Geographic structure of European anchovy: a nuclear-DNA study

Yanis Bouchenak-Khelladi, Jean-Dominique Durand, Antonios Magoulas, Philippe Borsa

To be cited as:
Bouchenak-Khelladi Y., Durand J.-D., Magoulas A., Borsa P. 2008. – Geographic structure of European anchovy: a nuclear-DNA study. Journal of Sea Research 59, 269-278 [doi:10.1016/j.seares.2008.03.001]

Geographic structure of European anchovy: a nuclear-DNA study

Yanis Bouchenak-Khelladi [a], Jean-Dominique Durand [b], Antonios Magoulas [c], Philippe Borsa [d,*]

[a] Department of Botany, Trinity College, Dublin, Ireland
[b] Institut de recherche pour le développement, Dakar, Senegal
[c] Institute of Marine Biology and Genetics, Hellenic Centre for Marine Research, Heraklion, Greece
[d] Institut de recherche pour le développement, Nouméa, New Caledonia

* Correspondence : IRD, Unité de recherche 128 "Biocomplexité des écosystèmes coralliens", BP A5, 98848 Nouméa, New Caledonia; *E-mail:* philippeborsa@yahoo.fr; *Tel.:* +687 260741; *Fax:* +687 264326

Abstract

Atlantic-Mediterranean anchovies were genetically characterized at two polymorphic nuclear loci (intron 6 of two creatine-kinase genes) and compared to reference *Engraulis albidus* and *E. encrasicolus* samples from the northern Western Mediterranean to provide new insights into their geographic structure. Northeastern Atlantic anchovy, represented by one sample from the Canary archipelago and one sample from the Alboran Sea, were genetically distinct from Mediterranean *E. encrasicolus* (Weir and Cockerham's $\hat{\theta}$=0.027-0.311), indicating geographic isolation from either side of the Almería-Oran oceanographic front. Generally smaller genetic differences were evident among anchovy populations from different sub-basins in the Mediterranean ($\hat{\theta}$=-0.019-0.116), the genetic differences between Black Sea and Ionian Sea / Aegean Sea anchovies being the strongest ($\hat{\theta}$=0.002-0.116). There was no evidence of the presence of *E. albidus* in our samples outside Camargue (northern shore of the Western Mediterranean). However, a sample from the southern Western Mediterranean appeared to be genetically intermediate between *E. albidus* and Mediterranean *E. encrasicolus*, indicating possible hybridization. Anchovy from the Benguela current system off southern Africa possessed allele frequencies characteristic of *E. albidus* at one locus and Northeastern Atlantic anchovy at the other locus, suggesting past introgression.

Keywords: Engraulis albidus; *E. encrasicolus*; Atlantic-Mediterranean divide; Intron-length polymorphism; Hybrid swarm; Introgression

1. Introduction

Marine species are often large-scale dispersers and usually broadcast spawners, two traits that promote high levels of gene flow in a habitat that is generally assumed to be dispersive. Potential barriers to gene flow in temperate marine species include the equatorial-tropical divide that separates the northern- and southern-hemisphere waters (Briggs, 1995), oligotrophic areas of the open ocean which are ecological deserts (Park, 1994), hydrological fronts which are the physical boundaries between different water masses (Tintore et al., 1988), and zones of upwelling. Gene flow may also be hampered by intrinsic barriers, such as assortative mating, ecological segregation (Dieckmann and Doebeli, 1999), gamete incompatibility (Palumbi, 1994) or hybrid inferiority (Vamosi and Schluter, 1999).

The European anchovy, a pelagic species, is widely distributed in the northeastern Atlantic and throughout the Mediterranean basin (Whitehead, 1990), and it is also present around the tip of southern Africa (Grant and Bowen, 1998). Unusually strong geographic structure in European anchovy has been noted (Grant, 2005; Magoulas et al., 2006). Potential barriers to gene flow in European anchovy include the warm equatorial waters between the temperate northeastern Atlantic and southern Africa, and the Almería-Oran oceanographic front (AOOF: Tintore et al., 1988) between Atlantic and Mediterranean surface waters. A significant part of the genetic (allozyme) differentiation in the Mediterranean occurs at the within-sub-basin scale [e.g. between the coastal lagoons and the open sea in the northern Western Mediterranean (Pasteur and Berrebi, 1985), or between the brackish waters of the extreme northern Adriatic and the oceanic waters of the central-southern Adriatic (Bembo et al. 1996a)], while allozyme frequencies between populations from different sub-basins may be similar [i.e. northern Western Mediterranean and South-Central Adriatic (Bembo et al., 1996b; Borsa, 2002)]. This apparent contradiction points to reproductive isolation between habitat-specific forms (Borsa, 2002). The European anchovy has thus been proposed to consist of two parapatric species, namely *Engraulis albidus* and *E. encrasicolus* (Borsa et al., 2004). Based on the scarce data available so far, *E. albidus* presumably inhabits the brackish waters and the coastal waters under the influence of river plumes in the northern Western Mediterranean and in the northern Adriatic Sea while *E. encrasicolus* occupies the oceanic waters of the continental shelf. The presence of two cryptic species in the area of distribution of European anchovy warrants the use of nuclear-DNA markers to detect eventual hybrids and / or heterozygote deficiencies that would point to admixture of two reproductively isolated entities.

In the present paper, we used allele-size variation at two intron loci to further characterize European anchovy populations, sampled over most of the species' area of distribution, along the *Engraulis albidus* / *E. encrasicolus* axis of genetic differentiation. In particular, we explored the geographic variability of *E. encrasicolus* populations between the Atlantic and the Mediterranean, and throughout the Mediterranean basin. Our aim was to provide a nuclear-DNA view of genetic variability in European anchovy, to

complement previous studies using mtDNA (Bembo et al., 1995; Magoulas et al., 1996; 2006) and allozymes (reviewed in Borsa, 2002).

2. Materials and methods

2.1. Samples

Seven hundred and thirty five European anchovy were collected between 1989 and 2002 from 13 locations, either coastal or oceanic, in the Atlantic, the Mediterranean, and the Black Sea (Fig. 1 A and B). Anchovy samples from Sète (Lio) in the Golfe-du-Lion, Chioggia (Adr) in the Adriatic Sea, Patras (Pat) in the Ionian Sea, Kavala (Kav) in the Aegean Sea, Varna (Bls) and Batumi (Geo) in the Black Sea, and the Canary archipelago (Can) in the Atlantic Ocean were from the collections of the Institute of Marine Biology of Crete, as concentrated DNA in TE buffer (1 *M* Tris, 10 m*M* EDTA, pH 8.1). Samples Lio, Adr, Pat, Kav, and Bls are sub-samples of, respectively, LION, ADR, PAT1, KAV1 and BLS of Magoulas et al. (1996), and samples Geo and Can are those of Magoulas et al. (2006), all scored for mitochondrial DNA variation. The samples from Mauguio (MAU) in Camargue, Annaba (ANN) in the southern Western Mediterranean, Oran (ORN) in the Alboran Sea, and from the Benguela current sytem (BEN) were preserved in alcohol immediately after collection. A sample from Cul-de-Beauduc in Camargue (CUL), which is the type locality for *Engraulis albidus* (Borsa et al., 2004), was chosen as reference for that species. A sample from off Sète, northern Western mediterranean (SET) was chosen as reference for *E. encrasicolus*. Sub-samples of MAU, BEN, CUL and SET have been screened for genetic variation by Borsa et al. (2004), under the names 'Mauguio', 'Benguela', 'Cul-de-Beauduc' and 'Sète', respectively. Sample 'Sète' included the neotype of *E. encrasicolus* (Borsa et al., 2004) and so did SET.

For samples CUL, SET, MAU, ANN, ORN and BEN, a piece of fin (about 5 mg dry weight) dissected from each individual was used as the source of DNA. DNA was extracted according to the phenol-chloroform protocol (Sambrook et al., 1989). The DNA pellet was then suspended in 150 µL of Ultra-Pure deionized water, and stored at –20 °C until genotyping using polymerase chain reaction (PCR).

2.2. Characterization of intron-length polymorphisms

Primer pairs *CK6-1F* (5' - CGACATTGTAATGATGTTACAATGA- 3') / *CK6-1R* (5' - ATTTCCTTTGGGTTGGCTCTTCTCT- 3') and *CK6-2F* (5' - CTCAGAACTACATACCAAACCAATG- 3') / *CK6-2R* (5'- ACTCACTGTAATTCTGAATAGAGCT- 3') [defined by Borsa et al. (2004)] were chosen to PCR-amplify locus-specific, length-polymorphic introns at loci *CK6-1* and *CK6-2* of the creatine-kinase multigenic family. Note that both *CK6-1 F/R* and *CK6-2 F/R* fragments are nested within the fragment framed by primers *CK6F* and *CK7R* of Chow and Takeyama (1998) [those initially used by Borsa et al. (2004)] which are not locus-specific.

The PCR was done in 10 µL reaction mix, containing 1.2 µL (5-20ng) diluted DNA

extract, 2.2 m*M* MgCl2, 0.2% bovine serum albumin, 0.074 m*M* dNTPs, 0.4 µ*M* forward primer (either *CK6-1F* or *CK6-2F*) labelled with dye 6-FAM (Sigma Genosys, London; absorption peak at 494 nm, re-emission peak at 525 nm), 0.4 µ*M* reverse primer (*CK6-1R* or *CK6-2R*, non-labelled), and 0.25 *U Taq*-polymerase (Promega, Madison WI, USA) in its buffer. The PCR program [comprising one initial denaturation step of 3 min at 94 °C, 35 cycles of denaturation (30 s at 94 °C), hybridization (30 s at 54 °C for locus *CK6-1*, 52°C for locus *CK6-2*), and extension (1 min at 72 °C), and a final extension step of 5 min at 72 °C] was run in a *Crocodile III* thermocycler (Appligène, Strasbourg, France).

Length polymorphism was observed after subjecting the heat-denaturated (for 5 min at 95 °C) PCR products to electrophoresis at 50 W for 3 hours (for locus *CK6-1*) or 6 hours and 30 min (for locus *CK6-2*) in 6% denaturing polyacrylamide gel (29:1 acrylamide/bis-acrylamide; Biorad Labs, Hercules, Canada) in 1X TBE buffer. DNA bands were visualized in an *FMBioII* fluorescence gel scanner (Hitachi Instruments, San José CA, USA) using a filter responding to the fluorescence specificity of dye 6-FAM.

Alleles at a locus were designated by arbitrary numbers increasing with allele size, the most common allele in *Engraulis albidus* being designated "*100*". The lengths of some alleles were obtained from their nucleotide sequences (locus *CK6- 1* allele *050*: 406 base pairs (bp) and allele *100*: 524 bp; locus *CK6- 2* allele *095*: 506 bp, allele *100*: 516 bp, and allele *105*: 525 bp).

2.3. Genetic data analysis

Genetic relationships among samples, each of which was defined by its allelic frequencies at both *CK6*-intron loci, were assessed by correspondence analysis (CA: Benzécri, 1982). CA is an ordination technique based on eigenanalysis of a contingency table, which differs from principal component analysis by deriving both row- and column-axes simultaneously, and by using the χ^2-distance metric instead of the Euclidean distance. CA was performed on matrices of sample allele frequencies using the AFC procedure implemented in BIOMECO (Lebreton et al., 1990). Eigenvalues of each axis were used as estimates of genetic differentiation between populations (Guinand, 1996).

Weir and Cockerham' s (1984) $\hat{f}$ and $\hat{\theta}$ estimates were computed using the FSTATS procedure implemented in GENETIX 4.02 (Belkhir et al., 2000). Parameter f quantifies the departure from Hardy-Weinberg genotype frequencies at a locus for each population, while parameter θ quantifies genetic differentiation between populations. The statistical significances of $\hat{f}$ - and $\hat{\theta}$ were tested by random permutations (procedure PERMUTATIONS of GENETIX 4.02). An estimated value was significantly different from 0 when its probability of occurrence, P, was <0.025 for $|\hat{f}|$ (bilateral test), and <0.05 for $\hat{\theta}$ (unilateral test). P was estimated as $(n+1)/(N+1)$ where n is the number of pseudo-values greater than or equal to the observed value, and N is the number of permutations.

The correlation between genotypes at the two *CK6*-intron loci, *R*ij (Weir, 1979), was used as a measure of genotypic disequilibrium. Random permutations of genotypes at a locus (procedure PERMUTATIONS of GENETIX 4.02) were used to generate a pseudo-distribution of *R*ij under the null hypothesis of genotypic equilibrium.

3. Results

3.1. Size polymorphism at loci CK6-1 and CK6-2

Length polymorphism at each of these two loci was of the Mendelian type, where no more than two DNA bands were observed for an individual, and most of all possible combinations of bands were observed in the total sample. The large size difference between the two most common *CK6-1* alleles, *050* (GENBANK AY486346) and *100* (GENBANK AY486347), was caused by a large (116 bp) insertion/deletion (indel). Length-variation at locus *CK6-2* was due to short indels (1-21 bp) as indicated by the alignment of alleles *095* (GENBANK AY486344), *100* (GENBANK AY486342) and *105* (GENBANK AY486343, GENBANK AY486345). Because of the limited resolution of polyacrylamide gels, alleles of similar sizes were pooled into compound alleles. Allele *c050* at locus *CK6-1* included the original *050* allele and two rarer alleles (*048* and *049*) that were distinct on some gels when run side-by-side but not otherwise. Two compound alleles were considered at locus *CK6-2*: *c102* (including size-alleles 101, *102*, and *103*), and *c105* (including size-alleles *104* and *105*).

At locus *CK6-1*, allele *100* was the most common in all samples (frequency >0.58; Table 1) and its frequency was 0.83 for the total sample. Allele *c050* ranked second, with a frequency of 0.11 in the total sample. All the other ten alleles (*001*, *010*, *030*, *052*, *090*, *095*, *098*, *105*, *110*, *120*) had frequencies ≤0.08 (Table 1). At locus *CK6-2*, the three most common alleles were *100*, *c102* and *c105*, with frequencies of, respectively, 0.44, 0.15, and 0.28 in the total sample. Allele *095*, also present in all samples, had a frequency of 0.06 in the total sample. The 7 other alleles at locus *CK6-2* (*010*, *090*, *097*, *098*, *107, 110* and *150*) were all rarer, with frequencies ≤0.06 (Table 1).

3.2. Hardy-Weinberg equilibrium

After adjustment of the level of type I error by sequential Bonferroni correction, only one ^*f*-value remained significantly different from 0 (for sample SET at locus *CK6-1*), indicating heterozygote deficiency. The null hypothesis of panmixia could not be rejected for any of the other samples, although high ^*f*-values were observed for Lio, Bls and ORN at locus *CK6-1*, and for ANN at locus *CK6-2* (Table 1).

The estimate of *R*ij in ANN was significant and was also the highest of the whole dataset (Table 1). No genotypic disequilibrium was detected in the other populations (Table 1).

3.3. Population geographic structure

Figures 1A and 1B give a preliminary geographic representation of allelic variation in European anchovy, where all alleles at a locus were pooled against its respective allele 100. An in-depth analysis of the genetic relationships among samples is provided by CA. The projection of samples on the plane defined by the first and second axes of CA

revealed genetically distinct groups (Fig. 2). As expected, Axis 1 (35.8% total inertia) separated the reference samples of *Engraulis albidus* (CUL, MAU) from *E. encrasicolus* (SET). All Mediterranean samples, except ANN, were close to SET along axis 1, opposite to a group including the samples from the northeastern Atlantic and the Alboran Sea ({Can and ORN}, hereafter called "Atlantic anchovy"). Axis 2 (23.3% total inertia) separated Atlantic anchovy from *E. albidus*. Axis 2 also separated Black Sea anchovy (samples Bls and Geo) from those from the Aegean Sea (Kav) and the Ionian Sea (Pat). Sample ANN appeared to be genetically intermediate between *E. albidus* and Mediterranean *E. encrasicolus*, whereas sample BEN appeared to be genetically intermediate between *E. albidus* and Atlantic anchovy (Figure 2). Axis 3 (not shown; 11.8% total inertia) confirmed the intermediate positions of samples ANN and BEN. Allele frequencies in ANN were intermediate between those of reference samples for *E. albidus* and Mediterranean *E. encrasicolus* at both loci, while BEN was close to *E. albidus* at *CK6-2* and close to Atlantic anchovy at *CK6-1* (Table 1).

Significant genetic differences were observed between the reference sample of *Engraulis albidus* [CUL, which was identical to MAU (two-locus $\hat{\theta}$ ($\hat{\theta}_{2L}$)=-0.008; P=1.00)] and that for *E. encrasicolus* (SET), sampled only a few tens of kilometers away in the Golfe-du-Lion. Samples CUL and MAU actually formed a distinct group within the total sample ($0.027<\hat{\theta}_{2L}<0.228$ for all comparisons with the other samples, all significant). Much lower $\hat{\theta}$ values were generally observed among all other Mediterranean anchovy samples, which were comparatively very distant geographically. All Mediterranean samples except CUL, MAU and ANN were not significantly different from SET. No genetic differences were detected between the two Black Sea samples, Bls and Geo ($\hat{\theta}_{2L}$=-0.007; P=0.89), as between the samples from the Aegean Sea (Kav) and the Ionian Sea (Pat) ($\hat{\theta}_{2L}$=-0.003; P=0.52). In Mediterranean *E. encrasicolus*, $\hat{\theta}$-values were not significantly different from 0 for the following three pairs of samples: SET/Lio, Kav/Pat and Bls/Geo (Table 2). The samples from the northeastern Atlantic (Can) and the Alboran Sea (ORN), which were genetically close to each other ($\hat{\theta}_{2L}$=0.001; P=0.38), were substantially different from all the other samples ($0.028<\hat{\theta}_{2L}<0.162$, with $0.001<P<0.06$). Overall, significant genetic differences ($\hat{\theta}_{2L}$=0.083-0.084; $P<0.01$) were found between anchovy populations west (sample ORN) and east (Lio and SET) of the AOOF. Samples BEN and ANN appeared differentiated from all the other samples ($0.027<\hat{\theta}_{2L}<0.277$, with $0.001<P<0.01$). This high degree of genetic differentiation generally was more obvious at locus *CK6-2* than at locus *CK6-1* (Table 2).

Samples which were not significantly different between them were pooled in Fig. 2. Kav/Pat pooled with Adr and SET/Lio further pooled with Kav/Pat/Adr on the one side, and with Geo/Bls on the other side (Fig. 2). In summary, three groups of samples [*E. albidus* (CUL/MAU), Mediterranean *E. encrasicolus* (SET/Lio/Kav/Pat/Bls/Geo), Atlantic anchovy (Can/ORN)], and two single samples with intermediate positions (BEN, ANN) were observed.

4. Discussion

4.1. Geographic isolation of Atlantic vs. Mediterranean E. encrasicolus

The hydrological characteristics of the surface waters of the Alboran Sea are much closer to those of the northeastern Atlantic than those of the Western Mediterranean, from which they are separated by the Almería-Oran oceanographic front (AOOF; Tintore et al., 1988). The AOOF may act as a barrier to the dispersion of larvae and even juvenile and adult anchovy. As expected under this scenario, the European anchovy population of Oran, in the Alboran Sea, was genetically close to that of the Canary archipelago while both populations showed significant allele-frequency differences with Western Mediterranean *Engraulis encrasicolus*. The geographic isolation between anchovy populations on either side of the AOOF is similar in intensity to that observed in Mediterranean mussel, *Mytilus galloprovincialis* (Quesada et al., 1995; Daguin and Borsa, 1999), Montagu's barnacle, *Chtamalus montagui* (Pannacciulli et al., 1997), seabass, *Dicentrarchus labrax* (Naciri et al., 1999), and sharpsnout seabream, *Diplodus puntazzo* (Bargelloni et al., 2005). It is possible that the AOOF coincides with a phylogeographic boundary (Borsa et al., 1997; Patarnello et al., 2007) corresponding to the location of secondary contact between formerly geographically isolated populations. Similar contact zones have been suspected elsewhere (Cunningham and Collins, 1994; Burton, 1998). Lending support to the latter hypothesis, a recent reanalysis of Atlantic / Mediterranean sea-bass population structure raised the prospect of a hybrid zone coinciding with the AOOF (Lemaire et al., 2005).

4.2. Geographic structure: Mediterranean E. encrasicolus

Temporally unstable, genetic heterogeneities on microgeographic scales have been ascribed to temporal variation in the genetic composition of anchovy (*Engraulis mordax*) recruits (Hedgecock et al., 1994), underlining the importance of temporal replicates in population genetic surveys of coastal pelagic fishes (Hedgecock, 1994; Carvalho and Hauser, 1998). Here, two samples (Lio and SET) were collected within the same area at an 8-year interval, spanning up to 5 generations if one considers a generation time of two years (Karaçam and Düzgünes, 1990). No differences in allele frequencies were detected between them, suggesting that allele frequencies in the *Engraulis encrasicolus* population of the northern Western Mediterranean are temporally stable, and enhancing the geographical significance of the genetic variation observed in *E. encrasicolus* across the Mediterranean.

The Sète population in the Western Mediterranean was found to be neither different from the Black Sea population nor the Aegean Sea population, while the latter two, which are geographically adjacent, were genetically distinct. The complex patterns of geographic structure inferred from the *CK6-1* and *CK6-2* data within the Mediterranean (present results) thus confirmed allozyme (Spanakis et al., 1989) and mtDNA results (Magoulas et al., 2006; Fig. 1C). Nevertheless, it is intriguing to observe that sample Adr

sampled in the northern Adriatic Sea was much closer to reference *E. encrasicolus* than *E. albidus* at loci *CK6-1* and *CK6-2*. Sample Adr is dominated by lineage-*B* mtDNA (Magoulas et al., 1996) as are other populations sampled in the northern Adriatic Sea, which were inferred to be *E. albidus* based on allozyme data (Borsa, 2002). As interlocus variance in introgression rate is an expected phenomenon in hybrid populations (Harrison, 1990; Barton and Gale, 1993), we speculate that these seemingly contradictory patterns may reflect differential introgression of the gene pool of either *E. albidus* or *E. encrasicolus* by genes of the other species. Further studies of genetic variation of anchovy in the Adriatic, based on an adequate sampling design and more genetic markers, should allow a future test of this hypothesis. Also, given the lack of precision concerning the actual sampling location of Adr (Magoulas et al., 1996), one cannot exclude that it was fished in the deeper, offshore waters of the Central Adriatic where anchovy stocks are typical *E. encrasicolus* (Borsa, 2002), rather than in the inshore waters of the Northern Adriatic.

4.3. Hybridization between genetically isolated populations?

Assuming that data from loci *CK6-1* and *CK6-2* reflect nuclear-genome wide structure, the intermediate genetic position of sample ANN between *Engraulis albidus* and Mediterranean *E. encrasicolus* may reflect either (1) the admixture in the sample of individuals of the two species, or (2) F1 between the two species, or (3) an established hybrid population or 'hybrid swarm' (Arnold, 1993). Depending on the hypothesis, different effects on genotype frequencies are expected. In the first case (admixture), heterozygote deficiency by Wahlund effect is expected; in the second case (F1 hybrids), heterozygote excesses are expected; in the last case (hybrid swarm), Hardy-Weinberg equilibrium frequencies are expected at a locus, but residual genotypic disequilibrium may be observed if the two marker loci are linked and the formation of the hybrid population is sufficiently recent. The results supported Hypothesis 3, suggesting that sample ANN could be from a hybrid population recently formed between *E. albidus* and *E. encrasicolus*. However, we cannot reject Hypothesis 1 either, because from all the populations surveyed, the Annaba population also had the highest level of heterozygote deficit at *CK6-2* which is the most discriminant locus between *E. albidus* and *E. encrasicolus*.

Significant genetic differences were observed between Benguela anchovy and Mediterranean *Engraulis encrasicolus*, a result that was not apparent in an earlier survey (Borsa et al., 2004), presumably because of limited sample sizes. Grant et al.'s (2005) cytochrome-*b* nucleotide sequence data (their table 4) show that the most common haplotype in Benguela anchovy ("*E. capensis*") is absent from the Western Mediterranean anchovy sample chosen as representative of *E. encrasicolus*. Grant et al. (2005) have also reported a genetic distance of *D* (Nei, 1978) = 0.049 based on allozyme frequencies, between *E. capensis* and *E. encrasicolus*, which amounts to about one million years reproductive isolation (Vawter et al., 1980). Grant and Bowen (2006) have hypothesized that southern African anchovy are a satellite population of European (Western Mediterranean) anchovy based on low haplotype diversities in southern African anchovy

and haplotype sharing with European anchovy. Here, Benguela anchovy appeared close to *E. albidus* at one intron locus and to northeastern Atlantic-Alboran Sea *E. encrasicolus* at the other locus. We speculate that the peculiar genetic composition of Benguela anchovy may result from past contact between the two latter, either before or after the transequatorial migration event that is thought to have brought northeastern Atlantic / Mediterranean anchovy to southern Africa (Grant and Bowen, 1998; Grant et al., 2005; Grant and Bowen, 2006). Alternatively, the Benguela anchovy may derive from an ancestral stock that also gave rise to both northeastern Atlantic *E. encrasicolus* and *E. albidus*.

4.4. Concurrent evidence of secondary contact

The average nucleotide divergence between the *A* and *B* lineages, which was estimated at between 3.7% for the whole mt DNA (Magoulas et al., 1996) and 2.2% from cytochrome *b* gene sequences (Grant et al., 2005), suggests reproductive isolation for ca. 1.1-1.85 million years (Grant 2005), a duration that may have been sufficient to initiate speciation (Avise et al., 1998). The presence of these two lineages is compatible with a scenario of former reproductive isolation and their co-existence in populations distributed from the Gulf of Biscay to the Aegean Sea indicates secondary contact. The geographic distribution of lineages *A* and *B* in European anchovy sampled across the northeastern Atlantic and the Mediterranean is discontinuous (Fig. 1C), in a fashion that partly parallels the patterns derived from the present nuclear-DNA data in *E. encrasicolus* (Atlantic / Mediterranean discontinuity; Western Mediterranean population genetically intermediate between Black Sea and Ionan Sea populations; isolation of the Black Sea population). This indicates correlation between nuclear and mtDNA variation, which strongly suggests incomplete re-homogenization between differentiated entities following secondary contact.

Acknowledgements

All the genotype data reported in this paper were obtained by Y.B.-K. in the course of his M.Sc. project, which was run at Station méditerranéenne de l'environnement littoral (Sète, France) under the supervision of P.B. and J.-D.D, in 2002. We are grateful to W.S. Grant for insightful discussions and comments, and to S.D. Connell, C. Lemaire, P. Luttikhuizen, D. Ponton, M. Raymond and five anonymous referees for comments on former versions of the paper. Ma. and Mo. Bouchenak-Khelladi, P. Fréon (IRD, Sète), Y. Guennégan (IFREMer, Sète), M. Houny (SoCoMaP, Grau-du-Roi), C. van der Lingen (M&CM, Cape Town), and J.-P. Quignard (Université Montpellier 2) helped with sampling; I. Calderón and J. Dietrich (IFREMer) contributed to the laboratory work. This research was funded by CNRS UMR 5000, IFREMer URM 16, and IRD UR 081.

References

Arnold, J., 1993. Cytonuclear disequilibria in hybrid zones. Annu. Rev. Ecol. Syst. 24, 521–554.

Avise, J.C., Walker, D., Johns, G.C., 1998. Speciation durations and Pleistocene effects on vertebrate phylogeography. Proc. R. Soc. Lond. B 265, 1707-1712

Bargelloni, L., Alarcon, J.A., Alvarez, M.C., Penzo, E., Magoulas, A., Palma, J., Patarnello, T., 2005. The Atlantic-Mediterranean transition: discordant genetic patterns in two seabream species, *Diplodus puntazzo* (Cetti) and *Diplodus sargus* (L.). Mol. Phyl. Evol. 36, 523-535.

Barton, N.H., Gale, K.S., 1993. Genetic analysis of hybrid zones. In: Harrison RG (ed.) Hybrid zones and the evolutionary process. Oxford University Press,New York, pp 13-45.

Belkhir, K., Borsa, P., Raufaste, N., Chikhi, L., Bonhomme, F., 2000. GENETIX version 4.02, logiciel sous WINDOWS^{TM} pour la génétique des populations. Laboratoire Génome et Populations, Université Montpellier 2, Montpellier.

Bembo, D.G., Carvalho, G.R., Cingolani, N., Arneri, E., Giannetti, G., Pitcher, T.J., 1996a. Allozymic and morphometric evidence for two stocks of the European anchovy *Engraulis encrasicolus* in Adriatic waters. Mar. Biol. 126, 529-538.

Bembo, D.G., Carvalho, G.R., Cingolani, N., Pitcher, T.J., 1996b. Electrophoretic analysis of stock structure in Northern Mediterranean anchovies, *Engraulis encrasicolus*. ICES J. Mar. Sci. 53, 115–128.

Bembo, D.G., Carvalho, G.R., Snow, M., Cingolani, N., Pitcher, T.J., 1995. Stock discrimination among European anchovies, *Engraulis encrasicolus*, by means of PCR-amplified mitochondrial DNA analysis. Fishery Bull. 94: 31–40.

Benzécri, J.-P., 1982. L'analyse des données. 2. L'analyse des correspondances. Dunod, Paris.

Borsa, P., 2002. Allozyme, mitochondrial-DNA, and morphometric variability indicate cryptic species of anchovy (*Engraulis encrasicolus*). Biol. J. Linn. Soc. 75, 261-269.

Borsa, P., Collet, A., Durand, J.-D., 2004. Nuclear-DNA markers confirm the presence of two anchovy species in the Mediterranean. C. R. Biol. 327, 1113-1123.

Borsa, P., Naciri, M., Bahri, L., Chikhi, L., García de León, F.J., Kotoulas, G., Bonhomme, F., 1997. Zoogéographie infraspécifique de la mer Méditerranée. Analyse des données génétiques populationnelles sur seize espèces atlanto-méditerranéennes (poissons et invertébrés). Vie Milieu 47, 295-305.

Briggs, J.C., 1995. Global biogeography. Elsevier, New York.

Burton, R.S., 1998. Intraspecific phylogeography across the Point Conception biogeographic boundary. Evolution 52, 734-745.

Carvalho, G.R., Hauser, L., 1998. Advances in the molecular analysis of fish population structure. Ital. J. Zool. 65 (suppl.), 21–33.

Chow, S., Takeyama, H., 1998. Intron length variation observed in the creatine kinase and ribosomal protein genes of the swordfish *Xiphias gladius*. Fisheries Sci. 64, 397-402.

Cunningham, C.W., Collins, T.M., 1994. Developing model systems for molecular biogeography: vicariance and interchange in marine invertebrates. In: Schierwater B, Streit B, Wagner GP,. DeSalle R (eds). Molecular ecology and evolution: approaches and applications. Birkhaüser Verlag, Basel, pp 405-433.

Daguin, C., Borsa, P., 1999. Genetic characterisation of *Mytilus galloprovincialis* Lmk. in North West Africa using nuclear DNA markers. J. Exp. Mar. Biol. Ecol. 235, 55-65.

Dieckmann, U., Doebeli, M., 1999. On the origin of species by sympatric speciation. Nature 400, 354-357.

Grant, W.S., 2005. A second look at mitochondrial DNA variability in European anchovy (*Engraulis encrasicolus*): assessing models of population structure and the Black Sea isolation hypothesis. Genetica 125, 293-309.

Grant, W.S., Bowen, B.W., 1998. Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. J. Hered. 89, 415-426.

Grant, W.S., Bowen, B.W., 2006. Living in a tilted world: climate change and geography limit speciation in Old World anchovies (*Engraulis*; Engraulidae). Biol. J. Linn. Soc. 88, 673-689.

Grant, W.S., Leslie, R.W., Bowen, B.W., 2005. Molecular genetic assessment of bipolarity in the anchovy genus *Engraulis*. J. Fish Biol. 67, 1242-1265.

Guinand, B., 1996. Use of a multivariate model using allele frequency distributions to analyse patterns of genetic differentiation among populations. Biol. J. Linn. Soc. 58, 173-195.

Harrison, R.G., 1990. Hybrid zones: windows on evolutionary processes. Oxf. Surv. Evol. Biol. 7, 69-128.

Hedgecock, D., 1994. Temporal and spatial genetic structure of marine animal populations in the California current. CalCOFI Rep. 35:73-81.

Hedgecock, D., Hutchinson, E.S., Li, G., Sly, F.L., Nelson, K., 1994. The central stock of northern anchovy (*Engraulis mordax*) is not a randomly mating population. CalCOFI Rep. 35, 121-136.

Karaçam H., Düzgünes E. 1990. Age, growth and meat yield of the European anchovy (*Engraulis encrasicolus*, L. 1758) in the Black Sea. Fisheries Res. 9, 181-186.

Lebreton, J.-D., Roux, M., Banco, G., Bacou, A.-M., 1990. BIOMECO (biométrie-écologie), logiciel d'écologie statistique pour PC et compatibles, version 3.9. Centre national de la recherche scientifique, Montpellier.

Lemaire, C., Versini, J.-J., Bonhomme, F., 2005. Maintenance of genetic differentiation across a transition zone in the sea: discordance between nuclear and cytoplasmic markers. J. Evol. Biol. 18, 70-80.

Magoulas, A., Castilho, R., Caetano, S., Marcato, S., Patarnello, T., 2006. Mitochondrial DNA reveals a mosaic pattern of phylogeographical structure in Atlantic and Mediterranean populations of anchovy (*Engraulis encrasicolus*). Mol. Phyl. Evol. 39, 734-746.

Magoulas, A., Tsimenides, N., Zouros, E., 1996. Mitochondrial DNA phylogeny and the reconstruction of the population history of a species: the case of the European anchovy (*Engraulis encrasicolus*). Mol. Biol. Evol. 13, 178-190.

Naciri, M., Lemaire, C., Borsa, P., Bonhomme, F., 1999. Genetic study of the Atlantic / Mediterranean transition in sea bass (*Dicentrarchus labrax*). J. Hered. 90, 591-596.

Nei, M., 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89, 583–590.

Palumbi, S.R., 1994. Genetic divergence, reproductive isolation, and marine speciation. Annu. Rev. Ecol. Syst. 25, 547-572.

Pannacciulli, F.G., Bishop, J.D.D., Hawkins, S.J., 1997. Genetic structure of populations of two species of *Chthamalus* (Crustacea: Cirripedia) in the north-east Atlantic and Mediterranean. Mar. Biol. 128, 73-82.

Park, T., 1994. Geographic distribution of the bathypelagic genus *Pareuchaeta* (Copepoda, Calanoida). Hydrobiologia 292/293, 317-332.

Pasteur, N., Berrebi, P., 1985. Structure génétique des populations naturelles d'animaux laguno-marins sur des bases de génétique biochimique et d'écologie (contrat no. 83.J.0024, Ministère de la recherche, Direction du développement scientifique, technologique et de l'innovation). Université des sciences et techniques du Languedoc, Montpellier.

Patarnello, T., Volckaert, F.J.C., Castilho, R. 2007. Pillars of Hercules: is the Atlantic-Mediterranean transition a phylogeographic break? Mol. Ecol. 16, 4426-4444.

Quesada, H., Zapata, C., Alvarez, G., 1995. A multilocus allozyme discontinuity in the mussel *Mytilus galloprovincialis*: the interaction of ecological and life-history factors. Mar. Ecol. Prog. Ser. 116, 99-115.

Rice, W.R., 1989. Analyzing tables of statistical tests. Evolution 43, 223-225.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory: New York

Spanakis, E., Tsimenides, N., Zouros, E., 1989. Genetic differences between populations of sardine, *Sardina pilchardus*, and anchovy, *Engraulis encrasicolus*, in the Aegean and Ionian Seas. J. Fish Biol. 35, 417-437.

Tintore, J., La Violette, P.E., Blade, I., Cruzado, A., 1988. A study of an intense density front in the eastern Alboran sea: the Almería-Oran front. J. Phys. Oceanogr. 18, 1384-1397.

Vamosi, S.M., Schluter, D., 1999. Sexual selection against hybrids between sympatric stickleback species: evidence from a field experiment. Evolution 53, 874-879.

Vawter, A.T., Rosenblatt, R., Gorman, G.C., 1980. Genetic divergence among fishes of the Eastern Pacific and the Caribbean: support for the molecular clock. Evolution 34: 705-711.

Weir, B.S., 1979. Inferences about linkage disequilibrium. Biometrics 35, 235-254.

Weir, B.S., Cockerham, C.C., 1984. Estimating *F*-statistics for the analysis of population structure. Evolution 38, 1358-1370.

Whitehead, P.J.P., 1990. Engraulididae. In: Quero J-C, Karrer C, Post A, Saldanha L (eds) Checklist of the fishes of the eastern tropical Atlantic (CLOFETA), vol. 1. UNESCO, Paris, pp. 228-229.

Table 1

Allele frequencies at creatine-kinase intron 6 loci *CK6- 1* and *CK6- 2* in European anchovy. *N*: sample size; ^*f*: estimator of Weir and Cockerham's (1984) *f*; *R*ij: genotypic correlation coefficient (Weir, 1979); *P*: probability of a value greater than or equal to the observed $|\hat{f}|$ or *R*ij, under the null hypothesis of panmixia [permutation tests (Belkhir et al., 2000): 1000 random permutations]. Abbreviations for samples as in legend to Fig. 1

Locus,	Sample												
Allele	CUL	MAU	Lion	SET	Adr	Pat	Kav	Bls	Geo	ANN	ORN	Can	BEN
CK6- 1													
120	-	-	-	-	-	0.01	-	-	-	-	-	-	-
110	-	-	-	0.01	-	0.01	0.02	-	-	0.01	-	-	-
105	-	-	-	-	-	-	-	0.02	-	-	-	-	-
100	0.93	0.93	0.93	0.85	0.74	0.73	0.78	0.94	0.95	0.86	0.58	0.61	0.75
098	-	-	-	-	-	-	0.01	-	-	0.02	-	-	-
095	-	-	-	-	-	0.01	-	-	-	-	-	-	-
090	-	-	-	-	-	0.02	-	-	0.01	-	-	0.01	-
052	-	-	-	-	-	-	0.02	-	-	-	-	0.01	0.01
c050	0.05	0.05	0.07	0.09	0.19	0.15	0.09	-	0.01	0.06	0.29	0.33	0.22
047	-	-	-	-	-	0.01	0.01	-	-	-	-	0.01	-
030	-	-	-	-	-	-	-	-	-	-	-	0.01	-
010	0.02	0.01	-	0.05	0.07	0.06	0.08	0.04	0.03	0.05	0.08	0.02	-
001	-	-	-	-	-	-	-	-	-	-	0.04	-	0.01
(*N*)	(112)	(46)	(15)	(95)	(52)	(53)	(58)	(56)	(55)	(49)	(24)	(49)	(49)
^*f*	-0.057	-0.048	1.000	0.283	-0.068	-0.052	-0.029	0.555	-0.025	-0.016	0.283	0.103	0.042
P	0.999	0.999	0.025	0.001	0.811	0.782	0.722	0.004	0.999	0.679	0.049	0.248	0.437
CK6- 2													
150	-	-	-	-	-	-	-	0.03	0.04	0.01	-	-	0.02
110	-	-	-	0.01	-	-	-	-	-	-	-	-	-
107	-	-	-	-	-	-	-	-	-	-	0.02	0.01	-
c105	0.10	0.10	0.43	0.49	0.47	0.36	0.46	0.48	0.44	0.16	0.23	0.13	0.03
c102	0.13	0.12	0.30	0.16	0.09	0.26	0.18	0.10	0.12	0.24	0.17	0.25	0.07
100	0.70	0.71	0.20	0.24	0.34	0.21	0.18	0.26	0.23	0.53	0.40	0.50	0.79
098	0.04	0.04	-	0.03	-	-	0.01	0.04	0.02	0.02	0.02	0.04	0.01
097	-	-	-	-	0.01	-	-	0.03	0.07	0.01	0.02	0.03	0.06
095	0.03	0.03	0.07	0.07	0.09	0.17	0.14	0.05	0.05	0.02	0.08	0.02	0.01
090	-	-	-	-	-	-	0.03	-	-	-	0.06	-	-
010	-	-	-	-	-	-	-	0.01	0.04	-	-	0.01	-
(*N*)	(115)	(46)	(15)	(93)	(46)	(29)	(36)	(50)	(53)	(49)	(24)	(48)	(48)
^*f*	-0.037	0.185	0.438	0.053	-0.188	0.119	0.146	0.164	0.127	0.233	-0.089	0.072	-0.024
P	0.763	0.044	0.010	0.242	0.986	0.194	0.086	0.032	0.054	0.013	0.886	0.258	0.712
*R*ij	0.048	0.105	0.228	0.081	0.133	0.189	0.119	0.126	0.094	0.262	0.172	0.193	0.099
P	0.942	0.678	0.734	0.489	0.449	0.215	0.694	0.187	0.640	0.020	0.610	0.132	0.552

Table 2

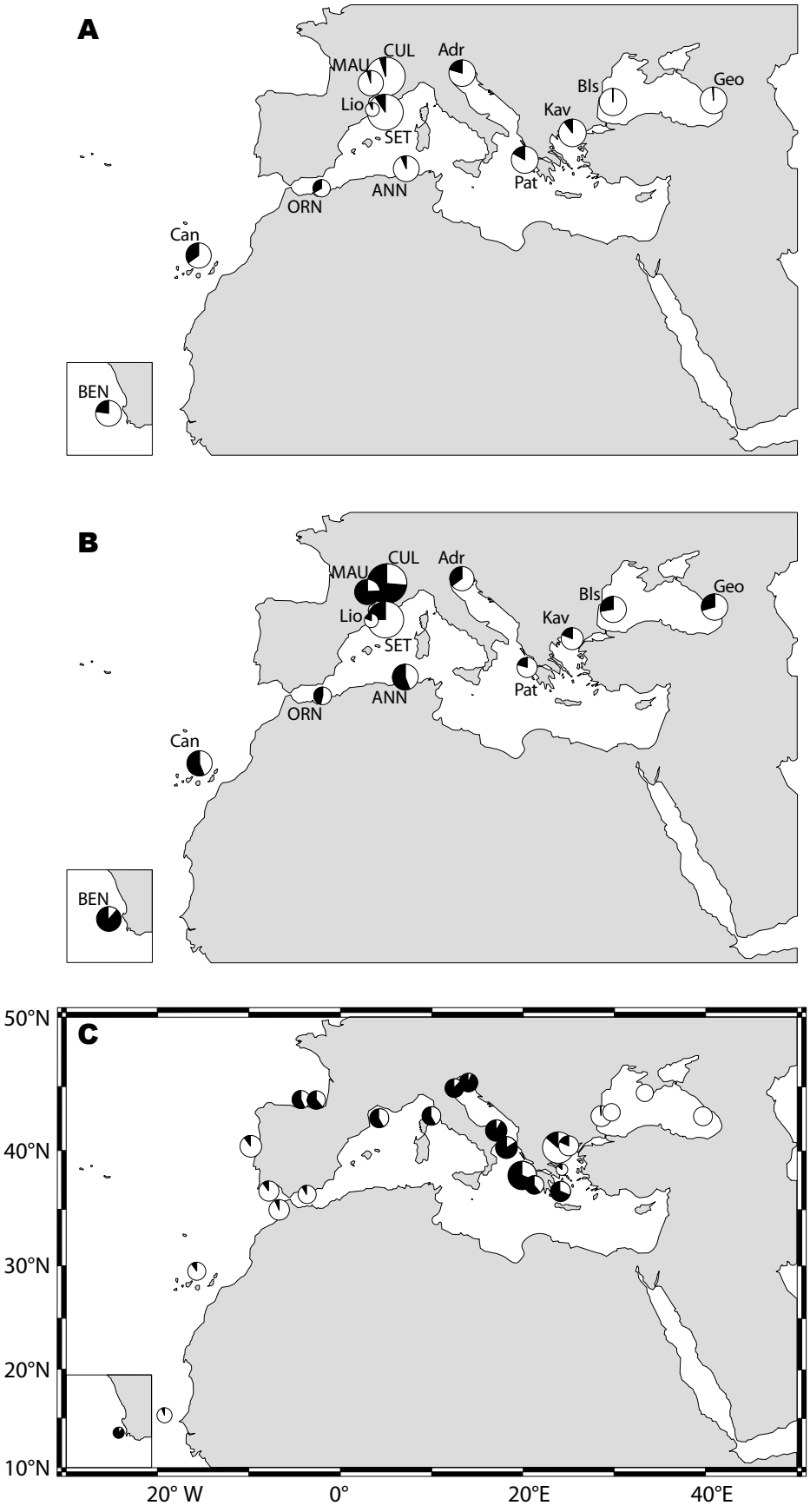
Population differentiation in European anchovy. Values of pairwise Weir and Cockerham's (1984) $\hat{\theta}$ at intron loci *CK6-1* (above diagonal) and *CK6-2* (below diagonal). Abbreviations for samples as in legend to Fig. 1

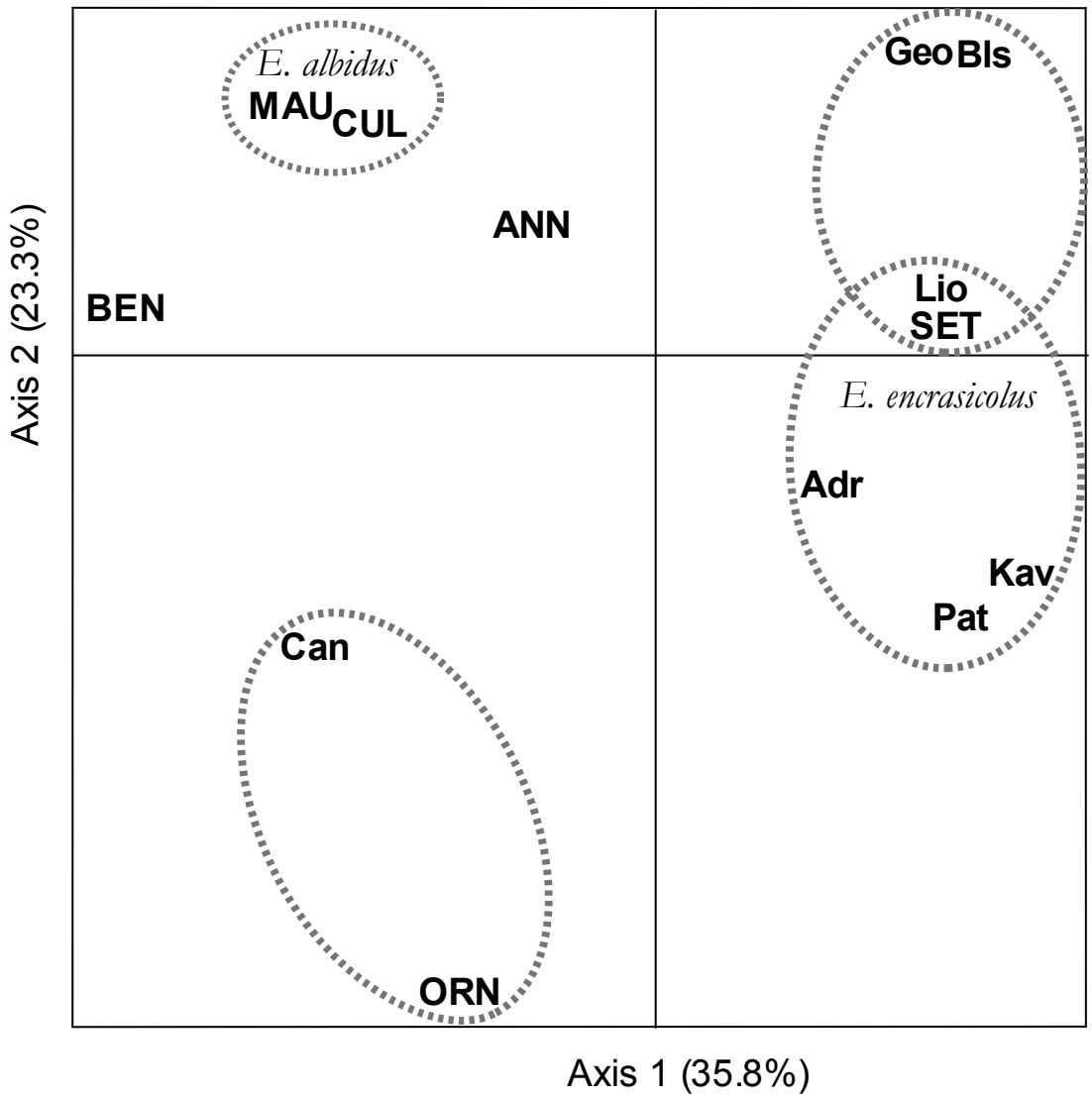
Sample	Sample												
	CUL	MAU	Lion	SET	Adr	Pat	Kav	Bls	Geo	ANN	ORN	Can	BEN
CUL		−0.007	−0.018	0.015	0.108	0.096	0.057	0.008	0.005	0.012	0.293	0.253	0.119
MAU	−0.008		−0.026	0.011	0.089	0.078	0.048	0.005	0.004	0.011	0.240	0.206	0.098
Lio	0.275	0.259		−0.007	0.056	0.048	0.026	−0.007	−0.003	−0.005	0.155	0.147	0.060
SET	0.247	0.232	−0.001		0.028	0.023	0.004	0.027	0.033	−0.007	0.139	0.134	0.039
Adr	0.205	0.190	0.027	0.004		−0.006	0.006	0.120	0.135	0.036	0.022	0.029	−0.002
Pat	0.246	0.227	−0.019	0.016	0.033		−0.001	0.101	0.116	0.027	0.027	0.036	0.004
Kav	0.278	0.258	−0.014	−0.003	0.015	−0.005		0.059	0.074	0.004	0.068	0.078	0.021
Bls	0.238	0.220	0.009	−0.004	−0.002	0.026	0.004		−0.007	0.018	0.283	0.251	0.142
Geo	0.238	0.218	0.002	0.001	0.006	0.018	0.002	−0.007		0.029	0.311	0.265	0.154
ANN	0.038	0.032	0.104	0.128	0.101	0.097	0.135	0.119	0.113		0.140	0.136	0.050
ORN	0.095	0.083	0.038	0.055	0.037	0.029	0.050	0.046	0.041	0.010		−0.011	0.028
Can	0.046	0.040	0.100	0.128	0.105	0.092	0.131	0.119	0.110	−0.010	0.009		0.023
BEN	0.016	0.012	0.381	0.316	0.280	0.331	0.358	0.309	0.301	0.095	0.168	0.102	

* value significant according to permutation test (1000 permutations under GENETIX; Belkhir et al., 2000). Significance level adjusted according to sequential Bonferroni procedure (Rice, 1989) by considering each locus independently, and 12 independent sets of 12 pairwise values at each locus

Fig. 1. Geographic variation in European anchovy. A. Allele frequencies at locus *CK6-1* (black: allele *100*; open: all other alleles). B. Allele frequencies at locus *CK6-1* (same settings as Fig. 1A). C. Mitochondrial haplotype frequencies, modified from Magoulas et al. (2006) and Grant et al. (2005) (black: lineage *B*; open: lineage *A*). [Area of pie diagram proportional to sample size. *CUL* Cul-de-Beauduc, Camargue, January 2001; *MAU* Mauguio lagoon, Camargue, May 2001; *Lio* off Sète, December 1992; *SET* 18 miles offshore from Sète, December 2000; *Adr* Chioggia, Adriatic Sea, November 1993; *Pat* Patras, Ionian Sea, 1989; *Kav* Kavala, Aegean Sea, 1989; *Bls* Varna, Black Sea, 1997; *Geo* Batumi, Black Sea, 1999; *ANN* off Annaba, Western Mediterranean, May 2001; *ORN* off Oran, Alboran Sea, February 2002; *Can* Canary archipelago, northeastern Atlantic, 1999; *BEN* Benguela Current, off southern Africa (inset), September 2000]

Fig. 2. Genetic variation at two intron loci in European anchovy. Projection on the principal plane of CA (axis 1 and axis 2) of thirteen samples from the Atlantic and the Mediterranean defined by their allelic frequencies at loci *CK6-1* and *CK6-2*. Samples have been lumped together (dashed lines) when pairwise $\hat{\theta}_{2L}$-values between them were not significant. Inertia values for each axis in brackets. Taxonomic labels placed near samples from type localities [respectively, CUL (Cul-de-Beauduc) for *Engraulis albidus* and SET (Sète) for *E. encrasicolus*]


A
CUL
Adr
MAU
Lio
SET
Bls
Geo
Kav
Pat
ANN
ORN
Can
BEN
B
CUL
Adr
MAU
Lio
SET
Bls
Geo
Kav
Pat
ANN
ORN
Can
BEN
C
50°N
40°N
30°N
20°N
10°N
20° W
0°
20°E
40°E


E. albidus
MAU
CUL
ANN
BEN
Geo
Bls
Lio
SET
E. encrasicolus
Adr
Kav
Pat
Can
ORN
Axis 2 (23.3%)
Axis 1 (35.8%)