Journal of Agricultural Science and Technology B 8 (2018) 131-143
doi: 10.17265/2161-6264/2018.02.007

Availability of Myoglobin as a Molecular Marker for Phylogenetic Relationships of Fish

Ming-Chih Huang[1], Yoshihiro Ochiai[2] and Shugo Watabe[3]

1. Department of Biological Sciences and Technology, National University of Tainan, 33, Sec.2, Shu-Lin St.,Tainan 700-05, Taiwan R.O.C.

2. Graduate School of Agricultural Science, Tohoku University, Aramaki, Aoba-ku, Sendai, Miyagi 980-0845, Japan

3. School of Marine Biosciences, Kitasato University, Minami-ku, Sagamihara, Kanagawa 252-0373, Japan

Abstract: In order to elucidate the phylogenetic relationship of fish, DNA and deduced amino acid sequences of myoglobin (Mb) were used for the phylogenetic analyses based on different approaches, namely, maximum likelihood (ML), neighbor joining (NJ), unweighted pair group method with arithmetic means (UPGMA) and maximum parsimony (MP) methods in comparison with the conventional molecular markers, mitochondrial cytochrome *b* (cyt-*b*) and cytochrome *c* oxidase subunit I (COI). The phylogenetic trees drawn based on Mb sequences were similar to those by the traditional classification based on the other molecular markers. The primary and secondary structures, as well as the modeled tertiary structures of Mbs were similar to each other, but were clearly distinguishable among those species. Such differences in structure would be associated with adaptation of Mb molecule to the physiological conditions of each species. These results suggest that Mb can be a molecular marker for the phylogenetic relationship of fish.

Key words: Myoglobin, mitochondrial cytochrome *b* (cyt-*b*), cytochrome *c* oxidase subunit I (COI), phylogeny.

1. Introduction

Fish show a large biodiversity in the strategies to adapt to respective inhabiting environments. More than 33,000 species of fish are known worldwide [1]. Among these fish, Osteichthyes is the largest class, followed by Chondrichthyes. Fish are highly diversed, in morphological, genetic, ecological, physiological and behavioral points of view, etc. Due to their high diversity, apparently similar species are likely to be different ones [2]. Species identification and classification of fish are difficult, but worth detailed investigation not only from biological viewpoints but also for effective utilization of marine bioresources.

To define the phylogenetical relationship of organisms, specific proteins or genes, especially some conservative or common ones are usually used as markers. For evolutionary molecular studies, genes contained in mitochondrial DNA are often used as molecular markers, because their nucleotide sequences are moderately diversed. Two genes in mitochondrial DNA encoding cytochrome *b* (cyt-*b*) and cytochrome *c* oxidase I (COI) are often used as molecular markers. Many studies on evolutionary taxonomy have indicated that cyt-*b* can clearly identify and specify organism species [3-8]. Usefulness of COI has been proposed by Hebert et al. [9].

Myoglobin (Mb), one of heme proteins, is found exclusively in muscle and responsible not only for oxygen storage in muscle but also for its pigmentation [10]. Vertebrates, including fish possess Mb with a few exceptions, such as some Antarctic species [11]. In fish, Mb can be found in striated muscles, such as skeletal muscle (especially dark muscle) and cardiac muscle [12]. Most fish Mbs consist of 147 amino acid residues with some exceptions, such as pufferfish *Tetraodon nigroviridis* (National Center for Biotechnical Information (NCBI) No. CAF31356) and

Corresponding author: Ming-Chih Huang, associate professor, research fields: aquatic biochemistry and genetics.

Takifugu rubripes (NCBI No. XP_003976095) Mbs with 146 residues and whale shark *Rhincodon typus* Mb (NCBI No. XP_020371173) with 149 residues.

Many attempts have been made to employ molecular markers, such as mitochondrial cyt-*b* and COI genes for phylogenetic analyses. Not all of them, however, have been successful for this purpose. The authors' research group has reported previously the availability of muscle tropomyosin for this purpose [13]. In this study, availability of Mb was evaluated as a new possible molecular marker for the purpose of precise classification and phylogenetic analysis of fish.

2. Materials and Methods

2.1 Alignment of Amino Acid Sequences

The sequence data of fish muscle Mb, cyt-*b* and COI were collected from the database in NCBI. Alignment of the deduced amino acid sequences was performed with European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI)—ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

Accession numbers of teleost Mbs examined in this study are as follows: channel catfish (*Ictalurus punctatus*, NP_001187526), greater amberjack (*Seriola dumerili*, BAG84239), milkfish (*Chanos chanos*, ABI97485), yellowtail amberjack (*S. lalandi*, BAH90800), Japanese amberjack (*S. quinqueradiata*, BAG84238), black rockcod (*Notothenia coriiceps*, AAG16646), zebrafish (*Danio rerio*, AAH56727), common carp (*Cyprinus carpio*, XP_018966946), bigeye tuna (*Thunnus obesus*, BAC82200), bluefin tuna (*T. thynnus thynnus*, AAG02105), yellofin tuna (*T. albacares*, AAG02112), longfin tuna (*T. alalunga*, AAG02106), Atlantic blue marlin (*Makaira nigricans*, AAG02107), Atlantic salmon (*Salmo salar*, ACM09229), Nile tilapia (*Oreochromis niloticus*, BAH19314), Mozambique tilapia (*O. mossambicus*, BAH22119), blue tilapia (*O. aureus*, BAH22118), rainbow trout (*Oncorhynchus mykiss*, BAI45225), walking catfish (*Clarias batrachus*, AGG38020), medaka (*Oryzias latipes*, XP_004065750), whale shark (*Rhincodon typus*, XP_020371173). The sperm whale (*Physeter catodon*, BAF03579), chicken (*Gallus gallus*, NP_001161224) and house mouse (*Mus musculus*, NP_001157520) sequences were used as outgroup.

Accession numbers of teleost cyt-*b*s are as follows: zebrafish (*D. rerio*, ALK26838), bigeye tuna (*T. obesus*, ADA69860), bluefin tuna (*T. thynnus thynnus*, BAC78542), yellowfin tuna (*T. albacares*, ADF43962), Atlantic blue marlin (*M. nigricans*, ABN47259), Atlantic salmon (*S. salar*, ACB30582), Nile tilapia (*O. niloticus*, ADA58756), Mozambique tilapia (*O. mossambicus*, ABB85081), blue tilapia (*O. aureus*, YP_003406702), rainbow trout (*O. mykiss*, AAK54370), walking catfish (*C. batrachus*, AIQ81032), medaka (*O. latipes*, BAV60900), whale shark (*R. typus*, YP_009002145), sperm whale (*P. catodon*, AGB56617), chicken (*G. gallus*, AAF73235) and house mouse (*M. musculus musculus*, YP_001686710).

Accession numbers of teleost COIs are as follows: channel catfish (*I. punctatus*, NP_612127), greater amberjack (*S. dumerili*, BAL52271), milkfish (*C. chanos*, NP_818802), yellowtail amberjack (*S. lalandi*, BAL52245), Japanese amberjack (*S. quinqueradiata*, BAL52232), black rockcod (*N. coriiceps*, AEH05440), zebrafish (*D. rerio*, NP_059333), common carp (*C. carpio,* NP_007084), bigeye tuna (*T. obesus*, YP_003587610), bluefin tuna (*T. thynnus thynnus*, AHH80744), yellowfin tuna (*T. albacares*, AIY51653), longfin tuna (*T. alalunga*, NP_955707), Atlantic blue marlin (*M. nigricans*, ACS34708), Atlantic salmon (*S. salar*, AAF61380), Nile tilapia (*O. niloticus*, ADA58747), Mozambique tilapia (*O. mossambicus*, AAT92269), blue tilapia (*O. aureus*, ADB43179), sperm whale (*P. catodon*, AGB56633), house mouse (*M. musculus*, NP_904330), rainbow trout (*O. mykiss*, NP_008292), walking catfish (*C. batrachus*, YP_009024095), medaka (*O. latipes*,

BAH84895), chicken (*G. gallus*, ADB06650), whale shark (*R. typus*, YP_009002135).

2.2 Analysis of Highly Variable Regions of Amino Acid Sequences

The amino acid sequences of highly variable regions in the above Mbs were analyzed by using Protein Variability Server (PVS) (http://imed.med.ucm.es/PVS/) [14]. The total of 20 fish species Mb with 146 and 147 amino acid residues (including rainbow trout *O. mykiss* and walking catfish *C. batrachus*) was analyzed.

2.3 Secondary Structure Prediction

Secondary structure prediction was performed using a software ExPASy proteomics tools/secondary structure prediction/GOR IV (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html) [15], a public server maintained by Swiss Institute of Bioinformatics. This analysis was carried out for only five common types of fish, such as medaka (*O. latipes*, XP_004065750), bluefin tuna (*T. thynnus thynnus*, AAG02105), zebrafish (*D. rerio*, AAH56727), Atlantic salmon (*S. salar*, ACM09229) and Mozambique tilapia (*O. mossambicus*, BAH22119).

2.4 Tertiary Structure Prediction

The prediction of tertiary structures was carried out by using SWISS-MODEL Automatic Modeling Mode (http://swissmodel.expasy.org/workspace/index.php?func=modelling_simple1). Simulation results were obtained by using Swiss-PdbViewer (http://spdbv.vital-it.ch/) [16]. For this analysis, the same species examined for the secondary structure prediction were investigated.

2.5 Phylogenetic Trees

The amino acid sequences of Mb (24 species), cyt-*b* (16 species) and COI (24 species) were subjected to phylogenetic tree construction, which was carried out with the maximum likelihood (ML) [17], neighbor joining (NJ) [18], unweighted pair group method with arithmetic means (UPGMA) [19] and maximum parsimony (MP) [20-22]. The four approaches were performed based on ClustalW, which generated paired alignments of all the sequences. Bootstrap majority consensus values on 1,000 replicates were calculated [23] and were indicated in percent at each branch node. All these programs were parts of Molecular Evolutionary Genetics Analysis ver. 7 (MEGA 7) [24]. Evolutionary distances were computed using the Jones-Taylor-Thornton (JTT) method for ML analysis [17], *p*-distance model for NJ and UPGMA [25], Subtree-Pruning-Regrating (SPR) for MP [25], and were expressed as the unit (the number of amino acid substitutions per site).

3. Results

3.1 Amino Acid Sequence Alignment and Identification of Highly Variable Regions

The alignments of amino acid sequences of Mbs from fishes, mammals and chicken are shown in Fig. 1. Total amino acid numbers of mammalian and chicken Mbs are 154, longer than those of most fish Mbs (147 residues). The first four residues, the 53rd, 122nd and 123rd residues are missing in fish Mbs. The eight helix segments A, B, C, D, E, F, G and H in the order from the N termini are indicated in the figure. Segment D is missing in fish Mbs. Rainbow trout and walking catfish Mbs were found to lack the 86th and 85th residues, respectively.

The analysis of highly variable regions for 20 fish Mbs by PVS showed similar pattern (data not shown). The 21st, 123rd and 131st residues showed high variations and the values of Shanon variability were 2.484, 2.364 and 2.421, respectively, followed by the 19th, 112th, 31st and 139th residues. The values were higher than 2 for all.

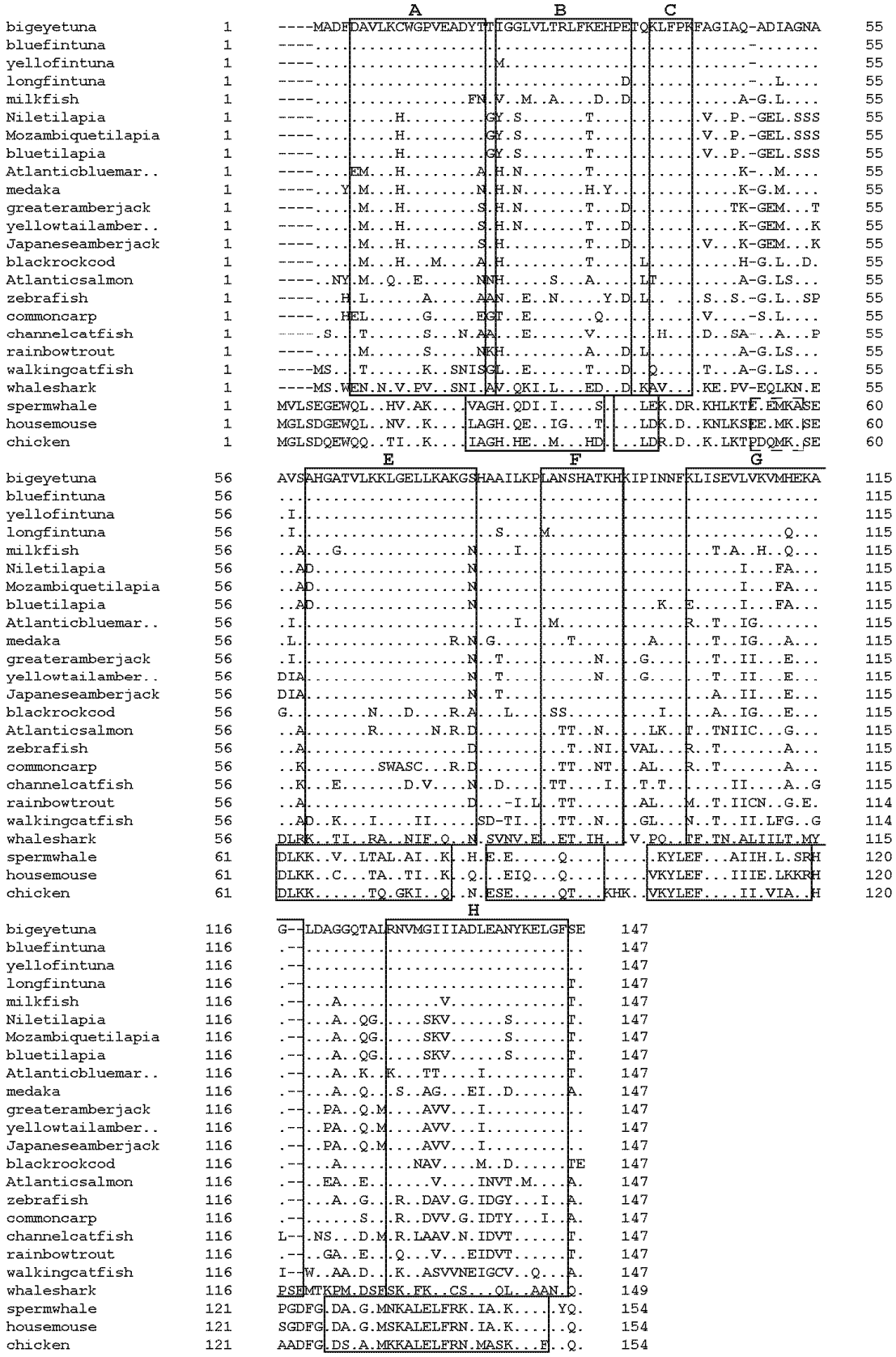

A
B
C
E
F
G
H
bigeyetuna
bluefintuna
yellofintuna
longfintuna
milkfish
Niletilapia
Mozambiquetilapia
bluetilapia
Atlanticbluemar..
medaka
greateramberjack
yellowtailamber..
Japaneseamberjack
blackrockcod
Atlanticsalmon
zebrafish
commoncarp
channelcatfish
rainbowtrout
walkingcatfish
whaleshark
spermwhale
housemouse
chicken

Fig. 1 Alignments of amino acid sequences of chicken, sperm whale, house mouse Mbs with those of other fish species.
Amino acid gap is indicated by a dashed line. The upper boxes indicated the α-helical segments A through H. For sperm whale, house mouse and chicken Mb, the α-helical segments A-H are lower boxes. Segment D is missing in fish Mbs.
The sequence data from the DDBJ/EMBL/GenBank databases are under the following accession numbers: channel catfish (*I. punctatus*, NP_001187526), greater amberjack (*S. dumerili*, BAG84239), milkfish (*C. chanos*, ABI97485), yellowtail amberjack (*S. lalandi*, BAH90800), Japanese amberjack (*S. quinqueradiata*, BAG84238), black rockcod (*N. coriiceps*, AAG16646), zebrafish (*D. rerio*, AAH56727), common carp (*C. carpio*, XP_018966946), bigeye tuna (*T. obesus*, BAC82200), bluefin tuna (*T. thynnus thynnus*, AAG02105), yellowfin tuna (*T. albacares*, AAG02112), longfin tuna (*T. alalunga*, AAG02106), Atlantic blue marlin (*M. nigricans*, AAG02107), Atlantic salmon (*S. salar*, ACM09229), Nile tilapia (*O. niloticus*, BAH19314), Mozambique tilapia (*O. mossambicus*, BAH22119), blue tilapia (*O. aureus*, BAH22118), rainbow trout (*O. mykiss*, BAI45225), walking catfish (*C. batrachus*, AGG38020), medaka (*O. latipes*, XP_004065750), whale shark (*R. typus*, XP_020371173). The sperm whale (*P. catodon*, BAF03579), chicken (*G. gallus*, NP_001161224), house mouse (*M. musculus*, NP_001157520) sequences were used as outgroup.

3.2 Secondary Structure Prediction

The highly variable residues were found in the random coil (the 21st one), α-helical regions (the 123rd and 131st ones), but were contained in the extended strands for salmon Mbs (the 131st one). The lower extent of variations was found in the 19th, 112th, 130th and 139th residues. The 21st, 123rd and 130th residues were the same among Mbs from the five fish species, but the 19th and 112th residues showed slight variations. The summary of secondary structures in the highly variable regions of Mbs is shown in Table 1.

3.3 Tertiary Structure Prediction

The tertiary structure prediction of fish Mbs performed by SWISS-MODEL and Swiss-PdbViewer revealed very similar tertiary structures among different fish species. Although slight species specificity in the secondary structure was recognized as described above, their tertiary structures well resembled each other. The two representative structures on tuna and salmon Mbs are shown in Fig. 2. Although the 21st, 123rd and 131st residues were highly variable among different fish species, these residues do not seem to affect the tertiary structures so much.

3.4 Phylogenetic Trees

Phylogenetic trees of Mb, cyt-*b* and COI were drawn by ML method, NJ method, MP method and UPGMA based on the amino acid sequences by the aid of MEGA 7. The results are shown in Figs. 3-5, respectively. Fig. 3 shows the analytical results obtained for Mb. The optimal tree with the sum of branch length is as follows: ML with the highest log likelihood (-3,452.08), NJ (3.293), UPGMA (3.255) and MP with the most parsimonious trees (645). Fig. 4 shows the results obtained for cyt-*b*, with the branch length as follows: ML with the highest log likelihood (-5,383.28), NJ (2.030), UPGMA (2.319) and MP with the most parsimonious trees (1,019). The results for COI are shown in Fig. 5, with the branch length as follows: ML with the highest log likelihood (-5,323.45), NJ (1.356), UPGMA (1.344) and MP with the most parsimonious trees (768). The phylogenetic trees of Mb, cyt-*b* and COI were used for bootstrap value check. The amino acid variations were calculated based on each different model, JTT method for ML, *p*-distance for NJ and UPGMA, SPR for MP. In particular, the alignment showed that tuna and tilapia Mbs are classified into the same clade. By the traditional taxonomy, both species belong to the order Perciformes. In the phylogenetic tree based on Mb sequences, these species were closer to each other compared to the trees based on cyt-*b* and COI.

4. Discussion

Amino acid sequences of conservative proteins (like cyt-*b* and COI) are usually used to infer distant phylogenetic relationships, such as early divergences near the root of the universal tree of life [26-28]. For

distant relationships of phylogeny, the use of nucleotide sequences can be problematic, because it is likely that the alignment is difficult, and the base frequencies may vary among species, and further the saturation of substitutions may have diluted phylogenetic information [29]. In some cases, use of amino acid sequences seems to be advantageous [30]. The nucleotide and amino acid sequences have been successfully applied to phylogenetic analyses, but in this study, amino acid base analyses are considered to be better than nucleotide base ones.

4.1 Properties of the Markers

Genes encoded by mitochondrial DNA are valuable for understanding the evolutionary relationships among individuals, populations and species. The cyt-*b* and COI genes are usually chosen as phylogenetic probes, because it is much easier to align protein-coding sequences that have evolved over the period spanning the origin of the species than to align either mitochondrial rDNA or non-coding sequences from the distant relatives.

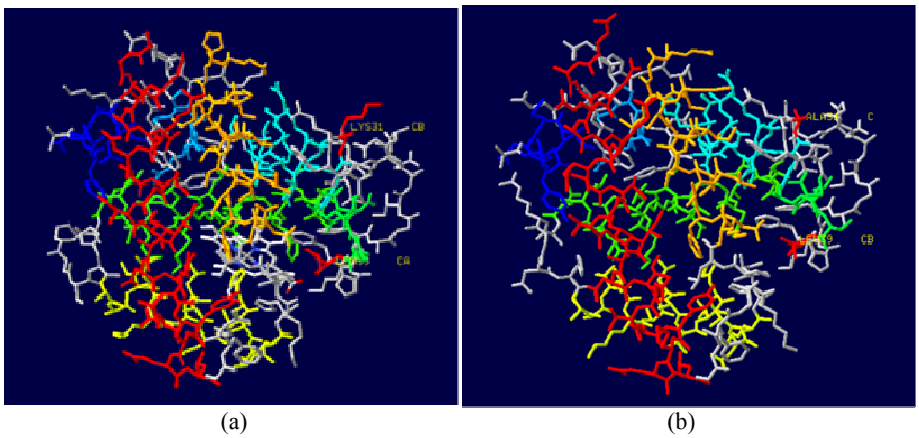
Cyt-*b* is the main subunit of transmembrane cytochrome complexes and functions as a member of electron transport chain. It is a kind of integral membrane protein that probably has eight transmembrane domains. Cyt-*b* is considered to be the most useful marker for exploring the relationships within families and genera. Fish cyt-*b*s generally consist of 380 amino acid residues.

On the other hand, COI is the main subunit of the cytochrome *c* oxidase complex. Cytochrome *c* oxidase is a key enzyme in aerobic metabolism and the member of the respiratory chain that catalyzes the reduction of oxygen to water. COI is recognized as an extremely useful DNA barcode, capable of accurate species identification in a very broad range of eukaryotic organisms [31-33].

Table 1 Highly variable area of amino acid position and secondary structure.

Species	Amino acid position (secondary structure/amino acid residue)						
	19	21	112	123	130	131	139
Bluefin tuna	e/T	c/I	h/H	h/T	h/G	h/I	h/N
Mozambique tilapia	c/T	c/Y	c/A	h/Q	h/S	h/K	h/S
Medaka	c/N	c/H	h/A	h/Q	h/A	h/G	h/D
Zebrafish	h/A	c/N	h/A	h/G	h/D	h/A	c/Y
Atlantic salmon	c/N	c/H	c/G	h/E	h/G	e/V	e/T

c: random coil; e: extended strand; h: α- helix.



(a) (b)

Fig. 2 Tertiary structure prediction of bluefin tuna (a) and Atlantic salmon (b) Mbs by ExPASy proteomics tool/Swiss-PdbViewer.

Different colors indicate different α-helical regions of Mbs.

Highly variable amino acid residues (positions 21, 123 and 131) are indicated in red with the letters in yellow.

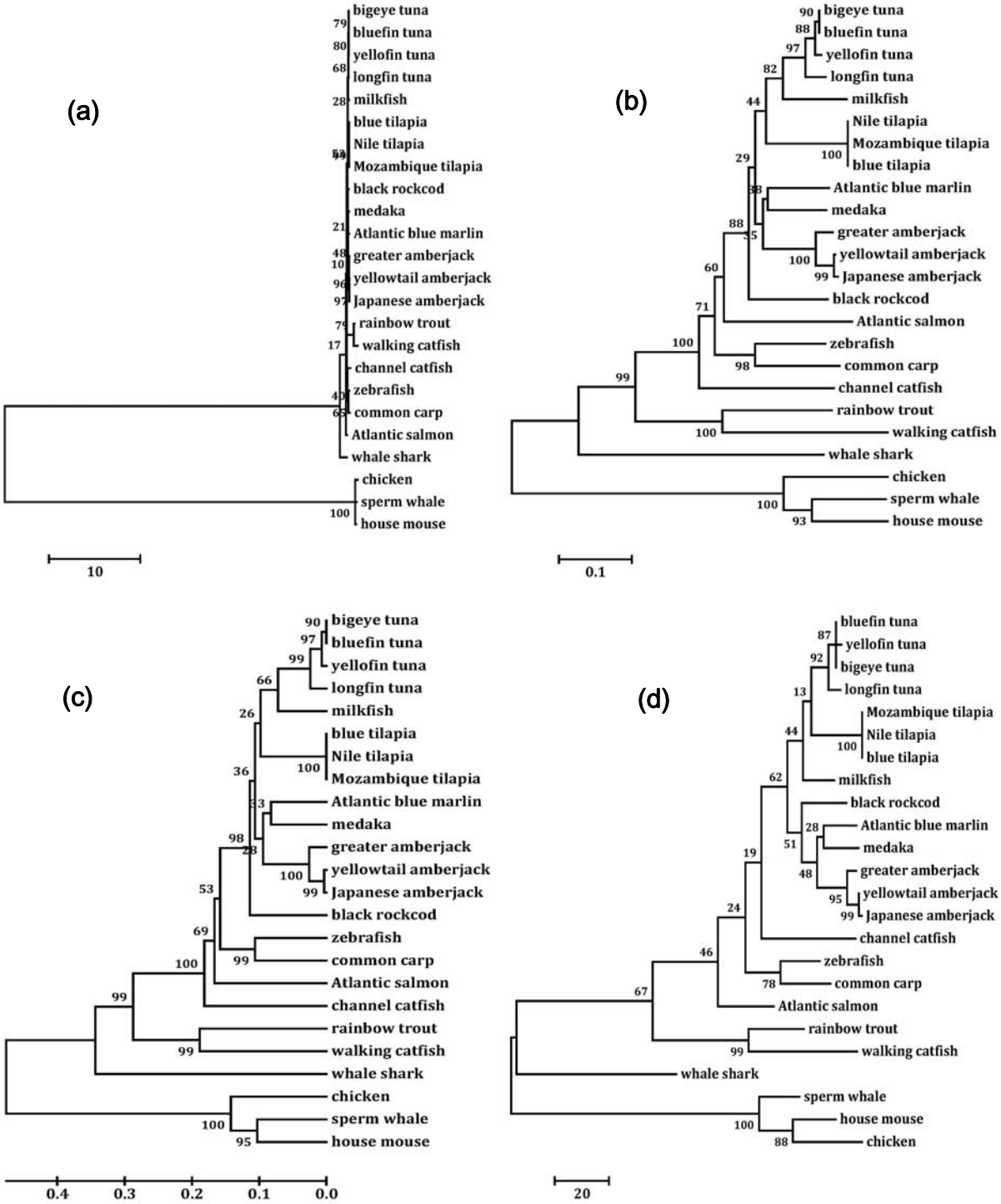


Fig. 3 Phylogenetic trees based on the amino acid sequences of Mbs from sperm whale, chicken, house mouse and various fish species (total of 24 species).

Deduced amino acid sequences were aligned using ClustalW, and the trees were constructed by the ML (a), NJ (b), UPGMA (c) and MP methods (d). The house mouse, chicken and sperm whale sequences were used as outgroup. The percentages of the replicated tree in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown at the nodes. The evolutionary distances were computed using the Poisson correction method. Evolutionary analyses were conducted in MEGA 7.

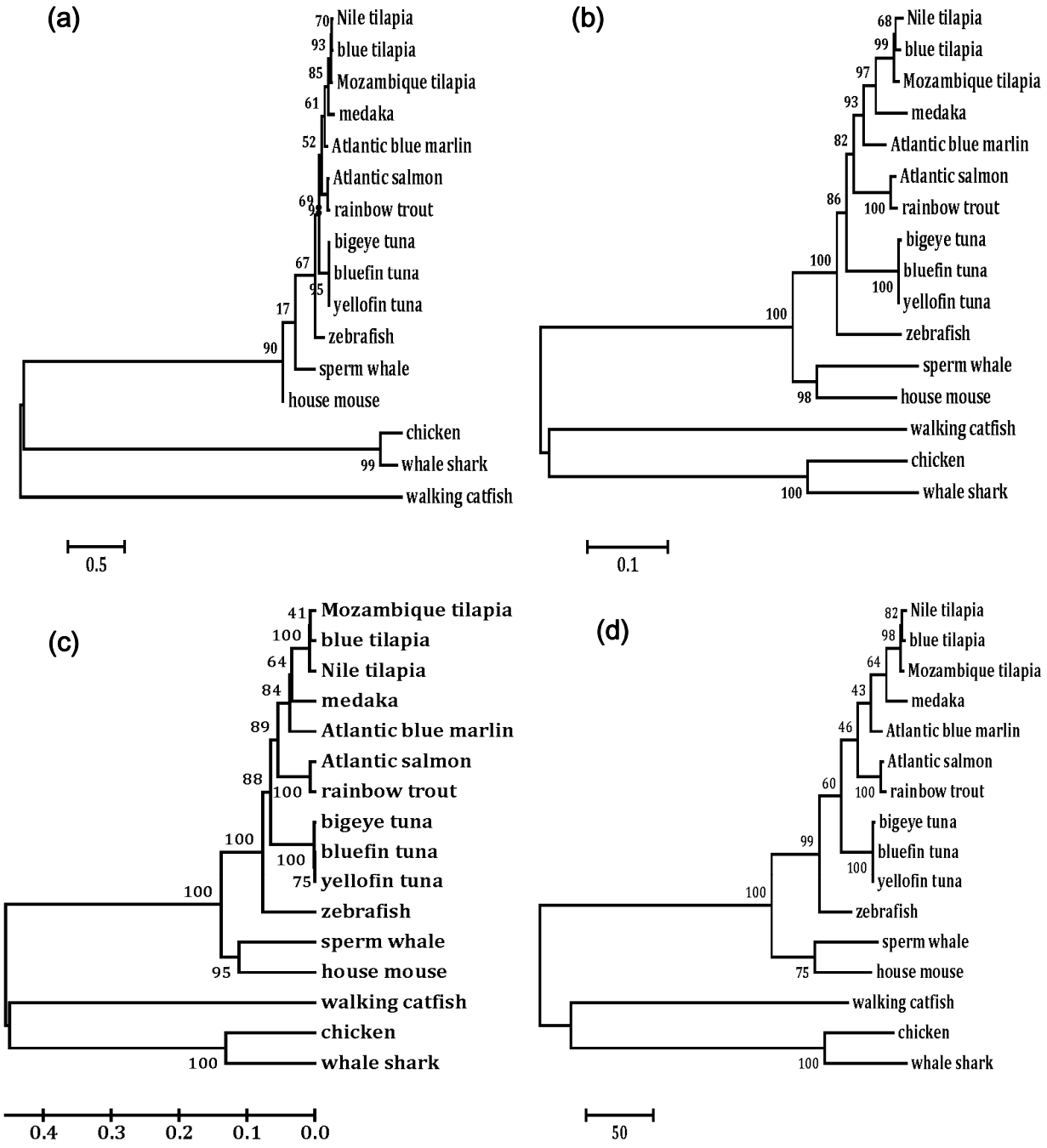


Fig. 4 Phylogenetic trees based on the amino acid sequences of cyt-*b*s from sperm whale, chicken, house mouse and various fish species (total of 16 species).

Deduced amino acid sequences were aligned using ClustalW, and the trees were constructed by the ML (a), NJ (b), UPGMA (c) and MP methods (d). The house mouse, chicken and sperm whale sequences were used as outgroup. The percentages of the replicated tree in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown at the nodes. The evolutionary distances were computed using the Poisson correction method. Evolutionary analyses were conducted with MEGA 7.

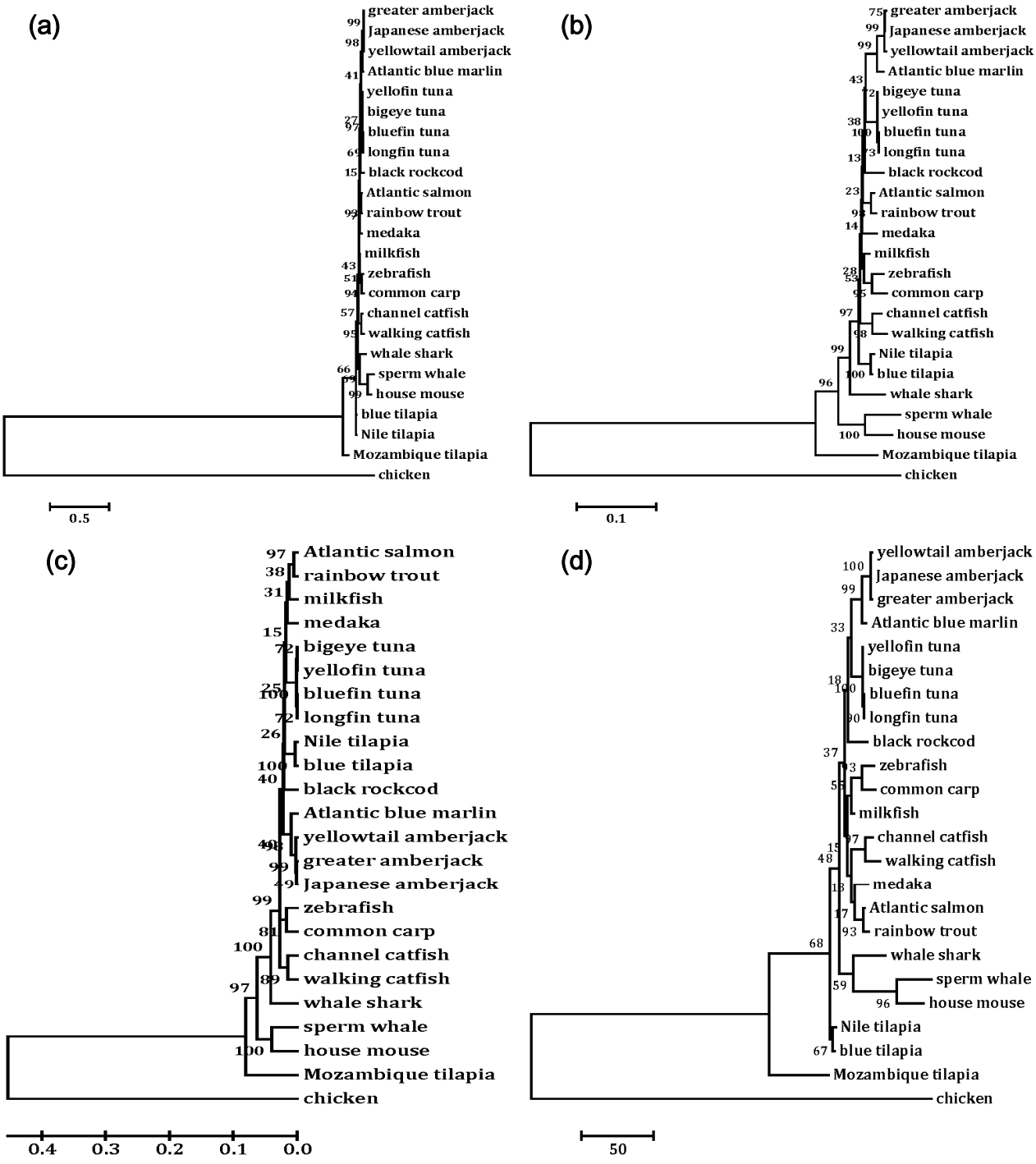


Fig. 5 Phylogenetic trees based on the amino acid sequences of COIs from sperm whale, chicken, house mouse and various fish species (total of 24 species).

Deduced amino acid sequences were aligned using ClustalW, and the trees were constructed by the ML (a), NJ (b), UPMGA (c) and MP methods (d). The house mouse, chicken and sperm whale sequences were used as outgroup. The percentages of the replicated tree in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown at the nodes. The evolutionary distances were computed using the Poisson correction method. Evolutionary analyses were conducted in MEGA 7.

The COI gene has been extensively studied in relation with the phylogeny of fish, since the sequences are highly conserved among species. COIs from various vertebrate taxa have been sequenced [32, 34-36]. Such well characterized COI genes have been proved to be a robust evolutionary marker also for the analysis of intraspecific and interspecific relationships in many marine fish species [32, 37]. The number of COI amino acid residues is generally 516, but those of tilapia and medaka examined in this study consist of 533 and 518 residues.

Mb is a less conserved protein. Unlike cyt-*b* and COI, Mb is not a mitochondrial member but exists in the sarcoplasm of muscle cells. Most fish Mbs consist of 146 amino acids residues. The numbers of amino acids in Mbs are much smaller than those of cyt-*b* (380) and COI (516), making it easier to compare the sequences.

In order to explore the evolutionary importance of biomarkers, phylogenetic tree based on the conventional taxonomy is an essential reference. The phylogenetic tree made by Bonde [38] was considered in this study.

4.2 Evolution of Mbs

Mb is a relatively compact globular protein, whose backbone structure consists of eight α-helical segments designated A through H from the N terminus. The heme resides in a hydrophobic "heme pocket" groove, and binds to the imidazole group of proximal histidine directly and that of distal histidine through an oxygen coordinate binding [39]. As the oxygen affinity level of Mb is higher than that of hemoglobin, Mb can store oxygen temporarily in muscle and transfer it to an electron transfer system. Out of 154 amino acids, more than 120 residues are present in the helical regions and the 32 amino acids are distributed all over the non-helical regions.

Fish Mbs are very unstable compared with those of higher vertebrates, and thus autoxidize and aggregate easily [40-43]. The primary and crystal structures of yellowfin tuna *T. albacares* Mb have been solved [44, 45]. The tertiary structures of fish Mbs are quite similar to those of mammalian counterparts [45, 46], although fish Mbs lack a D-helical [45].

Fish have changed Mb affinity for oxygen through adaption and evolution. Since Mb facilitates the oxygen transfer from the blood to tissues, environmental changes, like high water temperature or hypoxia tolerance could decrease oxygen affinity of Mb [47, 48]. Mbs also show cold-compensated metabolic demands at low temperatures [49]. It means that Mb plays an important role for the organisms to adapt environmental changes.

4.3 Structure Variance of Mbs

Apomyoglobin (apoMb) is a protein portion of Mb free from heme. It has eight α-helical segments, and the structure of each segment has been discussed [50]. The stability of segments A, G and H in a molten globule intermediate state was found to be important for unfolding of Mb [51]. These segments are readily structured just after translation and form a hydrophobic core, while helices B and E are involved in the core formation of apoMb [52]. Helix F is disordered by removal of heme [53]. The absence of the 53rd residue results in the disappearance of segment D [54]. The lengths of α-helical segments are comparable between whale and tuna Mbs, but tuna Mbs have shorter segments F and H [45]. Such differences would also affect the stability of Mbs.

4.4 Usefulness of Mb as a Molecular Marker

Based on the results obtained in the present study, Mb is considered to be a good marker for phylogenetic analysis of fish as shown in Fig. 3. The model of NJ method seems to be closer to the conventional taxonomy (Fig. 3b), where tilapias (Mozambique tilapia, Nile tilapia and blue tilapia) belonging to the family Cichlidae, the order Perciformes, are closer to tunas (family Scombridae).

In cyt-*b*s system, Atlantic salmon and rainbow trout branches were present in family Scombridae (Fig. 4), while, in the COI system, tilapia branches were far from family Scombridae (Fig. 5).

In this study, it was necessary to use the database in the web. The more fish species are to be analyzed, the more refined the model can be. Some fish species belonging to Acipenseriformes, Anguilliformes, Lophiiformes and Siluriformes were planned to be included in the present model, but the sequences from the limited numbers of fish species could be referred to. At present, it is difficult to find the species whose sequences of Mb, cyt-*b* and COI are available. For this reason, fish species examined were different for each phylogenetic analysis carried out in this study. The data for channel catfish (*I. punctatus*), greater amberjack (*S. dumerili*), milkfish (*C. chanos*), yellowtail amberjack (*S. lalandi*), Japanese amberjack (*S. quinqueradiata*), longfin tuna (*T. alalunga*), black rockcod (*N. coriiceps*) and common carp (*C. carpio*) were not available for the cyt-*b* trees (Fig. 4).

Compared with cyt-*b*s and COIs, Mbs provide advantage in their shorter sequences, which made it easier to identify the species or to estimate the relationship among fishes. The tertiary structures of Mbs were similar to each other among medaka, bluefin tuna, Mozambique tilapia, Atlantic salmon and zebrafish, although they were clearly distinguishable from each other in the primary structure (Fig. 1). The primary structures of Mbs were found to be highly variable at the 21st, 123rd and 131st residues, which are not reflected to their tertiary structures (Fig. 2). The phylogenetic trees drawn based on Mb sequences were similar to those by the conventional classification based on other markers (Figs. 4 and 5).

5. Conclusions

This study mainly focused on the development of a new marker for the problems encountered in fish species identification. The small-sized and conserved proteins can be excellent candidates for this purpose. Although many reports have discussed the theory of phylogenetic trees, the mathematical principles were not referred to in this study, but instead, a common model has been adopted to draw the trees. The present study showed that, although the Mb proteins are smaller (only 147 aa) compared to cyt-*b* (380 aa) or COI (516 aa), it is not inferior to the other two. This is the strongest advantage of Mb. The results obtained in this study suggest that Mb can be available as a useful molecular marker for the phylogenetic analysis of fish.

Acknowledgments

The funding of this study was provided in part by the National University of Tainan, Taiwan R.O.C.

References

[1] Froese, R., and Pauly, D., eds. 2018. FishBase. World Wide Web Electronic Publication. Accessed February 2018. http://www.fishbase.org.

[2] Itoi, S., Saito, T., Shimojo, M., Washio, S., and Sugita, H. 2007. "Identification of *Girella punctata* and *G. leonina* by PCR-RFLP Analysis." *ICES J. Marine Sci*. 64 (2): 328-31.

[3] Arnason, U., Bodin, K., Gullberg, A., Ledje, C., and Mouchaty, S. 1995. "A Molecular View of Pinniped Relationships with Particular Emphasis on the True Seals." *J. Mol. Evol.* 40 (1): 78-85.

[4] Lara, M. C., Patton, J. L., and Da Silva, M. N. 1996. "The Simultaneous Diversification of South American Echimyid Rodents (Hystricognathi) Based on Complete Cytochrome *b* Sequences." *Mol. Phylogenet. Evol*. 5 (2): 403-13.

[5] Faulkes, C. G., Bennett, N. C., Bruford, M. W., O'Brien, H. P., Aguilar, G. H., and Jsrvis, J. U. 1997. "Ecological Constraints Drive Social Evolution in the African Mole-Rats." *Proc. Biol. Sci.* 264 (1388): 1619-27.

[6] LeDuc, R. G., Perrin, W. F., and Dizon, A. E. 1999. "Phylogenetic Relationships among the Delphinid Cetaceans Based on Full Cytochrome *b* Sequences." *Mar. Mamm. Sci.* 15 (3): 619-48.

[7] Matthee, C. A., and Robinson, T. J. 1999. "Cytochrome *b* Phylogeny of the Family Bovidae: Resolution within the Alcelaphini, Antilopini, Neotragini and Tragelaphini." *Mol. Phylogenet. Evol.* 12 (1): 31-46.

[8] Castresana, J. 2001. "Cytochrome *b* Phylogeny and the Taxonomy of Great Apes and Mammals." *Mol. Biol. Evol.* 18 (4): 465-71.

[9] Hebert, P. D., Ratnasingham, S., and De Waard, J. R.

2003. "Barcoding Animal Life: Cytochrome *c* Oxidase Subunit I Divergences among Closely Related Species." *Proc. Biol. Sci.* 270 (Suppl. 1): S96-9.

[10] Faustman, C., and Phillips, A. 2001. "Measurement of Discoloration in Fresh Meat." In *Current Protocols in Food Analytical Chemistry,* edited by Wrolstad, R. E. New York: Wiley & Sons, Inc., F3.3.1-13.

[11] Di Prisco, G., Cocca, E., Parker, S. K., and Detrich, H. W. 2002. "Tracking the Evolutionary Loss of Hemoglobin Expression by the White-Blooded Antarctic Icefishes." *Gene* 295 (2): 185-91.

[12] Douglas, E. L., Peterson, K. S., Gyso, J. R., and Chapman, D. J. 1985. "Myoglobin in the Heart Tissue of Fishes Lacking Hemoglobin." *Comp. Biochem. Physiol. A Comp. Physiol*. 81 (4): 855-88.

[13] You, Y. C., and Huang, M. C. 2016. "The Tropomyosin Structure and Phylogenetic Analysis of Five Fish Species." *Int. J. Sci. Engin*. 6 (1): 23-42. (in Chinese with English abstract)

[14] Shannon, C. E. 1948. "The Mathematical Theory of Communication." *Bell Labs Tech. J*. 27 (3): 379-423.

[15] Garnier, J., Gibrat, J. F., and Robson, B. 1996. "GOR Method for Predicting Protein Secondary Structure from Amino Acid Sequence." *Methods Enzymol*. 266: 540-53.

[16] Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L., and Schwede, T. 2014. "SWISS-MODEL: Modeling Protein Tertiary and Quaternary Structure Using Evolutionary Information." *Nucleic Acids Res.* 42 (Web Server Issue): W252-8.

[17] Jones, D. T., Taylor, W. R., and Thornton, J. M. 1992. "The Rapid Generation of Mutation Data Matrices from Protein Sequences." *Comput. Applic. Biosci.* 8 (3): 275-82.

[18] Saitou, N., and Nei, M. 1987. "The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees." *Mol. Biol. Evol.* 4: 406-25.

[19] Sneath, P. H. A., and Sokal, R. R. 1973. *Numerical Taxonomy: The Principles and Practice of Numerical Classification.* San Francisco: Freeman, 573.

[20] Eck, R. V., and Dayhoff, M. O. 1966. *Atlas of Protein Sequence and Structure.* Silver Spring, MD: National Biomedical Research Foundation, 161-9.

[21] Fitch, W. M. 1971. "Toward Defining the Course of Evolution: Minimum Change for a Specified Tree Topology." *Syst. Biol*. 20 (4): 406-16.

[22] Fitch, W. M. 1977. "On the Problem of Discovering the Most Parsimonious Tree." *Am. Nat*. 111: 223-57.

[23] Felsenstein, J. 1985. "Confidence Limits on Phylogenies: An Approach Using the Bootstrap." *Evolution* 39 (4): 783-91.

[24] Kumar, S., Stecher, G., and Tamura, K. 2016. "MEGA 7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets." *Mol. Biol. Evol*. 33 (7): 1870-4.

[25] Nei, M., and Kumar, S. 2000. "Evolutionary Change of Amino Acid Sequence." In *Molecular Evolution and Phylogenetics,* edited by Nei, M., and Kumar, S. New York: Oxford University Press, 17-32.

[26] Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S., and Miyata, T. 1989. "Evolutionary Relationship of Archaebacteria, Eubacteria and Eukaryotes Inferred from Phylogenetic Trees of Duplicated Genes." *Proc. Natl. Acad. Sci. USA*. 86: 9355-9.

[27] Brown, J. R., and Doolittle, W. F. 1995. "Root of the Universal Tree of Life Based on Ancient Aminoacyl–tRNA Synthetase Gene Duplications." *Proc. Natl. Acad. Sci. USA* 92: 2441-5.

[28] Hashimoto, T., and Hasegawa, M. 1996. "Origin and Early Evolution of Eukaryotes Inferred from the Amino Acid Sequences of Translation Elongation Factors 1α/Tu and 2/G." *Adv. Biophys*. 32: 73-120.

[29] Yang, Z., and Roberts, D. 1995. "On the Use of Nucleic Acid Sequences to Infer Early Branchings in the Tree of Life." *Mol. Biol. Evol*. 12: 451-8.

[30] Hasegawa, M., and Hashimoto, T. 1993. "Ribosomal RNA Trees Misleading?" *Nature* 362 (6423): 795.

[31] Hebert, P. D., Stoeckle, M. Y., Zemlak, T. S., and Francis, C. M. 2004. "Identification of Birds through DNA Barcodes." *PLoS Biol.* 2 (10): e312.

[32] Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., and Hebert, P. D. 2005. "DNA Barcoding Australia's Fish Species." *Philos. Trans. R. Soc. Lond. B. Biol. Sci*. 360 (1462): 1847-57.

[33] Hajibabaei, M., Janzen, D. H., Burns, J. M., Hallwachs, W., and Hebert, P. D. 2006. "DNA Barcodes Distinguish Species of Tropical Lepidoptera." *Proc. Natl. Acad. Sci. USA* 103: 968-71.

[34] Brown, W. M. 1985. "The Mitochondrial Genome of Animals: Molecular Evolutionary Genetics." In *Molecular Evolutionary Genetics,* edited by MacIntyre, R. J. New York: Plenum Press, 95-130.

[35] Santos, S., Schneider, H., and Sampaio, I. 2003. "Genetic Differentiation of *Macrodon ancylodon* (Sciaenidae: Perciformes) Populations in Atlantic Coastal Waters of South America as Revealed by mtDNA Analysis." *Genet. Mol. Biol.* 26 (2): 151-61.

[36] Vinson, C., Grazielle, G., Schneider, H., and Sampaio, I. 2004. "Sciaenidae Fish of the Caete River Estuary, Northern Brazil: Mitochondrial DNA Suggests Explosive Radiation for the Western Atlantic Assemblage." *Genet. Mol. Biol.* 27 (2): 174-80.

[37] Zemlak, T. S., Ward, R. D., Connell, A. D., Holmes, B. H., and Hebert, P. D. N. 2009. "DNA Barcoding Reveals Overlooked Marine Fishes." *Mol. Eco. Res.* 9 (Suppl. 1):

237-42.

[38] Bonde, N. 1994. "Simplified Phylogenetic Tree of the Fishes." In *Quality and Quality Changes in Fresh Fish,* edited by Huss, H. H. Rome: Food and Agriculture Organization of United Nations, 10.

[39] Phillips, S. E. V., and Schoenborn, B. P. 1981. "Neutron Diffraction Reveals Oxygen-Histidine Hydrogen Bond in Oxymyoglobin." *Nature* 292: 81-2.

[40] Chow, C. J., Ochiai, Y., Watabe, S., and Hashimoto, K. 1985. "Effects of Freezing and Thawing on the Autoxidation of Bluefin Tuna Myoglobin." *Bulletin of the Japanese Society of Scientific Fisheries* 51 (12): 2073-8.

[41] Kitahara, Y., Matsuoka, A., Kobayashi, N., and Shikama, K. 1990. "Autoxidation of Myoglobin from Bigeye Tuna Fish (*Thunnus obesus*)." *Biochim. Biophys. Acta* 1038: 23-8.

[42] Chanthai, S., Ogawa, M., Tamiya, T., and Tsuchiya, T. 1996. "Studies on Thermal Denaturation of Fish Apomyoglobins Using Differential Scanning Calorimetry, Circular Dichroism and Fluorescence." *Fish. Sci*. 62 (6): 927-32.

[43] Madden, P. W., Babcock, M. J., Vayda, M. E., and Cashon, R. E. 2004. "Structural and Kinetic Characterization of Myoglobins from Eurythermal and Stenothermal Fish Species." *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.* 137 (3): 341-50.

[44] Watts, D. A., Rice, R. H., and Brown, W. D. 1980. "The Primary Structure of Myoglobin from Yellowfin Tuna (*Thunnus albacares*)." *J. Biol. Chem*. 255 (22): 10916-24.

[45] Birnbaum, G. I., Evans, S. V., Przybylska, M., and Rose, D. R. 1994. "1.70 Å Resolution Structure of Myoglobin from Yellowfin Tuna. An Example of a Myoglobin Lacking the D Helix." *Acta Cryst.* D50: 283-9.

[46] Fosmire, G. J., and Brown, W. D. 1976. "Yellowfin Tuna (*Thunnus albacares*) Myoglobin: Characterization and Comparative Stability." *Comp. Biochem. Physiol*. *B.* 55 (2): 293-9.

[47] Marcinek, D. J., Bonaventura, J., Wittenberg, J. B., and Block, B. A. 2001. "Oxygen Affinity and Amino Acid Sequence of Myoglobins from Endothermic and Ectothermic Fish." *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 280 (4): R1123-33.

[48] Anttila, K., Dhillon, R. S., Boulding, E. G., Farrell, A. P., Glebe, B. D., Elliott, J. A., Wolters, W. R., and Schulte, P. M. 2013. "Variation in Temperature Tolerance among Families of Atlantic Salmon (*Salmo salar*) Is Associated with Hypoxia Tolerance, Ventricle Size and Myoglobin Level." *J. Exp. Biol.* 216 (Pt 7): 1183-90.

[49] Lurman, G. J., Koschnick, N., Pörtner, H. O., and Lucassen, M. 2007. "Molecular Characterization and Expression of Atlantic Cod (*Gadus morhua*) Myoglobin from Two Populations Held at Two Different Acclimation Temperatures." *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* 148 (3): 681-9.

[50] Ueki, N., and Ochiai, Y. 2004. "Primary Structure and Thermostability of Bigeye Tuna Myoglobin in Relation of Those of Other Scombridae Fish." *Fish. Sci.* 70: 875-84.

[51] Luo, Y., and Baldwin, R. L. 2001. "How Ala→Gly Mutations in Different Helices Affect the Stability of the Apomyoglobin Molten Globule." *Biochemistry* 40 (17): 5283-9.

[52] Loh, S. N., Kay, M. S., and Baldwin, R. L. 1995. "Structure and Stability of a Second Molten Globule Intermediate in the Apomyoglobin Folding Pathway." *Proc. Natl. Acad. Sci. USA* 92: 5446-50.

[53] Fontana, A., Zambonin, M., De Laureto, P. P., De Fillipis, V., Clementi, A., and Scaramella, E. 1997. "Probing the Conformational State of Apomyoglobin by Limited Proteolysis." *J. Mol. Biol.* 266: 223-30.

[54] Hasan, M. M., Watabe, S., and Ochiai, Y. 2012. "Structural Characterization of Carangid Fish Myoglobins." *Fish Physiol. Biochem*. 38 (5): 1311-22.