K, Miya M, Inoue JG, Ishiguro NB, Nishida M, 2003. Mitochondrial genomics of Ostariophysan fishes: perspectives on phylogeny and biogeography. *Journal of Molecular Evolution*, 56: 464-472.

Sbisa E, Tanzariello F, Reyes A, Graziano P, Saccone C, 1997. Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. *Gene*, 205: 125-140.

Shadel GS, Clayton DA, 1993. Mitochondrial transcription initiation. *Journal of Biological Chemistry*, 268: 16083-16086.

Shadel GS, Clayton DA, 1997. Mitochondrial DNA maintenance in vertebrates. *Annual Review of Biochemistry*, 66: 409-435.

Smitherman RO, Dunham RA, Whitehead PK, 1996. Selection, hybridization and genome manipulation in Siluroidei. *Aquatic Living Resources*, 9: 93-102.

Sullivan JP, Lundberg JG, Hardman M, 2006. A phylogenetic analysis of the major groups of catfishes (Teleostei: Siluriformes) using rag1 and rag2 nuclear gene sequences. *Molecular Phylogenetics and Evolution*, 41: 636-662.

Swofford DL, 1998. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, MA.

Tavare S, 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. In: Miura RM, ed. *Some mathematical questions in biology: DNA sequence analysis*. American Mathematical Society, Providence, RI: 57-86.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG, 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, 25: 4876-4882.

Triantafyllidis A, Ozouf-Costaz C, Rab P, Suciu R, Karakousis Y, 1999a. Allozyme variation in European siluric catfishes, *Silurus glanis* and *Silurus aristotelis*. *Biochemical Systematics and Ecology*, 27: 487-498.

Triantafyllidis A, Abatzopoulos TJ, Economidis PS, 1999b. Genetic differentiation and phylogenetic relationships among Greek *Silurus glanis* and *Silurus aristotelis* (Pisces, Siluridae) populations, assessed by PCR-RFLP analysis of mitochondrial DNA segments. *Heredity*, 82: 503-509.

Triantafyllidis A, Krieg F, Cottin C, Abatzopoulos TJ, Triantafyllidis C, Guyomard R, 2002. Genetic structure and phylogeography of European catfish (*Silurus glanis*) populations. *Molecular Ecology*, 11: 1039-1055.

Walberg MW, Clayton DA, 1981. Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. *Nucleic Acids Research*, 9: 5411-5421.

Waldbieser GC, Bilodeau AL, Nonneman DJ, 2003. Complete sequence and characterization of the channel catfish mitochondrial genome. *DNA Sequencing*, 14: 265-277.

Wernersson R, Pedersen AG, 2003. RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Research*, 31: 3537-3539.

Xia X, Xie Z, 2001. DAMBE: software package for data analysis in molecular biology and evolution. *Journal of Heredity*, 92: 371-373.

Zuker M, Mathews DH, Turner DH, 1999. Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In: Barciszewski J, Clark BFC, eds. *RNA biochemistry and biotechnology*. Kluwer Academic Publishers, Boston: 11-43.