Dadkhah et al. BMC Ecology and Evolution (2023) 23:55
https://doi.org/10.1186/s12862-023-02166-2

BMC Ecology and Evolution

RESEARCH

Open Access

Characterization of the mitochondrial *Huso huso* genome and new aspects of its organization in the presence of tandem repeats in 12S rRNA

Khadijeh Dadkhah[1*], Ghodrat Rahimi Mianji[1], Ali Barzegar[2] and Ayoub Farhadi[1]

Abstract

Background The sturgeon group has been economically significant worldwide due to caviar production. Sturgeons consist of 27 species in the world. Mitogenome data could be used to infer genetic diversity and investigate the evolutionary history of sturgeons. A limited number of complete mitogenomes in this family were sequenced. Here, we annotated the mitochondrial *Huso huso* genome, which revealed new aspects of this species.

Results In this species, the mitochondrial genome consisted of 13 genes encoding proteins, 22tRNA and 2rRNA, and two non-coding regions that followed other vertebrates. In addition, H. huso had a pseudo-tRNA-Glu between ND6 and Cytb and a 52-nucleotide tandem repeat with two replications in 12S rRNA. This duplication event is probably related to the slipped strand during replication, which could remain in the strand due to mispairing during replication. Furthermore, an 82 bp repeat sequence with three replications was observed in the D-loop control region, which is usually visible in different species. Regulatory elements were also seen in the control region of the mitochondrial genome, which included termination sequences and conserved regulatory blocks. Genomic compounds showed the highest conservation in rRNA and tRNA, while protein-encoded genes and nonencoded regions had the highest divergence. The mitochondrial genome was phylogenetically assayed using 12 protein-encoding genes.

Conclusions In H. huso sequencing, we identified a distinct genome organization relative to other species that have never been reported. In recent years, along with the advancement in sequencing identified more genome rearrangements. However, it is an essential aspect of researching the evolution of the mitochondrial genome that needs to be recognized.

Keywords *Huso huso*, Mitogenome, Sturgeon, Caspian Sea, Tandem repeats

*Correspondence:
Khadijeh Dadkhah
kdadkhah29@gmail.com
[1] Laboratory for Molecular Genetics and Animal Biotechnology, Faculty of Animal Sciences and Fisheries, Sari Agricultural Sciences and Natural Resources University, Sari, Iran
[2] Department of Basic Sciences, Sari Agricultural Sciences and Natural Resources University, Sari, Iran

Background

Studying the molecular genetics of sturgeon could provide an excellent way to identify their origin and uncertain aspects. The *Acipenseridae* consists of four genera (*Acipenser*, *Huso*, *Pseudoscaphirhynchus*, and *Scaphirhynchus*), of which there are 25 species. Two genera, *Acipenser* and *Huso*, inhabit the Caspian Sea. Most sturgeons have been identified as endangered by the International Union for Conservation of Nature (IUCN) [1].

© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Studies have shown that caviar production has declined in Europe and North America, as well as in Russia and Iran [2]. Sturgeon is one of the most valuable groups of fish due to caviar production. Sturgeons are very important in genetic matters due to their 200 million-year-old history; it has been known as living fossil since the Jurassic era. Caspian sturgeon are listed in the CITES Convention. The sturgeon population is endangered for problems such as irregular and illegal fishing of sturgeon, environmental changes, climatic conditions, low rate of artificial reproduction, loss of natural environments, and pollution.

Molecular studies on the genetic structure could help to identify valuable information about this endangered fish. In determining the genetic structure of Caspian sturgeon, molecular techniques such as RFLP [3], RAPD [4], AFLP [5–7], microsatellite [8–12], and DNA sequencing [13, 14] have replaced traditional methods such as meristics and morphometrics. DNA sequencing is one of the most accurate methods of identifying information about sturgeon [15]. Mugue et al. [15] distinguished sturgeon at the species level with mitochondrial D-loop sequencing. In another study, mitochondrial D-loop sequencing was used as a complementary method to other laboratory methods [16]. Mitochondrial genome sequencing is a simple, immediate, and reliable method to identify species. The Cytb and control regions have been used mainly to identify sturgeon species. However, researchers have found that the Cytb gene does not have a good resolution for identifying these species [15]. The size of mitochondrial DNA is approximately 15–20 kb, and the mitochondrial genome contains several copies, is double-stranded, and has a circular molecule. It contains 13 protein-encoding genes, two genes encoding ribosomes, 22 genes encoding tRNAs, and two control regions. Because mitochondrial genomic DNA is inherited from the mother [17], it is suitable for evolutionary and historical studies [18]. Ludwig et al. [19] illustrated VNTR duplicate blocks in some sturgeon species in the control region in region 5' immediately after proline tRNA, which was three replicates with a size of 82 bp. Mugue et al. [20] also reported a repetitive sequence of 82 pairs of bases in the ship specie. Research has shown that repetitive blocks in the control region are related to genes close to CR (12S rRNA and tRNA) [21].

Ciftici et al. [22] recognized the presence of duplicate blocks in the D-loop regions and reported an 82–83 bp tandem repeat by sequencing in *Acipenser gueldenstaedtii*, *Acipenser stellatus*, and *huso huso* species. The frequency of multiple copies of the genome in a species may be attributed to the differences in mutation rates or a mechanism that may regulate mutation rates. A large number of studies have been performed on the teleosts fish of control regions that identified VNTR [19, 20, 22]. However, in this study, we also detected the presence of tandem repeats in the other two regions, 12S rRNA and tRNA. Previous research has shown excess tRNA in several fish species [21–26], but there have been no reports about the presence of repeats in 12 SrRNAs.

The vertebrate mitogenome is highly conserved; however, with increasing genome sequence data for fish, reports of rearrangement have been observed. The present study aims have been to draw a complete map of the mitochondrial genome of *Huso huso* and a complete sequencing of the mitochondrial genome, determination of genetic structure and phylogenetic relationships, and codon usage. Our attempt initially was to study the mitochondrial genome and its molecular mechanisms; in the following, we found novel aspects of rearrangement in this species. This study reveals that the mitochondrial genome of sturgeon could have differed in organization, gene content, and order. In addition, these results will provide a better perspective of understanding fish's evolution.

Results and discussion

Genome organization

The complete mitogenome of *Huso huso* was deposited in the Gene Bank (Accession number: MK213068). The *Huso huso* mitochondrial genome consisted of 16836 bp. This genome contained 37 encoding genes, 13 genes encoding proteins, 22tRNA and 2rRNA, and two OL and D-loop control regions, in which an additional tandem repeat tRNA- Glu was observed in Table 1.

Like many mitochondrial genomes, most genes were located in the heavy strand except ND6 and eight tRNA (tRNAGLN, tRNAala, tRNAGLU, tRNAser, tRNA tyr and tRNAcys, tRNA Asn and tRNA Pro) that were coded in the light strand. The D-loop (displacement loop) was located in the main non-coding region of the mitochondrial DNA molecule. The mitochondrial DNA could be replicated in two different ways, starting in the D-loop region. OL is the origin of light strand replication, which was identified in the WANCY region with a cluster of five tRNA similar to that of other vertebrates. There were two tRNA-Glu with repeated sequences, which could be seen in Fig. 1.

Nucleotide asymmetry of the strand is usually described by AT, and GC skews. GCskew is barely above zero (they are mostly negative). These results illustrate that the content of A is only slightly higher than that of T, whereas that of C is significantly higher than that of G. Skews are related to the difference in mutation pressures applied in light and heavy strings [27]. As a result, they are asymmetric and cause changes in mtDNA [18, 28, 29].

Table 1 Summary of gene features of *Huso huso*

Gene	Coding	Position		Size (bp)	Codon		Anticodon	Intergenic Nucleotides
	Strand	From	To		Start	Stop		
tRNA Phe	H	1	68	68			GAA	
12SrRNA	H	69	1081	1013				0
tRNA val	H	1082	1151	70			UAC	0
16SrRNA	H	1153	2854	1702				+1
tRNA leu	H	2855	2929	75			UAA	0
ND1	H	2930	3904	975	ATG	TAG		0
tRNA Ile	H	3914	3984	71			GAU	+9
tRNA Gln	H	3984	4054	71			UUG	-1
tRNA met	L	4054	4123	70			CAU	0
ND2	H	4124	5168	1045	ATG	TAG		0
tRNA trp	H	5169	5241	73			UCA	0
tRNA ala	H	5243	5312	70			UGC	+2
tRNA asn	H	5314	5386	73			GUU	0
OL	L	5387	5420	34				0
tRNA cys	L	5421	5487	67			GCA	0
tRNA tyr	L	5488	5558	71			GUA	0
COI	L	5560	7128	1569	GTG	TAA		+1
tRNA ser	H	7120	7192	73			UGA	-7
tRNA Asp	L	7198	7269	72			GUC	+7
COII	H	7284	7974	691	ATG	T		+14
tRNA lys	H	7975	8048	74			UUU	0
ATP8	H	8050	8217	168	ATG	TAA		+1
ATP6	H	8208	8891	684	ATG	TAA		-8
COIII	H	8891	9675	785	ATG	TAA		0
tRNA Gly	H	9676	9748	73			UCC	0
ND3	H	9749	10097	349	ATG	TAG		0
tRNA Arg	H	10098	10167	70			UCG	0
ND4 L	H	10168	10464	297	ATG	TAA		0
ND4	H	10458	11838	1381	ATG	T		-5
tRNA His	H	11839	11907	69			GUG	0
tRNA ser	H	11908	11975	68			GCU	0
tRNA leu	H	11976	12048	73			UAG	0
ND5	H	12049	13890	1842	ATG	TAA		0
ND6	L	13887	14408	522	ATG	TAG		-2
tRNA Glu	L	14409	14478	70			UUC	+4
tRNA Glu	L	14483	14552	70			UUC	0
cytb	H	14555	15695	1141	ATG	T		+2
tRNA thr	H	15696	15769	74			UGU	0
tRNA pro	L	15773	15842	70			UGG	+3
D-loop	H	15842	16836	994				0

Genome organization and evolutionary mechanism

A new type of mitochondrial genome organization has been found to contain an additional tRNA. This extra tRNA can be folded, creating a clover leaf structure of tRNA that also contains an anticodon.

This tRNA is very similar to the conventional tRNA sequence and is located between ND6 and Cytb. This issue is a new genome organization, and it could create new specific features in the genus *H. huso* that distinguish it from other sturgeon species. Changes in the tRNA of

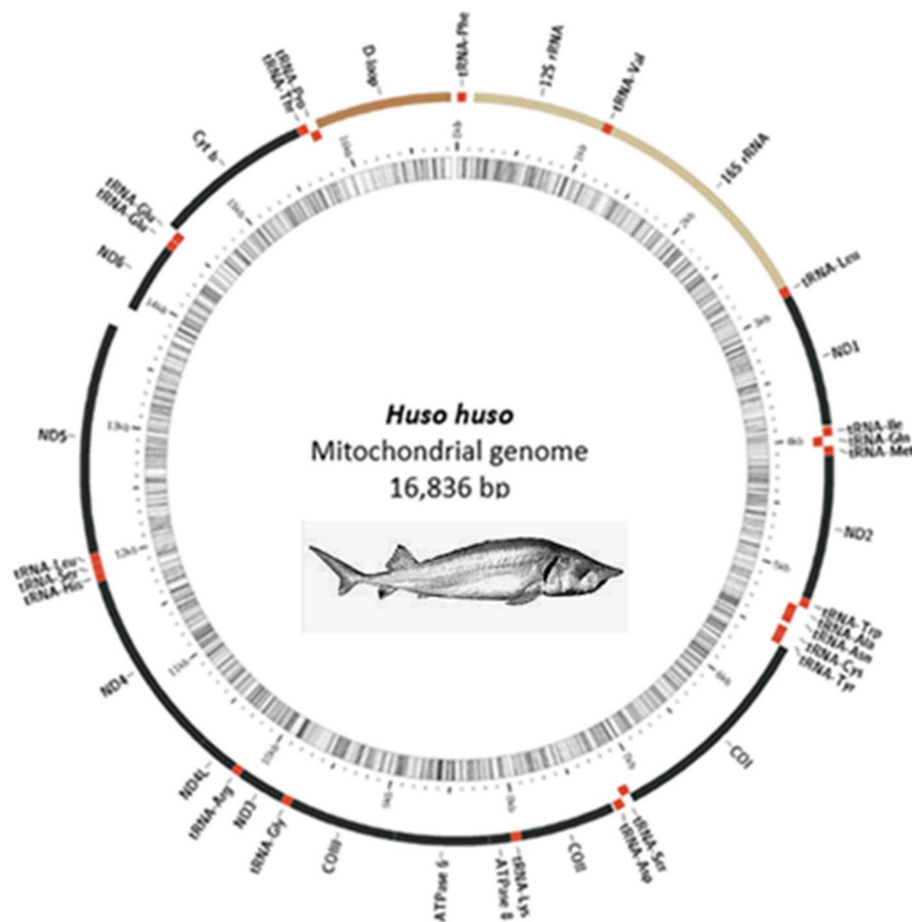

Fig. 1 The complete mitochondrial genomes in *Huso huso*. Protein coding, ribosomal RNA, and transfer RNA genes are shown using different colors. Genes encoded on the H-strand are in the outer region. Genes coded on the L-strand are in the inner region

other fish species have also been observed [21–25]. In a study of a fish species, additional tRNA-Ile was generated with anticodon mutations TAC to AAT. It was also seen in *Serranidae* that the extra tRNA-Asp in the light strand resulted from the rearrangement of the mitochondrial genome [30, 31].

This gene content of the fish mitochondrial genome is typically the result of gene amplification and causes diversity in species.

The rearrangement of the mitochondrial genome observed in fish usually can include translocation [32, 33], which is also due to the tandem duplication of gene regions. These results showed that extra tRNA is left in the mitochondrial genome after changes. Several abnormal fish mitochondrial genome tRNAs include extra tRNA-ser at the downstream ND5 in sea bass *Morone saxatilis* [24], extra tRNA met in *Pampus species* [22], the pseudo tRNA at the same position in parrotfish *Chlorurus sordidus* [23], extra tRNA Asn, pseudo tRNA ala in the WANCY cluster polar cod *Boreogadus saida* [21], and tRNA pro amplification in CR from *Antarctic notthenioig* [25] were found in previous results. These reports have shown that new tRNAs appear to have been seen in various places. This issue results from rearrangement and mispairing, which ultimately leaves one or more additional tRNA in the mitochondrial genomes of different fish species. Rearrangements in bony fish determine the complete nucleotide sequence of the mitochondrial genome [34], such that for a benthic fish, *Gonostoma gracile*, the entire mitochondrial genome includes 19 tRNA genes that exist in typical vertebrates. However, the gene sequence of tRNA Glu is different [35].

Protein coding genes

The cumulative length of *H. huso* mitochondrial protein-coding genes was 11406 bp, which was calculated as 67% of the total length of the mitochondrial genome. The genes encoded in mtDNA were highly compact and contained overlapping sections. Our overlap was

between 13 protein-encoding genes in this species between ATP8 and ATP6, which had an eight bp overlap, between ND4 L and ND4, which was five bp, and between ND5 and ND6, it was found to be two bp.

Most genes encoding mitochondrial proteins in this species began with the ATG primer, similar to many metazoa [36]. The COI gene only had one separate start codon, GTG. Among the coding genes for the protein, the Cytb, COII, and ND4 genes were terminated in T (Table 1). This incomplete codon is completed with the addition of poly A and finally will become TAA.

The calculation of available nucleotide abundance at each codon position in all 13 protein genes is shown in the table. The T nucleotide was mostly seen in the position of the second codon. Since triple codons are encoded with T in the second position for hydrophobic or hydrophobic residues, this observed deviation indicates a high ratio of hydrophobic residues between the coding proteins. The 12 genes encoding proteins in the heavy strand shared an anti-G deviation; it was approximately 10% or less at the third codon position. The deviation, primarily in the position of the third codon, is probably due to the selection pressure on the synonymous mutation in this position (Table S2).

Serine and leucine amino acids indicated the highest frequency in *Huso huso*, using six different codons, while some amino acids used 2 or 4 codons. After them, the amino acids alanine, threonine, glycine, proline, and valine showed the highest frequency with four codons. Of course, arginine also had four codons and was observed with a relatively high frequency. The rest of the amino acids used two codons, which displayed the same frequency (Fig. 2) (Table S3).

rRNA and tRNA

Evaluation of rRNAs in this species showed that the length for 12S rRNA was 1013 bp and for 16S rRNA was 1702 bp. The H strand encoded both rRNAs, and their nucleotide composition for 12S rRNA and 16S rRNA was A% 31.4, C% 27.5, T% 19.3, G% 21.7, and A% 35.5, C% 24.6, T% 19.6, G% 20.0, respectively. The deviation of the nucleotide composition is a strand-specific property in mtDNA, and this deviation observed is against G. Similar phenomena have been found in the mitochondrial genome of other fishes [37] and mammals [3] that are thought to be related to asymmetric replication of the H and L strand [38, 39], exceptionally in stem 16S rRNA and 12S rRNA, which is necessary for stability in the

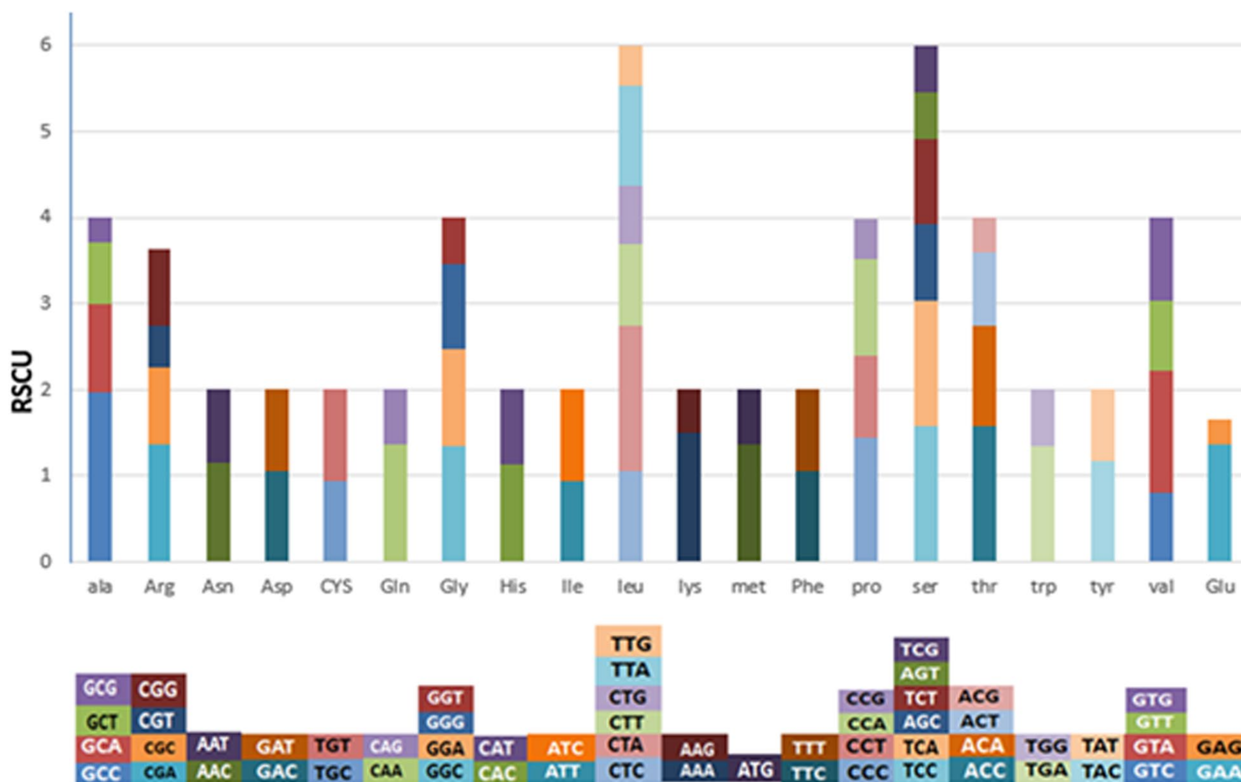

Fig. 2 Relative synonymous codon usage (RSCU) of the mitochondrial protein-coding genes and codon usage of *Huso huso* for whole genome sequencing

stem structure, and the deviation against G observed less [4, 40, 41].

tRNAThr and tRNAPro relative to cytochrome b differ from those determined in other vertebrates.

In region 5′ 12S rRNA, a 52-nucleotide VNTR sequence was observed with the following sequence, which could be the result of mispairing:

AGGCTTGGTCCTGGCCTTACTATCAATTTT AACCCAATTTACACATGCAAGT

This sequence could produce stems, loops, and a stable structure (Fig. S1). Of the tRNAs, 14 were encoded by the heavy strand, these tRNAs are 67 to 75 bp. All tRNAs except ser (AGY) could be converted to clover leaves, the structure of which could be determined by tRNA scanning. tRNAser was reduced in the dehydrooridine arm, transforming its second structure into a short clover leaf, similar to most metazoa [42] (Fig. 3).

Noncoding regions

Noncoding regions in mtDNA include OL and CR and several regions between genes. The size of the CR has been 994 bp, which is located downstream of tRNAper. This region is rich in AT and is considered the largest noncoding region.

CR contains several TASs, and the TAS sequence extends into several CR regions. This sequence is associated with CSB-conserved sequence blocks. The conserved CSB-1–3 sequence blocks are downstream of this region. The TAS region is rich in repeats and is the most variable part of the CR. It contains duplicate elements with a length of 82 bp. It contains a conserved TACAT motif, and the TAS motif is an inverse complement to ATGTA. The TAS motif can be paired with the CTAS motif, resulting in the formation of stable hairpin loops, which may also serve as a specific sequence signal to terminate mtDNA replication [43]. Repeat sequences were identified in this region in fish of different species [44–46].

A comparison of mtDNA sequences in sturgeon and conserved sequences of termination replication (TAS) sturgeons with changes in the number of consecutive duplicate sequences showed that they could form stable structures during mtDNA replication. In the mtDNA of sturgeons, the control region responsible for the termination of H-strand replication contains one to seven variable number tandem repeats (VNTRs) with a unit size of 78 to 83 bp [47]. The D-loop of sturgeon mtDNA differs from the human D-loop and contains more than one TAS; therefore, the termination of mtDNA replication cannot be explained by helicase activity alone. Additionally, it is not still determined why sturgeons exhibit a wide range of haplotypes differing in VNTR length and, correspondingly, in the number of TAS elements [47]. Kornienko et al. [47] identified VNTR regions containing highly conserved sequences that terminate with a CAT triplet in all sturgeon species studied. TAS elements are located in repeating units that constitute the VNTR. TAS nucleotide sequences are associated with the termination of mtDNA replication. An ineffective DNA repair system and a lack of protective histones in this organelle cause the mutation rate in mitochondrial DNA to be higher

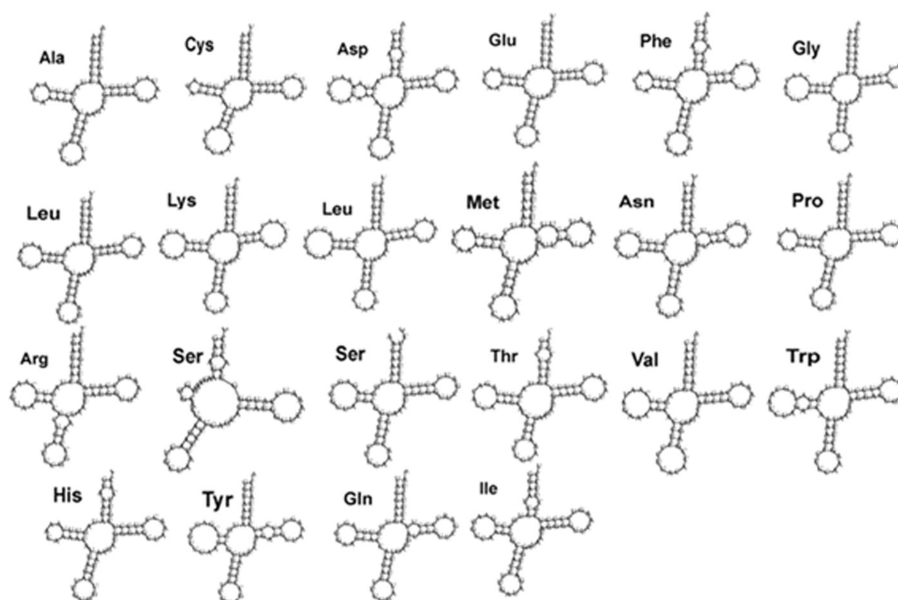

Fig. 3 The secondary structure of tRNA genes in *H.huso*

than in the DNA nucleus, resulting in VNTRs. Tandem repeats in this region increase the length of the D-loop, and a lower D-loop increases the rate of supercoil formation [43]. This issue has an effect on protein bonding and accessibility as well as on transcription and replication [48, 49] (Fig. 4).

Variation in repeat tandems in this region probably evolved through the process of illegitimate elongation [45]. It occurs during mtDNA replication and is possibly the mechanism mediated by slipped strands and mispairing [50].

We identified conserved CSBD and CSB-1, CSB-2, and CSB-3 sequence blocks with high similarity of CSB sequences from other fishes (Fig. 4). While five conserved block sequences have been reported in the conserved central domain in mammals in the control region, only 3 CSBs are typically found in bony fishes [51–53]. CSB-D is critical in maintaining regulatory performance in CR and is considered the most protected part of CR (95% of its sequence is conserved). The CSB 1, 2, and 3 sequences had the least conserving. Consensus sequences for CSB could be found here:

CSB-D, TACTGGCATCTGATTAAT;
CSB-1, TAATAGATAGTGAATGATATAATG ACATA;
CSB-2, CAAACCCCCTACCCCC;
CSB-3, TGTCAAACCCCAAAAGCA.

In addition, three conserved sequences were found downstream of the protected central conserved domain, and CSB-1 is an AT-rich region following a GACATA conserved motif.

The downstream CSBs are separated by a poly C and are identified by TTA or TA. CSB 2 consists of a sequence with a polyC stretch. These CSBs are involved in forming primer RNA for mtDNA amplification and play a vital role in the RNA switch in DNA synthesis, which begins in the OL region [29, 54].

The noncoding OL region is located in a cluster of five tRNA genes (trp (W), Ala (A), Asn (N), Cys (C), and Tyr (Y)). It is called the WANCY region. The OL region has 34 nucleotides and is identified by a stable loop stem structure with a GC-rich stem and a T-rich loop. Two common features of vertebrate OL are the 5' flanking region rich in pyrimidine (stem) and a motif (5′-CTT CCT-3′) found in the stem (Fig. 5).

Both may be related to the accuracy and efficiency of DNA replication in OL, as seen in the human mitochondrial genome [55].

Phylogenetic analysis

The control regions and the regions that caused the instability in the phylogenetic tree were discarded (the high homogeneity of the genes encoding the protein, due to their high capability in phylogenetic function, were used in the experiment). We performed partition maximum

Fig. 4 The D-loop region, along with the CSB-conserved blocks

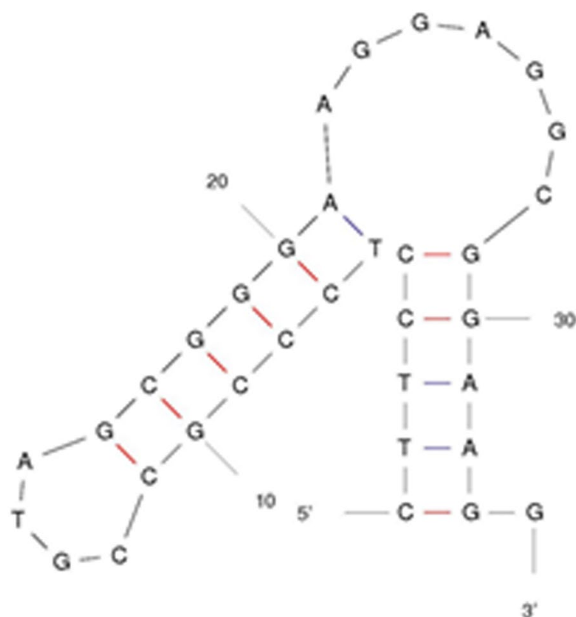

Fig. 5 The OL region was found in *Huso huso*

likelihood (ML) using the concatenated nucleotide sequences of the 12 protein-coding genes.

We excluded VNTRs because they are rapidly evolving sequences, which may lead to multiple substitutions at some sites. Therefore high heteroplasy would reduce the resolution of our phylogeny. The sturgeon species created two clades in the phylogenetic trees (Atlantic, Pacific).

Mitogenomic phylogeny has put *Acipenser stellatus* as a sister group of *Huso huso* with high support (bootstrap 100% and posterior probability 1.00). *Huso huso* similarly formed with high support (bootstrap 100%) a sister group to the remaining species of clade Atlantic. This topology is consistent with the phylogenies of the complete mitochondrial genome of Liao et al. [2], Li et al. [56], and Popovic et al. [57]. This topology is inconsistent with the findings of Mugue et al. [20]. Paraphilic species with *Huso huso*, including *Acipenser guelenstaedtii*, *Acipenser baerii*, *Acipenser nudiventris*, *Acipenser fulvescens*, *Acipenser ruthenus*, *Acipenser brevirotrum*. Our results agree with the taxonomy for sturgeons within the genus and reflected monophilic and paraphilic taxa (Fig. 6) [2, 30, and 61]. The analysis of Nedoluzhko et al. [58] and Sheraliev et al. [40] showed that in the Atlantic clad, *H. huso* was put as a sister group with all remaining *Acipenser* species. Also, Sheraliev et al. [40] confirmed that A. stellatus is closely related to *P. kaufmanni* (Fig. 6). Sheraliev et al. [40] showed that *H. huso* is not closely related to *A. stellatus*, and is an ancestor clade Atlantic species, these results contradict our findings and are consistent with some results [59]. « A. oxyrinchus" and *A. sturio* formed a separate branch in the Pacific clade, clustered as paraphily with the rest species in this clade. *Acipenser guelenstaedtii* and *Acipenser*

Fig. 6 Evolutionary relationship between *Huso huso* and other sturgeon species by a phylogenetic tree

baerii formed a group monophyletic, whereas *Acipenser nudiventris* and *Acipenser ruthenus* were together monophyletic in the clade Atlantic. *A. transmontanus* and *A. schrenckii* formed group monophyletic while *A. sinensis* and *A. dabryanus* were monophyletic in the clade Pacific. Polyodontidae was used as the outgroup because it was an ancient sturgeon. It is generally believed that the sturgeons originated from the ancient population in Europe, and the early diversity occurred in Asia [40]. Short mitochondrial gene fragments have revealed limitations in complex phylogenetic relationships in many lineages. Many informative sites from longer DNA sequences, such as the complete mitochondrial genome, allow deeper branches and higher levels of relationships to resolve complex relationships. Based on more genetic data, the mitochondrial genome will completely elucidate the phylogenetic relationships and determine a higher lineage level.

The substitution rates genes

More details showed that the divergence between substitutions in the mitochondrial genome of rRNA and tRNA was less than that in the protein coding genes, indicating the stability of the stem and loop in the rRNA and tRNA. The CR had the most divergence. The divergence ratio varies between the genes encoding the protein, and ATP8, ND3, ND6, Cytb, and ND4 L showed the highest conservation, but ND2 and COII showed the highest divergence. ND2 is a small protein-encoding gene that is amplified by a primer and can be used as a protein-encoding gene with good divergence in barcoding species is much more suitable than COI (the complete gene can be used instead of the partial gene). Applying these new genomic sequences to taxonomic tests of this species will be very useful. It will have achievements for phylogenetic analysis and the study of lineage rearrangements, conservation, and evaluation of biological studies.

We calculated the ratio of nonsynonymous substitutions (Ka) to the rate of synonymous substitution (Ks) (Fig. S2). Our results show that all 13 genes have a Ka/Ks ratio lower than one, indicating a strong selection signal for harmful mutations in all mitochondrial protein genes. However, the mean Ka/Ks ratio (0.1 to 0.001) showed a significant difference between individual genes. The highest rate (ND2, COII) indicates that the purification selection is under minimal pressure.

Conclusions

We identified the complete mitochondrial genome consisting of 22 tRNAs, two rRNAs, 13 protein-encoding genes, and two OL and D-loop control regions. The VNTR is found in three sites, and a VNTR is between ND6 and Cytb as pseudo-tRNA-Glu. This organization of the genome was separate from other previously reported sturgeon species. These results included increased tRNA in H. huso and a VNTR in the 12 SrRNAs.

The 13 genes encoding proteins show that they are less conserved than rRNA and tRNA. Divergence varies between genes. COII was the highest, while the ATP8 synthetase subunit was the lowest. Observing the divergence rates of these genes allows us to compare them for barcoding. Although COI is commonly used for DNA barcoding, this gene has observed low divergence, indicating that it cannot clearly distinguish very close species. The data analysis showed that ND2 is a better candidate for barcode identification in the grouping. It has a higher percentage of variable sites than COI. Therefore, it can differentiate between newly derived species.

In addition, the smaller size of ND2 makes it easier to use compared to COI. The sequence was amplified entirely using only one pair of primers. Thus, a complete gene instead of a partial gene can be used for barcoding.

Nevertheless, we observed a kind of diversity in this particular species, which included the presence of tandem repeats in the control region and 12S rRNA and tRNA-Glu, resulting from mispairing. There have been reports of pseudo tRNA in several species of fish, but there have been no reports about the presence of repeats in fish in 12S rRNA.

Methods

Sampling, PCR amplification, and sequencing

Samples of sturgeon were taken from the waters of Iran on the south coast of the Caspian Sea. Systematic studies were performed to select a random animal specimen to avoid exaggerating the effects and to conform to the conventional statistically significant criteria. Samples were received from three provinces along the south coast of the Caspian Sea (Mazandaran, Golestan, Gilan) in five fishing areas, and samples were collected from the caudal fin. Complete genomic DNA was obtained using the ammonium acetate method [60], and 16 pairs of universal primers overlapping mitochondrial genome fragments were amplified and then sequenced.

Primers and PCRs were performed based on the methods described by Shao et al. [61] (Table S1). One hundred µl of PCR products, along with 50 µl of each forward and reverse primer (10 pmol), were used to determine the sequences of DNA fragments using the Sanger method by the European company of Microsynth.

Assembling genome sequences and annotating

The sequencing results were manually corrected and edited using (ChromaSprov.1.42), and then the mitochondrial genome was searched for protein and rRNA using the BLAST tool. http://www.ncbi.nlm.nil.gov/

BLAST.cgi, and annotations were performed based on the mitochondrial genome alignment of closely related species in the GenBank database. The obtained sequences from 16 pieces were aligned with Clustal X and then edited with BioEdit software.

The overlapping sequences were achieved, and the results of the contigs were assembled using the SeqMan module of Lasergene 11.0 software (DNASTAR, Madison, WI, USA) and then mapped to reach the complete mitochondrial genome of Huso huso. Sequence results and combined annotations were used to map the species' genome. Most tRNA genes and their secondary structures were predictable by tRNA-scan [62]. The cloverleaf structure was detected by computer.

Sequence analysis

Nucleotide compositions were obtained using the DNASTAR program, and ATskew and GCskew were calculated using the formula [56].

$$\text{ATskew} = [A - T]/[A + T] \text{ and GCstew } = [G - C]/[G + C]$$

Tandem Repeat Finder was used to identify duplicate sequences. The complete mitochondrial genome sequencing of Ka/Ks, the ratio of protein-coding genes, was calculated by PAL2NAL online (http://www.bork.embl.de/pal2nal). The codon usage of 13 protein-coding genes was calculated using Mega 10 software.

Phylogenetic analysis

In this study, mitogenomes of additional species were obtained from GenBank and aligned using Clustal X. After removing their stop codons to draw the phylogenetic tree, the genes became concatenated. We used European, Chinese, Russian, and American sturgeon species for phylogeny. Sequences of 12 genes encoding proteins were aligned with Clustal X, and minor manual settings were used.

To compare the differences, probability tests were performed. Different models were used for comparison and Determination, and the best model selection based on BIC was determined for 12 protein-coding genes. Finally, GTR+I+G was selected as the best model, and 12 genes based on maximum likelihood and gamma distribution were evaluated using all sites and bootstrap 1000.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12862-023-02166-2.

Additional file 1: Fig. S1. The secondary structure of the variable number of tandem repeats (VNTR) in the 12S rRNA gene. **Fig. S2.** The average Ka/Ks ratio of 13 protein-coding genes Ka/Ks is the nonsynonymous substitution rate (Ka) ratio to the synonymous substitution rate (Ks). **Table S1.** Primer sequences and product size. **Table S2.** The average base composition of 13 protein-coding genes in Huso huso.

Acknowledgements

We thank the experts from the Caspian Sea Ecology Institute in Sari, who were involved in the sampling of sturgeon fish for this research. Caspian Sea Ecology Institute played no role in the study design, analyses, data interpretation, and manuscript writing.

Authors' contributions

KD Conceptualization, Formal analysis, Investigation, Methodology Writing – original draft, and Writing – review & editing. GR Conceptualization, Formal analysis, Investigation, Project administration, Supervision, and Writing – review & editing. AB Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, and Writing – review & editing. AF Conceptualization, Formal analysis, Investigation, Methodology, and Writing – review & editing.

Funding

This research is a part of the project to study "Molecular phylogeny assessment in sturgeon species of Caspian Sea using complete mitochondrial genome sequences," which was submitted. Samples are related to the approved project entitled Identification and isolation of microsatellite in Iranian sturgeon with number 2–76-11–86047 (Ministry of Jihad Agriculture, AREEO).This work was financially supported by a fund from Sari Agricultural Sciences and Natural Resources University (SANRU). Our gratitude is also extended to the Caspian Sea Ecology Institute in Sari personnel for the facilities to conduct this research.

Availability of data and materials

DNA sequence data generated and analyzed in this manuscript were deposited in a public database, NCBI (Accession number: MK213068). https://www.ncbi.nlm.nih.gov/nuccore/MK213068.

Declarations

Ethics approval and consent to participate

The study was conducted following the guidelines of the Iranian Council of Animal Care and approved by the ethics committee of Sari Agricultural Sciences and Natural Resources University with the Ethical code SANR1400-1.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 18 October 2022 Accepted: 8 September 2023
Published online: 26 September 2023

References

1. Dugo MA, Kreiser BR, Ross ST, Slack WT, Heise RJ, Bowen BR. Conservation and management implications of fine-scale genetic structure of Gulf sturgeon in the Pascagoula River. Mississippi J Appl Ichthyol. 2004;20(4):243–51. https://doi.org/10.1111/j.1439-0426.2004.00572.x.
2. Popović D, Baca M, Panagiotopoulou H. Complete mitochondrial genome sequences of Atlantic sturgeon, Acipenser oxyrinchus oxyrinchus, Gulf sturgeon, A. o. desotoi and European sturgeon A. sturio (Acipenseriformes: Acipenseridae) obtained through next generation sequencing. Mitochondrial DNA Part A. 2016;27(4):2549–51. https://doi.org/10.3109/19401736.2015.1038799.
3. Pourkazemi M, Skibinski DF A, Beardmore J. Application of mtDNA d-loop region for the study of Russian sturgeon population structure from Iranian coastline of the Caspian Sea. J Appl Ichthyol. 1999;15(4–5):23–8. https://doi.org/10.1111/j.1439-0426.1999.tb00199.x.
4. Timoshkina NN, Barmintseva AE, Usatov AV, Mugue NS. Intraspecific genetic polymorphism of Russian sturgeon Acipenser gueldenstaedtii. Russ J Genet. 2009;45:1098–106. https://doi.org/10.1134/S1022795409090117.

5. Congiu L, Fontana F, Patarnello T, Rossi R, Zane L. The use of AFLP in sturgeon identification. J Appl Ichthyol. 2002;18(4–6):286–9. https://doi.org/10.1046/j.1439-0426.2002.00376.x.
6. Xiao TQ, Lu CY, Li C, Cheng L, Cao DC, Sun XW. An AFLP-based approach for the identification of sex-linked markers in A mur sturgeon A cipenser schrenckii B randt, 1869. J Appl Ichthyol. 2014;30(6):1282–5. https://doi.org/10.1111/jai.12553.
7. Yarmohammadi M, Pourkazemi M, Ghasemi A, Hasanzadeh SM, Chakmehdouz F. AFLP reveals no sex-specific markers in Persian sturgeon (Acipenser persicus) or beluga sturgeon (Huso huso) from the southern Caspian Sea. Iran Prog Biol Sci. 2011;1(1):55–60.
8. Barmintseva AE, Mugue NS. Genetic Variation of the Siberian Sturgeon (*Acipenser baerii* Brandt, 1869) in Aquaculture. Russ J Genet. 2018;54(2):210–7. https://doi.org/10.1134/S1022795418020035.
9. Georgescu SE, Burcea A, Florescu I, Popa OG, Dudu A, Costache M. Microsatellite variation in Russian sturgeon (Acipenser gueldenstaedtii) from aquaculture. Sci Papers Anim Sci Biotechnol. 2014;47(1):73–6.
10. Georgescu S, Canareica O, Dudu A, Costache M. Analysis of the Microsatellite Variation in the Common Hybrid Between Russian Sturgeon (Acipenser Gueldenstaedtii Brandt and Ratzeburg, 1833) and Siberian Sturgeon (Acipenser Baerii Brandt, 1869) from Aquaculture. Transylv Rev Syst Ecol Res. 2013;15(2):117–24. https://doi.org/10.2478/trser-2013-0024.
11. May B, Krueger CC, Kincaid HL. Genetic variation at microsatellite loci in sturgeon: primer sequence homology in Acipenser and Scaphirhynchus. Can J Fish Aquat Sci. 1997;54(7):1542–7. https://doi.org/10.1139/f97-061.
12. Moghim M, Tan SG, Javanmard A, Pourkazemi M, Malar J. Inheritance of microsatellite loci and their application for pedigree analysis in the polyploid Persian sturgeon (*Acipenser persicus*). Zool Stud. 2012;51(8):1507–14.
13. Ogden R, Gharbi K, Mugue N, Martinsohn J, Senn H, Davey JW, Pourkazemi M, McEwing R, Eland C, Vidotto M, Sergeev A. Sturgeon conservation genomics: SNP discovery and validation using RAD sequencing. Mol Ecol. 2013;22(11):3112–23. https://doi.org/10.1111/mec.12234.
14. Çiftci Y, Eroğlu O, Firidin Ş. Mitochondrial cytochrome b sequence variation in three Sturgeon species (A. stellatus Pallas, 1771, A. gueldenstaedtii Brandt, 1833, H. huso Linnaeus, 1758) from the Black Sea Coasts of Turkey. Turkish Journal of Fisheries and Aquatic Sciences. 2013;13(2). https://doi.org/10.4194/1303-2712-v13_2_11.
15. Mugue NS, Barmintseva AE, Rastorguev SM, Mugue VN, Barmintsev VA. Polymorphism of the mitochondrial DNA control region in eight sturgeon species and development of a system for DNA-based species identification. Russ J Genet. 2008;44:793–8. https://doi.org/10.1134/S1022795408070065.
16. Boscari E, Barmintseva A, Pujolar JM, Doukakis P, Mugue N, Congiu L. Species and hybrid identification of sturgeon caviar: a new molecular approach to detect illegal trade. Mol Ecol Resour. 2014;14(3):489–98. https://doi.org/10.1111/1755-0998.12203.
17. Ladoukakis ED, Zouros E. Evolution and inheritance of animal mitochondrial DNA: rules and exceptions. Journal of Biological Research-Thessaloniki. 2017;24(1):1–7. https://doi.org/10.1186/s40709-017-0060-4.
18. Shoffner JM. Oxidative phosphorylation disease. The metabolic and molecular bases of inherited disease, vol. 1. New York: McGraw-Hill; 1995.
19. Ludwig A, May B, Debus L, Jenneckens I. Heteroplasmy in the mtDNA control region of sturgeon (Acipenser, Huso and Scaphirhynchus). Genetics. 2000;156(4):1933–47. https://doi.org/10.1093/genetics/156.4.1933.
20. Mugue N, Barmintseva A, Schepetov D, Shalgimbayeva G, Isbekov K. Complete mitochondrial genomes of the critically endangered Ship sturgeon Acipenser nudiventris from two seas. Mitochondrial DNA Part B. 2016;1(1):195–7. https://doi.org/10.1080/23802359.2016.1144103.
21. Breines R, Ursvik A, Nymark M, Johansen SD, Coucheron DH. Complete mitochondrial genome sequences of the Arctic Ocean codfishes Arctogadus glacialis and Boreogadus saida reveal oriL and tRNA gene duplications. Polar Biol. 2008;31:1245–52. https://doi.org/10.1007/s00300-008-0463-7.
22. Cui Z, Liu Y, Li CP, Chu KH. Species delineation in Pampus (Perciformes) and the phylogenetic status of the Stromateoidei based on mitogenomics. Mol Biol Rep. 2011;38:1103–14. https://doi.org/10.1007/s11033-010-0207-y.
23. Mabuchi K, Miya M, Satoh TP, Westneat MW, Nishida M. Gene rearrangements and evolution of tRNA pseudogenes in the mitochondrial genome of the parrotfish (Teleostei: Perciformes: Scaridae). J Mol Evol. 2004;59:287–97. https://doi.org/10.1007/s00239-004-2621-z.
24. Williams EP, Peer AC, Miller TJ, Secor DH, Place AR. A phylogeny of the temperate seabasses (Moronidae) characterized by a translocation of the mt-nd6 gene. J Fish Biol. 2012;80(1):110–30. https://doi.org/10.1111/j.1095-8649.2011.03158.x.
25. Zhuang X, Cheng CH. ND6 gene "lost" and found: evolution of mitochondrial gene rearrangement in Antarctic notothenioids. Mol Biol Evol. 2010;27(6):1391–403. https://doi.org/10.1093/molbev/msq026.
26. Zhuang X, Qu M, Zhang X, Ding S. A comprehensive description and evolutionary analysis of 22 grouper (Perciformes, Epinephelidae) mitochondrial genomes with emphasis on two novel genome organizations. PLoS One. 2013;8(8):e73561. https://doi.org/10.1371/journal.pone.0073561.
27. Hassanin A, Leger NE, Deutsch J. Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of Metazoa, and consequences for phylogenetic inferences. Syst Biol. 2005;54(2):277–98. https://doi.org/10.1080/10635150590947843.
28. Clayton DA. Replication of animal mitochondrial DNA. Cell. 1982;28(4):693–705. https://doi.org/10.1016/0092-8674(82)90049-6.
29. Clayton DA. Nuclear gadgets in mitochondrial DNA replication and transcription. Trends Biochem Sci. 1991;16:107–11. https://doi.org/10.1016/0968-0004(91)90043-U.
30. Cantatore P, Gadaleta MN, Roberti M, Saccone C, Wilson AC. Duplication and remoulding of tRNA genes during the evolutionary rearrangement of mitochondrial genomes. Nature. 1987;329(6142):853–5. https://doi.org/10.1038/329853a0.
31. Jacobs HT, Asakawa S, Araki T, Miura KI, Smith MJ, Watanabe K. Conserved tRNA gene cluster in starfish mitochondrial DNA. Curr Gen. 1989;15:193–206. https://doi.org/10.1007/BF00435506.
32. Inoue JG, Miya M, Tsukamoto K, Nishida M. Evolution of the deep-sea gulper eel mitochondrial genomes: large-scale gene rearrangements originated within the eels. Mol Biol Evol. 2003;20(11):1917–24. https://doi.org/10.1093/molbev/msg206.
33. Kong X, Dong X, Zhang Y, Shi W, Wang Z, Yu Z. A novel rearrangement in the mitochondrial genome of tongue sole, Cynoglossus semilaevis: control region translocation and a tRNA gene inversion. Genome. 2009;52(12):975–84. https://doi.org/10.1139/G09-069.
34. Wolstenholme DR. Animal mitochondrial DNA: structure and evolution. Int Rev Cytol. 1992;141:173–216.
35. Miya M, Nishida M. Organization of the mitochondrial genome of a deep-sea fish, *Gonostoma gracile* (Teleostei: Stomiiformes): first example of transfer RNA gene rearrangements in bony fishes. Mar Biotechnol. 1999;1(5):416–26.
36. Southern ŠO, Southern PJ, Dizon AE. Molecular characterization of a cloned dolphin mitochondrial genome. J Mol Evol. 1988;28:32–42. https://doi.org/10.1007/BF02143495.
37. Ortí G, Petry P, Porto JI, Jégu M, Meyer A. Patterns of nucleotide change in mitochondrial ribosomal RNA genes and the phylogeny of piranhas. J Mol Evol. 1996;42:169–82. https://doi.org/10.1007/BF02198843.
38. Reyes A, Gissi C, Pesole G, Saccone C. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. Mol Biol Evol. 1998;15(8):957–66. https://doi.org/10.1093/oxfordjournals.molbev.a026011.
39. Tanaka M, Ozawa T. Strand asymmetry in human mitochondrial DNA mutations. Genomics. 1994;22(2):327–35. https://doi.org/10.1006/geno.1994.1391.
40. Sheraliev B, Peng Z. Complete mitochondrial genome sequence and phylogenetic position of the Amu Darya sturgeon, Pseudoscaphirhynchus kaufmanni (Acipenseriformes: Acipenseridae). J Appl Ichthyol. 2020;36(4):389–92. https://doi.org/10.1111/jai.14043.
41. Wang HY, Lee SC. Secondary structure of mitochondrial 12S rRNA among fish and its phylogenetic applications. Mol Biol Evol. 2002;19(2):138–48. https://doi.org/10.1093/oxfordjournals.molbev.a004066.
42. Garey JR, Wolstenholme DR. Platyhelminth mitochondrial DNA: Evidence for early evolutionary origin of a tRNA ser AGN that contains a dihydrouridine arm replacement loop, and of serine-specifying AGA and AGG codons. J Mol Evol. 1989;28:374–87. https://doi.org/10.1007/BF02603072.
43. Saccone C, Pesole G, Sbisá E. The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern. J Mol Evol. 1991;33:83–91. https://doi.org/10.1007/BF02100199.
44. Broughton RE, Dowling TE. Evolutionary dynamics of tandem repeats in the mitochondrial DNA control region of the minnow Cyprinella

spiloptera. Mol Biol Evol. 1997;14(12):1187–96. https://doi.org/10.1093/oxfordjournals.molbev.a025728.
45. Buroker NE, Brown JR, Gilbert TA, O'hara PJ, Beckenbach AT, Thomas WK, Smith MJ. Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. Genetics. 1990;124(1):157–63. https://doi.org/10.1093/genetics/124.1.157.
46. Saitoh K, Hayashizaki K, Yokoyama Y, Asahida T, Toyohara H, Yamashita Y. Complete nucleotide sequence of Japanese flounder (Paralichthys olivaceus) mitochondrial genome: structural properties and cue for resolving teleostean relationship. J Hered. 2000;91(4):271–8. https://doi.org/10.1093/jhered/91.4.271.
47. Kornienko IV, Chebotarev DA, Makhotkin MA, Grigoriev VA, Ponomareva EN, Matishov GG. Termination of replication and mechanisms of heteroplasmy in sturgeon mitochondrial DNA. Mol Biol. 2019;53:107–17. https://doi.org/10.1134/S0026893319010060.
48. Raymakers C, Hoover C. Acipenseriformes: CITES implementation from Range States to consumer countries. J Appl Ichthyol. 2002;18(4–6):629–38. https://doi.org/10.1046/j.1439-0426.2002.00398.x.
49. Ruhanen H, Borrie S, Szabadkai G, Tyynismaa H, Jones AW, Kang D, Taanman JW, Yasukawa T. Mitochondrial single-stranded DNA binding protein is required for maintenance of mitochondrial DNA and 7S DNA but is not required for mitochondrial nucleoid organisation. Biochim Biophys Acta Mol Cell Res BBA-Mol Cell Res. 2010;1803(8):931–9. https://doi.org/10.1016/j.bbamcr.2010.04.008.
50. Fumagalli L, Taberlet P, Favre L, Hausser J. Origin and evolution of homologous repeated sequences in the mitochondrial DNA control region of shrews. Mol Biol Evol. 1996;13(1):31–46. https://doi.org/10.1093/oxfordjournals.molbev.a025568.
51. Guo X, Liu S, Liu Y. Comparative analysis of the mitochondrial DNA control region in cyprinids with different ploidy level. Aquaculture. 2003;224(1–4):25–38. https://doi.org/10.1016/S0044-8486(03)00168-6.
52. Lee WJ, Conroy J, Howell WH, Kocher TD. Structure and evolution of teleost mitochondrial control regions. J Mol Evol. 1995;41(1):54–66.
53. Liu M, Li JL, Ding SX, Liu ZQ. Epinephelus moara: a valid species of the family Epinephelidae (Pisces: Perciformes). J Fish Biol. 2013;82(5):1684–99. https://doi.org/10.1111/jfb.12112.
54. Doersen CJ, Guerrier-Takada C, Altman S, Attardi G. Characterization of an RNase P activity from HeLa cell mitochondria. Comparison with the cytosol RNase P activity. J Biol Chem. 1985;260(10):5942–9. https://doi.org/10.1016/S0021-9258(18)88920-7.
55. Hixson JE, Wong TW, Clayton DA. Both the conserved stem-loop and divergent 5'-flanking sequences are required for initiation at the human mitochondrial origin of light-strand DNA replication. J Biol Chem. 1986;261(5):2384–90. https://doi.org/10.1016/S0021-9258(17)35948-3.
56. Perna NT, Kocher TD. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J Mol Evol. 1995;41:353–8. https://doi.org/10.1007/BF00186547.
57. Liao X, Tian H, Zhu B, Chang J. The complete mitochondrial genome of Chinese sturgeon (Acipenser sinensis). Mitochondrial DNA Part A. 2016;27(1):328–9. https://doi.org/10.3109/19401736.2014.892101.
58. Nedoluzhko AV, Sharko FS, Tsygankova SV, Boulygina ES, Barmintseva AE, Krasivskaya AA, Ibragimova AS, Gruzdeva NM, Rastorguev SM, Mugue NS. Molecular phylogeny of one extinct and two critically endangered Central Asian sturgeon species (genus Pseudoscaphirhynchus) based on their mitochondrial genomes. Sci Report. 2020;10(1):722. https://doi.org/10.1038/s41598-020-57581-y.
59. Shen Y, Yang N, Liu Z, Chen Q, Li Y. Phylogenetic perspective on the relationships and evolutionary history of the Acipenseriformes. Genomics. 2020;112(5):3511–7. https://doi.org/10.1016/j.ygeno.2020.02.017.
60. MuCQown EC, Sloss BL, Sheehen RJ. Microsatellite analysis of genetic variation in sturgeon. New primer sequences for Scaphyhinchus and Acipenser. Trans Am Fish Soc. 2000;129:1380–8. https://doi.org/10.1577/1548-8659.
61. Shao W, Wei L, Li L, Wang H, Lin Z, Chen J. Universal DNA primers for amplification complete mitochondrial genome for sturgeons. Conserv Genet Resour. 2014;6:305–7. https://doi.org/10.1007/s12686-013-0112-5.
62. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. InNucleic acids symposium series 1999 Jan 1 (Vol. 41, No. 41, pp. 95–98). [London]: Information Retrieval Ltd., c1979-c2000.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions