Allozyme, mitochondrial-DNA, and morphometric variability indicate cryptic species of anchovy (Engraulis encrasicolus)

PHILIPPE BORSA*

Institut de Recherche pour le Développement (UR081), BPA5, Nouméa, New Caledonia

Received 22 April 2001; accepted for publication 2 November 2001

Previous surveys of population structure in the Atlantic-Mediterranean anchovy Engraulis encrasicolus L. have reported heterogeneity in morphology, allozyme frequencies, and mitochondrial DNA haplotype frequencies at a regional scale. In particular, two stocks of anchovy have been detected in the Adriatic Sea. In this paper, the available data is reviewed with the aim to relate genetic variation to geography at the widest possible geographical scale, for investigating the evolutionary mechanisms underlying stock structure in anchovy. Correspondence analysis of allozyme frequencies (24 samples, three polymorphic loci) compiled from the literature indicates three distinct entities in the Mediterranean Sea. Open-sea or oceanic anchovy populations are genetically different from inshore-water populations within a region (Nei's $\hat{G}_{ST} = 0.035–0.067$), while broadscale geographical variation is weak for each of these two habitat-specific forms ($\hat{G}_{ST} = 0.005–0.006$). Mitochondrial-DNA haplotype frequencies support the distinction between an inshore form and an oceanic form ($\hat{G}_{ST} = 0.067–0.107$), with virtually no genetic differences among oceanic populations across the Gulf of Biscay, the western Mediterranean and the Ionian Sea ($\hat{G}_{ST} = –0.001$). If natural selection on marker loci is unimportant, these results indicate the occurrence of two parapatric, genetically distinct, habitat-specific forms that are widely distributed throughout the Mediterranean Sea. Persistent allele and haplotype-frequency differences between these forms indicate reproductive isolation and the presence of an E. encrasicolus species complex in the Mediterranean. © 2002 The Linnean Society of London, Biological Journal of the Linnean Society 75: 261–269.

ADDITIONAL KEYWORDS: correspondence analysis – habitat specificity – isolation by distance – Mediterranean Sea – systematics.

INTRODUCTION

Because of the economic value of the anchovy, Engraulis encrasicolus L., and the need for efficient management of its fisheries, several genetic surveys designed to investigate stock structure have been carried out on this species, mostly in the Mediterranean Sea. Pasteur & Berrebi (1985) detected allozyme-frequency differences between anchovy populations in different habitats (open sea, brackish lagoon) in the Golfe-du-Lion. Spanakis et al. (1989) detected significant differences between populations in the Aegean Sea and the Ionian Sea. Bembo et al. (1996a) found significant differences between a putative south-central stock and a putative northern stock in the Adriatic Sea. In contrast, Tudela et al. (1999) reported no significant genetic structure (Nei’s $G_{ST} = 0.003$) among oceanic populations over an area, 1000 km wide, in the northern part of the western Mediterranean. These genetic differences are associated with morphological differences in the eastern Mediterranean (Spanakis et al., 1989; Bembo et al., 1996a). Tudela (1999), however, reported morphological differences against a background of genetic homogeneity among anchovy samples from the northern part of the western Mediterranean and suggested that the environment was the main determinant of morphological variation among anchovy populations. Mitochondrial (mt) DNA studies have also been conducted to investigate stock structure. Bembo et al.
detected heterogeneity in mtDNA haplotype frequency among anchovy samples around the Italian Peninsula. In another study, Magoulas et al. (1996) reported that European anchovy mtDNA belonged to one of two distinct haplogroups or phylads (A, B), distinguished from each other by 3.7% nucleotide divergence. Magoulas et al. (1996) suggested that the two phylads evolved in geographical isolation from each other and that their present coexistence in the Mediterranean Sea resulted from secondary contact. Significant heterogeneity is found in the distribution of the two haplogroups across the Mediterranean (Magoulas et al., 1996).

Such levels of genetic variation among populations are rare in the pelagic realm where, as a rule, only small genetic differences are detected across oceans (e.g. see Grant & Utter, 1984; Grant, 1984; Ward, 1995; Gold & Richardson, 1998; Grant & Bowen, 1998; Graves, 1998). The case of E. encrasicolus, therefore, is of potential interest to study the patterns and processes of genetic differentiation in pelagic fishes. However, so far no clear geographical picture of genetic variation is available in Atlantic-Mediterranean anchovy.

In this paper, the available morphological, allozyme, and mtDNA data for anchovy to: (1) further explore the patterns of genetic variation at a supra-regional scale, and (2) examine its systematic status are re-examined. The possibility of geographical patterns in genetic differentiation among populations was tested, and morphological and mtDNA correlates of allozyme-frequency variation searched for.

**METHODS**

Allozyme frequency data on E. encrasicolus were compiled from several sources in the literature (Pasteur & Berrebi, 1985; Spanakis et al., 1989; Bembo et al., 1996a,b; Tudela et al., 1999). The total data set was heterogeneous, with numbers of polymorphic loci scored (r) varying from two (Bembo et al., 1996a) to eight (Bembo et al., 1996b; Tudela et al., 1999). In this study, the locus is considered to be polymorphic when the frequency of the most common electromorph in the total sample was <0.95 in at least one sample. The geographical coverage of any study was generally restricted to a particular region: the Golfe-du-Lion (Pasteur & Berrebi, 1985; r = 6, N = 5, where N = number of samples); the Ionian and Aegean Seas (Spanakis et al., 1989; r = 4, N = 8); the Adriatic Sea (Bembo et al., 1996a; r = 2, N = 36); and the northern part of the western Mediterranean (Tudela et al., 1999; r = 8, N = 6). Only Bembo et al. (1996b; r = 8, N = 13) considered a range of samples, from the Gulf of Biscay in the north-eastern Atlantic and across the Mediterranean, but with only one sample from the western Mediterranean. The largest-possible combined data set, in terms of both geographical range (N = 24; Fig. 1) and number of polymorphic loci scored in common (r = 3; Table 1) combines data from Pasteur & Berrebi (1985), Bembo et al. (1996b) and Tudela et al. (1999).

Tudela et al. (1999) sampled one region, the Tyrrhenian Sea, in common with Bembo et al. (1996b), and the Golfe-du-Lion, in common with Pasteur & Berrebi (1985). Identities of the most common electro-
morphs across these three studies were established on the basis of similarity in relative electrophoretic mobility and regional frequencies. Identities of rarer electromorphs were uncertain, all faster (respectively, slower) migrating electromorphs at a locus were collapsed into a single ‘F’ (resp. ‘S’) class (Table 1). No diagnostic locus was found for any paired sample, either from Table 1, or from any of the allozyme surveys considered in this review.

Correspondence analysis (CA; Benzécri, 1982) is an ordination technique based on eigenanalysis of a contingency table, that differs from principal component analysis by deriving both row and column-axes simultaneously, and by using the χ²-distance metric instead of the Euclidean distance. CA was performed on matrices of allozyme frequencies per sample using the AFC procedure implemented in BIOMECO (Lebreton et al., 1990). Guinand (1996) showed that in this case the eigenvalues of each CA’s axis can be used as estimates of genetic differentiation between populations. CA allows the graphical representation of multiple–locus associations among samples through clustering in hyperspace. The CA algorithm allows the use of ‘supplementary elements’. This are either rows or columns of the data matrix, which are not taken into account for the calculation of the eigenvalues of an axis (unlike ‘active elements’), but which nevertheless can be placed in the graphical output. Supplementary elements are therefore useful as reference marks.

CA was also used for determining whether the

<table>
<thead>
<tr>
<th>Source</th>
<th>Sample</th>
<th>N</th>
<th>Electromorph</th>
</tr>
</thead>
<tbody>
<tr>
<td>PB85</td>
<td>Sè</td>
<td>32.3</td>
<td>0.05 0.85 0.10 0.23 0.77 0.00 0.09 0.91 0.00</td>
</tr>
<tr>
<td></td>
<td>Gr</td>
<td>38.7</td>
<td>0.12 0.87 0.01 0.25 0.75 0.00 0.00 1.00 0.00</td>
</tr>
<tr>
<td></td>
<td>Th</td>
<td>27.7</td>
<td>0.46 0.52 0.02 0.33 0.67 0.00 0.00 1.00 0.00</td>
</tr>
<tr>
<td></td>
<td>Ar</td>
<td>27.0</td>
<td>0.46 0.54 0.00 0.39 0.61 0.00 0.00 1.00 0.00</td>
</tr>
<tr>
<td></td>
<td>Ma</td>
<td>24.7</td>
<td>0.53 0.47 0.00 0.28 0.72 0.00 0.00 1.00 0.00</td>
</tr>
<tr>
<td>BE96</td>
<td>T1</td>
<td>50.0</td>
<td>0.47 0.52 0.01 0.24 0.76 0.00 0.03 0.97 0.00</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>48.7</td>
<td>0.42 0.54 0.04 0.24 0.76 0.00 0.00 1.00 0.00</td>
</tr>
<tr>
<td></td>
<td>A1</td>
<td>50.0</td>
<td>0.22 0.73 0.05 0.16 0.84 0.00 0.01 0.99 0.00</td>
</tr>
<tr>
<td></td>
<td>A2</td>
<td>50.0</td>
<td>0.22 0.76 0.02 0.22 0.78 0.00 0.01 0.97 0.02</td>
</tr>
<tr>
<td></td>
<td>V1</td>
<td>50.0</td>
<td>0.17 0.82 0.01 0.23 0.77 0.00 0.00 1.00 0.00</td>
</tr>
<tr>
<td></td>
<td>V2</td>
<td>49.3</td>
<td>0.17 0.78 0.05 0.30 0.70 0.00 0.00 1.00 0.00</td>
</tr>
<tr>
<td></td>
<td>I1</td>
<td>50.0</td>
<td>0.12 0.83 0.05 0.22 0.78 0.00 0.10 0.90 0.00</td>
</tr>
<tr>
<td></td>
<td>I2</td>
<td>50.0</td>
<td>0.23 0.77 0.00 0.14 0.86 0.00 0.00 1.00 0.00</td>
</tr>
<tr>
<td></td>
<td>S1</td>
<td>50.0</td>
<td>0.18 0.79 0.03 0.24 0.76 0.00 0.20 0.80 0.00</td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td>49.7</td>
<td>0.14 0.82 0.04 0.29 0.71 0.00 0.18 0.81 0.01</td>
</tr>
<tr>
<td></td>
<td>Ty</td>
<td>50.0</td>
<td>0.08 0.88 0.04 0.22 0.78 0.00 0.00 1.00 0.00</td>
</tr>
<tr>
<td></td>
<td>Ae</td>
<td>44.0</td>
<td>0.36 0.62 0.02 0.18 0.81 0.01 0.14 0.85 0.01</td>
</tr>
<tr>
<td></td>
<td>Bi</td>
<td>39.3</td>
<td>0.13 0.81 0.06 0.00 1.00 0.00 0.00 0.97 0.00</td>
</tr>
<tr>
<td>TU99</td>
<td>TA</td>
<td>100.0</td>
<td>0.12 0.84 0.04 0.31 0.68 0.01 0.04 0.96 0.00</td>
</tr>
<tr>
<td></td>
<td>GR</td>
<td>99.3</td>
<td>0.15 0.82 0.03 0.26 0.74 0.00 0.04 0.96 0.00</td>
</tr>
<tr>
<td></td>
<td>GL1</td>
<td>100.0</td>
<td>0.12 0.85 0.05 0.36 0.64 0.00 0.01 0.98 0.01</td>
</tr>
<tr>
<td></td>
<td>GL2</td>
<td>100.0</td>
<td>0.14 0.81 0.05 0.26 0.74 0.00 0.01 0.98 0.01</td>
</tr>
<tr>
<td></td>
<td>EL</td>
<td>99.7</td>
<td>0.10 0.87 0.03 0.26 0.74 0.00 0.02 0.98 0.00</td>
</tr>
<tr>
<td></td>
<td>LI</td>
<td>100.0</td>
<td>0.08 0.90 0.02 0.27 0.73 0.00 0.02 0.97 0.01</td>
</tr>
</tbody>
</table>

100Most common electromorph at a locus; Ffaster and Sslower compound electromorphs.

© 2002 The Linnean Society of London, Biological Journal of the Linnean Society, 2002, 75, 261–269
sample of mtDNA haplotypes analysed by Bembo et al. (1995) was heterogeneous, to eventually allow comparisons of the mtDNA data of these authors with those of Magoulas et al. (1996). Bembo et al. (1995) characterized individual mtDNA haplotypes by amplifying a 2.5-kb fragment of the ND5/6 gene using polymerase chain reaction (PCR) and cutting it using restriction enzymes. Table 1 in their study presents 53 composite haplotypes (actually 51, because two of these appear to have been counted twice) defined by restriction profiles for six enzymes. For each enzyme, each restriction profile was assigned a value of one (observed) or zero (not observed). The matrix of haplotypes × restriction profiles was then subjected to CA using the APE procedure of BIOMECO.

\[ G_{ST}, \text{Nei's (1973) parameter for the apportion of genetic diversity within and across subpopulations} \]
\[ (k = 1 \text{ to } s) \text{ was estimated over polymorphic loci as} \]
\[ G_{ST} = 1 - (\hat{H}_S/\hat{H}_T), \text{ where} \hat{H}_S \text{ equals the average gene diversity across loci (} j = 1 \text{ to } r \text{) in subpopulation} \]
\[ k (\hat{H}_S = 1/r \cdot \Sigma \hat{h}_{jk}), \text{ and} \hat{H}_T \text{ equals average gene diversity across loci in the total population} \]
\[ (\hat{H}_T = 1/r \cdot \Sigma \hat{h}_{jT}). \]

The gene diversities at a locus were estimated across loci in the total population (\( \hat{h}_{jk} = 2\Sigma x_{ijk}^2/(2n_j - 1) \cdot (1 - \Sigma x_{ijk}^2) \) and \( \hat{h}_{jT} = 2\Sigma x_{ij}^2/(2\Sigma x_{ij} - 1) \cdot (1 - 1/\Sigma x_{ij}^2 - \Sigma x_{ij} \cdot x_{ij}^2) \), where \( n_j \) was sample size. Sample-pairwise genetic distance estimates \( [\hat{G}_{ST}/(1 - \hat{G}_{ST}); \text{Rousset, 1997}] \) were plotted against geographical distances to test isolation by distance. The lack of significant correlation between genetic distance and geographical distance would lead to the rejection of the isolation-by-distance model.

**RESULTS**

Two main clusters of samples appeared in the correspondence analysis of compound electromorph frequencies (Table 1) of *E. encrasicolus* samples from the Gulf of Biscay, the northern part of the Western Mediterranean, the Tyrrenhenian Sea, Sicily, the Adriatic Sea, the Ionian sea and the Aegean Sea (Fig. 2). One cluster, ‘Group I’, included all samples from coastal habitats, from either brackish lagoons in the Golfe-du-Lion (samples Ar, Ma, and Th) or the extreme north of the Adriatic Sea (samples TI, T2). A second cluster (‘Group II’) comprised samples exclusively from the open-sea habitat (i.e. samples Bi, Gr, A1, A2, I1, Ty, V1, V2, EL, GL1, GL2, GR, LI, TA of Fig. 2). We included two samples from Sicily (S1, S2) together with sample Ae from the northern Aegean Sea into a third ensemble (‘Group III’). It was not clear whether samples S6 (Golfe-du-Lion) and II (Ionian Sea) belonged to Group II, or were slightly intermediate between Group II and Group III. Genetically intermediate samples may consist of mixtures of individuals from each group, or of hybrids. This could be tested from genotype counts: an excess of heterozygotes would suggest the presence of hybrids, and heterozygote deficiency would suggest Wahlund’s effect. Unfortunately, individual genotype data were not provided in any of the source articles to test these hypotheses. It is also possible that these samples belong to different genetic demes, or that the relative dispersion of samples in Group II reflect stochastic sampling variation. To check for internal stability of the clusters, CA was rerun on a portion of the data, as advised by Greenacre (1984). Thus considering a part of the data set [for instance, either all data of Pasteur & Berrebi (1985), or those of Tudela et al. (1999)] not as active, but as supplementary samples in the CA did not alter its outcome, only slightly diminishing the inertias of the two first axes (not shown). Importantly, Groups I-III defined as above remained.

Sample-pairwise genetic distances were plotted against the corresponding geographical distances (Fig. 3). Even though Mantel’s randomization test on the total matrix of pairwise data indicated a slight increase of genetic distance with geographical distance (\( P < 0.05 \)), suggesting an effect of isolation-by-distance, pairwise comparisons of samples from different groups (open circles in Fig. 3) formed a cluster distinct from paired comparisons within Groups I, II, or III (filled circles). Hence, most of the genetic variability shown in Figure 2 cannot be explained by isolation by distance. The following results and discussion will focus mainly on the distinction between Group I and Group II anchovies.
The results of CA inform us about data structure. To make inferences regarding population structure requires a confirmatory analysis. The distinction between Groups I and II was confirmed by a second CA on an independent dataset from the Adriatic Sea, consisting of 29 additional samples characterized at two polymorphic loci (Bembo et al., 1996a); six of a total of 35 samples analysed by these authors (Ancona 5/93 and 9/93, Trieste 5/93 and 9/93, and Vieste 4/93 and 9/93) were excluded from the analysis because they appeared to be the same as samples A1 and A2, T1 and T2, and V1 and V2, respectively, in Bembo et al. (1996b), which were already used for Figure 2. Full (resp. open) circles designate pairs of samples from a single (resp. two different) Group(s).

The results of CA inform us about data structure. To make inferences regarding population structure requires a confirmatory analysis. The distinction between Groups I and II was confirmed by a second CA on an independent dataset from the Adriatic Sea, consisting of 29 additional samples characterized at two polymorphic loci (Bembo et al., 1996a); six of a total of 35 samples analysed by these authors (Ancona 5/93 and 9/93, Trieste 5/93 and 9/93, and Vieste 4/93 and 9/93) were excluded from the analysis because they appeared to be the same as samples A1 and A2, T1 and T2, and V1 and V2, respectively, in Bembo et al. (1996b), which were already used for Figure 2. Full (resp. open) circles designate pairs of samples from a single (resp. two different) Group(s).

The results of CA inform us about data structure. To make inferences regarding population structure requires a confirmatory analysis. The distinction between Groups I and II was confirmed by a second CA on an independent dataset from the Adriatic Sea, consisting of 29 additional samples characterized at two polymorphic loci (Bembo et al., 1996a); six of a total of 35 samples analysed by these authors (Ancona 5/93 and 9/93, Trieste 5/93 and 9/93, and Vieste 4/93 and 9/93) were excluded from the analysis because they appeared to be the same as samples A1 and A2, T1 and T2, and V1 and V2, respectively, in Bembo et al. (1996b), which were already used for Figure 2. Full (resp. open) circles designate pairs of samples from a single (resp. two different) Group(s).

The results of CA inform us about data structure. To make inferences regarding population structure requires a confirmatory analysis. The distinction between Groups I and II was confirmed by a second CA on an independent dataset from the Adriatic Sea, consisting of 29 additional samples characterized at two polymorphic loci (Bembo et al., 1996a); six of a total of 35 samples analysed by these authors (Ancona 5/93 and 9/93, Trieste 5/93 and 9/93, and Vieste 4/93 and 9/93) were excluded from the analysis because they appeared to be the same as samples A1 and A2, T1 and T2, and V1 and V2, respectively, in Bembo et al. (1996b), which were already used for Figure 2. Full (resp. open) circles designate pairs of samples from a single (resp. two different) Group(s).

The results of CA inform us about data structure. To make inferences regarding population structure requires a confirmatory analysis. The distinction between Groups I and II was confirmed by a second CA on an independent dataset from the Adriatic Sea, consisting of 29 additional samples characterized at two polymorphic loci (Bembo et al., 1996a); six of a total of 35 samples analysed by these authors (Ancona 5/93 and 9/93, Trieste 5/93 and 9/93, and Vieste 4/93 and 9/93) were excluded from the analysis because they appeared to be the same as samples A1 and A2, T1 and T2, and V1 and V2, respectively, in Bembo et al. (1996b), which were already used for Figure 2. Full (resp. open) circles designate pairs of samples from a single (resp. two different) Group(s).

The results of CA inform us about data structure. To make inferences regarding population structure requires a confirmatory analysis. The distinction between Groups I and II was confirmed by a second CA on an independent dataset from the Adriatic Sea, consisting of 29 additional samples characterized at two polymorphic loci (Bembo et al., 1996a); six of a total of 35 samples analysed by these authors (Ancona 5/93 and 9/93, Trieste 5/93 and 9/93, and Vieste 4/93 and 9/93) were excluded from the analysis because they appeared to be the same as samples A1 and A2, T1 and T2, and V1 and V2, respectively, in Bembo et al. (1996b), which were already used for Figure 2. Full (resp. open) circles designate pairs of samples from a single (resp. two different) Group(s).

The results of CA inform us about data structure. To make inferences regarding population structure requires a confirmatory analysis. The distinction between Groups I and II was confirmed by a second CA on an independent dataset from the Adriatic Sea, consisting of 29 additional samples characterized at two polymorphic loci (Bembo et al., 1996a); six of a total of 35 samples analysed by these authors (Ancona 5/93 and 9/93, Trieste 5/93 and 9/93, and Vieste 4/93 and 9/93) were excluded from the analysis because they appeared to be the same as samples A1 and A2, T1 and T2, and V1 and V2, respectively, in Bembo et al. (1996b), which were already used for Figure 2. Full (resp. open) circles designate pairs of samples from a single (resp. two different) Group(s).

The results of CA inform us about data structure. To make inferences regarding population structure requires a confirmatory analysis. The distinction between Groups I and II was confirmed by a second CA on an independent dataset from the Adriatic Sea, consisting of 29 additional samples characterized at two polymorphic loci (Bembo et al., 1996a); six of a total of 35 samples analysed by these authors (Ancona 5/93 and 9/93, Trieste 5/93 and 9/93, and Vieste 4/93 and 9/93) were excluded from the analysis because they appeared to be the same as samples A1 and A2, T1 and T2, and V1 and V2, respectively, in Bembo et al. (1996b), which were already used for Figure 2. Full (resp. open) circles designate pairs of samples from a single (resp. two different) Group(s).

The results of CA inform us about data structure. To make inferences regarding population structure requires a confirmatory analysis. The distinction between Groups I and II was confirmed by a second CA on an independent dataset from the Adriatic Sea, consisting of 29 additional samples characterized at two polymorphic loci (Bembo et al., 1996a); six of a total of 35 samples analysed by these authors (Ancona 5/93 and 9/93, Trieste 5/93 and 9/93, and Vieste 4/93 and 9/93) were excluded from the analysis because they appeared to be the same as samples A1 and A2, T1 and T2, and V1 and V2, respectively, in Bembo et al. (1996b), which were already used for Figure 2. Full (resp. open) circles designate pairs of samples from a single (resp. two different) Group(s).
Focusing on Groups I and II in the Golfe-du-Lion and Adriatic Sea regions, $G_{ST}$ intra-Group across regions = 0.005–0.006 ($P = 0.050$); $G_{ST}$ inter-Groups in a region = 0.035–0.067 ($P < 0.001$). [Null-hypothesis probabilities of occurrence of $G_{ST}$-values larger or equal to the observed value were estimated by permutation tests (GENETIX; Belkhir et al., 1996).]

Significant heterogeneity in mtDNA haplotype frequency among anchovy populations across the Mediterranean has been reported (Bembo et al., 1995; Magoulas et al., 1996). Five (Trieste, Tyrrhenian, Ionian, Sicily, Aegean) of the seven samples (adding Ancona and Vieste) of Bembo et al. (1995) are useful to the present work as each can be assigned with confidence to one of the populations whose samples (respectively, $T1$ or $T2$, Ty, $I1$ or $I2$, $S1$ or $S2$, and $Ae$) were characterized at allozyme loci by Bembo et al. (1996b), hence ascribed to either Group I, II, or III. Samples Ancona and Vieste could both belong to either Group I or Group II depending on the sampling date (Bembo et al., 1996a; this Fig. 4). As sampling dates have not been given in Bembo et al. (1995), these two samples cannot be used for comparisons with allozymes. The 51 composite mtDNA haplotypes detected in 140 individuals from all seven Mediterranean samples clustered into two distinct clusters when subjected to CA (Fig. 5). One cluster spanned a small area on the plane defined by factorial axes one and two and comprised 21 haplotypes. This contributed 95% of the haplotypes found in the northern Aegean Sea. The other cluster consisted of a diffuse ensemble of 30 haplotypes, to which belonged 80% of the haplotypes found in the northern Adriatic (Trieste). The wide dispersion of the points in this cluster means that the molecular relationship of the haplotypes was less than within the other cluster. From these observations it can be deduced that the first cluster corresponds to phylad A of Magoulas et al. (1996) that is dominant in the Black Sea (99%) and the Aegean Sea (85%) and also exhibits a star-like phylogeny. The second cluster corresponds to phylad B that is dominant in the northern Adriatic Sea (86%) and also exhibits a more diffuse phylogeny. Three groups of samples can be distinguished in the study of Bembo et al. (1995), as in Magoulas et al. (1996), according to the respective proportions of A and B haplotypes. One group included all three Adriatic Sea samples (Trieste, Ancona, Vieste) where the frequency of A was $\leq 25\%$. A second group included samples from the Tyrrhenian Sea, the Ionian Sea, and Sicily, with 50%, 45%, and 45% A haplotypes, respectively. A third group corresponded to the northern Aegean Sea (95% A). From the restricted information available, it appears that Group I (sample Trieste) harbours a majority of haplogroup-B mitochondria, while approximately 50% of mitochondria in Group II (samples Tyrrhenian, Ionian) are of the A type. Group III samples were variable, with Sicily having 45% A mitochondria, while Aegean reached 95%.

Combining the results of Magoulas et al. (1996) with those of Bembo et al. (1995; present re-analysis) shows that Group II anchovy have similar haplotype frequencies in the Gulf of Biscay (40% A; $N = 48$), the Golfe-du-Lion (42% A; $N = 50$), the Tyrrhenian sea (50% A; $N = 20$), and the Ionian Sea (45% A; $N = 20$). This translates into $G_{ST} = -0.001$. Using either the northern Adriatic sample of Magoulas et al. (1996) (14% A; $N = 65$), or the Trieste sample of Bembo et al. (1995) (20% A; $N = 20$) as representing Group I anchovy yielded $G_{ST} = 0.067–0.107$ between the two groups. This result appeared to closely match allozymes.

**DISCUSSION**

Local variation in morphology has been documented in Clupeoids, including *E. encrasicolus* (Lee & Juge, 1965; Quignard et al., 1973; Spanakis et al., 1989, and references therein; Bembo et al., 1996a; Tudela, 1999). Although various subspecies of *E. encrasicolus* has been distinguished on the basis of morphology, they have not been retained in the current taxonomy of anchovies (Whitehead et al., 1984; Spanakis et al., 1989). Population genetic surveys have also yielded evidence of unusually high levels of genetic heterogeneity in this pelagic fish (Pasteur & Berrebi, 1985; Spanakis et al., 1989; Bembo et al., 1995, 1996a; 1996b; Magoulas et al., 1996; Grant et al. in Grant & Bowen, 1998). None of these studies, however, have raised the possibility that *E. encrasicolus* may consist of genetically distinct forms that occupy different habitats.

Allozyme frequency differences between anchovy samples collected at regular intervals over a 2-year period throughout the Adriatic yielded evidence for separate, temporally stable ‘stocks’ (Bembo et al., 1996a; Carvalho & Hauser, 1998). The geographical limit between stocks, as delineated by a drop in electromorph frequency for locus IDHP-2, coincided with a change in hydrology between shallow (<50 m) waters of the northern Adriatic and the deep sea waters of the southern-central Adriatic (Bembo et al., 1996a). Remarkably, this boundary also corresponded to another delineation based on G3PDH-2 electromorph frequencies (Bembo et al., 1996a). Although the northern Adriatic region is in geographical continuity with the open sea, hydrological data show that this region is a distinct, wide shallow area of brackish water. The northern Adriatic region is strongly affected by variation in freshwater discharge from rivers and is subjected to variable climatic forcing, both related to strong spatial and temporal variability of trophic conditions (Fonda-Umani et al., 1992; Vichi et al., 1999).
of reproductive isolation between the two groups, which, therefore, correspond to biological species. The maintenance of Groups I and II as distinct species may not be due to reproductive traits per se, but perhaps to ecological adaptations that select disruptively against intermediates or hybrids (Mallet, 1995) on either side of the hydrographic boundary between inshore waters and oceanic waters.

Group I (inshore) and Group II (oceanic) anchovies can be distinguished by their morphology (Bembo et al., 1996a; Fig. 4), and also have different colour patterns, different counts of vertebrae, different maximal lengths, and different growth rates as inferred from otoliths (Lee & Juge, 1965; Quignard, 1978; Levi et al., 1994; Bembo et al., 1996a). The morphological distinction between a lagoonal form and an offshore form has also been noted in Tunisia (Quignard et al., 1973) but no sample from this region has ever been analysed using allozymes. Currently, it is not possible to determine the genetic affinities of Tunisian anchovies to Groups I-III defined from Figure 1.

The mtDNA data available to date are not sufficient to investigate whether haplotype frequencies are geographically homogeneous for each habitat-specific form. This was suggested, however, for Group II anchovies when combining the results of Magoulas et al. (1996) with those of Bembo et al. (1995). MtDNA data also confirm that northern Aegean Sea and Sicily anchovy belong to different populations. The coexistence of three differentiated groups of populations (according to both allozyme and mtDNA haplotype frequencies) but only two mitochondrial clades suggests that the evolution of mitochondrial lineages did not parallel that of the nuclear gene pools. Such a complex pattern raises the possibility that previously geographically isolated populations have undergone secondary contact and introgressed (Magoulas et al., 1996), but only partly. Alternatively, one may hypothesize that mtDNA diversity reflects ancestral variation that was apportioned differently between demes. Further surveys of genetic variation in European anchovy should be designed to specifically address these issues.

In conclusion, the present review highlights the usefulness of broadscale geographical surveys of genetic variation for interpreting patterns of genetic structure at a regional scale. The emerging result was that E. encrasicolus appears to consist of at least two biological species (Group I, Group II) that are both present in the Golfe-du-Lion and in the Adriatic Sea. Group I anchovy occupy the inshore habitat of the two regions, while Group II anchovy are found offshore in the oceanic waters of the Biscay Gulf, the western Mediterranean, the central and southern Adriatic Sea, and the Ionian Sea. I speculate that the form sampled in the northern Aegean Sea and in Sicily (Group III)
might belong to a third species, as preliminarily suggested from morphometric, allozymic and mtDNA data (Spanakis et al., 1989; this review). Further population studies of anchovy warrant further insight into the questions of geographical differentiation and genetic architecture in pelagic fishes in relation to habitat specialization and historic events, as extensively debated in recent years (e.g. Bowen & Avise, 1990; Miya & Nishida, 1997; Grant & Bowen, 1998; Graves, 1998; Turan et al., 1998; Nesbo et al., 2000). From an applied perspective, updated taxonomy and reassessment of population structure within each species will be indispensable tools for the management of anchovy fisheries in the Mediterranean Sea and perhaps elsewhere in the eastern Atlantic.

ACKNOWLEDGEMENTS

Thanks to F. Bonhomme, V. Castric, W. S. Grant, D. Ponton and J.-P. Quignard who made insightful comments and suggestions for this article. W. S. Grant also offered his help in editing a former manuscript. This is a contribution of Laboratoire Génome Populations Interactions (UMR 5000 CNRS) and UR081/DRV (‘Génome / Population / Environnement chez les poissons tropicaux’) of Institut de Recherche pour le Développement.

REFERENCES


