



Genetic Variation Among Certain Fish Species in Terms of Evolution and Lineages

Hussein A. Saud* & Ilham J. J. Alshami

Department of Fisheries and Marine Resources, College of Agriculture, University of Basrah, Iraq

*Corresponding author: hussein.saud@uobasrah.edu.iq

Received 4th April 2021; Accepted 23th June 2021; Available online 4 November 2021

Abstract: Three tests of phylogenetic including likelihood-joining tree, neighbour-joining tree, and minimum evolution tree have been used based on *sox3* gene. Phylogenetic analysis was used to detect the genetic affinity and common ancestors for selected species that belong to the same or different families. This study showed the most appropriate methods for testing the genetic affinity among species and the methodology of each test according to the requirement of molecular applications. Secondary RNA predicted structure and minimum free energy were also included in this study because of their contribution to the detection of the orthologous gene and variance in RNA folding among species related to the different families. The genetic distance in the studied populations was calculated to know the most appropriate way to find out the genetic similarity among the studied species. The low distance-variance value of each group indicated significant genetic affinity among the species of the same family, this result is more consistent with the test of maximum-likelihood tree indicating the validity of this test to measure the genetic affinity among species that have common ancestors.

Keywords: Fishes, Marine, Fresh Water, Phylogenetic, Evolution, MFE.

Introduction

Historically, the knowledge about fish evolution was not adequately available as studies on tetrapods, because of the difficulties of the experiment in the aquatic environment (Ybazeta & Santini, 2004). Even the major fish lineages that have phylogenetic relationships still ambiguous in consanguinity, the reason can be explained by the large number of taxa belonging to many species of rich fish clades, for example, there are over 16000 species belong to Acanthomorpha, and more than 6000

species belong to Ostariophysii (Froese & Pauly, 2021).

In terms of evolutionary biology, conserved sequences of fishes for many genes are considered more identical among species. Highly conserved region in genes is of interest to researchers, as it is involved in many fields of molecular applications, such as phylogeny, evolutionary biology and other. Knowledge of evolution and variation of races among species are studied through an anatomical, phenotypic

and phylogenetic terms. Phylogenetic analysis is considered recent study that confirm or reveal the confusion that is not confirmed by the comparative phenotypic and anatomical studies. (Hill, 2006).

The knowledge gained from ancestral condition can be obtained from methods that used in comparative biology, but it is still the ancestor-descendant relationships among organisms can be achieved only in the genetics approaches (Nelson, 1969). The systematics of phylogenetic became more accepted now and useful in the study of classification of organisms that never attained by Darwin (Betancur *et al.*, 2017). Information on the phylogenetic relationship is very important and contributes to the explanation of adaptive different characteristics in the organisms life (McDowall, 1997). It assumes that is the best way to paradigm the implying evolution. The new molecular phylogeny can provide an excellent model to support the taxonomic groups that cannot be predicted by former studies (Betancur *et al.*, 2017). In addition, phylogenies are widely used for historical relations among species and for evolutionary divergence for taxon generated from ancestral lines (Kirkpatrick & Slatkin, 1993).

Predicting secondary RNA shape and minimum free energy (MFE) have included in this study to gain clear additional evidence about the variances in RNA among the lineages. Folding RNA is a necessary process to make it more reactive and functional, for example, enable it to mediate interaction, determine the protein binding, catalysis process as well as has more stability, from previous studies the prediction optimal RNA folding has

a lower value (more stable) than incidental folding (Clote *et al.*, 2005).

Here, the goal depends on selecting a conserved region of genes among selected species to evaluate and compare the likelihood of genetics identical with dissimilarities among species, with identify the evolutionary relationship and lineage using different ways of test including phylogeny tree and predicted secondary RNA structures. Since the division of species into the clades differs between the type of phylogenetic tree structure, this study focuses on selecting the most appropriate phylogenetic tree through choosing species belonging to the same genus and others from different taxonomic families. The results were supported via performing the statistical analyses of genetic variation between groups.

Materials & Methods

A total of 11 species in seven families of fishes were selected viz. *Oryzias latipes* (Temminck & Schlegel) (Adrianichthyidae), *Carassius auratus* (Linnaeus), *Cyprinus carpio* L. (Cyprinidae), *Danio rerio* (Hamilton) (Danionidae), *Gadus morhua* L. (Gadidae), *Poecilia mexicana* Steindachner and *Poecilia reticulata* Peters (Poeciliidae), *Kryptolebias marmoratus* Poey (Rivulidae), *Salmo trutta* L. (Salmonidae), *Acanthopagrus latus* Houttuyn (Sparidae). *Culter alburnus* Basilewsky (Xenocyprididae). The layout of higher taxa and scientific names of fishes were verified according to Froese & Pauly (2021). Genome references of these species were obtained from United States National Library of Medicine (USNLM); National Centre for Biotechnology Information (NCBI) (Table1) (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Through comparison among phylogenetic trees,

it can be determining the suitable phylogenetic tree test of fish species through comparison of closely related or dependent species belonging to the same families and others are not the same families. This research was carried out on January 2021.

Table (1): The accession number of *sox3* gene for studied species obtained from the National Center for Biotechnology Information (NCBI).

Fish species	Accession number
<i>Acanthopagrus latus</i>	XM_037076959.1
<i>Carassius auratus</i>	EF174418.1
<i>Culter alburnus</i>	MT127625.1
<i>Cyprinus carpio</i>	LN590686.1
<i>Danio rerio</i>	AB117960.2
<i>Gadus morhua</i>	XM_030369253.1
<i>Kryptolebias marmoratus</i>	KF887913.1
<i>Oryzias latipes</i>	XM_011479911.3
<i>Poecilia mexicana</i>	XM_015001832.1
<i>Poecilia reticulata</i>	LR880654.1
<i>Salmo trutta</i>	LR584428.1

For alignment sequences, assembling of sequences were aligned using Bioedit Sequence Alignment software application (Hall, 1999). All unequal sequences (894-903bp) for studied samples were trimmed to obtain equal sequence for all species before designing phylogenetic trees. Mega X software application has been used for construction of phylogenetic trees

among selected species, the analyses was described by Horiike (2016). It was selected three types of phylogenetic trees to show the variance diversity among species including Maximum Likelihood, neighbor-joining, and minimum evolution, and then they were compared with UPGMA tree structure. Variance mean distances were calculated for statistical issues.

The mean diversity of the entire population and UPGMA were computed using Mega X software to adopt the appropriate phylogenetic tree from the three types of phylogenetic constructing. The P-value of Tajima's test was calculated to find the significant variance among species within the same clade according to the type of tree structure.

Secondary RNA predicted structure and MFE were analysed using RNAfold (M-Fold) Web Server version 2.4.17 (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>), method was described by Mathews (2006).

Results & Discussion

Sox3 gene has been used in this study to analyse genetic variation and phylogenetic relationship among chosen species for many reasons, because it has single exon and size of coding region relatively moderate ranged 894 -903bp among studied species. It was found that amino acids of proteins in the coding region for *sox3* among studied species highly similar and conserved, whereas the dissimilarities represented very small variances (Fig. 1).

The variation of sequence can possibly be considered a good source of phylogenetic information. Open reading frame, in general,

perhaps is not sufficient for providing adequate information in the examination of much related species (Kocher & Stepien, 1997). However, both intron and exon supply data about phylogenies and lineages among species, exon generally contains the conserved region among orthologous genes. Here, it was demonstrated that the near relative similarity among species using secondary RNA predicted structure representing impression about how these species are located under the common ancestor.

In the other hand, these structures can determine the conserved region in genes. It is also possible that the results of the present study can observe the convergent range of the MFE for species belong same family.

RNA prediction shape can also contribute to detect the orthologues genes, where the variances in RNA folding increased as the organism is not belonging to the same family or the taxonomic order (Zuber *et al.*, 2017).

It was very obvious that Phylogenetic tree with Maximum likelihood tree represented the actual data than other methods, where *C. carpio*, *C. auratus* and *C. alburnus* are in the same clade. Similarly, *P. mexicana* and *P. reticulata* and then *K. marmoratus* and *O. latipes* came from common ancestor (Fig. 2). Genetically, it was demonstrated the genetic affinity between *K. marmoratus* and *O. latipes* through previous study using *O. latipes* antisense probe for *in situ* hybridization. In the study of Mourabit *et al.* (2014) medaka (*O. latipes*) *ntl* probe can successfully stain the Mangrove killifish (*K. marmoratus*) notochord by *in situ* hybridization. Both species are belonging to different families, however orthologous genes are highly identical and this convergence was confirmed through the test of

the neighbour-joining tree, where these two species were placed in the same clade (Fig. 3).

On the other hand, this may be due to the existence of common origins between these two species in a historical period somewhat relatively, and this is what the test of minimum evolution tree showed (Fig. 4), this test is typical to prove the shared clade between the related species. Also, it can be determined that these descendants came from one ancestor within a wide range of clade. It was clearly showed that all studied species located at the same clade within wide range except *A. latus* (outgroup). Literally from the above, each phylogenetic tree test gives a certain impression of the extent of the genetic affinity between different species belonging to different families, confirming that they have common origins in a period of time before.

Practically, it can take advantage of a phylogenetic tree in many applications of molecular biology not only in the concept of evolution but in the field of RNA antisense (probe) making or in genomic mapping. RNA structure groups can be contributed how they affect sequence evolution mainly in loops, they could grant clearly knowledge about used RNA based phylogenetic approaches (Moss, 2018).

The genetic distances in the studied population showed significant genetic similarity for species among the cyprinid group (e.g. 0.034 for *C. carpio* and *C. auratus*), and between *P. mexicana* and *P. reticulata* (0.004) (Table 2). By examining the phylogeny tree analysis, it was found that the most consistent shape with the data of the mean distance variance is the maximum-likelihood tree. This analysis placed all species belonging to the

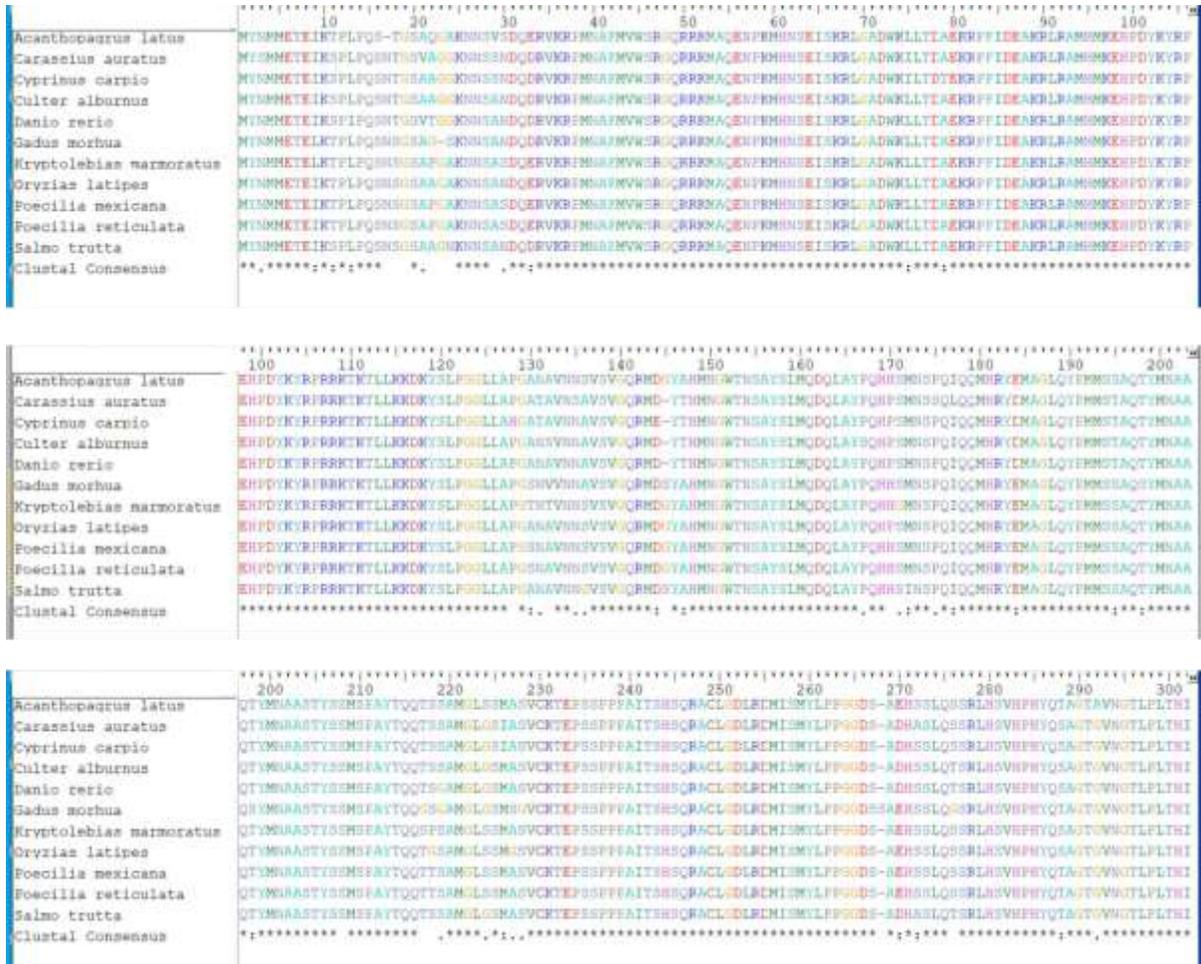


Fig. (1): The amino acid sequence of sox3 protein in studied species showed the conserved region (*) where the identical amino acid in this domain 100%, (, :) some species have different amino acids.

same families within the same clade, as the case of the Cyprinidae and the Poeciliidae families. While the other phylogeny test of the trees had separated *P. mexicana* and *P. reticulata*. This suggests that the maximum-likelihood tree is more suitable for measuring the divergence among species genetically. UPGMA tree (Fig. 5) was the closest in its construction with maximum-likelihood tree where the poeciliid species located within the same clade and this resemblance was not

observed with other methods. Tajima's test that statistically computes the variance between two measures of genetic variation was applied to confirm the statistical confidence of the species falling within the clade. Tajima's test showed that there were no significant differences among the species of *K. marmoratus*, *O. latipes*, *P. mexicana* and *P. reticulata* ($P = -0.40177$), similarly, there was no significant variance ($P = -0.28250$) among species in the

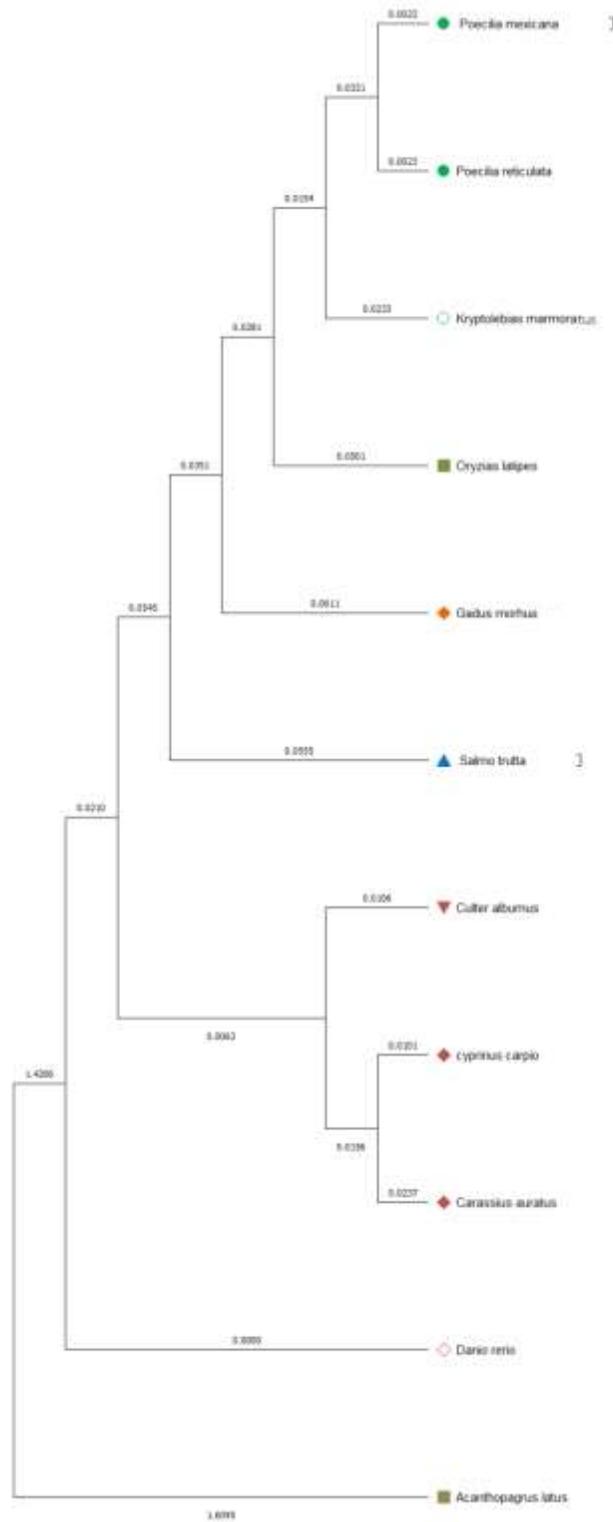


Fig. (2): Phylogenetic tree of studied species analysed by maximum-likelihood test, *Acanthopagrus latus* represents an outgroup among the selected species. Similar symbols and colours represent species that belong to the same family.

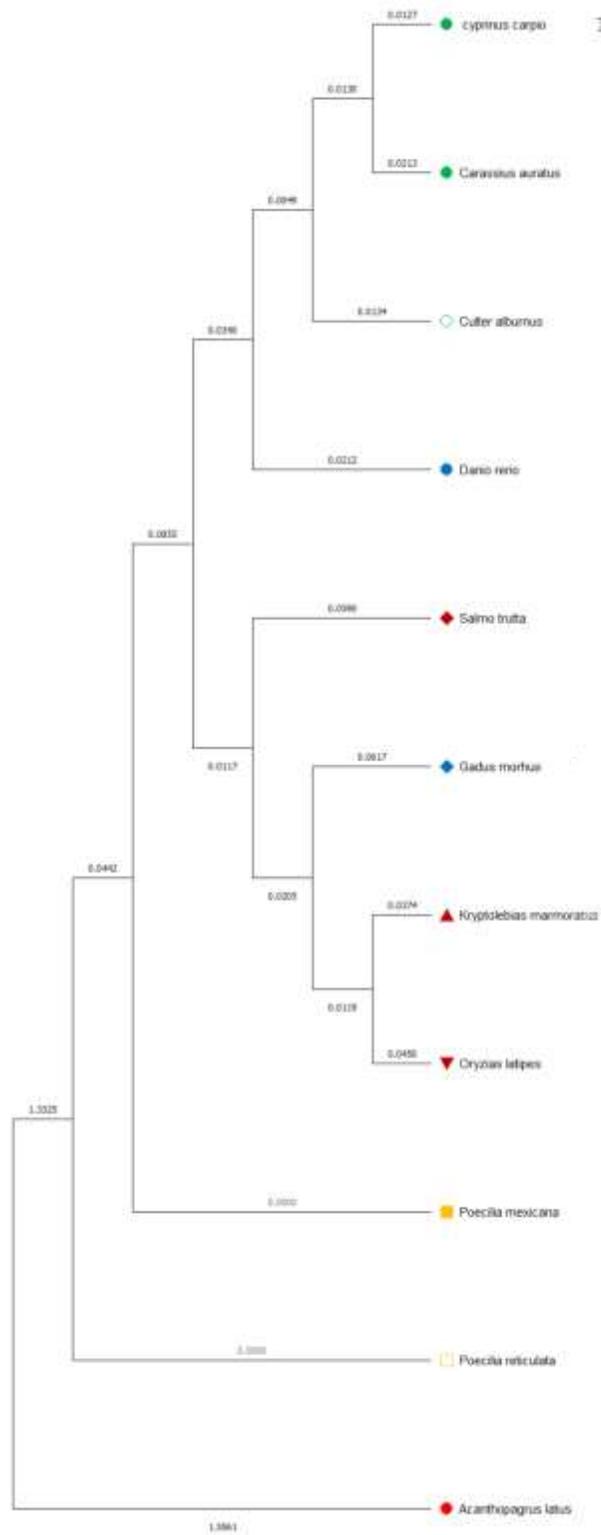


Fig. (3): Phylogenetic tree of studied species analysed by neighbour-joining test, *Acanthopagrus latus* represents an outgroup among the selected species. Similar symbols and colours represent species that belong to the same family.

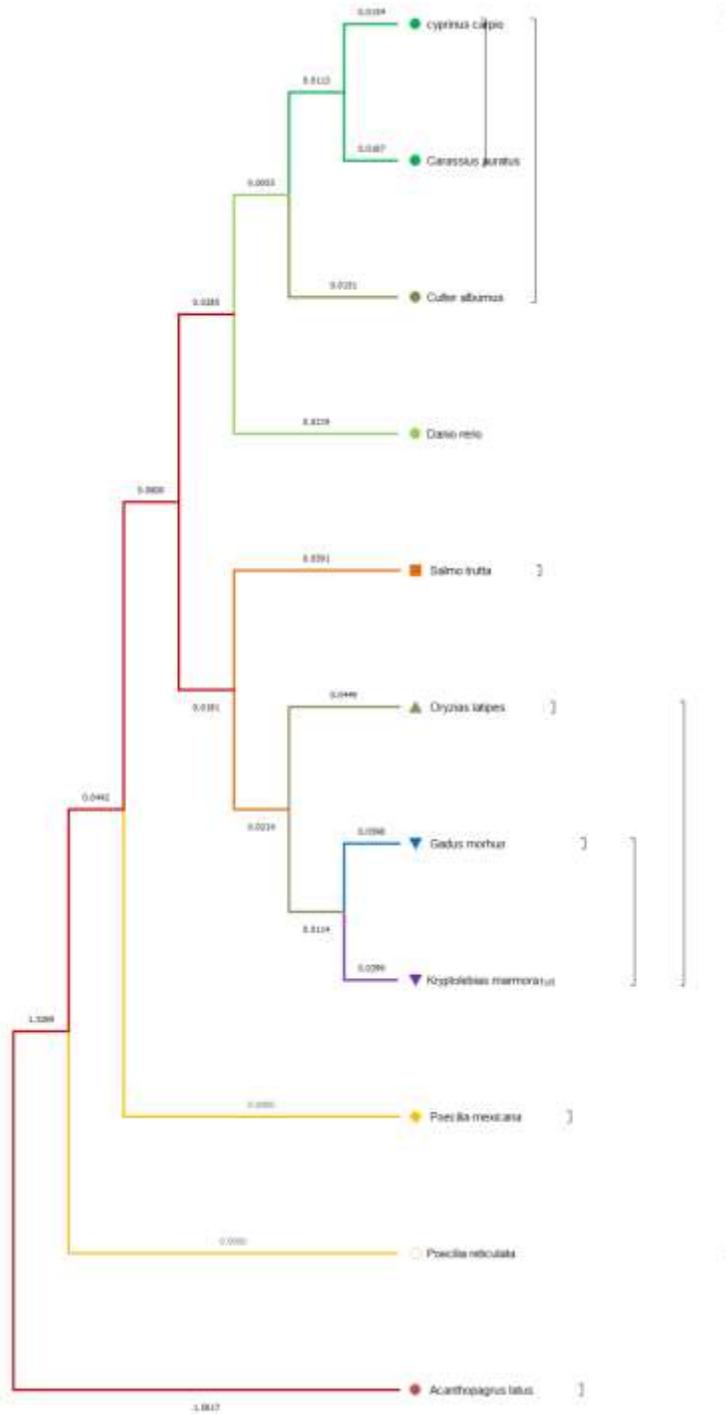


Fig. (4): Phylogenetic tree of studied species analysed by minimum-evolution test, *Acanthopagrus latus* represents an outgroup among the selected species. Similar symbols and colours represent species that belong to the same family.

Table (2): The mean distance value of pairwise in studied groups.

	<i>A. latus</i>	<i>G. morhua</i>	<i>C. auratus</i>	<i>C. alburnus</i>	<i>C. carpio</i>	<i>D. rerio</i>	<i>K. marmoratus</i>	<i>O. latipes</i>	<i>P. mexicana</i>	<i>P. reticulata</i>
<i>A. latus</i>										
<i>G. morhua</i>	0.1147252943									
<i>C. auratus</i>	0.1514285186	0.1589888130								
<i>C. alburnus</i>	0.1383579344	0.1479031606	0.0520958867							
<i>C. carpio</i>	0.1441691469	0.1566044026	0.0343448283	0.0364298852						
<i>D. rerio</i>	0.1423102244	0.1438676041	0.0655931269	0.0388284921	0.0542912181					
<i>K. marmoratus</i>	0.0630615588	0.0997663153	0.1522395051	0.1457139977	0.1517951575	0.1543636479				
<i>O. latipes</i>	0.0939420800	0.1299927599	0.1587911565	0.1550216028	0.1596434056	0.1621903012	0.0834064285			
<i>P. mexicana</i>	0.0653865408	0.1205707089	0.1536537365	0.1486549197	0.1519231911	0.1517656483	0.0564499557	0.0871847087		
<i>P. reticulata</i>	0.0678282872	0.1181325044	0.1564201013	0.1513895425	0.1546781648	0.1545173887	0.0564932490	0.0872534696	0.0044397608	
<i>S. trutta</i>	0.1161489045	0.1386648873	0.1415996710	0.1182593427	0.1331349493	0.1267557892	0.1336513342	0.1373763485	0.1381031008	0.1407824566

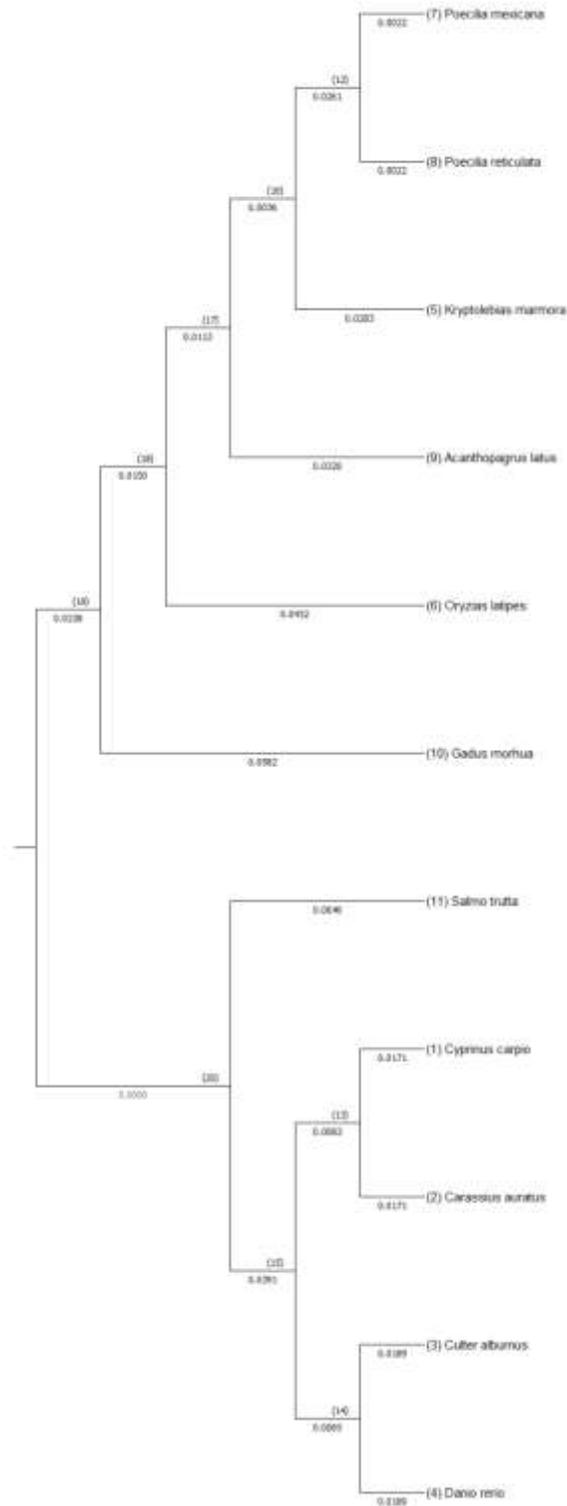


Fig. (5): UPGMA phylogenetic tree, *Acanthopagrus latus* represents an outgroup among the selected species. Similar symbols and colours represent species that belong to the same family.

Table (3): Tajima's test of studied groups.

Group	Tajima's D	Statistical significance
<i>S. trutta</i> , <i>K. marmoratus</i> , <i>O. latipes</i> , <i>G. morhua</i>	-1.06454	(Sig.)***, P < 0.001
<i>K. marmoratus</i> , <i>O. latipes</i> , <i>P. mexicana</i> , <i>P. reticulata</i>	-0.40177	NS, P > 0.10
<i>C. carpio</i> , <i>C. auratus</i> , <i>C. alburnus</i> , <i>D. rerio</i>	-0.28250	NS, P > 0.10

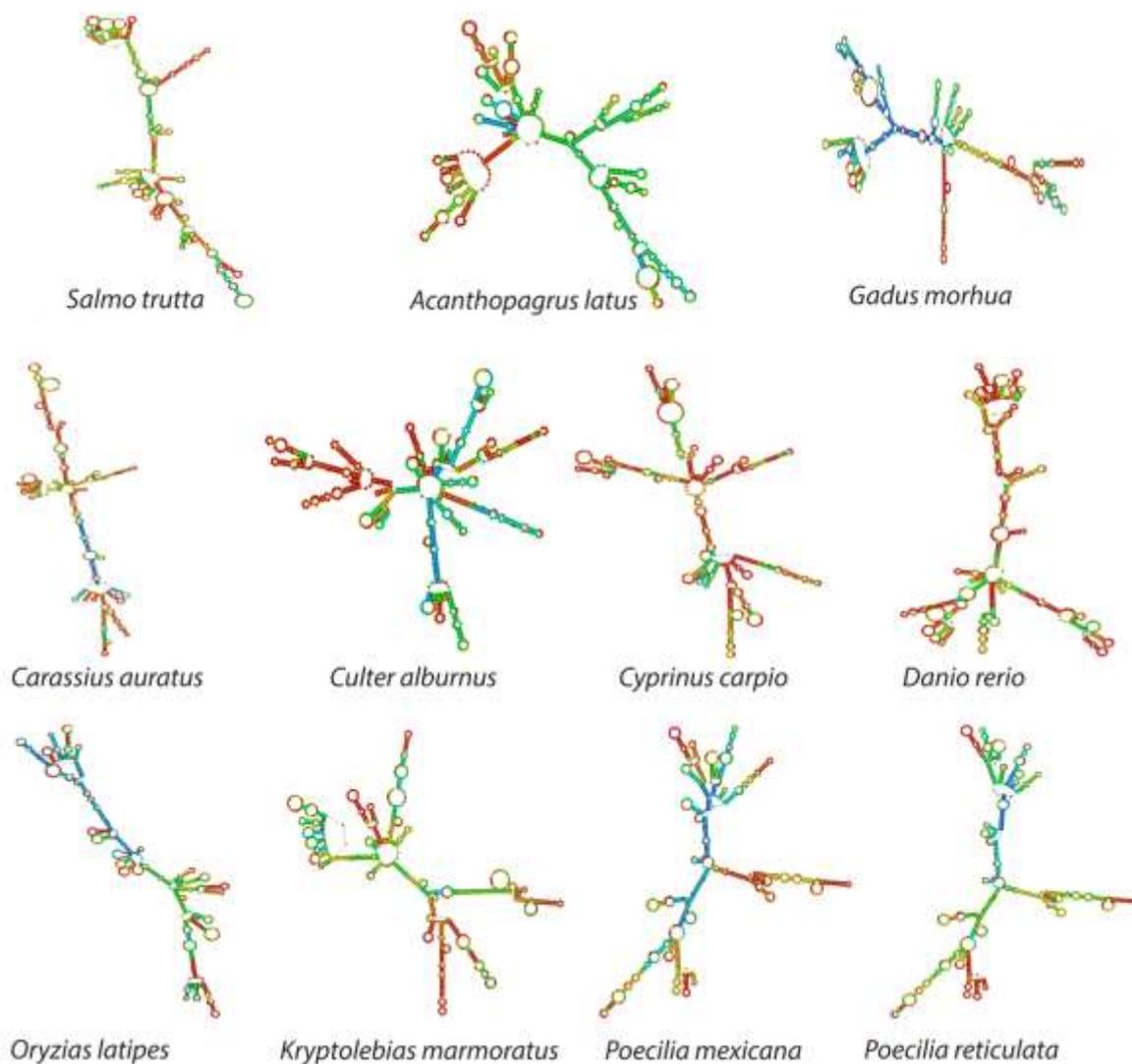


Fig. (6): The predicted structure of secondary RNA for minimum free energy (MFE) in studied species.

cyprinid family (*C. carpio*, *C. auratus*, *C. alburnus*, *D. rerio*), whereas there was highly significant difference ($P = -1.06454$) for the species of *S. trutta*, *K. marmoratus*, *O. latipes*, *G. morhua* that belong different families. This confirms that *P. mexicana* and *P. reticulata* are genetically very close, likewise, the same issue applies to cyprinid species (Table 3). From the mentioned above. It can consider that

maximum-likelihood tree is more reliable to divide the individuals of genetically related species into their clades that are taxonomically existed to the same family or genus.

Secondary RNA predicted structure is also good model providing the genetic affinity through determining the high closeness between species belong the same family. *P. mexicana* and *P. reticulata* have a convergent form in terms of structure and amount of free energy (Fig. 6). Species belong cyprinids have minimum free energy around -286.40 to -304.10 kcal.mol⁻¹ with free energy of the thermodynamic ranged -300.13 to -315.91 kcal.mol⁻¹ respectively, and these species belong to order Cyprinodontiformes including *K. marmoratus*, *P. mexicana* and *P. reticulata* have convergent MFE ranged between -334.90

to 338.20 kcal.mol⁻¹ with free energy of the thermodynamic ranged -348.77 to -350.18 kcal.mol⁻¹ (Table 4).

The current finding is that RNA prediction shape can also contribute to detect the orthologues genes. The variances in RNA folding increased as the organism is not belonging to the same family or the taxonomic order. Also, the analysis of phylogeny and RNA predicted structure can grant an impression about gene classification within the family and knowledge about gene behaviour. Variance in mRNA sequence make changes in MFE, sometimes these changes that caused by mutation can tend to decrease MFE subsequently a decrease in the stability of mRNA and protein expression.

Table (4): Minimum free energy (MFE) in the typical RNA folding for studied species.

Species	Order	Family	Minimum free energy kcal.mol ⁻¹	Free energy of the thermodynamic kcal.mol ⁻¹
<i>Salmo trutta</i>	Salmoniformes	Salmonidae	-290.50	-302.26
<i>Acanthopagrus latus</i>	Perciformes	Sparidae	-319.70	-332.14
<i>Gadus morhua</i>	Gadiformes	Gadidae	-340.60	-353.57
<i>Carassius auratus</i>	Cypriniformes	Cyprinidae	-286.40	-300.13
<i>Culter alburnus</i>	Cypriniformes	Xenocyprididae	-292.70	-306.78
<i>Cyprinus carpio</i>	Cypriniformes	Cyprinidae	-292.60	-306.70
<i>Danio rerio</i>	Cypriniformes	Danionidae	-304.10	-315.91
<i>Oryzias latipes</i>	Beloniformes	Adrianichthyidae	-321.70	-337.11
<i>Kryptolebias marmoratus</i>	Cyprinodontiformes	Rivulidae	-334.90	-348.83
<i>Poecilia mexicana</i>	Cyprinodontiformes	Poeciliidae	-335.30	-348.77
<i>Poecilia reticulata</i>	Cyprinodontiformes	Poeciliidae	-338.20	-350.18

Conclusions

In conclusion, it can be considered that the maximum-likelihood tree one of the most

suitable methods among other phylogenetic trees to distinguish genetically closest species that layout within the same clade.

Conflict of interest

The authors declared that they have no conflict of interest.

Orchid

H. A. Saud: <https://orcid.org/0000-0003-0002-6318>

I. J.J. Alshami: <https://orcid.org/0000-0001-5972-2698>

Acknowledgements

We thank Prof. Dr. Asaad Y. Ayied, College of Agriculture, University of Basrah for his assistance in bioinformatics analysis.

References

- Betancur-R, R., Wiley, E.O., Arratia, G., Acero, A., Bailly, N., Miya, M., Lecointre, G., & Ortí, G. (2017). Phylogenetic classification of bony fishes. *BMC Evolutionary Biology*, 17, 162. <https://doi.org/10.1186/s12862-017-0958-3>
- Clote, P., Ferré, F., Kranakis, E., & Krizanc, D. (2005). Structural RNA has lower folding energy than random RNA of the same dinucleotide frequency. *RNA*, 11, 578-591. <https://doi.org/10.1261/rna.7220505>
- Froese, R. & Pauly, D. (2021). Fish Base World Wide Web electronic publication. www.fishbase.org, version 06/2021.
- Hall, T. A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, 95-98. <https://ci.nii.ac.jp/naid/10030689140/#cit>
- Hill, R. V. (2006). Comparative anatomy and histology of *Xenarthran osteoderms*. *Journal of Morphology*, 267, 1441-1460. <https://doi.org/10.1002/jmor.10490>
- Horiike, T. (2016). An introduction to molecular phylogenetic analysis. *Reviews in Agricultural Science*, 4, 36-45. https://doi.org/10.7831/ras.4.0_36
- Kirkpatrick, M., & Slatkin, M. (1993). Searching for evolutionary patterns in the shape of a phylogenetic tree. *Evolution*, 47, 1171-1181. <https://doi.org/10.2307/2409983>
- Kocher, T. D., & Stepien, C. A. (1997). *Molecular systematics of fishes*. Academic press, Harcourt, Brace & company.314pp.
- Mathews, D. H., (2006). RNA secondary structure analysis using RNA structure. *Current Protocol in Bioinformatics*, 13, 12.6. <https://doi.org/10.1002/0471250953.bi1206s13>
- McDowall, R. M. (1997). The evolution of diadromy in fishes (revisited) and its place in phylogenetic analysis. *Reviews in Fish Biology and Fisheries*, 7, 443-462. <https://doi.org/10.1023/A:1018404331601>
- Moss, W. N. (2018). The ensemble diversity of non-coding RNA structure is lower than random sequence. *Advancing Research Evolution Science*, 3, 100-107. <https://doi.org/10.1016/j.ncrna.2018.04.005>
- Mourabit, S., Moles, M. W., Smith E., van Aerle, R., & Kudoh, T. (2014). Bmp suppression in mangrove killifish embryos causes a split in the body axis. *PLoS ONE* 9, e84786. <https://doi:10.1371/journal.pone.0084786>
- Nelson, G. J. (1969). Origin and diversification of teleostean fishes. *Annals of the New York Academy of Sciences*. 167, 18-30. <https://doi.org/10.1111/j.1749-6632.1969.tb20431.x>
- Ybazeta, G., & Santini, F. (2004). Patterns and processes in the evolution of fishes: An Introduction to the Symposium. *Integrative and Comparative Biology*, 44, 331-332. <https://doi.org/10.1093/icb/44.5.331>
- Zuber, J., Sun, H., Zhang, X., McFadyen, L., & Mathews, D. (2017). A sensitivity analysis of RNA folding nearest neighbor parameters identifies a subset of free energy parameters with the greatest impact on RNA secondary structure prediction. *Nucleic Acids Research*, 10, 6168-6176. <https://doi.org/10.1093/nar/gkx170>

التغيرات الوراثية بين الانواع السمكية من مفهوم التطور والسلالات

حسين عبد سعود والهام جبار جليل الشامي

قسم الاسماك والثروة البحرية، كلية الزراعة، جامعة البصرة، العراق

المستخلص: تم استخدام ثلاثة اختبارات في شجرة التطور الوراثي والتي تمثلت بشجرة الانضمام المحتملة ، وشجرة الانضمام للأفراد المتقاربة ، وشجرة التطور الأدنى بين الافراد. تم استخدام جين SOX3 للكشف عن التقارب الجيني والاسلاف المشتركة بين الانواع المنتخبة من الاسماك والتي تنتمي الى نفس العائلات واخرى من عوائل مختلفة. اوضحت الدراسة الحالية ان انسب الاختبارات لكشف التقارب الجيني بين الانواع هو اختبار شجرة الانضمام المحتمل اما باقي الاختبارات فكانت تشير الى دلائل اخرى على سبيل المثال وجود السلف المشترك والمسافة الزمنية للسلف المشترك بين الانواع. تضمنت الدراسة ايضا رسم الشكل المتوقع للحامض النووي الرايبوزي والحد الأدنى من الطاقة الحرة اللازمة في تشكيل الاواصر لإنتاج الشكل المتوقع للحامض الرايبوزي ومدى علاقة شكله المتوقع بالتقارب بين الانواع من الناحية الجينية وكذلك كمية الطاقة الحرة بين الانواع . بينت قيم التباين للمسافة الجينية المنخفضة لكل مجموعة إلى اقصى تقارب وراثي بين الأنواع التابعة لنفس العائلة ، وهذه يتناسب مع اختبار شجرة الاحتمالية القصوى التي تظهر ملائمة هذا الاختبار لقياس التقارب الجيني بين الأنواع المشتركة الأسلاف.

كلمات مفتاحية: اسماك، بحرية، مياه عذبة، علاقة فايولوجينية، تطور, MFE.