Evolutionary history and phylogenetic relationship between *Auxis thazard* and *Auxis rochei* inferred from *COI* sequences of mtDNA

Girish Kumar* and Swaraj Priyaranjan Kunal

Biological Oceanography Division (BOD),
National Institute of Oceanography (NIO),
Dona Paula, Goa 403004, India
E-mail: girishkumar.nio@gmail.com
E-mail: swar.mbt@gmail.com
*Corresponding author

S.K. Shyama

Department of Zoology,
Goa University,
Taleigao Plateau, Goa 403206, India
E-mail: skshyama@gmail.com

Abstract: Tunas of the genus *Auxis* are cosmopolitan species and the smallest members of the tribe Thunnini, the true tunas. In the present study, *COI* sequences of mtDNA were employed to examine the evolutionary history and phylogenetic relationship between *A. thazard* and *A. rochei*. A total of 29 *COI* sequences were retrieved from NCBI. Historic demographic analyses of sequence data showed that *A. thazard* has undergone sudden population expansion in the past while population size of *A. rochei* has been remain constant for long period. Non-significant value of Tajimas's *D* ($P = 0.22400$) and Fu's *FS* ($P = 0.21400$) test fail to reject the null hypothesis of neutral evolution for *A. rochei*. Phylogenetic analyses of nucleotide sequences demonstrated separate clusters for both species and are strongly supported by 98% bootstrap value. The results of the present study suggest the recent founding of *A. thazard* in world ocean while *A. rochei* represents the ancestral species.

Keywords: tuna; phylogenetic; evolutionary history; population expansion; *COI* sequence; mtDNA; diagnostic markers; *Auxis thazard*; *Auxis rochei*.

Reference to this paper should be made as follows: Kumar, G., Kunal, S.P. and Shyama, S.K. (2013) 'Evolutionary history and phylogenetic relationship between *Auxis thazard* and *Auxis rochei* inferred from *COI* sequences of mtDNA', *Int. J. Bioinformatics Research and Applications*, Vol. 9, No. 6, pp.604–613.

Biographical notes: Girish Kumar is PhD scholar at National Institute of Oceanography, Goa (India). He received his masters in Computer Science and Genetics from Maharishi Dayanand University, Rohtak (Haryana). His research interest is in area of population genetics and bioinformatics.

Swaraj Priyaranjan Kunal received his masters in Marine Biotechnology from Goa University. He is currently pursuing his PhD from National Institute of Oceanography, Goa (India). His main fields of interests are population genetics and bioinformatics.

S.K. Shyama is Associate Professor in Department of Zoology, Goa University. He received his PhD from Mangalore University. His main areas of research are genetic toxicology and cytogenetics. He is PhD co-guide of Girish Kumar.

1 Introduction

Tuna is an important commercial fish in the family Scombridae, spending their entire lives relatively near the surface of tropical, subtropical and temperate seas and oceans (Collette and Nauen, 1983). The word tuna refers to any of the 14 species of the tribe Thunnini in 4 genera: *Thunnus* (8 species); *Katsuwonus* (1 species); *Euthynnus* (3 species); *Auxis* (2 species).

Tunas of the genus *Auxis* are currently recognised as two distinct species, the narrow-corseleted *Auxis thazard* (Lacepe'de 1800) and the wide-corseleted *Auxis rochei* (Risso 1810) Collette and Aadland, 1996; Collette et al., 2001). Morphologically, they are differentiated primarily by the width of the corselet under the origin of the second dorsal fin and by the anterior extent of the dorsal scale-less area above the pectoral fin. In *A. thazard*, the corselet has five or fewer scales under the second dorsal fin, and the dorsal scale-less area extends anterior to the tip of the pectoral fin. On the contrary, *A. rochei* has six or more scales and the dorsal scale-less area does not reach the tip of the pectoral fin (Collette and Aadland, 1996).

However, morphological identification at early life stages is problematic because they are too small to have developed distinguishing morphological characteristics. Proper specific identification is essential for early life history studies. Molecular markers can provide a means for positive identification when morphological identification is uncertain or impossible (Morgan, 1975; Graves et al., 1988; Bartlett and Davidson, 1991; McDowell and Graves, 2002; Hyde et al., 2005; Perez et al., 2005). Techniques such as polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) analysis have been used to identify species of the scombrid tribes Thunnini and Sardini (Chow et al., 2003) as well as the species of the genus Thunnus (Chow and Inoue, 1993; Takeyama et al., 2001). In addition, sequencing of a mitochondrial gene region has been used to identify Thunnus species (Bartlett and Davidson, 1991; Ram et al., 1996; Quintero et al., 1998; Terol et al., 2002).

Mitochondrial DNA (mtDNA) is commonly used in population genetic surveys and molecular phylogenetic studies due to its high abundance in the cell, high mutation rate, and maternal inheritance (Curole and Kocher, 1999). A considerable progress in the sequencing of complete mtDNA genomes (mitogenomes) has been observed during past couple of years, making it useful genetic tools in resolving persistent controversies over higher-level relationships of teleosts (Inoue et al., 2001; Miya et al., 2001; Lavoue et al., 2005). MtDNA has also been used to study the evolutionary history in many tuna species (Ely et al., 2002; Chiang et al., 2006; Boustany et al., 2008). Although whole mtDNA

sequence is available for both *A. thazard* and *A. rochei* (Catanese et al., 2007), no attempt has been made to study the evolutionary history of two species. Present study provides the first attempt to study the historic demography of two species based on cytochrome oxidase subunit I (*COI*) sequences of mtDNA.

In present study, *COI* sequences have been employed as diagnostic DNA markers in *Auxis spp*. identification and to infer phylogenetic relationship and evolutionary history of the two species in world ocean.

2 Material and methods

2.1 Data source

A total of 29 *COI* sequences, *A. thazard* (18) and *A. rochei* (11) were retrieved from National Centre for Biotechnology Information (NCBI). Sequences of both the species were aligned separately as well as together using the ClustalW algorithm (Thompson et al., 1994) as implemented in MEGA5 (Tamura et al., 2011).

2.2 Nucleotide sequence analyses

The aligned sequences were imported into software program DnaSP 4.0 (Rozas et al., 2003) to calculate number of haplotypes and polymorphic sites. In addition, conserved DNA sequences were obtained by invoking the conserved DNA region option of DnaSP. Nucleotide composition of sequence data was calculated using program MEGA5 (Tamura et al., 2011). Nucleotide and haplotype diversity, molecular diversity indices, such as transitions, transversions, substitutions, and indels were obtained by software program Arlequin 3.11 (Excoffier et al., 2005).

2.3 Evolutionary divergence analyses

The estimates of evolutionary divergence were inferred from the sequence data by calculating the overall genetic distance between and within species. Analyses were conducted using the Maximum Composite Likelihood model (Tamura et al., 2004). For the analyses, all positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in software program MEGA5 (Tamura et al., 2011).

2.4 Evolutionary history

Arlequin 3.11 (Excoffier et al., 2005) was used to calculate the historic demographic parameters θ_0 (population before expansion), θ_1 (population after expansion) and τ (relative time since population expansion). Tau (τ) value can be transformed to estimate the actual time (T) since population expansion using formula $T = \tau/2\mu$ where μ is the mutation rate per site per generation. In the present study, the mutation rate of 3.6×10^{-8} mutations per site and year was applied for the *COI* sequences as this rate has been reported for the mtDNA sequences in teleosts (Donaldson and Wilson, 1999). Tajima's *D* statistical test (Tajima, 1989) and *FS* test of Fu (Fu, 1997) were carried out to examine

whether two species are at genetic equilibrium. Furthermore, Harpending's raggedness index (*Hri*) (Harpending, 1994) and sum of squared deviations (*SSD*) was calculated using Arlequin to test whether the sequence data deviate significantly from the expectations of population expansion model.

2.5 Phylogenetic relationship

Phylogenetic relationships among *COI* sequences were assessed using the neighbour-joining tree (Saitou and Nei, 1987) based on the Kimura two-parameter model (Kimura, 1980) using program MEGA5 (Tamura et al., 2011). All positions containing alignment gaps and missing data were eliminated by invoking the pairwise-deletion option for indels. The robustness of statistical support for the tree branch was determined by 1000 bootstrap replicates (Felsenstein, 1985). Only nodes with bootstrap support of greater than 50% are shown in the final tree.

3 Results and discussion

3.1 COI sequence characteristics

Final data set of aligned *COI* sequences after excluding the gap and missing data contained 455 sites for *A. thazard* while 419 sites were present in *A. rochei* (Table 1). There were no insertions or deletions of nucleotides in the alignment of *COI* sequences from the two species. Overall 35 variable sites (21, parsimony informative), constituting 18 haplotypes were detected among *COI* sequences (Table 1). Multiple sequence alignments of both the data sets as well as overall polymorphic sites from the aligned *COI* sequence were presented in supplementary material. Nucleotide sequence variation within species was low (*A. thazard* = 0.006, *A. rochei* = 0.009) whereas interspecific sequence variation was 0.021. The mean nucleotide composition of the sequences in both species was almost identical (*A. thazard*: A, 23.8%; C, 28.7%; G, 18.7%; and T, 28.8%; *A. rochei*: A, 24%; C, 27.8%; G, 18.7%; and T, 29.5%). The observed transitions outnumbered transversions in both the species (*A. thazard* and *A. rochei*) by a mean ratio of 2.60 and 3.33 respectively (Table 2). A stretch of 84bp conserved DNA sequences from *A. thazard* and 89bp from *A. rochei* were obtained from the sequence analysis (Table 3). These conserved sequences can be used as species specific diagnostic marker. Both nucleotide and haplotype diversity was high in *A. rochei* as compared to *A. thazard* (Table 2). Nucleotide and haplotype diversity are generally higher in older populations. Thus, the present study indicates that *A. rochei* is ancestral species of genus Auxis.

Table 1 Sequence polymorphism in *A. thazard* and *A. rochei*: number of sites excluding gap and missing data (*n*); monomorphic sites (*m*); variable sites (*v*); singleton sites (*s*); parsimony informative sites (*pi*)

Species	*(n)*	*(m)*	*(v)*	*(s)*	*(pi)*
A. thazard	455	436	19	17	2
A. rochei	419	406	13	8	5
Total	419	384	35	14	21

Table 2 Estimates of genetic diversity between *COI* sequences of *A. thazard* and *A. rochei*: number of sequences (*n*); number of haplotypes (*nh*); haplotype diversity (*h*); and nucleotide diversity (π); number of transitions (*ti*); number of transversions (*tv*); number of substitutions (*sbs*)

Species	*(n)*	*(nh)*	*(h)*	*(π)*	*(ti)*	*(tv)*	*(sbs)*
A. thazard	18	11	0.8562	0.08288	13	5	18
A. rochei	11	7	0.8727	0.13816	10	3	13

Table 3 Conserved *COI* sequences in *A. thazard* and *A. rochei*

Species	*Position*	*Conserved sequences*
A. thazard	102 to 185	ACCAAATCTACAATGTAATCGTTACGGCCCATGCCTTCGTAA TGATTTTCTTTATAGTAATGCCAATTATGATTGGAGGGTTCG
A. rochei	542 to 630	GCTGTCCTTCTCCTTCTATCACTCCCAGTTCTTGCCGCTGGCATT ACAATGCTCCTAACAGACCGAAACCTAAATACAACCTTCTTCGA

3.2 Evolutionary history

The observed mismatch distributions of *A. thazard* for *COI* sequences were unimodal (Figure 1). Tajima's *D* test of selective neutrality and Fu's *FS* test was highly negative and significant (Table 4), suggesting a sudden population expansion, which is supported by the non-significant sum of squared deviation ($P = 0.3873$) and Harpending's raggedness index ($P = 0.6453$). Rapid population expansion was also supported by large differences in θ_0 and θ_1 of *A. thazard* (Table 4). In contrast, mismatch analyses for *A. rochei* was characterised by a multimodal pattern (Figure 2), which is not congruent with the sudden growth expansion model. Both Tajima's *D* and Fu's *FS* tests failed to reject neutrality for this species (Table 4) suggesting that the effective population size of *A. rochei* has been large and stable for a long period. The tau (τ) value of sequence data for *A. thazard* and *A. rochei* was estimated to be 2.5098 and 9.3848 (95% confidence interval) respectively. Following the equation $T = \tau/2\mu$ and mutation rate 3.6×10^{-8} per site and year, it was estimated that the time since the expansion occurred was approximately 311,085 years ago for *A. rochei* whereas corresponding to $\tau = 2.5098$, the estimated time for population expansion for *A. thazard* was 76,612 years ago, which reflects the recent founding of *A. thazard* in world ocean.

Table 4 Demographic parameters of *A. thazard* and *A. rochei* based on mtDNA *COI* sequence data. Mismatch distribution parameters τ, θ_0, θ_1, Tajima's *D* test and Fu's *FS* values, Harpending's Raggedness index (*Hri*), and sum of squared differences (*SDD*)

Species	*(τ)*	*(θ_0)*	*(θ_1)*	*Tajima's D*	*Fu's FS*	*(Hri)*	*(SSD)*
A. thazard	2.5098	0.27070	99,999	–2.1789**	–5.9464***	0.1038	0.0161
A. rochei	9.3848	1.07754	99,999	–0.7909	–1.1524	0.0783	0.0295

Notes: *, **, *** Significant at $P < 0.05$, $P < 0.01$, and $P < 0.001$ respectively.

Figure 1 Mismatch distribution based on mtDNA *COI* sequences of *A. thazard*. Exp, expected distribution; Obs, observed distribution under a model of population expansion

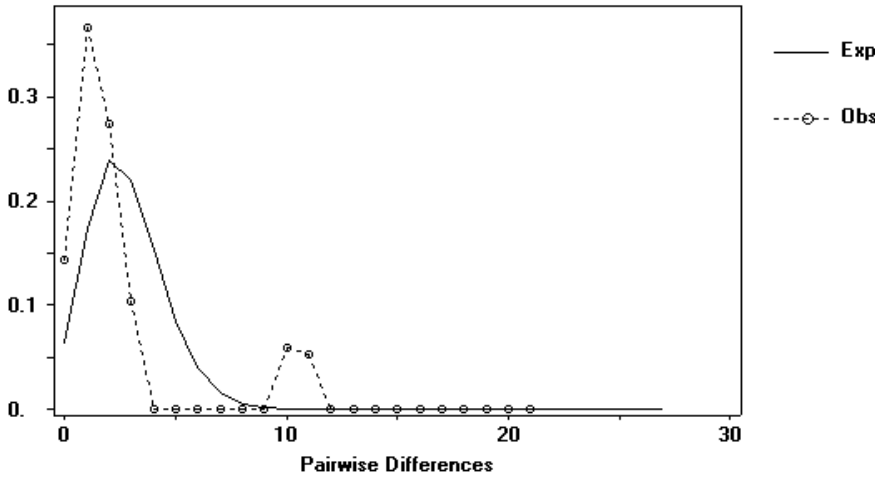
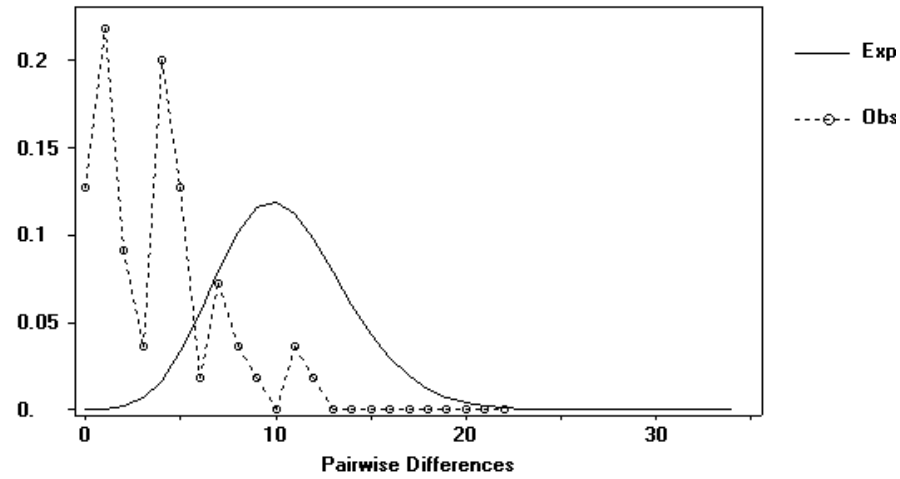


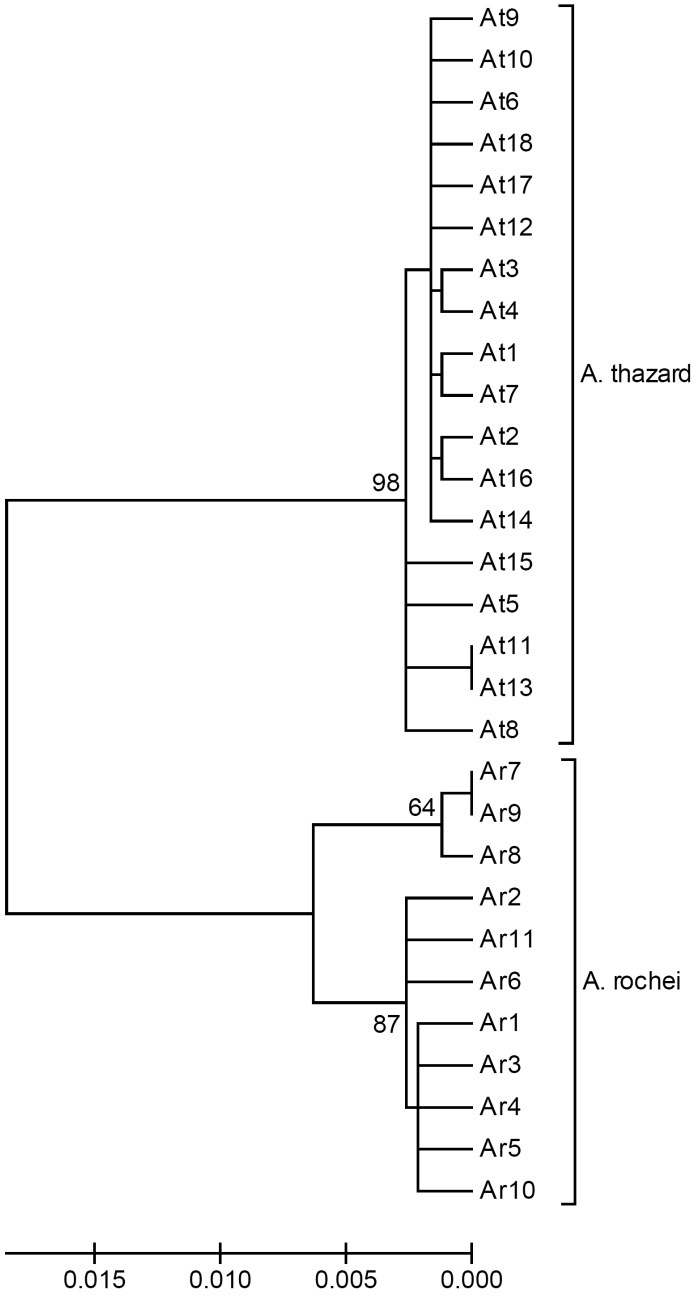
Figure 2 Mismatch distribution based on mtDNA *COI* sequences of *A. rochei*. Exp, expected distribution; Obs, observed distribution under a model of population expansion



3.3 Phylogenetic relationship

Phylogenetic analysis of nucleotide sequences demonstrated that *COI* sequences can be grouped into two distinct clusters (Figure 3). Cluster one represent the *A. thazard* while second cluster belongs to *A. rochei*. The two clusters are strongly supported by 98% bootstrap value. *A. rochei* represents the ancestral species and overall genetic distance between two species is 0.021.

Figure 3 Neighbour-Joining tree of mtDNA *COI* sequences of *A. thazard* and *A. rochei*. Bootstrap supports of >50% in 1000 replicates are shown



4 Conclusion

In conclusion, present study provides the first evidence of phylogenetic relationship and evolutionary history of *A. thazard* and *A. rochei* based on *COI* sequences of mtDNA. The results of this study demonstrate that *A. rochei* represent the ancestral species of genus Auxis. In addition, conserved DNA sequences obtained from sequence information of *COI* allowed the generation of species specific diagnostic marker for identification of *A. thazard* and *A. rochei*.

References

Bartlett, S.E. and Davidson, W.S. (1991) 'Identification of Thunnus tuna species by the polymerase chain reaction and direct sequence analysis of their mitochondrial cytochrome b genes', *Canadian Journal of Fisheries and Aquatic Sciences*, Vol. 48, pp.309–317.

Boustany, A.M., Reeb, C.A. and Block, B.A. (2008) 'Mitochondrial DNA and electronic tracking reveal population structure of Atlantic bluefin tuna (*Thunnus thynnus*)', *Marine Biology*, Vol. 156, pp.13–24.

Catanese, G., Infante, C. and Manchado, M. (2007) 'Complete mitochondrial DNA sequences of the frigate tuna *Auxis thazard* and the bullet tuna *Auxis rochei*', *DNA Sequence*, Vol. 1, pp.1–8.

Chiang, H-C., Hsu, C-C., Lin, H-D., Ma, G.C., Chiang, T-Y. and Yang, H-Y. (2006) 'Population structure of bigeye tuna (*Thunnus obesus*) in South China Sea, Philippine Sea and Western Pacific Ocean inferred from mitochondrial DNA', *Fisheries Research*, Vol. 79, pp.219–225.

Chow, S. and Inoue, S. (1993) 'Intra- and inter-specific restriction fragment length polymorphism in mitochondrial genes of Thunnus tuna species', *Bulletin of National Research Institute of Fisheries Science*, Vol. 30, pp.207–225.

Chow, S., Nohara, K., Tanabe, R., Itoh, T., Tsuji, S., Nishikawa, Y., Uyeyanagi, S. and Uchikawa, K. (2003) 'Genetic and morphological identification of larval and small juvenile tunas (Pisces: Scombridae) caught by a mid-water trawl in the western Pacific', *Bulletin of Fisheries Research Agency*, Vol. 8, pp.1–14.

Collette, B.B. and Aadland, C.R. (1996) 'Revision of the frigate tunas (Scombridae, Auxis), with descriptions of two new subspecies from the eastern Pacific', *Fishery Bulletin*, Vol. 94, pp.423–441.

Collette, B.B., Reeb, C. and Block, B. (2001) 'Systematics of the tunas and mackerels (Scombridae)', *Fish Physiology*, Vol. 19, pp.1–33.

Collette, 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 Fisheries Synopsis*, No. 125, pp.1–137.

Curole, J.P. and Kocher, T.D. (1999) 'Mitogenomics: digging deeper with complete mitochondrial genomes', *Trends in Ecology and Evolution*, Vol. 14, pp.394–398.

Donaldson, K.A. and Wilson, R.R. (1999) 'Amphi-Panamaic geminates of snook (Percoidei: Centropomidae) provide a calibration of the divergence rates in the mitochondrial DNA control region of fishes', *Molecular Phylogenetics and Evolution*, Vol. 13, pp.208–213.

Ely, B., Stoner, D.S., Alvarado Bremer, J.R., Dean, J.M., Addis, P., Cau, A., Thelen, E.J., Jones, W.J., Black, D.E., Smith, L., Scott, K., Naseri, I. and Quattro, J. (2002) 'Analyses of nuclear ldhA gene and mtDNA control region sequences of Atlantic northern bluefin tuna populations', *Marine Biotechnology*, Vol. 4, pp.583–588.

Excoffier, L., Laval, G. and Schneider, S. (2005) 'Arlequin version 3.11: an integrated software package for population genetics data analysis', *Evolutionary Bioinformatics*, Vol. 1, pp.47–50.

Felsenstein, J. (1985) 'Confidence limits on phylogenies: an approach using the bootstrap', *Evolution*, Vol. 39, pp.783–791.

Fu, Y.X. (1997) 'Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection', *Genetics*, Vol. 147, pp.915–925.

Graves, J.E., Simovich, M.A. and Schaefer, K.M. (1988) 'Electrophoretic identification of early juvenile yellowfin tuna, *Thunnus albacares*', *Fishery Bulletin*, Vol. 86, pp.835–838.

Harpending, R.C. (1994) 'Signature of ancient population growth in a low resolution mitochondrial DNA mismatch distribution', *Human Biology*, Vol. 66, pp.591–600.

Hyde, J.R., Lynn, E., Humphreys, R. Jr., Musyl, M., West, A.P. and Vetter, R. (2005) 'Shipboard identification of fish eggs and larvae by multiplex PCR, and description of fertilized eggs of blue marlin, shortbill spearfish, and wahoo', *Marine Ecology Progress Series*, Vol. 286, pp.269–277.

Inoue, J.G., Miya, M., Tsukamoto, K. and Nishida, M. (2001) 'A mitogenomic perspective on the basal teleostean phylogeny: resolving higher-level relationships with longer DNA sequences', *Molecular Phylogenetics and Evolution*, Vol. 20, pp.275–285.

Kimura, M. (1980) 'A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences', *Journal of Molecular Evolution*, Vol. 16, pp.111–120.

Lavoue, S., Miya, M., Inoue, J.G., Saitoh, K., Ishiguro, N.B. and Nishida, M. (2005) 'Molecular systematics of the gonorynchiform fishes (Teleostei) based on whole mitogenome sequences: implications for higher-level relationships within the Otocephala', *Molecular Phylogenetics and Evolution*, Vol. 37, pp.165–177.

McDowell, J.R. and Graves, J.E. (2002) 'Nuclear and mitochondrial DNA markers for specific identification of istiophorid and xiphiid billfishes', *Fishery Bulletin*, Vol. 100, pp.537–544.

Miya, M., Kawaguchi, A. and Nishida, M. (2001) 'Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences', *Molecular Biology and Evolution*, Vol. 18, pp.1993–2009.

Morgan, R.P. (1975) 'Distinguishing larval white perch and striped bass by electrophoresis', *Chesapeake Science*, Vol. 16, pp.68–70.

Perez, J., Alvarez, P., Martinez, J.L. and Garcia-Vazquez, E. (2005) 'Genetic identification of hake and megrim eggs in formaldehyde-fixed plankton samples', *ICES Journal of Marine Science*, Vol. 62, pp.908–914.

Quintero, J., Sotelo, C.G., Rehbein, H., Pryde, S.E., Medina, I., Perez-Marin, R.I., Rey-Mendez, M. and Mackie, I.M. (1998) 'Use of mtDNA direct polymerase chain reaction (PCR) sequencing and PCR-restriction fragment length polymorphism methodologies in species identification of canned tuna', *Journal of Agriculture and Food Chemistry*, Vol. 46, pp.1662–1669.

Ram, J.L., Ram, M.L. and Baidoun, F.F. (1996) 'Authentication of canned tuna and bonito by sequence and restriction site analysis of polymerase chain reaction products of mitochondrial DNA', *Journal of Agriculture and Food Chemistry*, Vol. 44, pp.2460–2467.

Rozas, J., Sánchez-DelBarrio, J.C. and Messeguer-Rozas, R. (2003) 'DnaSP, DNA polymorphism analyses by the coalescent and other methods', *Bioinformatics*, Vol. 19, pp.2496–2497.

Saitou, N. and Nei, M. (1987) 'The neighbor-joining method: a new method for reconstructing phylogenetic trees', *Molecular Biology and Evolution*, Vol. 4, pp.406–425.

Tajima, F. (1989) 'Statistical method for testing the neutral mutation hypothesis by DNA polymorphism', *Genetics*, Vol. 10, pp.512–526.

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', *Journal of Fish Biology*, Vol. 58, pp.1646–1657.

Tamura, K., Nei, M. and Kumar, S. (2004) 'Prospects for inferring very large phylogenies by using the neighbor-joining method', *Proceedings of the National Academy of Sciences (USA)*, Vol. 101, pp.11030–11035.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) 'MEGA5: molecular evolutionary genetics analysis using likelihood, distance, and parsimony methods', *Molecular Biology and Evolution*, Vol. 28, No. 10, pp.2731–2739.

Terol, J., Mascarell, R., Fernandez-Pedrosa, V. and Perez-Alonso, M. (2002) 'Statistical validation of the identification of tuna species: bootstrap analysis of mitochondrial DNA sequences', *Journal of Agriculture and Food Chemistry*, Vol. 50, pp.963–969.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) 'CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice', *Nucleic Acids Research*, Vol. 22, pp.4673–4680.