Population genetic structure of the white sardine, *Sardinella albella*, in the Persian Gulf and Sea of Oman by analysis of mitochondrial control region

Rahimi P. 1*; Rezvani Gilkolaie S.2; Ghavam Mostafavi P.1; Jamili Sh.2; Rahnema M.3

Received: March 2015 Accepted: March 2016

Abstract
Several studies on the white sardine: *Sardinella albella*, have focused on the identification of stock composition and behavior. In this study population genetic structure and historical demography of *S. albella* along the cost of the Persian Gulf and Sea of Oman were investigated with a 500-bp segment of mt-DNA control region. In total 40 samples were collected from 3 locations: Jask in Sea of Oman, Qeshm in Strait of Hormuz and Lengeh in the Persian Gulf during 2012-2013. 33 haplotype were obtained none of which were presented in all sampling sites. Analysis of molecular variance (AMOVA) indicated low genetic differentiation among regions ($F_{ST}$=0.024, $p<0.05$).The average pair wise differences between regional population were small but significant (0.0158-0.165). Molecular variance explained by differences among three regions was significantly different from zero but the $F_{ST}$ did not show clear phylogeographic isolation. This observation can support the conclusion that *S. albella* has a widespread dispersal potential.

Keywords: *Sardinella albella*, Clupeidae, Sardine Fish, Persian Gulf and Sea of Oman

1-Department of Marine Biology, Graduate School of Marine Science and Technology, Science and Research Branch, Islamic Azad University, Tehran, Iran.
2-Iranian Fisheries Science Research Institute, Agricultural Research, Education and Extension Organization, Tehran, Iran.
3-Department of Basic Science and Medicine, Zanjan Branch, Islamic Azad University, Zanjan, Iran
* Corresponding author's Email: prahimi1975@gmail.com
Introduction
The genetic structure of fish populations is important not only because of fundamental interest in biotic evolution (Tudela et al., 1999) but also for the management of fisheries (Roldan et al., 2000). So the biological and ecological knowledge about their populations can be relevant in devising suitable management strategies.

Sardines have a very important ecological role in marine ecosystems because they are first consumers in marine food chain and they are the main food of valuable species as tuna. So a decrease in their population will decline fishing of these species (Whithead, 1988). There are 10 genera of Clupeidae in the south of Iran. Sardinella albella is one of the most abundant species inhabiting the tropical and subtropical coastal regions in Indo –pacific from east of Africa, Red Sea, Persian Gulf, Sea of Oman and Indian ocean up to Indonesia and new Gionea (Whithead, 1985) (Fig. 1). In Iran Sardinella albella lives in estuaries along the north coast of the Persian Gulf and Sea of Oman. Adults make schools that can be maintained as a dark span during the day but at night schools scatter and migrate to deep areas to feed. In the breeding season, adults migrate to coastal areas from mid April to the end of June. Their eggs and larva are pelagic (Van Zailing et al., 1993).

The life history of Sardinella albella and their passive and active dispersal mechanism have a direct influence on population structure (Liu et al., 2009). Because of the importance of the identification of stock structure in stock assessment and harvest management (Whithead, 1985), in this study we investigated the population of Sardinella labella in 3 sites located along the north coast of the Persian Gulf and Sea of Oman. The mitochondrial control region serves as the origin of replication for the mitochondrial genome and is usually more variable than other coding genes (Avis, 1994; Parker et al., 1998). So it is an ideal marker for assessing genetic structure of recently diverged or closely related species or populations (Avis, 1994; Rand, 2000). We used partial sequence of mitochondrial control region (D-Loop) to test the hypothesis that indicates: there are no significant genetic differences between sampling zones.

Materials and methods
Sampling
The white sardines used in the present study were collected between autumn 2012 and summer 2013 from commercial landing in Sea of Oman (Jask), strait of Hormoz (Qeshm) and the Persian Gulf (Lengeh). Geographic locations and sample sites are given in Fig. 2.

The dorsal fin of each individual was obtained and preserved in 98% ethanol for genomic DNA extraction.
Figure 1: The map of global distribution of *Sardinella albella* (Whithead, 1985).

Figure 2: Sampling sites in the Persian Gulf and Sea of Oman.
DNA extraction, amplification, and sequencing

Genomic DNA was isolated using the phenol-chloroform standard method. The primers reported by Somante et al. 2000 were used to amplify a fragment of the control region (D-Loop) of mtDNA, (Fig. 3).

Polymerase chain reaction (PCR) was performed in a total volume of 25 μL reactions containing 1×PCR buffer, 200 mM dNTPs, 1.5 mM MgCl₂, 0.3 mM of each primer, 0.5 units Taq DNA polymerase (Fermentaze) and 1 μL of DNA. PCR cycling consists of an initial denaturation at 94°C for 5 min, followed by 30 cycles of 45 sec at 95°C, 45 sec at 52°C, and 2 min at 72°C, with a final elongation of 5 min at 72°C. After visualization on a 1% agarose gel, the appropriate PCR products were sent to the micro gen lab for sequencing. The primers used for sequencing were the same as those for PCR amplifications.

Data analyses

The sequences for all individuals were aligned and edited by BioEdit software (Hall, 1999) and Mega5 (Tamura et al., 2011) which uses Clustal w algorithm. Molecular diversity indices such as number of haplotypes, polymorphic sites, transition, transversions, and indels were obtained by using the program ARLEQUIN 3.5 (Excoffier et al., 2011). Population genetic structure was analyzed by using the analysis of molecular variation (AMOVA), which estimates the proportion of genetic variation within and among populations. Genetic differentiation between populations was assessed by comparing the average number of pairwise differences between population, within populations (PiX and PiY) and the corrected average pairwise difference (PiXY-(PiX+PiY)/2) using the program ARLEQUIN 3.5 (Excoffier et al., 2011). The significance of differentiation between pairs of populations was tested by exact test (Raymond and Rousset, 1995) using 10,000 Markov chain steps. Molecular pairwise F̂ST was estimated to evaluate genetic differentiation between pairs of data set. The demographic history was investigated by examine the mismatch distributions (Rogers and Harpending, 1992; Ray et al., 2003; Excoffier, 2004). The parameters of demographic expansion τ, θ₀, θ₁ were estimated. τ index is the time since population expansion expressed as unites of mutational time (Rogers and Harpending, 1992). θ₀ and θ₁ are the mutation parameters before and after population growth. The overall validity of the estimated demographic model is tested by the distribution of a tested statistic (SSD) (the sum of squared differences) between the observed and estimated mismatched distribution. To test whether the D-loop data confirm to neutrality's expectations, the Tajima's D (Tajima, 1989) and Fu's Fs (Fu, 1997) statistics were used. The significance of neutrality was tested with a coalescent simulation program (1000 simulations) as implemented in ARLEQUIN 3.5 (Excoffier et al., 2011).
Results

Genetic diversity
A 500 bp fragment of the control region (D_Loop) of 40 specimens was sequenced. 15 polymorphic sites defined 33 haplotypes among specimens (GenBank accession no. are from KP714734 tile KP714763). This high genetic variability is evident by the large haplotype diversity. No haplotype was presented in all zones (Table 1). Average values of pair wise differences among populations were between 3.25 and 5.01 and corrected average of pair wise difference ranged from 0.051 to 0.1313 (Table 2). These results demonstrated that nucleotide diversity is low and haplotype diversity is high in Lengeh and both nucleotide diversity and haplotype diversity are high in Jask (Table 1). However in Qeshm the haplotype diversity is high and the nucleotide diversity is low.

Geographical structure of genetic variation
The genetic difference was detected in the average number of the pairwise differences (Table 2). The average number of pairwise differences and corrected number of pairwise differences showed minor but significant differentiation between localities (Table 2). The most difference is between Qeshm and Lengeh. AMOVA test ($k_2$ p, $p<0.05$) of genetic diversity showed low variation between populations and high variation within them and fixation index of genetic diversities ($F_{ST}$) between 3 localities was low too (Table 3). Pairwise $F_{ST}$ between Lengeh and Qeshm is low but is more than that between Jask and Qeshm (Table 4).

Patterns of historical demography
None of the mismatched analyses deviated significantly ($p>0.05$) from expectations under the sudden expansion model, and hence analysis of demographic patterns was suitable (Table 5). The mismatch distribution for both Lengeh and Qeshm was unimodal, and closely matched the expected distribution under the sudden expansion modal. But the mismatch distributions for Jask were multimodal and much smoother, suggesting that this population has not always remained stationary (Fig.4). All Tajima’s D and Fu’s $F_s$ values were negative ($p<0.05$) suggesting the effects of populations growth (Table 5).

Figure 3: Structure of control region of mtDNA and the primers have been used from Samonte et al., (2000).
Table 1: Sample size (n), number of haplotype (nh), haplotype diversity (H), nucleotide diversity (π) and standard deviation (SD) for each group.

<table>
<thead>
<tr>
<th>Stocks</th>
<th>N</th>
<th>nh</th>
<th>H± SD</th>
<th>π±SD</th>
<th>Mean number of pair wise differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jask</td>
<td>13</td>
<td>10</td>
<td>0.9615±0.0496</td>
<td>0.01287±0.0073</td>
<td>6.052±3.082</td>
</tr>
<tr>
<td>Qeshm</td>
<td>12</td>
<td>11</td>
<td>0.9848±0.0403</td>
<td>0.00808±0.0043</td>
<td>3.789±2.05</td>
</tr>
<tr>
<td>Lengeh</td>
<td>15</td>
<td>11</td>
<td>0.8952±0.0704</td>
<td>0.00541±0.0034</td>
<td>2.536±1.44</td>
</tr>
</tbody>
</table>

Table 2: Pair wise population differences for mitochondrial control region sequences of *Sardinella albella* (*p*<0.05). Above diagonal: average number of pair wise differences between populations (P{XY}). Diagonal elements: average number of pair wise differences within population (P{X}). Below diagonal: corrected average number of pairwise difference (P{XY}-(P{X}+P{Y})/2).

<table>
<thead>
<tr>
<th></th>
<th>Lengeh</th>
<th>Qeshm</th>
<th>Jask</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lengeh</td>
<td>2.535</td>
<td>3.294</td>
<td>4.347</td>
</tr>
<tr>
<td>Qeshm</td>
<td>0.1313</td>
<td>3.789</td>
<td>5.010</td>
</tr>
<tr>
<td>Jask</td>
<td>0.0513</td>
<td>0.0874</td>
<td>6.055</td>
</tr>
</tbody>
</table>

Table 3: Hierarchical analysis of molecular variance (AMOVA) of mtDNA control region haplotypes of *Sardinella albella* (*k*<sub>2</sub> *p*<0.05).

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>df</th>
<th>Sums of squares</th>
<th>Variance component</th>
<th>Percentage of variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among populations</td>
<td>2</td>
<td>5.345</td>
<td>0.04876 va</td>
<td>2.35</td>
</tr>
<tr>
<td>Within populations</td>
<td>37</td>
<td>74.93</td>
<td>2.02513 vb</td>
<td>97.65</td>
</tr>
<tr>
<td>Total</td>
<td>39</td>
<td>80.275</td>
<td>2.07389</td>
<td></td>
</tr>
</tbody>
</table>

Fixation index: *F*<sub>ST</sub>: 0.024

Table 4: *F*<sub>ST</sub> estimates between geographical populations of *Sardinella albella* based on mitochondrial control region sequences haplotypes (*p*<0.05).

<table>
<thead>
<tr>
<th></th>
<th>Lengeh</th>
<th>Qeshm</th>
<th>Jask</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lengeh</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Qeshm</td>
<td>0.044</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Jask</td>
<td>0.01656</td>
<td>0.0158</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 5: Results of Tajima's D, Fu's Fs, mismatch distribution test (SSD), corresponding *p*-value and the mismatch distribution parameter estimates based on the *Sardinella albella* mtDNA control region in three populations (*p*<0.05).

<table>
<thead>
<tr>
<th>Samples</th>
<th>Tajima's D</th>
<th>Fu's Fs</th>
<th>SSD</th>
<th>p</th>
<th>τ</th>
<th><em>θ</em></th>
<th><em>θ</em>&lt;sub&gt;1&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lengeh</td>
<td>-1.282</td>
<td>0.076</td>
<td>-5.514</td>
<td>0.000</td>
<td>0.0215</td>
<td>0.09</td>
<td>2.5</td>
</tr>
<tr>
<td>Qeshm</td>
<td>-1.282</td>
<td>0.097</td>
<td>-7.736</td>
<td>0.000</td>
<td>0.0111</td>
<td>0.56</td>
<td>3.2</td>
</tr>
<tr>
<td>Jask</td>
<td>-1.197</td>
<td>0.114</td>
<td>-4.898</td>
<td>0.006</td>
<td>0.0106</td>
<td>0.55</td>
<td>3.7</td>
</tr>
</tbody>
</table>
Figure 4: Mismatch distributions of the *Sardinella albella* mtDNA control region in populations of Lengeh, Qeshm and Jask.

**Discussion**

The genetic structure of marine organisms can be attracted by many poorly understood variables such as physical oceanographic factors including currents, tides, and violent storms. The biological factors such as local recruitment, larval transport potential, or reproductive strategy can influence the populations too (Palumby, 1994). Some barriers such as large distances, massive river outflows or unsuitable habitats can separate the distribution of some marine fish populations (Liu *et al.*, 2009). Some studies have provided evidence of long – distance dispersal linking populations at the scale of ocean basins (Bowen *et al.*, 2001) and the others indicate that clear divisions have been found over relatively short geographic distances (Guarnieo *et al.*, 2002; Hoffman *et al.*, 2005; Bernal – Ramirez *et al.*, 2003; Carlin *et al.*, 2003; Rodriguez, 2011). The meaningful divergence of genetic signals can be maintained for widely distributed fishes whose populations might have been subjected to the effects of such complex variables (Ward *et al.*, 1994; Aripin and Showers, 2000; Atarhouch *et al.*, 2006).

**Historical demography and genetic diversity**

Populations of fishes often show high haplotypes but medium to low nucleotide diversities (Vaughan *et al.*, 1992; Grant and Bowen, 1998; Dudgeon *et al.*, 2000; Fauvelot *et al.*, 2003; Chen *et al.*, 2004 :). In 1997 Bowen and Grant investigated the phylogeography of the sardines
(Sardinops spp.) as a widely distributed species in temperate upwelling zones and assessed biogeography models for them. Bowen and Avise (1990) studied the influence of zoogeographic factors and life-history patterns of Atlantic menhaden, Brevoortia tyrannus. In 2004, Bay et al. investigated genetic structure of Chlorurus sordidus and found high genetic diversities in it. The high genetic diversities of these species were ascribed to their long stable evolutionary history (Grant and Bowen, 1998).

In this study the mismatch distribution (Fig.4) for both Lengeh and Qeshm were clearly unimodal and none of the mismatch distributions (Table 5) deviated significantly (p>0.05) from expectations under sudden expansion model (Rogers and Harpending, 1992) or there is a high rate of migration between near demes (Ray et al., 2003; Excoffier, 2004). But the mismatch distribution of Jask population is bimodal and suggested that the population is in equilibrium (Rogers and Harpending, 1992). Although Fu’s value was negative and significantly deviated from a neutral rate, the Tajima’s D was negative but not significant (p>0.05). Population expansion and mutation rate heterogeneity have opposite effects on Tajima’s D statistics (Aris-Brosou and Excoffier, 1996). The negative amount of Tajima’s D and Fu’s Fs is due to genetic expansion of population in the 3 localities. The high amount of θ1 can confirm expansion of genetic diversity. The result of increase in genetic diversity is stability of the population. These results demonstrated that our data fit well with the model of spatial expansions of population.

High haplotype diversity and low nucleotide diversity (Table 1) was found in the hyper variable portion at the 3’ end of the mitochondrial control region among the three localities. According to Grant and Bowen’s (Grant and Bowen, 1998) populations in 3 localities have all experienced rapid population expansion after a period of low effective population size. This criterion enhances the retention of new mutation (Avise et al., 1984). The results were consistent with mismatch distribution analysis and the neutrality test (Table 5). However the high haplotype diversity and low nucleotide diversity in Qeshm may be result in migration of some groups for feeding in the sampling site (Grant and Bowen, 1998).

Population genetic structure
Pelagic marine fishes generally show low levels of genetic diversity among geographic regions because of high dispersal potential during planktonic eggs, larval or adult history stages as a migratory habit and absence of physical barriers between ocean basins or adjacent continental margins (Palumbi, 1994; Grant and Bowen, 1998; Hewitt, 2000). In 2006 Atarhouch et al. investigated populations of Moroccan sardines (Sardinia pilchardus) and find that the genetic diversity can be the result of adaptation of fish to the environment and stability of
populations, number of effective reproductive females and dimension of population. Mangulas et al. (2006) investigated the population structure in Atlantic and Mediterranean anchovy (Engraulis encrasicolus) by studying mitochondrial DNA. The results revealed a mosaic pattern of phylogeographical structures for this species and also diversity among populations became low and that within populations become high. Species characters and environmental conditions influence gene flow between localities and cause differences in phenotypic and genotypic characters. Rodriguez et al. (2011) investigated population structure of Pacific sardine (Sardinops sagax) and found that high ability of migration of adults and the California current system, which disperses eggs and larva, may be due to weak phylogeographic structures in this species. The adults of Sardinella albella have horizontal migrations and prefer to vertically migrate to coastal margins for breeding and their larvae and juveniles are regarded as pelagic. So the high genetic diversity in Sardinella labella populations shows stability because of the creation of large schools. In contrast to the high genetic diversity within populations, low genetic diversity was detected among them (Table 3). These criteria show gene flow between localities. Oceanic and sea currents can effect dispersal of fishes. Among populations, differences might be the result of passive dispersal of eggs and larva and active dispersal of adults or existence of gene flow between populations (Bowen and Grant, 1997; Grant and Bowen, 1998; Hewitte, 2000). There is in current water along north coastal areas of the Persian Gulf from the Sea of Oman (Hunter, 1983; Pous et al., 2004) that can be the cause of passive transports of eggs & larva from the Sea of Oman to the Persian Gulf and the occurrence of gene flow between them. Better environmental conditions in the Sea of Oman, and the monsoons of the Indian Ocean, resulted in differences in morphological characters between animals of the Sea of Oman and the Persian Gulf (Emara, 1990; Alabdessalam, 1995; Cole, 1998; Feron, 1999; Aripin, 2000; Salarpouri et al., 2003; FAO, 2010; Salarpouri et al., 2014). In our study morphological characters showed meaningful differences between sampling sites (Rahimi et al., 2013) but investigation of genetic populations of S. albella revealed no meaningful genetic differentiation between populations and Tajima’s D and Fu's Fs show genetic expansion in 3 localities. High amounts of \( \theta_1 \) also demonstrate expansion of genetic diversity, so these results demonstrated that S. abella populations may be having stability and widespread dispersal in the Persian Gulf and the Sea of Oman. However the high haplotype diversity and low nucleotide diversity in Qeshm may be the result of migrations of some groups for feeding in the sampling site (Grant and Bowen, 1998). It seems that observed differences in morphological characters are only in the phenotype level and the
populations of *Sardinella albella* in the Persian Gulf and Oman Sea have no genetical differences. On the other hand low observed genetic variation (Fst=0.02) did not have enough ability to separate the populations in the sample sites. So our hypotheses: there are no significant genetic differences between sampling zones, can be accepted. But more investigation with other gene targets must be done to approve this conclusion. These results can be useful to manage the source of this fish in the Persian Gulf and the Sea of Oman.

**Acknowledgements**

We would like to specially thank Salarpouri, Laloei and Taghavi in the Iranian Fisheries Science Research Institute (IFRSRI) and Azad University of Tehran, Science and Research Branch and Azad University of Zanjan in sample collection and experimental works.

**References**


Bowen, B.W. and Grant, W.S., 1997. Phylogeography of the sardines (*Sardinops* spp.): assessing biogeographic models and population histories in temperate


Guarnieo, I., Franzellitti, S., Ungaro, N., Tommasini, S., Piccinetti, C.
and Tinti, F., 2002. Control region haplotype variation in the central Mediterranean common sole indicates geographical isolation and population structuring in Italian stocks. Journal of Fish Biology, 60, 1459-1474.


Molecular Phylogenetics and Evolution, 39, 734-746.


Rahimi, P., Rezvani, S., Mostafavi, P.G., Jamili Sh. and Rahnema, M., 2013. Study of morphological variation in different populations of white sardines (Clupeidae: Sardinella albella) in the Persian Gulf and Oman Sea. First Iranian Ichthyology Conference. Isfahan University, Iran.


**Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S., 2011.** MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, 28 (10), 2731-2739


