Received: 28 September, 2011
IOTC-2011-WPNT01-12

PHYLOGENETIC RELATIONSHIPS OF COASTAL TUNAS INFERRED FROM MITOCHONDRIAL DNA SEQUENCES IN THE CYTOCHROME C OXIDASE I (COI) GENE - A STUDY ON DNA BARCODING

V.K.Mudumala[1], V. S. Somvanshi[1] and W. S. Lakra[2]

[1]Fishery Survey of India, Mumbai,
[2]Central Institute of Fisheries Education, Mumbai, India

ABSTRACT

India has rich diversity of fish resources which are to be explored and exploited on sustainable basis. To accomplish the task there is a need to identify the species through DNA barcoding which is the ultimate tool for species identification. Tunas are highly migratory fish having commercial importance and some are extremely important. As tuna meat demand in the world market has witnessed an increasing trend, research on tuna resources and their exploitation patterns have also been actively pursued. Tuna and tuna like fishes belonging to the Tribe Thunnini, family Scombridae have four genera *viz Thunnus*, *Euthynnus*, *Katsuwonus* and *Auxis* with 13 species. The present study is attempted to know the sequence diversity in the mitochondrial Cytochrome c oxidase I (COI) gene as a tool to understand the differences among the genus and species occurring in the North-Western Indian EEZ *Viz*. *Thunnus tonggol*, *Euthynnus affinis*, *Auxis thazard* and *Auxis rochei.* Sequence divergences revealed that there is not much variation within the species and there reported large variations between the species. While analysing phylogenetic relationships of these coastal tunas using NJ (Neighbour - joining) method indicated shallow intra-specific and deep inter-specific divergence. The study provides an understanding of phylogenetic relationships existing among these tuna species and also helpful in retrieving reliable information about the species and would provide requisite solution to the current problem of species identification.

Keywords : Phylogenetic relationships, DNA barcoding, Cytochrome c oxidase I (COI),Sequence divergence

INTRODUCTION

Global fish diversity is reported to have 27,977 species belonging to 62 orders and 515 families Nelson (2006), wherein India's contribution is about 2487 (Anon, 2007) species forming 8.9% of total fish diversity of the world. Among them 1010 species are freshwater, 113 species are brackish water, 1364 species are marine. In Indian waters, due to excessive fishing pressure, some of the fishery resources have been over exploited and still a few are remaining under exploited. The ability to accurately identify species is fundamental to ecological research especially when

pertaining to studies of ecology and biological diversity. A tool to evaluate the assessment of richness and turnover across landscapes with hyper diverse taxa has become imperative more with the receding timeframe available against the rate of biodiversity loss. It can rightly be called as a "Race against Time". It therefore becomes absolutely essential for the taxonomists to identify, describe and categorize the organisms available in nature. Recently the use of DNA is directed to meet these objectives. The DNA based technologies have been suggested as the best option to "bridge" the gap between available taxonomic expertise and the need for the identification capability (Tautz *et al.* 2003). Researchers are keen to investigate and describe unrecognized diversity by way of inclusion of molecular information into taxonomic aspects. Tautz *et al.* (2003) made the case for a DNA-based taxonomic system. DNA sequence analysis has been used for more than three decades to assist species identification, but different sequences have been used for different taxonomic groups and in different places. Hebert *et al*.(2003) proposed that a single gene sequence woud be sufficient to differentiate all and proposed the use of mitochondrial gene cytochrome c oxidase I (COI) gene as a global bio-identification system for all animals.

Exploration and exploitation of the fishery resources not only augment the food production but also provide the information on the availability and their distribution. As some of the resources are reported to be falling under the status of extinction, endangered and threatened, it is necessary to identify and discover the diversity properly so as to take up the measures for the development, conservation and management. To accelerate the inventory and analysis of diversity, "The DNA Barcoding" has been evolved as a powerful tool. It will help in identifying species without any chance for doubt and be useful as a complement to the traditional, morphological taxonomy. Any biological research depends upon species diagnosis and the sole prospect for a sustainable identification capability lies in the construction of systems that employ DNA sequences as taxon "Barcodes". Genomic approaches to taxon identification exploit diversity among DNA sequences to identify organisms. These sequences can be viewed as genetic "Barcodes" that are embedded in every cell according to Kurtzman (1994). The identification of diversity by using molecular markers has recently been proposed and demonstrated though the use of mitochondrial genes viz. Cytochrome C oxidase I (COI) gene, Cytochrome b, 16s, 12s

and rRNA genes etc. These DNA barcodes provide a practical, standardized, species level identification and can be used in biodiversity assessment.

The use of nucleotide sequence differences in a single gene to investigate evolutionary relationships was first applied by Woese and Fox (1977). There are two separate tasks to which DNA barcodes are currently being applied, the first task involves the use of DNA data to distinguish between species (species identification) and the second towards the use of DNA data to discover new species (equivalent to species delimitation, species description). The Barcoding helps in recognizing known species and retrieving information about them and also in speedy discovery of the diversity yet to be recognized and named. Barcoding also provide vital new tools for appreciating and managing the changing biodiversity. The studies in the past on mtDNA variation required large tissue samples and time consuming protocol for isolation of mtDNA. But the use of the mtDNA primers and PCR amplification of the selected regions have made the examination of mtDNA much faster and easier.

The past research has established the utility of mitochondrial DNA sequences in differentiating closely related species. Mark *et al.* (2005) illustrated that greater sequence differences among closely related species average 5 to 10 fold finger in mitochondrial than nuclear gene and shorter segments of mtDNA. Intra-specific variation in mtDNA is low in most of the species. Even the barcode locality is limited to one gene, sequences from diverse organisms can be compared easily.

Tuna and tuna-like fishes belonging to the Tribe *Thunnini*, family Scombridae are sub classified into four genera viz *Thunnus, Euthynnus, Katsuwonus* and *Auxis* with 13 species. FAO had recently recognized that the species *Thunnus thynnus* (Northern bluefin tuna) was thought to be separate Atlantic and Pacific stocks. Serdey (2002) reported that the Northern bluefin tuna (NBT) of the Atlantic and Pacific stocks are thought to be sub- species are actually two separate species and brought all the 14 species in the Scombridae family. DNA barcoding will be helpful in avoiding such taxonomic ambiguity.

The present study attempts to study the phylogenetic relationship existing among the four Indian Coastal Tuna species (*Thunnus tonggol, Euthynnus affinis, Auxis thazard* and *Auxis rochei*) by using the DNA sequences.

The molecular tools developed in the biotechnology identify the species at molecular level and are poised to revolutionize the identification system. The present

work has been attempted to barcode the coastal tunas occurring in the North western Indian EEZ.

Materials and Methods

Samples of *Thunnus tonggol, Euthynnus affinis, Auxis thazard* and *Auxis rochei* were collected from the Sassoon Dock fishing harbour, Mumbai and were brought to the Fishery Survey of India (FSI) laboratory. A piece of tissue cut (approx 50 mg) from the caudal peduncle region above the lateral line was preserved in 95% ethyl alcohol. The samples for the four species were taken to NBFGR, Lucknow for further analysis. A total of 20 samples for four tuna species (5 samples per species) were collected for the investigations. The DNA barcoding analysis was carried out by following the protocols/procedure designed for DNA isolation, Electrophoresis, Polymerase Chain Reaction (PCR) amplification, Sequencing and Sequence analysis. For DNA extraction most frequently used method is the phenol-chloroform method by Ruzzante et. al (1996) and was applied for this study.. Determination of quality and quantity for DNA isolated was carried out by electrophoresis method after RNAase treatment to know the quality and concentration and were checked on 0.7% of Agarose Gel. After extraction, Polymerase Chain Reaction (PCR), technique was applied with a small fragment of deoxyribonucleic acid (DNA) for rapid cloning, or duplication to produce multiple DNA copies. PCR was used to identify individuals species identification, evolution studies etc.

CO-I gene specific universal primers (SIGMA) were chosen for the analysis and are given below:

Fish F: 5'-TCAACCAACCACAAAGACATTGGCAC-3'

Fish R: 5'-TAGACTTCTGGGTGGCCAAAGAATCA-3'

Visualization of Amplified Products of Cytochrome C-Oxidase Subunit-I Gene on Agarose Gel was done. To know the Quality and quantity test of PCR product by Agarose gel was attempted. Then Standard 100 bp ladder DNA was ran simultaneously to estimate the length of bands.

The DNA Sequencing Was attempted to determine the order of the nucleotides of a gene. The chain termination method by Sanger et al, (1977) was applied to allow DNA sequences of several kb in length to determine in the minimum of time. The DNA sequence is the first and most basic type of information to be obtained about a cloned gene. For DNA Sequence Analysis DNASTAR Laser gene software has been

used for analysis DNA sequence which consists of a core functionality built on the DataManager and SeqBuilder. ClustalW a multiple sequence alignment program for DNA or proteins was applied to retrieve biologically meaningful multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be seen. Evolutionary relationships was seen via viewing Cladograms or Phylograms. After sequence alignment sequence divergences were calculated using the Kimura two parameter(K2P) distance model(Kmura,1980). Neighbour-joining (NJ)trees of K2P distances were created to provide a graphic representation of patterning of divergence between the species. Later the species under study bootstrapping was performed. The nucleotide composition of all the fishes have been studied using MEGA 3.1 software and the pair wise distances and intra-specific, inter-specific distances, within group divergences have been performed with K_2P model of complete deletion.

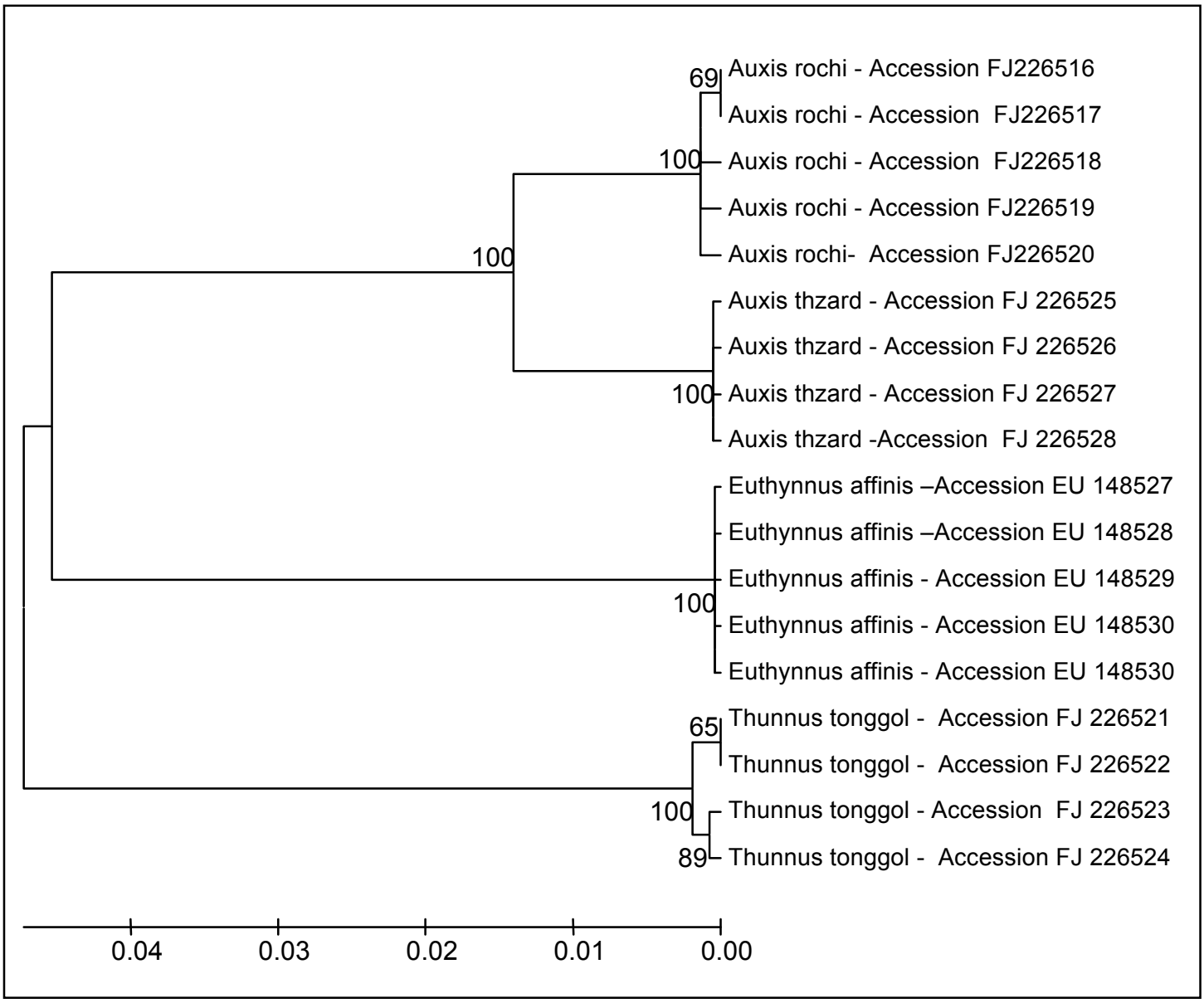

Fig 1 : Phylogenetic relationship of coastal tunas using Neighbour-Joining (NJ) Method

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Euthynnus affinis WL M790																		
2. Euthynnus affinis WL M789	0.000																	
3. Euthynnus affinis WL M474	0.000	0.000																
4. Euthynnus affinis WL M473	0.000	0.000	0.000															
5. Euthynnus affinis WL M472	0.003	0.003	0.003	0.003														
6. Auxis thzard 06902F	0.084	0.084	0.084	0.084	0.088													
7. Auxis thzard 06902G	0.086	0.086	0.086	0.086	0.089	0.002												
8. Auxis thzard 06902H	0.086	0.086	0.086	0.086	0.089	0.002	0.003											
9. Auxis thzard 06902I	0.084	0.084	0.084	0.084	0.088	0.000	0.002	0.002										
10. Auxis rochi 1317A	0.094	0.094	0.094	0.094	0.098	0.027	0.028	0.028	0.027									
11. Auxis rochi 1317B	0.096	0.096	0.096	0.096	0.100	0.030	0.031	0.031	0.030	0.003								
12. Auxis rochi 1317C	0.094	0.094	0.094	0.094	0.098	0.027	0.028	0.028	0.027	0.000	0.003							
13. Auxis rochi 1317D	0.093	0.093	0.093	0.093	0.096	0.025	0.027	0.027	0.025	0.002	0.005	0.002						
14. Auxis rochi 1317E	0.093	0.093	0.093	0.093	0.096	0.028	0.030	0.030	0.028	0.005	0.005	0.005	0.003					
15. Thunnus tonggol 1318B	0.093	0.093	0.093	0.093	0.097	0.097	0.099	0.099	0.097	0.088	0.090	0.088	0.086	0.090				
16. Thunnus tonggol 1318C	0.095	0.095	0.095	0.095	0.099	0.099	0.101	0.100	0.099	0.090	0.092	0.090	0.088	0.092	0.003			
17. Thunnus tonggol 1318D	0.093	0.093	0.093	0.093	0.097	0.097	0.099	0.099	0.097	0.088	0.090	0.088	0.086	0.090	0.000	0.003		
18. Thunnus tonggol 1318E	0.097	0.097	0.097	0.097	0.101	0.100	0.102	0.102	0.100	0.092	0.093	0.092	0.090	0.093	0.005	0.002	0.005	

Fig 2 : NJ Bootstrapping Test of Phylogeny (Pair wise distance calculation)

RESULTS

The simplest test of species identification by DNA barcode is to examine first whether any sequences are found in two species. Although sequences were not shared by species, sequence variation did occur in some species.

In the present study 3 genera were selected namely *Thunnus*, *Euthynnus* and *Auxis*. Under these genera 4 species of coastal tunas *Thunnus tonggol, Euthynnus affinis, Auxis thazard* and *Auxis rochei* have been chosen for the study. Of the 20 specimens (5 individuals per species) analyzed, 18 samples could only be amplified with the selected primer sets. The nucleotide composition of all the fishes have been studied using MEGA 3.1 software and the pair wise distances and intra-specific, inter-specific distances, within group divergences have been performed with K_2P model of complete deletion.

Nucleotide Composition

The percentage-wise nucleotide composition of coastal tunas have been computed and are presented in the Table 1 and the same also shown in Fig.3.3. The average composition was ‘A’ 24.0%, ‘T’ 30.2%, ‘G ’18.4% and ‘C’ 27.4%. The value of ‘AT’ (54.2%) is much higher than ‘GC’ (45.8 %). Among the three genera *Euthynnus affinis* shows the highest ‘AT’ content of 57.6 % and lowest ‘GC’ content 42.4%.

Table 1 : Nucleotide composition of coastal tunas

Name of the Fish	A %	T %	G %	C %	AT %	GC %
Auxis thazard	23.6	29.0	18.5	28.9	52.6	47.4
Auxis rochei	23.5	29.9	18.6	28.0	53.4	46.6
Euthynnus affinis	25.2	32.4	17.5	24.9	57.6	42.4
Thunnus tonggol	23.8	29.2	18.9	28.1	53.0	47.0
Average	**24.0**	**30.2**	**18.4**	**27.4**	**54.2**	**45.8**

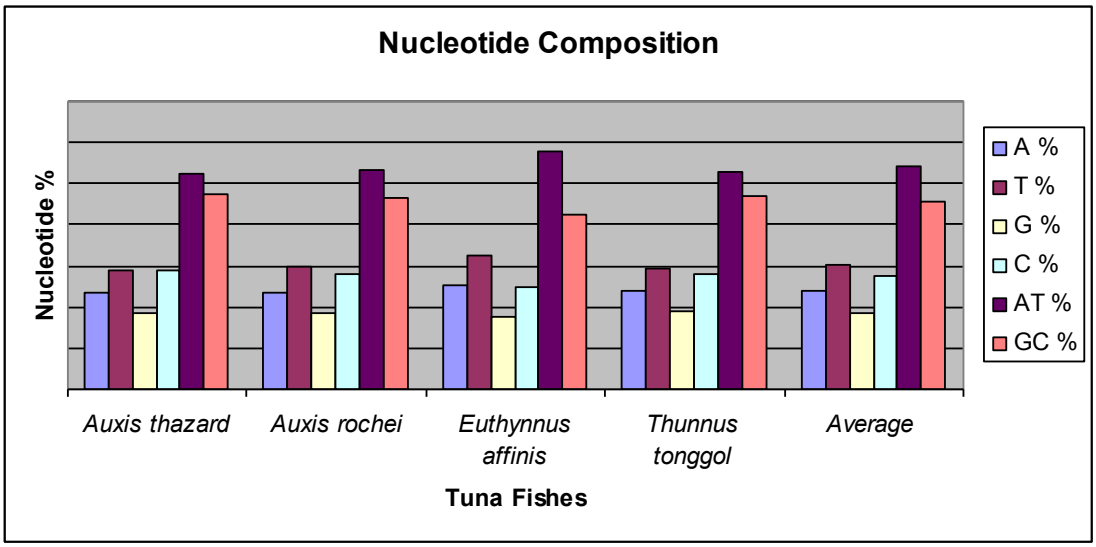

Fig. 3 : Percentage-wise Nucleotide composition of coastal tunas

The K_2P distances

The K_2P distances (Kimura's two parameter distance,1980) were analyzed to know the intra and inter specific variations of the species. The matrix of pair wise distances among the different fishes are presented in Fig.3.2. The overall mean K_2P distance among all the 18 individuals is 0.105.

The K_2P distances within the different 4 species of tunas of the present study are presented in Table 2 and shown in the Fig.

Table 2 : K_2P distances of four coastal tuna species

Species of Fish	Within Group distance
Euthynnus affinis	0.057
Auxis thazard	0.002
Auxis rochei	0.003
Thunnus tonggol	0.003

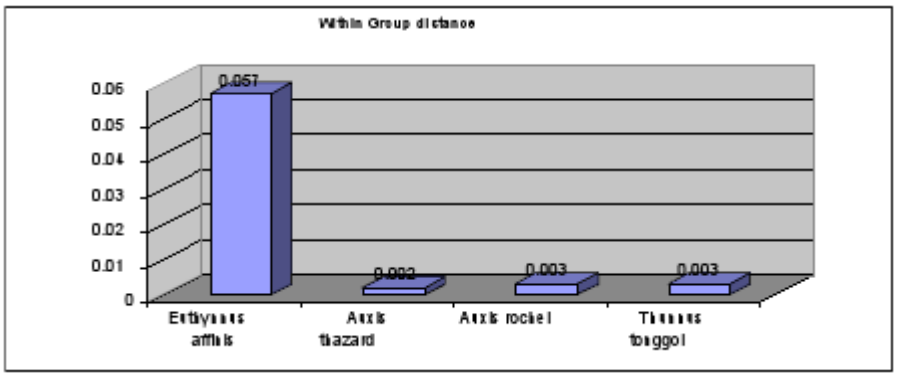

Fig. 4: Graphical presentation of K_2P distances of four coastal tuna species

The within group distances here can be considered as intra-specific distances, as each group is a single species with 4-5 individual specimens. The intra-specific K_2Pdistance within *Euthynnus affinis* is very high (0.057), when compared to all other three groups (0.002 to 0.003), while it is lowest in *Auxis thazard* (0.002).

The average **between group nucleotide distances** are

[1] *Euthynnus_affinis*

[2] *Auxis_thazard*

[3] *Thunnus_tonggol*

[4] *Auxis_rochei*

[1 2 3 4]

[1]

[2] 0.177

[3] 0.185 0.096

[4] 0.177 0.028 0.086

The mean value of Inter specific (between species / within a genus) divergence through K_2P distance reported between *Thunnus* and *Euthynnus* was 0.185, *Thunnus* and Auxis was 0.086 to 0.096. The range of divergence between species (inter-specific) reported was from 0.028 to 0.185.

When the between group distances (inter-specific/ within genus) were observed the group *Thunnus tonggol* is more distant to all other three groups (0.177 to 0.185) and the same can be observed in the phylogenetic tree also. While all the three groups *Euthynnus affinis, Auxis rochei* and *Auxis thazard* were depicted from the single branch with two sub branches for *Auxis* and *Euthynnus* genera, the genus *Thunnus* originates from a separate branch, which is again sub-divided in to two sub-branches with 69% and 85% bootstrap support, though they all are different individual specimens of the same species. The genus *Auxis* is again sub-divided in to two sub-branches as *A. thazard and A.rochei* species with 100% boot strap value, while *A.rochei* showed a sub-branch with two individuals with 69% bootstrap support.

The resulting phylogenetic tree indicated three different clusters. Among them *Auxis rochei* 2 individuals are showing divergence. In the cluster of *Thunnus tonggol* presence of 2 different sub groups and showing divergence between the same species and may be considered as sub species. There is no divergence observed in *Euthynnus affinis* as all of them clustered together (Fig.1). The sequence data were submitted to National Centre for Biotechnological Information (NCBI). The NCBI had authenticated data and allotted an accession numbers which were show in the phylogenetic tree. The phylogenetic relationship of these coastal tunas using NJ method (Neighbor joining) it showed shallow intra specific and deep inter specific divergences.

Discussion:

Thus, DNA barcoding is a novel system designed to provide rapid, accurate, species identifications by using short, standardized gene regions as internal species tags, which provide a DNA barcode for identifying species. A major goal of DNA barcoding is to enable the non taxonomist majority of biologists and indeed, anyone to access taxonomic information directly, while allowing professional taxonomists to focus on generating more such knowledge. DNA barcoding is positioned to aid the inventory of life by accelerating species discovery, by testing current taxonomic hypotheses, and by making species identifications more easily available. It was emphasized that DNA barcodes do not aim to recover phylogenetic relationships, even though they signal certain evolutionary relationships (Ward *et al.*, 2005), instead they seek to identify known species and to aid the discovery of new ones.

In the present study four species of Tuna were sequenced. The benchmark standard for any taxonomic system is its ability to deliver accurate species identification. COI barcoding for species identification is far more powerful than protein fingerprinting. One of the pre requisites for the success of any evolutionary method is the inter disciplinary approach involving details of geographical data such as place of fish sample collection, season, time, migratory factors, morphological identification coupled with the COI based identification system. Moreover, the alignment of COI sequences is straight forward, as indels were uncommon, reinforcing the results of earlier work showing the rarity of indels in this gene. Apart

from their ease of acquisition and alignment, the COI sequences possessed, a high level of diversity.

Chow *et al.* (2003) studied the Intra and inter specific nucleotides sequence variation of rDNA First Internal Transcribed Spacer (ITS1) and analyzed using all eight species of the genus *Thunnus* plus two out group species within the same family *viz.,* Skipjack tuna, *Katsuwonus pelamis* and Stripped bonito, *Sarda orientalis*. Intra specific nucleotide sequence variation in ITS1 including intra-genomic variation was low ranging from 0.003 to 0.014 (Kimura's two parameter distance K_2P) whereas variation between species within the genus *Thunnus* ranged from 0.009 to 0.05. The Atlantic and Pacific northern bluefin tuna *Thunnus thynnus thynnus* and *Thunnus thynnus orientalis*, recently proposed to be distinct species where found to share mainly identical ITS1 sequences (mean K2P = 0.006) well within the range of intra specific variation. The Northern bluefin tuna appeared to be a sister group to albacore tuna *Thunus alalunga*, with another *Thunnus* species in a distinct clade. The ITS 1 phylogeny was constant with mt DNA phylogeny in clustering the three tropical *Thunnus* species (*T. albacares*, *T. atlanticus* and *T. tonggol*). Southern bluefin tuna, *Thunnus maccoyii* and bigeye tuna *Thunnus obesus* showed a closer affinity to this tropical tuna group than to the northern bluefin tuna and albacore. The results obtained in his study, with regard to the K_2P distance in the present study also ranged between 0.002 to 0.102 and some of the species K_2P distances are comparable to the studies of Chow *et al.*

Persis (2009) studied the CO1 sequence divergence of marine fishes of Andhra coast and reported that, the K_2P distances between the two species of Tunas, *Euthynnus affinis* (EU541330) and *Auxis thazard* (EU541329) was 0.26% . The nucleotide composition of present study is 'A' 24.0%, 'T '30.2%, 'G '18.4% and 'C' 27.4%. The value of 'AT' (54.2%) is much higher than 'GC' (45.8 %). Among the three genera *Euthynnus affinis* shows the highest 'AT' content of 57.6% and lowest 'GC' content 42.4% in the north-west coast region. Whereas, the nucleotide composition reported for *Euthynnus affinis* was only 52.6% of 'AT' and 47.4% of 'GC'(Persis, 2009) in the upper east coast.

In the present study, emphasis is laid on the population of Tuna from the North - western Indian Exclusive Economic Zone. Samples of four species of Tuna namely

Thunnus tonggol, *Euthynnus affinis, Auxis thazard* and *Auxis rochei*. An attempt has been made to study the phylogenetic relationship existing among the four Indian Coastal Tuna species as mentioned above. The target of the barcoding study is to gather information on the intra-specific and inter-specific relationship among coastal Tuna in the Indian Coast

The resulting phylogenetic tree indicated three different clusters for three genera. Among them *Auxis rochei* 2 individuals are showing divergence in *Auxis rochei* and in *Thunnus tonggol* also prescence of 2 different sub groups and showing divergence between the same species and further studies may be taken up to look in to chances of separating them in to sub-species.

The use of modern molecular genetics techniques, combined with taxonomic expertise, provides a very powerful approach to solving existing taxonomic dilemmas, and allows new insights into the relatedness and evolution of fish species (Pegg *et al.*, 2005). The COI divergence studies have manifold advantages, in addition to species identification, in fisheries management, ecological and biodiversity conservation and bio-monitoring.

REFERENCES

ANDREW SERDEY 2002. One fin, two fins, red fins, blue fins – some problems of nomenclature and taxonomy affecting legal instruments governing tuna and other highly migratory species. *International Fisheries Bulletin*.Focus,No.5.

ANON. 2007**.** Annual Report. National Buraeu of Fish Genetic Resources.Lucknow,

CHOW, T. H S, TSUZUKI, H AND MATSUNAGA 2003 Mitochondrial DNA sequence variation within and between Thunnus tuna species and its application to species identification. Journal of Fish Biology 58: 1646-1657.

CHOW.S, T.NAKAGAWA, N SUZULA, H.TAKEYAMA AND T. MATSYNAGA 2006.Phylogenetic relationships among Thunnus species inferred from DNA ITS1 sequence. Journal of Fish Biology Vol.68 Issue of 24 – 35.

JOSEPH S, NELSON, 2006, Fishes of the world, 4th edition. John wiley & sons Inc; 624pp.

KURTZMAN, C.P. (1994) Molecular taxonomy of the yeasts. Yeast 10, 1727 1740.

KIMURA M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences.J. Mol. Evol. 16, 111-120.

MARK, S, PAUL E. WAGGONER. AND JESSE. H. AUSUBEL. 2005. Barcoding life, illustrated. Vol.3

PAUL D.N. HEBERT, ALINA CYWINSKA, SHELLEY L. BALL AND JEREMY R. DEWAARD, 2003 Biological identifications through DNA barcodes. *Proc. R. Soc. 2004 Lond.* B, 270 : 313-321.

PEGG G.G**,** SINCLAIR B., BRISKEY L.AND WILLIAM J. ASPDEN. 2005. MtDNA barcode identification of fish larvae in the southern Great Barrier Reef, Australia. *Recent Advances in the study of fish eggs and larvae M.P. Olivar and J.J. Govoni.*

PERSIS,M. 2009.Thesis on Studies On The Diversity Of Marine Fishes Of Kakinada Based On Cytochrome C Oxidase Sub Unit-I (CoI) Sequence Variation, Submitted to Andhra University, Visakhapatnam, Andhra Pradesh.

RUZZANTE, D.E. 1996. General Methods for isolation of DNA by phenol/chloroform extraction and ethanol precipitation. Dept. of chemistry and biochemistry. University of Oklahoma, Norman, Oklahoma.

SANGER,F., S. NICKLEN, AND A. R. COULSON, 1977 DNA sequencing with chain terminating inhibitors, Proc Natl Acad Sci U S A. December; 74(12): 5463–5467

TAUTZ , D. ARCTANDER, P. MINELLI, A. THOMAS, R.H AND VOLGER, A.P 2003 : A plea for DNA taxonomy. *Trends Ecol. Evol*: 18:70-74.

WARD R.D., ZEMLAK T.S., INNES B.H., LAST P.R., HEBERT P.D.N., 2005. DNA barcoding Australia's fish species. *Philos. Trans. R. Soc. Lond.* B 360: 1847–1857.

WOESE, C.R AND G. E . FOX. 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proc. Natl. Acad. Sci., USA 74: 5088 – 5090.