

PHYLOGENETIC ANALYSIS OF ENDEMIC FISH SPECIES OF OHRID LAKE

DOI: <http://dx.doi.org/10.18509/GBP.2018.16>
UDC: 597.552.51.063.7(497.771:285)

Emilija Georgievska
Ljubinka Sandjakoska
Chandra Kishore
Ninoslav Marina
Joice Sophia Ponraj

University of Information Science and Technology, St. Paul the Apostle, Ohrid 6000,
Republic of Macedonia

ABSTRACT

Computer-assisted analysis in biogeography is becoming one of the major research topics in the physical geography. Interestingly, the bioinformatics software tools provide good models for better understanding of a biological process. We have utilized them in finding the relationship between Ohrid lake endemic species of *Salmo Letnica* and other eight fishes from the Salmonidae family that constitute three subfamilies. The study of phylogenetics with considerable research interest in the phylogenetic relationships among the endemic fish species of Ohrid Lake is the main focus of this research article. Salmonidae family is found to be a big family constituting three subfamilies including Salmoninae family. The main research of the present work is aimed at the fishes from the Salmoninae family with more focus on the *Salmo* genus. Eight types of fishes from the mentioned genus were used in the process of creation of two different phylogenetic trees namely Neighbor-Joining and Unweighted Pair Group Method with Arithmetic Mean (UPGMA) phylogenetic tree. In order to create the trees, the accessions codes for the mitochondrial DNA sequences adopted from the National Center for Biotechnology Information website were imported into MATLAB for further analysis. The resulting trees were investigated further to understand clearly the correlation of *Salmo Letnica* with the other eight fishes.

Keywords: Salmonidae family, *Salmo* genus, mitochondrial DNA sequences, Neighbor-Joining Phylogenetic Tree, UPGMA Phylogenetic Tree

INTRODUCTION

Phylogenetic that includes the study of evolutionary relatedness and biogeography among various groups of organisms (species, populations) based on their genomic sequences has evolved much interest among researchers [1]. With the aid of phylogenetic trees, the evolutionary interrelationships among various species or other entities that are believed to have a common ancestor can be shown [2]. The nodes in a phylogenetic tree with descendants signify the most recent common ancestor of the descendants and the edge lengths represents the time [3]. Phylogenetic analyses have become a significant candidate in understanding biodiversity, evolution, ecology, and genomes [4].

The fishes from the Salmonidae Family are known to dominate as a top predator in cold water of the Northern hemisphere, but they are also introduced to many other areas in the

world feeding on smaller fish and aquatic insects [5]. They are considered as ray-finned fish belonging to the Salmoniformes order which is derived from the fish Atlantic Salmon and the genus *Salmo* [6]. The salmonids are slender fishes with forked tails and rounded scales. They exist in different sizes ranging from 13 cm (smallest) to 2 m (largest) and have mouths consisting of a single row of sharp teeth. Salmonids having anadromous lifecycle spend most of their life at sea but the reproduction process occurs in fresh water and many of them die after spawning [7]. The most important ecological factor limiting the life of salmonids is water temperature (4-8 degrees Celsius), which also means the rhythm of life cycle of the members of the salmonids is slower than in other freshwater fishes in Europe.

A wide variety of techniques has been employed to infer the systematics of the Salmonidae that include ecological [8], morphological, cytogenetic, ontogenetic, DNA restriction site differences and direct sequencing studies. The availability of new technology has greatly increased the speed with which new information can be generated. Presently, it has become easier to determine the mitochondrial sequences of DNA directly thereby providing precise characters which are comparable in diverse organisms.

The studies about the *Salmo letnica* present a significant interest for the researchers of the region and widely. Up to now, the studies have aimed at the ecology, dietary regime, reproduction, etc. On the other hand, there are very few studies can be found regarding the evolutionary relatedness. The main purpose of this study is to determine and explain the evolutionary ties between *Salmo Letnica* and a number of other fishes from the same family. In the present work, a phylogenetic analysis of fish belonging to the Salmonidae family was made. This family is divided into three subfamilies: whitefish (Coregoninae), graylings (Thymallinae) and salmon (Salmoninae). The Salmoninae subfamily has seven genus and of them is the *Salmo* genus which is analyzed in many research articles [9] and the members of *Salmo* genus will be discussed in detail.

MATERIALS AND METHODS

Lake Ohrid represents part of the natural border between southwestern Macedonia and Eastern Albania which is one of the deepest and oldest lakes under conservation of UNESCO (United Nations Educational, Scientific and Cultural Organization) from 1979 and it is considered as a living fossil museum [10]. The lake has a very unique aquatic ecosystem that makes it a special shelter for great number of inhabited endemic species. The number of the endemic species is estimated to be greater than two hundred covering the entire food-chain. This study is dedicated to one of those species– the Ohrid Trout complex *Salmo Letnica* [11]. Among the *Salmo Letnica*, seven other fishes from the genus *Salmo* (graphically represented in Figure 1) were included in this research and are explained below:

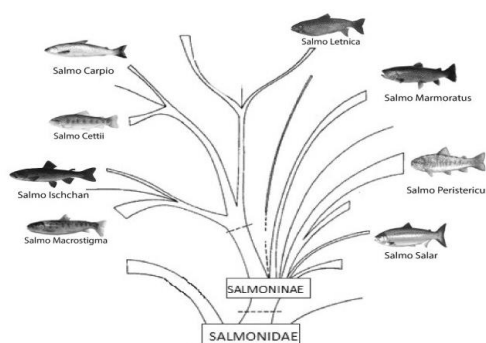


Figure 1. Graphical representation of each fish from the Salmonidae Family used in the present research

1. *Salmo letnica* [11]: It is also known by the name Ohrid Trout and is one of the endemic species in Ohrid lake. Despite the Ohrid lake, it is also found in the river Black Drim/Drin extending in Republic of Macedonia and Albania. Locally, the fish is known by the name “Koran” in Albania, and “Охридска пастрмка” (Ohridska Pastramka) in Macedonia. The fish lives at a depth of 60-80 m within the lake. Moreover, it is assigned as critically endangered in IUCN (International Union for Conservation of Nature) Red List of Threatened Species. It is characterized by different breeding areas and breeding times in which they are thought to be reproductively isolated from each other.

2. *Salmo carpio* [12]: It is widely known by the name carpione and is an endemic fish of Lake Garda in Italy. There were enormous efforts taken to inhabit this fish in many other lakes in Italy but each of them was unsuccessful. *Salmo carpio* reaches length up to 50 cm and live at a depth of 100-200 m. Apart from the mating season, they are silvery with a few black spots on the body and the number of spots gets bigger in mating season. Their maximum life span is five years.

3. *Salmo cettii* [13]: It is also known by the name Mediterranean trout which is mainly found in the Mediterranean region of Sicily, Sardinia, Corsica and on the Italian mainland in the Magra drainage. The fish lives in streams and karstic resurgences. They reach a maximum length of 40 cm and its population has decreased by around 30% in the last decade. This fish is also a part of the IUCN Red List of Threatened Species.

4. *Salmo ischchan* [14]: It is also known by the name Sevan Trout and is an endemic fish from the Lake Sevan in Armenia. In Armenia, it is well-known by the name “ishkhan” which reaches a maximum length of 90 cm breeding within the lake. This fish is now endangered owing to the various number of competitors introduced in the lake.

5. *Salmo macrostigma* [15]: This is an endemic fish species found in Algeria-Northwest Africa living in freshwater growing up to a maximum length of 60 cm.

6. *Salmo marmoratus* [16]: It is known by the name Marble Trout with a long body having very large head (25% of the body length) and because of that it is called “glavatica” in Bosnia and Herzegovina (glava=head). The fish has red spots that merge with the rest of its pigment and their natural habitats are rivers with temperature around 15 °C.

7. *Salmo peristericus* [11]: It is mainly found in Prespa Lake in Former Yugoslav Republic of Macedonia (FYROM). Four populations of the fish are known: one in the Agios Germanos stream in Greece, and the three others in the rivers Brajcinska, Kranska and Leva Reka in FYROM mainly living in the mountain stream. The population is present in 24.5 km out of 34.5 km and their habitat is fragmented by some impassable waterfalls with part of the species range covered by a national park. *Salmo Peristericus* is also listed in the IUCN Red List of Threatened Species.

8. *Salmo salar* [12]: It is naturally found in the Atlantic Ocean and rivers that flow in it which is also found in North Pacific due to human introduction. This fish is also known by different names such as silver salmon, outside salmon, bay salmon and black salmon. All of the above-described fishes belong to the Salmoniformes order that has one family-Salmonidae. This family has three subfamilies of Salmoninae, Coregoninae and Thymallinae where Salmoninae forms the major part of the tree. Salmoninae subfamily includes seven genera: Hucho, *Salmo*, Eosalmo, *Oncorhynchus*, *Brachymystax*, *Salvethymus* and *Salvelinus*. The eight fishes studied in this research belong to *Salmo* genus.

The phylogenetic analysis is usually performed in the following five steps as detailed below:

STEP 1: Choosing appropriate markers for the phylogenetic analysis

- STEP 2: Multiple sequence alignments
 STEP 3: Selection of an evolutionary model
 STEP 4: Phylogenetic reconstruction
 STEP 5: Evaluation of the phylogenetic tree

As a phylogenetic marker can be used for inferring the evolutionary relationship among the taxa. A phylogenetic marker can be DNA, RNA, protein etc. For analyzing the phylogenetic relationships, both coding and non-coding genetic regions can be used. In our analysis of the eight fishes, we have adopted the mitochondrial DNA (mtDNA). That is, DNA located in mitochondria and cellular organelles within eukaryotic cells that has the ability to convert chemical energy from food into a form that cells can use. With mtDNA comparisons between extinct and closely related non extinct species can be done. In order to get the accession codes for the mitochondrial D-loop sequences, we have used the National Center for Biotechnology Information (NCBI) website [17]. The provided mitochondrial D-loop sequences were loaded into MATLAB, and using the functions provided within the Bioinformatics Toolbox they were pairwise aligned before computing the distances and constructing the phylogenetic trees. For the construction of the trees we used distance based methods such as Unweighted Pair Group Method with Arithmetic Mean (UPGMA) and NJ.

RESULTS AND DISCUSSION

The importance of phylogenetics has been systemized and the previous studies concluded that phylogenetic analysis may be considered to be a highly reliable bioinformatics tool. The mitochondrial D-loop sequences have been stored primarily in order to be used for further creation of the phylogenetic trees. In the present work, two types of phylogenetic trees are provided: Neighbor-Joining Phylogenetic Tree and UPGMA Phylogenetic Tree which form the group of distance matrix methods. A common characteristic for both of the methods is calculating the distance between the leaves (or 'taxa') and are better for continuous characters. We have chosen them as the fastest method. First of all, as already explained, we have created a MATLAB structure with information about the sequences. This step involved the use of the accession codes for the mitochondrial D-loop sequences from different types of fishes of Salmonidae Family from the Salmo genus as presented in Table 1. Consequently, the get genbank function from the Bioinformatics Toolbox was employed inside a for-loop to retrieve the D-loop sequences from the NCBI data repository that are further loaded into MATLAB.

Table 1. GenBank accession number of each species

S.No.	Species	GenBank accession number
1	'Salmo_Letnica'	'AY598352';
2	'Salmo_Carpio'	'KJ834850';
3	'Salmo_Cettii'	'KJ834911';
4	'Salmo_Ischchan '	'JF795537';
5	'Salmo_Macrostigma'	'JN807337 ';
6	'Salmo_Marmoratus'	'KJ834801 ';
7	'Salmo_Peristericus '	'HE863710 ';
8	'Salmo_Salar'	'EU643691 ';

Building and Exploring UPGMA Phylogenetic tree

This method follows a clustering procedure. Here, each species represents a cluster on its own and the closest two clusters are joined together. By taking the average, the distance of the joint pair is recalculated. This process is repeated until a single cluster is created connecting all species. This algorithm does not aim to reflect evolutionary descent, but it assigns equal weight on the distance and assumes a randomized molecular clock. We have computed pairwise distances using the 'Jukes-Cantor's formula [18]. Since the sequences are not pre-aligned, seqpdist will pairwise align them before computing the distances. The UPGMA tree is build using seqlinkage function. The resulting UPGMA tree is depicted in Figure 2. Interestingly, it is seen that *Salmo Ischchan* and *Salmo Marmoratus* share a most recent common ancestor and also *Salmo Carpio* and *Salmo Cettii* share a most recent common ancestor. The common ancestor of *Salmo Ischchan* and *Salmo Marmoratus* is most closely related to the fish *Salmo Peristericus*. On the other hand, the most recent common ancestor of *Salmo Cettii* and *Salmo Carpio* is most closely related to the fish *Salmo Marmoratus*.

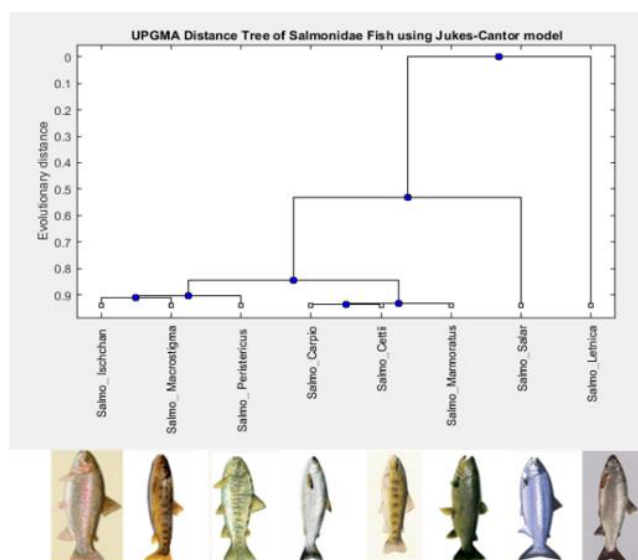


Figure 2. UPGMA Distance Tree of Salmonidae Fish using Jukes-Cantor model

Those common ancestors at some point of the evolutionary distance create one subtree with a common ancestor shared among the six previously mentioned fishes. This new common ancestor is most closely related to the fish *Salmo Salar* and it forms another subtree between 0.6 and 0.5 patristic distance. The patristic distance represents the path length between species calculated from the hierarchical clustering distances. Finally, we can conclude that the fish *Salmo Letnica* shares the most distant ancestor with the other seven fishes on careful investigation of the tree and is confirmed that it genetically differs the most. At this moment, we can find the closest species to the *Salmo Letnica* (entry 1 in Table 1) as follows: The Bioinformatics Toolbox provides some functions for selecting the nodes within given patristic distance. The function `h_all` selects the nodes (leaves and branches) within a patristic distance set by us, and `h_leaves` selects only the leaf nodes. In this case, we used the `h_leaves` function and selected the leaf nodes within 0.6 patristic distance to the *Salmo Letnica* leaf. The corresponding tree can be seen in Figure 3 and it is clearly noticed that this pruned tree differs from the previous UPGMA tree in one thing. The fish *Salmo Salar* is not included in it due to the fact that this fish represents a node from the phylogenetic tree that is below 0.6 patristic distances.

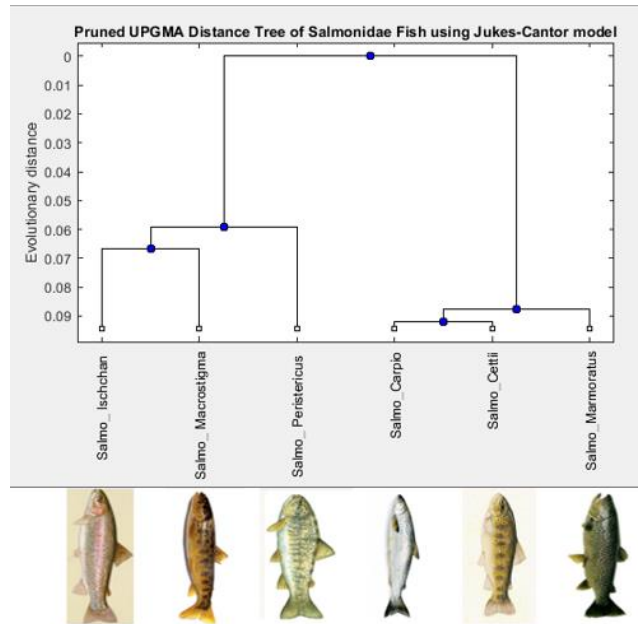


Figure 3. Pruned UPGMA distance Tree of Salmonidae Fish using jukes-Cantor model

Building Neighbor-Joining Phylogenetic tree:

This algorithm does not make the assumption of the molecular clock and adjust for the rate variation among branches. It begins with an unresolved star-like tree, where each pair is evaluated in order to be joined. Afterward, the sum of all branch lengths is calculated of the resultant tree. As a closest neighbor is considered the pair that has the smallest sum, and is thus joined. A new branch is inserted between them and the rest of the tree with the branch length is recalculated. This process goes on until only one terminal is present. When analyzing homologous sequences between species, the alternative tree topologies are important to consider. By using the seqneighjoin function a neighbor-joining tree can be built. For the construction of the neighbor-joining trees, the pairwise distances that were already calculated with seqpdist function are used. This method accomplishes clusters using the minimum evolution method. The Neighbor-Joining Phylogenetic tree is presented in Figure 4.

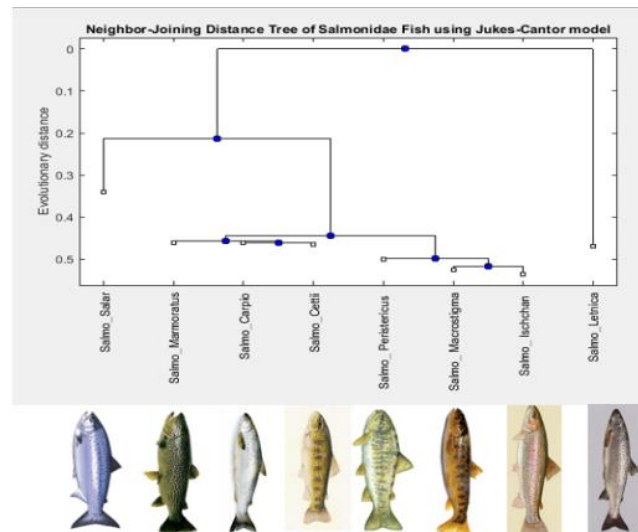


Figure 4. Neighbor-Joining Distance Tree of Salmonidae Fish using Jukes-Cantor model

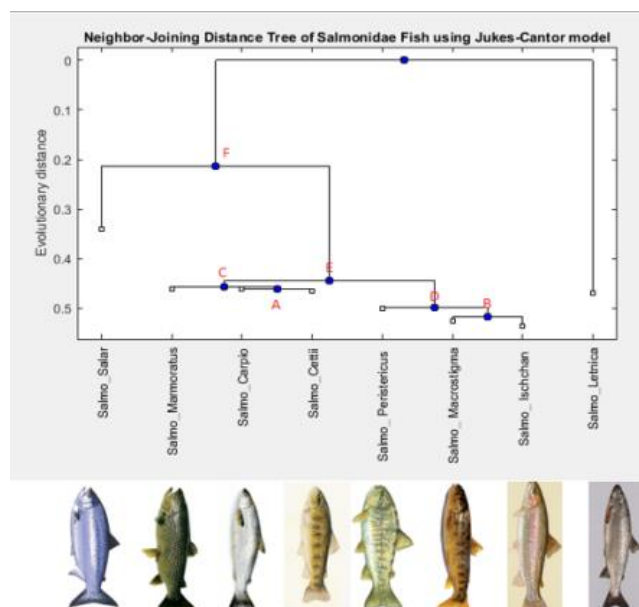


Figure 5. Neighbor-Joining Tree of Salmonidae Fish using Jukes-Cantor model

CONCLUSION

The present work described the study of evolutionary relationships among the fishes from the genera *Salmo* determined based on mitochondrial d-loop sequences. On careful analysis of both UPGMA and Neighbor-Joining trees, it can be concluded that there are two pairs of so-called “sister” nodes. One pair is *Salmo_Carpio* and *Salmo_Cetti* and the other pair is *Salmo_Macrostigma* and *Salmo_Ischchan*. It can be realized that they have a closer common ancestor with each other than with any other type of fishes used in this study (Figure 5). There are two subtrees to mention which includes *Salmo_carpio* and *Salmo_cetti* as subtree A and *Salmo_Macrostigma* and *Salmo_Ischchan* forms subtree B. The two fishes from the subtree A shares the most common history with the fish *Salmo_Marmoratus* and the tree type of fish share a common ancestor at some point forming another subtree C. The two fishes from the subtree B shares the most common history with the fish *Salmo_Peristericus* which in turn shares a common ancestor to leading to subtree D. It can be further seen that the six types of fishes mentioned above create a larger subtree (subtree E) with a common ancestor for all of them. *Salmo_Salar* joins this subtree at the point F representing a common ancestor for the seven of them. Additionally, *Salmo_Letnica* represents the outgroup of trees and it shares some common ancestor at the root of the tree with all of the other seven types of fishes. It can also be concluded that phylogenetic analysis using trees is found to be useful and powerful technique for determination of the relationships between species.

REFERENCES

- [1] Li, Pan, et al. Molecular phylogenetics and biogeography of the mint tribe Elsholtzieae (Nepetoideae, Lamiaceae), with an emphasis on its diversification in East Asia. *Scientific Reports* 2017.7.1 : p. 2057.
- [2] Pagel, M., Inferring evolutionary processes from phylogenies. *Zoologica Scripta*, 1997. 26(4): p. 331-348.

- [3] Page, R.D., Visualizing phylogenetic trees using TreeView. *Current Protocols in Bioinformatics*, 2002: p. 6.2. 1-6.2. 15.
- [4] Feder, M.E. and T. Mitchell-Olds, Evolutionary and ecological functional genomics. *Nature reviews genetics*, 2003. 4(8): p. 649-655.
- [5] Eschmeyer, W.N.P. and R. John, *Encyclopedia of fishes: a comprehensive illustrated guide by international experts*. 1998.
- [6] Matveev, V., H. Nishihara, and N. Okada, Novel SINE families from salmons validate Parahucho (Salmonidae) as a distinct genus and give evidence that SINEs can incorporate LINE-related 3'-tails of other SINEs. *Molecular biology and evolution*, 2007. 24(8): p. 1656-1666.
- [7] Quinn, T.P., *The behavior and ecology of Pacific salmon and trout*. 2011: UBC press.
- [8] Vladykov, V.D., review of salmonid genera and their broad geographical distribution. *Transactions of the Royal Society of Canada*, 1963. 1(N): p. 459-&.
- [9] Georgijev, S., study of salmo genus (pisces, salmonidae) on the balkan peninsula. *Croatian Journal of Fisheries*, 2011. 69(1): p. 11-20.
- [10] Cuculeski, N. and V. Cuculeski, *Comparative study with Bled*. 2016.
- [11] Karaman, S., *Pisces macedoniae*. 1924: Hrvatska Štamparija.
- [12] Von Linné, C., *Systema naturae: per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis*. Vol. 2. 1759: Salvius.
- [13] Rafinesque, C.S., *Indice d'ittologia siciliana: ossia catalogo metodico dei nomi latini, italiani, e siciliani dei pesci, che si rinvencono in Sicilia: Opuscolo del signore CS Rafinesque-Schmaltz*. 1810: del Nobolo.
- [14] Kessler, K., *The Aralo- Caspian Expedition. IV. Fishes of the Aralo- Caspio- Pontine ichthyological region*. St. Petersburg. *Naturgeschichte der Fische Islands*, 1877.
- [15] Duméril, A., Note sur une truite d'Algérie (*Salar macrostigma*, A. Dum.). *Revue et Magazin de Zoologie*, 1858. 10: p. 396-399.
- [16] Cuvier, G., *Revolutions of the Surface of the Globe*. 1829, London: Whittaker, Treacher and Arnot.
- [17] Coordinators, N.R., et al., Database resources of the national center for biotechnology information. *Nucleic acids research*, 2014. 42(Database issue): p. D7.
- [18] Jukes, T.H. and C.R. Cantor, Evolution of protein molecules. *Mammalian protein metabolism*, 1969. 3(21): p. 132.