Independent tuna length frequency and genotypic data sets suggest multiple breeding units in the Indian Ocean: Are the data correlated?

Sudath T. Dammannagoda[1], Shyama C. Ratnasiri[2], David A. Hurwood[1], and Peter B. Mather[1]

1 School of Earth, Environmental and Biological Sciences, Faculty of Science and Engineering, Queensland University of Technology, GPO Box2434, 2 George Street, Brisbane Q4001, Australia.

2 Department of Accounting, Finance and Economics, Griffith Business School, Griffith University, Southport Q4222, Australia.

Tuna are large marine, pelagic fish that are widely distributed across the world's oceans. The general perception about large species of tuna has been that they consist of homogeneous stocks across large geographical whole ocean basins which are supported by reported examples of trans-oceanic movement. This perception and a general lack of detailed knowledge about tuna stocks has led to many tuna fisheries being managed at ocean-wide scale as single stocks. Recent studies on some tuna species that has employed otolith, electronic tagging and genetic approaches however, have found evidence for heterogeneous stocks and multiple management units.

Length frequency analysis has been applied for a long time for fish stock assessment (Pauly 1987) although only relatively few studies have used length frequency analysis as a direct method for studying wild stock structure (e.g.: Nishida 1992; Morita & Koto 1970). Even among these cases, studies of 'length frequency modes' for stock identification purposes are rare.

A Length frequency distribution analysis (LFDs) of 313 million YFT individuals harvested from the Indian Ocean that covered the period; 1952 to 2008, shows that YFT captured in the Indian Ocean consistently express two distinct length modes (Fig 1). Fine scale analysis of LFDs at smaller geographical scales (5^0X5^0 grids) consistently showed the presence of both length modes across the Indian Ocean, with individuals from 'small length' length mode being more common above the equator, while in the southern hemisphere the 'large length' mode was dominant.
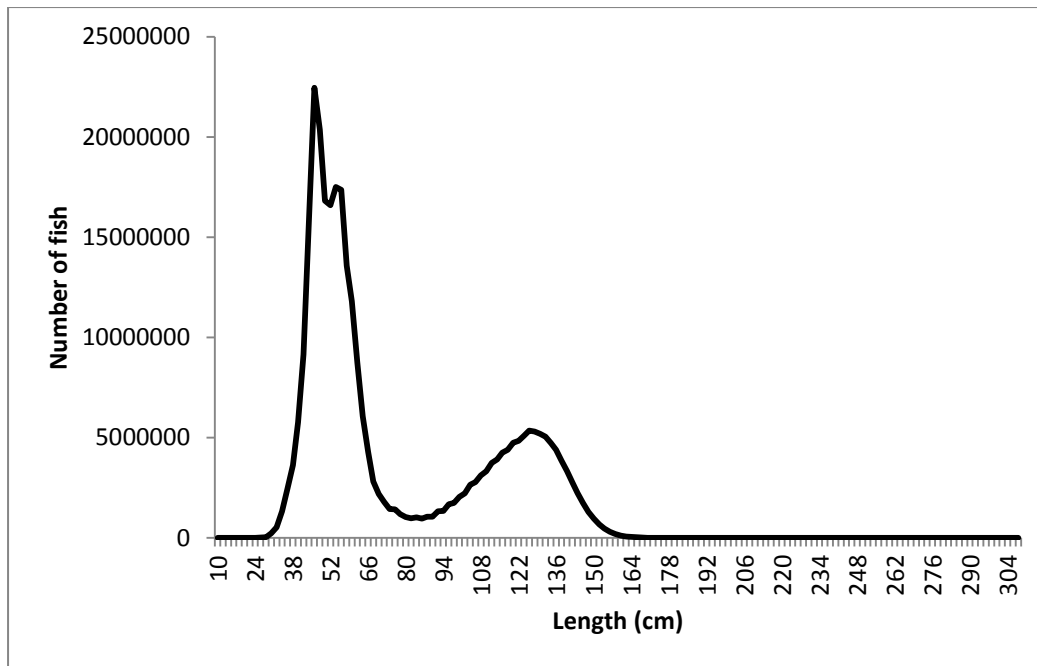
Figure 1. YFT LFD data from the entire Indian Ocean. This includes all locations and all gear types from 1952 to 2008 (n = 313,016,441)

Some fisheries scientists have interpreted this LFD pattern that is typical of purse seine (two length modes) data as resulting from a combination of gear selectivity and relative YFT vulnerability to this fishing method. According to this interpretation, they have argued that the small length mode caught by purse seine fishermen were juvenile YFT, while large individuals were adults that were equally vulnerable (Ariz et al. 2002). They have argued that low vulnerability of individuals in the size range between 60cm to 100cm to purse seine fishing must be related to differences in behavioural patterns and/or differential movement patterns.

An alternative hypothesis for this LFD pattern is that these two length modes could represent two discrete breeding units. Biologically, a discrete reproductive population of organisms will usually conform to a normal distribution for both qualitative and quantitative external phenotypic traits, and body length will reflect this trend. This is because selection overtime that matches phenotype to environmental parameters will reach a point where stabilising selection reduces genetic variation around the mean and favours mature individuals with intermediate and optimum size phenotypes. Where such an analysis shows however, that multiple modes are present in the sampled population, the interpretation of

this pattern (e.g. for length) can be that different modes may represent different 'biological populations', given that large representative sample sizes of each have been examined (Beaumont et al. 2010). In addition, according to the central limit theorem, statistical comparisons have shown that if more than a single mode exists within a single distribution for a particular trait (e.g.: length), the two modes probably represent different 'biological populations', given that very large representative sample sizes were taken (Buchanan-Wollaston & Hodgson, 1929; Harding 1949).

A recent genetic analysis of YFT diversity confined to the north western Indian Ocean also suggested that there were two non-interbreeding YFT groups in the study area. In this study, 394 YFT individuals from six fishing grounds around Sri Lanka and from a single site in the Maldive Islands (Fig 2) were screened for variation in a mitochondrial DNA (mtDNA) ATP*ase* gene fragment and at three (nDNA) microsatellite loci. Details of the genetic methodologies and data analysis employed have been described previously in Dammannagoda et al. 2008. Specifically Dammannagoda et al. (2008) used analysis of a multi-locus microsatellite data set from 394 YFT and the statistical software package STRUCTURE Version 2.3.3 (Pritchard et al. 2000) to test for presence of multiple YFT breeding units/stocks in this region. STRUCTURE uses a model-based full Bayesian Markov Chain Monte Carlo (MCMC) approach that clusters individuals to minimise Hardy-Weinberg disequilibrium and gametic phase disequilibrium between loci within groups.

Multiple STRUCTURE runs were performed to evaluate the reliability of the results, and the estimated number of populations was then determined from posterior probabilities of K estimated, assuming that they had originated from 1 to 7 populations. Individuals were assumed to have been correctly assigned to a population when their q-value (i.e. their posterior probability of belonging to an original population) was at least 80% for the population (Pritchard et al. 2000). STRUCTURE analysis of these data show that YFT individuals separated into two genetic clades (red vs green) (Fig 3). The genetic data suggests therefore that sets of genotypes or non-interbreeding YFT genetic groups were present in the study area.
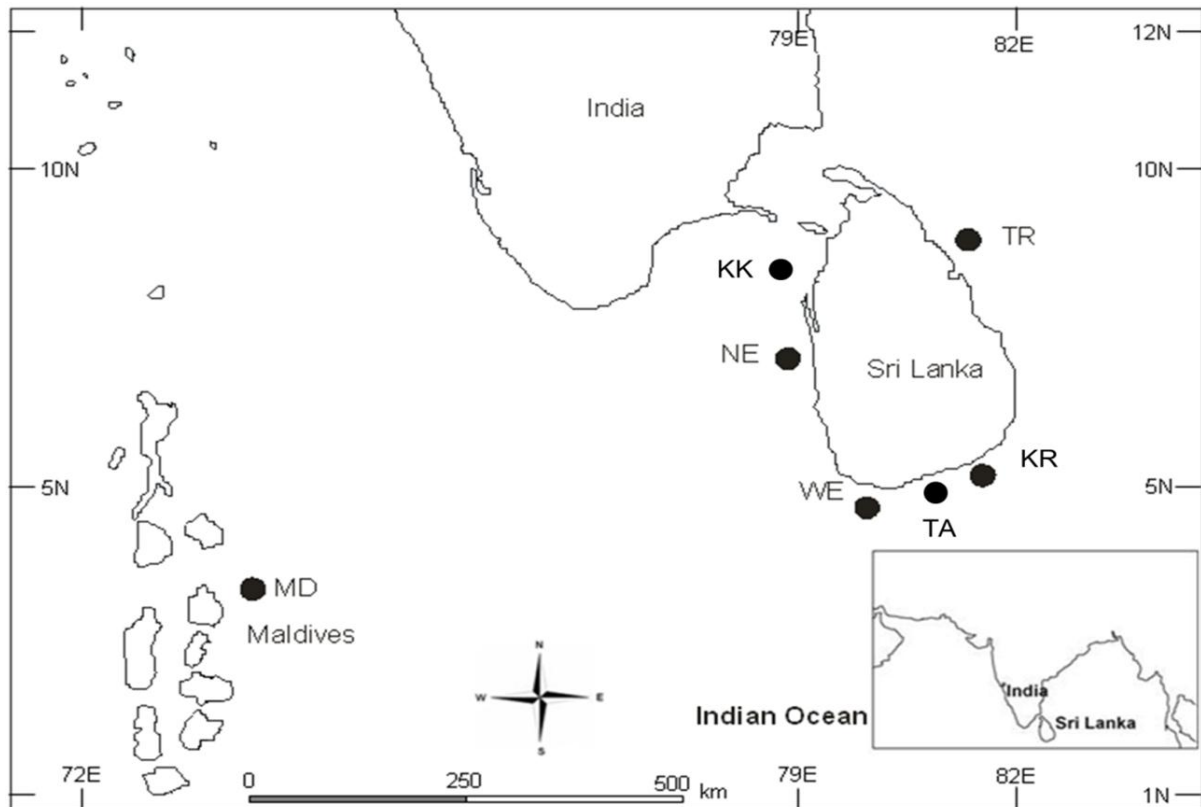
Figure 2. Sampling sites for yellowfin tuna in the north western Indian Ocean (Dammannagoda et al. 2008).
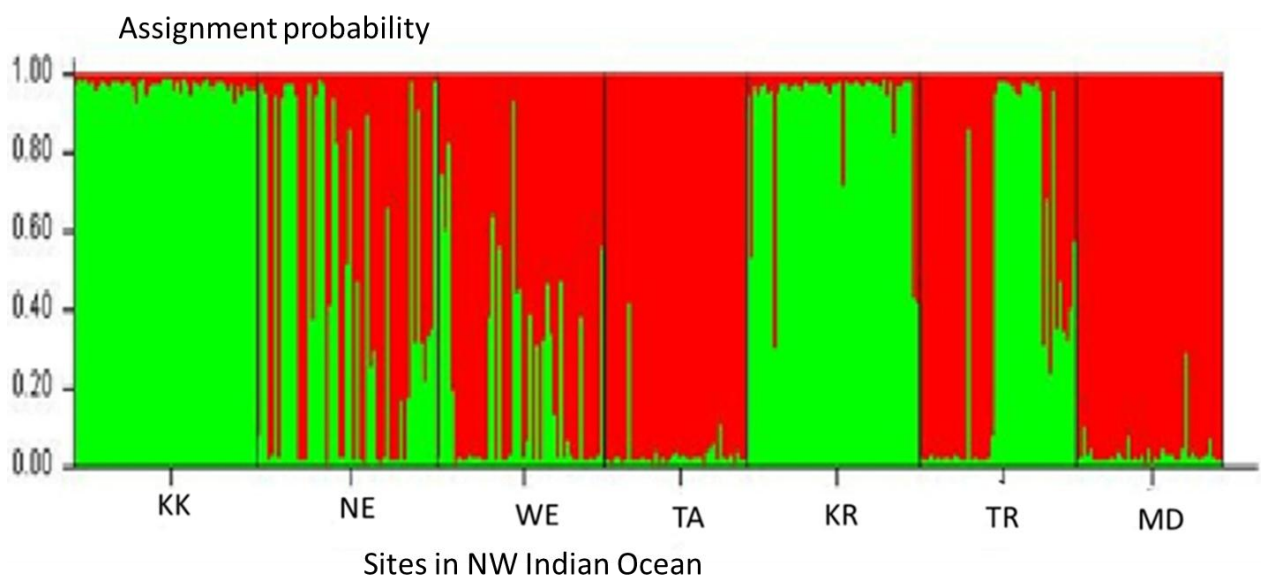

Figure 3. Bayesian clustering analyses (K = 2) for YFT (394 individuals; 3 loci) performed using STRUCTURE.

Of Interest also was that posterior length Frequency Distribution (LFD) of YFT in the same study area in the NW Indian Ocean like the broader data set from the IOTC also showed two length modes (Fig 4).
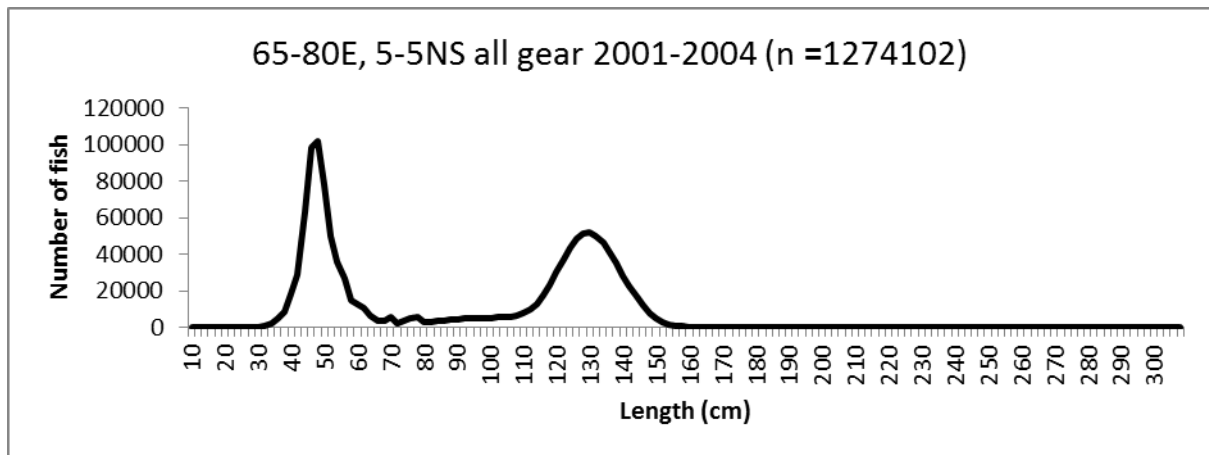

Figure 4. LFD of YFT from the north western Indian Ocean

During sampling of YFT for the genetic study, we observed two size class YFT groups (i.e. referred as large size and small size), and both small size and large size YFT individuals were sexually mature. The small size YFT group therefore cannot simply be juveniles of the large length YFT group. From this pattern, we hypothesised, that the two length modes may represent two non-interbreeding populations of YFT that coexist in NW region of the Indian Ocean.

Thus, two independent data sets, LFD for the whole Indian Ocean (from IOTC) and genetic for the NW Indian Ocean (Dammannagoda et al. 2008) have shown an identical trend. To examine this apparent coincidence we compared evidence from the genetic stock structure study of YFT undertaken in the north western Indian Ocean that suggested two independent genetic groups (Dammannagoda et al. 2008) and where two length modes were also observed with the IOTC data for the whole region. To address this question, representatives of known individual length from each length mode class in the Dammannagoda et al. 2008 were screened genetically using multi-locus genotyping and Bayesian cluster analysis to test whether they contained different sets of genotypes and if the genetics groups correlated with the length groups.

Of the 394 YFT screened in the original study, length data were available for 170 individuals of which 91 individuals exceeded 80cm in length and were categorised as belonging to the large length LFD mode group with the remaining individuals classified as small length mode group. STRUCTURE analysis of these data show that individuals separated into two genetic clades (red vs green) based on their individual sizes (Fig 5). LFD for these individuals are presented in Figure 6, with each individual's genetic cluster indicated in red or green. Some individuals could not be assigned to a genetic cluster with above 80% assignment probability, and therefore, individuals only possessing 70% or higher probability of assignment to a particular clade were selected for inclusion in the length frequency distribution. All YFT below 80cm in length formed one genetic group (in red), while YFT individuals above 80cm belonged to the alternative genetic group (green) except for 6 individuals. If the size relationship with genetic group pattern holds up then some YFT individuals from both genetic groups must be present in the small length mode, because individuals must grow to maturity regardless of which genetic group they belong to. In the large length mode group however, YFT should only be from a single genetic clade. In our study, only 6 individuals (out of a total of 170 individuals screened – 3.53%) that were above 80cm in size were assigned genetically to the small length mode group. This error rate, although very small, could result from genotyping error, length recording error, or may be a natural phenomenon where a few rare individuals that belong to the small size genetic group occasionally can grow to an extreme size. In general however, the data show consistently two YFT length modes that represent two genetically discrete groups at the 96% confidence level. Results of the current study suggest that there is a strong correlation between length mode and identified genetic group (red vs green). The probability of this strong correlation by chance is very low.
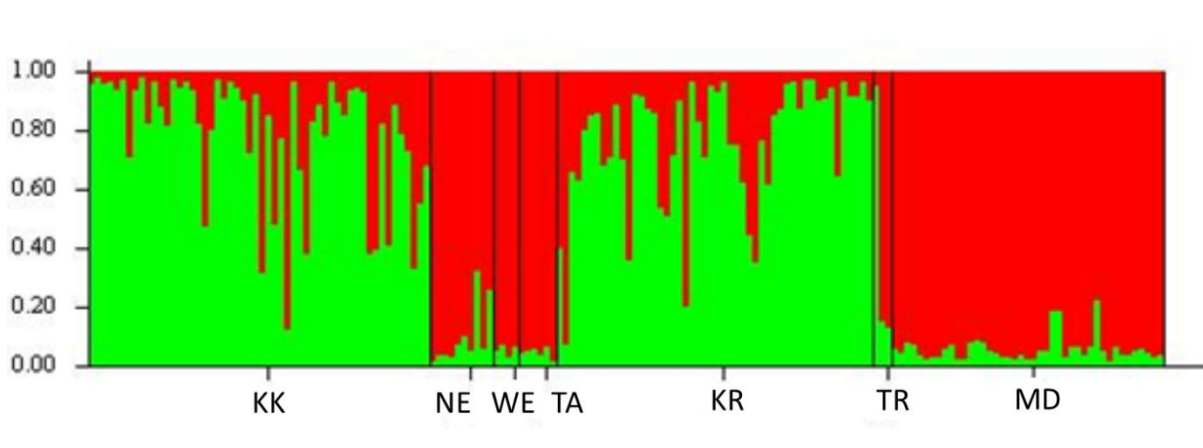
Figure 5. Bayesian clustering analyses (*K* = 2) of 170 YFT of known body length performed using STRUCTURE.
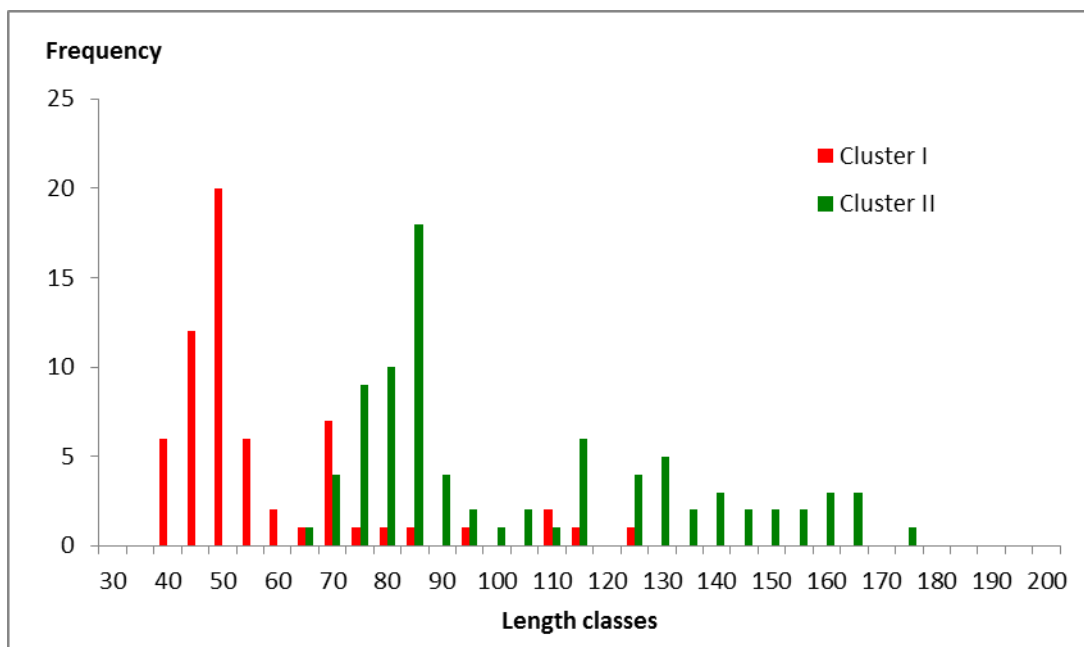

Figure 6. LFD of genetically screened YFT. Red bars represent YFT that belong to the 'small length genetic group', and green bars represent the 'large length genetic group'.

Other scombrid species also show similar bimodal/multimodal length frequency patterns and populations that are structured genetically. For example, a recent genetic study of skipjack tuna (SJT) (*Katsuwonus pelamis*) in the western Indian Ocean identified two divergent, non-interbreeding groups as shown from both mtDNA and nDNA (Dammannagoda et al. 2011), and a length frequency analysis of SJT in the same study area

over the same sampling time period with two length modes (IOTC data base-http://www.iotc.org/English/meetings/wp/wpbcurrent.php).

Thus, two independent data sets of two species that combined LFD, and genetic data at a regional scale (i.e. in the north western Indian Ocean) both shows a strong correlation between length modes and discrete genotype sets. These observations/correlations require more attention and should be extended to determine if a similar genetic correlation is apparent more widely across the Indian Ocean for YFT. Ultimately, this analysis can be used to manage YFT stocks across the Indian Ocean more effectively in the future.

References

Ariz, J., Pallares, P., Delgado de Molina, A., Fonteneau, A., and Santana, J.C. (2002). Analyses of the catches by weight category of yellowfin tuna (*Thunnus albacares*) undertaken by the purse seine fleets in the Indian Ocean from 1991 to 2000. *IOTC Proceedings No. 5 (WPTT02-25*): 375-387.

Beaumont A., Boudry P., Hoare K. (2010). Biotechnology in Fisheries and Aquaculture. Wiley-Blackwell, UK.

Buchanan-Wollaston, H.J. & Hodgson, W.C. (1929). A new method of treating frequency curves in fishery statistics, with some results. Journal Conseil International Pour L'exploration de la mer. 4: 207-225.

Dammannagoda ST, Hurwood DA, Mather PB. (2011).Genetic analysis reveals two stocks of skipjack tuna (*Katsuwonus pelamis*) in the north western Indian Ocean. Canadian journal of Fisheries and Aquatic Sciences. 68 (2). 210-223.

Dammannagoda ST, Hurwood DA, Mather PB (2008). Evidence for fine geographical scale heterogeneity in gene frequencies in yellowfin tuna *(Thunnus albacares*) from the north Indian Ocean around Sri Lanka - Journal of Fisheries Research. 90. 147-157.

Harding, J.P. (1949). The Use of Probability Paper for the Graphical Analysis of Polymodal Frequency Distributions. Journal of the Marine Biological Association of the United Kingdom. 28: 141-153.

Morita Y., and Koto T. (1970). Some considerations on the population structure of yellowfin tuna in the Indian Ocean based on the long line fishery data. Bulletin of the Far Seas Fisheries Research Laboratory. 4: 125-140.

Nishida, T. (1992). Considerations of stock structure of yellowfin tuna (*Thunnus albacares*) in the Indian Ocean based on fishery data. Fisheries Oceanography. 1: 143– 152.

Pauly, D. (1987). A Review of the ELEFAN system for analysis of length frequency data in fish and aquatic invertebrates, p. 7-34. *In* D. Pauly and G.R. Morgan (eds.) Length based methods in fisheries research. ICLARM Conference Proceedings 13. P468. International Centre for Living Aquatic Resources Management, Manila, Philippines, and Kuwait Institute for Scientific Research, Safat, Kuwait.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multi locus genotype data. *Genetics*. 155: 945-959.