Supplementary Material on the Methodology Part VII

*The **TRANSBORAN project**. TRANSboundary population structure of Sardine, European hake and Blackspot seabream in the AlBORAN Sea and adjacent waters: a multidisciplinary approach.*

Final report on sardine genetics

Coordinators of genetics: Alessia Cariani (UniBo), Kenza Mokhtar-Jamaï (INRH), Carolina Johnstone (IEO-CSIC).

Authors of the Report: Carolina Johnstone[1] and Montse Pérez-Rodríguez[2].

[1]Centro Oceanográfico de Málaga, Instituto Español de Oceanografía (IEO), Consejo Superior de Investigaciones Científicas (CSIC). Email: carolina.johnstone@ieo.es. [2]Centro Oceanográfico de Vigo, Instituto Español de Oceanografía (IEO), Consejo Superior de Investigaciones Científicas (CSIC). Email: montse.perez@ieo.es.

INTRODUCTION

A fish stock, from a genetic perspective, is understood as a reproductively isolated unit or population that is genetically different from other populations (or stocks). Stock boundaries can be studied establishing genetic diversity of populations. Genetic variation can be neutral or adaptive if it is influenced by environmental change. Neutral markers are the best choice to investigate connectivity or gene flow, migration and dispersal. Currently, two types of markers, microsatellites as neutral and SNPs as non-neutral, are being mostly used to decipher fish stock identity (Cuéllar-Pinzón et al., 2016).

Sardine (*Sardina pilchardus*; Walbaum, 1792) is a small pelagic marine species of commercial interest that inhabits waters of the central and north eastern Atlantic Ocean and the Mediterranean Sea (Fig. 1). Efforts to genetically characterize sardine stocks from the western Mediterranean go back more than two decades, with the first studies using allozymes. The initial studies were based on 15 loci encoding 11 allozymes and found a discontinuity at the enzymatic level between the Alboran Sea and other locations in the Spanish western Mediterranean and French Gulf of Lyon (Ramon and Castro, 1997). This study concluded sardine populations of the western Mediterranean are non panmictic but semi-independent isolated units, and detected the signal of the influence of the Almeria Oran front in isolating populations of the Alboran sea. Allozymes were further used to characterize sardines collected along the Atlantic and Mediterranean Moroccan coast and found a genetic break in the area of the Bay of Agadir (Chlaida et al., 2006, 2009). Mitochondrial DNA sequences are used to investigate the evolutionary history of species, and for sardine, they suggest past isolation of sardine populations in two distinct groups and expansion since the Pleistocene (Atarhouch et al., 2006).

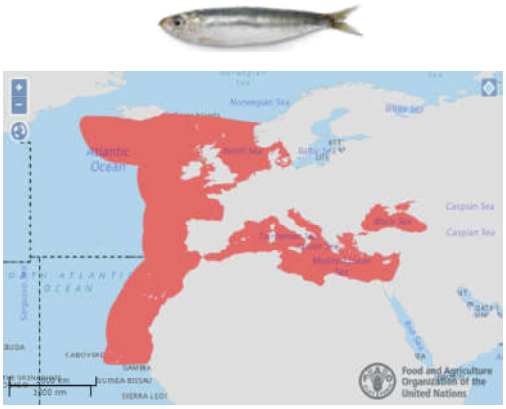

Figure 1. Image of *S. pilchardus* (above) and its geographic distribution according to FAO (Food and Agriculture Organisation (FAO), 2021).

To determine population genetic structure in sardine the first microsatellite loci were described almost 15 years ago (Gonzalez and Zardoya, 2007a). The genetic diversity at eight loci, SAR1.12, SAR1.5, SAR2.18, SAR9, SAR19.B3, SAR19B5, SARA2F and SARA3C, was analyzed in sardines collected in Atlantic and Mediterranean coasts finding weak differences consistent with a high level of gene flow between these populations and a pattern of isolation-by-distance (Gonzalez and Zardoya, 2007b). Another study assessed diversity at four microsatellite loci (Sp2, Sp7, Sp8 and SpI5), and in the control region of mitochondrial DNA in sardine populations spanning the Atlantic coast of Morocco, finding homogeneity with low but significant genetic differentiation following an isolation-by-distance pattern (Baibai et al., 2012). Five microsatellite loci, Sp2, Sp7, Sp8, SpI5 and Sp22, were studied in sardines collected

throughout its geographic range, from the North Sea to the Mediterranean Sea, detecting weak significant genetic structuring, pointing to a lack of genetic population structure (Kasapidis et al., 2012). Locus Sp22 seemed to be under hitchhiking selection, exhibiting a latitudinal cline with changes in allele frequencies from the Alboran Sea to the western Mediterranean and from the east Atlantic coast to the Azores and Madeira (Kasapidis et al., 2012). Sardines collected in eastern and central Mediterranean were characterized at eight microsatellite loci (SAR1D06, SAR9, SAR1.5, SAR1.12, Sp10, SpI5, SpI7 and SpIII93), and no genetic structure was found in the Adriatic sea (Ruggeri et al., 2012, 2013), in accordance with previous studies that point to a unique sardine population within the Mediterranean Sea.

MATERIALS AND METHODS

Sampling design for genetics

Sardines were collected for the *TRANSBORAN* project in 17 sampling ports belonging to 7 geographical subregions (Fig. 2).

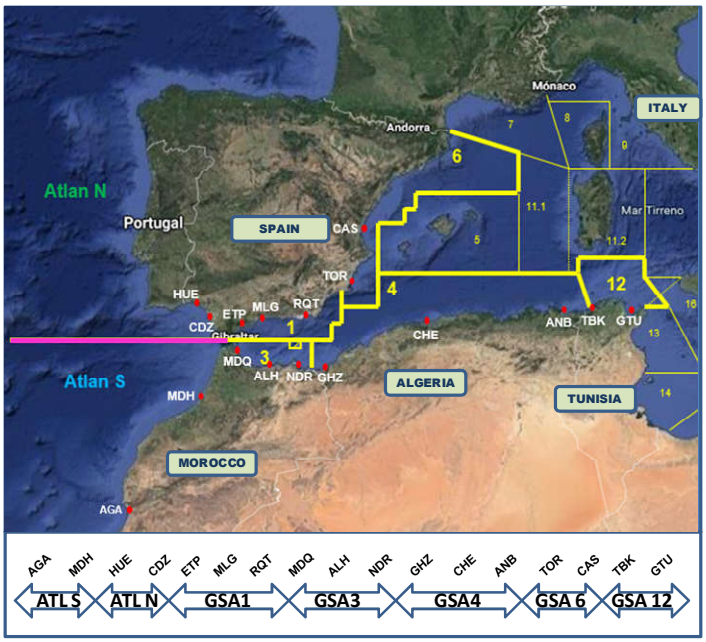

Figure 2. TRANSBORAN project sampling of sardine. Red dots indicate a total of 17 ports where sardines were collected. Geographical subregions managed by GFCM are indicated with yellow lines and those managed by ICES or CECAF are separated by a pink line. The image is a modification of an image authored by Teresa Pérez.

For genetics, tissue was dissected from fresh or frozen specimens, preserved in vials with ethanol 96%, and stored at 4°C or frozen at -20°C until shipping to the facilities of IEO at Vigo. The number of samples collected in each port is indicated in Table 1.

Table 1. Collection of sardine samples for genetics

GSA	FRO	Port	Code	N
Atl S	CECAF	Agadir	AGA	50
		Medhia	MDH	50
Atl N	ICES	Huelva	HUE	50
		Cádiz	CDZ	50
GSA3	GFCM	M´diqu	MDQ	50
		AlHoceima	ALH	50
		Nador	NDR	50
GSA1	GFCM	Estepona	ETP	50
		Málaga	MLG	50
		Roquetas	RQT	45
GSA3	GFCM	M´diqu	MDQ	50
		AlHoceima	ALH	50
		Nador	NDR	50
GSA4	GFCM	Ghazaouet	GHZ	52
		Cherchell	CHE	57
		Annaba	ANB	55
GSA6	GFCM	Torrevieja	TOR	50
		Castellón	CAS	50
GSA12	GFCM	Tabarka	TBK	50
		G. of Tunis	GTU	50

Genetic analysis of sardines

A portion of sardine muscle (approx. 100 mg) was used for extraction of genomic DNA from a total of 859 sardines. The FENOSALT method (Pérez, M., Presa, 2011) and Tissue gDNA isolation kit (NZYTech) were used to extract genomic DNA.

Twenty microsatellites loci were selected from the literature (Supplementary information, Table S1). Different temperatures of annealing were used in order to optimize the amplification efficiency for each loci separately and to establish the common amplification conditions in a multiplex PCR. Each locus was first amplified individually testing different annealing temperatures (TaS in Table S1). To group microsatellite loci in multiplex PCR (mPCR) using four fluorophore dyes (6-FAM, VIC, NED, PET) an *in silico* approximation was carried out with the Multiplex Manager 1.0® software (Holleley and Geerts, 2009). Microsatellite markers with non-specific amplification (several bands) or without amplification band were excluded from the design. Two mPCR were designed to amplify 12 microsatellite markers combining 6 loci per reaction. PCR constant conditions were 1 X PCR buffer (10x NH4 5x MyTaq™ Reaction Buffer Bioline), 10 mM di-nucleotide triphosphate (dNTP NZYSet, NZYTech), 0.1 to 0.3 µM of each primer and 0.2 U of polymerase (BIOTAQ ™ DNA polymerase, Bioline). The PCR amplification conditions were: 5 min at 95 ° C, 40 cycles consisting of 20 sec at 95 ° C, 25 sec at annealing temperature 55 ºC for mPCR 1 and 59 ºC for the mPCR 2, 45 sec at 72 ° C, and at an extension step at 72 ° C for 7 min.

Table 2. Multiplex PCRs designed to amplify nine microsatellite loci of *S. pilchardus*.

PCR[a]	Loci	Motif	Primer (5´-3´)[b]	Size range (bp)[c]	Reference
mPCR1	Sp8	$(CA)_n$	F: *VIC®ACGTCACAGTTCCCCACTG R : ACTGGCTGAGGAGGATGATG	109-211	(Kasapidis et al., 2012)
				110-160	(Cilli, 2009)
				111-237	(Baibai et al., 2012)
	SpIII93	$(ATCT)_n$	F: *FAM™TAAGCAGACGCGAAACTGAA R:CTTGCGACCTGACGTGATTA	170-292	(Ruggeri et al., 2013)
	Sp2	$(AG)_n/(TG)_n$	F: *NED™CGAGGCCTGATAGAAACCC R : AACCACGGTCAGTTCTCCAG	138-254	(Kasapidis et al., 2012)
				146-238	(Cilli, 2009)
				122-274	(Baibai et al., 2012)
	SAR1D06	$(TG)_n$	F:*PET™CGGCTATTCTTAGACTAGGTG R: CCCCATCAGCAATGAATAAG	120-158	(Ruggeri et al., 2012, 2013)
mPCR2	SpI5	$(TATC)_n$	F:*FAM™TGGCCTGTGATCTACAGTATGG R :CAATTTTGCCAGTTGTTTAGTC	123-183	(Ruggeri et al., 2012, 2013)
	SpI7	$(AGAT)_n$	F:*VIC®TGCTTTACTTCATTCCGTTGAA R: TCACATCATCACAACAAACACC	117-141	(Ruggeri et al., 2012, 2013)
	SARB-A07	$(GA)_n$	F:*NED™CTCCTCACTCAGCCGCTAAGGA R:GGGTAACATTTCGGCAAGTGCT	68-136	(Ruggeri et al., 2013)
	SAR1.12	$(GT)_n$	F:*NED™TGAGAATCACAGAATCTGAGCA R : CTGGAAGCTCTTGGCATCTT	166-292	(Gonzalez and Zardoya, 2007a, 2007b)
				183-273	(Ruggeri et al., 2013)
	Sp22	$(CA)_n/(GA)_n$	F:*PET™GGTCACGTTGGGTATTTGCTAT R : GTCTGGAACTGGGTGCTCTC	194-232	(Kasapidis et al., 2012)
				198-230	(Cilli, 2009)

[a]mPCR, multiplex PCR. [b]F: forward, R: reverse. Forward primers labelled with VIC®, FAM™, NED™, or PET™, as indicated. [c]Reported size range in literature (bp, base pairs).

Three microsatellite markers showing more than one band after PCR amplifiation were finally excluded from the designed mPCR. The final design allowed amplification of nine microsatellite loci in two mPCR (Table 2). Fragment analysis of the products of mPCR1 and mPCR2 was performed at CACTI (UVigo) using GeneScanTM 500 LIZ (Applied Biosystems, Foster City, CA, USA) as size standard. Alleles were called at 9 loci in 859 sardines with software GeneMarker v2.7.0 (SoftGenetics) by at least two independent readers. The established internal threshold for missing data per individual was 22% (two loci with missing data). A percentage of samples were replicated to confirm correct allele calling. The approximate number of electropherogram files generated for allele calling was 2500.

Statistical analysis of sardine genetic diversity

Microsatellite genetic data for 805 sardines at 9 loci were examined with GenAlEx software v6.5 (Peakall and Smouse, 2012) to obtain allele frequency and estimate genetic diversity through observed (H_O) and expected (H_E) heterozygosity. Genotyping scoring errors (null alleles, stuttering and drop-out) were assessed with MICRO-CHECKER (Van Oosterhout et al., 2004). The inbreeding coefficient F_{IS} (Weir and Cockerham, 1984) and compliance to equilibrium of Hardy–Weinberg (HW) were studied with GENEPOP 4.7 (Raymond and Rousset, 1995; Rousset, 2008). Probability values were corrected (Weir, 1996) due to low counts for certain genotypes with the Markov Chain Monte Carlo (MCMC) approximation (involving 10,000 dememorization steps, 1,000 batches and 10,000 iterations per batch).

Partitioning of genetic diversity at microsatellite loci within subpopulations relative to the total population was studied with Wright's F statistic F_{ST} obtained with GenAlEx v6.5 (Peakall and Smouse, 2012), and the associated probability (p) based on a 999 data permutation test. The number of migrants (Nm) was obtained with GenAlEx v6.5 (Peakall and Smouse, 2012) and GENEPOP 4.7 (Raymond and Rousset, 1995; Rousset, 2008) when corrected for sample size.

Clustering of genetic diversity was investigated with three different approaches, Principal Coordinate analysis (PCoA), Discriminant Analysis of Principal Components (DAPC) and Bayesian algorithms to proportionally assign individuals to inferred population clusters. PCoA was performed with GenAlEx software v6.5 (Peakall and Smouse, 2012) based on the individual-by-individual pairwise genetic distances averaged for each population. In the DAPC approach we used the R package adegenet (Jombart, 2008) in RStudio (RStudio Team, 2019) with R free software (R Team, 2018) to study differences among clusters identified from genotyped sardines. STRUCTURE v2.3.4 software (Pritchard et al., 2000) was used for the Bayesian approach assuming an ancestry admixture correlated allele frequency model (Falush et al., 2003), and considering prior sampling location information. Running parameters were set to a burn-in of 2.5×10^4 followed by a MCMC simulation of 5×10^4 runs simulating K 1 to 7 or 17 populations in 25 iterations. STRUCTURE results were explored with Structure Harvester (Earl and vonHoldt, 2012) and Clumpak (Kopelman et al., 2015) was used to detect the consensus solutions for the K clusters that best fit the Bayesian algorithm used by STRUCTURE.

RESULTS

Characterization of microsatellite loci and genetic diversity

Nine microsatellite loci of *S. pilchardus* were characterized in 859 sardines, and after exclusion of genotypes with missing data, a total of 805 multilocus genotypes were further analysed. Table 3 shows the number (N) of sardines genotyped at each loci, and the total number of alleles (Na) counted. Data were screened with MICRO-CHECKER (Van Oosterhout et al., 2004) for scoring errors and null alleles were highly probably in two loci, SAR-1.12 and SpIII93, for all sampled ports (Table 3). Null alleles were mainly excluded for loci SpI7, SARB-A07, Sp22, Sp8, Sp2 and SAR1D06. The results for locus SpI5 indicated scoring errors with null alleles may have occurred in certain, but not all, sampled ports. Locus Sp22 was the least variable with a total Na of 22, whereas loci SAR-1.12 and SpIII93 were the most polymorphic with a total Na counted of 64.

Table 3. Characterization of sardine microsatellite loci.

Pop[a]	SpI5	SpI7	SARB-A07	SAR-1.12	Sp22	Sp8	SpIII93	Sp2	SAR1D06
AGA	49/20	49/10	42/24	43/28	48/9	49/27	49/20	49/30	49/17
MDH	49/17	49/11	45/30	46/25	48/8	48/27	43/26	46/31	40/14
HUE	49/22	49/11	47/28	46/27	48/12	49/23	41/30	47/30	43/13
CDZ	48/21	48/9	48/26	41/24	48/7	48/27	48/27	48/34	48/13
ETP	40/20	40/10	40/24	28/22	39/10	40/28	38/23	39/29	36/11
MLG	47/24	47/9	47/26	38/26	47/10	47/30	45/22	46/32	46/16
RQT	44/19	43/10	41/25	42/27	44/7	44/28	43/24	44/26	44/18
MDQ	46/18	45/11	44/21	41/26	45/9	46/29	46/26	45/28	45/17
ALH	47/18	45/7	44/22	42/25	46/10	47/27	47/29	47/30	47/16
NDR	48/24	48/7	47/27	42/25	48/12	48/27	48/27	48/32	48/16
GHZ	51/16	51/9	51/26	49/26	51/10	51/26	45/31	45/31	47/14
CHE	49/22	49/8	43/22	40/25	48/11	49/24	49/31	49/30	49/16
ANB	54/22	54/**6**	45/27	49/26	49/10	54/31	46/32	52/32	49/15
TOR	48/25	48/**5**	48/28	42/28	48/10	48/25	47/31	48/29	48/16
CAS	46/22	46/**6**	46/29	43/25	46/13	46/25	46/28	46/29	46/14
TBK	46/20	46/**4**	46/26	42/25	46/9	46/26	46/28	46/27	46/16
GTU	44/24	44/**5**	44/26	36/25	44/12	44/28	44/31	44/27	44/15
Total Na	53	25	39	64	22	43	64	54	27

[a]Pop refers to each port where sardines were collected. Number of sardines (N) and number of alleles (Na) characterized for each microsatellite loci in the 17 ports under study. Each cell indicates data for N/Na, and are underlined when probable presence of null alleles was detected. The last row indicates the total Na counted for each locus.

Deviations from expected proportions in Hardy–Weinberg (HW) equilibrium imply non-random mating, selection for certain genotypes, mutations, or small population sizes. Under HW equilibrium assumptions, the difference between expected and observed heterozygosities can be measured with the inbreeding coefficient F_{IS} (Weir and Cockerham, 1984). Table 4 shows the values obtained for F_{IS} and the probabilities associated to HW equilibrium. In all ports significant deviations from HW proportions and homozygote excess (positive F_{IS}) were obtained for loci SAR-1.12 and SpIII93, in accordance with the presence of null alleles reported for these loci.

Table 4. Marker assessment of sardine microsatellite loci.

Pop[a]	SpI5	SpI7	SARB-A07	SAR-1.12	Sp22	Sp8	SpIII93	Sp2	SAR1D06
AGA	0.130	0.097	0.066	0.514***	0.008	-0.013	0.232**	0.003	0.056
MDH	0.083*	-0.088	-0.020	0.516***	0.150	0.024	0.265**	0.032	-0.023
HUE	0.189**	-0.044	0.036	0.350***	0.094	0.056	0.149**	-0.013	-0.005
CDZ	0.170*	0.020	0.018	0.477***	0.103	0.008	0.244***	0.076	0.167*
ETP	0.077	-0.004	0.077*	0.471***	0.106	0.136*	0.247***	0.018	0.007
MLG	0.053	-0.039	-0.008	0.498***	-0.024	0.003	0.296**	0.053	0.157*
RQT	0.172*	-0.022	0.073*	0.466***	0.101	0.010	0.236***	-0.018	0.016
MDQ	0.200	0.071	-0.057	0.285***	0.100	0.071	0.175**	0.046	0.062
ALH	0.063*	-0.098	0.038	0.655***	0.027	0.027	0.123*	0.038*	-0.023
NDR	0.133*	0.186*	-0.013	0.621***	0.091	0.003	0.198**	0.110**	0.046
GHZ	0.079*	0.027	-0.055	0.545***	0.074	0.098*	0.185**	0.035*	-0.060
CHE	0.120**	0.133	0.020	0.409***	0.067	0.100*	0.216***	0.022	0.068
ANB	0.232	-0.095	-0.036	0.619***	0.046	0.006	0.249***	0.022	-0.018
TOR	0.097	-0.082	0.013	0.574***	0.020	0.059	0.381***	0.020	0.062*
CAS	0.144	0.219	-0.007	0.507***	0.078	0.000	0.268***	0.009	0.008
TBK	0.066	0.073	0.032	0.617***	0.058	-0.052	0.158*	0.031	0.025
GTU	0.090	0.145	0.045	0.496***	-0.143	0.003	0.195***	0.010	0.085

[a]Pop refers to each port where sardines were collected. Each cell indicates the value of the inbreeding coefficient FIS calculated according to (Weir and Cockerham, 1984), followed by the significance level (* $p < 0.05$, ** $p < 0.01$, or *** $p < 000.1$) of the probability p value obtained with the exact probability test for Hardy–Weinberg equilibrium calculated by the Markov chain method (10,000 dememorization, 1,000 batches, 1,0000 iterations per batch).

At each sampling site, or "pop" (referring to population), the multilocus genotypes characterized at nine microsatellite loci were used to estimate genetic diversity as heterozygosity (Table 6), showing very similar values for the 17 populations analyzed.

Table 6. Estimation of sardine genetic diversity per sampling port.

Pop[a]	N	Na	Ne	H_O	H_E	uH_E
AGA	47	21	11	0.731	0.830	0.839
MDH	46	21	11	0.749	0.835	0.844
HUE	47	22	11	0.761	0.834	0.843
CDZ	47	21	11	0.719	0.836	0.845
ETP	38	20	11	0.748	0.851	0.863
MLG	46	22	11	0.745	0.839	0.848
RQT	43	20	11	0.742	0.835	0.845
MDQ	45	21	10	0.756	0.837	0.846
ALH	46	20	10	0.749	0.830	0.839
NDR	47	22	12	0.722	0.841	0.850
GHZ	49	21	12	0.747	0.832	0.840
CHE	47	21	11	0.742	0.842	0.851
ANB	50	22	12	0.739	0.840	0.849
TOR	47	22	12	0.713	0.829	0.838
CAS	46	21	12	0.743	0.845	0.854
TBK	46	20	12	0.740	0.828	0.837
GTU	43	21	12	0.752	0.828	0.838

[a]Pop refers to each port where sardines were collected. Each cell indicates the average value per port. N, number of individuals analyzed; Na, number of different counted alleles; Ne, number of effective alleles according to allele frequencies; H_O, observed heterozygosity; H_E; expected heterozygosity according to allele frequencies; uH_E, unbiased expected heterozygosity according to sample size.

Structuring of population genetic diversity and migration

Structuring of genetic diversity was assessed from sardine multilocus genotypes grouped in 17 populations (ports, Tables 7 and 8) or 7 populations (geographical subregions, Tables 9 and 10), either considering all nine microsatellite loci characterized (Table 7 and 9), or from the data for seven loci (Table 8 and 10), excluding SAR1.12 and SpIII93, the two loci with null alleles in all populations. The total expected genetic diversity was partitioned between populations under the assumption of equilibrium, calculating through pairwise comparisons the general or most frequently used fixation index F_{ST}, originally established by Wright, ranging between 0 and +1. Non-significant close to zero values of F_{ST} exclude structuring of genetic diversity between the pair of compared populations.

Table 7. F_{ST} pairwise comparison of sardine genetic diversity at nine microsatellite loci characterized in 17 populations (ports).

	AGA	MDH	HUE	CDZ	ETP	MLG	RQT	MDQ	ALH	NDR	GHZ	CHE	ANB	TOR	CAS	TBK	GTU	
AGA																		**AGA**
MDH	0.007																	**MDH**
HUE	0.009**	0.007																**HUE**
CDZ	0.014**	0.008*	0.008*															**CDZ**
ETP	0.012**	0.007	0.007	0.007														**ETP**
MLG	0.011**	0.008	0.008*	0.007	0.008													**MLG**
RQT	0.013**	0.007	0.007	0.006	0.007	0.007												**RQT**
MDQ	0.012**	0.008	0.007	0.008	0.006	0.007	0.007											**MDQ**
ALH	0.013**	0.008	0.007	0.007	0.006	0.009*	0.007	0.006										**ALH**
NDR	0.012**	0.008	0.007	0.008	0.006	0.007	0.007	0.007	0.006									**NDR**
GHZ	0.013**	0.008	0.006	0.007	0.006	0.007	0.008	0.007	0.007	0.006								**GHZ**
CHE	0.012**	0.008	0.007	0.005	0.006	0.006	0.006	0.008	0.007	0.006	0.006							**CHE**
ANB	0.018**	0.011**	0.010**	0.008*	0.008	0.009**	0.010**	0.008*	0.007	0.007	0.008*	0.007						**ANB**
TOR	0.023**	0.015**	0.013**	0.010**	0.011**	0.011**	0.012**	0.011**	0.011**	0.009*	0.009**	0.010**	0.007					**TOR**
CAS	0.021**	0.011**	0.012**	0.007	0.008	0.009*	0.008*	0.009**	0.008	0.008	0.008*	0.008*	0.005	0.006				**CAS**
TBK	0.024**	0.015**	0.013**	0.009*	0.011**	0.010**	0.011**	0.012**	0.011**	0.010**	0.008*	0.009**	0.006	0.007	0.006			**TBK**
GTU	0.020**	0.012**	0.009**	0.007	0.009	0.008*	0.009*	0.009**	0.008	0.008	0.007	0.008	0.006	0.007	0.006	0.008		**GTU**
	AGA	**MDH**	**HUE**	**CDZ**	**ETP**	**MLG**	**RQT**	**MDQ**	**ALH**	**NDR**	**GHZ**	**CHE**	**ANB**	**TOR**	**CAS**	**TBK**	**GTU**	

Significance levels indicated with * ($p < 0.05$) and ** ($p < 0.01$). All values non significant after Bonferroni correction (α 0.05) for multiple comparisons. The inset square surrounds pairwise comparisons between ports from the Alboran Sea geographical subregions GSA1 and GSA3.

Table 8. F_{ST} pairwise comparison of sardine genetic diversity at seven microsatellite loci characterized in 17 populations (ports).

	AGA	MDH	HUE	CDZ	ETP	MLG	RQT	MDQ	ALH	NDR	GHZ	CHE	ANB	TOR	CAS	TBK	GTU	
AGA																		**AGA**
MDH	0.007*																	**MDH**
HUE	0.008*	0.005																**HUE**
CDZ	0.017**	0.008*	0.008*															**CDZ**
ETP	0.013**	0.006	0.006	0.007														**ETP**
MLG	0.013**	0.007	0.007	0.006	0.008													**MLG**
RQT	0.014**	0.007	0.007	0.005	0.007	0.006												**RQT**
MDQ	0.013**	0.008	0.006	0.008*	0.005	0.006	0.007											**MDQ**
ALH	0.014**	0.006	0.007	0.007	0.006	0.007	0.006	0.006										**ALH**
NDR	0.013**	0.007	0.006	0.008*	0.005	0.006	0.007	0.006	0.004									**NDR**
GHZ	0.015**	0.007	0.006	0.007	0.006	0.007	0.008	0.006	0.006	0.006								**GHZ**
CHE	0.013**	0.006	0.006	0.005	0.006	0.005	0.006	0.008	0.006	0.006	0.005							**CHE**
ANB	0.021**	0.010**	0.011**	0.008*	0.007	0.006	0.009	0.007	0.007	0.006	0.007*	0.007						**ANB**
TOR	0.027**	0.014**	0.013**	0.010**	0.010**	0.009**	0.011**	0.011**	0.009**	0.009*	0.009**	0.009**	0.005					**TOR**
CAS	0.025**	0.012**	0.013**	0.008*	0.008*	0.008*	0.009**	0.010*	0.008*	0.008**	0.008**	0.009**	0.004	0.006				**CAS**
TBK	0.027**	0.014**	0.014**	0.008	0.010**	0.009**	0.010**	0.011**	0.009**	0.009*	0.008*	0.009*	0.005	0.006	0.006			**TBK**
GTU	0.023**	0.011**	0.010**	0.007	0.008	0.008	0.008**	0.009	0.007	0.007	0.007	0.007	0.006	0.006	0.006	0.007		**GTU**
	AGA	**MDH**	**HUE**	**CDZ**	**ETP**	**MLG**	**RQT**	**MDQ**	**ALH**	**NDR**	**GHZ**	**CHE**	**ANB**	**TOR**	**CAS**	**TBK**	**GTU**	

Significance levels indicated with * ($p < 0.05$) and ** ($p < 0.01$). All values non significant after Bonferroni correction (α 0.05) for multiple comparisons. The inset square surrounds pairwise comparisons between ports from the Alboran Sea geographical subregions GSA1 and GSA3.

Table 9. F_{ST} pairwise comparison of sardine genetic diversity at 9 loci in 7 populations.

	AtlS	AtlN	GSA1	GSA3	GSA4	GSA6	GSA12	
AtlS								**AtlS**
AtlN	0.006**							**AtlN**
GSA1	0.006**	0.002						**GSA1**
GSA3	0.006**	0.003	0.002					**GSA3**
GSA4	0.007**	0.003	0.003	0.002				**GSA4**
GSA6	0.014**	0.007**	0.006**	0.006**	0.004**			**GSA6**
GSA12	0.014**	0.006**	0.005**	0.006**	0.003	0.003		**GSA12**
	AtlS	**AtlN**	**GSA1**	**GSA3**	**GSA4**	**GSA6**	**GSA12**	

Significance levels indicated with * ($p < 0.05$) and ** ($p < 0.01$). All values non significant after Bonferroni correction (α 0.05) for multiple comparisons. The inset square surrounds pairwise comparisons between the Alboran Sea geographical subregions GSA1 and GSA3.

Table 10. F_{ST} pairwise comparison of sardine genetic diversity at 7 loci in 7 populations.

	AtlS	AtlN	GSA1	GSA3	GSA4	GSA6	GSA12	
AtlS								**AtlS**
AtlN	0.006**							**AtlN**
GSA1	0.006**	0.002						**GSA1**
GSA3	0.006**	0.003	0.002					**GSA3**
GSA4	0.008**	0.003	0.002	0.002				**GSA4**
GSA6	0.016**	0.008**	0.006**	0.006**	0.004**			**GSA6**
GSA12	0.015**	0.006**	0.005**	0.005**	0.003	0.003		**GSA12**
	AtlS	**AtlN**	**GSA1**	**GSA3**	**GSA4**	**GSA6**	**GSA12**	

Significance levels indicated with * ($p < 0.05$) and ** ($p < 0.01$). All values non significant after Bonferroni correction (α 0.05) for multiple comparisons. The inset square surrounds pairwise comparisons between the Alboran Sea geographical subregions GSA1 and GSA3.

Structuring of genetic diversity assessed through F_{ST} supports existing gene flow between the sampling locations under study, with connectivity within the Alboran Sea and adjacent waters. Significant pairwise comparisons is always observed in comparisons with the port of Agadir (AGA) or with the south Atlantic region (AtlS). Values of F_{ST} increase for pairwise comparisons between AGA or AtlS and the sampling ports or geographic subregions of the further Mediterranean, ports TOR and CAS in GSA6, and ports TBK and GTU in GSA12. These results indicate a possible barrier to gene flow in the south Atlantic (AtlS) and between the ports of AGA and MDH, or a pattern of isolation-by-distance. The Canary Current may be the hydrographical factor contributing to this barrier to gene flow.

Considering the high levels of connectivity within the region the estimation of migrants between sampling sites can assist in the study of population structure. Table 11 shows the number of migrants (Nm) estimated per locus from multilocus genotypes grouped in 17 populations (ports) or 7 populations (geographical subregions or GSA). Locus Sp22 has the lowest estimated Nm. Microsatellite loci Sp22 has been proposed to be under hitchhiking selection (Kasapidis et al., 2012), and the microsatellite marker that appears to be driving structuring. Another approach to study migration considers private alleles (Barton and Slatkint, 1986), and corrects for size the number of migrants estimated. Multilocus genotypes grouped in 17 populations (sampled ports) have a mean sample size of 46 and the number of migrants estimated is 15. When genotypes are grouped in 7 populations (geographical subregions) the mean sample size is 111 and the number of migrants estimated is 26. This high level of migration between populations contributes to gene flow and population connectivity, in agreement with the results obtained for F_{ST}.

Table 11. Number of migrants.

Locus	17 Pops	7 Pops
SpI5	19	52
SPI7	13	17
SARB-A07	24	62
SAR-1.12	13	31
Sp22	6	7
Sp8	25	60
SpIII93	15	27
Sp2	23	57
SAR1D06	18	37
Mean	17	39

Clustering of sardine genetic diversity

According to F_{ST} pairwise comparisons and number of migrants, the sardine populations studied are highly connected. Clustering of genetic diversity may allow discerning weak barriers to gene flow. Principal Coordinate analysis (PCoA) was performed to visualize the patterns of genetic variation based on the individual-by-individual pairwise genetic distances averaged for each population. Considering multilocus genotypes at 7 microsatellite loci of 805 sardines collected in 17 populations (ports), PCoA retains 77.65% variance in the first two axis (Fig. 3). Grouping the data in 7 populations (geographical subregions) retains up to 95% variance (Fig. 4). The results of PCoA again illustrate a high level of gene flow between the areas where the sardines were collected, with no clear barriers between different ports or geographical subregions, with geographically nearby populations appearing separated. An example is the observed distance between ports of GSA1 Estepona (ETP), Málaga (MLG) and Roquetas (RQT) in Fig. 3.

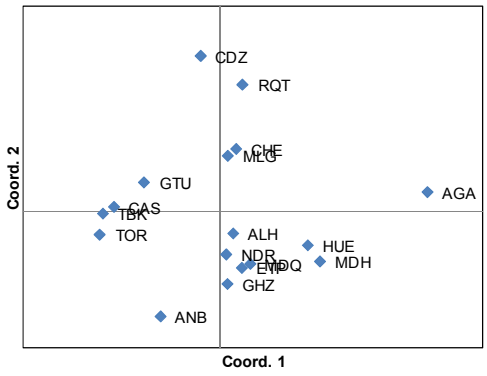

Figure 3. Principal Coordinate Analysis (PCoA). Analysis performed with genetic data for 805 sardines genotyped at 7 microsatellite loci and grouped in 17 sampling ports. The percentage of variation explained by the first two axes is 67.01% and 10.64% respectively.

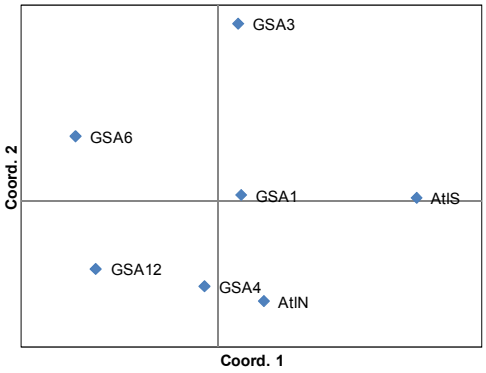

Figure 4. Principal Coordinate Analysis (PCoA). Analysis performed with genetic data for 805 sardines genotyped at 7 microsatellite loci and grouped in 7 geographical subregions. The percentage of variation explained by the first two axes is 90.17% and 4.82% respectively.

A different approach to cluster genetic diversity is DAPC, that was used to assess differences among inferred clusters, again grouping data in 17 populations (ports, Fig. 5) or 7 populations (geographical subregions, Fig. 6).

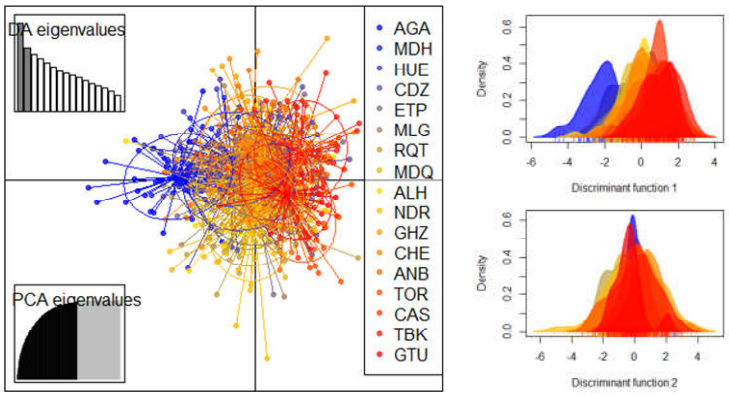

Figure 5. Discriminant Analysis of Principal Components (DAPC). Analysis performed with genetic data for 805 sardines genotyped at 7 microsatellite loci and grouped in 17 sampling ports. The proportion of conserved variance is 99.5%. The eigenvalues for the first two vectors is 35.32 and 28.95.

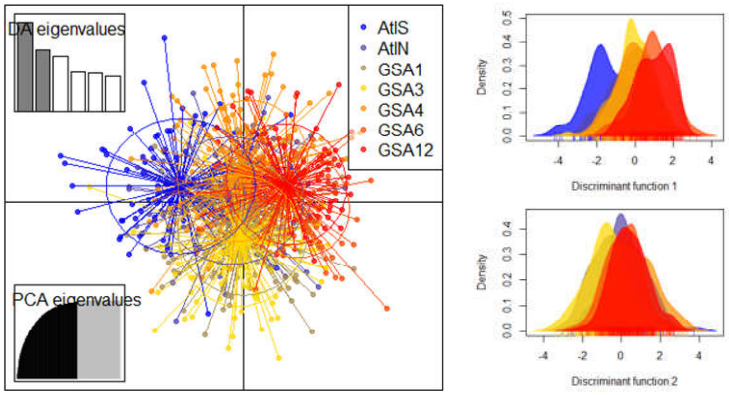

Figure 6. Discriminant Analysis of Principal Components (DAPC). Analysis performed with genetic data for 805 sardines genotyped at 7 microsatellite loci and grouped in 7 geographical subregions. The proportion of conserved variance is 99.5%. The eigenvalues for the first two vectors is 78.01 and 58.22.

DAPC is capable of capturing the weak structuring of genetic diversity between AGA (Fig. 5), or the south Atlantic geographical subregion (AtlS) (Fig. 6), and the rest of sampled areas analyzed, as obtained for F_{ST} pairwise comparisons (Tables 7-10).

Structuring of genetic diversity was further investigated through assignment of individual sardines to a number of *K* clusters using Bayesian algorithms. Multilocus genotypes of sardines at 7 to 9 microsatellite loci clustered in *K* 7 or 17 were explored assuming an admixture model and considering prior sampling location. There was no consistency among the different simulations, even when replicate analyses were run.

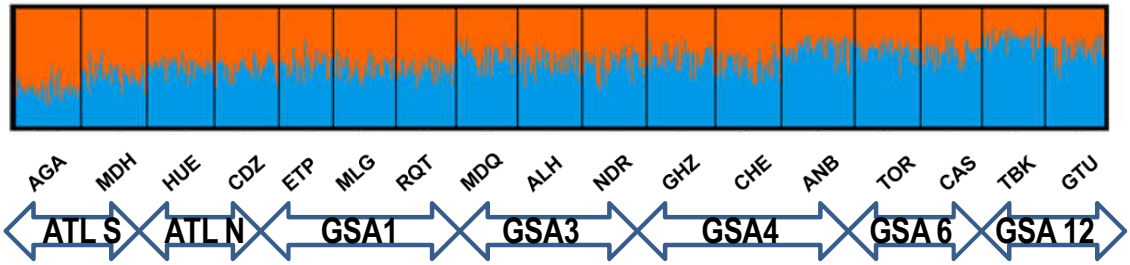

Figure 7. Consensus solution for clustering of genetic diversity through Bayesian algorithms. A simulation was run for *K* 1 to 7 with the genetic data of 805 sardines genotyped at 6 microsatellite loci strictly in HW. STRUCTURE software was used considering an ancestry admixture correlated allele frequency model, and prior sampling location information. Structure Harvester exploration of results indicated the best *K* was 2, and Clumpak produced the most frequent consensus solution for *K* = 2, proportionally assigning each of the 805 individuals to two populations (blue or orange).

To strictly comply with HW assumptions for STRUCTURE software, only six microsatellite loci (SpI7, SARB-A07, Sp22, Sp8, Sp2 and SAR1D06) were used. The results obtained support a best *K* of two clusters according to the statistic ΔK (deltaK) proposed by (Evanno et al., 2005). The consensus solution for the *K* clusters that best fit the Bayesian algorithm is shown in Figure 7, and is consistent with a high level of gene flow in the studied area and a weak barrier to gene flow between the ports of AGA and MDH. The fact that clustering of genetic diversity with a Bayesian approach did not provide a coherent picture between simulations is in agreement with an absence of structure (*K* =1) and a high level of gene flow, as pointed by F_{ST} pairwise comparisons, number of migrants, PCoA and DAPC.

CONCLUSIONS AND RECOMMENDATIONS

Selected microsatellite markers

Twenty microsatellite loci were selected from the literature and tested for sardine multiplex PCR. Nine were finally included in two multiplex PCR reactions, mPCR 1 (Sp8, SpIII93, Sp2, SAR1D06) and mPCR2 (SpI5, SpI7, SARB-A07, SAR1.12, and Sp22).

A total of 805 sardines collected in 17 ports in the Alboran Sea and adjacent waters were characterized for the 9 microsatelite loci. Microsatellite loci SpI5, SpI7, SARB-A07, Sp22, Sp8, Sp2 and SAR1D06 were in HW and displayed low null allele frequencies. Locus SpI5 displayed a higher null allele frequency and was not in HW in all the studied ports. Two markers, SpIII93 and SAR1.12 did not comply to HW requirements and had null alleles.

Microsatellite locus Sp22 appears to be under hitchhiking selection as reported in the literature.

Genetic diversity estimated through number of alleles (Na), and observed and expected heterozygosities (H_O and H_E) was similar to the values reported in the literature.

Structuring and Connectivity

No evidence of genetic structure between sardine populations of the Alboran Sea was found.

No evidence of strong genetic structure between sardine populations of the Alboran Sea and adjacent Mediterranean or Atlantic populations was found.

Weak structuring between populations of the south Atlantic (Agadir) and other sardine populations was found, suggesting a barrier to gene flow may exist between Medhia and Agadir in the south Atlantic Moroccan coast, in accordance to an isolation-by-distance pattern reported in the literature.

There is evidence of connectivity between sardine populations of the Alboran Sea.

There is a high level of gene flow between sardine populations from the Atlantic Ocean and the Mediterranean Sea.

Future perspectives

The spatial organization of sardine populations can be further assessed using geographic and genetic information, and may contribute to investigate the suggested isolation-by-distance pattern.

Inclusion of sardine samples collected outside the study area may complete the picture of the spatial organization of sardine populations through microsatellite loci characterization.

Inclusion of larval samples would contribute to assess structuring in highly connected areas.

Characterization of other type of molecular markers may contribute to further investigate the evolution and present structure of sardine populations. Evolutionary aspects could be studied sequencing mitochondrial DNA and influence of environmental factors through non neutral SNPs. Analysis of neutral SNPs would complete the picture concerning structuring of sardine genetic diversity in the area studied in the TRANSBORAN project.

REFERENCES

Atarhouch, T., Rüber, L., Gonzalez, E. G., Albert, E. M., Rami, M., Dakkak, A., et al. (2006). Signature of an early genetic bottleneck in a population of Moroccan sardines (Sardina pilchardus). *Mol. Phylogenet. Evol.* 39, 373–383. doi:10.1016/j.ympev.2005.08.003.

Baibai, T., Oukhattar, L., Quinteiro, J. V., Mesfioui, A., Rey-Mendez, M., and soukri, A. (2012). First global approach: Morphological and biological variability in a genetically homogeneous population of the European pilchard, Sardina pilchardus (Walbaum, 1792) in the North Atlantic coast. *Rev. Fish Biol. Fish.* 22, 63–80. doi:10.1007/s11160-011-9223-9.

Barton, N. H., and Slatkint, M. (1986). A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. *Heredity (Edinb)*. doi:10.1038/hdy.1986.63.

Chlaida, M., Kifani, S., Lenfant, P., and Ouragh, L. (2006). First approach for the identification of sardine populations Sardina pilchardus (Walbaum 1792) in the Moroccan Atlantic by allozymes. *Mar. Biol.* 149, 169–175. doi:10.1007/s00227-005-0185-0.

Chlaida, M., Laurent, V., Kifani, S., Benazzou, T., Jaziri, H., and Planes, S. (2009). Evidence of a genetic cline for Sardina pilchardus along the Northwest African coast. *ICES J. Mar. Sci.* 66, 264–271. doi:10.1093/icesjms/fsn206.

Cilli, E. (2009). Spatial and temporal population structure and species traceability in fishery resources of the Mediterranean Sea.

Cuéllar-Pinzón, J., Presa, P., Hawkins, S. J., and Pita, A. (2016). Genetic markers in marine fisheries: Types, tasks and trends. *Fish. Res.* 173, 194–205. doi:10.1016/j.fishres.2015.10.019.

Earl, D. A., and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi:10.1007/s12686-011-9548-7.

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* 14, 2611–2620. doi:10.1111/j.1365-294X.2005.02553.x.

Falush, D., Stephens, M., and Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164, 1567–1587. doi:10.1080/13235818.1996.10673675.

Food and Agriculture Organisation (FAO) (2021). Sardina pilchardus. Available at: http://www.fao.org/fishery/species/2910/en.

Gonzalez, E. G., and Zardoya, R. (2007a). Isolation and characterization of polymorphic microsatellites for the sardine Sardina pilchardus (Clupeiformes: Clupeidae): Primer note. *Mol. Ecol. Notes* 7, 519–921. doi:10.1111/j.1471-8286.2006.01640.x.

Gonzalez, E. G., and Zardoya, R. (2007b). Relative role of life-history traits and historical factors in shaping genetic population structure of sardines (Sardina pilchardus). *BMC Evol. Biol.* 7. doi:10.1186/1471-2148-7-197.

Holleley, C. E., and Geerts, P. G. (2009). Multiplex Manager 1.0: A cross-platform computer program that plans and optimizes multiplex PCR. *Biotechniques*. doi:10.2144/000113156.

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405. doi:10.1093/bioinformatics/btn129.

Kasapidis, P., Silva, A., Zampicinini, G., and Magoulas, A. (2012). Evidencia para selección hitchhiking de microsatélite en sardina europea (Sardina pilchardus) y las implicaciones para inferir la estructura del stock. *Sci. Mar.* 76, 123–132. doi:10.3989/scimar.03366.29B.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., and Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* 15, 1179–1191. doi:10.1111/1755-0998.12387.

Peakall, R., and Smouse, P. E. (2012). GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28, 2537–2539. doi:10.1093/bioinformatics/bts460.

Pérez, M., Presa, P. (2011). "FENOSALT: un método sintético para la extracción de ADN de peces y moluscos.," in *Métodos y Técnicas en investigación marina.*, ed. G. Garcia-

Estévez, J. M., Olabarria, C., Pérez, S., Rolán-Álvarez, E., Rosón (MADRID: Editorial Tecnos).

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.

R Team (2018). A Language and Environment for Statistical Computing. *R Found. Stat. Comput.* 2, https://www.R-project.org. Available at: http://www.r-project.org.

Ramon, M. M., and Castro, J. A. (1997). Genetic variation in natural stocks of Sardina pilchardus (sardines) from the western Mediterranean Sea. *Heredity (Edinb).* 78, 520–528. doi:10.1038/hdy.1997.81.

Raymond, M., and Rousset, F. (1995). GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. *J. Hered.* doi:10.1093/oxfordjournals.jhered.a111573.

Rousset, F. (2008). GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Mol. Ecol. Resour.* doi:10.1111/j.1471-8286.2007.01931.x.

RStudio Team (2019). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. URL http://www.rstudio.com/. *RStudio, Inc.*, 10. Available at: http://www.rstudio.com/.

Ruggeri, P., Splendiani, A., Bonanomi, S., Arneri, E., Cingolani, N., Santojanni, A., et al. (2012). Temporal genetic variation as revealed by a microsatellite analysis of european sardine (Sardina pilchardus) archived samples. *Can. J. Fish. Aquat. Sci.* 69, 1698–1709. doi:10.1139/F2012-092.

Ruggeri, P., Splendiani, A., Bonanomi, S., Arneri, E., Cingolani, N., Santojanni, A., et al. (2013). Estructura de stock de sardina pilchardus en el mar adriático mediante el análisis de marcadores microsatélites. *Sci. Mar.* 77, 565–574. doi:10.3989/scimar.03843.26A.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., and Shipley, P. (2004). MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes*. doi:10.1111/j.1471-8286.2004.00684.x.

Weir, B. S. (1996). Genetic Data Analysis II. By Bruce S. Weir, Sunderland, Massachusetts. Sinauer Associates, Inc.445 pages. ISBN 0-87893-902-4. *Genet. Res.* doi:10.1017/s0016672300034121.

Weir, B. S., and Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution (N. Y).* 38, 1358. doi:10.2307/2408641.

SUPPLEMENTARY INFORMATION.

Table S1. Summary information of the microsatellite markers designed for and tested "in vivo" in the *Sardina pilchardus* populations.

Microsat Loci	Repeat motif	Product size (bp)	Primer Sequences 5'-3'	Tm (°C)	Genbank Accession Number	Reference	Allele number	H_O / H_E	N of specimens / N of Locations	average N/location
SAR-9	(GT)17	183-273	F : AGGATGTGATGTCCATGAAGAAG R : ATTGCCTGCACTGAACA	55	EF012615	González et al 2007a	22	0,82/0,93	50/2	25
						González et al 2007b	48	0,85/0,93	433/9	48
						Ruggeri et al 2012			475/10	48
						Ruggeri et al 2013	40		377/7	54
SAR-1.5	(GT)11AT(GT)15	128–208	F : AGCTAAAAAGAAAACACACAG R:CCTTCATGACCCAAGGTGA	54	EF012616	González et al 2007a	22	0,84/0,94	50/2	25
						González et al 2007b	41	0,83/0,94	433/9	48
						Ruggeri et al 2012			475/10	48
						Ruggeri et al 2013	32		377/7	54
SAR-1.12	(GT)17	166–292	F : TGAGAATCACAGAATCTGAGCA R : CTGGAAGCTCTTGGCATCTT	54	EF012617	González et al 2007a	30	0,70/0,92	50/2	25
						González et al 2007b	61	0,66/0,94	433/9	48
		183-273		55		Ruggeri et al 2013	52		377/7	54

SAR-2.18	(GT)15	176–242	F : CTGCGTGTGAATGTAGTCTG R : CCATTTCTCACCATTTCTT	55	EF012618	González et al 2007a	24	0,80/0,93	50/2	25
						González et al 2007b	41	0,78/0,95	433/9*	48
SAR-A2F	(GT)48	185–269	F : GCAGTAGTGAACCATCCCCTA R : GACACACGCGTACCACCA	63	EF012619	González et al 2007a	31	0,90/0,96	50/2	25
						González et al 2007b	54	0,90/0,95	433/9	48
SAR-A3C	(GT)21	101–311	F : GGGGCTTTTCATTATTCATCAG R : CCTGTGCCCAAATGACACT	63	EF012620	González et al 2007a	45	0,66/0,98	50/2	25
						González et al 2007b	94	0,58/0,97	433/9	48
SAR-19B3	(GA)21	118–197	F : CGTTTGATCCCATCTGAAA R : TCTTGGCTCTTTGGCTTCT	58	EF012621	González et al 2007a	25	0,66/0,94	50/2	25
						González et al 2007b	60	0,56/0,96	433/9	48
SAR-19B5	(GT)48	147–337	F : AATGGCTTTCATCTGACCTTG R : TGGGACGACACCTTTAGTGAT	58	EF012622	González et al 2007a	42	0,79/0,97	50/2	25
						González et al 2007b	81	0,75/0,98	433/9	48
Sp22	(CA)3GA(CA)6 GA(CA)8	194-232	F : GGTCACGTTGGGTATTTGCTAT R : GTCTGGAACTGGGTGCTCTC	49	AY241284	Kasapidis et al 2012.	19	0,64/0,67	1540/21	73
		198-230		50		Thesis E. Cilli.	11		300/10	30

SpI5	(TATC)8TC(TATC)2	135-197	F :TGGCTGTGCATGTAAGTCTGT R :CAATTTTGCCAGTTGTTTAGTC	49	HM031962	Kasapidis et al 2012	31	0,76/0,88	1540/21	73
		117-193		56		Baibai et al 2012	12	0,67/0,85	384/11	35
		123-183	F: TGGCCTGTGATCTACAGTATGG R: CCTTTTGATAGCCCTGACACA	55		Ruggeri et al 2012	39		475/10	48
						Ruggeri et al 2013			377/7	54
Sp2	(AG)8/(TG)16	138-254	F : CGAGGCCTGATAGAAACCC R : AACCACGGTCAGTTCTCCAG	56	AJ639616	Kasapidis et al 2012	58	0,93/0,95	1540/21	73
		146-238		59		Thesis E. Cilli 2009	31		300/10	30
		122-274		54		Baibai et al 2012	26	0,03/0,96	384/11	35
Sp7	(AC)14	124-212	F : GCACAGGCGCTTACACAC R : TGTGACACCAGGCAGAGC	62	AJ639618	Kasapidis et al 2012	41	0,93/0,92	1540/21	73
	(CA)14	125-169		62		Thesis E. Cilli 2009	18		300/10	30
	(AC)14	112-262		61		Baibai et al 2012	17	0,80/0,90	384/11	35
Sp8	(CA)22	109-211	F : ACGTCACAGTTCCCCACTG R : ACTGGCTGAGGAGGATGATG	62	AJ639619	Kasapidis et al 2012	60	0,93/0,94	1540/21	73
		110-160		62		Thesis E. Cilli 2009	23		300/10	30

		111-327		63		Baibai et al 2012	24	0,83/0,94	384/11	35
Sp10	(GT)10/ (GT)11	124-194	F :ACCAGGGTGGGTGAGATTAC R :TGGAAACCCTCTGGACACAT	56	AY241279	Thesis E. Cilli 2009	21		300/10	30
	(TG)16	148-248	F: GCAAAAGTGCTCGAAGACG R: CGCTTTTGTTGGCTAAAACAT	58		Ruggeri et al 2012			475/10	48
						Ruggeri et al 2013	38		377/7	54
Sp45	(CA)20	160-216	F : TAGGAATTGAGCCATGACCTTT R :TGAAAACACCACTCATCAGGAC		AY241288	Thesis E. Cilli 2009	18		300/10	30
Sp5	(AC)17	84-90	F : ATAAGAGTTACGGGCCACCC R :TATGCCTACTGGCCTCAATG	57	AJ639617	Thesis E. Cilli 2009	33		300/10	30
SAR1D06	(TG)18	120-158	F: CGGCTATTCTTAGACTAGGTG R: CCCCATCAGCAATGAATAAG	50	AY636123	Ruggeri et al 2012			475/10	48
						Ruggeri et al 2013	23		377/7	54
SpI7	(AGAT)8	117-141	F: TGCTTTACTTCATTCCGTTGAA R: TCACATCATCACAACAAACACC	52	HM031963	Ruggeri et al 2012	9		475/10	48
						Ruggeri et al 2013	12		377/7	54
SpIII93	(ATCT)9	170-292	F:TAAGCAGACGCGAAACTGAA R:CTTGCGACCTGACGTGATTA	58	HM031964	Ruggeri et al 2013	52		377/7	54
SARB-A07	(GA)12	68-136	F:CTCCTCACTCAGCCGCTAAGGA R:GGGTAACATTTCGGCAAGTGCT		AY636114	Ruggeri et al 2013	39		377/7	54