Thesis for the degree of master of fisheries science

Molecular characterization of The mitochondrial genome of *Silurus soldatovi* from Nakdong River, Republic of Korea

by

Mahmoud Aboalyamin Soliman

KOICA-PKNU International Graduate Program of Fisheries Science

Graduate School of Global Fisheries

Pukyong National University

February 2020

Molecular characterization of the mitochondrial genome of *Silurus soldatovi* from Nakdong River, Republic of Korea

낙동강에서 채집된 *Silurus soldatovi* 미토콘드리아 게놈의 분자생물학적 특성연구

Advisor: Prof. KIM Hyun Woo

by

Mahmoud Aboalyamin Soliman

A thesis submitted in partial fulfillment of the requirement
for the degree of

Master of Fisheries Science
in KOICA-PKNU International Graduate Program of Fisheries Science,
Graduate School of Global Fisheries,
Pukyong National University

February 2020

Molecular characterization of the mitochondrial genome of *Silurus soldatovi* from Nakdong River, Republic of Korea

A dissertation

by

Mahmoud Aboalyamin Soliman

Approved by:

(Chairman) Prof. KIM Young Mog

(Member) Prof. PARK Wongyu

(Member) Prof. KIM Hyun-Woo

February 21, 2020

Table of Contents

List of Figures

List of Tables

Molecular characterization of the mitochondrial genome of *Silurus soldatovi* from Nakdong River, Republic of Korea

Mahmoud Aboalyamin Soliman

KOICA-PKNU International Graduate Program of Fisheries Science

Graduate School of Global Fisheries

Pukyong National University

Abstract

The entire mitochondrial genome structure of Soldatov's catfish, *Silurus soldatovi* has been characterised. The samples were gathered from a tributary of River Nakdong, South Korea. The specimens were subjected to bioinformatics assembly concerning next generation sequencing (NGS) declaims to establish mitogenome. The intact circular mitogenome of *S. soldatovi* was 16525 base pairs or bps long and determined 13 canonical protein-coding genes, 2 rRNAs as well as 22 tRNAs. Further, the mitogenome encompasses 2 non-coding regions, a control region designated as *D-loop* (890 bp long), an origin locus for light strand replication designated as O_L (31 bp long). These regions were well conserved in the mitogenome of *S. soldatovi*. Additionally, the A+T content was found to be 56.20 percent

while the G+C content was a little less and found to be 43.80 percent. The complete gene sequence of mitogenome of *S. soldatovi* was found to be matching with other Silurids. As many as 28 genes were positioned on the H strand, while the remaining nine genes were located on the L strand. Excluding COX1 gene with the initiation codon GTG, 12 protein-coding genes were anticipated to have ATG as the start codon. When compared to other presently identified mitogenome sequences of the family Siluridae, *Silurus soldatovi* exhibited the greatest similarity of 99.38 percent with Chinese haplotype *S. soldatovi* (NC022723) and of 97.17 percent with Chinese haplotype *Silurus asotus* (JX087351).

Keywords: Next generation sequencing, *Silurus soldatovi*, Mitochondrial genome

List of Abbreviations

µL Microliters

µM Micromoles

COI Cytochrome C oxidase subunit I

DMSO Dimethyl Sulfoxide

dNTPs Deoxynucleotide triphosphates

Kb Kilobytes

NGS Next generation sequencing

ORFs Open-ended reading frames

OL Light –Strand Replication Origin

PBS Phosphate Buffer Saline

PCGs Protein-coding genes

PCR Polymerase Chain Reaction

RAPDs Randomly Amplified Polymorphic DNA

RE Restriction Enzymes

rRNA Ribosomal Ribonucleic acid

TRFs Tree reading frames

tRNA Transfer Ribonucleic acid

TE Tris-EDTA

Introduction

Scientific management of fish species demands diverse information such as population size, age, reproduction, outcomes of fishing, the genetic structure for statistical analysis (Beaumont, Boudry, & Hoare, 2010). Genetic information is imperative to evaluate the genetic diversity of fish species for their sustainability (Okumuş & Çiftci, 2003). The potential of genetic information applications for recognizing and classifying fish populations has been accredited since ages. In the early 1930s, L. Von Toth (de Ligny, 1969), specified his efforts to establish blood group variation in fish in order to distinguish herring populations. Following this, in the 1950s and 1960s, a number of laboratories initiated research involved concerning analysis and explanation of blood group distinction in fish. Another approach extensively used today and has significantly influenced the genetic approach encompasses analysis of genetically determined electrophoretic distinction of proteins. Sick (1961), one of the pioneers in fish studies, defined haemoglobin variants in whiting (*Gadus melangus*) and cod (*Gadus morhua*). An extension to such methods comprise the investigation of enzymes using histochemical stains which is responsible for the evolution of population genetic studies of animals including fish (Ward, 2000). Such findings gained prominence and were encouraged by the ground-breaking studies of *Drosophila* by Lewontin & Hubby (1966) and on humans by Harris (1966), thereby revealing the ubiquity of such discrepancies. Evidently, protein electrophoretic

variation is rather an indirect measure of genetic dissimilarity; amino-acid substitutions cause variations, which eventually influence DNA modifications. Recognition and purification of restriction enzymes (RE), capable of generating nick in the DNA strand at a specific site paved the way for direct DNA alterations. Reports dealing with nuclear DNA were found to be confined to examination of minisatellite variability. These minisatellites are cyclic array or repeats of short oligonucleotide with less than 65 bps, the procedure is signified as DNA fingerprinting (Ward, 2000). Polymerase Chain Reaction (PCR) technology is regarded as the most advanced and significant scientific development in biological science. The process involves a rapid amplification of desired DNA template to generate specific DNA sequences. The procedure requires very little DNA and therefore it amplifies short fragments of a longer DNA segments. DNA amplification is carried out for diverse processes specially to assess the DNA sequence for variability of size and also for mutation analysis. Consequently, the procedure is exploited to perform the genetic analysis of fish. Techniques like RAPDs (Randomly Amplified Polymorphic DNA), AFLP (amplified fragment length polymorphisms), EPICs (exon-primed intron-crossing amplification), microsatellites (this involves repeats of 1-4 nucleotide base pairs and are usually less than 300 bp in aggregation) (Ward, 2000).

1.1. Application of DNA Markers

DNA markers technique is exploited to observe genetic variation in the entire genome of a population. It is gaining wider acceptance to study variation in population genetics as well as to understand the dissimilarities prevailing in a given population. The process gives a deeper understanding towards the existence of divergences (Askari, Shabani, & Kolangi Miandare, 2013).

In order to study a wider population and to understand the prevailing differences, several DNA markers have been developed comprising microsatellites, mtDNA, RAPDs, SNP (Single nucleotide base pair). The fish population also revealed a greater propensity of genetic variation when the fish genome was examined using these DNA markers. Therefore, several novel attempts are being performed to enhance the efficacy of DNA markers and other procedures like genome mapping and species identification. Major advancement has been observed for DNA barcoding and Expressed Sequence Tags (EST), which were imperative for the development in DNA sequencing procedures for numerous aquaculture species (Askari et al., 2013).

1.2. Nuclear DNA Polymorphism

The procedure of nuclear DNA polymorphism is capable of providing wider opportunities to study the mechanisms of evolution. However, when employed with nuclear DNA markers, there are some hitches, encompassing recombination, heterozygosity, gene□ specific variation in rate and history, selection (non□neutrality), low divergence, insertion/deletion polymorphism, polytomy, PCR and sequencing difficulties etc.

Nonetheless, some of the nuclear genes are not present in mitochondrial DNA markers (Zhang & Hewitt, 2003). Such specific DNA markers could be exploited to study the complete genome sequence.

An enormous proportion of nuclear DNA markers engaged in population analysis are generally the noncoding regions (referred as introns), since they are regarded as the regions with more variables when compared to the coding regions (referred as exons). However, for mitochondrial DNA markers, utilize exons or the coding regions enormously, the mitochondrial DNA is smaller in size and is devoid of intronic sequences.

Besides, nucleotide substitutions, insertion or deletions (indels) frequently comprise a huge portion of the perceived polymorphism. Thus, arrangements of indels (gaps or modification due to insertion in aligned genetic structures) encompass an effective phylogenetic information that cannot be disregarded. However, major existing phylogenetic inferring approaches do not utilize such gaps proficiently, therefore, gaps are either overlooked or treated as ambiguities and obscurities (Zhang & Hewitt, 2003).

For nuclear DNA, the problem is more intricate, considering the fact that diverse genomic sections may bear distinct rates of evolution and eventually moulds distinctive evolutionary forces. A few of these forces encompass population logjam together with random drift that may portray genome□wide impact. However, another approach depicting regional influence is recombination or selection. In case of nuclear markers, the relationship among the gene tree and the population of the species tree dangles on the interaction between these forces. Of these, the recombination or selection exercises a greater or pre-eminent influence

on the recovery of an appropriate gene tree, it is intricate to recover the correct gene tree and not to refer to the population tree. Currently, knowledge and understanding about the genetic and evolutionary perspectives of genomic sequences engaged or retained as molecular markers are confined only to a few model organisms while for others it is a long way to go (Zhang & Hewitt, 2003).

Compared to the nuclear genes, the mitochondrial DNA (mtDNA) is extensively exploited for phylogenetic studies since the features of coding matter conservation, rapid evolution, maternal inheritance as well as low intensities of intermolecular genetic recombination (Cui et al., 2017). Considering the fact that mtDNA dissimilarities are a more sensitive gauge of population singularities comprising hybridization and bottlenecks compared to the nuclear DNA owing to the smaller size of mtDNA (Okumuş & Çiftci, 2003). In a similar manner, sex-specific variances in gene flow could be exhibited by associating nuclear DNA with mitochondrial DNA (Okumuş & Çiftci, 2003). Owing to the fact that mitochondrial DNA is physically alienated from the rest of cellular DNA, moreover, it is moderately convenient to isolate (Carvalho & Pitcher, 2012). mtDNA is a circular molecule, 16kb long; furthermore, the fish mitochondrial genome contains 13 genes responsible for coding protein molecules specifically 2 rRNAs and 22 tRNAs, together with two non-coding sequences. Moreover, the mitochondrial genome of the fish also contains the site of origin for the light strand replication designated as O_L as well as the control region designated as the *D-loop*. The genomic sequences of mitochondrial genes are conserved and are arranged in a similar fashion as in vertebrates for as many as 37 genes, while the two non-coding

regions are similar from hagfish to eutherian mammals. In a similar manner, one non-coding segment of DNA, which is known to act as the initiation site for mitochondrial DNA and RNA transcript is also found to be conserved (Beaumont et al., 2010).

Cytochrome B (Cyt B) or cytochrome C oxidase I (COI), partial DNA sequences are extensively used genetic markers for the identification of species (Nicolas et al., 2012), Conversely, the rationality of exploiting COI as standard barcode is subjected to dialog, and researchers have discovered it is less effectual for some taxon units (Dai et al., 2012).

Furthermore, it is evident that the entire mitochondrial genome sequences are persuasive for the evolutionary as well as bio-geographical study. The complete information of mitochondrial DNA sequences play a vital role in distinguishing and identifying the genetic structure and complete biology of the organism including aquatic life-forms (Chauhan & Rajiv, 2010; Okumuş & Çiftci, 2003; Yáñez, Houston, & Newman, 2014).

Evidently, the entire mitochondrial DNA sequence is attained through the next generation sequencing technology (NGS). Initially, only short sequences of mtDNA were targeted using NGS technology, with technological advancement, the entire mitochondrial genome was subjected to investigation and evaluation (Hahn, Bachmann, & Chevreux, 2013). Such technological improvement and development of sequencing technologies has brought revolution in the form of reduced sequencing cost, augmentation in sequencing capacity as well as enabling individual laboratories to be equipped with the procedures to sequence the genome of their model organisms as well as increasing the possibility of resequencing (Varshney, Nayak, May, & Jackson, 2009).

The Soldatov's catfish, *Silurus soldatovi* is a non-migratory, vital, freshwater fish species, bearing commercial significance in aquaculture commerce and industry. *S. soldatovi* is broadly dispersed in the Far East predominantly in Heilongjiang River and the Ussuri River between Russia and China (Bogutskaya, Naseka, Shedko, Vasileva, & Chereschnev, 2008) (Fig. 1).

S. soldatovi is similar to *S. asotus* . Whereas, *S. asotus* is identified by the anterior edge of the pectoral spine being prominently serrated; and the tooth band on vomer continuous (Chu, Zheng, & Dai, 1999). We here determined and characterized the complete mitogenome of *S. soldatovi* (MN171302), which was collected from Nakdong River. The complete mitogenome of *S. soldatovi* would provide useful information for its management in Korea.

Fig. 1. The Soldatov's catfish, *Silurus soldatovi* habitat (www.geologypage.com).

Materials and Methods

2.1. Sample collection and preservation

The sample was collected from a tributary of Nakdong River, South Korea (E128°06`27.73", N35°32`20.96") in 2018 (Fig. 2). The collected specimen was instantly stored in 96 percent ethanol (DUKSAN PURE CHEMICALS, KOREA) and kept in a deep freezer (at -20°C) in the Department of Marine Biology, Pukyong National University, South Korea.

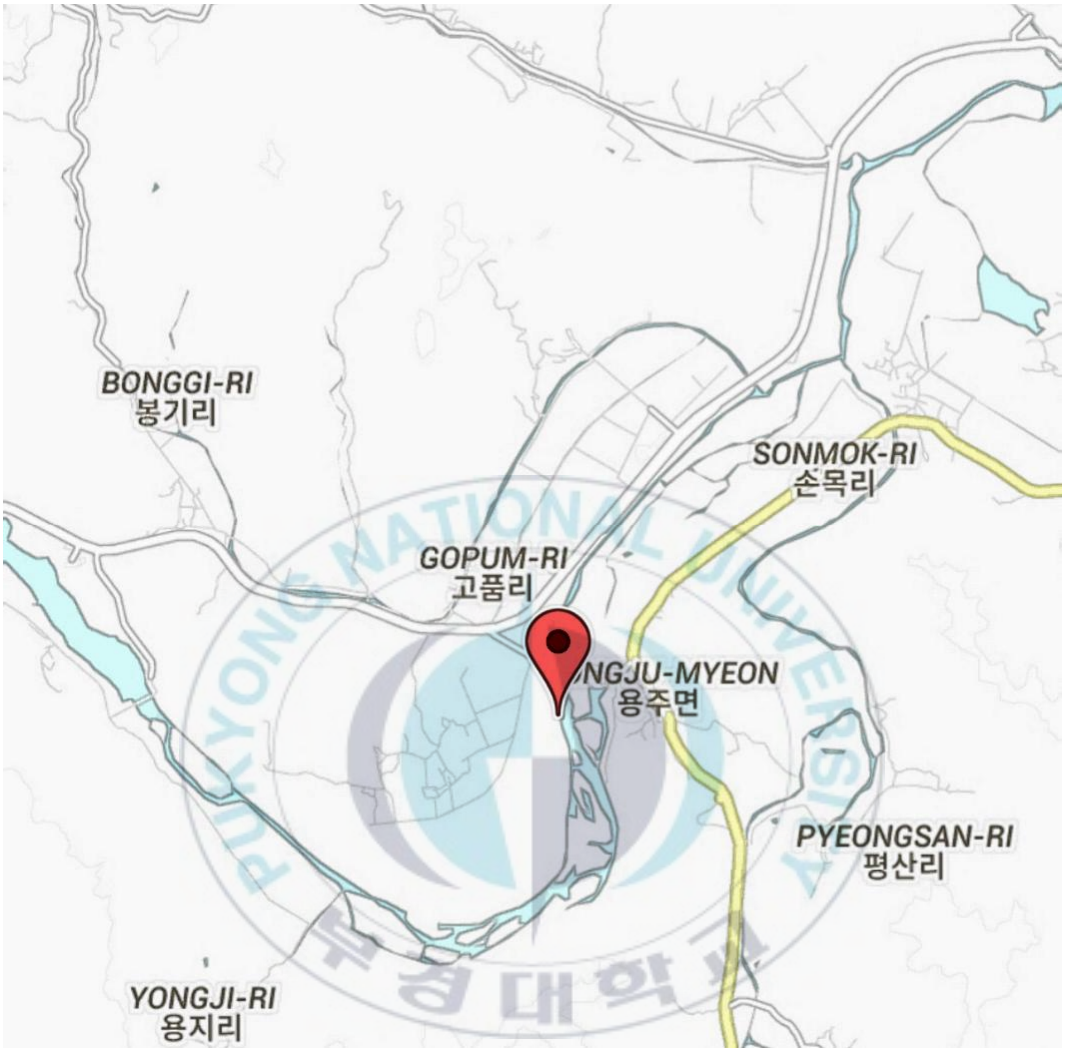

Fig. 2. Sample collection site (maps.google.com).

2.2. Genomic DNA extraction

From the collected sample, genomic DNA was isolated using DNeasy® Blood and Tissue Kit (Qiagen, Germany), following the protocol mentioned in the kit. 200mg muscle tissue was taken from the specimen and held in 1X lysis buffer (600µl), this was followed by a homogenization process using motorized Tissue lyser II (Qiagen, Hilden, Germany). Subsequently, Proteinase K (30.0 U/mg) and Sodium Dodecyl Sulfate 2% (SDS) (Biosesang, Korea) were added. This was followed by incubation of sample at 60°C for 12 hours for complete lysis of tissue. After the incubation interval, 600µl of binding buffer designated as the GC buffer was added followed by the addition of 300µl of Isopropanol. The column tube was used to filter supernatant using centrifugation at 13,000 rpm for 1 minute. The column was washed twice with 500µl washing buffer1 and buffer2. The genomic DNA was then eluted using 100µl of TE buffer. With the use of the NanoDrop spectrophotometer, ND-1000 (Thermo Scientific, Waltham, MA, USA), and 1% agarose gel electrophoresis the intensity of the isolated genomic DNA was determined. The isolated genomic DNA was stored at -20°C for further studies and examinations.

2.3. Polymerase chain reaction and sequencing

Analysis of the COI region of the specimen was performed using PCR along with a pair of universal PCR primers, BCH and BCL (Folmer, Hoeh, Black, & Vrijenhoek, 1994). Further, the quality check of the primers was confirmed using OligoAnalyser 3.1 (http://sg.idtdna.com/calc/analyzer), while synthesis was done commercially (Bioneer Co. Korea). 20µl of the PCR reaction mixture contained 11.2µl autoclaved or sterilized

deionized water, 2µl each of 10X Ex Taq buffer and deoxynucleotide triphosphates (dNTPs) mixture (2.5µM Takara, Japan), 0.6µl Dimethyl Sulfoxide (DMSO), 1µl each primer (0.5µM forward and reverse), 0.2µl Ex Taq DNA polymerase (Takara, Japan). To this 2.0µl genomic DNA was added as a template.

The PCR was programmed with the following parameters:

- The opening or the initial denaturation at 94°C for 3 minutes duration.
- 35 cycles of denaturation at 94°C for 30 seconds.
- Annealing for 30 seconds at 50°C.
- Extension at 72°C for 45 seconds.

The splitting of the PCR yields was accomplished using 1.5% agarose gel electrophoresis after staining with the loading star (DynEBio, Cat. No. A750, Republic of Korea). AccuPrep® Gel purification kit was used as per the protocol of the company (BIONEER, Korea) to purify the PCR products targeting the COI partial sequences. With the help of the dideoxynucleotides analog method (BIONEER, Korea), the DNA sequence of PCR products was commercially determined. Analysis and identification of the procured DNA were carried out using the program Basic Local Alignment Search Tool (BLAST) (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

2.4. Full mitochondrial genome isolation

The determination of the entire mitochondrial DNA sequence of *S. soldatovi* was performed employing the next-generation sequencing (NGS) platform. Isolation of the mitochondrial DNA (mtDNA) from the fish specimen was done using the mitochondrial DNA isolation kit (Abcam, USA) as per instructions mentioned in the kit. The procedure involved dissection of the specimen to take out 200mg muscle tissue of the specimen, followed by washing with ice-cold Phosphate Buffer Saline (PBS) for three times, the tissue was then centrifuged at 600 x g at 4°C for 5 minutes. The supernatant was removed, and the tissue sample was resuspended in 1.0ml of 1X Cytosol Extraction Buffer. This was followed by incubation on ice for 10 minutes. This was followed by the tissue-homogenization in an ice-cold Dounce tissue grinder. To check the efficacy of tissue homogenization process, 2µl of the homogenized suspension was pipetted onto a coverslip and observed under the microscope to determine if cells are intact or crushed. The homogenate was then transferred to a new 1.5ml microcentrifuge tube followed by centrifugation at 700 x g at 4°C for 10 minutes, this aids in removing nuclei as well as intact cells. Again, the supernatant was collected in a new 1.5 ml tube followed by centrifugation at 10,000 x g at 4°C for 30 minutes. After discarding the supernatant, the pellet was resuspended in 1ml 1X cytosol extraction buffer, this was followed by centrifugation at 10,000 x g at 4°C for 30 minutes. The supernatant was discarded, the pellet collected was isolated mitochondria. The lysis of mitochondria was carried out using 30µl mitochondrial lysis buffer, held for 10 minutes on ice. The procedure was followed by the addition of 5µl

enzyme mix and then incubated in a water bath maintained at 50°C for 90 minutes duration or till the solution turns out to be clear. To this, 100µl absolute ethanol was added, mixed and retained for 10 minutes at -20°C. This step was followed by centrifugation using a microcentrifuge at the maximum speed at room temperature for 5 minutes. Pellet was collected and the supernatant was discarded, the pellet was regarded as mitochondrial DNA. The mtDNA pellet was washed twice using 1ml of 70 percent ethanol, traces of ethanol were removed with the help of pipette tip; for 5 minutes, the mtDNA was air-dried. The procedure was followed by resuspension of DNA in 20µl TE buffer. The extracted DNA was then kept at -20°C for future studies. The Fragmentation of the purified mitochondrial DNA into smaller sizes (~350 bp) was performed using Covaris M220 Focused-Ultrasonicator (Covaris Inc., USA).

The formation of a library was carried out by means of TruSeq® sample preparation kit v2 (Illumina, USA); utilizing the fragmented sequences, the quality and the quantity of the designed library was measured with 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Further, the NGS sequencing was carried out using the Illumina Miseq platform 2x300 bp pair ends (Illumina, USA); while the mitogenome assembly was done with the help of Geneious® 11.0.2 software (Kearse et al., 2012).

Results

3.1. Gene organization in the mitochondrial DNA

The acquired partial COI sequence of the sample exhibited 99.61 percent and 97.03 percent nucleotide sequence similarity with the *S. soldatovi* (AB860299) and *S. asotus* (JX087351) from China. For *S. soldatovi* (GenBank accession number MN171302), the entire circular mitogenome was found to be 16,525 bp long, while the complete A+T content of the *S. soldatovi* mitogenome was found to be 56.20 percent; furthermore, the gene arrangement of the organism was instituted akin to other Siluriform fish. When the relationship between purine and pyrimidine was compared, the C content was found to be the highest while the G content was the lowest (Miya et al., 2003). This pattern was practically similar to Silurids which displays approximately 16 percent G composition (Alam et al., 2019; Alam, Andriyono, Hossain, Eunus, & Kim, 2019; Chen et al., 2012; Wang et al., 2015).

The present study unearthed that the mitogenome of *S. soldatovi* possesses Adenine as the highest content 5046 (30.53 percent), followed by Cytosine 4635 (28.04 percent), while Thymine content was reported to be 4244 (25.68 percent) and Guanine content was found as 2603 (15.75 percent).

Table 1. The Gene organization of complete mitochondrial genome of *Silurus soldatovi*

Gene	Position		Size (bp)	Codon		Intergenic nucleotide (bp)	Strand
	Start	Stop	Nucleotide	Start	Stop		
tRNA Phe	1	71	71			0	H
12S rRNA	72	1021	950			0	H
tRNA Val	1022	1096	75			0	H
16S rRNA	1097	2770	1674			0	H
tRNA Leu	2771	2845	75			0	H
ND1	2846	3820	975	ATG	TAA	0	H
tRNA Ile	3821	3894	74			2	H
tRNA Gln	3893	3963	71			0	L
tRNA Met	3964	4030	67			-1	H
ND2	4032	5076	1045	ATG	T--	0	H
tRNA Trp	5077	5148	72			0	H
tRNA Ala	5149	5219	71			0	L
tRNA Asn	5220	5292	73			0	L
OL	5293	5323	31			0	
tRNA Cys	5324	5394	71			-3	L
tRNA Tyr	5398	5467	70			-1	L
COX1	5469	7019	1551	GTG	TAG	1	H
tRNA Ser	7019	7090	72			-3	L
tRNA Asp	7094	7167	74			-12	H
COX2	7180	7870	691	ATG	T--	1	H
tRNA Lys	7870	7945	76			0	H
ATP8	7946	8113	168	ATG	TAA	10	H
ATP6	8104	8787	684	ATG	TAA	1	H
COX3	8787	9570	784	ATG	T--	1	H
tRNA Gly	9570	9644	75			1	H
ND3	9644	9992	349	ATG	T--	2	H
tRNA Arg	9991	10066	76			3	H
ND4L	10064	10360	297	ATG	TAA	7	H
ND4	10354	11734	1381	ATG	T--	0	H
tRNA His	11735	11804	70			0	H
tRNA Ser (UCN)	11805	11871	67			-1	H
tRNA Leu	11873	11947	75			1	H
ND5	11947	13773	1827	ATG	TAA	4	H
ND6	13770	14288	519	ATG	TAA	0	L
tRNA Glu	14289	14357	69			-1	L
Cyt b	14359	15496	1138	ATG	T--	0	H
tRNA Thr	15497	15567	71			3	H
tRNA Pro	15565	15635	71			0	L
Control Region (D-loop)	15636	16525	890			0	H

3.2. Protein-coding genes

The gathered and identified sequences of mitochondrial DNA were further annotated or interpreted and evaluated. Protein-coding genes were identified using NCBI ORF finder (http://www.ncbi.nlm.nih.gov/orffinder/), subsequently, deciphered and decoded by a coalition of homologous genes of the other established and published Telestos mitogenomes. Eventually, as many as 13 protein-coding genes were recognized in the mitogenome of *S. soldatovi,* of these, 12 genes were determined on the heavy strand (H), while only ND6 gene was found to be located on the light strand (L) (Fig. 4). A variation in base-pair is observed in protein-coding genes comprising 168 bp (ATP8) to 1827 bp (ND5), which is deliberated to accumulate 11409 bp in length demonstrating around 69.04 percent of the entire mitogenome of *Silurus soldatovi* (Table 1). The initiation codon of 12 protein-coding genes (PCGs) was ATG, while COX1 starts with GTG.

In *Silurus soldatovi*, the mitogenome possess six open-ended reading frames (ORFs) specifically ND1, ATP6, ATP8, ND5, ND4L, and ND6 termination or end codon was TAA; while COX1 comprises TAG as the end codon. Other protein-coding genes contained incomplete (TA-/T--) end or termination codons especially in ND2, ND3, COX2, COX3, ND4 and *Cyt b*. Five PCGs including COX3, ATP6, ND3, ND4L, and ND6 enclosed overlapped nucleotides extending from 1-10, although five PCGs like COX1, COX2, ND1, ND2, and *Cyt b* comprised intergenic spacers (1-6 bp) nucleotides.

On the contrary, ND1, ND6, ND4, ND2, and *Cyt b* had null or zero intergenic nucleotide spacer. In the case of *Silurus soldatovi* 13 PCGs overlaps of tree reading frames (TRFs) are

acknowledged on the analogous strand specifically ND5 and ND6 overlap by 4 nucleotides. On the other hand, ND4 and ND4L displayed an overlap of 7 nucleotides, while ATP6 and ATP8 show an overlap of 10 nucleotides (Table 1). The complete terminating or stop codon bearing TAA was present in the six genes, whereas, in the case of the COX1 gene, the terminating codon was TAG.

In six genes namely ND2, ND3, COX2, COX3, ND4 and *Cyt b* bear (TA- or T--) as an incomplete stop codon. As displayed in Table 1, the negative (-) numerals reveal overlapping nucleotide between two genes, the positive (+) numerals specify the intergenic space between two genes in the mitogenome, while zero (0) showed that neither overlap nor space exists between genes.

3.3. Ribosomal RNA and transfer RNA genes

Boore et al., (2005) identified two ribosomal RNAs (12S rRNA and 16S rRNA) genes as well as 22 transfer RNAs (tRNA) genes from the mitogenome of *S. soldatovi* akin to any other bony fish. The small 12S rRNA gene was positioned between $tRNA^{Phe}$ and $tRNA^{Val}$; while the large 16S rRNA was located between $tRNA^{Val}$ and $tRNA^{Leu}$. These genes are known to be 950 bp and 1674 bp long respectively, accounting for an accumulative 2624 bp (15.88 percent) of the entire mitogenome.

The twenty-two tRNA genes (Table 1) of *S. soldatovi* mitogenome were found to be scattered among other genes, comprised 1586 bp diverged in sizes ranging from 67 bp to 76 bp in length, accounting for 9.6 percent approximately of the entire mitogenome. Further analysis revealed that of 22 genes, eight tRNA genes were positioned on the L-strand

whereas, the remaining 14 tRNA genes were located on the H-strand (Fig. 4). Five tRNA genes exhibited overlap nucleotides ($tRNA^{Ile}$, $tRNA^{Gln}$, $tRNA^{Gly}$, $tRNA^{Arg}$, and $tRNA^{Thr}$) that diverse from 1-4 bp. On the other hand, two tRNAs ($tRNA^{Tyr}$ and $tRNA^{Asp}$) encompassed intergenic spacer extending from 1-12 nucleotides, while the remaining fifteen tRNA genes comprised zero intergenic nucleotides. Nevertheless, in the case of *Silurus soldatovi*, the secondary putative structure of 22 tRNA genes displayed huge resemblance as it is in the mitogenomes of the other animals.

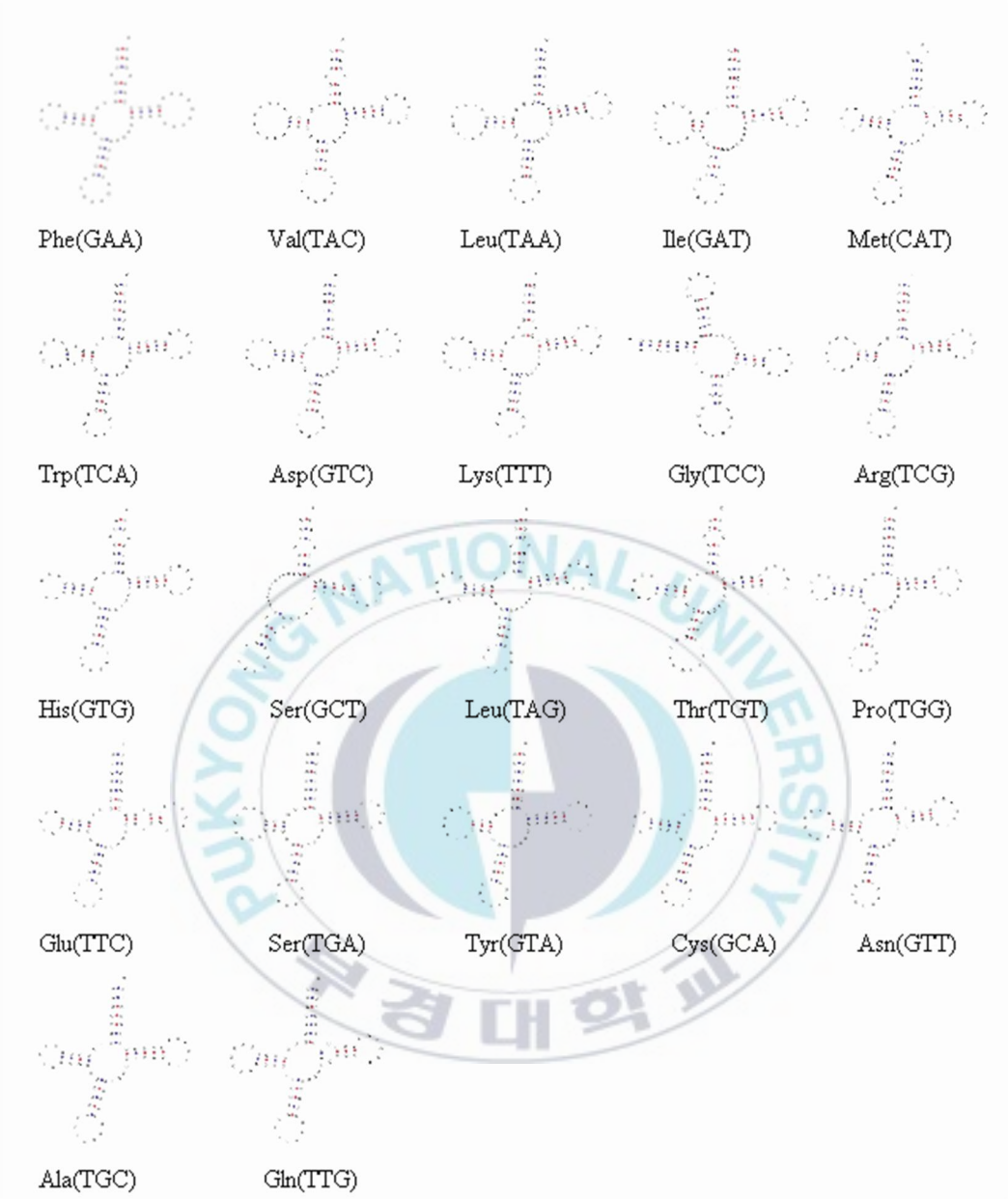

Fig. 3. A putative secondary structure of twenty-two tRNA genes of *Silurus soldatovi* mitogenome.

3.4. Non-coding region

Two non-coding regions namely the origin of the light strand replication O_L and *D-loop* control region were discovered in the mitochondrial genome of *Silurus soldatovi*. The O_L was positioned between two tRNA genes ($tRNA^{Asp}$ and $tRNA^{Cys}$) and was reported to be 31 bp long while the control region (*D-loop*) was located between two tRNA genes ($tRNA^{Pro}$ and $tRNA^{Phe}$) and was found to be of 890 bp, contributing approximately 5.40 percent of the entire mitogenome (Fig. 4).

The control region of mtDNA of *S. soldatovi* was smaller when compared with the established size of the control region of mtDNA of the Chinese haplotype of *Silurus soldatovi* (NC022723, 891 bp), moreover, it was found to be identical to the Japanese haplotype of *Silurus asotus* (NC015806, 890 bp) (Nakatani et al., 2011; C. Wang et al., 2016).

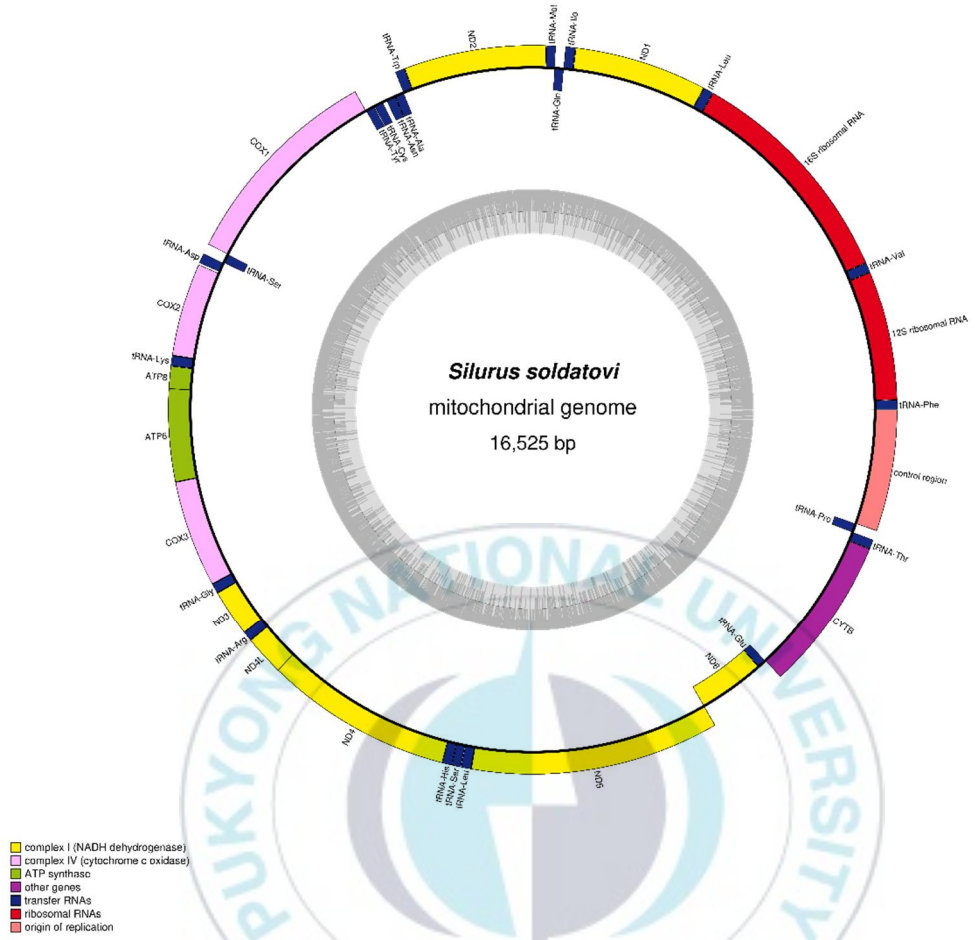

Fig. 4. The Mitochondrial gene map of *Silurus soldatovi*.

3.5. Construction of the Phylogenetic tree for *Silurus soldatovi*

MEGA7 program of the Minimum Evolution algorithm (Kumar et al., 2016) was used to construct a phylogenetic tree of the Silurid mitogenomes. The known mitogenomes in the GenBank database revealed that the specimen of *S. soldatovi* taken from Nakdong River displayed the greatest nucleotide sequence identity (99.38 percent) with the Chinese haplotype of *S. soldatovi* (NC022723).

This resemblance was followed by identity with the Chinese haplotype of *S. asotus* (JX087351), which displayed 97.17 percent identity (Fig. 5). Considerably, the two Chinese haplotypes of *S. asotus* (JX087351 and JN116720) were found to be more diligently associated with two *S. soldatovi* (MN171302 and NC022723) compared to their Japanese and other Chinese haplotypes (JX256247 and NC015806). Furthermore, it was revealed that the genetic distance between the Korean haplotype of *S. soldatovi* (MN171302) as well as the Chinese haplotype of *S. asotus* (JX087351) was found to be 0.032. On the other hand, the genetic distance between two *S. asotus* from China (JX087351 and JX256247) was found to be 0.054. This demonstrates greater or superior intraspecies variation.

With the presently reported entire mitogenomes of the family Siluridae, a phylogenetic tree was created using the MEGA 7.0 software through the Minimum Evolution (ME) algorithm using 1,000 bootstrap replications. The GenBank accession figures were exhibited, this was followed by the scientific name of each species.

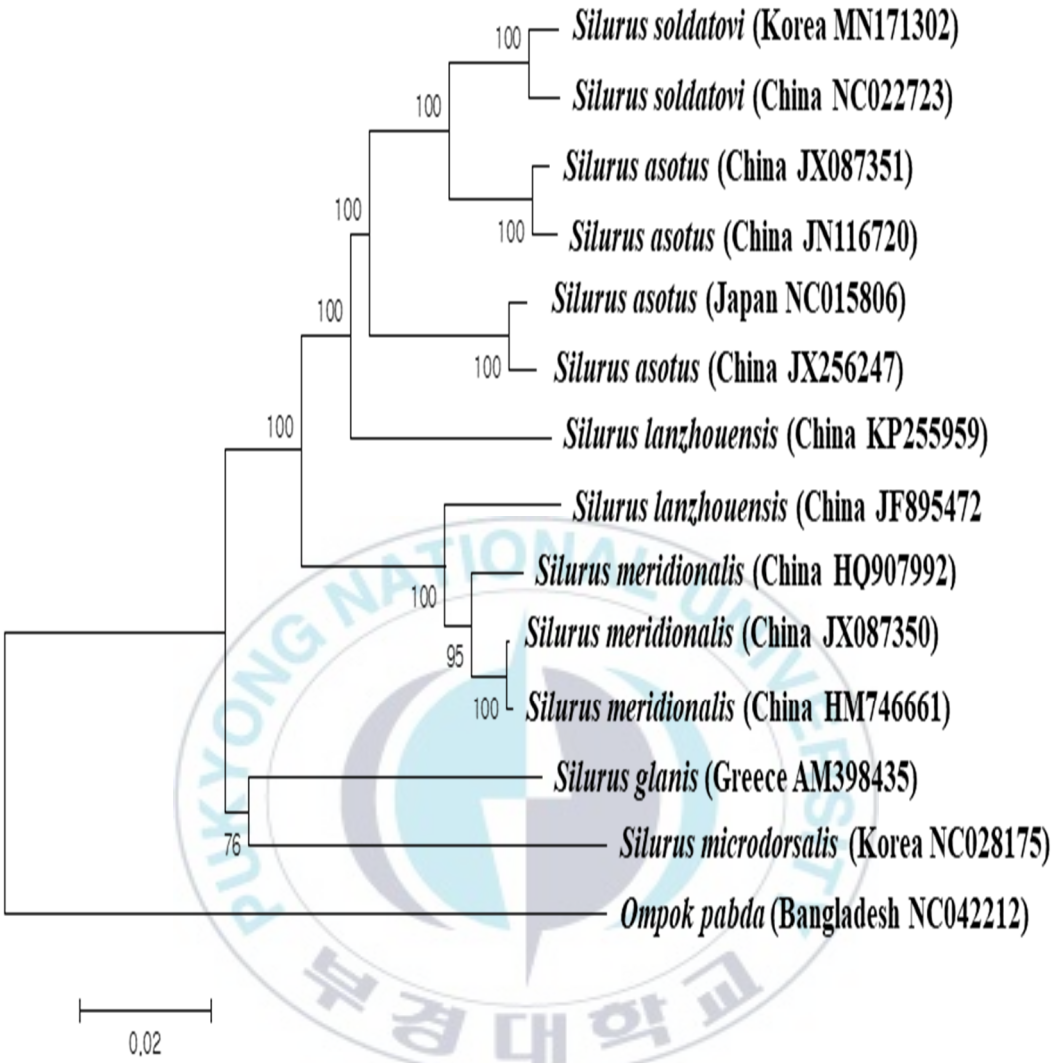

Fig. 5. The Phylogenetic relationship of *Silurus soldatovi* in the family siluridae

Discussion

The gene organization as well as the gene order of 37 genes was alike and matching with the sequence and order of the reported species of Siluridae (Alam et al., 2019; Nakatani et al., 2011; Wang et al., 2015). The structural organization and translation direction of the mitogenome was predictably akin to other published vertebrates mitogenome (Boore, Macey, & Medina, 2005; Miya & Nishida, 1999).

The complete mitogenome (16,525 bp long) of *Silurus soldatovi* (MN171302) was found to be shorter than the nearest species, the Chinese haplotype of *S. soldatovi* (16527 bp, NC022723); while it was found to be longer than *Silurus asotus* (16521 bp, NC015806). The overall base sequence or composition of the *Silurus soldatovi* mitogenome was found as- Adenine: 30.53 percent, Thymine: 25.68 percent, Guanine 15.75 percent and Cytosine: 28.04 percent. This picture reveals anti-guanine biases similar to other vertebrates. Consequently, the overall G+C content (43.80 percent) was less as compared to the A+T content (56.20 percent), indicating greater purine and pyrimidine in the A+T content. In the mitochondrial genome the tRNAs were bundled in three conserved forms (IQM, WANCY, and HSL) (Satoh et al., 2016). The IQM (Isoleucine, Glutamine, and Methionine) cluster was detected between ND1 and ND2, while the WANCY (Tryptophan, Alanine, Asparagine, Cysteine, and Tyrosine) group was positioned between ND2 and COX1. The HSL (Histidine, Serine, and Leucine) bundle was sited between the ND4 and ND5 genes

in the mitogenome of *S. soldatovi*. Thus, indicating that the whole mitochondrial genome study is a vital tool in biological progression, owing to a comparatively fast pace in base substitution (Martin & Palumbi, 1993), for understanding maternal inheritance (Birky, Fuerst, & Maruyama, 1989), and comparatively ease of extraction, protection, and manipulation (Dowling, 1990). For the conservation of freshwater fish reserves, it is imperative to use mitochondrial DNA in diverse ways for instance, in order to analyse the genetic discrepancies within the population, to ascertain the evolutionarily deviating group of population and to monitor and examine the conservation worth of population from phylogenetic or evolutionary interpretations (Moritz, 1994). The distinction of the phylogenetic centered on diverse genes can be created (ND4, ND5, COX1, and *cyt-b*) as well as beneficial for considering the evolutionary pace in different species in a distinctive genus. The consequences recommended that the wide-ranging mitochondrial genome could afford trustworthy information for molecular-based research involving environmental DNA evaluation in diverse water bodies (Kelly et al., 2014; Yamamoto et al., 2017).

Conclusion

The present study characterized the complete mitochondrial genome arrangement of *Silurus soldatovi*, collected from Nakdong River, South Korea. The mitogenome size of *S. soldatovi* was found to be 16,525 bp. The structure of the mitochondrial genome was identical with other fish. The phylogenetic evaluation utilizing the complete mitochondrial genome exhibited the utmost similarity with the Chinese haplotype of *Silurus soldatovi* (NC022723) and *Silurus asotus* (JX087351) bearing 99.38 percent and 97.17 percent nucleotide sequence distinctiveness respectively. The genetic distance concerning the Korean haplotype of *S. soldatovi* (MN171302) and the Chinese haplotype of *S. asotus* (JX087351) was found to be 0.032, while the genetic distance between two *S. asotus* from China (JX087351 and JX256247) was found to be 0.054 unveiling greater intraspecies variation therefore, further re-examination of these two species is required to establish a better understanding of their evolutional correlation. The concluded genetic outcomes will be utilized as novel tools, these are advantageous for additional thoughtfulness of the evolutionary and phylogenetic categorization of Silurids species. Furthermore, they could be utilized for environmental DNA examination studies supported by the GenBank database for the freshwater fish species in Korean waters.

Acknowledgement

All praise to Allah the Almighty who taught man that which he knew not, and for giving me the grace, and favor to carry out this study. Secondly, thanks for everyone contributed to making me who I am today, and for my father Mr. Aboalyamin Soliman. I would like to express my appreciation, and respect to my supervisor Professor KIM Hyun Woo for his precious support during the whole study program, from which I have acquired a lot regarding my title. I could not have finished this study without the full and sincere support of my precious colleges Md. JOBAIDUL Alam and Sapto Andriyono, and last but not least, I would like to extend my thanks and appreciation to all the staff in KOICA-PKNU office especially Professor KANG Kyoung Mi for her kind attitude and help on the academic aspect of the study.

References

Alam, J., Andriyono, S., Hossain, A., Eunus, A., & Kim, H.-W. 2019. The complete mitochondrial genome of a Pabdah catfish, *Ompok pabda* (Hamilton, 1822). Mitochondrial DNA Part B, 4(1): 507-508.

Askari, G., Shabani, A., & Kolangi Miandare, H. 2013. Application of molecular markers in fisheries and aquaculture. J Animal Sci, 2: 82-88.

Beaumont, A. R., Boudry, P., & Hoare, K. 2010. Biotechnology and genetics in fisheries and aquaculture.

Birky, C., Fuerst, P., & Maruyama, T. 1989. Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. Genetics, 121(3): 613-627.

Bogutskaya, N. G., Naseka, A. M., Shedko, S. V., Vasileva, E. D., & Chereschnev, I. 2008. The fishes of the Amur River: updated check-list and zoogeography. Ichtyological Exploration of Freshwaters, 19(4): 301-366.

Boore, J. L., Macey, J. R., & Medina, M. 2005. Sequencing and comparing whole mitochondrial genomes of animals, Methods in Enzymology, Vol. 395: 311-348: Elsevier.

Carvalho, G. R., & Pitcher, T. J. 2012. Molecular genetics in fisheries: Springer Science & Business Media.

Chauhan, T., & Rajiv, K. 2010. Molecular markers and their applications in fisheries and aquaculture. Advances in Bioscience and Biotechnology, 1(04): 281.

Chen, D.-X., Chu, W.-Y., Liu, X.-L., Nong, X.-X., Li, Y.-L., Du, S.-J., & Zhang, J.-S. 2012. Phylogenetic studies of three sinipercid fishes (Perciformes: Sinipercidae) based on complete mitochondrial DNA sequences. Mitochondrial DNA, 23(2): 70-76.

Chu, X., Zheng, B., & Dai, D. 1999. Fauna Sinica: Osteichthyes: Siluriformes: Science Press, Beijing.

Cui, L., Dong, Y., Liu, F., Gao, X., Zhang, H., Li, L., Cen, J., & Lu, S. 2017. The first two complete mitochondrial genomes for the family Triglidae and implications for the higher phylogeny of Scorpaeniformes. Scientific Reports, 7(1): 1553.

Dai, Q.-Y., Gao, Q., Wu, C.-S., Chesters, D., Zhu, C.-D., & Zhang, A.-B. 2012. Phylogenetic reconstruction and DNA barcoding for closely related pine moth species (*Dendrolimus*) in China with multiple gene markers. PLoS One, 7(4): e32544.

Dowling, T. 1990. Nucleic acide II. Restriction site analysis. Molecular Systematics: 250-319.

Folmer, O., Hoeh, W., Black, M., & Vrijenhoek, R. 1994. Conserved primers for PCR amplification of mitochondrial DNA from different invertebrate phyla. Molecular Marine Biology and Biotechnology, 3: 294-299.

Hahn, C., Bachmann, L., & Chevreux, B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Research, 41(13): e129-e129.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., & Duran, C. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, 28(12): 1647-1649.

Martin, A. P., & Palumbi, S. R. 1993. Body size, metabolic rate, generation time, and the molecular clock. Proceedings of the National Academy of Sciences, 90(9): 4087-4091.

Miya, M., & Nishida, M. 1999. Organization of the mitochondrial genome of a deep-sea fish, Gonostoma gracile (Teleostei: Stomiiformes): first example of transfer RNA gene rearrangements in bony fishes. Marine Biotechnology, 1(5): 416-426.

Miya, M., Takeshima, H., Endo, H., Ishiguro, N. B., Inoue, J. G., Mukai, T., Satoh, T. P., Yamaguchi, M., Kawaguchi, A., & Mabuchi, K. 2003. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. Molecular phylogenetics and evolution, 26(1): 121-138.

Moritz, C. 1994. Applications of mitochondrial DNA analysis in conservation: a critical review. Molecular Ecology, 3(4): 401-411.

Nakatani, M., Miya, M., Mabuchi, K., Saitoh, K., & Nishida, M. 2011. Evolutionary history of Otophysi (Teleostei), a major clade of the modern freshwater fishes: Pangaean origin and Mesozoic radiation. BMC Evolutionary Biology, 11(1): 177.

Nicolas, V., Schaeffer, B., Missoup, A. D., Kennis, J., Colyn, M., Denys, C., Tatard, C., Cruaud, C., & Laredo, C. 2012. Assessment of three mitochondrial genes (16S, Cytb, CO1) for identifying species in the Praomyini tribe (Rodentia: Muridae). PLoS One, 7(5): e36586.

Okumuş, İ., & Çiftci, Y. 2003. Fish population genetics and molecular markers: II-molecular markers and their applications in fisheries and aquaculture. Turkish Journal of Fisheries and Aquatic Sciences, 3(1): 51-79.

Varshney, R. K., Nayak, S. N., May, G. D., & Jackson, S. A. 2009. Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends in Biotechnology, 27(9): 522-530.

Wang, C., Qiu, J., Peng, X., Ai, W., Huang, X., Liu, W., & Chen, S. 2016. The complete mitochondrial genome of *Polydactylus sextarius* (Teleostei, Mugiliformes). Mitochondrial DNA Part A, 27(5): 3344-3345.

Wang, K., Xu, J., Cui, J., Li, Q., Xu, P., & Sun, X. 2015. Complete mitochondrial genome of Northern Sheatfish (*Silurus soldatovi*). Mitochondrial DNA, 26(6): 891-892.

Ward, R. 2000. Genetics in fisheries management. Hydrobiologia, 420(1): 191-201.

Yáñez, J. M., Houston, R. D., & Newman, S. 2014. Genetics and genomics of disease resistance in salmonid species. Frontiers in Genetics, 5: 415.

Zhang, D. X., & Hewitt, G. M. 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. Molecular Ecology, 12(3): 563-584.