Mitochondrial DNA Length Variation and Heteroplasmy in Natural Populations of the European Sea Bass, *Dicentrarchus labrax*

Donatella Cesaroni, Federica Venanzetti, Giuliana Allegrucci, and Valerio Sbordoni

Dipartimento di Biologia, Università Tor Vergata, Roma, Italy

Mitochondrial DNA length variation in *Dicentrarchus labrax* is the outcome of the simultaneous variation in the copy number of two tandem repeat arrays located in the D-loop region. The repeated sequences, named R17 and R48, are 17 and 48 bp long, respectively. On the basis of 209 individuals collected in eight Mediterranean localities, length variation and its partitioning between the two repeat arrays have been characterized and quantified by means of frequency distributions of repeat copy number and genetic diversity estimates. The median copy numbers are 4 and 11 for the R17 and R48 repeat arrays, respectively. For both repeat arrays, continuity in step variation and the occurrence of both frequent and rare arrangements was observed. More than 50% of the individuals are heteroplasmic for up to four variants. Within populations, heteroplasmy ranges between 33.3% and 70.0% of the individuals. A large proportion of the total gene diversity (29.1%) occurs within individuals and the greatest proportion (66.6%) is found between individuals within populations, while only 4.3% is due to variation between populations. Although the interpopulation component of the diversity is rather low, the significant differentiation of the two Eastern populations from all the other Western samples suggests the occurrence of a phylogeographic component of the genetic divergence between the Mediterranean populations, which is only recognizable on a large geographic scale.

**Introduction**

The control region (D-loop) is the most variable region of mitochondrial DNA (mtDNA) because it has the highest average rate of nucleotide substitution and because of the variation in copy number of tandemly repeated sequences. These repeat arrays have been regarded as primarily responsible for the length variation of the mtDNA molecule in many animal species (Brown 1983; Moritz, Dowling, and Brown 1987; Rand 1994; Meyer 1994). In fish, length variation is not a rare phenomenon, despite the small size range of the mtDNA molecule (Billington and Hebert 1991; Meyer 1994). In most cases where length variation has been characterized it is due to variation in the copy number of tandem repeats in the control region (Bentzen, Legget, and Brown 1988; Mulligan and Chapman 1989; Virgin, Proenca, and Grossfeld 1989; Buroker et al. 1990; Arnason and Rand 1992; Brown, Beckenbach, and Smith 1992). In fish, as in other animals, high levels of variation in the copy numbers of tandem repeats in the control region are commonly associated with high levels of heteroplasmy (Rand 1994), which may be up to 100%, as in *Morone americana* (Mulligan and Chapman 1989) and *Gadus morhua* (Arnason and Rand 1992). Several models have been developed to explain the origin and the evolutionary meaning of this kind of variation, its evolutionary rate, its selective or neutral maintenance, and the mechanisms of homogenization between repeats and of transmission of variants both within and between individuals (Rand and Harrison 1989; Buroker et al. 1990; Wilkinson and Chapman 1991; Arnason and Rand 1992; Brown, Beckenbach, and Smith 1992; Rand 1993, 1994; Hoelzel et al. 1994).

The mitochondrial genome of the European sea bass, *Dicentrarchus labrax*, is about 18 kb in size, one of the largest mtDNA molecules in fishes. MtDNA length variations have been reported among individuals, and this polymorphism has been found to be entirely due to length variation of the D-loop region (Venanzetti et al. 1994). On the basis of DNA sequences from three D-loop length variants, the following repeated sequence organization has been observed (Cecconi, Giorgi, and Mariottini 1995): (1) an array of 17-bp sequence repeats with a variable copy number, (2) an array of 48-bp tandemly repeated sequences with a variable copy number, and (3) an "imperfect" 17-bp repeat (\(\Psi R2\)) occurring only once in the D loop region, which separates the 17-bp repeat array from the 48-bp one. Both repeat arrays are made up of closely related sequences; copies of the 17-bp tandem repeats differ from each other by only three nucleotides, while copies within the 48-bp repeat clusters vary by only one nucleotide.

In this paper we analyze the pattern of length variation of this mtDNA region in eight Mediterranean populations, from southern France to Greece and Egypt, by quantifying the levels of intra- and interpopulation variation. The aim is to characterize and quantify the length variation in a large geographical sample of individuals in order to assess its modality of expression, its partitioning between the two categories of repeat arrays, the distribution of repeat copy numbers, the frequency of heteroplasmy, and the differences between individuals and populations. Given the peculiar organization of the D-loop region of the European sea bass, which is due to the coexistence in the same molecule of two distinct repeat arrays, we specifically intended to (1) verify if the pattern of variation is similar for the two repeat arrays, (2) compare to those patterns observed when only one repeat array occurs in the D-loop region, (3) check for heteroplasmy and its mechanisms, and (4) test if such a hypervariable marker might be useful to detect differentiation between populations.
Materials and Methods

Sampling

We analyzed 209 individuals from eight Mediterranean localities. Six samples came from natural populations (FCT: Fiumicino, Latium, Italy; SET: Sète, Gulf of Lyons, France; EGL: Bardawil, Sinai, Egypt; MUR: Muravera, Sardinia, Italy; PRE: Prevost, Gulf of Lyons, France; SBB: Sabaudia, Latium, Italy). Two additional samples came from reared stocks derived from indigenous populations (MSL: Menidi, Gulf of Arta, Central Greece, F₁ offspring of wild specimens from Messolongi; SCL: Marsala, Sicily, Italy, F₂ offspring of wild specimens).

DNA Isolation, Amplification and Electrophoresis

DNA extractions followed a standard phenol-chloroform procedure (Allegrecci et al. 1995). For each individual, we used about 20 ng of total genomic DNA to amplify by polymerase chain reaction (PCR) the variable portion of the D-loop region. Primer sequences were derived from Cecconi, Giorgi, and Mariotti (1995). They are complementary to unique sequences flanking the tandem arrays located at positions 284–302 (CCCCAGTGCACACAAACGCGC) and positions 1509–1490 (TTGTCCTGAACTAACAGCC) of the D. labrax D-loop sequence (EMBL data library accession no. X81472). PCR conditions were as follows: melt at 94°C for 1 min, anneal at 55°C for 1 min, and extend at 72°C for 2 min and 30 s; the cycle was repeated 30 times. PCR products were separated by electrophoresis through ethidium-bromide-stained 1%–1.5% agarose gels in Tris-borate buffer (TBE; Sambrook, Fritsch, and Maniatis 1989), and their sizes were scored in comparison to molecular weight standards. The amplified fragments varied in size from 800 to 1,500 bp. Since several individual amplifications produced more than one band, three additional tests were carried out to reduce the probability of scoring PCR artifacts: (1) all fragments from individual PCR reactions with multiple bands were isolated using low-melting-point agarose gels and reamplified separately; if the single-fragment reamplification produced the entire multiple pattern, that fragment was discarded; (2) one specimen from each genotype class that showed multiple bands at the standard annealing temperature, was reamplified twice using annealing temperatures of 50°C and 57°C; if the pattern was identical, the individual was used in the study; (3) specimens from the same genotype classes were also double digested with endonucleases Xba I and Mlu I, which cut in unique sites flanking the repeat arrays; if all the fragments decreased in size by the digested portions as expected, the individual was included in the study.

Estimates of the Number of Tandem Repeats

To ascertain the numbers of copies of the 17-bp sequences (R17) and 48-bp sequences (R48), each fragment was double digested with Nde I, an endonuclease recognizing a site within each R48, and Xba I, which cleaves the amplified region at a unique site 163 bp away from the 3’ end. After digestion the products were visualized on 2% TBE agarose gels and compared to molecular weight standards. Each residual DNA fragment was composed by the 5’ end of the entire fragment, which was constantly 290 bp long, plus a fragment which was as many times 17 bp long as the number of the occurring R17 repeats. The number of R48 repeats was inferred comparing the fragment sizes before and after digestion.

Statistical Analyses

The statistical analysis, i.e., frequency distributions, χ² tests, t-tests, and correlation tests, was performed by using standard procedures included in the STATISTICA package (version 4.5, StatSoft, Inc., Tulsa, Okla., 1993).

Gene diversity within an individual (Kᵢ), within a population (Kᵢ), and within the total sample (K₀) were calculated according to the general formula K = 1 – Σxᵢ, where xᵢ is the frequency of the iᵗʰ length variant (Birky, Fuerst, and Maruyama 1989). Total variation was apportioned into hierarchical components according to hierarchical C statistics (Lewontin 1972): the proportion of variation within individuals, Cᵢ = (mean Kᵢ)/K₀, among individuals within the population, Cᵢp = (mean Kᵢ – mean Kᵢ)/K₀, and among populations within the total sample, Cᵢp = (Kᵢ – mean Kᵢ)/K₀. Heterogeneity G-tests (Sokal and Rolhf 1981, p. 721) were performed in order to evaluate the heterogeneity of Kᵢ and Kᵢ estimates from the eight samples. G for heterogeneity (Gₘ) and pooled G (Gₚ) were computed, and a total G (Gₚ) was partitioned into contributions due to individual samples. Results from G-tests could provide an indirect evaluation of the reliability of the subsequent C estimates.

Divergence estimates between length variants were calculated on the basis of a method developed by Nei (1987) for DNA length polymorphism caused by insertion and deletion. According to Nei (1987, p. 266), the number of gap nucleotides (nucleotides in the gaps) per nucleotide site (gᵢ) between the iᵗʰ and jᵗʰ haplotypes is given by gᵢj = mᵢj/gᵢj, where mᵢj is the number of gap nucleotides between the iᵗʰ and jᵗʰ haplotypes, and gᵢj is the total number of nucleotides compared, including gap nucleotides. By considering the D-loop length variants as sequences of “repeat” units instead of sequences of nucleotides, we evaluated “the number of gap repeats per repeat site” with an analogous formula. The gᵢ values were calculated between all possible pairwise comparisons of fragments, and differences in frequency distributions of the values from individuals within and between populations were tested by t-tests, adjusting the significance levels for multiple tests using the sequential Bonferroni technique (Rice 1989). Average g values within and between populations were also calculated, and interpopulation estimates were corrected for within-population polymorphism according to the formula: g₉ = gₓᵧ – 0.5(gₓₓ + gᵧᵧ), where gₓₓ is the mean estimate in pairwise comparisons of individuals between X and Y populations, and gₓᵧ are the mean estimates among individuals within these respective populations (Avise 1994, p. 96). Squared Mahalanobis distances between centroids of populations and their significance were calculated in the space defined.
FIG. 1.—Frequency distributions of the sizes of the variants and the copy numbers of R17 and R48 repeated sequences deriving from the analysis of 354 DNA fragments produced by PCR amplification of the mtDNA D-loop regions of 209 individuals of *Dicentrarchus labrax*.

FIG. 2.—Bivariate histogram showing the range of the simultaneous variation in copy number for the R17 and R48 tandem repeats. Each bar represents the relative frequency of a certain combination of R17 and R48 within the total European sea bass sample.
Table 1

<table>
<thead>
<tr>
<th>Sample statistics</th>
<th>EGL</th>
<th>FCT</th>
<th>MSL</th>
<th>MUR</th>
<th>PRE</th>
<th>SBB</th>
<th>SCL</th>
<th>SET</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>29</td>
<td>28</td>
<td>20</td>
<td>34</td>
<td>19</td>
<td>27</td>
<td>32</td>
<td>20</td>
<td>209</td>
</tr>
<tr>
<td>Total fragments</td>
<td>44</td>
<td>45</td>
<td>40</td>
<td>67</td>
<td>29</td>
<td>57</td>
<td>57</td>
<td>57</td>
<td>524</td>
</tr>
<tr>
<td>No. of length variants</td>
<td>34</td>
<td>18</td>
<td>17</td>
<td>47</td>
<td>77</td>
<td>73</td>
<td>30</td>
<td>31</td>
<td>104</td>
</tr>
<tr>
<td>Average size of fragments (bp)</td>
<td>1.050</td>
<td>1.177</td>
<td>1.040</td>
<td>1.155</td>
<td>1.197</td>
<td>1.086</td>
<td>1.189</td>
<td>1.145</td>
<td>1.132</td>
</tr>
<tr>
<td>Median R17 copy number</td>
<td>5</td>
<td>8</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Median R48 copy number</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>11</td>
<td>13</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>No. of distinct genotypes</td>
<td>26</td>
<td>17</td>
<td>11</td>
<td>32</td>
<td>18</td>
<td>19</td>
<td>22</td>
<td>19</td>
<td></td>
</tr>
</tbody>
</table>

**Heteroplasmy**

No. of homoplasmic individuals | 16  | 16  | 16  | 12  | 10  | 18  | 15  | 7   | 100 |
No. of individuals with 2 mtDNAs | 11  | 9   | 9   | 9   | 13  | 8   | 8   | 11  | 10  |
No. of individuals with 3 mtDNAs | 2   | 3   | 4   | 7   | 1   | 1   | 4   | 2   | 24  |
No. of individuals with 4 mtDNAs | 0   | 0   | 1   | 2   | 0   | 0   | 2   | 1   | 6   |

% Heteroplasmy | 44.8 | 42.9 | 70.0 | 64.7 | 47.4 | 47.4 | 33.3 | 53.1 | 65.0 | 52.1 |

Note: Three-letter symbols refer to the different geographic samples detailed in the Materials and Methods section.

all of the 354 scored fragments ($r = 0.170, P < 0.002$) or excluding eight outliers ($r = -0.106, P < 0.05$). This negative correlation suggested the occurrence of size constraints on the mtDNA molecule.

**Heteroplasmy**

More than 52% of the individuals were heteroplastic for up to four mtDNA length variants. Individuals carrying two length variants were clearly more frequent than individuals with three or four variants. Heteroplasmy within populations ranged between 33.3% and 70.0% (table 1). To test if there were differences between homoplasmic and heteroplasmic individuals in the sizes of variants, we carried out a correlation analysis between the numbers and the relative lengths of the variants. This test produced a statistically significant correlation coefficient ($r = 0.183, P = 0.001$), suggesting that long variants occur more frequently in heteroplasmic than in homoplasmic individuals. In addition, a t-test for unpaired samples showed that the means of the variant sizes are significantly different in homoplasmic and heteroplasmic individuals ($t = 3.87; df = 352; P = 0.0001$). Actually, length variants sized more than 1,400 bp were found only in heteroplasmic individuals. Looking at the frequency distributions of the R17 and R48 repeat copy numbers, a statistically significant difference was found for the R48 repeats ($t = 3.13, df = 352, P = 0.002$). Statistical significance for the R17 repeats was less strong ($t = 1.94, df = 352, P = 0.053$) than for the R48 repeats.

In most cases (97 out of 109), the size differences between variants within heteroplasmic individuals were due to differences in copy number of only one of the two repeat arrays. The variation in copy number of the R48 repeat was more frequent (94 cases) than that of the R17 repeat (27 cases). Only the Sardinian sample (MUR) showed a relatively high frequency of simultaneous variation of the two kinds of repeats in the same specimen (in 8 out of 22 heteroplasmic individuals). About 66% of the heteroplasmic individuals showed an increment in size between variants equal to the size of only one repeat, either R17 or R48. The remaining 34% showed variants with increased sizes, as expected from multiple repeat copies, without intermediate steps.

**Genetic Diversity and Genetic Differentiation Between Populations**

Table 2 gives estimates of gene diversity for this region of mtDNA within an individual ($K_i$), within a population ($K_c$), and within the total sample ($K_b$). These estimates were obtained for the whole variable region.

Table 2

<table>
<thead>
<tr>
<th>Haplotype diversity</th>
<th>EGL</th>
<th>FCT</th>
<th>MSL</th>
<th>MUR</th>
<th>PRE</th>
<th>SBB</th>
<th>SCL</th>
<th>SET</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean $K_p$ per population</td>
<td>0.219</td>
<td>0.233</td>
<td>0.397</td>
<td>0.373</td>
<td>0.246</td>
<td>0.173</td>
<td>0.303</td>
<td>0.355</td>
</tr>
<tr>
<td>$K_c$</td>
<td>0.964</td>
<td>0.867</td>
<td>0.893</td>
<td>0.973</td>
<td>0.958</td>
<td>0.935</td>
<td>0.947</td>
<td>0.961</td>
</tr>
<tr>
<td>Mean $K_p$ per population</td>
<td>0.040</td>
<td>0.069</td>
<td>0.022</td>
<td>0.163</td>
<td>0.079</td>
<td>0.018</td>
<td>0.029</td>
<td>0.050</td>
</tr>
<tr>
<td>$K_c$</td>
<td>0.838</td>
<td>0.724</td>
<td>0.711</td>
<td>0.888</td>
<td>0.842</td>
<td>0.738</td>
<td>0.763</td>
<td>0.776</td>
</tr>
<tr>
<td>Mean $K_p$ per population</td>
<td>0.713</td>
<td>0.171</td>
<td>0.385</td>
<td>0.311</td>
<td>0.167</td>
<td>0.154</td>
<td>0.786</td>
<td>0.304</td>
</tr>
<tr>
<td>$K_c$</td>
<td>0.850</td>
<td>0.716</td>
<td>0.821</td>
<td>0.866</td>
<td>0.839</td>
<td>0.805</td>
<td>0.890</td>
<td>0.887</td>
</tr>
<tr>
<td>Genotype diversity</td>
<td>0.958</td>
<td>0.880</td>
<td>0.855</td>
<td>0.967</td>
<td>0.942</td>
<td>0.930</td>
<td>0.941</td>
<td>0.943</td>
</tr>
</tbody>
</table>
and for the two repeat arrays separately. A comparison of the $K_r$, $K_s$, and $K_c$ values clearly indicates that most of the total variation is already expressed within a population. The homogeneity of the $K_r$ and $K_c$ estimates deriving from different geographic samples was tested by means of heterogeneity $G$-tests. Both values of $G$ for heterogeneity ($G_H$) were not significant ($K_r$: $G_H = 6.06$, $df = 7$, $P > 0.5$; $K_c$: $G_H = 4.52$, $df = 7$, $P > 0.5$). Partitioning the total $G$ ($G_T$) into contributions due to individual samples showed that none of them was significant. The results from the $G$-tests validated the use of mean $K_r$ and $K_c$ estimates to calculate $C$ values (apportionment of the total diversity into hierarchical components). In terms of relative diversity, the hierarchical $C$ statistics showed that 29.1% of the total diversity lies within individuals, 66.6% among individuals within the same population, and only 4.3% among populations.

Notwithstanding the very low percentage of variation occurring among populations, we tested if geographically distinct samples were differentiated on the basis of this mtDNA region. A general formulation for DNA length polymorphism caused by deletion and insertion (Nei 1987) was modified and used to measure divergence between length variants. Pooling together both repeat arrays, which are contiguous along mtDNA, the presence or absence of a given repeat was used to calculate the total number of gap repeats per repeat site ($g_{ij}$) between all pairwise combinations of the scored fragments. Figure 3 shows the frequency distributions of $g_{ij}$ estimates in pairwise comparisons of fragments within (dark bars) and between (light bars) populations. Although divergence between intra- and interpopulation estimates largely overlap, the $t$ statistics showed that most of the compared distributions were significantly different; this was also true when the significance levels were adjusted using the sequential Bonferroni test. In particular, the Egyptian (EGL) and Greek (MSL) populations were found to be significantly different from all the other Western populations, except between EGL and SBB (central Italy). Tests of the significance of the squared Mahalanobis distances (based on R17 and R48 copy numbers) between populations produced similar results, with distances between EGL and MSL and all the other populations being statistically significant ($P < 0.001$).

The principal coordinate analysis was used to visualize the relationships between populations using two estimates of genetic differentiation: the $g_{ij}$ values and the squared Mahalanobis distances. Both scatterplots in figure 4 are based on the first two principal axes and...
show that the two Eastern populations (EGL, MSL) are located in the opposite direction from all the Western populations except SBB. The separation of the two Eastern populations from all the others is more evident when the squared Mahalanobis distances are considered (fig. 4B).

Discussion

Length Variation and Copy Number of the Repeat Arrays

Major size differences in animal mtDNA are often due to variation in the copy numbers of tandemly repeated sequences. The molecular characterization and sequencing of the D. labrax mtDNA control region has revealed unusual features, such as the occurrence of two different repeat arrays which allow the expression of a great deal of variation (Venanzetti et al. 1994; Cecconi, Giorgi, and Mariottini 1995). A similar D-loop region organization has been found in rabbit mtDNA, where two sets of tandem repeats (20 bp and 153 bp), both variable in copy numbers, generate high levels of heterogeneity (Mignotte et al. 1990). This D-loop organization probably involves mechanisms leading to variation which are more complex than in the other vertebrates.

In D. labrax, analyzing the copy number variation as expressed by each repeat array separately, we obtain results comparable to those observed for other vertebrates. The frequency distribution of the R17 copy number is clearly skewed to few repeats (fig. 1). Analogous repeat copy number frequency distributions have been observed for the 40-bp repeat sequences of the cod, Gadus morhua (Arnason and Rand 1992), and the 82 bp repeat sequences of the white sturgeon, Acipenser transmontanus (Buroker et al. 1990; Brown, Beckenbach, and Smith 1992), where the length variation is accounted for by a single repeat array. On the other hand, the patterns of variation of the R48 and R17 repeat arrays in D. labrax are quite different from one another (fig. 1), since they differ in the shape of the copy number frequency distribution (R48 = normal; R17 = skewed), the minimum copy number (R48 = 6; R17 = 1), and the median copy number (R48 = 11; R17 = 4). The shape of the frequency distribution of the R48 copy numbers is comparable to that observed in the evening bat, Nycticeius humeralis, where an 81-bp sequence is repeated five to eight times (Wilkinson and Chapman 1991).

If the R17 and R48 arrays behave differently, and mtDNA length polymorphism in D. labrax is the outcome of their simultaneous variation, several questions arise: Is the variation expressed within each array similar? Are there mechanisms which reduce or increase the potentiality for repeat array variation? Which kinds of basic changes within repeat arrays are frequent, and which are rare?

Heteroplasmy

High frequency of heteroplasmy due to tandemly repeated sequences has been generally explained by high rates of recurrent mutation for insertions and deletions (indels) (Rand and Harrison 1989; Buroker et al. 1990; Brown, Beckenbach, and Smith 1992; Arnason and Rand 1992). The levels of heteroplasmy in D. labrax (table 1) are comparable to those observed in other animals, especially fish, with tandemly repeated sequences in the D-loop region (Rand 1993). However, the percent heteroplasmy is different among geographically different samples, with a relatively wide range of estimates around an average value of 52.1%, indicating that the incidence of heteroplasmy could differ significantly between populations, as is already known for the white sturgeon (Brown, Beckenbach, and Smith 1992). This result strongly suggests the need to include more than one population sample when assessing heteroplasmy levels within a species.

In the white sturgeon, a significant difference in the distribution of repeat copy numbers between heteroplasmic and homoplasmic individuals has been found (Brown, Beckenbach, and Smith 1992). This difference is due to a strong bias toward large molecules in heteroplasmic fishes. They never carry, as the predominant variant, the single repeat molecule, which is otherwise largely dominant in homoplasmic fishes. In D. labrax, we observe a similar phenomenon: heteroplasmic individuals have a higher number of large variants than homoplasmic individuals. The large size of these variants depends only on variations in the R48 copy number.
However, the R48 frequency distributions for heteroplasmic and homoplasmic individuals in *D. labrax* are not clearly as separate as in sturgeon.

The hierarchical apportioning of diversity for D-loop length variants has shown that a large proportion of the total variation (29.1%) occurs within individuals. If we consider recurrent mutation for indels as the primary process generating heteroplasy, the European sea bass should have high mutation rates. A high mutation rate in a tandem array may produce length variants which are detectable within individuals, because the generation of variation overcomes the loss due to random drift during cell division (Arnason and Rand 1992).

When gene diversity within individuals is estimated separately as length variation due to R17 and R48 repeats, respectively (table 2), the mean value of $K_g$ (gene diversity within a population) is significantly lower for the R17 repeats (0.062) than for the R48 repeats (0.248). The apportionment of diversity within individuals is different as well, since the proportion of the total diversity within individuals is $C_i = 0.072$ for the R17 repeats and $C_i = 0.285$ for the R48 repeats. This result suggests that the two repeat arrays behave differently during the mechanism of replication. The competitive displacement model developed by Burorker et al. (1990) to explain length variants and heteroplasm genesis in white sturgeon could also explain the gain or loss of repeat units observed in the *D. labrax* D-loop region (Cecconi, Giorgi, and Mariottini 1995). However, competing hypotheses about possible different roles and implications for short and long repeats of *D. labrax* mtDNA have still to be experimentally tested.

Most heteroplasmic sea basses show changes in their variants which are increments of the copy number of the R17 or R48 repeats. The hypothesis of recurrent mutation for indels provides a reasonable explanation for this outcome because it allows for the introduction of incremental changes to the copy numbers during the replication mechanisms, which might maintain heteroplasm. On the other hand, some individuals do not show simple incremental changes between variants, and the differences may be large. For example, we find variants within individuals differing up to five times the size of one R48 repeat, with no intermediate steps, or variants differing simultaneously for the number of the R17 and R48 repeats. This situation might be considered comparable to the "bimodal distribution" of length variants observed in heteroplasmic sturgeons (Brown, Beckenbach, and Smith 1992), where rare individuals show high proportions of quite different variants. A possible explanation for the occurrence of heteroplasmic individuals carrying very different variants could be incidental biparental mtDNA inheritance. However, there are only a few cases where the paternal contribution has been demonstrated: in anchovy (Magoulas and Zouros 1993), in mice (Gyllensten et al. 1991), in *Drosophila* (Kondo et al. 1990), and in mussels (Hoeh, Blakley, and Brown 1991). On the other hand, it is still uncertain whether heteroplasmic individuals always transmit all the size variants to their offspring (Mulligan and Chapman 1989), although studies on the evening bat supported an affirmative answer (Wilkinson and Chapman 1991).

Value of the Genetic Diversity for mtDNA Length Variants in European Sea Bass

Even if the hierarchical apportioning of diversity shows that the greatest proportion (66.6%) is attributed to variation between individuals within populations, differences in mtDNA variants between populations are clearly apparent by looking at the data summarized in table 1 and the R17–R48 copy number arrangements. The residual 4.3% of diversity, which lies between populations, seems to be sufficient to produce significant differences between groups of populations, since both the means and the distributions of divergence estimates between individuals sampled in the same and in distinct localities are often significantly different (fig. 3). The principal coordinate analyses suggest that two main groups of populations may be distinguished, although the genetic distance geometry is not fully concordant with the geographic origin. A group of Western Mediterranean samples (SCL, PRE, SET, MUR, FCT) are displaced on one side of the scatterplots in figure 4, whereas on the opposite side there are the samples collected near the Eastern coasts (EGL, MSL) with the addition of the outsider Tyrrenian population (SBT). Actually, the Greek and Egyptian samples share typical repeat array arrangements; 8 out of the 17 length variants found in MSL individuals occur also in EGL samples. Moreover, both population samples have higher frequencies of the smallest variant sizes than do the Western samples. They are also the only two samples which are significantly different from all the others based on the squared Mahalanobis distances and on the differences between intra- and interpopulation divergence ($g$ values). The differentiation of these Eastern populations suggests the occurrence of a phylogeographic component which is distinguishable only on a large scale. However, it is likely that its role in shaping the pattern of mtDNA length variation in the European sea bass is relatively minor. The distinctiveness of the Eastern populations is also consistent with other molecular studies carried out on the same samples. Both allozymes (Allegucci, Fortunato, and Sbordoni 1997) and RAPD markers (Caccone et al. 1997) suggest an east west trend of genetic differentiation within the Mediterranean populations of *D. labrax*. In addition, a study of mtDNA sequence variation based on fragments of the cytochrome b gene points to differences in the haplotype distribution between Eastern and Western Mediterranean populations of *D. labrax* (Patarnello et al. 1993).

The extremely high level of intrapopulation polymorphism found in the D-loop region overcomes the possibility to appreciate significant differences between adjacent populations. As Rand (1994) stressed, such a pattern of mtDNA length variation is of little value in distinguishing among populations because high mutation rates may produce convergent or parallel acquisition of length variants in distinct haplotypes. On the other hand, nucleotide polymorphisms in repeat arrays (RAPS: repeat associated polymorphisms; Rand 1992)
could be a useful tool to study evolutionary relationships between populations from the spread of point mutations through copies of repeats in tandem arrays (Rand 1994). Sequences of the _D. labrax_ D-loop region from three individuals indicate the occurrence of both nucleotide polymorphisms in tandem repeat arrays and order shuffling between repeats (Cecconi, Giorgi, and Mariottini 1995). The sequencing of a greater number of repeats could provide an in-depth look at the dynamics of the homogenization of the tandem arrays and the occurrence of concerted evolution.

**Acknowledgments**

We are grateful to Adalgisa Caccone and Carlo Fortunato, who extracted all the DNA samples used in this study together with G.A. for a companion paper on RAPD-PCR variation. Thanks are also due to Paolo Mariottini for providing PCR primers and to François Bonhomme for providing French samples. We thank Adalgisa Caccone for her critical reading of the manuscript and helpful comments, and two anonymous reviewers for their constructive comments. The work presented here is part of a research project on the genetic characterization of the European sea bass supported by the National Research Council (C.N.R.), Special Project RAISA, Subproject No. 3. We are also indebted to Stefano Cataudella for help in planning and collecting samples during the preliminary phases of the research project.

**LITERATURE CITED**


JEFFREY R. POWELL, reviewing editor
Accepted February 3, 1997