Multiple SNP Markers Reveal Fine-Scale Population and Deep Phylogeographic Structure in European Anchovy (Engraulis encrasicolus L.)

Iratxe Zarraonaindia1, Mikel Iriondo1, Aitor Albaina1, Miguel Angel Pardo2, Carmen Manzano1, W. Stewart Grant3, Xabier Irigoien4, Andone Estonba1

1 Laboratory of Genetics, Department of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country (UPV/EHU), Sarriena azuoa z/g, Leioa (Bizkaia), Spain, 2 AZTI-Tecnalia, Food Research Unit, Parque Tecnológico de Bizkaia, Astondo Bidea, Edif. 609, Derio (Bizkaia), Spain, 3 Commercial Fisheries Division, Alaska Department of Fish and Game, Anchorage, Alaska, United States of America, 4 AZTI-Tecnalia, Marine Research Division, Herrera Kaia Portualde z/g, Pasaia (Gipuzkoa), Spain

Abstract

Geographic surveys of allozymes, microsatellites, nuclear DNA (nDNA) and mitochondrial DNA (mtDNA) have detected several genetic subdivisions among European anchovy populations. However, these studies have been limited in their power to detect some aspects of population structure by the use of a single or a few molecular markers, or by limited geographic sampling. We use a multi-marker approach, 47 nDNA and 15 mtDNA single nucleotide polymorphisms (SNPs), to analyze 626 European anchovies from the whole range of the species to resolve shallow and deep levels of population structure. Nuclear SNPs define 10 genetic entities within two larger genetically distinctive groups associated with oceanic variables and different life-history traits. MtDNA SNPs define two deep phylogroups that reflect ancient dispersals and colonizations. These markers define two ecological groups. One major group of Iberian-Atlantic populations is associated with upwelling areas on narrow continental shelves and includes populations spawning and overwintering in coastal areas. A second major group includes northern populations in the North East (NE) Atlantic (including the Bay of Biscay) and the Mediterranean and is associated with wide continental shelves with local larval retention currents. This group tends to spawn and overwinter in oceanic areas. These two groups encompass ten populations that differ from previously defined management stocks in the Alboran Sea, Iberian-Atlantic and Bay of Biscay regions. In addition, a new North Sea-English Channel stock is defined. SNPs indicate that some populations in the Bay of Biscay are genetically closer to North Western (NW) Mediterranean populations than to other populations in the NE Atlantic, likely due to colonizations of the Bay of Biscay and NW Mediterranean by migrants from a common ancestral population. Northern NE Atlantic populations were subsequently established by migrants from the Bay of Biscay. Populations along the Iberian-Atlantic coast appear to have been founded by secondary waves of migrants from a southern refuge.

Introduction

European anchovies, Engraulis encrasicolus, are widely distributed in near-shore pelagic waters in the Eastern Atlantic from the North Sea and into the Mediterranean and Black Sea and as far south as southern Africa. Populations in the North East (NE) Atlantic and Mediterranean are partitioned into several spawning groups that are isolated from one another by complex shorelines and oceanic regimes [1,2]. Several biochemical and molecular studies have shown that many of these partitions are congruent with genetic differences between populations [3–13]. In some areas, small-scale structure appears among anchovy populations in the Bay of Biscay [12,13], the North Western (NW) Mediterranean [14–16], Adriatic Sea [6] and Black Sea [17]. It is uncertain whether these fine-scale differences are due to adaptive divergence [5,17], or to cryptic species diversity [14–16]. The analysis of mitochondrial (mt) DNA has further resolved two deep maternal lineages in European anchovies that appear to reflect ancient isolations, dispersals and colonizations [9,10,18]. These lineages, or mtDNA phylogroups, vary greatly in frequency from one location to another, but are not strictly associated with the geographical population groups delimited by nuclear molecular markers [5–13]. The historical events that influenced the geographical distributions of these phylogroups have been a topic of debate [9,19]. Phylogroup A occurs at a high frequency in the Black and Aegean seas and in the eastern Atlantic, but only at...
intermediate frequencies at most locations in the Mediterranean and NE Atlantic [10]. In contrast, phylogroup B occurs at moderate frequencies at many locations [10]. A mtDNA haplotype frequency shift occurs between Mediterranean and Atlantic populations [9,10], and this shift is echoed by a shift in microsatellite allele frequencies [13]. The mechanisms leading to high abundances of phylogroup A in disjunct locations are still unresolved. So far, no morphological or adaptive differences have been detected between individuals in phylogroups A and B [11].

While migration, genetic drift and adaptive selection may explain much of the genetic structure of anchovy populations, deeper genetic partitions with coalescences dating to tens of thousands of years more likely reflect ancient climate events during the Pleistocene ice-ages. During each ice age, continental glaciers expanded across North America and Eurasia, leading to lower ocean temperatures and sea levels [20]. These drops in sea level gave rise to expanded shorelines, and in some places to the appearance of land barriers that severed connections between marine populations [21]. As temperatures dropped, populations likely tracked optimal habitats as they shifted toward lower latitudes [22]. Populations moved poleward and dispersed into suitable coastal habitats as mid-latitude waters warmed and coastal barriers receded with rising sea levels. These ocean-climate shifts greatly influenced anchovy populations in the NE Atlantic and Mediterranean and potentially imprinted the genetic structures of contemporary populations. Hence, the examination of genetic variability can potentially provide insights into the population histories of European anchovies.

Previous studies of European anchovies have been limited in two ways. First, most studies of European anchovies have been regional, where samples have been restricted to one of the many sub-basins in the Mediterranean Sea. Populations in the North Sea and Bay of Biscay and along the Iberian-Atlantic coast have received less attention. The analysis of these populations will greatly influence anchovy populations in the NE Atlantic and potentially imprinted the genetic structures of contemporary populations. Hence, the examination of genetic variability can potentially provide insights into the population histories of European anchovies.

Materials and Methods

a) Samples and DNA Extraction

Anchovies were provided either by commercial vessels or by oceanographic institutes that collected the samples during scientific acoustic surveys (BIOMAN, PELGAS, ECOCADIIZ, ECOMED, PELACUS). All surveys followed local regulations and guidelines for such research. Anchovies were collected following fishing without unnecessary suffering of the animals and following usual procedures. No experimentation with animals was performed. No other ethical issues applied to the present research project.

A total of 626 anchovies (mean N = 30) were collected from the entire geographic range of this species and its spawning areas (Table 1, Fig. 1). Five locations (N = 188) were sampled in the Mediterranean Sea along an east-west axis. Several locations were sampled in the Eastern Atlantic (N = 443), including 127 fish from the North Sea and English Channel (6 locations), 106 fish from the Bay of Biscay (3 locations) and 210 fish from Galicia to South Africa (7 locations). Most samples were collected from 2007 to 2009. Genomic DNA was extracted from muscle tissue using the DNeasy 96 Tissue Kit (Qiagen), following manufacturer’s protocol. A specific protocol was used (www.ambion.com/techlib/misc/genomicDNA_manual.html) for tis-
sues preserved in RNAlater® (Applied Biosystems), as was the case for North Sea and English Channel samples.

b) SNP Selection and Genotyping

The 90 novel SNPs, including 75 nuclear (n) DNA and 15 mtDNA SNPs, described by Zarraonaindia et al. included in Molecular Ecology Resources Primer Development Consortium et al. [31] were considered for this study. Five nuclear markers exhibiting significant departures from Hardy-Weinberg expectation (HWE) [31] were discarded. Genotypic Disequilibrium tests, within and among sequenced fragments, were performed for the remaining 70 SNPs using GENEPOP 4.0 [32] with rejection probabilities of $P<0.001$. Linked SNPs were phased into haplotypes using the Bayesian statistical method implemented in PHASE 2.1 [33]. Haplotypes were reconstructed by location to avoid biases from population structuring. Missing genotypes were classified as null genotypes to avoid haplotype reconstruction errors. Only one SNP or haplotype block was chosen for each sequenced fragment to ensure maximum independence among the markers. Markers showing the largest heterozygosities per fragment were given preference. This yielded 49 SNP markers, including 26 individual SNPs and 10 haplotype blocks (Table S1). We additionally tested for natural selection with BayeScan [34] in samples grouped by geographical area. We calculated Bayes factors (BF) to identify candidate loci by using 20 pilot runs of 5000 iterations and an additional burn-in of 50,000 iterations, for a total of 100,000 iterations (sample size of 5000 and thinning interval of 10). Locus specific BFs are the ratio of the posterior probabilities of selection and neutrality, given the data. The marker representing haplotype block KLN-332-(144-444) (Table S1) presented a log10BF value >2 corresponding to decisive evidence for selection (BayeScan v2.0 User Manual). Hence, it was discarded from the final SNP panel of 47 nDNA SNPs, representing 35 independent markers including 26 individual SNPs and 9 haplotype blocks. We used FSTAT 2.9.3 [35] to estimate the number of alleles, expected and observed heterozygosity and $F_{IS}$ for each sample.

The 626 anchovies from 20 localities were screened for the 47 nDNA SNPs panel and 15 mtDNA SNPs with TaqMan® OpenArray™ genotyping system. DNA concentrations and reactions for amplification and detection of the SNPs followed the TaqMan® OpenArray™ Genotyping System User Guide. Genotypes were scored using Autocaller 1.1 (Applied Biosystems).

c) Statistical Analysis

Nuclear SNPs were examined in three ways. First, population structure was inferred from individual assignments based on the analysis of the nuclear SNPs using the Bayesian model-based clustering algorithm implemented in STRUCTURE 2.3.2 [36]. This program uses Hardy-Weinberg proportions and gametic disequilibrium to cluster individuals into $K$ groups. Ten independent runs were conducted for each value of $K$ ($K=1–10$) using 500,000 Markov Chain Monte Carlo iterations, after a burn-in of 50,000 iterations. We assumed a mixed ancestry model and correlated allele frequencies [37] and used a locprior model. We used CLUMPP 1.1.2 [38] to determine optimal assignments of individuals to clusters, maximizing similarity between the different STRUCTURE replications. The most likely number of populations ($K$) that best explained the pattern of genetic variability was evaluated using the approach of Evanno et al. [39]. Graphical output of individual membership coefficients in each cluster was created with DISTRUCT 1.1 [40].

Second, we estimated genetic distances between samples and used these distances in additional analyses. Reynolds genetic distance [41] was estimated between samples from allele frequencies over loci using POPULATION 1.2.28 [42]. This matrix was used to construct a Neighbour-Joining (NJ) tree for which topological confidence was evaluated with 1000 bootstrap
replicates. Isolation by distance was tested by searching for a correlation between genetic and geographical distance matrices. Geographical distances (km) were calculated as the shortest path between sample locations inside the 1000 m isobaths and by considering the known spatial distribution of the European anchovy. The matrix was obtained using the spatial analysis tool Path Distance implemented in ArcGIS 9.2. Matrix comparisons by Mantel’s method were carried out with the program ZT [43], and significance was determined with 10,000 iterations.

Lastly, genetic differentiation among samples was estimated with the unbiased fixation index ($F_{ST}$) [44] using FSTAT [35,45]. Confidence intervals for $F_{ST}$ were determined by jackknifing, and statistical significance was determined with 15,000 permutations. We used the Bonferroni correction probabilities for multiple tests [46]. Population groups were defined by non-significant values of mean $F_{ST}$ between samples and by significant values of $F_{ST}$ with other populations [30].

The number of SNP mtDNA haplotypes ($n_h$), haplotype diversity ($h$), and Tajima’s $D$ [47] and Fu’s $F_S$ [48] measures of neutrality were estimated with DNAsp 5 [49]. The statistical significances of $D$ and $F_S$ were tested with 1000 coalescent simulations. Mutational relationships among haplotypes were represented by a median-joining (MJ) network [50] constructed with NETWORK 4.2 [http://www.fluxus-technology.com]. The SNPs Cyt-b-318, D-loop-323 and D-loop-336 were not used to construct the network, because their high levels of polymorphism would reduce the resolution among haplotypes.

**Results**

a) Nuclear DNA Analysis

STRUCTURE analysis of 626 anchovies yielded the most likely partition of $K=2$. Membership in these two groups varied latitudinally (Fig. 2). One group included samples from the North Sea and English Channel, the Bay of Biscay and the Mediterranean (excluding the Alboran Sea). The other group was represented by anchovies in South Africa and encompassed samples from eastern Atlantic locations extending from Galicia to South Africa, but also including the Alboran Sea.

A similar pattern of population differentiation appeared in genetic distances between populations. The NJ tree of Reynolds genetic distances also showed two major groups supported by a 90% bootstrap value (Fig. 3). The first group consisted of NE Atlantic samples, including those from the North Sea, Bay of Biscay and the Mediterranean, except the Alboran Sea. The second group consisted of Atlantic samples arranged latitudinally in the tree from Galicia to South Africa, but also including the Alboran Sea. The test for isolation by distance (IBD) among populations was significant (Mantel’s test: $r = 0.771, \ P = 0.0001$; matrix $20 \times 20$). IBD among the NE Atlantic samples from the Bay of Biscay to the Canary Islands (including the Alboran Sea) was highly significant ($r = 0.903, \ P = 0.0007$; matrix $7 \times 7$). The addition of the distantly located sample from South Africa lowered the correlation between geographical distance and genetic distance ($r = 0.709, \ P = 0.0001$, matrix $8 \times 0$).

Pairwise $F_{ST}$ ranged from 0.04, between two samples from the Bay of Biscay, to 0.318 between samples from Denmark and South Africa, and averaged 0.078$\pm$0.012 (Table 1). No significant $F_{ST}$ values appeared between samples collected in different years at nearby locations. For example, $F_{ST}$ between samples from Delta (2007) and nearby Tarragona (2009) in the NW Mediterranean and between samples from the North Sea (2006–2009) were not significant. Based on these pairwise $F_{ST}$ values, the original 20 samples were pooled into 10 groups, which showed within-group homogeneity but between-group heterogeneity. The largest amount of divergence between these 10 groups was $F_{ST}$ = 0.270$\pm$0.053 between the South African and Aegean Sea samples (Table 2). Importantly, a significant amount of divergence ($F_{ST}$ = $0.013 \pm 0.007$) was detected between samples from nearby locations in the Bay of Biscay (BISC-1: Bisc-5029; and BISC-2: Bisc-5001), which were placed into two different pooled groups. Both pooled groups (BISC-1 and BISC-2) showed a larger amount of divergence from the populations in the North Sea and English Channel pooled group (NSEC), ($F_{ST}$ = $0.031 \pm 0.013$ and $F_{ST}$ = $0.027 \pm 0.008$, respectively) than from NW Mediterranean samples (NWMD) ($F_{ST}$ = $0.020 \pm 0.011$ and $F_{ST}$ = $0.024 \pm 0.013$, respectively).

b) Mitochondrial DNA Analysis

The 15 mtDNA SNPs defined 55 haplotypes, but only 23 haplotypes after excluding the highly polymorphic SNPs, Cyt-b-318, D-loop-323 and D-loop-336. The MJ Network showed two phylogroups, A and B, separated by 3 mutational steps, including a transition, a transversion and an indel (Fig. 4). Haplotype diversity in phylogroup A ($h = 0.825, \ SD = 0.012$) was similar to haplotype diversity in phylogroup B ($h = 0.829, \ SD = 0.014$). Phylogroup A displayed a star-like haplotype network, and both Tajima’s $D$ and Fu’s $F_S$ were negative, but not significant ($D = -0.605; \ P > 0.10$ and $F_S = -19.950; \ NS$). In contrast, the phylogroup B network was more reticulated, with two predominant haplotypes. Neither $D$ ($D = 0.171; \ P > 0.10$) nor $F_S$ ($F_S = -15.085; \ NS$) was significant. The mismatch distribution for both phylogroups was unimodal (not shown). In contrast, the two phylogroups together showed a bimodal mismatch distribution, with a positive and significant Tajima $D$ value ($D = 3.155, \ P < 0.01$). Both phylogroups appeared in each sample, but at different frequencies (Table 3). Phylogroup A was common in the Atlantic, from Galicia to the Canary Islands and in the Alboran

![Figure 2. Individual clustering analysis obtained by STRUCTURE analysis of 626 European anchovies for $K=2$. Each vertical bar represents an individual and the different geographic areas that were sampled are separated by vertical black lines. The color proportions of each bar correspond to the individual’s estimated membership fractions to each of the clusters (cluster membership coefficient).](http://www.plosone.org/doi/10.1371/journal.pone.0042201.g002)
and Aegean seas. Phylogroup B predominated in the North Sea and English Channel, Adriatic Sea and southern Africa. Phylogroups A and B occurred at similar frequencies in samples from the NW Mediterranean and the Bay of Biscay. Moderate haplotype diversities appeared in samples from Galicia (\(h = 0.606\)) and South Africa (\(h = 0.669\)), whereas the largest diversities appeared in the Bay of Biscay (Bisc-5029: \(h = 0.929\); Bisc-5020: \(h = 0.912\)) and English Channel (\(h = 0.913\)).

**Discussion**

The present study combines the use of mitochondrial and nuclear SNP markers, together with a large coverage of the distributional range of the European anchovy to provide a deeper understanding of the factors shaping the genetic population structure of the species and to disentangle historical from ecological factors reflected in contemporary populations. On the one hand, the analysis of maternally inherited, non-recombining mtDNA complements nuclear SNPs and provides insights into ancient events that have produced deep genetic/genomic partitions in European anchovies. On the other hand, nuclear SNPs enhance the power to detect population divergences resulting from genetic drift.

When comparing the resolving power of nDNA SNPs with other types of markers, such as microsatellites, the use of 4–12 nuclear SNPs is expected to have the same resolving power as
were 7–8 times larger between samples than for microsatellite genotypes [13]; the average microsatellite haplotype block size in anchovy samples was 0.029. These SNP markers provided enough power to test for possible genetic differences between anchovies showing life-history patterns similar to microsatellite genotypes [51–53]. In the present study, 47 nuclear SNP markers provided 35 independent makers have been used and among them 9 markers consisted of haplotype blocks. SNP markers provided $F_{ST}$ values that were 7–8 times larger between samples than $F_{ST}$ values estimated with 7 microsatellite loci [13]; the average microsatellite $F_{ST}$ between the Bay of Biscay, Gulf of Cadiz and Mediterranean anchovy samples was 0.029, whereas SNP markers provided enough power to detect significant differences among the samples of both regions. Hence, our study achieved greater resolution of population structure with SNP markers than would have been possible with microsatellites.

## Table 2. $F_{ST}$ values between the 10 homogeneous population groups of European anchovies.

|               | NSEC | BISC1 | BISC2 | CIAT | SIAT | CAN | SAF | NWMD | ADR | AEG |
|---------------|------|-------|-------|------|------|-----|-----|------|-----|-----|
| NSEC          | 0.013| 0.008 | 0.016 | 0.023| 0.033| 0.042| 0.014|      | 0.022| 0.041|
| BISC1         | 0.031| 0.007 | 0.014 | 0.020| 0.026| 0.035| 0.011| 0.019| 0.036|
| BISC2         | 0.027| 0.013 | 0.008 | 0.014| 0.023| 0.033| 0.013| 0.019| 0.036|
| CIAT          | 0.063| 0.053 | 0.022 | 0.005| 0.011| 0.017| 0.020| 0.017| 0.023|
| SIAT          | 0.114| 0.103 | 0.069 | 0.020| 0.015| 0.020| 0.029| 0.025| 0.030|
| CAN           | 0.169| 0.149 | 0.110 | 0.045| 0.026| 0.011| 0.034| 0.032| 0.031|
| SAF           | 0.247| 0.241 | 0.198 | 0.105| 0.073| 0.042| 0.051| 0.046| 0.053|
| NWMD          | 0.046| 0.020 | 0.024 | 0.056| 0.098| 0.137| 0.235| 0.008| 0.013|
| ADR           | 0.041| 0.042 | 0.039 | 0.053| 0.085| 0.141| 0.233| 0.014| 0.011|
| AEG           | 0.092| 0.086 | 0.086 | 0.079| 0.106| 0.163| 0.270| 0.038| 0.027|

Pairwise $F_{ST}$ values between population represented below diagonal and pairwise standard error (SE) above diagonal. All comparisons in the table were significant (p < 0.05) after Bonferroni correction of rejection probabilities. Population group abbreviations: NSEC (Kiel, Denmark, Germany, Scotland 1 and 2, English Channel); BISC-1 (Bisc-5029 and Bisc-5020); BISC-2 (Bisc-5001); CIAT (Central Iberian-Atlantic: Galicia, Aveiro, Portugal-S); SIAT (Southern Iberian-Atlantic: Gulf of Cadiz, Alboran Sea); CAN (Canary Islands); SAF (South Africa); NWMD (North-western Mediterranean: Tarragona, Delta); ADR (Adriatic Sea); and AEG (Aegean Sea).

doi:10.1371/journal.pone.0042201.0002

On geographical scales of hundreds of kilometers, genetic data resolved at least ten homogeneous population groups (Table 2) that likely arose after the last glacial maximum as a consequence of limited gene flow between groups, or may reflect secondary contact between previously isolated populations. These genetic groups differ in some cases from the stocks defined for management in the species.

### Population Structure and Management Implication

On large geographical scales, two genetic groups of populations were identified (Fig. 2, 3) that appear to be associated with different oceanic regimes. One group includes populations in the North Sea, Bay of Biscay and Mediterranean Sea that inhabit wide-shelf areas characterized by larval-retention mechanisms. These anchovies spawn and overwinter in more oceanic offshore areas [54–59]. The ecology of anchovies in this group reflects the classic ‘ocean triads’ model of a productive environment that is also conducive to egg and larval retention [1]. The other group included populations extending from Galicia on the Iberian Peninsula to at least the Canary Islands, but also included southern Africa. Populations in these areas inhabit narrow-shelf waters associated with upwelling. In contrast to the first group, these anchovies spawn and overwinter in coastal areas to avoid offshore advection driven by upwelling [60–64]. Populations from the later group are characterized by isolation by distance (IBD), common among populations of small- to medium-sized pelagic fishes with life histories similar to those of European anchovies [65,66]. The prevalence of IBD in European anchovies and in these marine species implies an approach to equilibrium between migration and genetic drift [67].

On geographical scales of hundreds of kilometers, genetic data resolved at least ten homogeneous population groups (Table 2) that likely arose after the last glacial maximum as a consequence of limited gene flow between groups, or may reflect secondary contact between previously isolated populations. These genetic groups differ in some cases from the stocks defined for management in the species.

Genetic homogeneity among anchovy samples in the North Sea-English Channel area and genetic divergence between this northern-most population group (NSEC, Table 2) and those in the Bay of Biscay anchovies indicate that North Sea-English Channel anchovy populations should be managed as an independent stock unit. Anchovies have not been historically abundant in the North Sea region, but abundances have increased in the last few years, possibly in response to oceanic warming [58,68] and may support fisheries in the future.

Regarding southern European populations, the genetic division found in the east Iberian Atlantic area appears to correspond to morphological differences. Caneco et al. [69] reported morphometric differences between anchovies from the Gulf of Cadiz and the Portugal area, suggesting that these differences may be due to adaptation to environmental differences between areas. Asynchronous abundances also indicate demographic independence between populations in these two areas. Large populations in Galicia and Portugal historically supported large harvests until the
early 1960s when these populations declined [70,71]. A southern center of abundance is located in the Gulf of Cadiz, which presently supports a large fishery. However, currently these two groups (Portugal and Gulf of Cadiz anchovies) are managed as a single stock (ICES Division IXa; Fig. 1) [72].

In the Mediterranean, our sampling does not allow us to define the geographical boundaries between the four genetically distinct groups (Portugal and Gulf of Cadiz anchovies) as managed by a single stock (ICES Division IXa; Fig. 1) [72].

Figure 4. Median Joining Network of haplotypes in European anchovies defined by 12 of the 15 mitochondrial DNA SNPs. (Cytb-318, Dloop-323 and Dloop-336 were given a weight of 0). Phylogroups A and B were separated by 3 mutational steps, while other haplotypes were separated by 1 mutational step. Numbers along the branches specify the mutated SNP: 1) CYTB-60, 2) CYTB-156, 3) CYTB-318, 4) CYTB-516, 5) CYTB-534, 6) Dloop-323, 7) Dloop-336, 8) Dloop-466, 9) Dloop-568, 10) mt-12S-358, 11) mt-12S-390, 12) mt-12S-434, 13) mt-165-1176, 14) mt-165-1180, 15) mt-165-1227.

doi:10.1371/journal.pone.0042201.g004

spawning, larval retention and growth, the ‘ocean triad hypothesis’ proposed by Agostini & Bakun [1] based on a consideration of oceanic processes. A close relationship between anchovies in Alboran Sea and Atlantic anchovies has been reported (Table 2, Fig. 2, 3), suggesting that the Almeria-Oran front is a barrier to dispersal for anchovies [10–12]. The latter genetic similarity throws a new light on the harvest management of anchovies in the Alboran area. Under the current management policy, Alboran Sea anchovies are grouped with NW Mediterranean anchovies. However, the results of our study of SNPs and a previous study
of allozymes [12] show that the anchovies in the Alboran Sea are more closely related to populations in the adjacent Gulf of Cadiz. These two stocks together, therefore, represent a more meaningful management unit and should be treated as a single stock. These two stocks together, therefore, represent a more closely related to populations in the adjacent Gulf of Cadiz.

The analyses of multiple nDNA and mtDNA SNP markers confirm and expand previous population genetic and phylogeographic hypothesis for European anchovy populations. The nuclear SNPs yielded larger values of $F_{ST}$ than did microsatellite markers and, hence, provided a greater amount of power to detect population structure. The results of our study and of several other studies demonstrate hierarchical genetic differences between populations on different temporal and spatial scales. Some divergences between populations in the Mediterranean and in the Atlantic can be explained by random drift in populations isolated by current gyres, by frontal systems, or by distance. However, some small-scale differences between populations in the NE Atlantic show a closer relationship to populations in the NW Mediterranean. This geographically unusual relationship is confirmed here with nuclear SNP markers. The relationship cannot be explained by contemporary gene flow, because genetically differentiated populations inhabit the intervening areas along a potential dispersal route. This similarity more likely reflects ancient founding events by colonists from a common ancestral population. The hypothesis of a common shared ancestor had been previously suggested [9,10,13,19], but the inclusion of a large number of samples in the NE Atlantic and the analysis of a large number of markers in our study allows us to refine this hypothesis.

Other regional relationships also appear to reflect ancient dispersals and founder events. For example, small values of $F_{ST}$ between northern North Sea-English Channel and some populations in the Bay of Biscay indicate a close genetic relationship between these groups that may reflect historical biogeographic relationships. Present-day northern populations and those in the Bay of Biscay have increased in abundance since the 1990s [58]. Beare et al. [30] proposed that the recent expansion of anchovies in the North Sea was due to a northern shift in the distributions of southern populations because of ocean-climate warming. However, Petitgas et al. [68] countered that population growth in the North Sea was an expansion of remnant populations in response to a widening of their thermal habitat and to higher levels of ocean productivity. The small, but significant, differences between northern populations and those in the Bay of Biscay support the latter hypothesis and suggest that the progenitors of populations in northern waters, which became available for colonization only after 7000 years ago, likely dispersed from previously established populations in the Bay of Biscay.

Our genetic results also confirm dispersals of anchovies from European waters across the equator to southern Africa. The major European mtDNA lineages also occur in southern African anchovies [18], and allozyme frequencies in southern African populations are similar to those in European populations [75]. While allozyme diversities in southern African anchovies are as large, or larger, as those in European populations [6,7,76], both nuclear SNPs and mtDNA SNPs show lower diversity in our study. Based on a single sample of European anchovies from the NW Mediterranean, Grant & Bowen [18] suggested that similarities in allozyme and cytochrome-b frequencies with Mediterranean anchovies indicated a Mediterranean origin of present-day populations of southern African anchovies. However, our results for populations in the NE Atlantic show a closer relationship between some NE Atlantic and southern African anchovies.

**Conclusions**

The analyses of multiple nDNA and mtDNA SNP markers confirm and expand previous population genetic and phylogeographic hypothesis for European anchovy populations. The nuclear SNPs yielded larger values of $F_{ST}$ than did microsatellite markers and, hence, provided a greater amount of power to detect population structure. The results of our study and of several other studies demonstrate hierarchical genetic differences between populations on different temporal and spatial scales. Some divergences between populations in the Mediterranean and in the Atlantic can be explained by random drift in populations isolated by current gyres, by frontal systems, or by distance. However, some small-scale differences between populations in the Bay of Biscay and in other areas may be due to adaptive divergences in life-history traits. These life-history contrasts are
ideal starting points for genomic studies to understand the genetic basis of adaptation.

Our multi-locus approach also revealed two major groups of European anchovies that are associated with different oceanic systems. One group inhabits productive wide-shell ecosystems with physical mechanisms that enhance nutrient enrichment, an abundance of larval food and egg and larval retention, the ‘ocean triads’ [1]. Because these ecosystem features are conducive to population growth and local retention, these anchovies tend to spawn and overwinter in offshore oceanic areas, without being lost to the system. In contrast, the other group inhabits narrow-shell areas that are dominated by upwelling. Anchovies in these populations spawn and overwinter in coastal areas to avoid offshore advection during upwelling. While the ecologies of these two groups have been described in the literature, we show here that these ecological groups are also genetically differentiated from each other to a small extent.

The analysis of the mtDNA resolves two deeply separated phylogroups in European anchovies. The frequencies of these lineages vary among populations and give insights into patterns of dispersal and colonization during the Pleistocene climate cycles. One conundrum is the genetic similarity between some anchovy lineages and gives insights into patterns of each other to a small extent.

Our multi-locus approach also revealed two major groups of European anchovies: population structure and management implications. ICES J Mar Sci 63: 1625–1644.

We are grateful to colleagues who provided samples from the following institutes: Instituto Español de Oceanografía (IEO, Malaga); Institut de Ciències del Mar (CMIMA-CSIC, Barcelona); Department of Biology from the University of Genit; Biodiversity and Conservation Group of the Center for Marine Sciences (CCMAR, Portugal); Institute of Fisheries and Sea Research (INRB/IPIMAR, Portugal); Department of Environmental Affairs (DEA, South Africa); Leibniz Institute of Marine Sciences from Kiel; Fisheries Research Services (FRS) from the Marine Laboratory of Aberdeen; Institute for Marine Resources and Ecosystem Studies (IMARES, Netherlands); Federal Research Institute of Nutrition and Food Max Rubner-Institut (MRIF); Marine Biological Association (MBA, Plymouth) and the Food Research Unit and Marine Research Unit of AZTI-Tecnalia. The authors thank I. Miguel and F. Rendo (Sequencing and Genotyping Service of the University of the Basque Country, SGIker) for their technical assistance.

Author Contributions
Conceived and designed the experiments: IZ XI AE. Performed the experiments: IZ. Analyzed the data: IZ MI. Contributed reagents/materials/analysis tools: IZ MI AA MAP XI AE. Wrote the paper: IZ MI AA CM XI AE WSG.

References
1. Agostini VN, Bakun A (2002) ‘Ocean triads’ in the Mediterranean Sea: physical mechanisms potentially structuring reproductive habitat suitability (with example application to European anchovy, Engraulis encrasicolus). Fish Oceanogr 11: 129–142.
2. Palomo J (1992) Spawning of anchovy Engraulis encrasicolus in the Northwestern Mediterranean relative to hydrographic features in the region. Mar Ecol Prog Ser 79: 215–223.
3. Kuhlma OV, Kuhlma VV (1984) Genetic differentiation and reproductive relationship between the Black Sea race and the Azov Sea race of the European anchovy. II. Genetic differences between the Black/Sea race and the Azov Sea race and their interracial heterogeneity. Genetica 20: 309–313.
4. Pasteur N, Berrebi P (1985) Structure génétique des populations naturelles d’animaux larvaires 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.
5. Spanakis E, Tsimenides N, Zouros E (1989) Genetic differences between populations of sardine, Sardinus pilchardus, and anchovy, Engraulis encrasicolus, in the Aegean and Ionian seas. J Fish Biol 35: 417–417.
6. Bembo DG, Carvalho GR, Cingolani N, Arneri E, Giannetti G, et al. (1996a) Allozymic and morphometric evidence for two stocks of the European anchovy (Engraulis encrasicolus) calls for further investigation of its genetic structure and biogeography. ICES J Mar Sci 66: 2176–2182.
7. Borsa P (2002) Allozyme, mitochondrial-DNA, and morphometric variability indicate cryptic species of anchovy (Engraulis encrasicolus). Biol J Linn Soc 75: 261–269.
8. Borsa P, Collet A, Durand JD (2004) Nuclear-DNA markers confirm the presence of two anchovy species in the Mediterranean. C R Biol 327: 1113–1123.
9. Bouchenak-Khelladi Y, Durand JD, Magoulas A, Borsa P (2008) Geographical structure of European anchovy: A nuclear-DNA study. J Sea Res 59: 269–278.
10. Ivanova PP, Dobrovolov IS (2006) Population-genetic structure on European anchovy (Engraulis encrasicolus) calls for further investigation of its genetic structure and biogeography. ICES J Mar Sci 63: 2176–2182.
11. Grant WS, Bowern BW (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.
12. Totschnig K (1995) Late Devensian and Holocene shorelines of the British Isles and North Sea from models of glacio-hydro-isostatic rebound. J Geol Soc 152: 667–683.
13. Marko PB, Hoffman JM, Emme SA, McGovern TM, Keever C, et al. (2010) The ‘expansion-contraction’ model of Pleistocene biogeography: rocky shores suffer a sea change?. Mol Ecol 19: 146–169.
14. Totschnig K (1995) Late Devensian and Holocene shorelines of the British Isles and North Sea from models of glacio-hydro-isostatic rebound. J Geol Soc 152: 437–440.
15. 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 Phylogenet Evol 39: 734–746.
16. Kristoffersen JB, Magoulas A (2008) Population structure of anchovy Engraulis encrasicolus L. in the Mediterranean Sea inferred from multiple methods. Fish Res 91: 167–193.
17. Totschnig K (1995) Late Devensian and Holocene shorelines of the British Isles and North Sea from models of glacio-hydro-isostatic rebound. J Geol Soc 152: 437–440.
18. Marko PB, Hoffman JM, Emme SA, McGovern TM, Keever C, et al. (2010) The ‘expansion-contraction’ model of Pleistocene biogeography: rocky shores suffer a sea change?. Mol Ecol 19: 146–169.
19. Totschnig K (1995) Late Devensian and Holocene shorelines of the British Isles and North Sea from models of glacio-hydro-isostatic rebound. J Geol Soc 152: 437–440.
51. Ryynänen HJ, Tonteri A, Vasemägi A, Primmer CR (2007) A comparison of single-nucleotide polymorphisms in inferences of population history. Trends Ecol Evol 18: 249–256.

52. Jones B, Walsh D, Werner L, Fiumera A (2009) Using blocks of linked single nucleotide polymorphisms as highly polymorphic genetic markers for parentage analysis. Mol Ecol Resour 9: 467–497.

Genetic Structure of European Anchovy

25. Miklejdoj CD, Montooth KL, Rand DM (2007) Positive and negative selection on the mitochondrial genome. Trends Genet 25: 259–263.

26. Karl SA, Toonen RJ, Grant WS, Bowen BW (2012) Common misconceptions in molecular ecology. Echoes of the modern synthesis. Mol Ecol doi: 10.1111/j.1365-294X.2012.05576.x.

27. Sunnucks P (2000) Efficient genetic markers for population biology. Trends Ecol Evol 15: 199–205.

28. Brumfield RT, Beehl P, Nickerson DA, Edwards SV (2003) The utility of single-nucleotide polymorphisms in inferences of population history. Trends Ecol Evol 18: 249–256.

29. Guichoux E, Lagache I, Waguer S, Chaumeil P, Leger P, et al. (2011) Current trends in microsatellite genotyping. Mol Ecol Resour 11: 591–611.

30. Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Mol Ecol 15: 1419–1439.

31. Molecular Ecology Resources Primer Development Consortium. Zarronaimidea I, Alabbage A, Brion M, Manzano C, Pardo MA, et al. (2012) Permanent Genetic Resources added to Molecular Ecology Resources Database 1 October 2011–30 November 2011. Mol Ecol Resour 12: 374–376.

32. Rousset F (2000) Genepop’007: a complete reimplementation of the Genepop software for Windows and Linux. Mol Ecol Resour 8: 103–106.

33. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68: 978–989.

34. Foll M, Gaggiotti OE (2008) A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. Genetics 190: 977–993.

35. Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices, Version 2.9.3. [http://www2.unil.ch/popgen/softwares/fstat.htm].

36. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

37. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure by using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567–1587.

38. Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23: 1801–1806.

39. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14: 2611–2620.

40. Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. Mol Ecol Resour 4: 137–138.

41. Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics 103: 767–779.

42. Langella O (1999) Populations 1.2.28. CNRS UPR9034.

43. Bonnet E, Van de Peer Y (2002) ZT: a software tool for simple and partial population structure. Mol Ecol Resour 2: 1801–1806.

44. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism. Mol Ecol Resour 9: 451–459.

45. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585–595.

46. Weir BS (1996) Genetic Data Analysis II. Sinauer Associates, Sunderland, MA.

47. Zardoya R, Castiño R, Granda C, Favre-Krey E, Caetano S, et al. (2004) Differential population structuring of two closely related fish species, the mackerel (Scomber scombrus) and the club mackerel (Scomber scombrus), in the Mediterranean Sea. Mol Ecol 13: 1785–1798.

48. Ruzzante DE, Mariani S, Bekkevold D, Andre C, Mosegaard H, et al. (2006) Morphometric variation among anchovy Engraulis encrasicolus I, Albaina A, Iriondo M, Manzano C, Pardo MA, et al. (2012) Permanent Genetic Resources added to Molecular Ecology Resources Database 1 October 2011–30 November 2011. Mol Ecol Resour 12: 374–376.

49. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism. Bioinformatics 25: 1451–1452.

50. Bandeir HJ, Forster P, Rohd A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Ecol Evol 16: 37–48.

51. Rynanen HJ, Tonteri A, Vasemägi A, Primmer CR (2007) A comparison of biallelic markers and microsatellites for the estimation of population and trends in microsatellite genotyping. Mol Ecol Evol 16: 37–48.

52. Meiszner SI, Taylor BI, Archer FI, Martinson A, Taylor BI, et al. (2011) Sperm whale population structure in the eastern and central North Pacific inferred by the use of single-nucleotide polymorphisms, microsatellites and mitochondrial DNA. Mol Ecol Resour 11: 278–299.

53. Motos L, Souto A, Valencita V (1996) The spawning environment of the Bay of Biscay anchovy (Engraulis encrasicolus L.). Sci Mar 60: 117–149.

54. Cuttitta A, Carini V, Patini B, Bonanno A, Basilone G, et al. (2003) Anchovy egg and larval distribution in relation to biological and physical oceanography in the Strait of Sicily. Hydrobiologia 503: 117–120.

55. Mesnick SI, Taylor BI, Archer FI, Martinson A, Taylor BI, et al. (2011) Sperm whale population structure in the eastern and central North Pacific inferred by the use of single-nucleotide polymorphisms, microsatellites and mitochondrial DNA. Mol Ecol Resour 11: 278–299.

56. Zardoya R, Castiño R, Granda C, Favre-Krey E, Caetano S, et al. (2004) Morphometric variation among anchovy Engraulis encrasicolus I, Albaina A, Iriondo M, Manzano C, Pardo MA, et al. (2012) Permanent Genetic Resources added to Molecular Ecology Resources Database 1 October 2011–30 November 2011. Mol Ecol Resour 12: 374–376.

57. Pestana G (1989) Manancial IberoAtlanítico de Sardinha (Sardinia pilchardus Walb) na sua avaliação e medidas de gestão. Ph.D. thesis. Univ. Lisboa.

58. ICES (2010) Report of the Working Group on Anchovy and Sardine (WGANSA). ICES document CM 2010/ACOM: 16.

59. Baldo F, Drake P (2002) A multivariate approach to the feeding habits of small fishes in the Guadalquivir estuary. J Fish Biol 61: 21–32.

60. Drake P, Arias AM, Baldo F, Cuesta JA, Rodriguez A, et al. (2002) Spatial and temporal variation of the nekton and hyperbenthos from a temperate European estuary with a regulated freshwater inflow. Estuaries 25: 451–468.

61. Baldo F, Boulan A, González-Ortégano E, Baldo F, Vilas C, et al. (2007) Spatial-temporal distribution of early life-stages of European anchovy Engraulis encrasicolus I, within a European temperate estuary with regulated freshwater inflow: effects of environmental variables. J Fish Biol 70: 1698–1709.

62. Caneco B, Silva A, Morais A (2004) Anchovy spawning in the Mira estuary (southwestern Portugal). J Fish Biol 67: 1242–1265.

63. Re´ P (1996) Anchovy spawning in the Mira estuary (southwestern Portugal). Sci Mar 60: 141–153.

64. Aldanondo N, Cotano U, Tiepolo M, Boyra G, Irigoien X (2010) Growth and larval distribution in relation to biological and physical oceanography in the Bay of Biscay based on otolith microstructure and chemistry. Fish Oceanogr 19: 196–208.

65. Zardoya R, Castiño R, Granda C, Favre-Krey E, Caetano S, et al. (2004) Anchovy spawning in the Mira estuary (southwestern Portugal). Sci Mar 60: 141–153.

66. Pimentel de Castro E, Peña A, Exposito H, Trivino SE, et al. (2011) Sperm whale population structure in the eastern and central North Pacific inferred by the use of single-nucleotide polymorphisms, microsatellites and mitochondrial DNA. Mol Ecol Resour 11: 278–299.