

Cytochrome oxidase I (COI) sequence conservation and variation patterns in the yellowfin and longtail tunas

Swaraj Priyaranjan Kunal*

Biological Oceanography Division,
National institute of Oceanography,
Dona Paula, Goa 403 004, INDIA
Email: swarajpkunal@gmail.com

*Corresponding author

Girish Kumar

Biological Oceanography Division,
National institute of Oceanography,
Dona Paula, Goa 403 004, INDIA
Email: geniusgirish@gmail.com

Abstract

Tunas are commercially important fishery world wide. There are at least 13 species of tuna belonging to three genera, out of which genus *Thunnus* has maximum eight species. Based on their availability, they can be characterized as oceanic such as *Thunnus albacares* (yellowfin tuna) or coastal such as *Thunnus tonggol* (longtail tuna). Although these two are different species, but morphological differentiation can only be seen in mature individuals, hence misidentification may result in erroneous data set, which ultimately affect conservation strategies. The mitochondrial DNA cytochrome oxidase c subunit 1 (COI) gene is one of the most popular markers for population genetic and phylogeographic studies across the animal kingdom. The present study aims to study the sequence conservation and variation in mitochondrial Cytochrome oxidase I (COI) between these two species of tuna. COI sequence analysis of yellowfin and longtail revealed the close relationship between them in *Thunnus* genera. The present study is first direct comparison of mitochondrial COI sequences of these two tuna species.

Key words: Cytochrome oxidase I; Mitochondrial DNA; *Thunnus albacares*; *Thunnus tonggol*.

Biographical notes:

Swaraj Priyaranjan Kunal is a PhD student at NIO Goa. He gained his MSc in Marine Biotechnology from Goa University, Goa, India. His research interests are in the areas of marine biotechnology, bioinformatics and evolutionary biology.

Girish Kumar is a PhD student at NIO Goa. He gained his MSc in Genetics from MDU, Rohtak, India. His research interests are in the areas of , molecular phylogenetics, population genetics and evolutionary biology.

Introduction

Tunas are salt water (Oceanic as well as coastal) fishes, categorized in three genus of family Scombridae. Species belonging to genus *Thunnus* are called true tunas. Therefore *Thunnus albacares* (yellowfin tuna) and *Thunnus tonggol* (longtail tuna) are among species of true tuna. *Thunnus albacares* has a global distribution around the major oceanic waters between 40° N and 40° S. An epipelagic, oceanic species, generally found above and below thermocline. The thermal boundaries of occurrence are in the range of 18-31 C°, and so this species is usually confined to upper 100m of the water column [1](Collete and Nauen, 1983). It is regarded among major contributor to the global marine fisheries and constitutes approx 27% of all tuna catches worldwide [2](FAO, 2008). As far as Indian waters are concerned, it is the most dominant species among larger pelagic fishes, contributing approx 19% (13228 metric tonnes of all tuna catches [3](IOTC 2010).

Thunnus tonggol is second smallest species of the *Thunnus* genus. It has a narrow coastal distribution in tropical and temperate waters of Indo-Pacific region. *Thunnus tonggol* is a commercially important fishery with its great demand in export market. Global catch rose to 233,830 t in 2009 [4](FAO, 2010), while in Indian waters it is 9% (6111 metric tonne) of all tuna catches, mostly distributed around north-west coast of India [3](IOTC, 2010).

Generally the morphological traits to differentiate between these two tuna species appear once the individual matures; hence misidentification of juveniles of these two species is unavoidable [5](Chow and Kishino, 1995). Molecular genetic markers had been used for tuna species identification and as well as phylogenetic relationship studies [6-12](Alvarado Bremer et al., 1997; Takeyama et al., 2001; Terol et al., 2002; Lin et al., 2005; Chow et al., 2006; Paine et al., 2008; Vinas et al., 2009). Mitochondrial DNA (mtDNA) is maternally inherited and has a relatively fast mutation rate. It is for this reason, there appears to be significant variation in mtDNA sequences between species and comparatively small variance within species [13](Moritz et al., 1987). In particular, the COI gene has been used so far to identify species of a tribe [14](Wang et al., 2011). The COI genes are preferred over other mitochondrial genes as it has lower mutation rates [15](Yi et al., 2002) and the third position of the codons shows a high incidence of nucleotide substitution as compared to other protein genes [16](McClellan, 2000).

The present sequence analysis of COI genes addresses the sequence conservation and variation patterns in the yellowfin and longtail tunas sequence conserved patterns as well as evolutionary distance analysis within the thunnini tribe. The Phylogenetic analysis was also performed for studying phylogenetic relationships among members of the tribe. This evolutionary information will provide new insight to study and identification of true tunas.

Material and methods

Data source

COI sequences of yellowfin tuna (*Thunnus albacares*) and longtail tuna (*Thunnus tonggol*) were retrieved from NCBI nucleotide database and processed further to remove unnamed and putative sequences. The final data sets contain 42 sequences of *Thunnus albacares* and 14 sequences of *Thunnus tonggol*. Both the sequences data sets were aligned separately in Clustal X 2.0 [17] and terminal analigned regions were spliced out.

Sequence conservation analysis

Conserved regions in the *COI* sequence alignments of both the data sets (yellowfin and longtail) were mapped manually by careful observation. In spite of mono, di and tri nucleotide conserved regions being present in the alignment, minimum 15 nucleotide stretches were taken in account for mapping conserved region. BioEdit version 7.0.5 [18] was used for finding the conserved regions. Composition distances of sequences, i.e., differences in nucleotide composition for all pair of sequences in the given sequence alignments, excluding gap regions, were calculated. Nucleotide substitution sites, i.e., transition – transversion sites in the alignments were mapped and their frequencies, transition – transversion rate, ratios of purines and pyrimidines were calculated, all positions containing gaps and missing data were eliminated from the dataset. Transition – transversion bias were estimated by Tamura and Nei's method [19](Tamura et al., 2004). All calculations were conducted in MEGA4 [20](Tamura et al., 2007)

COI sequence variation analysis

Polymorphic sites i.e., sites with two, three and four variants in the alignments of each data set were analyzed, in this study gap regions were excluded and insertion-deletion polymorphism was studied based on the non-overlapping and overlapping indel sites by using DNAsp software [21](Librado and Rojas, 2009).

Sequence distance estimation

COI sequences of Thunnini tribe, intraspecific pair wise distances were calculated using the Maximum Composite Likelihood method in MEGA4 [20](Tamura et al., 2007). Codon positions included were 1st+2nd+3rd+Non-coding. All positions containing gaps and missing data were eliminated from the data set.

Phylogenetic relationship

The aligned thunnini tribe *COI* sequences were taken to study the phylogenetic analysis, and the remaining unaligned sequences were flushed. The evolutionary history was inferred using the Neighbor-Joining method [22](Satiou and Nei, 1987). The percentage of replicate trees in which the associated taxa clustered together was tested in the bootstrap test by 1000 replicates [23](Felsenstein, 1985). The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated from the dataset.

Results and Discussion

Yellowfin and Longtail Tuna COI sequence conservation

COI sequence conservation analysis showed 9 conserved regions in yellowfin and 13 conserved regions in longtail tuna. All the conserved regions with respect to its positions were depicted in Table 1. Multiple sequence alignments of both the data sets were presented in Supplementary material.

The nucleotide frequencies of yellow-fin *COI* are 0.241 (A), 0.287 (T/U), 0.283 (C), and 0.189 (G). The transition/transversion rate ratios are $k_1 = 1.899$ (purines) and $k_2 = 3.877$ (pyrimidines). The overall transition/transversion bias is $R = 1.527$. There were a total of 652 positions in the final dataset. Longtail-tuna *COI* nucleotide frequencies are 0.235 (A), 0.299 (T/U), 0.277 (C), and 0.189 (G). The transition/transversion rate ratios are $k_1 = 61.21$ (purines) and $k_2 = 80.702$ (pyrimidines). There were a total of 596 positions in the final dataset. The overall transition/transversion bias is $R = 36.256$, where

$$R = [A*G*k_1 + T*C*k_2]/[(A+G)*(T+C)]$$

Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated from the datasets.

COI sequence polymorphism

The COI genes of yellowfin and longtail tuna are highly conserved. The sequence polymorphic site analysis of the both species reveals that the analysed regions consists 13 polymorphic sites, in which 10 are single variant sites and 3 are two variant sites in yellowfin tuna sequences. Long tail tuna sequence analysis shown that there were 22 polymorphic sites, in which 20 sites are single variant sites and 2 sites are two variant sites. A detailed picture of polymorphic sites was presented in Table 2.

Thunnini tribe COI evolutionary distances

The COI sequence pair of thunnini tribe evolutionary distances ranges from 0.000 to 0.144. *Katsuwonus pelamis* and *Thunnus maccoyii* sequence pair distances is 0.144, this suggests that both the species are distantly related within the family. The comparative sequence pair distances analysis within the *Thunnus* genus showed that the sequence pair distances are low. The sequence pair distances ranges from 0.002 to 0.017. *Thunnus albacares* (yellowfin) and *Thunnus tonggol* (longtail) COI sequence was closely related and pairwise distance observed were 0.002. Intra species pairwise distances of *Thunnus* genus were highlighted in the Table 3.

Evolutionary relationship

Phylogenetic relationships among species of thunnini tribe (family: Scombridae) were inferred based on COI genes. The data set was analyzed under Neighbor-Joining method. In the phylogram all the tuna species are well clustered (Fig1). The phylogram reveals two major clades i.e., *Thunnus* being the major one while rest all in other. *Thunnus* genera recovered as monophyletic group in the phylogram having yellowfin and longtail tunas in a closely clustered and easily identifiable sub-clade. Both the species place in the phylogram was marked in the phylogram for better understanding. *Thunnus alalunga* species is distantly placed in the clade one; it shows high sequence variable species within the *Thunnus* genus. The second major clade found to have *Auxis*, *Euthynnus* and *Katsuwonus* genera. In this clade all the genera were recovered well and *Katsuwonus pelamis* species was showing close relationship with *Auxis* genera.

Conclusion

Mitochondrial *COI* gene sequences analysis of yellowfin and longtail tunas showed that both species are closely related and having the least evolutionary distance within the *Thunnus* genera. The studies also showed the sequence variation is more in longtail tuna COI sequences. The elucidated conserved region sequences in both the species have potential implications as marker sequences which can be used for molecular identification at genus level. Further phylogenetic studies showed genus *Thunnus* clustered and retrieved well in the phylogram, a clear evidence of conservation of this genus within the thunnini tribe and provide a strong base for molecular evolutionary analysis.

Acknowledgements

The authors acknowledge Department of Science and Technology, New Delhi, India for the award of fellowship. We are grateful to the Director of the National Institute of Oceanography, India for valuable support and encouragement. The author SPK is thankful to the Head, Department of Biotechnology, Acharya Nagarjuna University, India for providing valuable suggestions for the study. This is NIO's contribution no. xxxx

References:

- 1 Collete, B. B., and Nauen, C. E. (1983) FAO species catalogue, vol. 2. Scombrids of the world. An annotated and illustrated catalogue of tunas, mackerels, bonitos, and related species known to date. *FAO Fish Synop.* **125**, 1-137.
- 2 FAO. (2008) Fisheries statistical collection: Global Tuna catches by stock.
- 3 Vijayakumaran, K., and Sijo, P. V. (2010) Update on tuna fisheries in India. IOTC-2010-SC-12
- 4 FAO Report. (2010) The State of World Fisheries and Aquaculture.
- 5 Chow, S., and Kishino, H. (1995) Phylogenetic relationships between tuna species of the genus *Thunnus* (Scombridae: Teleostei): inconsistent implications from morphology, nuclear and mitochondrial genomes. *J. Mol Evol.* **41**, 741–748
- 6 Alvarado Bremer, J.R., Naseri, I., and Ely, B. (1997) Orthodox and un-orthodox phylogenetic relationships among tunas revealed by the nucleotide sequence analysis of the mitochondrial control region. *J. Fish Biol.* **50**, 540-554.
- 7 Takeyama, H., Chow, S., Tsuzuki, H., and Matsunaga, T. (2001) Mitochondrial DNA sequence variation within and between tuna *Thunnus* species and its application to species identification. *J Fish Biol.* **58**, 1646–1657.
- 8 Terol, J., Mascarell, R., Fernandez-Pedrosa, V., Perez-Alonso, M. (2002) Statistical validation of the identification of tuna species: bootstrap analysis of mitochondrial DNA sequences. *J Agric Food Chem.* **50**, 963–969.
- 9 Lin, W. F., Shiau, C. Y., and Hwang, D. F. (2005) Identification of four *Thunnus* tuna species using mitochondrial cytochrome b gene sequence and PCR-RFLP analysis. *J. Food Drug Anal.* **13**, 382–387.
- 10 Chow, S., Nakagawa, T., Suzuki, N., Takeyama, H., and Matsunaga, T. (2006) Phylogenetic relationships among *Thunnus* species inferred from rDNA *ITS1* sequence. *J Fish Biol.* **68**, 24–35.
- 11 Paine, M. A., McDowell, J. R., Graves, J. E. (2007) Specific identification of Western Atlantic Ocean scombrids using mitochondrial DNA cytochrome c oxidase subunit I (*COI*) gene region sequences. *Bull Mar Sci.* **80**, 353–367.

- 12 Viñas, J., and Tudela, S. (2009) A Validated Methodology for Genetic Identification of Tuna Species (Genus *Thunnus*). *PLoS ONE* **4**(10), e7606. doi:10.1371/journal.pone.0007606.
- 13 Moritz, C., Dowling, T. E., and Brown, W. M. (1987) Evolution of Animal Mitochondrial DNA: Relevance for Population Biology and Systematics. *An. Rev. Ecol. Syst.* **18**, 269-292.
- 14 Wang, J. F., Jiang, L. Y., and Qiao, G. X. (2011) Use of a mitochondrial COI sequence to identify species of the subtribe Aphidina (Hemiptera, Aphididae). *ZooKeys* **122**: 1–17. doi: 10.3897/zookeys.122.1256.
- 15 Yi, S., Ellsworth, D. L., and Li, W. H. (2002) Slow molecular clocks in old world monkeys, apes and humans. *Mol. Biol and Evol.* **19**, 2191-2198.
- 16 McClellan, D. A. (2000) The codon-degeneracy model of molecular evolution. *J. Mol. Evol.* **50**, 131-140.
- 17 Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.
- 18 Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* **41**:95-98.
- 19 Tamura, K., Dudley, J., Nei, M., & Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. and Evol.* **24**, 1596-1599.
- 20 Tamura, K., Nei, M., & Kumar, S. (2004) Prospects for inferring very large phylogenies by using the neighbour-joining method. *Proceedings of the National Academy of Sciences (USA)*. **101**, 11030-11035.
- 21 Librado, P., and Rozas, J. (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics.* **25**, 1451-1452.
- 22 Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406-425.
- 23 Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution.* **39**, 783-791.

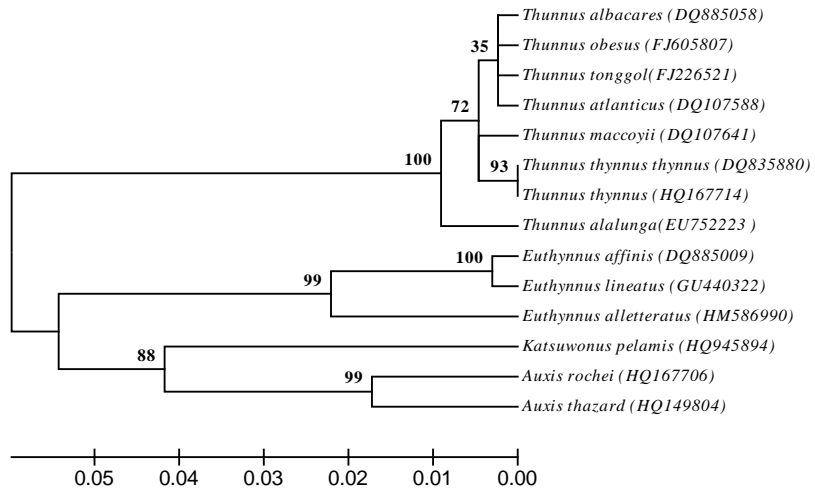


Figure 1: Phylogenetic tree of thunnini tribe.

Table 1: Conserved *COI* patterns in the yellowfin and long tail fin tunas.

S. No	Yellowfin		Longtail	
	Position	Sequence	Position	Sequence
1	5 to 65	-TATCTAGTATTCGGTGCATGAGCTGGAATAGTT GGCACGGCCCTTAAGCTTGCTCATCCGAG-	3 to 21	-TTTATCTAGTATTCGGTGC-
2	74 to 255	-AGCCAACCAGGTGCCCTTCTTGGGGACGACCAG ATCTACAATGTAATCGTTACGGCCCATGCCTTCG TAATGATTTTCTTTATAGTAATACCAATTATGAT TGGAGGATTTGGAAACTGACTTATTCCTCTAATG ATCGGAGCCCCGACATGGCATTCCACGAATGA ACAACATGAGCTT-	76 to 100	-CCAACCAGGTGCCCTTCTTGGGGAC-
3	257 to 288	-TGACTCCTTCCCCCTCTTTCCTTCTGCTCCT-	107 to 212	-ATCTACAATGTAATCGTTACGGCCCATGCCTTC GTAATGATTTTCTTTATAGTAATACCAATTATGA TTGGAGGATTTGGAAACTGACTTATTCCTCTAAT GATCG-
4	295 to 339	-TTCAGGAGTTGAGGCTGGAGCCGGAACCGGTTG AACAGTCTACCC-	214 to 234	-AGCCCCGACATGGCATTCCC-
5	356 to 429	-CTGGCCCACGCAGGGGCATCAGTTGACCTAACT ATTTTCTCACTTCACTTAGCAGGGTTTCTCAA TTCTTGG-	238 to 261	-AATGAACAACATGAGCTTCTGACT-
6	431 to 465	-GCAATTAACCTCATCACAACAATTATCAATATG AA-	263 to 329	-CTTCCCCCTCTTTCCTTCTGCTCCTAGCTTCT TCAGGAGTTGAGGCTGGAGCCGGAACCGTTGAA-
7	476 to 555	-ATTTCTCAGTATCAAACACCACTGTTTGTATGA GCTGTACTAATTACAGCTGTTCTTCTCCTACTTT CCCTTCCAGTCCT-	331 to 396	-AGTCTACCCTCCCCTTGCCGGCAACCTGGCCCA CGCAGGGGCATCAGTTGACCTAACTATTTTCTC-
8	557 to 630	-GCCGCTGGTATTACAATGCTCCTTACAGACCGA AACCTAAATACAACCTTCTTCCACCTGCAGGAG GGGGAGA-	398 to 426	-CTTCACTTAGCAGGGGTTTCTCAATTCT-
9	632 to 652	-CCAATCCTTACCAACACCTA-	436 to 457	-TTAACTTCATCACAACAATTAT-
10			459 to 491	-AATATGAAACCTGCAGCTATTTCTCAGTATCAA-
11			493 to 533	-CACCACTGTTTGTATGAGCTGTACTAATTACAG CTGTTCTT-
12			535 to 550	-CTCCTACTTTCCTTC-
13			576 to 597	-TGCTCCTTACAGACCGAAACCT-

Table 2: Sequence polymorphism in yellowfin and longtail tuna.

Species	No. of Sequences	Monomorphic Sites	One Variable Sites	Two Variable Sites	Total Polymorphic Sites
Yellowfin	42	639	10	3	13
Longtail	14	574	20	2	22

Table 3: Sequence pairwise distances of *COI* gene of Thunnini tribe family.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
[1]														
[2]	0.004													
[3]	0.002	0.006												
[4]	0.134	0.134	0.132											
[5]	0.114	0.117	0.117	0.115										
[6]	0.119	0.122	0.121	0.099	0.044									
[7]	0.116	0.119	0.119	0.096	0.042	0.006								
[8]	0.126	0.127	0.124	0.087	0.118	0.101	0.104							
[9]	0.116	0.117	0.114	0.085	0.125	0.118	0.115	0.035						
[10]	0.006	0.010	0.008	0.139	0.119	0.119	0.121	0.124	0.114					
[11]	0.006	0.010	0.008	0.139	0.119	0.119	0.121	0.124	0.114	0.000				
[12]	0.012	0.017	0.010	0.127	0.122	0.117	0.114	0.124	0.114	0.014	0.014			
[13]	0.008	0.012	0.010	0.144	0.119	0.124	0.126	0.126	0.121	0.010	0.010	0.016		
[14]	0.002	0.006	0.004	0.132	0.112	0.117	0.114	0.124	0.114	0.008	0.008	0.014	0.010	

[1] *Thunnus albacares*

[4] *Katsuwonus pelamis*

[7] *Euthynnus lineatus*

[10] *Thunnus thynnus thynnus*

[13] *Thunnus maccoyii*

[2] *Thunnus obesus*

[5] *Euthynnus alletteratus*

[8] *Auxis thazard*

[11] *Thunnus thynnus*

[14] *Thunnus tonggol*

[3] *Thunnus atlanticus*

[6] *Euthynnus affinis*

[9] *Auxis rochei*

[12] *Thunnus alalunga*