New complete mitogenome datasets and their characterization of the European catfish (Silurus glanis)

Kinga Székvári a,b, Zoltán Szabolcsi c,d, Barbara Kutasy b,e, Géza Hegedűs f, Eszter Virág b,f,g,*

a Festetics Doctoral School, Hungarian University of Agriculture and Life Sciences, Gödöllő, Hungary
b Department of Plant Science and Biotechnology, Szent István University, i.e. the Predecessor of the Festetics Bioinnovation Group, Keszthely, Hungary
c Department of Animal Sciences, Georgikon Faculty, Szent István University, Keszthely 8360, Hungary
d Research Centre for Natural Sciences, Institute of Enzimology, Budapest, Hungary
e Department of Plant Physiology and Plant Ecology, Hungarian University of Agriculture and Life Sciences Georgikon Campus Keszthely, Keszthely, Hungary
f EduCoMat Ltd., Keszthely, Hungary
g Department of Molecular Biotechnology and Microbiology, Institute of Biotechnology, Faculty of Science and Technology, University of Debrecen, Hungary

ABSTRACT

We present new complete mitogenome sequences of Silurus glanis (S. glanis) from 4 samples such as male and female individuals from two countries (Hungary, Czech Republic). The complete mitochondria were determined from genome sequencing by using illumina MiSeq platform resulting in long, 300 bp, paired-end reads. De novo assembly was performed resulting in one nod (scaffold) covering the total mitochondria in each sample. The mitochondrial genomes were circular, double-stranded molecules of 16,524 bp in length and consisted of 13 protein-coding genes (PCGs), 2 ribosomal RNA genes, 22 transfer RNA genes, and 1 control region. These sequences were deposited in the NCBI GeneBank under the accession numbers (MW796040, MW796041, MW796042, MW796043) and compared with the only available S. glanis mitochondrial genome (NC_014261.1) sequenced by unidentified technology and showed 99% similarity. We found in
• We provided 4 new whole mitochondrial genomes of the European catfish (*Silurus glanis*, *S. glanis*). Continued expansion of mitochondrial genome databases to include both a greater number of species and increased representation of populations from throughout their range will provide an improved basis for analysis.

• Our data will be useful for *S. glanis* species monitoring, phylogenetic, population, and evolutionary studies.

• Illumina long read has been chosen and applied for the NGS sequencing methodology. The sequencing using paired-end reads of 300bp were uniquely used for whole mitochondrial genome sequencing in the case of the species of the Teleostei group. Higher coverage of the nucleotide positions and proper quality values of the assemblies (see N50 values) resulted in a more accurate whole mitochondrial genome of our samples. These mitogenomes were suggested and provided as new reference sequences for further studies.
1. Data Description

*S. glanis* is the largest-bodied European freshwater fish, Inhabitant of Native in Eastern Europe and western Asia. This species is now extensively dispersed and introduced in several countries to the west and south of its endemic range. The *S. glanis* belongs to the family Siluridae, a group of freshwater fish indigenous to Europe, Asia, and Africa. There are 107 species from 12 genera in this family. Among the 18 *Silurus* species, two are native to Europe: wels catfish and Aristotle’s catfish (*S. aristotelis*). European catfish is the largest-bodied fish of the order Siluriformes and can attain a maximum length of 500 cm, although it more commonly reaches 300 cm [1].

The circular mitogenomes of *S. glanis* (GeneBank accession numbers MW796040, MW796041, MW796042, MW796043) were 16,524 bp in length, in all 4 samples, which contained 37 genes (13 protein-coding, 22 tRNAs, 2 rRNAs) and one control region displacement loop (D-loop) (Fig. 1, Table 2, Supplementary 1). The genes encoded by the mitogenome are characteristic to the vertebrate mitochondrial genome. The organization of the genes also tends to be conserved among vertebrates for 37 genes and the D-loop, which are arranged in the same order from hagfish to eutherian mammals [2–6]. Information for each individual is presented in Table 1. The representative complete mitogenome map in Fig. 1. The 4 mitochondria showed 99% similarity, twelve of 13 PCGs contained the typical ATG as a start codon, however, the gene COI started with GTG. Similar data were found by Vittas; Wu et al.; Zeng et al. [7–9]. 6 genes (nad5, nad4L, atp6, atp8, COI, nad1) of 13 PCGs ended in TAA for the stop codon. 3 genes ended in TAG (nad2, nad3, nad6). 4 genes (cytb, nad4, COII, COIII) ended in only a T residue. Such immature stop codon is completed via post-transcriptional polyadenylation [10].

**Fig. 1.** *S. glanis* mitochondrial genome. Genes for proteins and rRNAs are shown with standard abbreviations, outside of the circle. Genes for tRNAs are designated by a single letter for the corresponding amino acid, inside of the circle. Before the genes, it shows their position in seq4.
Table 1
Reported mitogenome samples.

| Sample ID | GeneBank accession number | Origin          | Gender | GC%   | AT%  |
|-----------|---------------------------|-----------------|--------|-------|------|
| seq1      | MW796040                  | Czech Republic  | female | 44.86 | 5514 |
| seq2      | MW796041                  | Czech Republic  | male   | 44.86 | 5514 |
| seq3      | MW796042                  | Czech Republic  | female | 44.86 | 5514 |
| seq4      | MW796043                  | Hungary         | male   | 44.87 | 5513 |

Table 2
Reference genome gene organization.

| Gene/Element | Abbreviation | Position | Size(bp) | Startcodon | Stopcodon |
|--------------|--------------|----------|----------|------------|-----------|
| NADH dehydrogenase subunit 4L | ND4L | 1–297 | 297 | ATG | TAA |
| NADH dehydrogenase subunit 4 | ND4 | 291–1671 | 1381 | ATG | T* |
| tRNA\(^{His}\) | H | 1672–1741 | 70 | |
| tRNA\(^{Ser}\) | S | 1742–1807 | 66 | |
| tRNA\(^{Leu}\) | L | 1811–1883 | 73 | |
| NADH dehydrogenase subunit 5 | ND5 | 1884–3710 | 1827 | ATG | TAA |
| NADH dehydrogenase subunit 6 | ND6 | 3707–4225 | 519 | ATG | TAG |
| tRNA\(^{Glu}\) | E | 4226–4294 | 69 | |
| cytochrome b | CYTB | 4296–5433 | 1138 | ATG | T* |
| tRNA\(^{Thr}\) | T | 5434–5504 | 61 | |
| tRNA\(^{Pro}\) | P | 5503–5572 | 70 | |
| control region | D-loop | 5573–6459 | 887 | |
| tRNA\(^{Asp}\) | F | 6460–6529 | 70 | |
| 12S ribosomal RNA | 12S | 6530–7482 | 953 | |
| tRNA\(^{Val}\) | V | 7483–7554 | 72 | |
| 16S ribosomal RNA | 16S | 7555–9232 | 1678 | |
| tRNA\(^{Leu}\) | L | 9233–9307 | 75 | |
| NADH dehydrogenase subunit 1 | ND1 | 9308–10282 | 975 | ATG | TAA |
| tRNA\(^{Ile}\) | I | 10284–10355 | 72 | |
| tRNA\(^{Gln}\) | Q | 10355–10425 | 71 | |
| tRNA\(^{Met}\) | M | 10425–10493 | 69 | |
| NADH dehydrogenase subunit 2 | ND2 | 10494–11540 | 1047 | ATG | TAG |
| tRNA\(^{Asp}\) | W | 11539–11609 | 71 | |
| tRNA\(^{His}\) | A | 11612–11680 | 69 | |
| tRNA\(^{Asn}\) | N | 11682–11754 | 73 | |
| tRNA\(^{Cys}\) | C | 11789–11854 | 66 | |
| tRNA\(^{Glu}\) | Y | 11858–11927 | 70 | |
| cytochrome c oxidase subunit I | COI | 11929–13479 | 1551 | GTG | TAA |
| tRNA\(^{Ser}\) | S | 13480–13550 | 71 | |
| tRNA\(^{Ala}\) | D | 13555–13626 | 72 | |
| cytochrome c oxidase subunit II | COII | 13641–14331 | 691 | ATG | T* |
| tRNA\(^{Leu}\) | K | 14332–14405 | 74 | |
| ATP synthase F0 subunit 8 | ATP8 | 14407–14574 | 168 | ATG | TAA |
| ATP synthase F0 subunit 6 | ATP6 | 14565–15248 | 684 | ATG | TAA |
| cytochrome c oxidase subunit III | COIII | 15248–16031 | 784 | ATG | T* |
| tRNA\(^{Cys}\) | G | 16032–16104 | 73 | |
| NADH dehydrogenase subunit 3 | ND3 | 16105–16455 | 351 | ATG | TAG |
| tRNA\(^{Arg}\) | R | 16454–16524 | 71 | |

In seq1 and seq3, which are female samples, control region, trnP, trnT, CYTB, ND5, trnL, trnS, trnH, ND4, ND4L, trnR, ND3, trnG, COX3, ATP6, ATP8, trnK, COX2, trnD, COX1, trnW, ND2, trnM, trnL, ND1, trnL, trnL, trnV, trnS, trnF were encoded by the H-strand, trnQ, trnA, trnN, trnC, trnY, trnY, ND6, trnE, were encoded by the L-strand. On the other hand, in seq2 and seq4, which are males, trnQ, trnA, trnN, trnC, trnY, trnS, ND6, trnE, were encoded by the H-strand, control region, trnP, trnT, CYTB, ND5, trnL, trnS, trnH, ND4, ND4L, trnR, ND3, trnG, COX3, ATP6, ATP8, trnK, COX2, trnD, COX1, trnW, ND2, trnM, trnL, ND1, trnL, trnL, trnV, trnS, trnF were encoded by the L-strand.
Table 3
The observed amino acid changes in the 13 protein-coding genes in the examined *S. glanis* samples. Black and red colors indicate the changes of NC_0142611 to examined amino acids at the position of the gene sequence. Each indicated amino acid changes were detected uniformly in the four newly identified mitogenomes. The Val154Ile substitution in the cox2 gene was observed only in seq3.

| Protein coding genes | Amino acid changes |
|----------------------|--------------------|
| nad1                 | -                  |
| cyt b                | Val145Ile          |
| nad6                 | Val42Ile           |
| nad 5                | Asn76Ser, Val211Ile, Ile538Met, Thr599Ala |
| nad4                 | Ser71Cys, Phe73Leu, Val76Leu, Ala338Thr, Pro339His |
| nad4L                | Phe53Ser, Cys56Phe, Met61Ile, Arg62Leu |
| nad3                 | Pro36Ser, Asp39Glu, Ser44Pro, His55Arg, Ser56Phe |
| cox3                 | Pro140Ser, Gly142Val |
| atp6                 | -                  |
| atp8                 | -                  |
| cox1                 | Met1Val, Gln111Leu, His260Tyr |
| cox2                 | Ile76Val, Val154Ile* |
| nad2                 | Ser179Leu, Pro326Leu |

Amino acid sequences were compared to the only available *S. glanis* mitogenome (NC_014261.1). From the 13 PCGs, we found differences in 10, which are nad2, nad3, nad4, nad4L, nad5, nad6, cyt b, COI, COII, COIII. Atp6, atp8, and nad1 were the same in the 4 examined samples and in the *S. glanis* mitochondrial genome (NC_014261.1) sequence as well. In cyt b and nad6 was 1, in nad2 and COIII were 2, in COI were 3, in nad4L and nad5 were 4, in nad4 and nad3 were 5 amino acid changes comparing to the only available *S. glanis* sequence (NC_014261.1). COII showed differences in seq3 there were 2 amino acid changes, in seq1, seq2, and seq4 was only 1. The amino acid changes are shown in Table 3 (Supplementary 2.). Sequencing multiple mitochondrial genomes from the same species, *S. glanis* revealed varying levels of intraspecies genetic variation.

The phylogenetic relationship of *S. glanis* was compared with previously analyzed mitogenomes of other Siluriformes, Silurus, Kryptopterus, Ompok, and Pterocryptis genus in Siluridae family and on two other genera, Ictalurus in Ictaluridae, Pangasianodon in Pangasiidae.
Fig. 2. Phylogenetic tree of *Silurus glanis* with other catfishes. Based on the mitochondrial 12S rRNA.

family. The phylogenetic tree is shown in Fig. 2. The phylogenetic location of *S. glanis* was the closest to *S. microdorsalis*. Similar results were found by Park et al.; Yang et al. [11,12].

2. Experimental Design, Materials and Methods

*S. glanis* samples were collected from Hungary and Czech Republic rivers. The genomic DNAs were extracted with Thermo Scientific™ GeneJET Genomic DNA Purification Kit from caudal fin and stored at −70 °C. After the extraction, we checked the DNA concentration and purity by agarose gel electrophoresis and spectrophotometric quantification. Four *S. glanis* total gDNA samples were de novo sequenced and analyzed. Mitogenomes were reconstructed in silico in each sample. Libraries were prepared using the Nextera XT DNA Sample Preparation Kit according to the manufacturer’s protocol [13], unless otherwise stated. Sequencing reactions were carried out using the MiSeq v2 (2 × 301 bp) chemistries (Illumina). Similar sequencing technique was used by Austin et al.; Tabassum et al.; Alam et al. [14–16]. The raw reads were cleaned by the trimming of adaptor sequences, empty reads, and ambiguous nucleotides (‘N’ at the end of the reads). The reads obtained were then assembled using the SPAdes (St Petersburg genome assembler) assembly toolkit containing various assembly pipelines based on the Bruijn Graph [17,18]. Total genome sizes of four individuals were approximately between 800–810 Mb, with
4383–4388 scaffolds and N50 varied between 3.1–3.4 Mb. The predicted genome size was corresponding to the most closely related Silurus asotus (831–1411 Mb) [19] and other Siluriformes, whose genome sizes vary from 599 Mb in Bagarius yarrelli [20] to 1200 Mb in Clarias batrachus [21]. The longest individual scaffolds were 9–9.5 Mb. Mitochondrial genomes were separated into individual scaffolds: NODE_3 (seq1), NODE_4 (seq3), NODE_6 (seq4), NODE_8 (seq2) with the same length 16,524 and k-mer coverage for the last (largest) k values used were 46.985597, 56.102578, 65.455701 and 50.835330. The sequencing coverage varied between 140–150x of the four mitochondrial genomes. The base composition was GC 44, 86% and AT 55, 14% in the samples from Czech Republic and GC 44, 87% and AT 55, 13% in the sample from Hungary. The mitogenome contigs were identified by BLAST+ [22] alignments to the previously published S. glanis mitochondrial genome (NC_014261.1). For sequence comparison, we used NCBI BLAST [23] and Geneious 9.0.5 [24]. The phylogenetic analysis was performed using Geneious Tree Builder, the Alignment type was Global alignment with free end gaps, the Genetic Distance Model was Tamura-Nei, and the Tree built Method was Neighbor-Joining. The analysis is based on the mitochondrial 12S rRNA, because this gene sequence is frequently used in molecular taxonomy and phylogeny [25–28]. For phylogenetic analysis, nucleotide sequences were downloaded from the NCBI database.

Ethics Statement

This study is based on non-living animal experiments, only tissue samples. Do not require an ethics statement.

Declaration of Competing Interest

The authors report no conflict of interest. The authors alone are responsible for the content and writing of the paper. The first and second authors participated in equal proportions of the work.

CRediT Author Statement

Kinga Székvári: Visualization, Investigation; Zoltán Szabolcsi: Conceptualization, Methodology; Barbara Kutasy: Data curation, Writing – original draft; Géza Hagedűs: Software, Resources; Eszter Virág: Software, Resources, Supervision, Writing – review & editing.

Acknowledgment

We express our thanks to the EduCOMat Ltd, Hungary to perform bioinformatics analysis and support the publication.

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.dib.2021.107418.

References

[1] G.H. Copp, et al., Voracious invader or benign feline? A review of the environmental biology of European catfish Silurus glanis in its native and introduced ranges, Fish Fish. 10 (3) (2009) 252–282, doi:10.1111/j.1467-2979.2008.00321.x.
