

Short Communication

Development and characterization of thirty-two microsatellite markers for the anchovy, *Engraulis ringens* Jenyns, 1842 (Clupeiformes, Engraulidae) via 454 pyrosequencing

**Sandra Ferrada-Fuentes^{1,2}, Ricardo Galleguillos¹
Cristian B. Canales-Aguirre^{3,4} & Victoria Herrera-Yañez^{1,5}**

¹Laboratorio de Genética y Acuicultura, Departamento de Oceanografía
Universidad de Concepción, Concepción, Chile

²Programa de Doctorado en Sistemática y Biodiversidad, Universidad de Concepción
Concepción, Chile

³Centro i~mar, Universidad de Los Lagos, Puerto Montt, Chile

⁴Núcleo Milenio de Salmónidos Invasores (INVASAL), Concepción, Chile

⁵Programa de Magister en Ciencias m/ Zoología, Universidad de Concepción
Concepción, Chile

Corresponding author: Cristian B. Canales-Aguirre (cristian.canales@ulagos.cl)

ABSTRACT. The anchovy, *Engraulis ringens*, is an economically valuable fish, and the most heavily exploited resource in the Humboldt Current System. To analyze its genetic structure and diversity, microsatellite markers were developed using 454 pyrosequencing. A total of 27,352 reads containing di-, tri-, tetra-, penta-, and hexanucleotide microsatellite repeat units were identified from 136,537 reads. Among 80 loci containing more than six repeat motifs, 32 primer sets (40%) produced reproducible PCR products, and all of these loci were polymorphic. Some loci showed deviations from *HWE* and possible null allele's presence, results of an excess of homozygotes. In an analysis of 45 individuals from one *E. ringens* population, the number of alleles per locus ranged from 2 to 33, observed heterozygosity ranged from 0.171 to 0.976, and the probability of identity values ranged from 0.006 to 0.513. These microsatellites will be useful for numerous ecological studies focused on this important pelagic fish; including the examination of population genetic structure, estimating effective population size and providing information for fisheries management.

Keywords: *Engraulis ringens*, Peruvian anchovy, genetic diversity, Humboldt Current System, Chile.

Microsatellite loci are widely used to understand how the genetic diversity of marine organisms is distributed along the geographic space. In addition, this information can be useful for management of marine resources (Canales-Aguirre *et al.*, 2010a, 2010b, 2016; Galleguillos *et al.*, 2011, 2012; Ferrada-Fuentes *et al.*, 2014). Traditional Short Sequence Repeat (SSR) development is time-consuming and involves laborious iterations of genomic DNA library screening with SSR probes required to isolate microsatellite-containing sequences (Castoe *et al.*, 2012). Next-generation sequencing technologies are remarkably well-developed and are widely used for genome sequencing, transcriptome sequencing, and genome deep-sequencing in animals (Ferrada-Fuentes *et al.*, 2014; Plough & Marko, 2014). The use of new technologies, like Illumina, 454 or SOLiD sequencing, to create and

characterize microsatellite loci, quickly obtaining a large number of microsatellite loci without the need to clone a library (Castoe *et al.*, 2012).

The anchovy, *Engraulis ringens* Jenyns, 1842, is a small pelagic fish exploited for its high commercial value in the Humboldt Current System. This species sustains the world's largest single-species fishery with 6.5 million ton landed per year on average over the last decade (Bertrand *et al.*, 2008). The anchovy plays a key ecological role in the Humboldt Current System because it is the major prey of predators such as fish, marine mammals and seabirds (Espinoza & Bertrand, 2008). Thus, the anchovy is essential for the maintenance of the integrity of this ecosystem (Espinoza & Bertrand, 2008). To date, little is known about their genetic diversity and their spatial genetic structure. Only two polymorphic genes (*i.e.*, Calmodulin and

Internal Transcribed Spacers, ITS1) have been used, however, these showed low polymorphisms at the population level (Ferrada *et al.*, 2002). To contribute to filling this gap in knowledge (looking for more polymorphic molecular markers), in this study we report the isolation and characterization of 32 polymorphic loci for *E. ringens*. These microsatellite loci were developed as a tool for estimating genetic diversity and population genetic structure in this species, with the goal of providing baseline information for management plans aimed at their protection.

The samples used in this study were collected in accordance with the national legislation of the country (Chile). We did not kill fishes for the purpose of this study. We obtained tissue samples after samples were fished from authorized purse seine fishing commercial vessels. No specific approval is required for this vertebrate. Total genomic DNA was extracted from four individuals of *Engraulis ringens* collected in the Arauco Gulf, Chile (36°55.2'S, 73°22.8'W), using Nucleospin Tissue Kit (Machery and Nagel), following manufacturer protocol. DNA samples were checked with the Bioanalyzer Agilent Model 2100. The enriched library was built using a range of 500 ng to 1 µg of DNA, the GS Rapid Library Preparation kit and a single lane run on a Roche 454 GS Junior system were used to sequence a part of the genome. The NGS was performed at OMICS Solutions (<http://omics-solutions.cl>). Sequencing generated a total of 80.7 Mb of quality-filtered data, corresponding to 136,537 non-redundant reads. The MISA 4.0 software (<http://pgrc.ipk-gatersleben.de/misa/>) was used to search for repeated motifs (di-, tri-, tetra-, penta-, and hexanucleotide), and primers were designed using Primer3 (<http://bioinfo.ut.ee/primer3-0.4/>). A total of 27,352 reads with microsatellites were detected, resulting in 13,211 reads with primers.

Genomic DNA from 45 individuals was used to test 80 microsatellites. Polymerase chain reaction (PCR) amplifications were performed in a 10 µL volume (10 mM Tris pH 8.4, 50 mM KCl, 25 µg mL⁻¹ BSA, 0.4 µM unlabeled reverse primer, 0.04 µM of forward primer (fluorescently labelled 6-FAM, NED, PET or VIC), 3 mM MgCl₂, 0.8 mM dNTPs, 0.5 units Taq Polymerase (Invitrogen), and 20 ng DNA template using an Applied Biosystems GeneAmp 9700. A touchdown thermal cycling program (Don *et al.*, 1991) encompassing a 10°C span of annealing temperatures ranging between 65-55°C was used for all loci. Touchdown cycling parameters consisted of an initial denaturation step of 5 min at 95°C followed by 20 cycles of 95°C for 30 s, highest annealing temperature of 65°C (decreased 0.5°C per cycle) for 30 s, and 72°C for 30 s; and 20 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 30 s, with a final extension at 72°C for 5 min. PCR

products were run on an ABI-3130xl sequencer using GS-500 (LIZ) as an internal size standard. Microsatellites were analyzed using Peak Scanner v1.0 (Applied Biosystems). Unambiguous scoring was possible for 32 polymorphic loci. We assessed the genetic diversity of the 32 polymorphic loci in 45 specimens collected from the Arauco Gulf, Chile.

Characteristics of the loci are provided in Table 1. We estimated the number of alleles per locus (N_A), observed and expected heterozygosity (H_O and H_E), and the probability of identity (P_I) using GenAIEx v6.5 (Peakall & Smouse, 2012). The presence of null alleles was evaluated using MicroChecker v.2.2.3 (Van Oosterhout *et al.*, 2004). Tests for deviations from Hardy-Weinberg equilibrium (HWE) and for linkage disequilibrium were conducted using Genepop v4.0 (Rousset, 2008). Parameters for linkage disequilibrium were as follows: 1000 dememorization steps, 100 batches, and 1000 iterations per batch.

Thirty-two of the tested primer pairs amplified high-quality PCR product exhibiting polymorphism. We observed a low percentage of amplification failure (maximum of 26% 44µER: Table 1) among the samples used in this study.

The number of alleles per locus ranged from 2 (50µER) to 33 (53µER), observed heterozygosity ranged from 0.171 (38µER) to 0.976 (50µER), and the probability of identity values ranged from 0.006 (53µER) to 0.513 (38µER). No linkage disequilibrium was found between any pair of loci, indicating that the markers were independent.

Twelve loci showed moderate polymorphisms (2-9 alleles) and 20 loci were highly polymorphic (10-33 alleles). The number of alleles and observed heterozygosity obtained in *E. ringens* was higher than previously reported in marine fishes ($N_A = 20$ and $H_O = 0.790$) (DeWoody & Avise, 2000). High N_A and H_O are common in species that have a large population size and/or high mutation rate (higher than 10⁻⁴ mutations/gene/generation) (Jarne & Lagoda, 1996). After Bonferroni correction for multiple comparisons, 12 loci showed significant deviations from expectations under HWE . There are several explanations for the deficits of heterozygosity in these 12 loci. Rico *et al.* (1997) reviewed the possible causes of excess homozygosity, evaluating various hypotheses concerning genotyping errors, the presence of null alleles, the Wahlund effect (Wahlund, 1927) inbreeding, assortative mating, and/or selection. Deviations from HWE associated with an excess of homozygotes seem to be common in fish species (O'Connell & Wright, 1997). Moreover, the occurrence of null alleles has been commonly reported during the characterization of microsatellite loci and in population genetics studies (Dakin & Avise, 2004).

Table 1. Characterization of 32 polymorphic microsatellite loci developed for *Engraulis ringens* from central Chile. Size range: indicates the range of observed alleles in base pairs and not includes primer; N_i : number of individuals genotyped; N_A : number of alleles observed; H_o and H_E : observed and expected heterozygosity, respectively; P_I : probability of identity for each locus. HWE : exact P -values of Hardy-Weinberg Equilibrium test; †: significant deviations from Hardy-Weinberg expectations after Bonferroni corrections (corrected $\alpha = 0.001$). Loci with the possible presence of null alleles are in bold.

Locus (accession no)	Primer pair sequence (5' - 3')	Dye	Repeat motif	Size range (bp)	N_i	N_A	H_o	H_E	P_I	P -value (HWE)
6 μ ER (KY073503)	F: TGGGTTGATAAATAGACTAGA R: AGTATTAACACTTGTAGGTGC	6-FAM	(ATGA)5	175-218	35	6	0.771	0.677	0.150	0.913
7 μ ER (KY073509)	F: GGATGATATTTCTCACTTTG R: GTTTTTACACTCTAAATGTC	VIC	(CACT)5	226-255	37	4	0.297	0.408	0.414	0.004
9 μ ER (KY073517)	F: GATAAAAAGCACTGCTGTATT R: ACTAATGAATGTTAAGCAGTC	NED	(GTCT)5	257-320	37	8	0.730	0.787	0.078	0.444
11μER (KY073523)	F: GTC AAGGAAAACAGTTTATT R: AGAGCACAATAGAAGITGATA	PET	(TATT)5	168-273	38	6	0.474	0.667	0.144	0.009
17μER (KY073515)	F: TTAGTATATGGGTATGIGTCC R: C AAGATTCACTATGTAAGC	6-FAM	(TGA)7	155-218	35	14	0.200	0.875	0.027	0.000†
18μER (KY073496)	F: AAACACTACACTCATGAAGTGC R: GAGTCTACATGTGTAAGTTCG	VIC	(AC)12	113-255	44	20	0.750	0.905	0.016	0.000†
21μER (KY073495)	F: ATGTACAACCTCCAAAATCT R: ATTACTGGTATGAAATGAGTG	PET	(CTA)8	117-229	41	18	0.366	0.911	0.014	0.510
22 μ ER (KY073505)	F: GTGTGTATGTTTCTTTTCAA R: TCTCTATGGGACTTTAACATA	6-FAM	(CAGA)6	273-296	43	6	0.674	0.659	0.166	0.003
27 μ ER (KY073519)	F: GCTTCTGGATGTTTTAGAT R: AACACTATCTGACAAGTACACA	VIC	(TGTC)6	149-257	44	4	0.455	0.513	0.347	0.833
28 μ ER (KY073499)	F: CATGGTTTTAAATCTGTGAC R: TACAAATGAGAGCAAATACA	NED	(TTTA)6	239-297	41	12	0.854	0.857	0.036	0.972
25 μ ER (KY073504)	F: CATCAAAGTTATTCACCTCAC R: GAATTCTTCTAACTGACACT	PET	(TCAT)6	145-287	35	7	0.886	0.646	0.180	0.310
44μER (KY073512)	F: TACACAAGACGTTTCAGAGT R: TACATACAACCTTGGAGACT	6-FAM	(TGG)9	129-160	33	11	0.273	0.884	0.025	0.000†
36μER (KY073518)	F: GGTATATTTGATGGCATT R: CAGAGTACTCTTTGAGTGTG	VIC	(TGGTG)5	142-194	39	7	0.205	0.464	0.308	0.000 †
38μER (KY073510)	F: GACAAGAGATTAACATTACCA R: AATAACTGTAAGTCGCTTTG	NED	(TTACA)5	147-158	41	3	0.171	0.298	0.513	0.003
35 μ ER (KY073521)	F: CTCAGTGAACAAGTCACT R: TAAATACATGCTTAAGAGTCC	PET	(GCAAG)5	147-174	40	4	0.425	0.457	0.332	0.060
75μER (KY073497)	F: CTGTAATATCCACTCAAAGAT R: TTTTTCACAGTATAATGCTG	6-FAM	(ACACAG)9	172-358	39	16	0.359	0.923	0.014	0.000†
45 μ ER (KY073502)	F: ATAAAAGTTGAGGCTGTTT R: GACTTTGAAGACAGCTGTAC	VIC	(ATTC)7	162-203	44	9	0.795	0.765	0.088	0.460
48 μ ER (KY073520)	F: CCAATAGTTCAGTACCAG R: ACAGTAAGCTAGAGTATCCAG	NED	(GCT)10	136-175	43	12	0.860	0.832	0.048	0.995
39μER (KY073522)	F: GTTGTGAGAAGCTTTAGTCA R: GATAAGAAATACACAGGAAAG	PET	(TTCC)5	143-193	44	10	0.636	0.829	0.048	0.032
50 μ ER (KY073514)	F: GTCAGGGGTGTAATAATAA R: TCCACTTCTATTATGTTATGG	6-FAM	(ATCGT)6	157-163	42	2	0.976	0.500	0.375	0.000†
53μER (KY073498)	F: TGTAGAAGAAAGTGACAGAGA R: GTTTATTGGTGTGAGTCATT	VIC	(GA)17	136-235	45	33	0.822	0.943	0.006	0.033
61μER (KY073501)	F: GAGATTTAGCAGACAGATGT R: TAGTATTCTAACACGTAGCT	NED	(AG)20	190-226	45	11	0.756	0.880	0.027	0.258
49 μ ER (KY073508)	F: ATCAATATGCTAATGCTGTC R: GGACTTTTTAGCATCAGTAT	PET	(ACAGC)6	126-171	44	10	0.773	0.830	0.050	0.427
65μER (KY073511)	F: TCTGTGACTTCTGTAAGTACAG R: GTGTGTGAGGTTTAGATGTG	VIC	(GT)21	139-252	43	18	0.721	0.880	0.026	0.000†
67μER (KY073525)	F: GTTCATTAATAAGCAGAAGAG R: TTCACCAAGATATTACTCACT	PET	(CACGCA)6	232-297	42	14	0.690	0.822	0.048	0.001
63μER (KY073524)	F: TTCATTATCACAGCTAGTAGC R: GAACGTATAAAGAGGAGAGAT	PET	(CTTCT)8	118-208	42	14	0.643	0.848	0.039	0.000†
72μER (KY073506)	F: TTTTCTTACATTAGCACAG R: AATTTGTAGTACAGCTGTGTC	NED	(ACACAT)5	187-308	45	11	0.444	0.660	0.174	0.001
73 μ ER (KY073516)	F: ACTTTGAGTCTGGAATAAAGT R: CCATAGATTAGAGGACAATAA	NED	(TGA)17	106-185	44	20	0.932	0.915	0.014	0.678
69μER (KY073513)	F: GGCCTACATTAATAACATACT R: CTAATGTGGGAATATAGTGAG	PET	(TAA)15	122-174	45	18	0.756	0.915	0.013	0.000†
26μER (KY073500)	F: ACTACAGTAACTCATGATGG R: TGGAAATACAGTAGAGTAGGTG	NED	(TGGA)6	133-229	39	15	0.333	0.842	0.040	0.000†
75μER (KY073497)	F: CTGTAATATCCACTCAAAGAT R: TTTTTCACAGTATAATGCTG	6-FAM	(TTAGGG)5	223-323	39	16	0.359	0.911	0.014	0.000†
24μER (KY073494)	F: ATAGTAGGCCACACTCACTC R: ATGACATCATTGTGAGAAGT	VIC	(TCAC)6	201-250	44	10	0.500	0.819	0.056	0.000†

Therefore, the low occurrence of amplification failure in these loci may indicate that a significant excess of homozygotes may be a common characteristic in fish populations resulting from biological phenomena instead of the presence of null alleles. We recommend evaluating these hypotheses before using disequilibrium loci in population studies.

These new loci will provide tools for examining population structure of this species and aid in better understanding of the ecology and conservation of *E. ringens*. This information will help determine appropriate fisheries management action along the Humboldt Current System coast for this highly exploited species.

ACKNOWLEDGMENTS

This work forms part of the FIPA 2015-22, and DIUC 212.113.083-1.0. Cristian B. Canales-Aguirre is supported by Nucleo Milenio INVASAL funded by Chile's government program, Iniciativa Científica Milenio from Ministerio de Economía, Fomento y Turismo.

REFERENCES

- Bertrand, S., E. Díaz & M. Lengaigne. 2008. Patterns in the spatial distribution of Peruvian anchovy (*Engraulis ringens*) revealed by spatially explicit fishing data. *Prog. Oceanogr.*, 79: 379-389.
- Canales-Aguirre, C.B., S. Ferrada & R. Galleguillos. 2010a. Isolation and characterization of microsatellite loci for the jack mackerel (*Trachurus murphyi* Nichols, 1920). *Conserv. Genet.*, 11: 1235-1237.
- Canales-Aguirre, C.B., S. Ferrada, C.E. Hernández & R. Galleguillos. 2010b. Population structure and demographic history of *Gerypteris blacodes* using microsatellite loci. *Fish. Res.*, 106: 102-106.
- Canales-Aguirre, C.B., S. Ferrada-Fuentes, R. Galleguillos & C.E. Hernández. 2016. Genetic structure in a small pelagic fish coincides with a marine protected area: seascape genetics in Patagonian Fjords. *PLoS ONE*, 11: e0160670.
- Castoe, T.A., A.W. Poole, A.P.J. de Koning, K.L. Jones, D.F. Tomback, S.J. Oyler-McCance, J.A. Fike, S.L. Lance, J.W. Streicher, E.N. Smith & D.D. Pollock. 2012. Rapid microsatellite identification from Illumina paired-end genomic sequencing in two birds and a snake. *PLoS ONE*, 7: e30953.
- Dakin, E.E. & J.C. Avise. 2004. Microsatellite null alleles in parentage analysis. *Heredity*, 93: 504-509.
- DeWoody, J.A. & J.C. Avise. 2000. Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *J. Fish. Biol.*, 56: 461-473.
- Don, R.H., P.T. Cox, B.J. Wainwright, K. Baker & J.S. Mattick. 1991. Touchdown PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.*, 19(4): 4008.
- Espinoza, P. & A. Bertrand. 2008. Revisiting Peruvian anchovy (*Engraulis ringens*) trophodynamics provides a new vision of the Humboldt Current System. *Prog. Oceanogr.*, 79: 215-227.
- Ferrada, S., K. Hernández, R. Montoya & R. Galleguillos. 2002. Estudio poblacional del recurso anchoveta (*Engraulis ringens* Jenyns, 1842) (Clupeiformes, Engraulidae), mediante análisis de ADN. *Gayana.*, 66: 243-248.
- Ferrada-Fuentes, S., R. Galleguillos, C.B. Canales-Aguirre, C.N. Love, K.L. Jones & S.L. Lance. 2014. Development and characterization of thirty-three microsatellite markers for the Patagonian sprat, *Sprattus fuegensis* (Jenyns, 1842), using paired-end Illumina shotgun sequencing. *Conserv. Genet. Resour.*, 6: 833-836.
- Galleguillos, R., C.B. Canales-Aguirre & S. Ferrada. 2012. Genetic variability in jack mackerel *Trachurus murphyi* Nichols: new SSRs loci and application. *Gayana*, 76: 67-71.
- Galleguillos, R., C.B. Canales-Aguirre, S. Ferrada & A. Barrera. 2011. SSRs in *Octopus mimus*: development and characterization of nine microsatellite loci. *Rev. Biol. Mar. Oceanogr.*, 46: 491-494.
- Jarne, P. & P.J. Lagoda. 1996. Microsatellites, from molecules to populations and back. *Trends Ecol. Evol.*, 11: 424-429.
- O'Connell, M. & J.M. Wright. 1997. Microsatellite DNA in Fishes. *Rev. Fish. Biol. Fish.*, 7: 331-363.
- Peakall, R. & P.E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*, 28: 2537-2539.
- Plough, L.V. & P.B. Marko. 2014. Characterization of microsatellite loci and repeat density in the gooseneck barnacle, *Pollicipes elegans*, using next-generation sequencing. *J. Hered.*, 105: 136-142.
- Rico, C., K.M. Ibrahim, I. Rico & G.M. Hewitt. 1997. Stock composition in North Atlantic populations of whiting using microsatellite markers. *J. Fish. Biol.*, 51: 462-475.
- Rousset, F. 2008. Genepop'007: a complete re-implementation of the Genepop software for Windows and Linux. *Mol. Ecol. Resour.*, 8: 103-106.
- Van Oosterhout, C., W.F. Hutchinson, D.P.M. Wills & P. Shipley. 2004. Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes*, 4: 535-538.

Wahlund, V. 1927. Zusammensetzung von populationen und kor relation serscheinungen vomstand punkt der vererbungslehreausbetrachtet. Hereditas, 11: 65-106.

Received: 31 March 2017; Accepted: 21 June 2017