Molecular Ecology Notes

Final version MEN 06/467: October 18th, 2006

PRIMER NOTE

**Isolation and characterization of polymorphic
microsatellites for the sardine, *Sardina pilchardus*
(Clupeiformes: Clupeidae)**

E G. GONZALEZ* and R. ZARDOYA*

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

Corresponding author:
Elena Guacimara González
Museo Nacional de Ciencias Naturales, CSIC
José Gutiérrez Abascal, 2
28006 Madrid
Spain
Tel: +34-91-4111328 ext 1130
Fax: +34-91-5645078
Email: elenag@mncn.csic.es

Keywords: Clupeidae, *Sardina pilchardus*, microsatellite, marine pelagic fish
Running title: Novel microsatellites for *Sardina pilchardus*

Abstract

The sardine (*Sardina pilchardus*, Walbaum 1792) is a small pelagic fish species that sustains an economically important fishery in the eastern Atlantic Ocean. Eight perfect microsatellite loci of sardine were characterized from a partial genomic library enriched for CA repeat sequences. All loci revealed high levels of polymorphism, with the number of alleles per locus ranging between 22-45, and an average observed heterozygosity of 0.772. The newly described loci can be employed to determine population genetic structure of the sardine, and may aid in fishery management of this species.

1 The European sardine is a small pelagic fish, which inhabits the Mediterranean Sea,
 2 and the northeastern Atlantic Ocean from the North Sea to Senegal (Parrish *et al.*,
 3 1989), and has traditionally sustained economically important fisheries, particularly in
 4 Morocco and Spain. Two different subspecies has been recognized based on
 5 morphological characters (Parrish *et al.*, 1989) and further supported by mitochondrial
 6 control region sequence data (Atarhouch *et al.*, 2006). *S. pilchardus pilchardus* is
 7 found in the northern Atlantic Ocean whereas *S. pilchardus sardine* is found in the
 8 Mediterranean Sea, and off the Atlantic Ocean coast of Morocco. The validity of
 9 these subspecies awaits further confirmation from nuclear data.

10 Here, we present the isolation and characterization of eight polymorphic
 11 microsatellites loci that should represent a useful molecular tool for the study of
 12 sardine genetic variability within its distribution both at fine and large geographical
 13 scales. These nuclear markers will allow contrasting phylogeographic patterns
 14 previously inferred based on mitochondrial data (Atarhouch *et al.*, 2006), confirming
 15 the validity of the two recognized subspecies, and delimiting fishery stock boundaries
 16 and putative instances of overexploitation.

17 An enriched partial genomic DNA library was generated from a single specimen of
 18 *S. pilchardus* following Ostrander *et al.* (1992) and Hamilton *et al.* (1999). Genomic
 19 DNA was digested with *Rsa* I and *Xmn* I (New England Biolabs), and ligated to the
 20 SNX adaptor (Hamilton *et al.*, 1999). The digested-ligated DNA was enriched for the
 21 CA motif using 100 pmol of biotinylated (CA)₁₂ probe bound to streptavidin magnetic
 22 beads. Enriched DNA was then ligated into a pGEM-T vector (Promega), and
 23 transformed into *E. coli* XL-10 Gold ultra competent cells (Stratagene). Approximately
 24 800 white recombinants were transferred onto Hybond-N+ nylon membrane
 25 (Amersham), screened with a biotin-labelled probe (CA)₁₂, and detected by a
 26 chemiluminescent method (CDP-Star detection kit, Sigma). A total of 150 positive
 27 colonies were selected, and re-screened by performing a PCR amplification using the
 28 SNX and (CA)₁₂ unlabelled probe as primers. Inserts of 66 recaptured positive clones
 29 (average length= 450 bp $\pm$ 127) were sequenced on an ABI Prism 3700 DNA Analyzer
 30 (Applied Biosystems) using the ABI Prism BigDye Terminator Cycle Sequencing
 31 Ready Reaction kit (V3.0) (Applied Biosystems), and following manufacturer's
 32 instructions. Half of the positive clones were excluded for primer designing because
 33 they contained either long repetitions (not appropriate for genotyping), compound
 34 motifs or minisatellite repeat sequences. Of the remaining half (GenBank accession
 35 numbers EF012615-EF012648), primers were designed for nine loci using PRIMER3
 36 software (Rozen, Skaletsky, 1998).

37 Tests for polymorphism were carried out on two populations collected off the
 38 coasts of northern Spain (Pasajes de San Pedro, Cantabric Sea; N = 25) and
 39 southwestern Morocco (Dakhla, Atlantic Ocean; N = 25). PCR conditions were
 40 optimized for nine microsatellite loci. PCR amplification consisted of 35 cycles of

denaturation at 95°C for 45 s, annealing at 54-63°C for 30s, and extension at 72°C for 30s. Cycles were followed by a final extension at 72°C for 20 min. PCR reactions contained approximately 5 ng of sample DNA, 0.8 U of *Taq* DNA polymerase (Eppendorf), 0.5 µM of each primer, 0.2 mM of each dNTP, and 1x*Taq* buffer (10 mM Tris-HCl, pH 8.3, 50mM KCl, and 1.5 mM MgCl₂) in a total volume of 10 µl.

Forward primers were labelled with fluorescent dyes for genotyping. PCR amplified products were run on an ABI Prism 3700 DNA Analyzer (250-500 LIZ size standard; Applied Biosystems), and allele scoring was performed using GeneMapper v3.7 (Applied Biosystems). For each locus, number of alleles (N_A), as well as observed (H_o) and expected heterozygosities (H_e) (Nei, 1987) were calculated using GENETIX 4.02 (Belkir *et al.*, 2000). Similarly, F_{IS} statistic estimations that detect deviations from Hardy-Weinberg equilibrium (HWE), and the linkage disequilibrium test were performed for each locus using the program GENEPOP version3.3 (Raymond, Rousset, 1995). Significance of both analyses was tested with the specified Markov chain Montecarlo method of Guo and Thompson (1992) that was run in 1000 batches of 2000 iterations each with the first 500 iterations discarded before sampling. P values from multiple comparisons were corrected by applying a Bonferroni correction (Rice, 1989).

One (SAR2.24) out of nine loci resulted monomorphic, and the rest showed high levels of polymorphism with the number of identified alleles ranging from 22 to 45 per locus (average = 30). No evidence of linkage disequilibrium was found for any of the corresponding exact tests. The amount of genetic variability, in terms of observed heterozygosity, was high (total average per locus, $H_T = 0.772$, ranging from $H_o=0.660$ to $H_o=0.900$) (Table 1). After Bonferroni correction, six F_{IS} estimates remained significant, which indicates their departure from HWE due to heterozygote excess. This result may be due to a Wahlund effect as expected when pooling samples of individuals from different populations. Analyzing the two population samples separately with the program MICRO-CHECKER (Van Oosterhout *et al.*, 2004) resulted in the detection of HWE disequilibrium due to null alleles in four loci (SAR1.12, SARA3C, SAR19B3 y SAR19B5) only at the Dakhla sampling site.

Acknowledgments

We thank M.J. San Sebastián and T. Atarhouch for kindly providing sardines from Pasajes de San Pedro and Dakhla, respectively. E.G.G. is sponsored by a predoctoral fellowship of the Ministerio de Educación y Ciencia (MEC). This work received financial support from a project of the MEC to R. Z. (REN 2001-1514/GLO).

References

- Atarhouch T, Ruber L, Gonzalez EG, *et al.* (2006) Signature of an early genetic bottleneck in a population of Moroccan sardines (*Sardina pilchardus*). *Mol Phylogenet Evol* **39**, 373-383.
- Belkir K, Borsa P, Chickhi L, Raufaste N, Bonhomme F (2000) *GENETIX 4.04 Logici el sous Windows TM, pour la Génétique des Populations* Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier, France.
- Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques* **20**, 1004-1006, 1008-1010.
- Hamilton MB, Pincus EL, Fleischer RC (1999) Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *Biotechniques* **27**, 500-515.
- Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Parrish RH, Serra R, Grant WS (1989) The Monotypic Sardines, *Sardina* and *Sardinops* - Their Taxonomy, Distribution, Stock Structure, and Zoogeography. *Canadian Journal of Fisheries and Aquatic Sciences* **46**, 2019-2036.
- Raymond M, Rousset F (1995) GENEPOP 3.3: population genetic software for exact test and ecumenism. *J. Hered.* **86**, 248-249.
- Rice WR (1989) Analyzing tables of statistical tests. *Evolution* **43**, 223-225.
- Rozen S, Skaletsky HJ (1998) PRIMER3. Whitehead Institute for Biomedical Research. Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html.
- Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535-538.

Table 1 Summary data for microsatellite developed from sardine (*Sardina pilchardus*)*

Locus	Repeat motif	T ^a (°C)	Dye	Primers sequence 5'-3'	N	N _A	Allele range	H _O	H _E	F _{IS}	GenBank No.
SAR9	(GT) ₁₇	55	PET	F: AGGATGTGATGTCCATGAAGAAG R: † <u>GTTCTT</u> ATTGCCTGCACTGAACA	50	22	183-273	0.82	0.93	0.12	EF012615
SAR1.5	(GT) ₁₁ AT(GT) ₁₅	54	VIC	F: AGCTAAAAAGAAAACACACAG R: † <u>GTTTCTT</u> CCCTTCATGACCCAAGGTGA	50	22	128-208	0.84	0.94	0.11	EF012616
SAR1.12	(GT) ₁₇	54	6-FAM	F: TGAGAATCACAGAATCTGAGCA R: † <u>GTTTCTT</u> CTGGAAGCTCTTGGCATCTT	50	30	166-292	0.70	0.92	0.24	EF012617
SAR2.18	(GT) ₁₅	55	6-FAM	F: CTGCGTGTGAATGTAGTCTG R: † <u>GTTTCTT</u> CCATTTCTCACCATTCTT	50	24	176-242	0.80	0.93	0.14	EF012618
SARA2F	(GT) ₄₈	63	VIC	F: GCAGTAGTGAACCATCCCCTA R: † <u>GTTTCTT</u> GACACACGCGTACCACCA	50	31	185-269	0.90	0.96	0.06	EF012619
SARA3C	(GT) ₂₁	63	6-FAM	F: GGGGCTTTTCATTATTCATCAG R: † <u>GTTTCTT</u> CCCTGTGCCCAAATGACACT	50	45	101-311	0.66	0.98	0.33	EF012620
SAR19B3	(GA) ₂₁	58	NED	F: CGTTTGATCCCATCTGAAA R: TCTTGGCTCTTTGGCTTCT	50	25	118-197	0.66	0.94	0.30	EF012621
SAR19B5	(GT) ₄₈	58	VIC	F: AATGGCTTTCTCTGACCTTG R: TGGGACGACACCTTTAGTGAT	50	42	147-337	0.79	0.97	0.19	EF012622

* N= number of individuals assayed, N_A= number of alleles per locus, expected (H_E) and observed (H_O) heterozygosities, F_{IS}= Wright's statistics. Bold F_{IS} values are significant probability estimates after Bonferroni correction (Rice 1989). † Underlined bases were added to 5' end of the reverse primers to promote adenylation by *Taq* DNA polymerase