Development of gene-associated single nucleotide polymorphisms for Japanese anchovy *Engraulis japonicus* through cross-species amplification

Iratxe Montes¹ · Mikel Iriondo¹ · Carmen Manzano¹ · Andone Estonba¹

Received: 8 February 2017 / Accepted: 13 July 2017 / Published online: 14 November 2017
© The Author(s) 2017. This article is an open access publication

Abstract Cross-species amplification is a fast and cost-effective approach for the discovery of genetic markers in non-model species that have a phylogenetically close species with developed markers. The present study employed this method to discover new single nucleotide polymorphism (SNP) markers in the Japanese anchovy *Engraulis japonicus*. We selected 482 mitochondrial and nuclear SNPs described in the European anchovy *Engraulis encrasicolus*; primer and probe sets developed for this species were used for genotyping 35 individuals of *E. japonicus*. A total of 451 SNPs (93.6%) were successfully cross-amplified, of which 176 SNPs (36.5%) were polymorphic in *E. japonicus*. The mean expected heterozygosity was 0.102, and no marker deviated significantly from Hardy–Weinberg proportions. Two mitochondrial SNPs were identified as a set of species diagnosing markers since, compared to four other species from the genus *Engraulis*, they formed a haplotype only present in *E. japonicus*. The new SNP set presented here constitutes the first SNP-type marker developed for *E. japonicus*, which is

suitable for a range of applications including traceability and population genetics studies.

Keywords *Engraulis encrasicolus* · European anchovy · Next-generation sequencing · Population genetics · Transcriptome · Transferability

Introduction

The Japanese anchovy *Engraulis japonicus* is a small pelagic schooling fish and plankton feeder distributed in the north-western Pacific Ocean. This species has a high economic impact and is heavily fished in China, Japan, Korea and Taiwan [1, 2], where *E. japonicus* larvae are utilized as a food resource known as *shirasu* (whitebait in Japanese). There is increasing market demand and a huge fishing effort for this resource [3], thus due to the risks that could arise from non-sustainable fishing, some governments have taken measures to reduce larval exploitation, e.g., the Taiwan Government announced a seasonal closure to the harvest in 1999 [4]. Potential overexploitation of this species could be avoided by studying its population genetic structure, which provides essential information for sustainable fishery management [5].

While its closely related species, the European anchovy *Engraulis encrasicolus*, has been the target of many population genetics studies [6–19], *E. japonicus* has a mitochondrial genome sequence of 16,675 bp [20, 21] and 20 nuclear microsatellite markers [22, 23], which has limited the number of genetic studies on it [4, 24–27]. Development of genetic markers in *E. japonicus* will help detail its population genetic structure.

Cross-species amplification uses primers and probes developed for a species to sequence and discover genetic

Electronic supplementary material The online version of this article (doi:10.1007/s12562-017-1134-9) contains supplementary material, which is available to authorized users.

✉ Iratxe Montes
iratxe.montes@ehu.eus

Mikel Iriondo
m.iriondo@ehu.eus

Carmen Manzano
carmen.manzano@ehu.eus

Andone Estonba
andone.estonba@ehu.eus

¹ Department of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country (UPV/EHU), Leioa, Spain

markers in its closely related species [28]. This approach is a good choice for developing genetic markers in species with no reference genome but abundant genomic resources available in closely related species. This is the case for *E. japonicus*, since hundreds of genetic markers are available in the closely related *E. encrasicolus*, including mitochondrial DNA (mtDNA) sequences [29], microsatellites [30], and nuclear (nSNP) and mitochondrial single nucleotide polymorphisms (mtSNPs) discovered from cloning, PCR and subsequent comparative sequencing [31], and from transcriptome sequencing [32]. Any of these classes of markers could be cross-amplified in *E. japonicus* to develop new genetic markers for this species. However, the increasing throughput of sequencing and SNP genotyping [28, 33, 34] has confined microsatellites to specific applications, making SNPs the workhorse of molecular ecology and one of the current promising classes of genetic markers.

The aim of this study was to develop SNP markers for *E. japonicus* through cross-species amplification of hundreds of previously discovered SNPs in the nuclear and mitochondrial genome of *E. encrasicolus*. The developed SNP panel could be used to investigate population genetic structure and processes of local adaptation in this species.

Materials and methods

A total of 482 SNPs previously developed for *E. encrasicolus* [31, 32], comprising 467 nSNPs and 15 mtSNPs, were selected for designing primers and probes (Table S1). SNPs were genotyped with TaqMan OpenArray Genotyping System (Life Technologies, CA) (see below) at the Sequencing and Genotyping Service (SGIker) of the University of the Basque Country (UPV/EHU). Information on all the primers/probes is included in Table S1.

The mtSNPs and nSNPs were genotyped for 35 and 22 *E. japonicus* individuals, respectively. Additionally, all the SNPs were genotyped for 22 *E. encrasicolus* individuals as control samples. *E. japonicus* individuals were obtained from a bag of dried anchovy snack purchased at a Japanese supermarket by collaborators. As indicated in the product label, individuals comprising the snack were fished in the Hiuchi-nada area of the Seto Inland Sea, Japan. The 22 *E. encrasicolus* individuals were taken from one haul fished in 2012 in the Bay of Biscay (see sample 14 in [35]) during scientific surveys. These specimens were stored in pure ethanol until DNA extraction.

A piece of muscle tissue (from dried *E. japonicus* and ethanol-fixed *E. encrasicolus* tissue) was used for genomic DNA extraction by a phenol/chloroform method [36]. DNA quantity and purity were measured using Nanodrop ND-8000 spectrophotometer and Qubit fluorometer (Thermo Fisher Scientific, MA). Reactions for SNP amplification/

detection were done with a constant DNA concentration (25 ng/μl) following the instructions of TaqMan OpenArray Genotyping System. Scoring of individual genotypes was performed using TaqMan Genotyper software version 2.1 (Life Technologies). After default clustering, genotyping calls in the scatter plot were reviewed and manually adjusted for producing final cluster assignments.

Based on genotyping calls, SNPs were classified into five categories previously defined by Montes et al. [32]: no signal (no amplification), disperse (< 80% of individuals assigned to a cluster), paralogous sequence variants or multi-site variants (PSV/MSV) ($\geq 99\%$ of heterozygous individuals), monomorphic [minor allele frequency (MAF) < 0.01], and polymorphic (MAF ≥ 0.01) (Fig. 1). Then, conversion and validation rates were obtained. SNPs with a reliable genotyping signal for more than 80% of genotyped individuals were defined as converted SNPs; therefore, conversion rate was estimated as the ratio between the number of monomorphic, PSV/MSV and polymorphic SNPs (Fig. 1c–e), and the number of genotyped SNPs. Then, polymorphic SNPs were considered as the validated ones (transferred or cross-amplified SNPs), therefore, validation rate was estimated as the ratio between the number of polymorphic SNPs and the number of genotyped SNPs. The subsequent analyses were conducted for polymorphic nSNPs. Expected heterozygosity (H_e) was estimated using GeneClass2 [37]. Hardy–Weinberg equilibrium probability test for each nSNP and genotypic linkage disequilibrium between all nSNP pairs were run in Genepop version 4.0 [38] with 10,000 dememorization runs, 1000 batches and 5000 iterations per batch. The false discovery rate method was used to correct interaction p -values for multiple comparisons [39] using the spreadsheet available as supplementary material in Pike [40].

Taking into account the commercial origin of *E. japonicus*, a test for ensuring species assignment (no mislabeling) was carried out. For this purpose, a total of 157 sequences of the mitochondrial cytochrome *b* gene (*cytB*) (Table S2) from five *Engraulis* species, including *E. encrasicolus* and *E. japonicus*, were downloaded from the National Center for Biotechnology Information (NCBI) nucleotide database (<https://www.ncbi.nlm.nih.gov/nucleotide/>). Sequences were aligned with the ClustalW application [41] bundled in BioEdit software [42]. Then, the genotyped mtSNPs in *cytB* were assessed for species identification.

Results

According to nucleotide variations in *cytB*, the five *Engraulis* species had different haplotypes from each other (Table 1).

Results showed high conversion and validation rates for *E. japonicus*. A conversion of 93.6% was obtained with 451

Fig. 1 Genotyping cluster examples for each of the five single nucleotide polymorphism (SNP) categories considered: **a** no signal, **b** disperse, **c** paralogous sequence variants or multi-site variants (PSV/MSV), **d** monomorphic, **e** polymorphic. Each circle corresponds to a genotyped individual. Filled circles correspond to *Engraulis japonicus* individuals, and open circles represent *Engraulis encrasicolus* reference individuals. Light-blue squares are negative controls (water), yellow circles represent non-amplified samples, black circles are undetermined individuals for which genotyping was not possible due to poor amplification, red circles are homozygous samples for allele 1, dark blue circles are homozygous samples for allele 2, green circle cluster represents heterozygous samples

Table 1 Nucleotide bases at each single nucleotide polymorphism (SNP) site in the mitochondrial cytochrome *b* gene (*cytB*) for five anchovy species

Species	<i>n</i>	ss1583998495		ss1583998210		ss1583998454		ss1583998457		ss1583998494	
		14994 ^a		15091 ^a		15263 ^a		15461 ^a		15479-80 ^a	
<i>Engraulis japonicus</i>	28 (35)	A	56	G	53	G	61	T	46	GG	43
<i>Engraulis encrasicolus</i>	112 (22)	A	108	G/A	124/4	A	134	T/C	129/5	CA/GG	104/23
<i>Engraulis anchoita</i>	7	G	2	C	7	G	7	T	7	AC	7
<i>Engraulis mordax</i>	5	A	1	C	5	T	5	T	5	AC	5
<i>Engraulis ringens</i>	5	G	2	C	5	A	5	T	5	AC	5

For each SNP, the observed variant base and the number of variant individuals for each species are given. Numbers in parentheses indicate genotyped individuals

n Number of sequences for each species derived from the National Center for Biotechnology Information (NCBI) nucleotide database (Table S2), *ss* NCBI accession number for each SNP

^aSNP position in a reference sequence (accession number AP009137.1) is shown

converted SNPs. Of the non-converted SNPs, 13 gave no signal and 18 were disperse (Table 2; Table S1). A similar conversion rate was obtained for mtSNPs (93.3%) and nSNPs (93.6%).

With respect to the validation rate, the 451 converted SNPs included three false (PSV/MSV), 272 monomorphic, and 176 polymorphic SNPs, yielding a validation rate of 36.5% (Table 2; Table S1). Validation rate was higher for nSNPs (37.0%) than for mtSNPs (20.0%). No linkage

disequilibrium was found among the 173 nSNPs, indicating that all nSNPs were independent. Mean H_e was 0.102, and no nSNP deviated significantly from Hardy–Weinberg proportions (Table S1).

Regarding species identification, combinations of alleles at five mtSNPs differentiated among *Engraulis* species, and the combination of only two mtSNPs was specific to *E. japonicus* (Table 1). Specifically, the combination of ss1583998210 and ss1583998454 SNPs formed

Table 2 Summary of statistics of SNP genotyping

Category	mtSNPs	nSNPs	Total SNPs
Failed SNPs			
No signal	–	13	13
	–	2.8%	2.7%
Disperse	1	17	18
	6.7%	3.6%	3.7%
False SNPs			
PSV/MSV	–	3	3
	–	0.6%	0.6%
Monomorphic	11	261	272
	73.3%	55.9%	56.4%
True SNPs			
Polymorphic	3	173	176
	20.0%	37.0%	36.5%
Total	15	467	482
Converted SNPs	14	437	451
	93.3%	93.6%	93.6%
Validated SNPs	3	173	176
	20.0%	37.0%	36.5%

The categorization of SNPs is explained in the text

mt Mitochondrial, n nuclear

GG haplotype in *E. japonicus*, which did not appear in any of the four other *Engraulis* species (Table 1). Therefore, the two diagnostic mtSNPs suggested that the genotyped individuals were genuine *E. japonicus*.

Discussion

Conversion rate is generally expected to be low in cross-species amplification studies, due to inter-specific mutations in genomic regions flanking the target SNP where primers/probes are designed for SNP genotyping [43]. This drawback

is especially remarkable in fish species, which are typically characterized by high mutation rates [44]. Consequently, only a few examples of cross-species SNP amplification are available in fish species (see Table 3). The present study showed one of the highest conversion rates (Table 3) likely due to two factors. First, amplified SNPs were transcriptome-derived SNPs that are embedded in coding regions. Therefore, flanking sequences are highly conserved between species and suitable for designing primers/probes that successfully cross-amplify in closely related species [45]. Most of the previous studies for non-model species (Table 3) have discovered SNPs through cross-species amplification based on the emulsion, paired isolation, and concatenation polymerase chain reaction (EPIC-PCR) method [46–48], in which primers are designed on exon regions to amplify the intron [49]. However, the EPIC-PCR has been ousted by transcriptome sequencing through high throughput sequencing, the strategy chosen in the present study, which allows the discovery of thousands of potential SNPs that can be used for cross-species amplification [50, 51]. The second factor that could explain the high conversion rate is related to the fact that the smaller the genetic distance between two taxa, the smaller the number of mutations in the genomic regions for designing primers/probes. The two anchovy species analyzed in the present study are phylogenetically close [52, 53] but, reference and target species in most of the previous cross-amplification studies were distantly related.

The validation rate obtained in the present study was also high (36.5%), and could even have been underestimated because of the limited number of individuals. The small sample size ($n = 22$ for nSNPs), together with its single geographical origin, are likely factors that downwardly biased the proportion of polymorphic SNPs in *E. japonicus*. Thus, the validation rate of the set of markers should increase by analyzing more individuals from multiple locations [54]. Nevertheless, the validation rate obtained in the present study was high compared to that of previous studies

Table 3 Summary of SNP discovery through cross-species amplification in fish species

Reference species	Target species	Tested SNPs	Converted SNPs ^a	Validated SNPs ^a	Reference
<i>Oncorhynchus mykiss</i> , <i>Salmo salar</i>	Three <i>Oncorhynchus</i> spp.	116	80 (69.0%)	53 (45.7%)	[46]
Nine teleost species	<i>Salmo salar</i>	69	27 (39.1%)	11 (15.9%)	[48]
	<i>Thymallus thymallus</i>	42	5 (11.9%)	5 (11.9%)	
	<i>Salmo trutta</i>	10	6 (60.0%)	4 (40.0%)	
	<i>Salvelinus alpinus</i>	10	3 (30.0%)	1 (10.0%)	
	<i>Clupea harengus</i>	578	572 (99.0%)	69 (11.9%)	
	<i>Clupea pallasii</i>	578	58 (10.0%)	10 (1.7%)	[50]
	<i>Engraulis encrasicolus</i>	578	58 (10.0%)	10 (1.7%)	
<i>Thunnus alalunga</i>	<i>Thunnus thynnus</i>	128	32 (25.0%)	23 (18.0%)	[47]
<i>Thunnus thynnus</i>	<i>Thunnus alalunga</i>	384	311 (81.0%)	93 (24.2%)	[51]
<i>Engraulis encrasicolus</i>	<i>Engraulis japonicus</i>	482	451 (93.6%)	176 (36.5%)	This study

^aTerms explained in the text

(Table 3). Albaina et al. [47] conducted a cross-species amplification of SNPs from albacore to Atlantic bluefin tuna, and obtained a validation rate of 18%; the reciprocal cross resulted in a validation success of 24.2% [51]. Cross-amplification success between Atlantic and Pacific herring, *Clupea harengus* and *Clupea pallasii* respectively, is even lower, 11.9% [50]. The high success obtained in the present study is likely due to the high polymorphic level of the anchovy species, and to the design of an appropriate SNP set, with SNPs from coding regions. Overall, the present study highlights the importance of exploiting the exponentially increasing amount of gene sequence data in public databases, which are a very useful basis for identifying new polymorphic markers in non-model organisms.

In conclusion, a set of 176 polymorphic SNPs was developed for *E. japonicus*, which represents an important and valuable genomic resource for this species. This set can be useful for further research on population genomics, which can provide valuable information on ecological factors and for conservation. In addition, two diagnostic SNPs were developed to identify *E. japonicus*, which could serve as molecular tools for food traceability.

Acknowledgments The authors thank SGIker of the University of the Basque Country (UPV/EHU) for technical support. The authors are indebted to Mr. Igor Velado for providing samples of *E. japonicus*. This research was supported by the Genomic-Resources Research Group of the Basque University System (IT558-10) supported by the Department of Education, Universities and Research of the Basque government. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions All authors contributed extensively to the work presented in this paper. I. M. and A. E. designed the study. I. M. carried out the laboratory work and analyzed data. I. M. interpreted the data and wrote the paper. M. I., C. M. and A. E. gave technical support and conceptual advice, and edited the manuscript.

Compliance with ethical standards

Data accessibility All the sequences of the primers and probes used are included in Table S1, and accession numbers from all sequences used for the species identification step are included in Table S2.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Young SS, Chiu TS (1994) Maturation of *Japanese anchovy*, *Engraulis japonica* T & S., from I-lan Bay, NE Taiwan. *Zool Stud* 33:302–309
- Chiu TS, Young SS, Chen CS (1997) Monthly variation of larval anchovy fishery in I-Lan Bay, NE Taiwan, with an evaluation for optimal fishing season. *J Fish Soc Taiwan* 24:273–282
- Yasue N, Takasuka A (2009) Seasonal variability in growth of larval *Japanese anchovy* *Engraulis japonicus* driven by fluctuations in sea temperature in the Kii Channel, Japan. *J Fish Biol.* doi:10.1111/j.1095-8649.2009.02238.x
- Yu HT, Lee YJ, Huang SW, Chiu TS (2002) Genetic analysis of the populations of *Japanese anchovy* (Engraulidae: *Engraulis japonicus*) using microsatellite DNA. *Mar Biotechnol* 4(5):471–479
- Reiss H, Hoarau G, Dickey-Collas M, Wolff WJ (2009) Genetic population structure of marine fish: mismatch between biological and fisheries management units. *Fish Fish.* doi:10.1111/j.1467-2979.2008.00324.x
- 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.* doi:10.1111/j.1095-8649.1989.tb02993.x
- Bembo DG, Carvalho GR, Cingolani N, Pitcher TJ (1996) Electrophoretic analysis of stock structure in northern Mediterranean anchovies *Engraulis encrasicolus*. *ICES J Mar Sci.* doi:10.1006/jmsc.1996.0011
- Borsa P (2002) Allozyme, mitochondrial-DNA, and morphometric variability indicate cryptic species of anchovy (*Engraulis encrasicolus*). *Biol J Linn Soc.* doi:10.1046/j.1095-8312.2002.00018.x
- Ivanova PP, Dobrovolov IS (2006) Population-genetic structure on European anchovy (*Engraulis encrasicolus*, Linnaeus, 1758) (Osteichthyes: Engraulidae) from Mediterranean Basin and Atlantic Ocean. *Acta Adriat* 47:13–22
- Magoulas A, Castilho R, Caetano S, Marcato S, Partanello T (2006) Mitochondrial DNA reveals a mosaic pattern of phylogeographical structure in Atlantic and Mediterranean populations of anchovy (*Engraulis encrasicolus*). *Mol Phylogenet Evol.* doi:10.1016/j.ympev.2006.01.016
- Sanz N, García-Marín JL, Viñas J, Roldán M, Pla C (2008) Spawning groups of European anchovy: population structure and management implications. *ICES J Mar Sci.* doi:10.1093/icesjms/fns128
- Zarraonaindia I, Iriondo M, Albaina A, Pardo MA, Manzano C, Grant WS, Irigoien X, Estonba A (2012) Multiple SNP markers reveal fine scale population and deep phylogeographic structure in European anchovy (*Engraulis encrasicolus* L.). *PLoS ONE.* doi:10.1371/journal.pone.0042201
- Zarraonaindia I, Pardo MA, 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.* doi:10.1093/icesjms/fsp187
- Borrell YJ, Piñera JA, Sánchez Prado JA, Blanco G (2012) Mitochondrial DNA and microsatellite genetic differentiation in the European anchovy *Engraulis encrasicolus* L. *ICES J Mar Sci.* doi:10.1093/icesjms/fsS229
- Tuncay SS, Bardakci F (2014) Population genetics of European Anchovy (*Engraulis encrasicolus* L.) in the seas of Turkey based on microsatellite DNA. *Front Mar Sci.* doi:10.3389/conf.fmars.2014.02.00154
- Viñas J, Sanz N, Peñarrubia L, Araguas RM, García-Marín JL, Roldán MI, Pla C (2014) Genetic population structure of European anchovy in the Mediterranean Sea and the Northeast Atlantic Ocean using sequence analysis of the mitochondrial DNA control region. *ICES J Mar Sci.* doi:10.1093/icesjms/fst132
- Ouazzani KC, Benazzou T, Tojo N, Chlaida M (2015) Mitochondrial phylogenomics and genetic population structure of anchovies (*Engraulis encrasicolus*) along the Moroccan coast using

- sequence analysis of the mitochondrial DNA cytochrome b. PeerJ PrePrints. doi:[10.7287/peerj.preprints.1543v1](https://doi.org/10.7287/peerj.preprints.1543v1)
18. Montes I, Iriondo M, Manzano C, Santos M, Conklin D, Carvalho GR, Irigoien X, Estonba A (2016) No loss of genetic diversity in the exploited and recently collapsed population of Bay of Biscay anchovy (*Engraulis encrasicolus* L.). Mar Biol. doi:[10.1007/s00227-016-2866-2](https://doi.org/10.1007/s00227-016-2866-2)
 19. Montes I, Zarraonaindia I, Iriondo M, Grant SW, Manzano C, Cotano U, Conklin D, Irigoien X, Estonba A (2016) Transcriptome analysis deciphers evolutionary mechanisms underlying genetic differentiation between coastal and offshore anchovy populations in the Bay of Biscay. Mar Biol. doi:[10.1007/s00227-016-2979-7](https://doi.org/10.1007/s00227-016-2979-7)
 20. Inoue JG, Miya M, Tsukamoto K, Nishida M (2001) Complete mitochondrial DNA sequence of the Japanese anchovy *Engraulis japonicus*. Fish Sci 67:828–835
 21. Zhang H, Xian W (2015) The complete mitochondrial genome of the larvae Japanese anchovy *Engraulis japonicus* (Clupeiformes Engraulidae). MDN. doi:[10.3109/19401736.2013.863297](https://doi.org/10.3109/19401736.2013.863297)
 22. Chiu TS, Lee YJ, Huang SW, Yu HT (2002) Polymorphic microsatellite markers for stock identification in Japanese anchovy (*Engraulis japonica*). Mol Ecol Notes. doi:[10.1046/j.1471-8286.2002.00142.x](https://doi.org/10.1046/j.1471-8286.2002.00142.x)
 23. Lin L, Zhu L, Liu SF, Su YQ, Zhuang ZM (2011) Polymorphic microsatellite loci for the Japanese anchovy *Engraulis japonicus* (Engraulidae). Genet Mol Res. doi:[10.4238/vol10-2gmr1085](https://doi.org/10.4238/vol10-2gmr1085)
 24. Kim JY, Cho ES, Kim WJ (2004) Population genetic structure of Japanese anchovy (*Engraulis japonicus*) in Korean waters based on mitochondrial 12S ribosomal RNA gene sequence. J Life Sci 14:938–950
 25. Cho ES, Kim JI (2006) Mitochondrial DNA polymorphism of the Japanese anchovy (*Engraulis japonicus* Temminck & Schlegel) collected from the Korean offshore and inshore waters. J Life Sci 16:812–827
 26. Oh TY, Kim JI, Seo YI, Cho ES (2009) The population genetic structure of the Japanese Anchovy (*Engraulis japonicus* Temminck & Schlegel) in the West, South and East Seas of Korea Based on Microsatellite DNA Analysis. J Life Sci. doi:[10.5352/JLS.2009.19.2.17](https://doi.org/10.5352/JLS.2009.19.2.17)
 27. Yu ZN, Kong XY, Guo TH, Jiang YY, Zhuang ZM, Jin XS (2005) Mitochondrial DNA sequence variation of Japanese anchovy *Engraulis japonicus* from the Yellow Sea and East China Sea. Fish Sci 71:299–307. doi:[10.1111/j.1444-2906.2005.00964.x](https://doi.org/10.1111/j.1444-2906.2005.00964.x)
 28. Seeb JE, Carvalho G, Hauser L, Naish K, Roberts S, Seeb LW (2011) Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. Mol Ecol Res. doi:[10.1111/j.1755-0998.2010.02979.x](https://doi.org/10.1111/j.1755-0998.2010.02979.x)
 29. Lavoué S, Miya M, Saitoh K, Ishiguro NB, Nishida M (2007) Phylogenetic relationships among anchovies, sardines, herrings and their relatives (*Clupeiformes*), inferred from whole mitogenome sequences. Mol Phylogenet Evol 43(3):1096–1105
 30. Pakaki V, Magoulas A, Kasapidis P (2009) New polymorphic microsatellite loci for population studies in the European anchovy, *Engraulis encrasicolus* (L.). Mol Ecol. doi:[10.1111/j.1755-0998.2009.02681.x](https://doi.org/10.1111/j.1755-0998.2009.02681.x)
 31. Molecular Ecology Resources Primer Development Consortium, Abreu AG, Albaina A, Alpermann TJ, Apkenas VE, Bankhead-Dronnet S, Bergek S, Berumen ML, Cho CH, Clobert J, Coulon A, De Feraudy D, Estonba A, Hankeln T, Hochkirch A, Hsu TW, Huang TJ, Irigoien X, Iriondo M, Kay KM, Kinitz T, Kothera L, Le Hénaff M, Lieutier F, Lourdais O, Macrini CM, Manzano C, Martin C, Morris VR, Nanninga G, Pardo MA, Plieske J, Pointeau S, Prestegard T, Quack M, Richard M, Savage HM, Schwarcz KD, Shade J, Simms EL, Solferini VN, Stevens VM, Veith M, Wen MJ, Wicker F, Yost JM, Zarraonaindia I (2012) Permanent genetic resources added to Molecular Ecology Resources Database 1 October 2011–30 November 2011. Mol Ecol Res. doi:[10.1111/j.1755-0998.2011.03109.x](https://doi.org/10.1111/j.1755-0998.2011.03109.x)
 32. Montes I, Conklin D, Albaina A, Creer S, Carvalho GR, Santos M, Estonba A (2013) SNP discovery in European anchovy (*Engraulis encrasicolus* L.) by high-throughput transcriptome and genome sequencing. PLoS ONE. doi:[10.1371/journal.pone.0070051](https://doi.org/10.1371/journal.pone.0070051)
 33. De Wit P, Pespeni MH, Palumbi SR (2015) SNP genotyping and population genomics from expressed sequences—current advances and future possibilities. Mol Ecol. doi:[10.1111/mec.13165](https://doi.org/10.1111/mec.13165)
 34. Grover A, Sharma PC (2016) Development and use of molecular markers: past and present. Crit Rev Biotechnol. doi:[10.3109/07388551.2014.959891](https://doi.org/10.3109/07388551.2014.959891)
 35. Montes I, Zarraonaindia I, Iriondo M, Grant WS, Manzano C, Cotano U, Conklin D, Irigoien X, Estonba A (2016) Transcriptome analysis deciphers evolutionary mechanisms underlying genetic differentiation between coastal and offshore anchovy populations in the Bay of Biscay. Mar Biol 163:205. doi:[10.1007/s00227-016-2979-7](https://doi.org/10.1007/s00227-016-2979-7)
 36. Sambrook J, Russell DW (2006) Purification of nucleic acids by extraction with phenol: chloroform. Cold Spring Harb Protoc. doi:[10.1101/pdb.prot4455](https://doi.org/10.1101/pdb.prot4455)
 37. Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. J Hered. doi:[10.1093/jhered/esh074](https://doi.org/10.1093/jhered/esh074)
 38. Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol Ecol Res. doi:[10.1111/j.1471-8286.2007.01931.x](https://doi.org/10.1111/j.1471-8286.2007.01931.x)
 39. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Stat Methodol 57:289–300
 40. Pike N (2011) Using false discovery rates for multiple comparisons in ecology and evolution. Methods Ecol Evol 2:278–282. doi:[10.1111/j.2041-210X.2010.00061.x](https://doi.org/10.1111/j.2041-210X.2010.00061.x)
 41. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680
 42. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98
 43. Miller JR, Koren S, Sutton G (2010) Assembly algorithms for next-generation sequencing data. Genomics. doi:[10.1016/j.ygeno.2010.03.001](https://doi.org/10.1016/j.ygeno.2010.03.001)
 44. Knutsen H, Olsen EM, Jorde PE, Espeland SH, André C, Stenseth NC (2011) Are low but statistically significant levels of genetic differentiation in marine fishes 'biologically meaningful'? A case study of coastal Atlantic cod. Mol Ecol. doi:[10.1111/j.1365-294X.2010.04979.x](https://doi.org/10.1111/j.1365-294X.2010.04979.x)
 45. Malhi RS, Trask JS, Shattuck M, Johnson J, Chakraborty D, Kanthaswamy S, Ramakrishnan U, Smith DG (2011) Genotyping single nucleotide polymorphisms (SNPs) across species in Old World monkeys. Am J Primatol. doi:[10.1002/ajp.20969](https://doi.org/10.1002/ajp.20969)
 46. Smith CT, Elfstrom CM, Seeb LW, Seeb JE (2005) Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. Mol Ecol. doi:[10.1111/j.1365-294X.2005.02731.x](https://doi.org/10.1111/j.1365-294X.2005.02731.x)
 47. Albaina A, Iriondo M, Velado I, Laconcha U, Zarraonaindia I, Arrizabalaga H, Pardo MA, Lutcavage M, Grant WS, Estonba A (2013) Single nucleotide polymorphism discovery in albacore and Atlantic bluefin tuna provides insights into worldwide population structure. Anim Genet. doi:[10.1111/age.12051](https://doi.org/10.1111/age.12051)
 48. Ryyänänen HJ, Primmer CR (2006) Single nucleotide polymorphism (SNP) discovery in duplicated genomes: intron-primed exon-crossing (IPEC) as a strategy for avoiding amplification of

- duplicated loci in Atlantic salmon (*Salmo salar*) and other salmonid fishes. BMC Genomics. doi:[10.1186/1471-2164-7-192](https://doi.org/10.1186/1471-2164-7-192)
49. Li C, Riethoven JJM, Ma L (2010) Exon-primed intron-crossing (EPIC) markers for non-model teleost fishes. BMC Evol Biol. doi:[10.1186/1471-2148-10-90](https://doi.org/10.1186/1471-2148-10-90)
 50. Helyar SJ, Limborg MT, Bekkevold D, Babbucci M, Van Houdt J, Maes GE, Bargelloni L, Nielsen RO, Taylor MI, Ogden R, Cariani A, Carvalho GR, FishPopTrace Consortium, Panitz F (2012) SNP discovery using next generation transcriptomic sequencing in Atlantic herring (*Clupea harengus*). PLoS ONE. doi:[10.1371/journal.pone.0042089](https://doi.org/10.1371/journal.pone.0042089)
 51. Laconcha U, Iriando 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. doi:[10.1371/journal.pone.0128247](https://doi.org/10.1371/journal.pone.0128247)
 52. Jérôme M, Martinsohn JT, Ortega D, Carreau P, Verrez-Bagnis V, Mouchel O (2008) Toward fish and seafood traceability: anchovy species determination in fish products by molecular markers and support through a public domain database. J Agric Food Chem. doi:[10.1021/jf703704m](https://doi.org/10.1021/jf703704m)
 53. Chairi H, Rebordinos L (2014) A rapid method for differentiating four species of the Engraulidae (Anchovy) family. J Agric Food Chem. doi:[10.1021/jf405680g](https://doi.org/10.1021/jf405680g)
 54. Lachance J, Tishkoff SA (2013) SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it. BioEssays. doi:[10.1002/bies.201300014](https://doi.org/10.1002/bies.201300014)