Fisheries Research 177 (2016) 31–38

Contents lists available at ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres

High resolution SNPs selection in *Engraulis encrasicolus* through Taqman OpenArray

Gaetano Catanese[a,*], Iratxe Montes[b], Mikel Iriondo[b], Andone Estonba[b], Daniele Iudicone[a], Gabriele Procaccini[a]

[a] Department of Integrative Marine Ecology, Stazione Zoologica Anton Dohrn (SZN), Villa Comunale, 80121 Napoli, Italy
[b] Department of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country (UPV/EHV), 48080 Bilbao, Spain

ARTICLE INFO

Article history:
Received 28 July 2015
Received in revised form 13 January 2016
Accepted 15 January 2016

Keywords:
Anchovy
Mediterranean
Fst
SNPs
Engraulis encrasicolus

ABSTRACT

The European anchovy (*Engraulis encrasicolus*) is one of the most important species in fisheries, representing the majority of landings worldwide. Identification of genetic stocks and assessment of divergence among them, is critical in order to implement management strategies. Population genetic structure in the Mediterranean basin is not clear and has not been extensively investigated with highly informative markers for population analyses. In this work, we aimed to identify a small SNP panel to be utilized for fine scale population genetic analysis within the Mediterranean basin. In order to do that, we used a set of 424 species-specific SNPs for assessing differentiation among *E. encrasicolus* populations within the Mediterranean and between Atlantic and Mediterranean Sea. Hence, we applied a Fst ranking method, for selecting a SNP sub-set from the large 424 SNPs panel and we compared the results obtained with the two sets of markers. Population assignment power and patterns of population differentiation were comparable. Analyses revealed a clear distinction between anchovies in the Atlantic and Mediterranean areas, and lower differentiation among Mediterranean populations. Our approach was successful in selecting a 96 SNP subset with high resolution and cost effectiveness to genotyping that can represent a useful tool for population genetic studies and stock management in this economically important species.

1. Introduction

The European anchovy (*Engraulis encrasicolus*; Family: Engraulidae), one of the most important species in fisheries worldwide, is distributed in the whole Mediterranean Sea, including the Black and the Azov Seas, and also along the eastern Atlantic coast, from Norway to South Africa (Whitehead et al., 1988).

Together with sardines, anchovies represent the majority of landings in the world and in particular in the Mediterranean Countries (Whitehead et al., 1988). Mediterranean and Black Sea give 5% of the world's anchovy catch, which is equivalent to 563,000 t. The largest supplying countries of this area are Turkey, Italy, Georgia and Greece. In recent years, the high commercial demand has caused an excessive exploitation of the resource, forcing authorities to adopt measures for the development of management plans and for monitoring the stock status (UE Council Regulation No 1967/2006).

In order to implement management strategies, it is critical to identify spawning areas, to unravel population genetic structure and to assess genetic divergence among stocks. The distribution of spawning habitats along the coasts of the Mediterranean is only known for areas where detailed acoustic surveys have been conducted (Giannoulaki et al., 2013). Adults seem to have a limited dispersal from restricted areas of high productivity allowing that particular regions could harbor distinct populations or stocks. Previous studies using molecular markers, showed a complex population structure for European anchovies in the Atlantic Ocean, and in distinct areas of the Mediterranean Sea (Spanakis et al., 1989; Keskin and Atar, 2012; Bembo et al., 1996a,b; Kristoffersen and Magoulas, 2008; Grant, 2005; Borsa, 2002; Borsa et al., 2004). Mitochondrial DNA markers identified two major mitochondrial lineages within the Mediterranean and Black Seas: one was predominant in the Black Sea and the Aegean Sea, and the other much more frequent in the western basin (Magoulas and Zouros 1993; Magoulas et al., 1996). A recent secondary contact between previously isolated Atlantic and Mediterranean populations has been suggested (Magoulas et al., 2006). Nuclear markers such as allozyme loci (Borsa et al., 2002; Erdogan et al., 2009), introns

* Corresponding author.
E-mail address: gaetano.catanese@szn.it (G. Catanese).

http://dx.doi.org/10.1016/j.fishres.2016.01.014
0165-7836/

at loci CK6-1 and CK6-2 of the creatine-kinase multigenic family (Borsa et al., 2004; Bouchenak-Khelladi et al., 2008) and microsatellites (Zarraonaindia et al., 2009; Borrell et al., 2012; Oueslati et al., 2014) have also been used to assess genetic structure in Atlantic and Mediterranean Sea anchovy populations, detecting differences mainly due to allele frequency distribution.

Recent studies also utilized Single Nucleotide Polymorphisms (SNPs), which nowadays are among the most commonly used genetic markers in population genetics and molecular ecology. A first SNP panel, composed by 47 nuclear and 15 mitochondrial SNP markers (Molecular Ecology Resources Primer Development Consortium et al., 2012), was utilized for studying Mediterranean and Atlantic samples of *E. encrasicolus* (Zarraonaindia et al., 2012). The SNP panel utilized, detected low genetic distinction among most of the Mediterranean samples included in the analysis, with the exception of the sample from the Alboran Sea that grouped with Atlantic populations from the Gulf of Cadiz (Zarraonaindia et al., 2012). In a more recent study, a large number of SNPs (about 19,000) was discovered through a Next Generation Sequencing approach from transcriptomic and genomic libraries of *E. encrasicolus*. Authors selected and tested a 530 SNP panel in four Atlantic and one Mediterranean population, using a Taqman OpenArray technique. A total of about 420 independent markers were found, after quality filtering (polymorphism, LD and HWE), to be potentially useful for future genetic studies (Montes et al., 2013).

The reduction of statistical redundancy and the selection of the more informative loci from hundreds of SNPs is a strategic approach to obtain reduced SNP panels for studies of genetic fish population and paternity association, with high resolution power but lower costs (Hess et al., 2014; Larson et al., 2014). In this contest, new platforms based on 96.96 SNP Type Assay, provide flexible, cost effective and ease of technique to genotype simultaneously multiple samples. Various selection methods can be utilized, which result in significant differences in SNP composition among panels, with different power of population discrimination and assignment (e.g., Ozerov et al., 2013). One of the methods is based on the preliminary screening of putatively distinct populations using a larger set of candidate SNP markers and on the selection of informative loci for population differentiation, based on their Fst value. This method has been utilized in recent works, and has been considered efficient and reliable in respect to other approaches (Hohenlohe et al., 2011; Storer et al., 2012; Ozerov et al., 2013; Larson et al., 2014). How big a SNP panel should be to give consistent results is debated in the literature. Panels of 50–100 SNP markers are reported to allow accurate pedigree reconstruction (Anderson and Garza, 2006) and to track family lineages over two generations with reasonable power. Morin et al. (2009) state that ~30 SNPs should be adequate to detect moderate levels of differentiation, but studies aimed at detecting distinct demographic units may require 80 or more SNPs. A following study reports that SNP panels of 48 or 96 loci perform differently in assignment probability, with larger panels and panels created from Fst ranks performing better than the others (Storer et al., 2012). Obviously, SNP selection using Fst ranks is affected by the population panel utilized (Ozerov et al., 2013). Selected SNPs "will be biased toward the most differentiated populations and will not allow for assignment to more finely differentiated groups" (Helyar et al., 2011), if highly distinct populations are included in the selection and selected panels are utilized for less distinct populations. In contrast, population assignment accuracy of the SNP panel will be lower for highly differentiated populations, when ranking of candidate loci is obtained screening from poorly differentiated populations (Anderson, 2010).

The goal of the present study was to select a minimum set of 96 information-rich SNPs that meet a desired threshold for fine-scale prospective studies of anchovy in the Mediterranean basin, reducing costs and time of analysis but not affecting markers resolution in respect to the larger available panel. In order to do that, a Fst ranking method was applied to a larger SNP panel, previously selected from transcriptomic and genomic sequences. The genetic patterns of variability of *E. encrasicolus* populations within the Mediterranean Sea and between Mediterranean and a neighboring Atlantic population, have been compared between for the two panels to assess for possible problems related to ascertainment bias and selection methods utilized.

2. Material and methods

2.1. Sampling and SNPs genotyping

One hundred and eleven individuals, belonging to five populations, were analysed. Mediterranean samples were collected from distinct Geographical Sub-Areas (GSA): GSA 17 (Chioggia—CHI; Adriatic Sea, Italy) GSA 19 (Cirò Marina—CIR; Ionic Sea, Italy), GSA 10 (Cetara—CET; Tyrrhenian Sea, Italy), GSA 6 (Tarragona—TAR; Balearic Sea, Spain). Atlantic samples, collected in the FAO Fishing area 27, ICES sub-area IXa (Gulf of Cadiz—CAD; Atlantic Ocean, Spain), were also included in the analysis (Fig. 1; Table 1).

A portion of muscle tissue (1 g) from each individual was preserved in 95% ethanol for further DNA extraction. Total genomic DNA was isolated from 30 mg of tissue using Nucleospin Tissue kit (Macherey-Nagel, Düren, Germany). All DNA isolation procedures were performed following the manufacturer's protocol. Quality and concentration of the extracted DNA was checked using a NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, U.S.) and Qubit 2.0 fluorometer (Invitrogen/Thermo Fisher Scientific, Wilmington, U.S).

All individuals were screened for 424 SNPs, using TaqMan OpenArray SNP platform (Life Technologies), from the SNP panel isolated and validated by Montes et al. (2013). Genomic DNA (66 ng per sample) was used as template at the required DNA starting concentration (22 ng/μl). Subsequent reactions for the amplification and detection of the SNPs were carried out following TaqMan® OpenArray™ Genotyping System User Guide (Applied Biosystems, Carlsbad, U.S.) at the Sequencing and Genotyping Service (SGIker) of the University of the Basque Country (UPV/EHU).

2.2. Selection of reduced SNPs set

Genotyping data were visualized and examined with the TAQMAN GENOTYPER software v2.1 (Life Technologies). After default clustering was performed, data were visualized in the scatter plot and genotype calls were refined manually for producing the final cluster assignments. For each sample, the proportion of loci successfully genotyped using the 424 SNP set was calculated. In order to increase the power to differentiate populations in Central Mediterranean Sea, 96 SNPs were selected. We applied an initial quality-control based filtering-tool and we have excluded SNPs whose minor allele frequency (MAF) was <0.01. Additionally, we excluded those SNPs whose Hardy–Weinberg *P*-value was <0.001. Then, SNPs were ranked using pairwise Fst values among the three Mediterranean populations sampled along the Italian coasts and those SNPs showing the highest Fst values were retained.

Finally, we constructed the panel of the "best" 96 markers for prospective stock identification and stock assignment in Central Mediterranean Sea, transferring TaqMan® OpenArray™ genotyping technology to the more economic technology called Fluidigm (Fluidigm, South San Francisco, California). Thus, the transfer obliged us to exclude loci whose flanking regions did not allow for designing primer pairs with appropriate thermodynamic characteristics for amplification by Fluidigm 96.96 dynamic array.

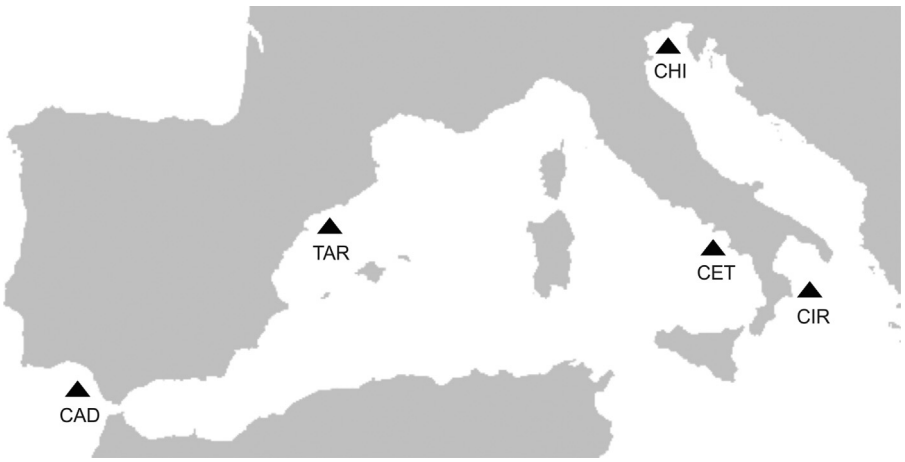

Fig. 1. Map of sampling locations, listed in Table 1.

Table 1
Sampling information and summary statistics. For each of the sampling locations the following values are given: number of individuals analyzed (*N*), location coordinates, and year of sampling, number of different alleles, (Na), and percentage of polymorphic SNPs. Values are reported for both 424 SNP and 96 SNP panels.

Sampling location	ID	*N*	Coordinates Lat; Long	Year	424 SNPs		96 SNPs	
					Na	% polym	Na	% polym
Mediterranean								
Cetara	CET	20	40°37′N; 14°43′E	May 2013	1.840	84.20	1.916	91.67
Chioggia	CHI	20	45°08′N; 12°25′E	April 2013	1.847	84.91	1.958	95.83
Cirò Marina	CIR	20	39°24′N; 17°11′E	April 2013	1.830	83.25	1.906	90.62
Tarragona	TAR	25	40°53′N; 01°10′E	March 2009	1.884	88.44	1.968	96.87
Atlantic								
Cadiz	CAD	26	36°32′N; 06°28′W	April 2009	1.906	90.57	1.958	95.83

2.3. Data analysis

The following analyses were performed on both 424 and 96 SNP panels. Results were visually compared and the Pearson correlation between population pairwise Fst values obtained with the two different panels was calculated using the software STATISTICA ver. 7.0 (StatSoft). The expected heterozygosity (He), observed heterozygosity (Ho), and MAF were evaluated by GENALEX 6.5 (Peakall and Smouse, 2012). The program GENEPOP 4.2 (Raymond and Rousset, 1995; Rousset, 2008) was used to estimate the deviations from Hardy-Weinberg equilibrium (HWE) and the exact test (10,000 dememorizations, 100 batches and 5000 iterations per batch) for statistically significant linkage disequilibrium (LD). The sequential Bonferroni method (Rice, 1989) for multiple comparisons was applied to correct the significance level.

BAYESCAN 2.1 (Foll and Gaggiotti, 2008) was used to detect outlier markers. The estimation was performed with 10 pilot runs, of 5000 iterations and an additional burn-in of 50,000 iterations (sample size of 5000 and thinning interval of 10).

Nei's genetic distance matrices (Nei, 1972) were calculated among samples using POPULATIONS v1.2.31 (Langella, 2000) and GENALEX 6.5. The significance of correlation between genetic and geographic distances, as well as between Fst values among population of the two panels, were tested with a Mantel test using GENALEX 6.5 and R v3.2.2 software (R Development Core Team, 2012). To enable calculations of Isolation by distance (IBD), geographical distances in km were determined by use of Google Earth. The calculations were performed by plotting pairwise corrected genetic distance and Fst against the natural logarithm of the geographical distance in km.

Fst and AMOVA values (10,000 permutations to test for significance) were assessed for genetic differentiation between all pairs of samples by the program ARLEQUIN v.3.5 (Excoffier and Lischer, 2010). A false discovery rate (FDR) control (Benjamini and Hochberg, 1995) was applied to detect type I error and to give the adjusted *P*-values in the Fst multiple comparisons. Bayesian clustering analysis, implemented in the software STRUCTURE version 2.3.4 (Pritchard et al., 2000), was utilized to infer main genetic clusters (*k*). For each value of *k*, 5 iterations were run using the admixture model, with a burn-in period of 50,000, followed by 100,000 Markov Chain Monte Carlo iterations for values of $k=1$ through $k=5$. We used the 'admixture' and 'correlated-allelic-frequencies' models with a location prior. STRUCTURE HARVESTER (Earl and vonHoldt, 2012), which uses the methods proposed by Evanno et al. (2005), was employed for estimating the putative number of distinct clusters. Individuals were assigned to a reference population using GENECLASS v2.0 software (Piry et al., 2004), using the Bayesian approach described by Rannala and Mountain (1997).

3. Results

The genetic analysis of Tyrrhenian (CET), Adriatic (CHI), Ionian (CIR), Tarragona (TAR) and Cádiz (CAD) samples based on the whole panel of 424 SNPs, showed MAF values ranging from 0.004 to 0.496. Only 21 loci were monomorphic while 10 loci showed MAF values lower than 1%.

With regard to the percentage of polymorphic loci and number of alleles over all loci (N_A), CIR and CAD showed the lowest and highest values, respectively (Table 1). The lowest expected (He) and observed (Ho) heterozygosities were found in CET, whereas the highest values were found in CAD (Table 2 and Supplementary material, S.1).

Excluding monomorphic loci within populations, significant deviations from HWE were detected, after Bonferroni corrections, in 15 loci for CET, 17 loci for CHI, 13 loci for CIR, 15 loci for TAR

Table 2
Expected (He) and observed (Ho) heterozygosity; number of loci with HWE deviations; and number of loci with private alleles, for the five populations analysed. Values are given for both SNPs sets.

ID	424 SNPs Set				96 SNPs Set			
	He	Ho	Loci with HWE deviations	Private alleles	He	Ho	Loci with HWE deviations	Private alleles
CET	0.251	0.241	7	4	0.270	0.260	8	1
CHI	0.262	0.254	10	1	0.312	0.310	4	–
CIR	0.254	0.246	7	1	0.271	0.270	9	–
TAR	0.265	0.262	10	1	0.310	0.305	3	–
CAD	0.313	0.290	8	12	0.349	0.316	4	–

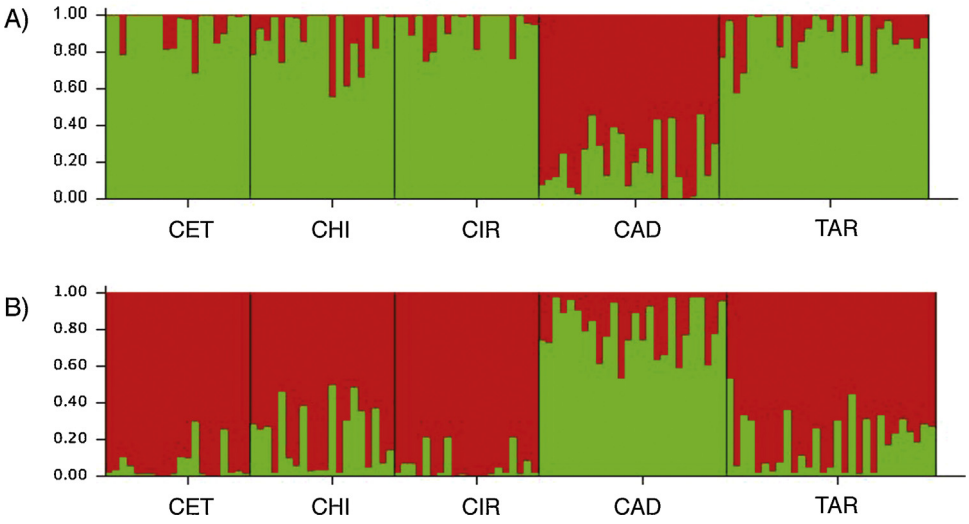

Fig. 2. STRUCTURE analysis with 424 SNP (A) and 96 SNP (B) panels. $K = 2$ for both panels.

and 31 loci for CAD (Table 2 and Supplementary Material, S.1). Private alleles (Pa) were identified in CET (Pa = 4), CHI (Pa = 1), CIR (Pa = 1), TAR (Pa = 1) and CAD (Pa = 12) (Table 2 and Supplementary material, S.1). Linkage disequilibrium (LD) was assessed for each pair of loci meeting HWE and 11 SNPs were in LD within five haplotypes. Finally, Bayescan software detected only one outlier locus, (14931_283), a candidate locus under the effect of natural selection. This SNP mainly differentiates CAD sample (frequency$_{14931_283*A}$ = 0.820) from the Mediterranean populations (frequency$_{14931_283*A}$ = 0.000-0.212).

3.1. Genetic differentiation among populations using the large set

The AMOVA analyses carried out applying the 424 SNP markers revealed significant values of variance (10.62%) among populations (Fst = 0.106; $P = 0.0000$). Only a 3.05% of variance (Fis = 0.034; $P = 0.0263$) was detected among individuals within populations, while it was 86.33% within individuals (Fit = 0.136; $P = 0.0000$).

Overall estimates of pairwise population genetic differentiation (Fst) ranged between −0.008 and 0.211. Atlantic and Mediterranean pairwise comparisons were always highly significant ($P < 0.001$), ranging from 0.184 to 0.211 (Table 3). Lower, but also significant estimates ($P < 0.05$), were revealed in most of the pairwise comparisons within the Mediterranean Sea. The only exception is represented by CET, that is not significantly distinct from CIR and TAR. No changes in significant P values were detected after FDR corrections (Table 3). Results of Mantel tests were qualitatively the same using pairwise Fst or Nei's genetic distances. The correlation between genetic and geographic distances among the Mediterranean populations TAR, CET and CIR was significant ($r = 0.9976$; $P = 0.031$), but the hypothesis of correlation between these two variables was rejected ($P > 0.05$) when tested for all populations.

STRUCTURE analysis of the 111 anchovy individuals clearly identified two population groups ($K = 2$; DeltaK = 6047.04; Supplementary material, S.2). The first group was formed by samples from Mediterranean Sea and the second by samples from Atlantic Ocean (Fig. 2a).

3.2. Selection of highly informative SNPs and comparison with large panel

The 96 SNP subset, with the locus-specific Fst values utilized for selection, is available as Supplementary Material (S.3). Using the 96 SNP panel, MAF values varied from 0.017 to 0.983, with only 2 loci (05837_299 and 16005_81) having values lower than 2%. Monomorphic loci were observed in all populations: 8 in CET; 4 in CHI; 9 in CIR; 3 in TAR and 4 in CAD. Based on the 96 SNP panel, CIR and TAR samples showed the lowest and highest percentage of polymorphic loci, respectively (90.62% and 96.87%). Value of number of alleles ranged from $N_A = 1.906$ in CIR, to $N_A = 1.968$, in TAR. Heterozigosity values were lowest in CET (He = 0.270, Ho = 0.260) and highest in CAD (He = 0.349, Ho = 0.316). Loci with significant deviations from HWE were detected for each population, while private alleles were detected in only one locus in CET (Table 2 and Supplementary material, S.1). No outliers were detected using the 96 loci panel.

AMOVA analysis showed that most of the genetic variation (12.16%) is attributable to differences among populations (Fst = 0.121; $P = 0.0000$). High variance was also present within individuals (84.58%; Fit = 0.154; $P = 0.0000$) whereas only 3.26% of the variation was present among individuals within the same population (Fis = 0.037; $P = 0.0303$).

Table 3
Pairwise FST values between populations of European anchovies, using the 424 SNPs set (above) and the 96 SNPs set (below).

	CET	CHI	CIR	TAR	CAD
CET		**0.010** (0.0126/0.018)	−0.00026 (0.681/0.681)	0.005 (0.0667/0.074)	**0.211** (0.0000/0.0000)
CHI	**0.038** (0.0000/0.0000)		**0.011** (0.0047/0.0094)	**0.008** (0.0152/0.019)	**0.184** (0.0000/0.0000)
CIR	**0.023** (0.0005/(0.0007)	**0.044** (0.0001/(0.0001)		**0.008** (0.0101/0.0168)	**0.209** (0.0000/0.0000)
TAR	**0.015** (0.0084/0.0084)	**0.011** (0.0380/0.0380)	**0.021** (0.0005/0.0005)		**0.191** (0.0000/0.0000)
CAD	**0.242** (0.0000/0.0000)	**0.178** (0.0000/0.0000)	**0.261** (0.0000/0.0000)	**0.178** (0.0000/0.0000)	

In bold: significant values of FST.
In parentheses *P* values and *FDR* corrected *P* values.

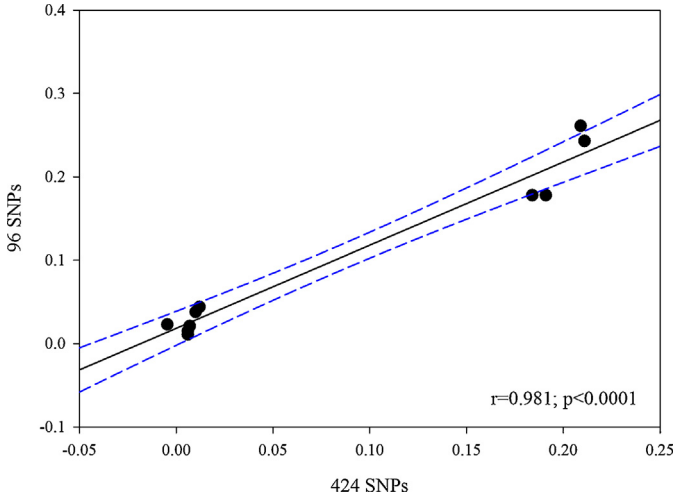

Fig. 3. Correlation of Fst values among population with 424 SNP and 96 SNP panels. Blue lines indicate the 95% confidence interval. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Pairwise Fst values are all significant (Table 3). The Atlantic population of Cadiz was highly differentiated from the Mediterranean populations, with pairwise Fst values ranging from 0.178 to 0.261 ($P < 0.001$). The lower Fst values were present between TAR and the other three Mediterranean populations. Fst values among Mediterranean populations were always higher than values recorded with the 424 SNP panel, as expected from the panel selection method (Table 3). The pairwise Fst values obtained with the two sets of markers showed a highly significant correlation ($r = 0.981$; $P < 0.0001$) (Fig. 3).

The Bayesian approach implemented in STRUCTURE (for 96 SNP set $K = 2$; DeltaK = 488.76; see Supplementary material, S.2) also showed similar results between the two datasets (Fig. 2b). Analysis were also carried out excluding the Atlantic populations, for both SNP panels. Two groups were also identified, with not clear geographically coherent population structure among the Mediterranean samples (data not shown).

Finally, two different threshold values were selected to compare assignment accuracy between the two panels. The two panels performed in a comparable manner, when a 90% assignment accuracy is applied. The specimens from Mediterranean populations presented correct assignment values ranging from 35 to 48% for the complete dataset and from 41 to 75% for the 96 SNP subset. A slightly higher percent of individuals are correctly assigned in CET and CIR, with the 96 SNP panel, while the proportion of correctly assigned individuals is higher for the large panel in CHI and TAR. When using a more relaxed threshold value (70%), the small panel increases the proportion of correctly assigned individuals and outperforms the large panel in CET, CHI and CIR. Both panels assigned the 100% of individuals with 100% of assignment accuracy in the Atlantic population of CAD (Fig. 4).

4. Discussion

The aim of this study was to select a highly discriminant set of SNP markers for accurate genetic studies of stocks of the European anchovy *E. encrasicolus*, reducing costs and time of analysis but not affecting markers resolution.

The importance of SNP selection in fishery science has been stated in many scientific projects and publications concerned with e.g., conservation applications (Hess et al., 2015), genetic stock identification (Ozerov et al., 2013; Larson et al., 2014) population assignment (Storer et al., 2012) and local adaptation (Limborg et al., 2012).

Two factors can affect results reliability in the process of SNPs selection. One is related with a possible ascertainment bias due to the material utilized in the selection of the original SNPs from which the small panels are derived. The second is related to the method utilized for selecting informative SNPs.

The possibility of ascertainment bias has been deeply analyzed by Albrechtsen et al. (2010) for human populations. The authors demonstrate that the level of bias strongly depends by the material utilized in the SNP selection and by the genetic measure utilized to infer differentiation among target populations. Our reduced panel originates from a larger SNP selection of 424 SNPs developed by Montes et al. (2013). Montes and collaborators utilized individuals collected in three distinct localities along an ample geographic gradient, potentially catching a high proportion of the variation existing in the Atlantic and Mediterranean distribution of the species. Two of the populations utilized in the initial SNP selection were also included in our analysis (i.e., Cadiz and Tarragona). The Spanish Mediterranean population resulted to be poorly differentiated from the Italian Mediterranean ones, suggesting that the variation existing among the Italian sites falls within the range of variation encompassed by the Montes et al. study, and reducing the possibility of ascertainment bias. The comparison of patterns of population differentiation between the two SNP panels also support the reliability of the small panel.

In their study, Albrechtsen et al. (2010) assess that the effects of a potential loss of MAF in the studied populations can also be reduced using measures of population subdivision that are not directly depended from allele frequency, such as Fst. This is one of the reasons why we decided to use the Fst ranking method for selecting the reduced SNP panel. It was demonstrated that SNP datasets selected by maximum diversity or Fst values are more informative and yield higher power than randomly selected SNPs (Hohenlohe et al., 2011; Storer et al., 2012; Ozerov et al., 2013; Larson et al., 2014). Small subsets of selected SNP can be used for a variety of purposes including: selection of breeding stock in species where individuals have a comparatively low value relative to the cost of high-density arrays, selection of replacement animals on commercial farms, parentage assignment, optimizing mate choice, and marker-assisted management (Saatchi and Garrick, 2014). The use of a panel of 96 SNPs provided very accurate identification of stocks from distinct populations of Chinook Salmon, showing its important applicability for harvest management and conservation of fisheries (Clemento et al., 2014; Larson et al., 2014). Reduced SNP

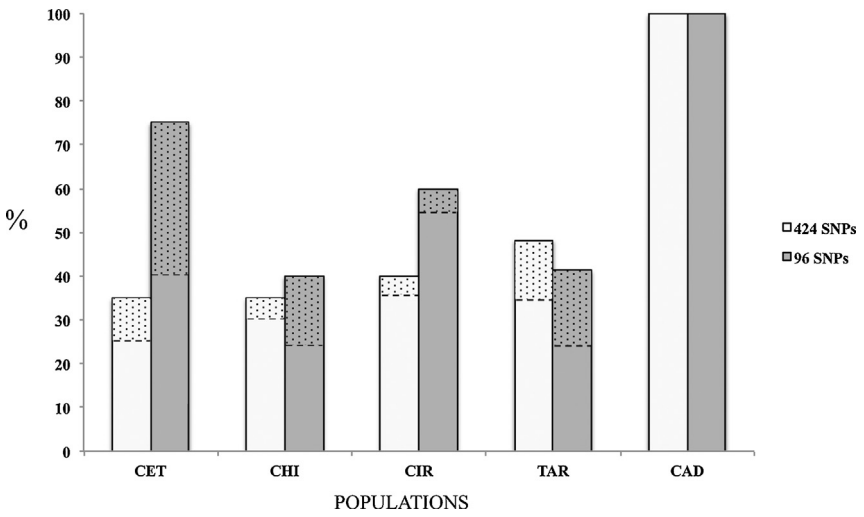

Fig. 4. Correct assignment of individuals in each population using 424 and 96 SNP panels. Proportion of individuals assigned with 90% probability are shown in solid colors; proportion of individuals assigned at 70–90% probability are shown in dotted colors.

panels selected on Fst and MAF performed successfully in parentage analysis, species identification and population structure analysis in the Pacific lamprey (Hess et al., 2014).

In our work, results show that SNPs selected ranking Fst values were efficient in capturing a large part of the genetic variation identified by a larger SNP panel. The two different SNP panels identified similar values of polymorphism, while the small set performs better in assigning individuals to their geographic origin, although at not stringent threshold probability values. Values of Ho detected for each population were very similar in relation to the two distinct panels of SNPs, with a small increase using the 96 SNP set.

Although obtained with a limited number of populations and individuals, general patterns of *E. encrasicolus* population structure obtained in our study were consistent with results from previous studies in the same region. Two genetic groups were mainly identified in our data set, highlighting the genetic separation existing between Atlantic and Mediterranean areas. Fst values between the two groups ranged from 0.184 to 0.211 for the 424 SNPs set. Fst values were higher with the 96 SNPs set and ranged from 0.178 to 0.261, two times larger than the values reported by Zarraonaindia et al. (2012) using 64 SNPs (Adriatic/North western Mediterranean vs. Southern Iberian-Atlantic). The Atlantic-Mediterranean disjunction has been also reported in numerous studies of *E. encrasicolus* (Magoulas et al., 2006; Borrell et al., 2012; Viñas et al., 2013; Zarraonaindia et al., 2009, 2012) as well as in many other species (Patarnello et al., 2007) and is consistent with isolation in the critical area corresponding to the point where surface waters link the Atlantic and Mediterranean forming an ecological barrier known as Almeria-Oran Front (Naciri et al., 1999; Bargelloni et al., 2005). Our SNPs set unequivocally distinguish Atlantic and Mediterranean samples, since all the individuals sampled in Cadiz are assigned with 100% of probability to the original population.

While a physical barrier separates predominantly Atlantic and Mediterranean populations, lower genetic differences are expected within the Mediterranean Sea. In this study, both the 424 and the 96 SNP panels showed little (but almost always significant) differentiation among the Mediterranean populations, including the Spanish population of Tarragona. Although apparently counterintuitive, low genetic distance values between Spanish and Adriatic coasts were previously reported in anchovies using a different set of SNP markers (Zarraonaindia et al., 2012). The genotyping of nuclear SNPs for population genetic structure of others pelagic species showed lower differentiation of populations within the Mediterranean Sea, but higher in relation with distant areas such as Atlantic Ocean, Indian Ocean and Pacific Ocean samples. In albacore (*Thunnus alalunga*), investigated by using 75 SNPs, four genetically homogeneous populations were detected for each basin, but the samples of Mediterranean Sea grouped all into a single cluster (Laconcha et al., 2015). Similarly, a study using 381 SNPs for European hake populations (*Merluccius merluccius*) showed that, using putatively neutral SNPs, in the Mediterranean the most significant comparisons involved only population samples from the Eastern basin. However, the Atlantic and the Mediterranean outlier loci revealed the presence of well-differentiated clusters within each basin, with Mediterranean samples being more heterogeneous. Clusters corresponding to the Western, Central and Eastern Mediterranean populations were attributed to two genetic breaks in the Mediterranean (the Siculo–Tunisian Strait and the south of the Peloponnese) (Milano et al., 2014).

Our study showed that in Mediterranean Sea all considered sample sites are significantly distinguishable each other by frequency distribution. The existence of two groups in Mediterranean Sea (eliminating CAD sample from the analysis) is coherent with previously cited studies. The lack of coherent geographic structure could be due to the sympatric co-occurrence of the two populations, subjected to different levels of mixing in the different stages of the anchovy life cycle.

Anchovies tend to spawn in the peak season (summer) predominantly in localized areas, mostly associated with point sources of nutrient enrichment that enhance productivity, such as river runoff or local upwelling. In the western/central Mediterranean, potentially suitable spawning areas are located in the Gulf of Lions and off the Catalan coast, the Alboran Sea and the Italian coasts of the Ligurian, Tyrrhenian, Adriatic Seas and Sicily Strait (Giannoulaki et al., 2013). In South and Central Tyrrhenian Sea as well as in Western Ionian Sea little data of spawning areas are available. Samples used in our work were collected in potentially spawning habitats and the range of body length of collected specimens (about 12–13 cm) suggests that they were all at maturity stage. As the season progresses, the population expands over wider areas in search of resources. This could promote long-distance dispersal from the spawning areas with consequent admixture or secondary contact of different anchovy stocks. Ecological selection has also been suggested between inshore (lagoon) and offshore habitats in western and central Mediterranean anchovy samples (Oueslati et al., 2014). This mechanism, over imposing possible patterns of dispersal and genetic drift, makes the comprehension of stock boundaries and dynamics in *E. encrasicolus* even more difficult to assess.

Our results open the field to further studies assessing stock boundaries and dispersal in the important fishery area of the Central Mediterranean basin, utilizing this reduced set of SNP. The present work highlights the usefulness of SNP reduced panels to examine genetic variation and population patterns at regional scale in small pelagic fishes, maintaining the efficacy and increasing the efficiency in relation to larger panels.

Acknowledgments

Authors thank the MIUR Italian Flagship project RITMARE for funding the research. The authors thank for technical support provided by sequencing and genotyping facilities from SGIker of UPV/EHU, especially to Dr. Irati Miguel and Dr. Fernando Rendo.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fishres.2016.01.014.

References

Albrechtsen, A., Nielsen, F.C., Nielsen, R., 2010. Ascertainment biases in SNP chips affect measures of population divergence. Mol. Biol. Evol. 27, 2534–2547.
Anderson, E.C., 2010. Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. Mol. Ecol. Resour. 10, 701–710.
Anderson, E.C., Garza, J.C., 2006. The power of single-nucleotide polymorphisms for large-scale parentage inference. Genetics 172, 2567–2582.
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. Phylogenet. Evol. 36, 523–535.
Bembo, D.G., Carvalho, G.R., Cingolani, N., Pitcher, T.J., 1996a. Stock discrimination among European anchovies, *Engraulis encrasicolus*, by means of PCR-amplified mitochondrial DNA analysis. Fish. Bull. 94, 31–40.
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.
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B (Methodol.) 57, 289–300.
Borrell, Y.J., Pinera, J.A., Sanchez, J.A., Blanco, G., 2012. Mitochondrial DNA and microsatellite genetic differentiation in the European anchovy *Engraulis encrasicolus* L. ICES J. Mar. Sci. 69, 1357–1371.
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 occurrence of two anchovy species in the Mediterranean. C. R. Biol. 327, 1113–1123.
Bouchenak-Khelladi, Y., Durand, J.D., Magoulas, A., Borsa, P., 2008. Geographic structure of European anchovy: a nuclear-DNA study. J. Sea Res. 59, 269–278.
Clemento, A.J., Crandall, E.D., Garza, J.C., Anderson, E.C., 2014. Evaluation of a single nucleotide polymorphism baseline for genetic stock identification of Chinook Salmon (*Oncorhynchus tshawytscha*) in the California current large marine ecosystem. Fish. Bull. 112, 112–130.
Earl, D.A., vonHoldt, B.M., 2012. STRUCTURE HARVESTER: a web site and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 2, 359–361.
Erdogan, Z., Turan, C., Koç, H.T., 2009. Morphologic and allozyme analyses of European anchovy (*Engraulis encrasicolus* L. 1758) in the Black, Marmara and Aegean Seas. Acta Adriat. 50, 77–90.
Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10, 564–567.
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.
Foll, M., Gaggiotti, O.E., 2008. A genome scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180, 977–993.
Giannoulaki, M., Iglesias, M., Tugores, M.P., Bonanno, A., Patti, B., De Felice, A., Leonori, I., Bigot, J.L., Ticina, V., Pyrounaki, M.M., Tsagarakis, K., Machias, A., Somarakis, S., Schismenou, E., Quinci, E., Basilone, G., Cutitta, A., Campanella, F., Miquel, J., Oñate, D., Roos, D., Valavanis, V., 2013. Characterizing the potential habitat of European anchovy *Engraulis encrasicolus* in the Mediterranean Sea, at different life stages. Fish. Oceanogr. 22, 69–89.
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.
Hess, J.E., Campbell, N.R., Docker, M.F., Baker, C., Jackson, A., Lampman, R., McIlraith, B., Moser, M.L., Statler, D.P., Young, W.P., Wildbill, A.J., Narum, S.R., 2015. Use of genotyping by sequencing data to develop a high-throughput and multifunctional SNP panel for conservation applications in Pacific lamprey. Mol. Ecol. Resour. 15 (1), 187–202.
Hohenlohe, P.A., Amish, S.J., Catchen, J.M., Allendorf, F.W., Luikart, G., 2011. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. Mol. Ecol. Resour. 11, 117–122.
Keskin, E., Atar, H.H., 2012. Genetic structuring of European anchovy (*Engraulis encrasicolus*) populations through mitochondrial DNA sequences. Mitochondr. DNA 23, 62–69.
Kristoffersen, J.B., Magoulas, A., 2008. Population structure of anchovy *Engraulis encrasicolus* L. in the Mediterranean Sea inferred from multiple methods. Fish. Res. 91, 187–193.
Laconcha, U., Iriondo, M., Arrizabalaga, H., Manzano, C., Markaide, P., Montes, I., Zarraonaindia, I., Velado, I., Bilbao, E., Goñi, N., Santiago, J., Domingo, A., Karakulak, S., Oray, I., Estonba, A., 2015. New nuclear SNP markers unravel the genetic structure and effective population size of Albacore Tuna (*Thunnus alalunga*). PLoS One 10 (6), e0128247, http://dx.doi.org/10.1371/journal.pone.0128247.
Langella, O., 2000. POPULATIONS 1.2.30: population genetic software, individuals or population distance, phylogenetic trees. http://bioinformatics.org/~tryphon/populations/.
Larson, W.A., Seeb, J.E., Pascal, C.E., Templin, W.D., Seeb, L.W., 2014. Single-nucleotide polymorphisms (SNPs) identified through genotyping-by-sequencing improve genetic stock identification of Chinook salmon (*Oncorhynchus tshawytscha*) from western Alaska. Can. J. Fish. Aquat. Sci. 71, 698–708.
Limborg, M.T., Helyar, S.J., De Bruyn, M., Taylor, M.I., Nielsen, E.E., Ogden, R., Carvalho, G.R., Consortium, F.P.T., Bekkevold, D., 2012. Enviromental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). Mol. Ecol. 21, 3686–3703.
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.
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.
Magoulas, A., Zouros, E., 1993. Restriction-site heteroplasmy in anchovy (*Engraulis encrasicolus*) indicates incidental biparental inheritance of mitochondrial DNA. Mol. Biol. Evol. 10, 319–325.
Milano, I., Babbucci, M., Cariani, A., Atanassova, M., Bekkevold, D., Carvalho, G.R., Espiñeira, M., Fiorentino, F., Garofalo, G., Geffen, A.J., Hansen, J.H., Helyar, S.J., Nielsen, E.E., Ogden, R., Patarnello, T., Stagioni, M., fishpoptrace Consortium Tinti, F., Bargelloni, L., 2014. Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*). Mol. Ecol. 23, 118–135.
Molecular Ecology Resources Primer Development Consortium, Zarraonaindia, I., Albaina, A., Iriondo, M., Manzano, C., Pardo, M.A., et al., 2012. Permanent genetic resources added to molecular ecology resources database 1 October 2011–30 November 2011. Mol. Ecol. Resour. 12, 374–376.
Montes, I., Conklin, D., Albaina, A., Creer, S., Carvalho, G.R., Santos, M., Estonba, A., 2013. SNP discovery in European anchovy (*Engraulis encrasicolus*, L) by high-throughput transcriptome and genome sequencing. PLoS One 8, e70051.
Morin, P.A., Martien, K.K., Taylor, B.L., 2009. Assessing statistical power of SNPs for population structure and conservation studies. Mol. Ecol. Resour. 9, 66–73.
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., 1972. Genetic distance between populations. Am. Nat. 106, 283–292.
Oueslati, S., Fadhlaoui-Zid, K., Kada, O., Augé, M.T., Quignard, J.P., Bonhomme, F., 2014. Existence of two widespread semi-isolated genetic entities within Mediterranean anchovies. Mar. Biol. 161, 1063–1071.
Ozerov, M., Vasemagi, A., Wennevik, V., Diaz-Fernandez, R., Kent, M., Gilbey, J., Prusov, S., Niemela, E., Vaha, J.P., 2013. Finding markers that make a difference: DNA pooling and SNP-arrays identify population informative markers for genetic stock identification. PLoS One 8 (12), e82434.
Peakall, R., Smouse, P.E., 2012. Genalex 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 19, 2537–2539.
Piry, S., Alapetite, A., Cornuet, J.M., Paetkau, D., Baudouin, L., Estoup, A., 2004. GENECLASS2: a software for genetic assignment and first generation migrants detection. J. Hered. 95, 536–539.
Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.
Patarnello, T., Volckaert, F.A., Castilho, R., 2007. Pillars of Hercules: is the Atlantic-Mediterranean transition a phylogeographical break? Mol. Ecol. 16 (21), 4426–4444.
R Development Core Team, 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0 http://www.R-project.org/.

Raymond, M., Rousset, F., 1995. GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. J. Hered. 86, 248–249.
Rannala, B., Mountain, J.L., 1997. Detecting immigration by using multilocus genotypes. Proc. Natl. Acad. Sci. U. S. A. 94, 9197–9201.
Rice, W.R., 1989. Analyzing tables of statistical tests. Evolution 43, 223–225.
Rousset, F., 2008. GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol. Ecol. Resour. 8, 103–106.
Saatchi, M., Garrick, D.J., 2014. Developing a Reduced SNP Panel for Low-cost Genotyping in Beef Cattle. Animal Industry Report, AS 660, ASL R2854. http://lib.dr.iastate.edu/ans_air/vol660/iss1/19.
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.
Storer, C.G., Pascal, C.E., Roberts, S.B., Templin, W.D., Seeb, L.W., Seeb, J.E., 2012. Rank and order: evaluating the performance of SNPs for individual assignment in a non-model organism. PLoS One 7, e49018.
Viñas, J., Sanz, N., Peñarrubia, L., Araguas, R.M., García-Marín, J.L., Roldan, M.I., Pla, C., 2013. Genetic population structure of European anchovy in the Mediterranean Sea and the Northeast Atlantic Ocean using analysis of the mitochondrial DNA control region. ICES J. Mar. Sci. 71, 391–397.
Whitehead, P.J.P., Nelson, G.J., Wongratana, T., 1988. FAO species catalogue Vol. 7. Clupeoid fishes of the world. An annotated and illustrated catalogue of the herrings sardines, pilchards, sprats, shads, anchovies and wolf-herrings. Part 2. Engraulididae. FAO Fish. Synop. 7, 305–579.
Zarraonaindia, I., Iriondo, M., Albaina, A., Pardo, M.A., Manzano, C., Grant, W.S., Irigoien, X., Estonba, A., 2012. Multiple SNP markers reveal fine-scale population and deep phylogeographic structure in European anchovy (*Engraulis encrasicolus* L.). PLoS One 7, e42201.
Zarraonaindia, I., Pardo, M.A., Iriondo, M., Manzano, C., Estonba, A., 2009. Microsatellite variability in European anchovy (*Engraulis encrasicolus*) calls for further investigation of its genetic structure and biogeography. ICES J. Mar. Sci. 66, 2176–2182.