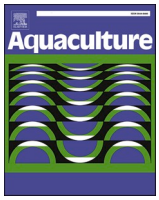
Aquaculture 550 (2022) 737836

Contents lists available at ScienceDirect

Aquaculture

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



Population structure and geographic origin assignment of *Mytilus galloprovincialis* mussels using SNPs

Ane del Rio-Lavín [a,*], Natalia Díaz-Arce [b], María Angélica Larraín [c], Cristián Araneda [d], Naiara Rodríguez-Ezpeleta [b], Elisa Jiménez [a], Miguel Ángel Pardo [a]

[a] AZTI, Food Research, Basque Research and Technology Alliance (BRTA), Parque Tecnológico de Bizkaia, Astondo Bidea, Edificio 609, 48160 Derio, Bizkaia, Spain
[b] AZTI, Marine Research, Basque Research and Technology Alliance (BRTA), Txatxarramendi ugartea z/g, 48395 Sukarrieta, Bizkaia, Spain
[c] Departamento de Ciencias de los Alimentos y Tecnología Química, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, Olivos 1007, Independencia, Santiago, Chile
[d] Departamento de Producción Animal, Facultad de Ciencias Agronómicas, Avda. Santa Rosa 11315, La Pintana, Santiago, Chile

ARTICLE INFO

Keywords:
Mytilus galloprovincialis
RAD-seq
SNPs
Origin assignment
Aquaculture
Traceability

ABSTRACT

Seafood traceability represents a major goal for regulators and fishing industries worldwide who seek to prevent commercial fraud, protect marine resources and ensure consumer safety. Genetic approaches can be used to trace the provenance of seafood based on the ability of DNA markers to assign samples back to their population of origin. Here, we have used thousands of genome-wide Single Nucleotide Polymorphism markers to provide a detailed genetic structure of the highly farmed Mediterranean mussel *Mytilus galloprovincialis* in part of its native (Atlantic and Mediterranean areas) and introduced ranges (South-eastern Pacific area). Also, we have assessed the power of the newly developed markers to assign samples to their geographic origin. Results showed a clear differentiation between the Atlantic and the Mediterranean *M. galloprovincialis* populations, with significant differences also observed between the Mediterranean and South-eastern Pacific individuals. In addition, we found 90–100% of individuals could be correctly assigned to the Atlantic or Mediterranean/South-eastern Pacific populations when using only 10 to 25 SNPs. Our results support the possibility of the development of an accurate and cost-effective origin assignment tool with global uses in aquaculture management, seafood traceability and food safety.

1. Introduction

Fisheries and aquaculture provide nutritious food and employment to many countries across the world. As the seafood market globalises, ensuring the traceability of this highly traded food commodity is becoming crucial. Tracing the geographical origin of fish and fish products is the key to prevent commercial fraud, to guarantee sustainable fisheries and aquaculture management and to enforce food safety regulations. This is not only essential for food authorities, but also for producers who aim at certifying their products and promoting consumer confidence. In the EU, seafood species identification is mandatory, as stated in the Council Regulation (EC) N° 1379/2013 on the common organization of the markets in fishery and aquaculture products, amending Council Regulations (EC) N° 1184/2006 and (EC) N° 1224/2009 and repealing Council Regulation (EC) N° 104/2000. These regulations require that seafood labels indicate provenance to assure transparency and traceability along the food chain. Being able to guarantee the provenance is particularly relevant for mussels due to their importance in international trade. Processing of mussels extends their limited shelf-life and thus allows an increase in the distance travelled from producers to consumers, significantly hindering their traceability.

Mussels constitute an important protein source worldwide available as fresh, frozen and processed products, whose global aquaculture production has increased over the last decades with a production of 2.11 million tonnes in 2018, valued in ~4519 million USD (FAO, 2020). Mussel aquaculture is mainly performed in semiculture systems where the spat are collected directly from the environment and can thus be considered natural populations. The European mussels comprise three species: *M. galloprovincialis*, *M. edulis* and *M. trossulus*. In the European Union (EU), mussels account for 34% of the total aquaculture production, being *M. edulis* and *M. galloprovincialis* the two main species harvested and cultivated (EUMOFA, 2019). Spain is the main

* Corresponding author.
E-mail address: adelrio@azti.es (A. del Rio-Lavín).

https://doi.org/10.1016/j.aquaculture.2021.737836
Received 23 July 2021; Received in revised form 9 December 2021; Accepted 14 December 2021
Available online 18 December 2021

M. galloprovincialis productor in the EU, mainly in the north western coast of the Iberian Peninsula, Galicia, producing more than 97% of the Spanish mussels. The native distribution of *M. galloprovincialis* covers the Mediterranean Sea, the Black Sea and the Atlantic coast from Portugal to France, but its current distribution is extensive worldwide, due to human-mediated introductions and its ability to rapidly adapt to varying environmental conditions (Westfall and Gardner, 2010; Gardner et al., 2016; Larrain et al., 2018; Zardi et al., 2018; Zbawicka et al., 2018). The presence of the invasive Northern hemisphere *M. galloprovincialis* in a major mussel producing country like Chile may pose a threat to the local aquaculture industry. Several previous studies (Toro et al., 2005; Westfall and Gardner, 2010; Tarifeño et al., 2012; Astorga et al., 2015; Larrain et al., 2018) have reported its presence in the central coast of Chile, in the BíoBio region, being recent anthropogenic activities such as shipping and aquaculture the likely source of invasion. Thus, tracing the geographical origin of this species is also crucial for the development and implementation of management strategies to mitigate the invasion and protect the sustainable exploitation of native bivalve species.

Genetic markers such as SNPs have become powerful tools for fish and fish product traceability (Martinsohn and Ogden, 2009). Geographic origin identification is based on the ability to assign a sample to a particular population, while excluding it from others, requiring the populations of interest to be sufficiently genetically distinct one from each other. This way, the degree of population divergence would dictate the selection of SNP markers for origin assignment (Ogden and Linacre, 2015). Several analyses based on SNP markers have been performed in *Mytilus* taxa, to differentiate populations between and within this species, either by using sequences available in public databases for SNP discovery (Zbawicka et al., 2012; Fraisse et al., 2015; Wenne, 2018), or, more recently, using new SNPs discovered through Restriction site-Associated DNA sequencing (RAD-seq) (Araneda et al., 2016). Compared to other markers, SNPs offer advantages due to their higher genome coverage across neutral and under selection regions (Morin et al., 2004; Helyar et al., 2011), and their applicability in highly fragmented DNA, a characteristic leveraging rapid and efficient genotyping of processed seafood products (Scarano and Rao, 2014).

A few studies have focused on describing genetic diversity and population structure of *M. galloprovincialis* along its native range where this species is extensively farmed. Paterno et al. (2019) provided an overall SNP-based genetic structure of *M. galloprovincialis* in the Mediterranean and Black seas, evidencing strong differences between samples from the two seas, whereas no differences were found between samples from the Western and Central Mediterranean areas. However, Atlantic samples were missing from those analyses, which would have been desirable to completely analyse the native range of this species. On the Iberian coast, a handful of studies have described population structure using allozymes (Quesada et al., 1995b), mitochondrial markers (Quesada et al., 1995a; Kijewski et al., 2011) and microsatellites (Diz and Presa, 2008). Although these studies support the existence of two Iberian *M. galloprovincialis* populations, Atlantic and Mediterranean, they are restricted by the type and number of markers used and lack of origin assignment analysis.

In the present study, we performed population genetic analyses of farmed *M. galloprovincialis* based on hundreds of samples from nine locations and thousands of SNP markers obtained by RAD-seq. Our aim is: (1) to assess the population structure of the farmed *M. galloprovincialis* in part of its native (Atlantic and Mediterranean areas) and introduced ranges (South-eastern Pacific area); and (2) to develop and evaluate a new SNP maker based tool to assign samples to their geographic origin.

2. Materials and methods

2.1. Sample collection, DNA extraction and RAD-seq libraries preparation and sequencing

A total of 222 adult samples of *M. galloprovincialis* (5–9 cm) were collected from nine different mussel farms from autumn 2018 to winter 2019 (Table 1, Fig. 1). From each mussel, a piece of mantle border tissue was excised and immediately stored in RNAlater® (Sigma Aldrich, USA). Genomic DNA extraction was performed using the Wizard® Genomic DNA Purification Kit (Promega, USA), starting from 50 mg of tissue and following manufacturer's instructions. Extracted DNA was eluted in sterile Milli-Q water and its concentration was determined by means of the Quant-iT dsDNA HS assay kit using a Qubit® 1.0 Fluorometer (Life Technologies, USA). DNA integrity was assessed by electrophoresis on agarose gel 0.8% (*w/v*). Species identification was performed by a real-time PCR assay targeting the polyphenolic adhesive protein gene using primers previously described by Dias et al. (2008), with a SYBR™ Green post-PCR melting curve analysis (del Rio-Lavín et al., 2021). This methodology shows specific peaks in dissociation curves with unique melting temperature (Tm) values for each species.

RAD-seq libraries were prepared following the method described by Etter et al. (2011). Briefly, 500 ng of starting DNA was digested with the *SbfI* restriction enzyme (Thermo Fisher, USA) and ligated to modified Illumina P1 adapters containing 5 bp unique barcodes. Pools of DNA from 32 individuals were sheared using the Covaris® M220 Focused-ultrasonicator™ Instrument (Life Technologies, USA) and size selected to 300–500 bp by AMPure XP Beads (Beckman, USA) (Bronner et al., 2014). After Illumina P2 adaptor ligation, each library was amplified using 12 PCR cycles. Each pool was paired-end sequenced (125 bp) on an Illumina HiSeq2500 (Illumina Inc., USA).

2.2. RAD-tag assembly and SNP calling

Raw reads were analysed using Stacks version 2.3b (Catchen et al., 2013). Detailed procedures followed in this study are shown in Fig. S1. Demultiplexing and quality filtering were performed using the "process_radtags" module, truncate ng all reads to 80 nucleotides to avoid the low-quality bases at the end of the reads. PCR duplicates were removed by applying the "clone_filter" module to the reads whose forward and reverse pairs passed filtering. Only samples with at least 500,000 quality passing forward reads were kept (Table S1). Stacks per individual were *de novo* assembled using the "ustacks" module with a minimum depth coverage to create a stack (m) of three and a maximum nucleotide mismatches (M) allowed between stacks of four. Only samples with between 6000 and 25,000 RAD loci and a minimum coverage of 10× were kept. A catalog of RAD loci was generated using the "cstacks" module, allowing a maximum number of mismatches between sample tags (n) of five and using a population map with the nine mussel production points. Stacks parameters (m, M and n) were selected following the method of *de novo* parameter optimization based on Paris et al. (2017). Matches of individual RAD loci to the catalog were searched using the "sstacks" module, data were stored by locus and SNPs were called using information from paired-end reads using the "tsv2bam" and "gstacks" modules.

From the generated catalog, genotype tables including all individuals (Dataset 1) or only the Mediterranean and South-eastern Pacific individuals (Dataset 2) were generated as follows. SNPs present in RAD loci found in at least 70% of the individuals across populations were selected and exported to PLINK, Variant Call Format (VCF) and Genepop formats using the "populations" module. Only SNPs located on the forward reads were considered. Using PLINK version 1.09b (Purcell et al., 2007), SNPs with a Minimum Allele Frequency (MAF) lower than 0.05, a genotyping rate lower than 0.9 and failing the Hardy-Weinberg equilibrium test at $P < 0.05$ in two or more of the sampling points were excluded. Only samples with a genotyping rate above 0.8 were retained.

Table 1
Mussel samples used for population genetic analyses. MED = Mediterranean area; ATL = Atlantic area; SEP = South-eastern Pacific area. Dataset 1 (all areas): ATL, MED and SEP. Dataset 2*: MED and SEP.

Area	Country / Region	Sampling point	Code	Latitude	Longitude	N° indiv.	N° indiv. Genotyped	N° indiv. Training set	N° indiv. Holdout set
ATL	Spain / Basque country	Mutriku	MUES	43° 18′40.2” N	2° 22′35.2” W	25	20	10	10
	Spain / Basque country	Mendexa	MEES	43°21′20.2” N	2° 26′53.4” W	25	22	9	13
	Spain / Galicia	Ría de Betanzos-Sada	SGES	43° 23′ 45.0” N	8° 17′ 40.0” W	23	10	4	6
	Spain / Galicia	Ría de Arousa	AGES	42° 29′ 36.0” N	8° 55′ 51.0” W	24	18	10	8
	Portugal / Algarve	Porto da Baleeira	PBPT	37°00′41.8” N	8°55′48.7” W	25	24	14	10
						122	**94**	**47**	**47**
MED*	Spain/ Catalonia	Delta del Ebro	DEES	40°33′16.3” N	0°32′27.8″ E	25	22	10 (9)	12 (13)
	Tunisia / Bizerte	Bizerte lagoon	LBTN	37°11′48.0″ N	9°51′23.0″ E	25	10	4 (7)	6 (3)
						50	**32**	**14 (16)**	**18 (16)**
SEP*	Chile / Biobío	Coliumo	COCL	36°22′16.4” S	72°57′25.2” W	25	18	10 (9)	8 (9)
	Chile / Biobío	Caleta Tumbes	TBCL	36°32′14.0” S	72°57′32.0” W	25	23	13 (12)	10 (11)
						50	**41**	**23 (21)**	**18 (20)**
					Dataset 1	**222**	**167**	**84**	**83**
					Dataset 2*	**100**	**73**	**37**	**36**

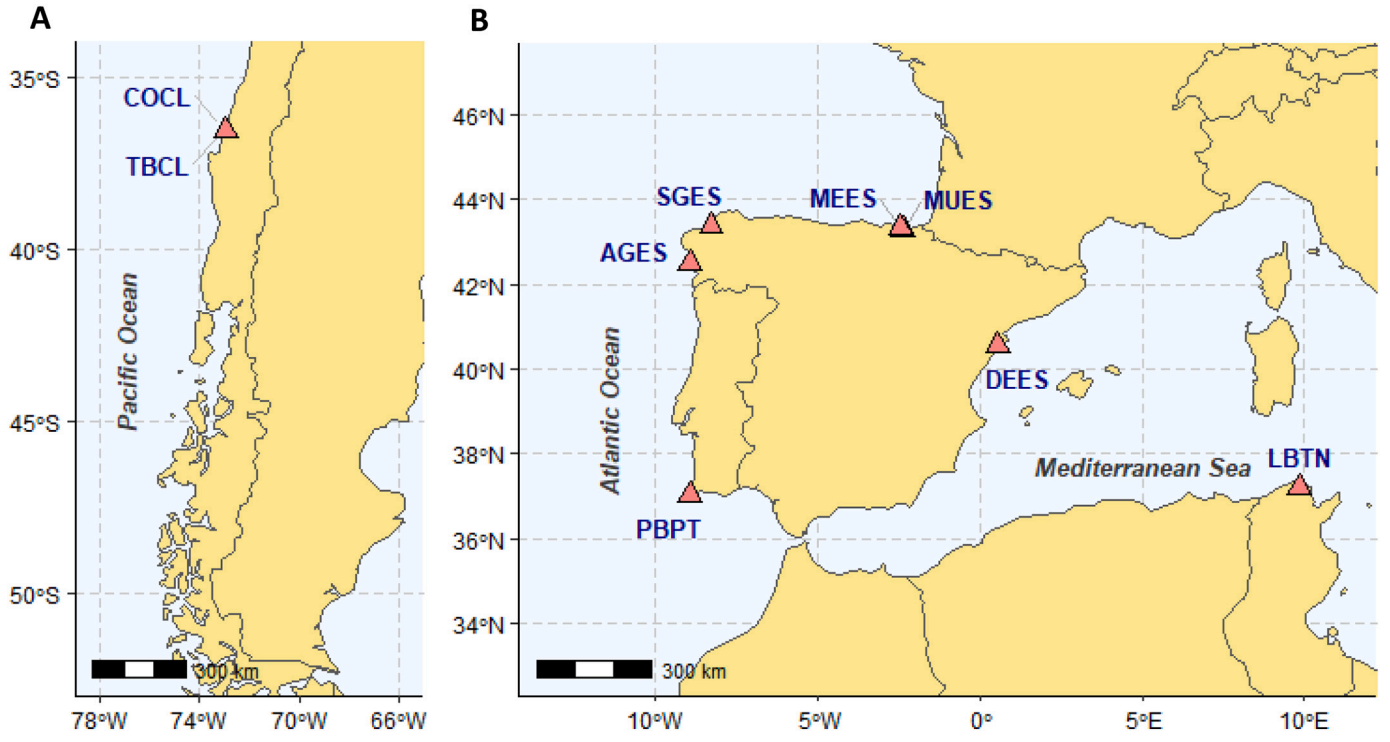


Fig. 1. Map showing the geographic location where the samples used for this study were collected. See Table 1 for definitions of abbreviations.

Genotype tables were exported to Structure and Genepop formats using PDGSpider 2.1.1.5 (Lischer and Excoffier, 2012).

2.3. Genetic diversity and population structure analysis

For population structure, the following analyses were performed for each dataset including only the first SNP per tag. Principal Component Analysis (PCA) was performed using the adegenet R package with no previous population assignment of samples (Jombart and Ahmed, 2011). ADMIXTURE (Alexander et al., 2009) was run setting 5000 bootstraps for each K value (number of potential ancestral populations, that ranged from 1 to the number of presumed populations +1). A first ADMIXTURE run was launched for each value of K to check the number of steps necessary to reach the default 0.001 likelihood value during the run. This information was used to set the “-c” parameter (steps to be fulfilled in each bootstrapped run) to assure convergence. The K value with the lowest associated error was identified using ADMIXTURE’s cross-validation procedure. Genetic differentiation between sampling areas was analysed through the total and pairwise F_{ST} values with a 95% confidence interval (CI) estimated running 10,000 bootstraps with the R package diveRsity (Keenan et al., 2013). Loci potentially influenced by positive selection in each dataset were identified using the R package pcadapt (Luu et al., 2016). A screeplot representing the percentage of variance explained by each PC was used to choose the number of

principal components (K), and SNPs with *p*-values (adjusted following Benjamini and Hochberg (1995)) below 0.05 were classified as genomic outliers.

2.4. Origin assignment informative loci selection

The selection of the most informative loci for geographic origin assignment was performed based on F_{ST} values calculated by GENEPOP 4.7.3 (Raymond and Rousset, 1995; Rousset, 2008) following Weir and Cockerham, 1984. Considering the genetic clusters obtained in this study, two different assignment analysis were performed in order to match individuals to one of the three ocean areas of interest: (1) Atlantic *vs* Mediterranean/ South-eastern Pacific individuals (using Dataset 1) and (2) Mediterranean *vs* South-eastern Pacific individuals (using Dataset 2). Genetic assignment was performed following the "Training, Holdout & Leave-one-out" (THL) method (Anderson, 2010) in order to avoid high-grading bias. Briefly, for each dataset, individuals were randomly divided in two groups (training and holdout) with equally distributed number of samples from each location (Table 1). Training samples were used to identify the loci with the highest discriminative power to assign samples their areas of origin based on F_{ST} values. Next, using both training and holdout samples as reference, assignment scores for each holdout individual was calculated with Geneclass2 (Piry et al., 2004) using Rannala and Mountain (1997) criterion (0.05 threshold) and following leave-one-out (LOO) procedure. Assignment power of the most discriminant SNPs was assessed by calculating percentages of correctly and incorrectly assigned holdout individuals to their origin. The assignment rate was calculated using two levels of assignment probability thresholds: 80% and 90%, and samples below these values were considered unassigned.

3. Results

All the 222 mussel samples analysed in this study were identified as *M. galloprovincialis*, all showing the dissociation curve peak corresponding to a melting temperature of 78.856 ± 0.16. Three samples from Galicia and five from the Basque Country, excluded from the analysis, were identified as *M. galloprovincialis* x *M. edulis* hybrids, meaning that hybrids were found in four of the nine production points.

3.1. RAD-tag assembly and SNP calling

The total number of read pairs retained after quality filtering was 288,996,135 (79%), with an average of 1,301,784 reads per individual. After PCR duplicate removal, with an average of 27.6% clones per individual, 201,648,224 were used for SNP identification. The mean number of RAD tags obtained per sample was 13,976. After individual and SNP filtering, the final datasets resulted in 167 samples and 959 SNPs (Dataset 1) and 1506 SNPs and 73 samples (Dataset 2) (Table S2).

3.2. Population structure and genetic clustering

PCA and ADMIXTURE analyses showed that mussels from the Atlantic area (Basque Country, Galicia and Algarve) are genetically distinct from those from the Mediterranean Sea (Ebro Delta and Bizerte) and South-eastern Pacific (Biobío) areas (Fig. 2A and 3A). ADMIXTURE analysis identified K = 2 as the most likely number of ancestral populations whose proportions differentiate two main *M. galloprovincialis* populations within the Iberian coast. Individuals from the Atlantic Ocean were clearly separated from the Mediterranean Sea and South-eastern Pacific populations by the first principal component explaining 5,55% of variance in allele frequency in the PCA analysis. Interestingly, the Mediterranean samples grouped closer to the South-eastern Pacific samples, than to the Atlantic samples. Although originated from the same ancestral populations, samples from Mediterranean and South-eastern Pacific locations have diverged and are genetically differentiated (Fig. 2B and 3B).

Total F_{ST} estimations revealed significant differences between the three above mentioned groups. The highest F_{ST} values were obtained between the Atlantic and the other two areas: 0.0959 (95% CI: 0.0882, 0.1043) with respect to the Mediterranean and 0.0723 (95% CI: 0.0660, 0.0795) with respect to the South-eastern Pacific. As expected, a lower value was obtained for Mediterranean with respect to the South-eastern Pacific F_{ST} statistic (0.0134) (95% CI: 0.0070, 0.0207). This genetic

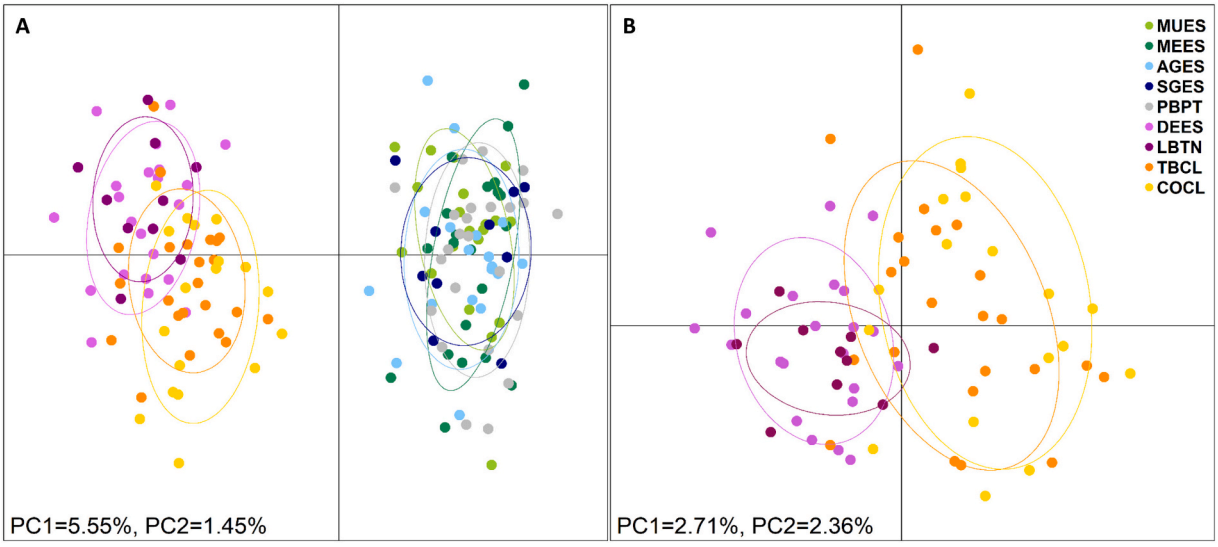


Fig. 2. Principal Component analysis (PCA) of mussel samples. (A) Individuals from all areas (Dataset 1): Atlantic (● MUES |● MEES|● SGES|● AGES |● PBPT), Mediterranean (● DEES|● LBTN) and South-eastern Pacific (● COCL|● TBCL). (B) Mediterranean (● DEES|● LBTN) and South-eastern Pacific (● COCL|● TBCL) individuals (Dataset 2). Each point represents one sample and colours denote its sampling site. Ovals represent 95% inertia ellipses.

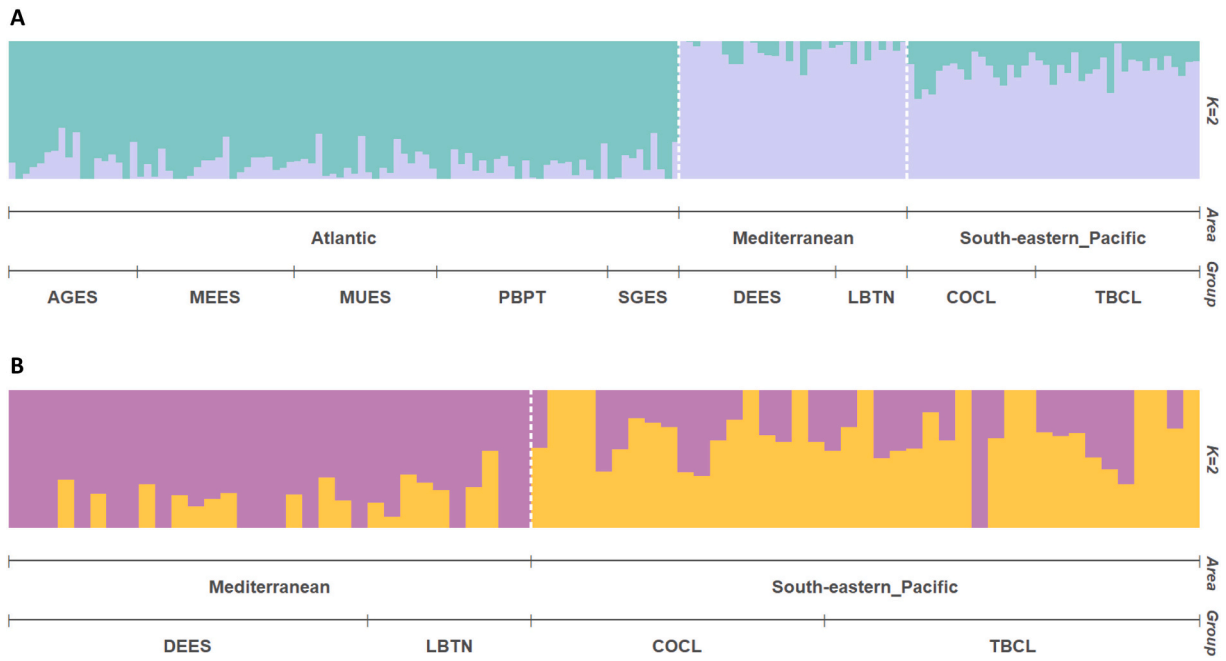


Fig. 3. Graphical representation of ADMIXTURE clustering approach, where each bar represents an individual and each colour its inferred membership to each of the two potential ancestral populations (K = 2). (A) Dataset 1 including all areas: Atlantic (MUES|MEES|SGES|AGES |PBPT), Mediterranean (DEES|LBTN) and South-eastern Pacific (COCL|TBCL). (B) Dataset 2 including only the Mediterranean (DEES|LBTN) and South-eastern Pacific individuals (COCL|TBCL). See Fig. S3 for K = 3.

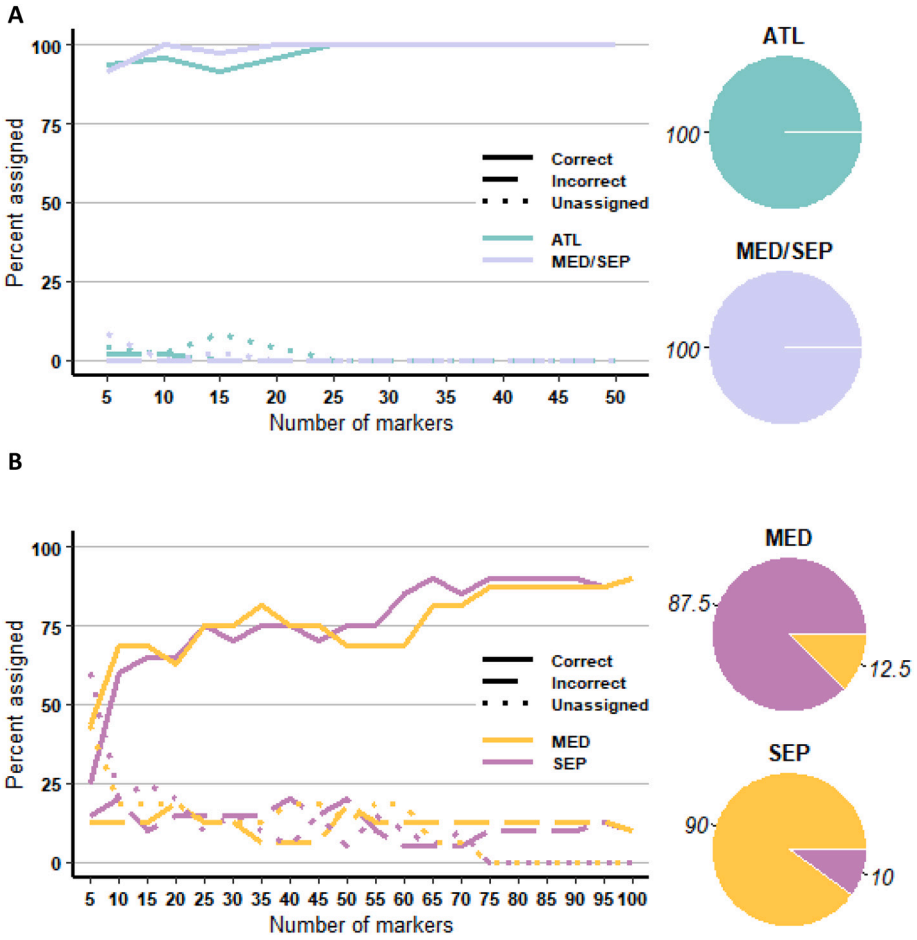


Fig. 4. Origin assignment of individuals following THL method. (Left) Progression of the percentage of correctly, incorrectly and unassigned Atlantic (ATL), Mediterranean (MED) and South-eastern Pacific (SEP) samples as the number of markers (SNPs) increases. (Right) Percentages of samples assigned to their origin using a 25 SNP panel for Dataset 1 (A) and a 75 SNP panel for Dataset 2 (B). Markers were selected and ranked according to their F_{ST} value. The assignment threshold used was 80% in both cases, see Fig. S2 for 90% assignment scores.

Table 2
Pairwise F_{ST} values matrix for all mussel farming points (using Dataset 1): Atlantic (MUES|MEES|SGES|AGES|PBPT), Mediterranean (DEES|LBTN) and South-eastern Pacific (COCL|TBCL). Values significantly different from zero at 95% CI are in bold.

	MUES	MEES	SGES	AGES	PBPT	DEES	LBTN	COCL
MEES	−0.0003							
SGES	0.0006	0.0020						
AGES	−0.0018	0.0006	−0.0021					
PBPT	0.0035	0.0020	0.0044	0.0023				
DEES	**0.0934**	**0.1005**	**0.1025**	**0.0935**	**0.1030**			
LBTN	**0.0873**	**0.0966**	**0.1016**	**0.0953**	**0.1020**	0.0047		
COCL	**0.0615**	**0.0664**	**0.0694**	**0.0656**	**0.0711**	0.0127	0.0139	
TBCL	**0.0712**	**0.0781**	**0.0773**	**0.0739**	**0.0823**	**0.0138**	**0.0171**	−0.0017

differentiation was also reflected in the pairwise F_{ST} estimations between the nine mussel sampling points (Table 2). Although F_{ST} values between the Easten Pacific COCL location and the Mediterranean populations were not significantly different from zero at 95% CI when using Dataset 1, a clear and significant genetic differentiation was observed when using the specific dataset developed for these two areas (Dataset 2) (Table S3). Outlier analyses run by pcadapt detected 40 and 9 SNPs showing clear evidence of positive selection (pvalue<0.05) for Dataset 1 and 2 respectively.

3.3. Traceability tool development

With the aim of developing a traceability SNP panel with the highest accuracy and lowest number of SNPs, the 100 most discriminant SNPs were selected using half of the individuals (Table S4). The remaining half, excluded from the SNP selection, were assigned using Geneclass2. The assignment was performed combining both, neutral and outlier loci information; being under positive selection 18 out of 100 SNPs for dataset 1 and 1 out of 100 SNPs for dataset 2 (Table S4). For Dataset 1, the percentage of correctly assigned samples when using 10 to 25 SNPs ranked based on F_{ST} criterion, resulted in between 90 and 100% of assignment for Atlantic and Mediterranean/South-eastern Pacific samples (Fig. 4A, Fig. S2). For Dataset 2, assignment success to the region of origin using a panel of 75 SNPs ranked according to de F_{ST} value resulted in 87.5% and 90% correct assignment of the Mediterranean and South-eastern Pacific origin samples, respectively (Fig. 4B, Fig. S2).

4. Discussion

4.1. Population structure of M. galloprovincialis

The analysis of the *M. galloprovincialis* from nine locations performed here has provided a robust population genetic structure of this species in part of its both native and introduced range. This has allowed the clear differentiation between two mussel lineages within *M. galloprovincialis*, the Atlantic and the Mediterranean, with significant differences observed between the ancestral Mediterranean native and the more recently introduced mussel in the South-eastern Pacific area.

The genetic differentiation between Atlantic and Mediterranean populations is concordant with previous studies performed along the Iberian Peninsula using allozymes (Quesada et al., 1995b), mitochondrial markers (Quesada et al., 1995a; Kijewski et al., 2011) and microsatellites (Diz and Presa, 2008). These studies place the Iberian phylogeographic break between Almeria and Alicante, which corresponds with the position of the well-defined Almeria-Oran oceanographic front (AOOF). The concordance between these four different marker types strongly supports genetic differentiation between the two Iberian *M. galloprovincialis* populations and proves the temporal stability of this structure, allowing the development of accurate geographic origin verification tools. The differentiation between these two lineages using SNP markers has also being proved when analysing a set of reference Northern hemisphere *M. galloprovincialis* individuals from Oristano (Italy) and Punta Camarinal (Spain) (Larrain et al., 2018; Zbawicka et al., 2018; Zbawicka et al., 2019; Popovic et al., 2020). The use of thousands of SNP markers throughout the genome, has increased the resolution in population structure revealing high levels of genetic differentiation (F_{ST} = 0.0867). Among Atlantic samples, our findings contradict the large proportion of significant pairwise F_{ST} values found when using six microsatellites (Diz and Presa, 2008), since no significant differences were depicted. This discrepancy could be explained by the higher number of genetic markers used in our study, which might provide a better resolution of the genetic differentiation.

We found that Mediterranean samples grouped closer to the South-eastern Pacific samples than to the Atlantic samples, which is explained by a relatively recent human-mediated introduction of Mediterranean *M. galloprovincialis* mussels on the coast of Chile (Westfall and Gardner, 2010; Tarifeño et al., 2012; Larrain et al., 2018). The presence of Mediterranean populations in South America has also recently been described by Lins et al. (2021), being genetically more similar to the Black Sea and South-eastern Mediterranean populations. Nevertheless, although originated from Mediterranean ancestral population, our analyses support significant differentiation between current Mediterranean and South American populations. This indicates that these populations have diverged over the years and are now genetically differentiated. Interestingly, Larrain et al. (2018) found an F_{ST} value of zero between Mediterranean (from Oristano, Italy) and South-eastern Pacific samples (from BioBio, Chile), potentially due to the fact that this previous study used lower number of SNP markers. The thousands of SNP markers developed here specifically for *M. galloprovincialis* mussels may have uncovered an existing genetic differentiation between these two populations. Within the Mediterranean samples, the complete lack of differentiation found agrees with the genetic structure observed between *M. galloprovincialis* mussels from Western Mediterranean by Paterno et al. (2019), where no differences were found between samples from the Western and Central Mediterranean.

4.2. SNP markers for M. galloprovincialis geographical origin assignment

Once the overall picture of *M. galloprovincialis* population structure has been settled, the geographic origin assignment power of SNP markers presented here paves the way for an accurate and cost-effective traceability tool development. Due to the high genetic differentiation observed between Atlantic and Mediterranean *M. galloprovincialis*, our study has shown that a correct assignment between 90 and 100% can be obtained when using 10 to 25 SNPs. As in our study, small SNP panels with high assignment power have previously shown to be useful for genetic traceability systems of other marine species. Proof of this are the small panels of SNP markers (n = 10–30) that were developed for the traceability of European commercial fish (cod, herring, sole and hake) with 93–100% of correct assignment (Nielsen et al., 2012; Ogden and Linacre, 2015). In this sense, Montes et al. (2017) also detailed the use of a SNPs-based test to determine the provenance of anchovies with an overall assignment percentage of 94.3% of the individuals analysed.

As expected, our SNP traceability panel was less effective distinguishing between mussels from the Mediterranean and South-eastern Pacific areas, which is consistent with the patterns observed in the

population structure analysis. A possible explanation could be the insufficient accumulation of genetic differences, either by genetic drift or adaptation to the local environment between these two populations, due to the relatively recent introduction of the population in the South-eastern Pacific. Nevertheless, the assignment score for the Mediterranean and South-eastern Pacific samples could be improved by increasing the number of samples used for SNP selection.

Previous origin assignment analysis of mussels in South America, where Atlantic and Mediterranean *M. galloprovincialis* samples were used as reference to identify species within *Mytilus* genus, observed lower assignment rates between these two European locations using higher number of SNPs (Larrain et al., 2018; Zbawicka et al., 2018). Larrain et al. (2018) used a 49-SNP panel to assign the Atlantic and Mediterranean populations with an 86.2% and 93.1% assignment success, respectively, suggesting mixture of individuals between these two European locations. Zbawicka et al. (2018) used a 51-SNP panel to assign Atlantic and Mediterranean samples with an 89.65% and 96.55% assignment success. In both studies, only two sampling sites were selected around the Iberian Peninsula, having 29 samples each of them. This lower assignment power can be explained by the fact that the SNPs selected for the panels aimed to identify the different *Mytilus* species located in the Southern Hemisphere and not to differentiate lineages within Northern Hemisphere *M. galloprovincialis*. As shown in our study, the use of a specific SNP panel developed precisely for origin assignment of *M. galloprovincialis,* shows a better performance in assigning the geographical origin than using SNPs shared by different species.

In summary, SNP panels represent a promising traceability tool for food authorities due to their easy laboratory transfer, their potential assignment power and their applicability in processed seafood products. However, it is worth mentioning that the spatial resolution of DNA based methods is often limited, since it is not possible to differentiate samples from different farming locations within the same geographical area. In these cases, a multidisciplinary approach combining DNA based techniques with other techniques that might reflect individuals' surroundings, such as isotope or trace element analysis, would improve the resolution determining the geographic location. In addition, the origin assignment markers presented here could also help to determine the source of invasion of this species in non-native areas and to understand the subsequent impacts in local shellfish production. Consequently, this will allow the development of management strategies to mitigate future invasions, ensuring the conservation of native biodiversity.

Our research supports the possibility of the development of an accurate and cost-effective origin assignment tool based on a reduced panel of SNPs which can be applied to fresh, frozen or processed mussel products. Developing genetic tools to assign seafood to their provenance will protect marine resources, guarantee sustainable aquaculture management, enforce current food safety regulations and prevent commercial fraud.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aquaculture.2021.737836.

Data accessibility

Demultiplexed and quality filtered RAD-tags used in this study are available at the National Center of Biotechnology Information (NCBI) Sequence Read Archive (SRA) with BioProject accession number PRJNA786797.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Our study was carried out thanks to the financial support of the SEA-TRACES (EAPA_87/2016) INTERREG Atlantic Area EU funded project and ORIME FEMP Project (00001-1NA2018-33) from the Department of Aquaculture and Fisheries from the Basque Government. Ane del Rio-Lavín is the recipient of a PhD grant from the Education Department of the Basque Government. Samples were kindly provided by IIM-CSIC (Spain), Consello Regulador Mexillón de Galicia (Spain), Fepromodel (Spain) and Universidad de Chile (Chile). This manuscript is contribution number 1083 from the Food Research Division of AZTI.

References

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664.
Anderson, E.C., 2010. Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. Mol. Ecol. Resour. 10, 701–710.
Araneda, C., Larrain, M.A., Hecht, B., Narum, S., 2016. Adaptive genetic variation distinguishes Chilean blue mussels (*Mytilus chilensis*) from different marine environments. Ecol. Evolution 6, 3632–3644.
Astorga, M.P., Cardenas, L., Vargas, J., 2015. Phylogenetic approaches to delimit genetic lineages of the *Mytilus* Complex of South America: how many species are there? J. Shellfish Res. 34 (919–930), 912.
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. 57, 289–300.
Bronner, I.F., Quail, M.A., Turner, D.J., Swerdlow, H., 2014. Improved protocols for Illumina sequencing. Curr Protoc Hum Genet 80, 1–42.
Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22, 3124–3140.
del Rio-Lavín, A., Jiménez, E., Pardo, M.Á., 2021. SYBR-Green real-time PCR assay with melting curve analysis for the rapid identification of *Mytilus* species in food samples. Food Control 130, 108257.
Dias, P.J., Sollelis, L., Cook, E.J., Piertney, S.B., Davies, I.M., Snow, M., 2008. Development of a real-time PCR assay for detection of *Mytilus* species specific alleles: application to a sampling survey in Scotland. J. Exp. Mar. Biol. Ecol. 367, 253–258.
Diz, A.P., Presa, P., 2008. Regional patterns of microsatellite variation in *Mytilus galloprovincialis* from the Iberian Peninsula. Mar. Biol. 154, 277–286.
Etter, P.D., Bassham, S., Hohenlohe, P.A., Johnson, E.A., Cresko, W.A., 2011. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. In: Orgogozo, V., Rockman, M.V. (Eds.), Molecular Methods for Evolutionary Genetics, pp. 157–178.
EUMOFA, 2019. The EU Fish Market. Publications office of the European Union, Luxembourg.
FAO, 2020. FAO yearbook. Fishery and Aquaculture Statistics 2018/ FAO annuaire. Statistiques des pêches et de l'aquaculture 2018/ FAO anuario. Estadísticas de pesca y acuicultura 2018. Roma/Rome.
Fraisse, C., Belkhir, K., Welch, J.J., Bierne, N., 2015. Local interspecies introgression is the main cause of extreme levels of intraspecific differentiation in mussels. Mol. Ecol. 25, 269–286.
Gardner, J.P., Zbawicka, M., Westfall, K.M., Wenne, R., 2016. Invasive blue mussels threaten regional scale genetic diversity in mainland and remote offshore locations: the need for baseline data and enhanced protection in the Southern Ocean. Glob. Chang. Biol. 22, 3182–3195.
Helyar, S.J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M.I., Ogden, R., Limborg, M.T., Cariani, A., et al., 2011. Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. Mol. Ecol. Resour. 11 (Suppl. 1), 123–136.
Jombart, T., Ahmed, I., 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27, 3070–3071.
Keenan, K., McGinnity, P., Cross, T.F., Crozier, W.W., Prodöhl, P.A., O'Hara, R.B., 2013. diveRsity: an Rpackage for the estimation and exploration of population genetics parameters and their associated errors. Methods Ecol. Evol. 4, 782–788.
Kijewski, T., Śmietanka, B., Zbawicka, M., Gosling, E., Hummel, H., Wenne, R., 2011. Distribution of *Mytilus* taxa in European coastal areas as inferred from molecular markers. J. Sea Res. 65, 224–234.
Larrain, M.A., Zbawicka, M., Araneda, C., Gardner, J.P.A., Wenne, R., 2018. Native and invasive taxa on the Pacific coast of South America: impacts on aquaculture, traceability and biodiversity of blue mussels (*Mytilus spp.*). Evol. Appl. 11, 298–311.
Lins, D.M., Zbawicka, M., Wenne, R., Poćwierz-Kotus, A., Molina, J.R.A., Alves, L.P., Rocha, R.M., 2021. Ecology and genetics of Mytilus galloprovincialis: a threat to bivalve aquaculture in southern Brazil. Aquaculture 540, 736753.
Lischer, H.E., Excoffier, L., 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28, 298–299.
Luu, K., Bazin, E., Blum, M.G.B., 2016. Pcadapt : an R package to perform genome scans for selection based on principal component analysis. Mol. Ecol. Resour. 17.
Martinsohn, J.T., Ogden, R., 2009. FishPopTrace—Developing SNP-based population genetic assignment methods to investigate illegal fishing. Forensic Science International: Genetics Supplement Series 2, 294–296.
Montes, I., Laconcha, U., Iriondo, M., Manzano, C., Arrizabalaga, H., Estonba, A., 2017. Reduced single nucleotide polymorphism panels for assigning Atlantic Albacore and

Bay of Biscay Anchovy individuals to their geographic origin: toward sustainable fishery management. J. Agric. Food Chem. 65, 4351–4358.
Morin, P.A., Luikart, G., Wayne, R.K., Group, S.W., 2004. SNPs in ecology, evolution and conservation. Trends Ecol. Evol. 19, 208–216.
Nielsen, E.E., Cariani, A., Mac Aoidh, E., Maes, G.E., Milano, I., Ogden, R., Taylor, M., et al., 2012. Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. Nat. Commun. 3, 851.
Ogden, R., Linacre, A., 2015. Wildlife forensic science: a review of genetic geographic origin assignment. Forensic Sci. Int. Genet. 18, 152–159.
Paris, J.R., Stevens, J.R., Catchen, J.M., 2017. Lost in parameter space: a road map for stacks. Methods Ecol. Evol. 8, 1360–1373.
Paterno, M., Bat, L., Souissi, J.B., Boscari, E., Chassanite, A., Congiu, L., Guarnieri, G., et al., 2019. A genome-wide approach to the phylogeography of the mussel *Mytilus galloprovincialis* in the Adriatic and the Black Seas. Front. Mar. Sci. 6.
Piry, S., Alapetite, A., Cornuet, J.M., Paetkau, D., Baudouin, L., Estoup, A., 2004. GENECLASS2: a software for genetic assignment and first-generation migrant detection. J. Hered. 95, 536–539.
Popovic, I., Matias, A.M.A., Bierne, N., Riginos, C., 2020. Twin introductions by independent invader mussel lineages are both associated with recent admixture with a native congener in Australia. Evol. Appl. 13, 515–532.
Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., et al., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575.
Quesada, H., Beynon, C.M., Skibinski, D.O., 1995a. A mitochondrial DNA discontinuity in the mussel *Mytilus galloprovincialis Lmk*: pleistocene vicariance biogeography and secondary intergradation. Mol. Biol. Evol. 12, 521–524.
Quesada, H., Zapata, C., Alvarez, G., 1995b. A multilocus allozyme discontinuity in the mussel *Mytilus galloprovincialis*: the interaction of ecological and life-history factors. Mar. Ecol. Prog. Ser. 116, 99–115.
Rannala, B., Mountain, J.L., 1997. Detecting immigration by using multilocus genotypes. Proc. Natl. Acad. Sci. U. S. A. 94, 9197–9201.
Raymond, M., Rousset, F., 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J. Hered. 86, 248–249.
Rousset, F., 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol. Ecol. Resour. 8, 103–106.
Scarano, D., Rao, R., 2014. DNA markers for food products authentication. Diversity 6, 579–596.
Tarifeño, E., Galleguillos, R., Llanos-Rivera, A., Arriagada, D., Ferrada, S., Canales-Aguirre, C.B., Seguel, M., 2012. Erroneous identification of the mussel, *Mytilus galloprovincialis* (Lamarck 1819) as the specie, *Mytilus chilensis* (Hupe 1854) in the Bay of Concepcion, Chile. Gayana (Concepción) 76, 167–172.
Toro, J.E., Ojeda, J.A., Vergara, A.M., Castro, G.C., Alcapan, A.C., 2005. Molecular characterization of the Chilean blue mussel (*Mytilus chilensis* Hupe 1854) demonstrates evidence for the occurrence of *Mytilus galloprovincialis* in southern Chile. J. Shellfish Res. 24, 1117–1121.
Wenne, R., 2018. Single nucleotide polymorphism markers with applications in aquaculture and assessment of its impact on natural populations. Aquat. Living Resour. 31, 2.
Westfall, K.M., Gardner, J.P.A., 2010. Genetic diversity of Southern hemisphere blue mussels (Bivalvia: Mytilidae) and the identification of non-indigenous taxa. Biol. J. Linn. Soc. 101, 898–909.
Zardi, G.I., McQuaid, C.D., Jacinto, R., Lourenço, C.R., Serrão, E.A., Nicastro, K.R., 2018. Re-assessing the origins of the invasive mussel *Mytilus galloprovincialis* in southern Africa. Mar. Freshw. Res. 69, 607.
Zbawicka, M., Drywa, A., Śmietanka, B., Wenne, R., 2012. Identification and validation of novel SNP markers in European populations of marine *Mytilus* mussels. Mar. Biol. 159, 1347–1362.
Zbawicka, M., Trucco, M.I., Wenne, R., 2018. Single nucleotide polymorphisms in native south American Atlantic coast populations of smooth shelled mussels: hybridization with invasive European *Mytilus galloprovincialis*. Genet. Sel. Evol. 50.
Zbawicka, M., Gardner, J.P.A., Wenne, R., 2019. Cryptic diversity in smooth-shelled mussels on Southern Ocean islands: connectivity, hybridisation and a marine invasion. Front. Zool. 16, 32.