

Submitted to:

*Molecular Phylogenetics and Evolution*

Revised version: August 3<sup>th</sup>, 2005

**Signature of an early genetic bottleneck in a population of  
Moroccan sardines (*Sardina pilchardus*)**

Touriya Atarhouch<sup>1</sup>, Lukas Rüber<sup>2</sup>, Elena G. Gonzalez<sup>2</sup>, Eva Albert<sup>2</sup>, Mohamed Rami<sup>1</sup>,  
Allal Dakkak<sup>1</sup>, and Rafael Zardoya<sup>2</sup>

<sup>1</sup>*Département de Parasitologie, Laboratoire de Biologie Moléculaire, BP 6202, Institut  
Agronomique et Vétérinaire Hassan II; 10101; Rabat, Morocco*

<sup>2</sup>*Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias  
Naturales, CSIC, José Gutiérrez Abascal, 2; 28006 Madrid, Spain*

Corresponding author:

Rafael Zardoya

Museo Nacional de Ciencias Naturales, CSIC

José Gutiérrez Abascal, 2

28006 Madrid

Spain

Tel: +34-91-4111328 ext 1129

Fax: +34-91-5645078

Email: rafaz@mncn.csic.es

**Abstract**

Fishery assessment models meant to determine sustainability of commercial marine fish failed to predict recent stock collapses due to overexploitation. One flaw of assessment models is that they strongly rely on catch and age-composition statistics, but largely ignore the genetic background of studied populations. We examined population genetic structure of sardine (*Sardina pilchardus*) in the centraleastern and northeastern Atlantic Ocean and Mediterranean Sea in order to aid fishery management of this heavily fished small pelagic species. We found that sardine has a striking mitochondrial control region, and sequenced a fragment of 387 bp of its 5'-end in 261 individuals collected off the coasts of Morocco (Dakhla, Tantan, Safi, Larache, Nador), Portugal (Quarteira), Spain (Pasajes, Barcelona), and Greece (Kavala). High levels of haplotypic diversity rendered a rather unresolved NJ phylogeny. The recovered tree had no phylogeographic structuring except for the clustering of 13 individuals of Safi. In contrast, individuals grouped together according to the presence or absence of a 13-bp insertion in the sequence.  $\Phi_{st}$  pairwise comparisons and molecular variance analyses supported genetic differentiation between the population of Pasajes (Bay of Biscay), and those of the Mediterranean Sea and Moroccan coast, with a contact zone around the Strait of Gibraltar. This result confirms the existence of two subspecies, *S. pilchardus pilchardus* and *S. pilchardus sardina* that were previously identified based on meristics and morphometry. Mismatch distribution analysis showed that sardine populations are expanding since the Pleistocene. Surprisingly, the population of Safi showed strong and statistically significant levels of genetic differentiation that could be related with isolation and genetic drift. Comparative analysis of the Safi population versus the rest including mismatch distributions, and a Bayesian skyline plot suggest that the Safi population likely underwent an early genetic bottleneck. The genetic singularity of the Safi population could have been responsible for the historical collapse of this sardine stock in 1970s.

## Introduction

Marine pelagic fishes of commercial interest are severely overfished, and their populations are in many cases being driven to collapse (Hutchings, 2000; Myers and Worm, 2003; Pauly *et al.*, 2003). While there is clear evidence that depletion of fish species has a negative impact on the marine ecosystem (Jackson *et al.*, 2001; Frank *et al.*, 2005), the long-term consequences in terms of loss of genetic diversity are largely unknown (Hauser *et al.*, 2002). Indeed, the paramount question on the geographic extent and population genetic structure of the stocks (*i.e.* populations with a single spawning ground to which adults return every year) remains largely unanswered for many commercial marine fish species. This is particularly worrisome since failure to detect population demes, and their genetic status, may lead to local overexploitation, extreme reduction of genetic variability (*i.e.* genetic bottleneck), and ultimately to severe depletions due to loss of fitness and the viability to withstand survival challenges (Utter, 1991; Cook *et al.*, 1997; Hutchings, 2000; Knutsen *et al.*, 2003).

The sardine (*Sardina pilchardus*, Walbaum 1792) is a small pelagic fish species of most important interest to fisheries in centraleastern and northeastern Atlantic Ocean. This cupleiform is found from the North Sea to Senegal, as well as in the Mediterranean, Marmara, and Black seas (Parrish *et al.*, 1989). The global catches reported for this species in the last years were around one million tons per year, being Morocco and Spain the countries with largest reported captures (FIGIS, 2004). Several studies based on meristics and morphometry attempted to determine the structure of sardine populations within its area of distribution (*e.g.* (Silva, 2003). However, results were in general not conclusive in terms of delimiting stock boundaries, and in some cases observations seemed contradictory. Studies based on differences in vertebral counts identified four stocks in the centraleastern and

northeastern Atlantic Ocean: The septentrional Atlantic stock, which extends from the North Sea (57°N) to the north coast of Spain (43°N); the meridional Atlantic group, which occurs off the Portuguese coasts until the Strait of Gibraltar (from 43°N to 36°N); the Moroccan stock, which extends from Cap Spartel (36°N) to Cap Juby (28°N); and the Saharian stock, which is distributed from Cap Juby to Cap Blanc in Mauritania (21°N) (Parrish *et al.*, 1989). A study (Belveze and Erzini, 1983) that was focused exclusively on Moroccan sardine populations supported a strong influence of the Canary current upwelling on the distribution of sardines, and hence, on the geographic boundaries of the stocks. Based on the results of this study, as well as on the integration of biological data on meristics, morphometrics, growth, migrations, parasites, distribution, and reproduction, three separate stocks were recognized off the Northwest African coast (FAO, 1998): Northern stock, defined between Cape Spartel and El Jadida (32°N); Central stock, identified between Safi (32°N) and Cap Bojador (26°N); and Southern stock, established southwards Cap Bojador. Of these, the Central stock sometimes is further subdivided into two fishing zones: Safi - Sidi Ifni and Sidi Ifni – Cap Bojador. Although, the central stock experienced a significant reduction in biomass in the 1970s, high recruitment levels in the 1980s brought back sardine abundance to the levels achieved before the collapse (Belveze and Erzini, 1983; Kifani, 1995).

Biological data need to be contrasted with genetic data in order to discern whether observed phenotypic differences represent independent stocks or local adaptations (Tinti *et al.*, 2002). As other marine pelagic fishes, sardines show schooling and migratory behavior, as well as great dispersal capabilities both at the larval and adult stages. Therefore, in the absence of barriers, the null hypothesis is that sardine populations should show extensive gene flow and little genetic structuring. The alternative hypothesis is that local recruitment may act against

gene flow, and eventually lead to genetic subdivision. So far, the few studies that have focused on population genetics of sardines showed low levels of genetic differentiation in the Aegean Sea (Spanakis *et al.*, 1989), the Spanish Mediterranean coast (Ramon and Castro, 1997), and the Adriatic Sea (Tinti *et al.*, 2002).

The Almeria-Oran oceanographic front (Tintoré *et al.*, 1988) has been shown to produce one of the main transitions for many marine organisms between the Mediterranean and the Atlantic Ocean (Naciri *et al.*, 1999; Bahri-Sfar *et al.*, 2000; Bargelloni *et al.*, 2003; Zardoya *et al.*, 2004). This oceanographic front also seems to have an effect on sardine population structuring. Yet, the transition between the Atlantic Ocean and the Mediterranean Sea for sardine is not so clearly defined as in other fish species. Instead, some authors (Andreu, 1969; Parrish *et al.*, 1989) have proposed based on phenotypic variation mainly in gill raker counts and head length the existence of two different subspecies that show a different distribution pattern with respect to the Atlantic Ocean-Mediterranean Sea transition: *Sardina pilchardus pilchardus* in the Eastern Atlantic from the North Sea to Southern Portugal, and *Sardina pilchardus sardina* in the Mediterranean Sea and off the Northwest African coast.

We examined genetic variability of sardine with particular emphasis on populations off shore Morocco using mitochondrial control region nucleotide sequence data. Phylogenetic analyses of sardine sequence data, and statistical comparison of haplotype frequencies among sardine populations were used 1) to test the null hypothesis of no population genetic structure for sardines; 2) to determine the taxonomic validity of the two proposed subspecies of sardine, and whether the boundaries of their distribution are determined by the Almeria-Oran front; and 3) to assess whether the severe decline of the Moroccan sardine central stock in the

1970s due to strong levels of fishing effort could be correlated with any previous genetic singularity of this stock.

## **Materials and Methods**

### *Sampling*

Samples were collected in five landing ports of Morocco (Dakhla, Tantan, Safi, Larache, and Nador), one of Portugal (Quarteira), one of Greece (Kavala), and two of Spain (Barcelona and Pasajes San Pedro). Sampling locations are shown in Figure 1. The date and number of samples collected per site are listed in Table 1. Samples were stored either in absolute ethanol or frozen at  $-20^{\circ}\text{C}$ .

### *DNA extraction, PCR amplification, and automated sequencing*

Total DNA was isolated from muscle tissue. After overnight digestion at  $55^{\circ}\text{C}$  in extraction buffer (EDTA 0.1 M, Tris-HCl pH 8.0 0.05 M, SDS 1%, and Proteinase K 0.2 mg/ml), DNA was purified using a standard phenol/chloroform extraction followed by an ethanol precipitation. Universal primers (Ostellari *et al.*, 1996) designed to amplify by PCR the 5'-end of fish mitochondrial control regions did not work on sardine samples. Therefore, the entire mitochondrial control region (Fig. 2; Genbank accession No DQ102482) of a specimen from Larache was amplified by PCR using two primers, LN20 (5'- ACCACTAGCACCCAAAGCTA-3') and HN20 (5'-GTGTTATGCTTTAGTTAAGC-3') (Chikhi *et al.*, 1997). A standard PCR reaction was performed under the following conditions: a single cycle of 5 min at  $94^{\circ}\text{C}$ , followed by 35 cycles of 60 S at  $94^{\circ}\text{C}$ , 60 S at  $50^{\circ}\text{C}$ , and 90 S at  $72^{\circ}\text{C}$ , with a final cycle of 5 min at  $72^{\circ}\text{C}$ . The PCR product was purified by ethanol precipitation, and sequenced with the LN20 and HN20 primers, using the BigDye Terminator Sequencing Ready Reaction v3.0 kit (Applied Biosystems) and following manufacturer's instructions on an ABI 3700 automated sequencer. Two

internal PCR primers (TYW 5'- CTGAGCATGGCACTTATCCCTAC-3' and TYH 5'- CCTGAAGTAGGAACCAGATG -3') were designed in order to complete the entire sequence.

A fragment of about 700 bp of the 5'-end of the sardine mitochondrial control region was PCR amplified in all sampled individuals using the LN20 and TYR (5'- CATAGCGTGGTTGCGACCACGGTT-3') primers, and the conditions above described, and the first 400 bp were sequenced with primer LN20. The new sardine partial mitochondrial control region sequences were deposited in Genbank under the accession numbers DQ139463-DQ139723.

#### *Phylogenetic, population genetic, and demographic analyses*

Nucleotide sequences were aligned using Clustal X v1.83 (Thompson *et al.*, 1997), and further revised by eye. Gaps resulting from the alignment were excluded from phylogenetic and population genetic analyses. Phylogenetic relationships were reconstructed with PAUP v4.0b10 (Swofford, 1998) using the neighbour-joining (NJ) method (Saitou and Nei, 1987). The Akaike information criterion (AIC) implemented in Modeltest v3.6 (Posada and Crandall, 1998) was used to select the evolutionary model that best fit the empirical data set. The inferred evolutionary model and parameters were used to estimate maximum likelihood distances. Because of the singularity of the sardine 5'-end control region, no clupeiform outgroups could be used, and the NJ tree was rooted using the midpoint rooting option. Robustness of the resulting tree was tested with bootstrapping (Felsenstein, 1985).

Population genetic statistics were estimated using Arlequin 2001 (Schneider *et al.*, 2000) and DnaSP 4.0 (Rozas *et al.*, 2003). Descriptive statistics such as the number of polymorphic sites ( $S$ ), haplotype diversity ( $H_d$ , (Nei, 1987), nucleotide diversity ( $\pi$ , (Nei, 1987) and the average number of pairwise nucleotide differences

( $k$ , (Tajima, 1983) were determined for each population. Pairwise haplotype divergences were estimated with Arlequin 2001 using the fixation index  $\Phi_{ST}$  (Excoffier *et al.*, 1992), which includes information on mitochondrial haplotype frequency (Weir and Cockerham, 1984), and genetic distances (Tamura Nei (Tamura and Nei, 1993) with gamma correction, TrN+G, the best fit model among those available in Arlequin 2001). Significance of pairwise population comparisons was tested by 20,000 permutations. In order to determine the amount of genetic variability partitioned within and among populations, an analysis of molecular variance (AMOVA) was performed in Arlequin 2001 (Excoffier *et al.*, 1992). Permutations (N=20,000) were used to construct null distributions and to test the significance of variance components (Guo and Thompson, 1992). In all multiple tests,  $P$ -values were adjusted using the sequential Bonferroni correction (Rice, 1989).

Historical demography of sardine populations was characterized using the mismatch distribution (Slatkin and Hudson, 1991; Rogers and Harpending, 1992; Schneider and Excoffier, 1999) as implemented in Arlequin 2001 (Schneider *et al.*, 2000) and DnaSP 4.0 (Rozas *et al.*, 2003). The observed distribution of the frequency of pairs of individuals who differ by a certain number of nucleotide differences data set was tested against constant population size and sudden population expansion models. The pattern of pairwise differences between haplotypes usually forms a unimodal wave in samples from expanding populations, whereas samples drawn from populations at demographic equilibrium yield a multimodal pattern of numerous sharp peaks (Slatkin and Hudson, 1991; Rogers and Harpending, 1992; Schneider and Excoffier, 1999). The validity of the assumed stepwise expansion model was tested with a parametric bootstrapping approach (1,000 permutations). Deviations from the sudden population expansion model were further tested using the Harpending's raggedness index ( $Hr_i$ ; (Harpending, 1994). To

test for deviations from neutrality (as would be expected under population expansion) we used Tajima's D (Tajima, 1989) and Fu's  $F_s$  (Fu, 1997) tests as implemented in DnaSP 4.0 (Rozas *et al.*, 2003).

To address the question of whether the genetic pattern exhibited by the Safi population could be the result of genetic bottleneck and drift, we performed several comparative analyses. The number of shared haplotypes between Safi and other populations was determined using Arlequin 2001. Mismatch distributions of pairwise nucleotide differences of the Safi population and the other populations were estimated and contrasted using DnaSP 4.0 (Rozas *et al.*, 2003). Past population dynamics of the Safi population were estimated with a Bayesian skyline plot model using BEAST v. 1.2 (Drummond *et al.*, 2005). The Bayesian skyline plot model generates a posterior distribution of effective population size through time using a Markov chain Monte Carlo (MCMC) sampling. The parameter  $m$  that represents the number of grouped intervals was set to 12. The MCMC analysis was run for  $20 \times 10^6$  generations (sampled every 1,000 iterations), of which the first 10% was discarded as burn-in. The substitution model used was GTR+G. The median and corresponding credibility intervals of the Bayesian skyline plot were depicted using Tracer v. 1.2.1 (available at <http://evolve.zoo.ox.ac.uk/software/>).

## Results

### *The sardine mitochondrial control region*

The entire mitochondrial control region of the sardine is 1642 bp in length, and as in most fishes, it is localized between the tRNA<sup>Pro</sup> and the tRNA<sup>Phe</sup> genes. The overall base composition of the L-strand of the control region is A: 28%, C: 24%, G: 21%, and T: 27%. In other vertebrates, this region includes the origin of H-strand replication, the sites of initiation of both H- and L-strand transcription, and several

motifs involved in the regulation of both processes. The main features of sardine mitochondrial control region are shown in Figure 2. The sardine mitochondrial control region can be divided into a conserved central domain that can be aligned with other fish mitochondrial control regions (Lee *et al.*, 1995), and the more variable 5'- and 3'-ends. Two complete conserved sequence blocks (CSB2 and CSB3) were identified at the 3'-end (Lee *et al.*, 1995). However, the most striking feature of the sardine mitochondrial control region is an extended 5'-end that includes a poly-T (18 bp) stretch, a reverse and complementary CSB-2 motif (CSB2'), and five direct and imperfect repeats (each of 23-bp) in tandem. This unusual 5' end would explain the failure of published universal fish primers (e.g. (Ostellari *et al.*, 1996) in PCR amplifying the sardine control region.

#### *High levels of intrapopulation variation*

A fragment of about 400 bp of the 5'-end of the sardine mitochondrial region was sequenced in 261 individuals corresponding to nine populations. An insertion of 13 bp was found at position 25 of the sardine mitochondrial control region in some individuals of all populations. The frequency of this insertion in the different populations is shown in Figure 1. After gap exclusion, the alignment was reduced to 387 sites. Of these, 230 were polymorphic, and 137 were parsimony informative. A total of 217 transitions and 55 transversions were scored. Polymorphic sites defined a total of 242 different haplotypes. The most abundant haplotype was found in six individuals. A total of 232 haplotypes were unique. As a result overall haplotype diversity was high ( $0.999 \pm 0.001$ ). Nucleotide diversity was between 0.017-0.019 in all populations except in Pasajes that was 0.014 and in Safi that was 0.023.

Population genetic statistics are listed in Table 1.

*Phylogenetic relationships and population structuring*

The GTR (General Time Reversible; (Lanave *et al.*, 1984) (A: 0.25, C: 0.29, G: 0.22, T: 0.24; A-C: 6.37, A-G: 113.78, A-T: 6.45, C-G: 12.68, C-T: 46.02, G-T: 1.00) + G ( $\alpha=1.13$ ) was the evolutionary model that best fit the empirical sequence data according to the AIC. NJ (using maximum likelihood distances) was used to recover phylogenetic relationships among studied sardine populations (Fig. 3). The reconstructed phylogeny showed limited resolution, and main clades lacked bootstrap support. The NJ phylogeny did not show any phylogeographic structuring. However, 13 out of 29 individuals from Safi were recovered together in the same clade (Fig. 3). On the other hand, individuals sharing the 13-bp gap at position 25 of their mitochondrial control region sequence grouped together to the exclusion of those having an insertion at that position (Fig. 3). The presence of two main types of haplotypes (with and without gap) may represent retention of an ancestral polymorphism.

Genetic differentiation among sardine populations was assessed using  $\Phi_{ST}$  pairwise comparisons (Table 2).  $\Phi_{ST}$  values were in general low. Of the 36 possible comparisons, only those involving Pasajes and Safi showed statistically significant values (Table 2). The pairwise comparisons between Safi and Nador, Pasajes and Dakhla, Pasajes and Larache, and Pasajes and Quarteira were not significant. These results were further confirmed using hierarchical AMOVA tests (Table 3). The analyzed populations showed overall significant levels of genetic structuring. A three-gene pool comparison that separated the population of Pasajes from that of Safi and from the rest resulted in significant values.

### *Historical demography*

The mismatch distribution including all samples was unimodal (Fig. 4), and perfectly adjusted to the distribution predicted by the growth-decline population model. Indeed, the Harpending's raggedness index was low further indicating a significant fit of the observed and expected distributions. The parameters of the expansion model and the Harpending's raggedness index are shown in Table 4. Fu's  $F_s$  and Tajima's  $D$  statistics rendered negative values that indicated an excess of low frequency haplotypes, and hence, a significant deviation from neutrality. Mismatch indices allowed us to estimate effective female population size and the time and rate of expansion. Estimated effective population size after expansion ( $\theta_1$ ) was about 3300 times higher than before expansion. Reported mutation rates for the fish mitochondrial control region vary between  $1.1 \times 10^{-7}$  (McMillan and Palumbi, 1997) and  $3.6 \times 10^{-8}$  (Donaldson and Wilson, 1999) per site per year. Since sardines reach sexual maturity between the first and second year of life (Caponio et al., 2004), the estimated time since population growth for all populations is  $0.54-1.64 \times 10^5$  years before present.

Safi is the only population that does not share any haplotype with other populations. However, Safi has similar levels of haplotype diversity to other populations and the highest levels of nucleotide diversity. Overall, the Safi population seems to show signs of strong genetic drift. To determine whether the Safi population underwent a genetic bottleneck in the past, we compared the mismatch distributions and related parameters of the different populations (Fig. 5). The means ( $\tau$ ) of the different mismatch distributions are nearly identical with the exception of that of Safi that suggests an older onset of its population growth. The estimated time since population growth for Safi is  $0.66-2.03 \times 10^5$  years before present. The resulting output of the MCMC analysis of the Safi population using the

Bayesian skyline plot model is summarized in Figure 6. According to the plot, the Safi population experienced a faster population growth around  $0.68-2.08 \times 10^5$  years before present that slowed down around  $0.45-1.39 \times 10^5$  years before present.

## **Discussion**

During the last 50 years, fishing capacity has steadily grown, and as a result some marine fish stocks started collapsing during the 1980s (Myers and Worm, 2003). Catch statistics demonstrate that depleted stocks exhibit little signs of recovery as much as 15 years after the collapse (Hutchings, 2000). However, there is a void of studies showing the genetic status of the populations before and after collapse (Hauser *et al.*, 2002). A population that is not under equilibrium is likely to be more sensible to environmental changes, parasites and diseases, or strong loads of fishing effort. On the other hand, depleted fish stocks can eventually recover past large population sizes, but this demographic effect not necessarily may be accompanied by restoring of the original genetic diversity because of random genetic drift. Genetic studies are important to delimit stocks, and to determine the relative importance of genetic bottlenecks as cause and effect of stock collapse under circumstances of heavy fishing.

Sardines are an economically very important fishery in the Atlantic coasts of Western Europe, Morocco, and Mauritania, as well as in the Mediterranean Sea, and thus, population dynamics of this species in terms of catch statistics has been carefully monitored in the last years in order to control landings and to manage the stocks. Although sardine stocks have experienced several significant population size reductions in the last 30 years, catch statistics and stock assessment models show that overall stocks are able to recover large population sizes (tens of millions), and therefore, it is assumed that there is no serious evidence of collapse for the species

(Beverton, 1990). This conclusion, however, needs to be corroborated by studies that establish the genetic status of the different stocks. Effective population sizes ( $N_e$ ) in marine fishes are normally several orders of magnitude smaller than population sizes due to strong bias in reproductive success, and size-dependent fecundity among other causes (Hauser *et al.*, 2002). A recent study showed that because of the small effective population size, an overexploited population of New Zealand snapper suffered a significant loss of genetic diversity in recent years (Hauser *et al.*, 2002). The history of the snapper fishery recalls that of sardine fishery off Morocco coasts, both having severe stock declines due to overfishing (Belveze and Erzini, 1983; Kifani, 1995; Hauser *et al.*, 2002). We have determined genetic variability of sardine stocks in the Atlantic Ocean and Mediterranean Sea with particular emphasis on one of the most important landings areas, the upwellings off Morocco coasts. The results obtained are not only relevant for the biology and phylogeography of the species but also for fishery management.

Phylogenetic and population genetic analyses reject the null hypothesis of panmixia for the studied sardine populations, and reveal significant genetic structuring in contrast to previous, more localized, genetic analyses of sardine stocks (Spanakis *et al.*, 1989; Ramon and Castro, 1997; Tinti *et al.*, 2002). The population of Pasajes is clearly distinct from Mediterranean Sea and most Moroccan populations. This result supports the existence of two different subspecies of sardine as proposed based on meristic studies (Parrish *et al.*, 1989). The sample from Pasajes would belong to the subspecies *Sardina pilchardus pilchardus*, whereas those of the Mediterranean Sea and off Moroccan Atlantic Ocean coasts would correspond to *Sardina pilchardus sardina*. The  $\Phi_{ST}$  pairwise comparisons involving Larache or Quarteira, versus the remaining populations (except Safi) were not significant. Hence, the Atlantic Ocean region next to the Gibraltar strait seems to be a contact

zone of both subspecies. Moreover, we can conclude that the Almeria-Oran front is not an efficient phylogeographic break (Bargelloni *et al.*, 2003) for sardine since there seems to be gene flow between populations of the Mediterranean Sea and those of the Moroccan Atlantic Ocean coast. These results prompt for a differential fishery management of sardine populations from European Atlantic Ocean coasts and those from the Mediterranean Sea and the Northwest African coast. Further genetic analyses of sardine populations off Mauritania coast are needed in order to determine the southern limit of *Sardina pilchardus sardina*.

Sardine fishery is particularly intensive off Morocco coasts, and led to a severe decline of sardine captures in the 1970s around Safi (Belveze and Erzini, 1983; Kifani, 1995). Our sequence data allow testing whether the local overfishing of the Safi population and the associated collapse of the population could be related to a special genetic status. Both in the recovered NJ phylogeny as well as in the population genetic analyses, Safi showed a significantly different haplotype frequency distribution coupled with higher nucleotide diversity, and pairwise nucleotide differences. These results reflect some degree of isolation of this population that prevents its admixture with the surrounding ones, and enhances an independent demographic history. Although highly speculative, the causes for the isolation of the Safi population may be related to oceanographic barriers (e.g. gyres), a hypothesis that needs to be tested further. Moreover, the presence of highly related haplotypes that are absent in other populations would likely indicate strong genetic drift acting on the Safi population. Mismatch distribution analysis show that sardine populations are expanding since the Pleistocene glaciations. The different mean of Safi mismatch distribution, and the Bayesian skyline plot are consistent with an earlier genetic bottleneck of the sardine population at Safi. Sardines are highly dependent on upwellings, and putative past acute variations on the Moroccan upwellings might

have played an important role in provoking this early genetic bottleneck. Perhaps, the special genetic status and conditions of the Safi population after the genetic bottleneck enhanced the negative effect of the strong fishing effort off the coasts of Morocco in the 1970s, and led to the historical collapse experienced by the sardine fishery at Safi. Obviously, the likely consequence of this most recent depletion in terms of yet another potential genetic bottleneck are still not visible in the mitochondrial control region sequences. In order to better understand population dynamics of the Safi population and to corroborate its genetic singularity, future studies will need to incorporate nuclear markers such as microsatellites to contrast mitochondrial patterns. Meanwhile, our results prompt for a special fishery management of the Safi population of sardines.

#### **Acknowledgments**

We thank Sabounji Mouhenisse, Hidane Kamal, María Jesús San Sebastián, Rita Castilho, Regina Cunha, Toni Lombarte, and Greg Kray for providing some of the samples. This work received partial financial support from AEI-MAE (Agencia Española de Cooperación Internacional- Ministerio de Asuntos Exteriores) project No 168/03/P to TA and RZ. LR was supported by a Swiss Science Foundation postdoctoral fellowship (823A-061218) and EGG by a MEC (Ministerio de Educación y Ciencia) predoctoral grant.

## References

- Andreu, B., 1969. Las branquiespinas en la caracterización de las poblaciones de *Sardina pilchardus*. Inv. Pesq. Bar. 33, 425-607.
- Bahri-Sfar, L., Lemaire C., Ben Hassine O.K., Bonhomme F., 2000. Fragmentation of sea bass populations in the western and eastern Mediterranean as revealed by microsatellite polymorphism. Proc. R. Soc. Lond. B 267, 929-935.
- Bargelloni, L., Alarcon J.A., Alvarez M.C., Penzo E., Magoulas A., Reis C., Patarnello T., 2003. Discord in the family Sparidae (Teleostei): divergent phylogeographical patterns across the Atlantic-Mediterranean divide. J. Evol. Biol. 16, 1149-1158.
- Belveze, H., Erzini K., 1983. The influence of hydroclimatic factors on the availability of the sardine (*Sardina pilchardus*, Walbaum) in the Moroccan Atlantic fishery. FAO Fish. Rep. 291, 285-328.
- Beverton, R.J.H., 1990. Small marine pelagic fish and the threat of fishing: are they endangered. J. Fish Biol. 37(suppl.A), 5-16.
- Caponio, F., Lestingi A., Summo C., Bilancia M.T., Laudadio V., 2004. Chemical characteristics and lipid fraction quality of sardines (*Sardina pilchardus* W.): influence of sex and length. J. Appl. Ichthyol. 20, 530-535.
- Chikhi, L., Agnese J.-F., Bonhomme F., 1997. Fortes differences des ADN mitochondriaux de populations de *Sardinella aurita* de la mer Méditerranée et de l'Atlantique Est. Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie 320, 289-297.
- Clement, M., Posada D., Crandall K.A., 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9, 1657-1660.
- Cook, R.M., Sinclair A., Stefansson G., 1997. Potential collapse of North Sea cod stocks. Nature 385, 521-522.

- Donaldson, K.A., Wilson R.R., 1999. Amphi-panamaic geminates of snook (Percoidei-Centropomidae) provide a calibration of divergence rates in the mitochondrial DNA control region of fishes. *Mol. Phylogenet. Evol.* 13, 208-213.
- Drummond, A.J., Rambaut A., Shapiro B., Pybus O.G., 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* 22, 1185-1192.
- Excoffier, L., Smouse P.E., Quattro J.M., 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131, 479-491.
- FAO, 1998. Report of the Ad Hoc Working Group on Sardine (*Sardina pilchardus* Walb.). CEECAF/ECAF SERIES/78/7,
- Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783-791.
- FIGIS, 2004. A world overview of species of interest to fisheries. *Sardina pilchardus*, Species Identification and Data Programme. FIGIS Species Fact Sheets. FAO-FIGIS
- Frank, K.T., Petrie B., Choi J.S., Leggett W.C., 2005. Trophic cascades in a formerly cod-dominated ecosystem. *Science* 308, 1621-1623.
- Fu, Y.-X., 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147, 915-925.
- Guo, S., Thompson E., 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48, 361-372.
- Harpending, R.C., 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum. Biol.* 66, 591-600.
- Hauser, L., Adcock G.J., Smith P.J., J.H. B.R., Carvalho G.R., 2002. Loss of microsatellite diversity and low effective population size in an overexploited

- population of New Zealand snapper (*Pagrus auratus*). Proc. Natl. Acad. Sci. USA 99, 11742-11747.
- Hutchings, J.A., 2000. Collapse and recovery of marine fishes. Nature 406, 882-885.
- Jackson, J.B., Kirby M.X., Berger W.H., Bjorndal K.A., Botsford L.W., Bourque B.J., Bradbury R.H., Cooke R., Erlandson J., Estes J.A., Hughes T.P., Kidwell S., Lange C.B., Lenihan H.S., Pandolfi J.M., Peterson C.H., Steneck R.S., Tegner M.J., Warner R.R., 2001. Historical overfishing and the recent collapse of coastal ecosystems. Science 293, 629-37.
- Kifani, S., 1995. Approche spatio-temporelle des relations hydroclimat-dynamique des espèces pélagiques en région d'upwelling: cas de la sardine du stock central marocain. Université de Bretagne Occidentale, France
- Knutsen, H., Jorde P.E., André C., Stenseth N.C., 2003. Fine-scaled geographical population structuring in a highly mobile marine species: the Atlantic cod. Mol. Ecol. 12, 385-394.
- Lanave, C., Preparata G., Saccone C., Serio G., 1984. A new method for calculating evolutionary substitution rates. J. Mol. Evol. 20, 86-93.
- Lee, W.-J., Conroy W., Howell H., Kocher T.D., 1995. Structure and evolution of teleost mitochondrial control regions. J. Mol. Evol. 41, 54-66.
- McMillan, W.O., Palumbi S.R., 1997. Rapid rate of control-region evolution in Pacific butterflyfishes (Chaetodontidae). J. Mol. Evol. 45, 473-484.
- Myers, R.A., Worm B., 2003. Rapid worldwide depletion of predatory fish communities. Nature 423, 280-283.
- 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., 1987. Molecular evolutionary genetics. Columbia University Press, New York.

- Ostellari, L., Bargelloni L., Penzo E., Patarnello P., Patarnello T., 1996. Optimization of single-strand conformation polymorphism and sequence analysis of the mitochondrial control region in *Pagellus bogaraveo* (Sparidae, Teleostei): rationalized tools in fish population biology. *Anim. Genet.* 27, 423-427.
- Parrish, R.H., Serra R., Grant W.S., 1989. The monotypic sardines, *Sardina* and *Sardinops*: their taxonomy, distribution, stock structure, and zoogeography. *Can. J. Fish. Aquat. Sci.* 41, 414-422.
- Pauly, D., Alder J., Bennett E., Christensen V., Tyedmers P., Watson R., 2003. The future of fisheries. *Science* 302, 1359-1361.
- Posada, D., Crandall K.A., 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14, 817-818.
- Ramon, M.M., Castro J.A., 1997. Genetic variation in natural stocks of *Sardina pilchardus* (sardines) from the western Mediterranean Sea. *Heredity* 78, 520-528.
- Rice, W.R., 1989. Analyzing tables of statistical tests. *Evolution* 43, 223-225.
- Rogers, A.R., Harpending H., 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* 9, 552-569.
- Rozas, J., Sánchez-DelBarrio J.C., Messeguer X., Rozas R., 2003. DnaSP, DNA polymorphism analysis by the coalescent and other methods. *Bioinformatics* 19, 2496-2497.
- Saitou, N., Nei M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406-425.
- Schneider, S., Excoffier L., 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics* 152, 1079-89.

- Schneider, S., Roessli D., Excoffier L., 2000. Arlequin ver 2.000: a software for population genetics data analysis. Genetics and Biometry Laboratory. University of Geneva, Geneva
- Silva, A., 2003. Morphometric variation among sardine (*Sardina pilchardus*) populations from the northeastern Atlantic and the western Mediterranean. ICES J. Mar. Sci. 60, 1352-1360.
- Slatkin, M., Hudson R.R., 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129, 555-62.
- Spanakis, E., Tsimenides N., Zouros E., 1989. Genetic differences between populations of sardine, *Sardina pilchardus*, and the anchovy, *Engraulis encrasicolus*, in the Aegean and Ionian Seas. *J. Fish Biol.* 35, 417-437.
- Swofford, D.L., 1998. PAUP\*: Phylogenetic Analysis Using Parsimony (\*and other methods), version 4.0. Sinauer Associates, Inc., Sunderland, MA, US.
- Tajima, F., 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105, 437-460.
- Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585-595.
- Tamura, K., Nei M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10, 512-526.
- Thompson, J.D., Gibson T.J., Plewniak F., Jeanmougin J., Higgins D.G., 1997. The Clustal X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876-4882.
- Tinti, F., di Nunno C., Guarniero I., Talenti M., Tommasini S., Fabbri E., Piccinetti C., 2002. Mitochondrial DNA sequence variation suggests the lack of genetic

heterogeneity in the Adriatic and Ionian stocks of *Sardina pilchardus*. *Mar. Biotechnol.* 4, 163-172.

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

Utter, F.M., 1991. Biochemical genetics and fishery management: an historical perspective. *J. Fish Biol.* 39, 1-20.

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

Zardoya, R., Castilho R., Grande C., Favre-Krey L., Caetano S., Marcato S., Krey G., Patarnello T., 2004. Differential population structuring of two closely related fish species, the mackerel (*Scomber scombrus*) and the chub mackerel (*Scomber japonicus*), in the Mediterranean Sea. *Mol. Ecol.* 13, 1785-98.

## Legends to Figures

Fig. 1. Map of sea areas around Morocco showing sampling localities of sardine. Inset: map of Western Europe showing additional sampling localities. Pie diagrams depict relative frequencies in each population of the mitochondrial control region sequences with (grey) and without (white) a 13 bp insertion.

Fig. 2. Main molecular features of the mitochondrial control region of the sardine.

Fig. 3. Neighbor-joining phylogram constructed using GTR+G distances. Populations are identified by different symbols. The presence (white box) or absence (grey box) of a 13 bp insertion in the mitochondrial control region sequences is map onto the phylogeny.

Fig. 4. Mismatch distribution for all pairwise combinations of the 261 analyzed sardine individuals. The observed distribution (bars) and those expected under the growth decline (solid lines) and constant (dashed lines) population models are represented.

Fig. 5. Mismatch distributions for each of the analyzed sardine populations. Dashed line corresponds to Safi. Grey lines represent other populations. Black solid line represents a combined data set with all populations except Safi.

Fig. 6. A Bayesian skyline plot ( $m=12$ ) estimated from an alignment of Safi sardine sequences. The solid line indicates the median estimate whereas the grey lines indicate the 95% credibility intervals. The initial mutation rate was set to 1, and years in the X axis were estimated using a mutation rate of  $1.1 \times 10^{-7}$  (McMillan and Palumbi, 1997)

**Table 1** Descriptive statistics of studied sardine populations\*

Population	Location	Date	<i>n</i>	No. Haplotypes	<i>S</i>	$H_{exp}$	<i>k</i>	$\pi$	$\theta$
Dakhla	23°N; 16°W	May 02	32	30	68	0.996±0.009	7.1±3.4	0.018±0.010	16.9±5.4
Tantan	28°N; 11°W	July 02	27	27	69	1.000±0.010	7.0±3.4	0.018±0.010	17.9±6.0
Safi	32°N; 9°W	May 02	29	28	70	0.997±0.010	8.9±4.2	0.023±0.012	17.8±5.8
Larache	35°N; 6°W	May 02	30	30	71	1.000±0.010	7.2±3.4	0.018±0.010	17.9±5.8
Nador	35°N; 3°W	May 02	30	29	62	0.998±0.009	7.4±3.6	0.019±0.010	15.6±5.1
Quarteira	37°N, 8°W	Sep 04	26	25	65	0.997±0.012	7.2±3.5	0.019±0.010	17.0±5.7
Barcelona	41°N; 2°E	Dec 02	27	27	56	1.000±0.010	6.6±3.2	0.017±0.009	14.5±4.9
Kavala	41°N; 24°E	Dec 02	30	29	63	0.998±0.009	6.9±3.3	0.018±0.009	16.5±5.2
Pasajes	43°N; 2°W	Dec 02	30	27	59	0.991±0.012	5.5±2.7	0.014±0.008	14.9±4.9

\* *n* = sample size, *S* = number of polymorphic sites,  $H_{exp}$  =gene diversity (Nei 1987), *k* = mean pairwise nucleotide differences (Tajima 1993),

$\pi$ = nucleotide diversity (Nei 1987),  $\theta$  =expected heterozygosity per site (Watterson 1975)

**Table 2** Pairwise  $\Phi_{ST}$  (below diagonal) and associated  $P$  values (above diagonal) among sardine populations

	Dakhla	Tantan	Safi	Larache	Nador	Quarteira	Barcelona	Kavala	Pasajes
Dakhla	—	0.715	0.000	0.706	0.122	0.708	0.779	0.342	0.003
Tantan	-0.004	—	0.000	0.835	0.243	0.905	0.867	0.331	0.001
Safi	0.080*	0.077*	—	0.000	0.002	0.000	0.000	0.000	0.000
Larache	-0.004	-0.006	0.076*	—	0.062	0.626	0.370	0.477	0.039
Nador	0.009	0.004	0.047	0.013	—	0.268	0.133	0.146	0.000
Quarteira	-0.004	-0.008	0.073*	-0.003	0.003	—	0.690	0.403	0.011
Barcelona	-0.006	-0.007	0.080*	0.001	0.009	-0.004	—	0.168	0.000
Kavala	0.001	0.001	0.080*	-0.001	0.008	0.000	0.007	—	0.000
Pasajes	0.026	0.031*	0.105*	0.013	0.047*	0.020	0.045*	0.040*	—

\* significant a  $p < 0.001$  (after Bonferroni correction)

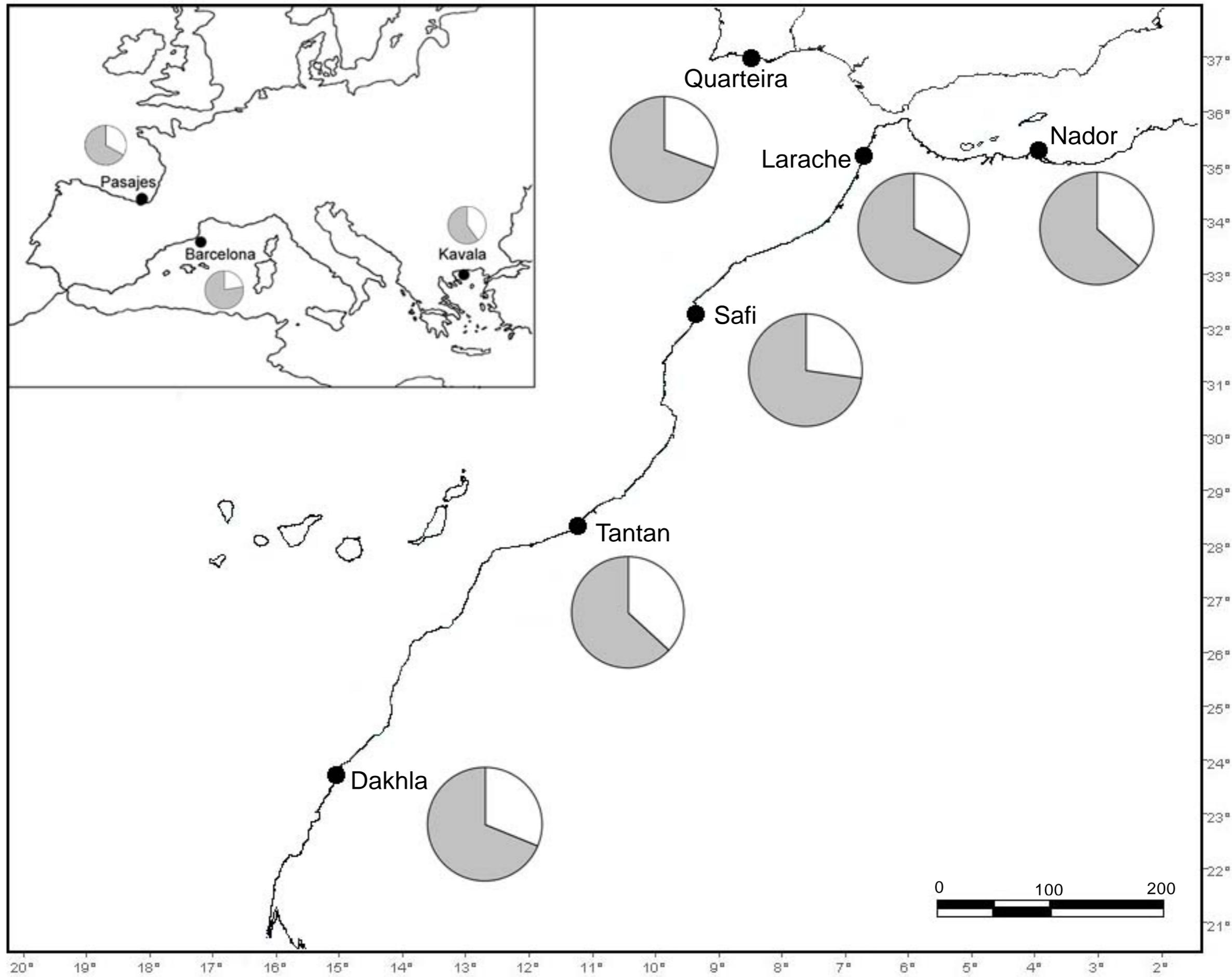
**Table 3** Genetic structuring of sardine populations

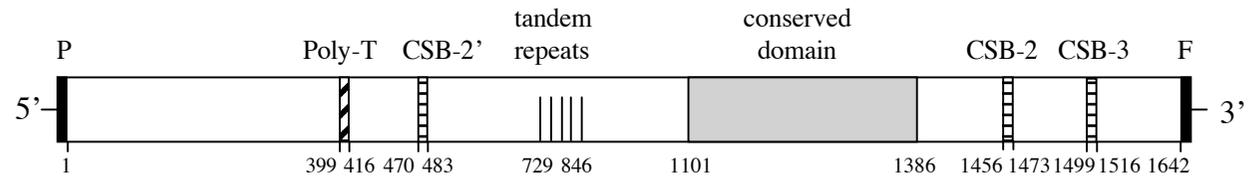
Structure tested	Observed partition		$\Phi$ Statistics	P
	Variance	% total		
1. One gene pool (Dakhla, Tantan, Safi, Larache, Nador, Quarteira, Barcelona, Kavala, Pasajes)				
Among populations	0.095	2.61	$\Phi_{ST} = 0.026$	0.000
Within populations	3.539	97.39		
2. Three gene pools (Dakhla, Tantan, Larache, Nador, Quarteira, Barcelona, Kavala)( Safi) (Pasajes)				
Among groups	0.220	5.85	$\Phi_{CT} = 0.058$	0.028
Within groups	0.002	0.05	$\Phi_{SC} = 0.001$	0.000
Within populations	3.539	94.10	$\Phi_{ST} = 0.059$	0.000

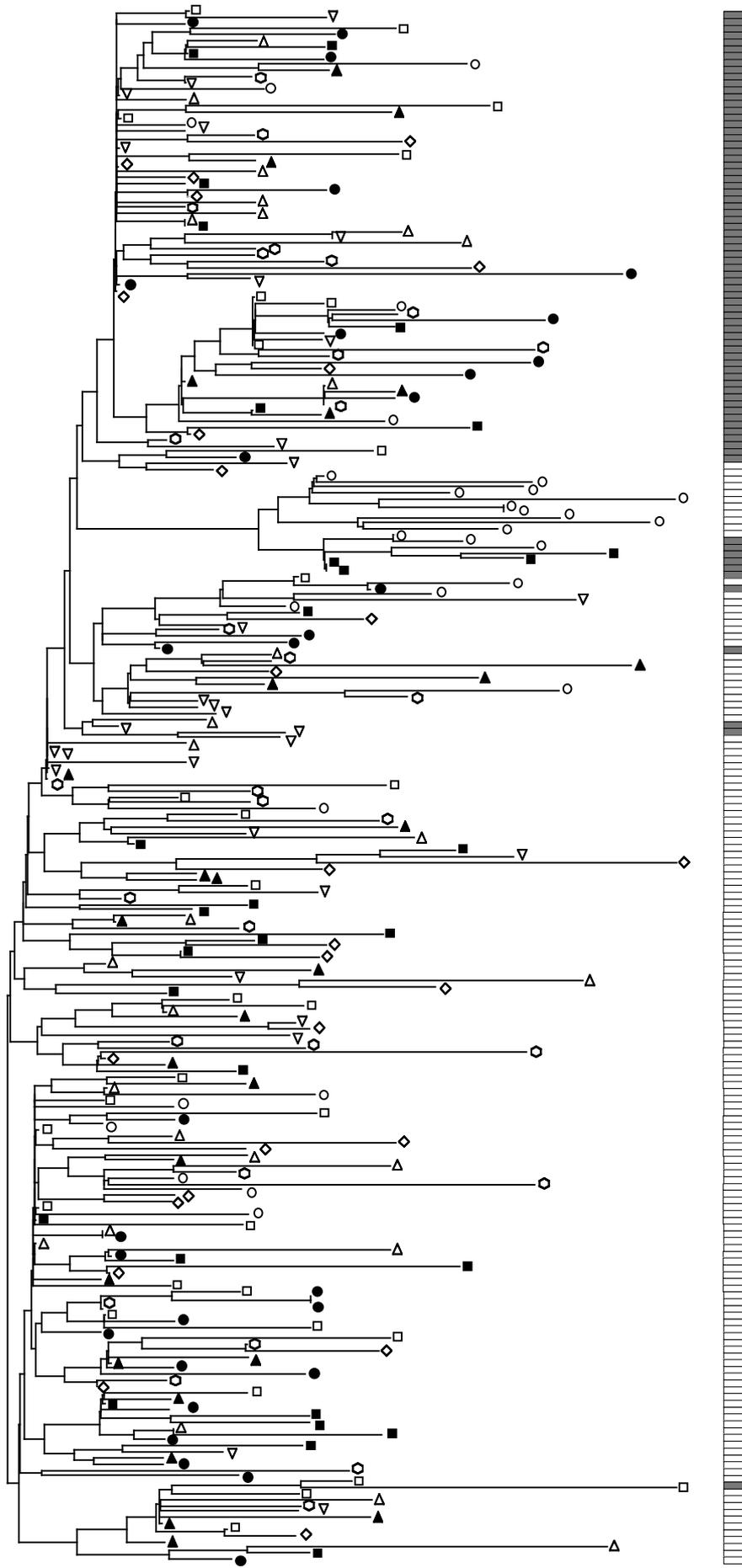
**Table 4** Statistical tests of neutrality, and demographic parameters estimates for sardine populations

	All samples
<b>Goodness of fit tests</b>	
Tajima $D$	-2.63*
Fu $F_s$	-594.40*
<b>Demographic parameters</b>	
$H_{ri}$	0.0075
$SSD$	0.00007
$S$	230
$\theta_0$	0.117
$\theta_1$	391.25
$\tau$	6.852

\* significant at  $p < 0.001$







□ Dakhla    △ Tantan    ○ Safi    ◻ Larache    ◇ Quarteira    ▽ Pasajes    ■ Nador    ▲ Barcelona    ● Kavala

