Genetic Variation of *Clupeonella engrauliformis* Populations Inferred from RFLP Analysis of Mitochondrial DNA D-loop Region on the Southern Coast of the Caspian Sea, Iran

F. LALOEI¹, A. EIMANIFAR²* and S. REZVANI³

¹Ecological Academy of Caspian Sea
PO Box 961, Khazar-abad, Sari, Iran

²Westocentral Asian Artemia Reference Center (WAARC)
PO Box 57157-1367, Urmia, Iran

³Iranian Fisheries Research Organization (IFRO)
PO Box 14155-6116, Tehran, Iran

Abstract

Analysis of the genetic structure of the *Clupeonella engrauliformis*, one of the bony fish species, with high commercial value, was undertaken to gain information regarding the differentiation of this species from two distinct geographical areas (Anzali and Amirabad) located at the western and eastern parts of the Southern Caspian Sea, Iran. Restriction fragment length polymorphism (RFLP) analysis was performed on PCR amplified DNA fragments containing the D-loop region of the *C. engrauliformis* mitochondrial DNA. A total of 100 individuals collected from two areas were surveyed with 18 endonucleases. Six out of 18 enzymes were polymorphic yielding a total of 12 composite haplotypes. Chi-square tests showed significant results in haplotype frequencies among populations and analysis of molecular variance (AMOVA) tests indicated that most of the variation was due to within populations (98.8 %) with $F_{ST}=0.01172$. Despite the existence of significant differentiation between haplotypes for mtDNA variation, no consistent pattern of genetic structuring was revealed and nucleotide divergence among *C. engrauliformis* populations was low. The results indicated that different genotypes exist in these areas, resulting in identification of genetic stocks of *C. engrauliformis*. Therefore,

---

*Corresponding author. Tel.: +98 441 372 1818; Fax: +98 441 436 2728
E-mail address: amineimanifar@yahoo.com
careful attention must be given to these matters for the effective cultivation, management and protection of \textit{C. engrauliformis} stock on the Southern coast of the Caspian Sea.

\section*{Introduction}

The Caspian Sea (Xazar), with a total area of 435,000 km$^2$, located in northern Iran, is the largest lake in the world. The Caspian Sea has diverse ecological settings due to large and small rivers that drain into the southern coast of the sea, and its connections with littoral countries and provinces in this area \cite{CEP2002}. The most popular parts belonging to Iran are the southern and western coasts. There are over 120 fish species in the southern part of the Caspian Sea, commercially divided into two groups of sturgeons and bony fishes. The bony fishes are also divided into Kilka and other species \cite{Naderi2004}. These valuable species, with their exclusive biological characteristics, provide a fundamental basis for aquaculture management.

Kilka, one of the economically valuable species in the Caspian Sea, plays an important role in the feeding regimes of other fish species such as sturgeons \cite{Sedov1983}. There are three species of Kilka in the Caspian Sea; namely, anchovy (\textit{Clupeonella engrauliformis}), big eye (\textit{C. grimmi}) and common Kilka (\textit{C. cultriventris}) \cite{Nikonorov1964}. These species are considered to be endangered in this region due to its biological over-fishing since 1998 \cite{Fazli2007}.

\cite{Fazli2007} reported that the three Kilka species declined to their lowest level in 1998. These changes occurred as an invasive species, \textit{Ctenophora mnemiopsis}, entered the Caspian Sea \cite{Ivanov2001}. This ctenophore was transported with ballast water from the Black Sea, and appeared in 1999 in the Caspian Sea \cite{Ivanov2000}. Kideys et al. (2001a; 2001b) reported that because \textit{C. mnemiopsis} is a voracious predator on zooplankton, the availability of food for Kilka (\textit{Clupeonella} spp., a zooplanktivorous fish) declined significantly in 2001. It has also been shown that \textit{C. mnemiopsis} caused the collapse of the Kilka stocks in the Caspian Sea \cite{Kideys2001a; Kideys2001b; Kideys2003; Kideys2004; Kideys2005}. These facts, and the ecological problems involved, caused a decrease in the biomass and annual catch of current Kilka stocks, especially anchovy (\textit{C. engrauliformis}) \cite{Fazli2007}.

Defining genetic diversity in aquaculture populations or stocks is essential for interpretation, understanding and effective management of
populations or stocks (Okumus and Ciftci 2003). With the development of polymerase chain reaction (PCR), restriction fragment length polymorphisms (RFLP) analysis of PCR-amplified segments of mtDNA has become a common method for population genetic studies (Hall and Nawrocki 1995; Hansen and Loeschcke 1996). Examination of genetic variation at the DNA level has enhanced analytical power considerably. Mitochondrial DNA does not recombine, is maternally inherited, and has a relatively fast evolutionary rate compared to a single copy nuclear DNA (Brown et al. 1979; Wilson et al. 1985). Thus, a great deal of attention should be given to the analysis of mtDNA in order to determine genetic differentiation among C. engrauliformis populations.

The RFLP technique has been used as a molecular tool for evaluating genetic variation and relationships in several aquatic species, including Artemia species (Bossier et al. 2004; Baxevanis et al. 2005; Eimanifar et al. 2006), Cyprinus carpio (Gross et al. 2002; Kohlmann et al. 2003; Lehoczky et al. 2005), Rutilus rutilus caspicus (Rezvani et al. 2007), Xiphias gladius (Chow et al. 1997), Silurus glanis (Triantafyllidis et al. 1999; Krieg et al. 2000), Engraulis encrasicolus (Bembo et al. 1995), Salmo trutta L. (Apostolidis et al. 1996; Hansen et al. 1997), sturgeon (Wolf et al. 1999; Rezvani 2000), Aphanopus carbo (Quinta et al. 2004), Trachurus trachurus, T. mediterraneus and T. picturatus (Karaiskou et al. 2004), and Nephrops norvegicus (Stamatis et al. 2004).

In the present study, genetic study of anchovy Kilka (C. engrauliformis) populations was undertaken because of its importance to fisheries organizations near the Caspian Sea, enabling them to make reliable decisions on conservation management of this stock. Therefore, the current study was aimed at providing more detailed information on the genetic variability among C. engrauliformis populations. To do that, we examined restriction fragment length polymorphisms (RFLPs) of the mtDNA D-loop region samples collected from two major parts of the natural distribution area, the western (Anzali) and eastern (Amirabad) regions of the Southern coast of the Caspian Sea.
Materials and Methods

Sample collections

Sampling was carried out in two distinct areas located in the western and eastern parts of the Southern coast of the Caspian Sea. The samples were preserved in alcohol (100 %, Merck) then transferred to the genetic laboratory. The detailed information about the location of samples and the numbers studies are shown in Table 1.

Table 1. List of Clupeonella engrauliformis samples and locations on Southern coast of Caspian Sea, and number of individuals analyzed for mtDNA markers

<table>
<thead>
<tr>
<th>Region code</th>
<th>Geographic region/GIS coordinates</th>
<th>Number of individuals analysed for mtDNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANZ</td>
<td>Anzali (western part) 49° 32´ 28” - 37° 36´ 54”</td>
<td>50</td>
</tr>
<tr>
<td>AMA</td>
<td>AmirAbad (Eastern part) 53° 14´ 24” - 36° 59´ 67”</td>
<td>50</td>
</tr>
</tbody>
</table>

mtDNA preparation and PCR amplification

Approximately 50 mg of alcohol-preserved fin tissue from each fish was placed in 500 µl STE (Sodium Tris EDTA) buffer, 30 µl SDS (10 % sodium dodecyl sulphate), and digested with 20 µl of proteinase K (10 mg/ml, Roche®) at 56°C overnight. The DNA was extracted according to the phenol-chloroform protocol and 100 % ethanol precipitation (Fevolden and Pogson 1997; Rezvani 2000). Extracted DNA was examined using 1 % agarose gels, stained with ethidium bromide (10 mg•ml⁻¹; Sigma®), to check its quality and quantity, then diluted to the appropriate concentration for PCR amplification. Mitochondrial DNA was analysed by restriction fragment length polymorphisms (RFLPs) performed on PCR amplified products. The primer used for D-loop amplification was: D-loop₁ (5’→3’, 20 mer, CCT AAT CTC TGG CGA CAC GC), D-loop₂ (5’→3’, 20 mer, GCT ACA CTA GCC ACA CAC TA) (Lee et al. 1996). Double-stranded DNA was amplified in 50 µl reaction volumes containing 1 unit of DNA polymerase from Thermus aquaticus (Taq polymerase), 5 µl of 10X reaction buffer (50 mM KCl, 10 mM Tris-HCl), 2 mM MgCl₂, 200 µM of each dNTP, 0.2 µM of each primer and approximately 50-100 ng of DNA. The PCR amplification conditions were as follows: one preliminary denaturation at 94°C (5 min), followed by strand denaturation at 94°C (1 min), annealing at 54°C (30 s) and primer extension at 72°C (1 min) repeated for 30 cycles and a final extension at 72°C (7 min). All products from the PCR
amplification were confirmed by electrophoresis and were subjected directly to digestion with restriction endonucleases. The mtDNA solutions were stored at -20°C until use.

**Restriction endonuclease digestion**

The PCR products were digested following the recommendations of the manufacturers. Eighteen enzymes (Acyl, Alul, Alw26I, BclI, DdeI, DpnI, Eco47I, Hae III, HpaII, Hhal, HinfI, HpaII, MaeII, MaeIII, MboI, MboII, MspI, NdeII, and TaqI) recognizing 1 to 5 nucleotides (bases) were used. Incubations at 37°C for 4-6 h with appropriate amounts of the enzymes digested the mtDNA. Digested fragments of the D-loop segment were electrophoretically separated on 6% polyacrylamide gels (PAGE) and stained with silver nitrate. The 50 bp DNA ladder (MBI Fermentas™, Iran) was used as molecular weight markers in order to measure the size of mtDNA fragments. A, B, C, etc. in the order of detection designated the cleavage patterns by each enzyme.

**Data analysis**

For each enzyme, variable restriction patterns were alphabetically designated as they were encountered. The presence or absence of restriction sites were inferred for each of the six enzymes from a series of restriction fragment patterns that differed by a single site. The site codes across the D-loop for a restriction enzyme were concatenated, and each fish was assigned a six-letter code that described its composite, multi-enzyme haplotype. A binary character state matrix consisting of the presence or absence of all the restriction sites in composite haplotypes was produced using the GENERATE program in REAP software package (McElroy et al. 1992). Haplotype (h) and nucleotide (π) diversities within populations were estimated according to Nei (1987). The average nucleotide diversity (π_{xy}) and nucleotide divergence (δ) among populations were estimated according to Nei and Tajima (1981) using REAP. The homogeneity of the haplotype frequencies within and between geographic regions was evaluated using the contingency X² test (Roff and Bentzen 1989). The significance level was obtained by 10000 Monte Carlo randomizations using the MONTE programme from the REAP package. In order to assess the extent of genetic differentiation at the different levels of the geographic hierarchy, the overall molecular variance was partitioned into components corresponding to the population divergence within and among regions, using the analysis of molecular variance model (AMOVA) (Excoffier et al. 1992) with Arlequin version 2.0 program package (Schneider et al. 2000). Significance of
the variance components and the fixation index $F_{ST}$ value were estimated using a permutation method.

**Results and Discussion**

The amplified segment of the D-loop had an approximate size of 1015 bp, respectively. Restriction enzymes revealed the 24.9 base pairs that were analyzed, which included 2.4 percent of the target gene (D-loop) surveyed. Six ($\text{HaeIII}$, $\text{HhaI}$, $\text{HinfI}$, $\text{MaeIII}$, $\text{MboI}$, and $\text{NdeII}$) out of eighteen restriction enzymes were polymorphic for the D-loop. Between 2 and 5 fragment variants were detected for each restriction enzyme. The RFLPs six enzymes in the mtDNA D-loop region of 100 individuals examined are summarized in Table 2. The number of cleavage patterns produced by site variation was two (A, B) in $\text{MboI}$, $\text{HaeIII}$, $\text{MaeIII}$, and $\text{NdeII}$ which were observed at two sampled regions, the exception being genotype B in $\text{NdeII}$ seen only in the AMA region. Three produced patterns (A, B, and C) in $\text{HhaI}$ and $\text{HinfI}$ were observed in both regions (Table 2).

Use of twelve surveyed enzymes yielded monomorphic patterns in all individuals examined. Digestion patterns from the restriction enzymes revealed 12 composite mtDNA genotypes (haplotypes) (Table 3). Haplotype frequencies of the sampled regions, haplotype and nucleotide diversity (%) are given in the same table.

The number of observed haplotypes within regions varied in which nine and seven haplotypes were seen at the AMA and ANZ regions, respectively (Table 3).

There were four shared haplotypes among samples, whose frequencies ranged from 0.02 to 0.38. Composite haplotype I (No.I) was abundant, representing 37% of the individuals analysed. Haplotype numbers V, VII, and XI were observed only in the Anzali region, while Haplotype numbers VI, VIII, IX, X, and XII were limited to the Amirabad region (Table 3).

A few rare haplotypes with low frequency occurred only once among all samples. These rare haplotypes were also detected in *Rutilus rutilus caspicus* populations on the southern coast of the Caspian Sea (Rezvani et al. 2007). It is understandable that the occurrence of multiple composite haplotypes within a population may reflect formation of new populations with diverse genetic characteristics. *C. engrauliformis* is considered to be a migratory fish species which moves at rotatory incidents to
Table 2. Approximate fragment sizes of restriction morphs for *C. engrauliformis* samples observed by digesting the mitochondrial DNA segment D-loop with six different restriction endonucleases. The number of cleavage patterns by site variation were three (A, B, and C) for enzymes surveyed.

<table>
<thead>
<tr>
<th>Restriction Endonucleases</th>
<th>Hae III</th>
<th>Hha I</th>
<th>Hinf I</th>
<th>Mae III</th>
<th>Mbo I</th>
<th>Nde II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restriction morph</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>Fragment size (bp)</td>
<td>525</td>
<td>525</td>
<td>860</td>
<td>500</td>
<td>460</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>490</td>
<td>330</td>
<td>155</td>
<td>360</td>
<td>400</td>
<td>260</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>160</td>
<td>155</td>
<td>155</td>
<td>200</td>
</tr>
<tr>
<td>Total length (bp)</td>
<td>1015</td>
<td>1015</td>
<td>1015</td>
<td>1015</td>
<td>1015</td>
<td>1015</td>
</tr>
<tr>
<td>Number of fragments</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>

bp = base pair
the southern part of the Caspian Sea. This fish migrates in spring to the middle part of the Caspian Sea, and in winter season goes to the southern part (Fazli et al. 2003). The ratio of males to females of anchovy Kilka in the population was closed to 1:1, with a slight majority of males in spring (50.3%) and that of females in autumn (53.3%) (Fazli 2007). Therefore, formation of a larger number of haplotypes in the AMA region might result from the migratory behavior of this species.

Table 3. Composite genotypes (haplotypes), haplotype frequencies, haplotype diversity including standard error, nucleotide diversity, and sample sizes of the two regions of *Clupeonella engrauliformis* population studied.

<table>
<thead>
<tr>
<th>Haplotype no.</th>
<th>Composite genotype</th>
<th>Regions*</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>AMA</td>
<td>ANZ</td>
</tr>
<tr>
<td>I</td>
<td>AAAAAA</td>
<td>18 (0.36)</td>
<td>19 (0.38)</td>
</tr>
<tr>
<td>II</td>
<td>AAAAAA</td>
<td>7 (0.14)</td>
<td>1 (0.02)</td>
</tr>
<tr>
<td>III</td>
<td>AABAAA</td>
<td>10 (0.2 )</td>
<td>4 (0.08)</td>
</tr>
<tr>
<td>IV</td>
<td>AAABAA</td>
<td>8 (0.16)</td>
<td>7 (0.14)</td>
</tr>
<tr>
<td>V</td>
<td>AAAABA</td>
<td>0</td>
<td>3 (0.06)</td>
</tr>
<tr>
<td>VI</td>
<td>AABABA</td>
<td>2 (0.04)</td>
<td>0</td>
</tr>
<tr>
<td>VII</td>
<td>BAAAAA</td>
<td>0</td>
<td>4 (0.08)</td>
</tr>
<tr>
<td>VIII</td>
<td>BBBBBB</td>
<td>3 (0.06)</td>
<td>0</td>
</tr>
<tr>
<td>IX</td>
<td>BBBBBB</td>
<td>2 (0.04)</td>
<td>0</td>
</tr>
<tr>
<td>X</td>
<td>AAAAAAB</td>
<td>1 (0.02)</td>
<td>0</td>
</tr>
<tr>
<td>XI</td>
<td>AACAAA</td>
<td>0</td>
<td>2 (0.04)</td>
</tr>
<tr>
<td>XII</td>
<td>AAAAAAC</td>
<td>1 (0.02)</td>
<td>0</td>
</tr>
</tbody>
</table>

| Number of haplotypes | 9 | 7 |
| Rate of haplotypes | 0.18 | 0.14 |
| Nucleotide diversity (%) | 1 | 0.8 |
| Haplotype diversity (%) | 81 | 76.1 |
| Standard error (%) | 3.42 | 3.83 |
| Total sample size | 50 | 50 |

*AMA and ANZ refer to sampling regions included in the Amirabad and Anzali regions.

1Number of haplotype occurrence.

2Rate of haplotype occurrence; Number of haplotype occurrence / sample size analysed.

Haplotypes are numbered by ranging in alphabetical order the digestion types of each restriction endonuclease. Note: Composite haplotypes reflect digestions with the following restriction endonucleases (left to right): *HaeIII, Hhal, Hinfl, MaelIII, Mbol and NdeII*.

The geographical distribution of *C. engrauliformis* is restricted mainly to both regions of the southern part of the Caspian Sea. Each of the three Kilka species has unique morphological characteristics and they can easily be differentiated from each other on this basis. However, sometimes populations with similar phenotypes may differ in terms of genetics (Barker et al. 1997). Morphological parameters are not reliable factors to
identify valuable stocks (Lin et al. 2002). Therefore, DNA data using molecular based markers provides a useful tool to get more information on the genetic structure and diversity of threatened fish populations or stocks (Cheng and Lu 2005).

The AMOVA analysis suggested that most of the genetic polymorphism (98.8 %) was due to differences within populations. The variation among populations contributes only 1.2 % of the total differences, and the FST value (0.01172) was not significant (P= 0.15), suggesting these samples from the two regions examined have not diverged in terms of genetics.

Genetic variability among two populations of *C. engrauliformis* is shown in table 3. Haplotype rate occurred at a rate of 0.14 and 0.18 in the two regions. Haplotype diversity (h) within each population was 81 (AMA region) and 76.1 (ANZ region), respectively, with an average of 78.5. Nucleotide diversity (π) was 1 % (AMA region) and 0.8 % (ANZ region), respectively, with an overall average of 0.9. The obtained values were at lower levels than those observed by Imsiridou et al. (1998) who observed 21 different composite haplotypes and relatively high nucleotide diversity within *Leuciscus cephalus* populations using the D-loop as a target gene region. Possible reasons for the observed low genetic variability are as follows: firstly, industrial pollution caused by Caspian Sea littoral countries; secondly, development of coastal communities with resulting sewage discharge into coastal water; thirdly, man-made activities with adverse effects on aquatic populations and coastal ecosystems of the Caspian Sea.

The contingency X2-test revealed significant outcomes in haplotype frequencies for the entire population sets (X2=10.9, P<0.005). Nucleotide divergence (δ) between the two regions (AMA and ANZ) was estimated to be 0.01 %. The low level of genetic divergence among tested samples agrees with statements made by Waters et al. (2000). We recognize that in each ecosystem in which fish movements from one region to another take place, total nucleotide divergence will be low. Physiological behaviour of this species shows, due to lack of physical barriers, the possibility of moving *C. engrauliformis* at both western and eastern areas of southern Caspian Sea can be occurred. In studies carried out by Waters et al. (2000) the level of genetic diversity of *Alosa sapidissima* populations at the Colombia river was lower than at other locations due to lack of gene flow of this species in this river in comparison with other rivers.

We concluded that the PCR-RFLP analysis adopted in the present study may be suited for the identification of *C. engrauliformis* genetic
stocks. The collected samples can be preserved in ethanol at the sampling site and stored for several years until identification is required. We suggest that comprehensive study of physiological features such as reproduction and nutrition of Kilka stocks will be useful for the effective cultivation, management and protection of *C. engrauliformis* populations.

Acknowledgements

We would like to express our heart-felt gratitude and strong appreciation to Prof. James S. Clegg at Bodega Marine Laboratory and Section of Molecular & Cellular Biology University of California, who improved the quality of this manuscript with their comments and suggestions. This work was supported by a research grant from the Iranian Fisheries Research Organization (IFRO).

References


