Food Research International 62 (2014) 104–110

Contents lists available at ScienceDirect

Food Research International

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

Traceability of mussel (*Mytilus chilensis*) in southern Chile using microsatellite molecular markers and assignment algorithms. Exploratory survey

María Angélica Larraín [a,*], Nelson F. Díaz [b], Cármen Lamas [b], Carla Uribe [a], Cristián Araneda [b]

[a] Universidad de Chile, Facultad de Ciencias Químicas y Farmacéuticas, Departamento de Ciencia de los Alimentos y Tecnología Química. Santiago, Chile.
[b] Universidad de Chile, Facultad de Ciencias Agronómicas, Departamento de Producción Animal. Santiago, Chile.

ARTICLE INFO

Article history:
Received 31 August 2013
Accepted 22 February 2014
Available online 1 March 2014

Keywords:
Mytilus
Mussels
Traceability
Microsatellites
Assignment test

ABSTRACT

The international seafood trade has adopted the food chain or "from farm to fork" concept in terms of standards and regulations regarding food quality, safety and authenticity, from primary production to the consumer. This has led to an increasing need for traceability, but administrative traceability systems (physical labeling, information recording and automatic data treatment) are not flawless and require validation through analytical procedures. Currently, DNA-based methods used for species identification and population genetics, coupled with allocation algorithms can be used to verify administrative traceability systems. We evaluated the potential of a panel of nine microsatellite markers combined with allocation algorithms for their ability to assign *Mytilus* individuals from southern Chile to their geographical origin, evaluating the performance of four assignment methods: genetic distance and frequency-based criteria and a Bayesian based method using prior information or not. The reallocation test showed that the Bayesian method with prior information performed best. When tested with a real traceability verification case, the frequency-based algorithm showed the best results, re-allocating individuals to their original population at least 6 times more often than individuals from other locations in a challenging scenario with low genetic differentiation among locations. In order to apply this allocation method for traceability purposes, it would be necessary to strengthen this SSR panel with more informative loci and complement it with SNP markers.

1. Introduction

Marine mussels are one of the most cultivated and marketed bivalves, with the *Mytilus* genus used widely in prepared food products. Mussel culture in southern Chile is an important economic activity, yielding 14.9% of the world's *Mytilidae* aquaculture production in 2010 (FAO, 2012). Most of our national landings are processed then destined for export.

Currently, availability and international trade of seafood are based on *Codex Alimentarius* provisions and other food standards related to quality, safety, sustainability, and regulations (CE.Nº104/2000; CE.Nº178/2002). These standards incorporate the food chain or "from farm to fork" concept (Moretti, Turchini, Bellagamba, & Caprino, 2003). Compliance with these standards and regulations has been enforced via several administrative traceability systems, including physical labeling, recording of information and automatic data entry to identify, among other elements, the components, origin, dates, transformations and processing conditions that the product underwent throughout the food chain.

DNA-based methods can verify the accuracy of traditional identification methods such as product labeling, determining the species present in foodstuffs including derived products and allowing for identification of different taxonomic levels, from single individuals to breeds, local populations and species, along the food chain (Negrini et al., 2008). Seafood traceability encompasses three levels (Ogden, 2008): identification of the species, geographical origin determination, and supply chain tracking and tracing. In order to develop DNA-based traceability in the *Mytilus* species, questions for at least the first two levels must be addressed: *What species is it?* and *Where was it caught?*. Next, suitable molecular markers must be selected, along with clustering and assignment methods to apply them, and then it must be evaluated whether the quality of these outputs shows an acceptable level of certainty for further application (Ogden, 2008).

Genetic traceability is based on genome studies, employing different DNA markers, that is, physically identifiable locations within a chromosome called "loci" (singular "locus"). These loci may be located in expressed regions of DNA (Woolaver, Nichols, Morton, & Stutchbury, 2013), or more often, in DNA segments with unknown coding function (Collard, Jahufer, Brouwer, & Pang, 2005; Hayes, Sonesson, & Gjerde, 2005). DNA markers show allelic variants that result in polymorphism-producing DNA mutations. These variants can be identified by molecular

* Corresponding author. Tel.: +56 2 2978 16 36. Address: Sergio Livingstone 1007, Independencia-8380492. Santiago, Chile.E-mail: mlarrain@uchile.cl (M.A.Larraín).
E-mail address: mlarrain@uchile.cl (M.A. Larraín).

http://dx.doi.org/10.1016/j.foodres.2014.02.016
0963-9969/

techniques used to classify the genotype, and these techniques can be applied to identify individuals, populations, species, or groups of interest. The use of DNA markers for traceability purposes has been explored for foodstuffs (Dalvit, De Marchi, & Cassandro, 2007; Schwägele, 2005) and seafood (Bossier, 1999). In mussels, the first level of traceability is currently fulfilled by genetic species identification methods, employed to investigate commercial fraud by species substitution (Fernández-Tajes et al., 2011; Santaclara et al., 2006). Inoue, Waite, Matsuoka, Odo, and Harayama (1995) developed a PCR-length polymorphism analysis that allowed for genetic identification of three *Mytilus* species (*Mytilus edulis*, *Mytilus trossulus* and *Mytilus galloprovincialis*), using a region of polyphenolic adhesive protein. Later, Santaclara et al. (2006) added a step based on PCR-RFLP (restriction fragment length polymorphism) analysis of the same gene, in order to differentiate between *Mytilus chilensis* and *M. galloprovincialis*. Furthermore, RAPD (random amplified polymorphic DNA) (Rego et al., 2002), FINS (forensically informative nucleotide sequencing) (Santaclara et al., 2006), and AFLP (amplified fragment length polymorphism) (Groenenberg et al., 2011) methodologies have been employed for *Mytilus* species identification, although to a lesser extent.

Geographical origin of the mussels, the second level of traceability, cannot be established merely by identifying the species, due to the worldwide distribution of *Mytilus* (Gérard, Bierne, Borsa, Chenuil, & Féral, 2008; Westfall & Gardner, 2013; Wonham, 2004) and hybrid zones (Dias et al., 2009; Gardner & Westfall, 2012; Kijewski, Wijsman, Hummel, & Wenne, 2009). Determining the geographical origin of the sample is equivalent to identifying its biological population (Ogden, 2008). To this end, various chemical markers (trace elements coupled with chemometric analysis) (Costas-Rodríguez, Lavilla, & Bendicho, 2010), proximate composition, mineral, free amino acids, fatty acids, and volatile compound analysis (Fuentes, Fernández-Segovia, Escriche, & Serra, 2009) have been used to characterize mussels from different Spanish regions, showing different degrees of correlation with geographical origin. DNA-based methods using different molecular markers have also been used for this purpose. These methods require detailed knowledge of the genetic diversity and population structure in a given geographical area in order to assign an individual to its source population (Blohm et al., 2007). To study genetic diversity and population structure in mussels, various DNA markers have been used, such as allozymes (Koehn, Milkman, & Mitton, 1976; Skibinski, Beardmore, & Cross, 1983), sequencing of the mitochondrial *cytochrome c oxidase subunit I* (*COI*) gene (Ghabooli et al., 2013), ISSR (inter simple sequence repeat) (Varela, González-Tizón, Mariñas, & Martínez-Lage, 2007), SNP (single-nucleotide polymorphism) (Zbawicka, Drywa, Śmietanka, & Wenne, 2012), and microsatellite or SSR (simple sequence repeats). SSR markers are tandemly repeated motifs of 1 to 6 DNA base pairs, which are abundantly distributed within genomes and usually characterized by a high number of alleles. In mussels, microsatellites have been used extensively in genetic diversity evaluation and population structure assessment by several authors in different geographical areas (Diz & Presa, 2008, 2009; Li, Wang, & Bai, 2009; Shields, Heath, & Heath, 2010; Wei, Wood, & Gardner, 2013a, 2013b).

Recent advances in statistical and mathematical methods for assigning individuals to populations, combined with hardware resources supporting computationally-demanding analysis, have yielded promising tools. Assignment methods use genetic information to ascertain population or cluster membership of individuals. Computational methods have been developed specifically to assign individuals to predefined (baseline populations) or non-predefined (clusters) categories (Manel, Gaggiotti, & Waples, 2005). These methods have been implemented using frequentist or likelihood methods (maximum likelihood or Bayesian methods). Frequentist approaches use statistical hypothesis testing and give a *p*-value derived from a predefined or simulated frequency distribution. Likelihood methods assume that observed data arises from a probabilistic model with unknown parameters and use data to estimate those parameters, in order to assess the degree of uncertainty associated with these estimates (Manel et al., 2005; Negrini et al., 2008). Another approach is to use Bayesian inference that derives the posterior distribution, taking into account prior probability and a "likelihood function" derived from the observed data, allowing for incorporation of existing knowledge into data analysis (Beaumont & Rannala, 2004).

At present, to determine the geographic origin in *Mytilus*, as well as other marine (cod, herring, sole, hake, gilthead sea bream, blue marlin, golden humped tench, and Asian sea bas) (De Innocentiis et al., 2005; Glover et al., 2010; Lo Presti, Kohlmann, Kersten, Gasco, & Di Stasio, 2010; Martinsohn & Ogden, 2009; Nielsen et al., 2012; Sorenson, McDowell, Knott, & Graves, 2013; Stewart, James, Roden, & Dutton, 2013; Yue et al., 2012) and terrestrial species (cattle, wild boars and pigs) (Caratti et al., 2010; Negrini et al., 2007; Negrini et al., 2008; Ramos, Megens, Crooijmans, Schook, & Groenen, 2011), molecular markers coupled with assignment methods are becoming strong traceability tools.

The current study was designed to evaluate the potential of microsatellite markers combined with allocation algorithms for assigning *Mytilus* individuals from southern Chile to their geographical origin, as a tool for verifying and validate administrative traceability systems.

2. Materials and methods

2.1. Mussel sample collection and SSR genotyping

Samples of mussels (n = 50 by location) were collected in southern Chile in six sites from 41°31′S; 72°20′W to 50°50′S; 74°00′W (Fig. 1), including one wild population and five seed collection centers. In these collection centers, larvae from the natural environment are fixed to mesh collectors to provide the *Mytilus* industry with seeds. All samples were collected sub-tidally. Shell size of 15–25 mm was considered as seed, and larger sizes as adult (Table 1). Approximately 50–100 mg of fixed mantle edge tissue was used for DNA extraction by the phenol–chloroform method, adapted to mussels as described in Larraín, Díaz, Lamas, Vargas, and Araneda (2012).

Nine SSR loci were assayed; for five of these, primers were taken from published sources: *Mgu1* and *Mgu3* (Presa, Pérez, & Diz, 2002), *MT203* and *MT282* (Gardeström, Pereyra, & André, 2007) and *Med737* (Lallias, Stockdale, Boudry, Lapegue, & Beaumont, 2009). Primers for the other four loci: *Mg15* (AY102082, F: ATCATGGTCTCACGGCAACCTT and R: GCGTCGAAACAGTGTTAGTCAACC), *Mg17* (AY102084, F: CCACTCAAAC GGTAGGAGGTTTCA and R: GGGCGGTTTAGCTGATGTTGGTA), *Mg56* (AF445374.1, F: GACCCCATACCAAAGCAAAC and R: TCGCCTTAGCAGGG TGAAAT) and *MIT02* (EH663073, F: GGACCATCTGCGTAGATTCA and R: GGTGGGCGTAGTTATTTCTG) were re-designed using the sequences published on the GenBank and AmplifX freeware program (http://crn2m.univ-mrs.fr/recherche/brue/jullien-nicolas/programmation/amplifx). Thermal profile was 95 °C for 5 min, followed by 35 cycles at 95 °C for 1 min, specific Tm for 20–40 s (Table 2), 72 °C for 50 s with a 10 min final extension step at 72 °C. PCR was performed in a total volume of 15 μL containing 1.5 μL of 10× PCR buffer, and standardized conditions for each microsatellite are described in detail in Table 2. A negative control with template DNA replaced with water was performed for each set of amplifications to diminish genotyping errors due to contamination. For genotyping, polyacrylamide gels (6%) with silver staining were used. For every gel, the size of amplified fragments was estimated from 10 bp DNA ladder (Invitrogen®) or Hyperladder V (BioLine®).

2.2. Data analysis

Micro-Checker (Van Oosterhout, Hutchinson, Wills, & Shipley, 2004) was used to test for the presence of null alleles, stuttering and large allele dropout. GENEPOP 4.0.10 (Raymond & Rousset, 1995; Rousset, 2008) was used to test genotypic distributions of conformance to

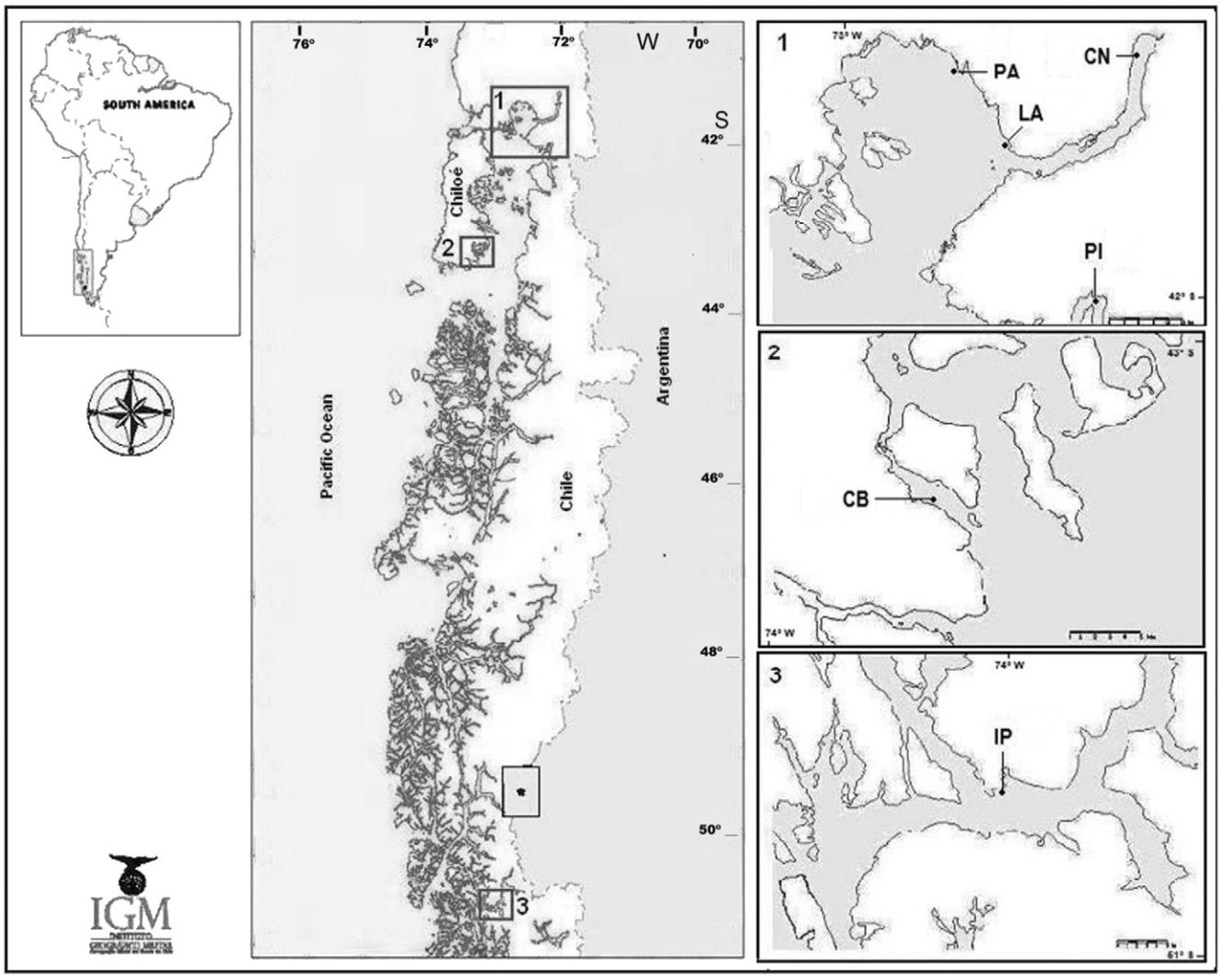

Fig. 1. Location of sampling sites in southern Chile. Sample codes are indicated in Table 1.

Hardy Weinberg (HW) expectations and to estimate Wright's fixation index (F_{ST}) according to Weir and Cockerham (1984).

Four assignment methods were assayed to assign reference populations as possible origin of individuals, based on multilocus genotypes. 1) The genetic distance-based criterion (Cornuet, Piry, Luikart, Estoup, & Solignac, 1999) was used to compute Nei's *Da* genetic distance (Nei, Tajima, & Tateno, 1983) between the individual to be allocated and each reference population, assigning it to the population with the lowest genetic distance. 2) The frequency-based method of Paetkau, Calvert, Stirling, and Strobeck (1995) was used to allocate the individual to the reference population with the most likely allele frequency. 3) The Bayesian-based method of Pritchard, Stephens, and Donnelly (2000) operating in an "unsupervised" mode, in which no prior information about the origin of individuals input into the model was used to simultaneously delineate clusters of individuals based on their multilocus genotypes by maximizing HW and linkage equilibrium between the members and assigning individuals to the identified clusters (Baudouin, Piry, & Cornuet, 2004; Manel et al., 2005). As clusters are

Table 1
Sampling location, source, sampling date and stage of development of mussel samples.

Code	Sampling location South latitude/west longitude	Source	Sampling date	Stage of development
1-PA	Piedra Azul 41° 32′ 55.35″/72° 46′ 14.35″	Seed collection center	24.06.2009	Seed
1-PI	Pichicolo 42° 02′ 23.76″/72° 35′ 27.17″	Seed collection center	24.06.2009	Seed
1-LA	Caleta La Arena 41° 41′ 00.00″″/72° 40′ 18.92″	Seed collection center	24.06.2009	Seed
1-CN	Canutillar 41° 31′ 13.90″/72° 20′ 15.69″	Seed collection center	24.06.2009	Seed
2-CB	Canal Coldita–Piedra Blanca 43° 14′ 48.82″/73° 41′ 42.77″″	Seed collection center	25.06.2009	Seed
3-IP	Isla Peel 50° 50′ 29.83″/74° 00′ 41.27″	Wild population	16.02.2009	Adult

Table 2
PCR conditions of the nine microsatellite loci used in genetic diversity and population structure studies of Chilean mussel (*Mytilus chilensis*).

	Mgu1	*Mgu3*	*MT203*	*MT282*	*Mg15*	*Mg17*	*Mg56*	*Med737*	*MIT02*
Component [concentration]									
MgCl [mM]	1.5	2.0	2.0	2.0	2.0	2.0	2.0	1.5	2.0
dNTP (each) [μM]	100	100	167	100	133	100	100	80	100
Primer F and R (each) [μM]	0.6	0.2	0.3	0.3	0.3	0.5	0.2	0.2	0.4
Taq [U]	0.5	0.3	0.3	0.3	0.4	0.4	0.5	0.4	0.5
DNA [ng]	40	30	30	40	40	40	40	40	40
Annealing conditions									
Temperature [°C]	54	59	59	58	50	60	67	63	53
Time [s]	40	30	20	20	20	30	20	30	30

abstract objects that are not necessarily coincident with sampling sites, allocation was considered correct when the individual was included in the cluster with the highest proportion of individuals from its location. 4) The same Bayesian method operating in a "supervised" mode, was used to incorporate the multilocus genotypes of individuals and prior information about the geographical origin of the samples. In this case, the baseline populations are effectively represented by sampling sites. The first two statistical criteria, genetic distance and allele frequency, were implemented using the software package GeneClass2, assigning a population of origin for each individual with an assignment threshold (p-value of 0.05), using the leave one out procedure (Piry, 2004). The population of origin was defined as that from which the individual was sampled. The probability of assigning an individual to a population was calculated using the likelihood method without associated probabilities, and expressed as a score (Piry, 2004). The Bayesian method was implemented in Structure 2.3.3 (Pritchard et al., 2000), assignments of individuals were conducted under the mixture model, which assumes that individuals were taken from groups that are not pure but rather have some degree of mixing. In both Bayesian methods, a 50,000 initial burn-in was used to minimize the effect of the starting configurations, followed by 100,000 Markov chain iterations; these values, recommended by Falush, Stephens, and Pritchard (2007), were adequate to achieve convergence with our data. In all cases, six repeated runs were performed, selecting the parameter K = 6, the number of locations analyzed.

The performance of all four assignment methods was evaluated in re-allocation assays, carried out by analyzing for each population: i) Sensitivity (S), calculated as the number of individuals correctly assigned to their original location, divided by the total number of individuals sampled from that location. This index reflects how good a test is at correctly assigning individuals to a location. ii) Specificity (E), calculated as the number of individuals properly excluded from the population, divided by the total number of individuals who do not belong to the population. This index reflects how good the test is at correctly excluding individuals who do not belong to that location (Altman & Bland, 1994; Loong, 2003). iii) Average probability assignment score (AP), calculated per location as the average of the likelihood of each successful re-assignment to the respective location. iv) Likelihood ratio (LR+), calculated as S / (1-E). This statistic summarizes how many times more (or less) likely it is that individuals belonging to a location will be assigned to this location as compared to individuals belonging to another location (Deeks & Altman, 2004).

3. Results

Micro-Checker did not detect large allele dropout, but stuttering was suggested for locus *Mgu1* and *Med737* in some sampling sites. In general, there were warnings for the presence of null alleles in all six locations across loci, with the exception of five locations for locus *MT282*, one location for *Med737*, and two locations for *MIT02*. The nine microsatellite loci were polymorphic in all locations, and 68 alleles were detected among the analyzed individuals. Allele frequencies are shown in Table S1 — Supplementary material. Correcting allele frequencies for null alleles among populations of *Mytilus* did not change the outcome of the statistical tests of differentiation (Diz & Presa, 2008). Other authors (Chapuis & Estoup, 2007; Hauser, Seamons, Dauer, Naish, & Quinn, 2006) showed that the presence of null alleles in microsatellite loci had a minor effect on the estimation of population differentiation, genetic distance determination, and assignment of individuals to populations (Carlsson, 2008); consequently, we decided not to correct the allele frequencies for the presence of null alleles.

The global F_{ST} value shows that the genetic differentiation between the six sites was 0.042 and indicates that 4.2% of total allele frequency variance lies among sample sites and is highly significant ($P < 0.001$). Therefore, 95.8% of the variance between allele frequencies is explained by variation within sites.

The genetic distance and the frequency-based methods implemented in GeneClass2 properly re-allocated 152 individuals (50.7%) to their localities of origin (Table 3). The Bayesian-based method implemented in Structure 2.3.3 considering only multilocus genotyping data (unsupervised mode) performed poorly, properly re-assigning only 105 individuals (35.0%). When information on the location of origin of the individuals was taken into account, correct assignment rose sharply to 286 individuals (95.3%) successfully re-allocated to the six sites of origin (Table 3). In general, average sensitivity (S) across allocation algorithms was low, with the exception of the Bayesian method with prior information ($S = 0.95$). Specificity was high for all methods ($E > 0.87$), with the highest value again for the Bayesian method with prior information ($E = 0.99$). The probability of assignment was poor ($PA = 0.19$) for the genetic distance-based method. The other three methods showed values <0.7, with the best performance once again for the Bayesian method using prior information. According to the likelihood ratio index, the Bayesian method in an "unsupervised" mode showed the lowest average LR+, re-allocating individuals to their original population at least 3.7 times more often than individuals from other locations (Table 3). The genetic distance and frequency based methods performed better, showing an average LR+ index of 5.4 and 6.0 respectively.

4. Discussion

Product labeling is the basic information source for determining the species present in a seafood product, identifying its geographical origin upstream in the supply chain (tracing), and following a product through the supply chain (tracking). This information is used by customers to make their choice and by authorities or manufacturers to verify legal compliance and make recalls when necessary. DNA-based techniques can be used to verify the accuracy of the labeling information and to audit administrative traceability systems (Yue et al., 2012). We focused on *Mytilus*, which are already analyzed using genetic species identification methods to investigate commercial fraud by species substitution (Fernández-Tajes et al., 2011; Santaclara et al., 2006). Geographic origin was assessed using a nine-microsatellite panel, along with different allocation algorithms, to test the assignment of individuals to populations in southern Chile.

Table 3
Performance of the assignment methods tested for the nine microsatellite loci in Chilean mussel (*Mytilus chilensis*) individuals from six locations.

Software Criterion Algorithm	Correct allocation			1-PA	1-PI	1-LA	1-CN	2-CB	3-IP	Average
	Number of individuals	(%)								
GeneClass2	152	50.7	S	0.48	0.36	0.80	0.32	0.56	0.52	0.51
Frequency			E	0.88	0.86	0.96	0.83	0.80	0.89	0.87
Paetkau et al. (1995)			AP	0.72	0.62	0.90	0.62	0.75	0.74	0.73
			LR+	4.00	2.57	20.00	1.88	2.80	4.73	6.00
GeneClass2	152	50.7	S	0.54	0.26	0.80	0.28	0.70	0.46	0.51
Genetic distance Nei et al. (1983)			E	0.88	0.89	0.94	0.83	0.73	0.94	0.87
Cornuet et al. (1999)			AP	0.19	0.18	0.20	0.18	0.19	0.19	0.19
			LR+	4.50	2.36	13.33	1.65	2.59	7.67	5.35
Structure 2.3.3	105	35.0	S	0.34	0.26	0.56	0.24	0.42	0.28	0.35
Bayesian without prior information			E	0.86	0.87	0.95	0.84	0.87	0.83	0.87
Pritchard et al. (2000)			AP	0.70	0.70	0.73	0.60	0.67	0.79	0.70
			LR+	2.43	2.00	11.20	1.50	3.23	1.65	3.67
Structure 2.3.3	286	95.3	S	0.92	0.96	1.00	0.94	0.96	0.94	0.95
Bayesian with prior information			E	0.99	1.00	1.00	0.98	0.99	1.00	0.99
Pritchard et al. (2000)			AP	0.94	0.91	0.98	0.91	0.96	0.94	0.94
			LR+	92.00	∞	∞	47.00	96.00	∞	–

S: Sensitivity.
E: Specificity.
AP: Average probability assignment score.
LR+: Likelihood ratio.

Population structure refers to the subdivision of a metapopulation into subgroups of individuals that are on average more related to each other than to other members of a wider population (Balding, 2006) as a result of the processes of genetic drift, natural selection, mutation, and gene flow. Divergent selection and genetic drift promote the generation of genetic structure while genetic flow homogenizes the subpopulations (Jones & Wang, 2012). The power of assignment depends closely on the level of population structure in the zone under study, reflected in levels of differentiation among locations. Wright's F_{ST} statistic is widely used to estimate the proportion of genetic diversity among populations and reflects different degrees of allele frequency divergence among locations, ranging from 0 (no difference in allele frequencies among locations) to 1 (maximum variance in allele frequencies among locations) (Holsinger & Weir, 2009; Meirmans & Hedrick, 2011). Levels of differentiation required for reasonable assignment success are usually cited as F_{ST} values of 0.05–0.10 (Hauser et al., 2006); in our results, global F_{ST} was lower (0.042). For traceability purposes, this finding poses a challenging scenario, as the power of the assignment is directly correlated with the genetic differentiation between populations. However, despite low genetic differentiation between the *Mytilus* sampling sites in southern Chile, our dataset performed better (50.1% correct assignments) than a panel of 13 SSR assayed in blue marlin between more separated locations (Atlantic and Pacific populations) (41.0% correct assignments) (Sorenson et al., 2013), but was overcome by a panel of 7 SSR assayed in golden humped tench between Italian lakes (64–92% correct assignations) (Lo Presti et al., 2010). This finding was possibly attributable to different connectivity conditions between relatively separate lakes and the sea, determining gene flow and larval dispersal. Also a panel of 9 SSR applied to Atlantic cod in Norway farms achieved 75% of correct assignations (Glover et al., 2010) and a panel of 16 SSR loci previously selected by its assignment power and used in farmed Asian sea bass showed 96.7% of correct assignments (Yue et al., 2012). This better results found in farmed fishes are not surprising due to the reported differences in allele frequencies between farms and commercial strains in cultivated fish species, as is the case for Atlantic salmon and rainbow trout (Glover, 2008; Glover, Hansen, & Skaala, 2009; Glover, Skilbrei, & Skaala, 2008). This effect does not occur in *Mytilus* aquaculture, where seeds are collected from the natural environment with no artificial selection.

In general, the weakest assignation method was the Bayesian without prior information, and the best was the Bayesian method including this information. However for traceability purposes, the true membership of individuals is usually the information being verified, and therefore would not be available in real analysis; therefore, this method is not an appropriate alternative for traceability purposes. Bayesian analysis has the potential to incorporate other information to complement genetic data, for example, oceanographic information, chemical markers, or behavioral data, to improve inference when the population structure is weak as with many marine species (Selkoe, Henzler, & Gaines, 2008). But routine use of such methods is not particularly simple, because Bayesian analysis involves numerical solutions using the Markov chain Monte Carlo (MCMC) approach, which is demanding in computation time and resources, and also requires a certain degree of experience to determine the simulation parameters (e.g., the number of initial simulations, or "burn-in period," the number of replicates, and the number of populations) (Baudouin et al., 2004).

Although distance-based criteria are less sensitive to deviations from HW equilibrium due to null allele, this method was not suitable because the average assignment probability was very low (0.19). Genetic distances were developed to differentiate between species, to visualize evolutionary changes over long periods of time based primarily on dismissing mutations and gene flow. Small changes occurring within a species, where the effect of the mutation is low or negligible compared to gene flow, do not allow for proper discrimination among populations of the same species (Flury, Weigend, Ding, Taubert, & Simianer, 2007; Toro, Fernandez, & Caballero, 2009).

In our situation, with six locations of similar sample sizes ($n = 50$), the prevalence or a priori probability of an individual belonging to each group is 17% (1/6). The method with the worst performance (Bayesian without prior information) correctly re-allocated (35%) at least twice the number of individuals expected purely by chance; moreover, the frequency-based method correctly re-assigned 50% of the individuals. This is a promising value considering the existing population structure in the zone, and considering that the assayed microsatellite panel has only nine loci and that assignment success increases with the number of loci. To increase assignment performance to the level required for traceability purposes, the panel must be improved by incorporating more microsatellite loci and adding SNP (single nucleotide polymorphisms) markers. Recently SNPs were described in the *Mytilus* species from European samples (Vera et al., 2010; Zbawicka et al., 2012) and in fish species, also these markers are used to generate forensically-validated panels for geographic origin assignment to investigate illegal fishing in the FishPopTrace project (Martinsohn & Ogden, 2009).

5. Conclusion

The frequency-based algorithm showed the best performance in matching *Mytilus* individuals from southern Chile to their geographical origin, reaching preliminary but promising values of 50% of correct assignment in a challenging scenario with low genetic differentiation among locations. In order to apply this allocation method for traceability purposes, the panel would need to be improved by incorporating more informative loci and complementing it with SNP markers.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.foodres.2014.02.016.

Acknowledgments

This research was supported by CONICYTFONDECYT Grant No. 1130302 and Universidad de Chile — Vicerrectoría de investigación — Domeyko — Alimentos Grant 2007-2010 . The authors thank: Dr. Carlos Varela – Universidad de Los Lagos – for contacts with shellfish cultivation industry, Eugenio Yokota and Julio Bañados from Granja Marina Chauquear, Marcela Cárcamo from Cultivos Qullaipe and Armando Salinas from Aguas del Sur S.A. who kindly allowed sampling in their farms; José Villarroel, for collecting the wild population sample; Andrea Bravo for helping with DNA extraction; and Juan Vidal from Instituto Geográfico Militar, for invaluable assistance with the preparation of maps. Also Dr. Edwar Fuentes who kindly supported in the discussion of classification methodologies..

References

Altman, D., & Bland, M. (1994). Statistics notes: Diagnostic tests 1: Sensitivity and specificity. *BMC*, *308*, 1552.

Balding, D. J. (2006). A tutorial on statistical methods for population association studies. *Nature Reviews Genetics*, *7*(10), 781–791.

Baudouin, L., Piry, S., & Cornuet, J. M. (2004). Analytical Bayesian approach for assigning individuals to populations. *Journal of Heredity*, *95*(3), 217–224.

Beaumont, M.A., & Rannala, B. (2004). The Bayesian revolution in genetics. *Nature Reviews Genetics*, *5*(4), 251–261.

Blohm, D., Bonhomme, F., Carvalho, G., Crosetti, D., Cross, T., Dahle, G., et al. (2007). Genimpact — Evaluation of genetic impact of aquaculture activities on native populations. Final scientific report. *6th Framework plan of the European Commission. A European network (EU contract n. RICA-CT-2005-022802)* (pp. 176).

Bossier, P. (1999). Authentication of seafood products by DNA patterns. *Journal of Food Science*, *64*(2), 189–193.

Caratti, S., Rossi, L., Sona, B., Origlia, S., Viara, S., Martano, G., et al. (2010). Analysis of 11 tetrameric STRs in wild boars for forensic purposes. *Forensic Science International: Genetics*, *4*(5), 339–342.

Carlsson, J. (2008). Effects of microsatellite null alleles on assignment testing. *Journal of Heredity*, *99*(6), 616–623.

CE.Nº104/2000. Reglamento del Consejo por el que se establece la organización común de mercados en el sector de la pesca y acuicultura. In (pp. L 17/22 ES). Diario oficial de las comunidades europeas, 2000.

CE.Nº178/2002. Reglamento del Parlamento Europeo y del Consejo por el que se establecen los principios y los requisitos generales de la legislación alimentaria, se crea la Autoridad Europea de Seguridad Alimentaria y se fijan los procedimentos relativos a la seguridad alimentaria. In (pp. L 31/ 31). Diario Oficial de las Comunidades Europeas, 2002.

Chapuis, M. -P., & Estoup, A. (2007). Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, *24*(3), 621–631.

Collard, B., Jahufer, M., Brouwer, J., & Pang, E. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*, *142*(1), 169–196.

Cornuet, J. -M., Piry, S., Luikart, G., Estoup, A., & Solignac, M. (1999). Comparision of methods employing multilocus genotypes to select or exclude populations of origin of individuals. *Genetics*, *153*, 1989–2000.

Costas-Rodríguez, M., Lavilla, I., & Bendicho, C. (2010). Classification of cultivated mussels from Galicia (Northwest Spain) with European Protected Designation of Origin using trace element fingerprint and chemometric analysis. *Analytica Chimica Acta*, *664*(2), 121–128.

Dalvit, C., De Marchi, M., & Cassandro, M. (2007). Genetic traceability of livestock products: A review. *Meat Science*, *77*(4), 437–449.

De Innocentiis, S., Miggiano, E., Ungaro, A., Livi, S., Sola, L., & Crosetti, D. (2005). Geographical origin of individual breeders from gilthead sea bream (Sparus auratus) hatchery broodstocks inferred by microsatellite profiles. *Aquaculture*, *247*(1–4), 227–232.

Deeks, J. J., & Altman, D.G. (2004). Diagnostic tests 4: Likelihood ratios. *BMJ*, *329*.

Dias, P. J., Dordor, A., Tulett, D., Piertney, S., Davies, I. M., & Snow, M. (2009). Survey of mussel (*Mytilus*) species at Scottish shellfish farms. *Aquaculture Research*, *40*(15), 1715–1722.

Diz, A. P., & Presa, P. (2008). Regional patterns of microsatellite variation in *Mytilus galloprovincialis* from the Iberian Peninsula. *Marine Biology*, *154*(2), 277–286.

Diz, A. P., & Presa, P. (2009). The genetic diversity pattern of *Mytilus galloprovincialis* in Galician Rías (NW Iberian estuaries). *Aquaculture*, *287*(3–4), 278–285.

Falush, D., Stephens, M., & Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Molecular Ecology Notes*, *7*(4), 574–578.

FAO (2012). FishStatJ version 2.0.0 software for fishery statistical time series. Rome: FAO Fisheries and Aquaculture Department, Fisheries information and Statistics Unit.

Fernández-Tajes, J., Longa, A., García-Gil, J., Chiu, Y. -W., Huang, Y. -S., Méndez, J., et al. (2011). Alternative PCR - RFLP methods for mussel *Mytilus* species identification. *European Food Research and Technology*, *233*(5), 791–796.

Flury, C., Weigend, S., Ding, X., Taubert, H., & Simianer, H. (2007). Haplotype kinship for three populations of the Goettingen minipig. *Genetics Selection Evolution*, *39*(2), 159–179.

Fuentes, A., Fernández-Segovia, I., Escriche, I., & Serra, J. A. (2009). Comparison of physicochemical parameters and composition of mussels (*Mytilus galloprovincialis* Lmk.) from different Spanish origins. *Food Chemistry*, *112*(2), 295–302.

Gardeström, J., Pereyra, R. T., & André, C. (2007). Characterization of six microsatellite loci in the Baltic blue mussel *Mytilus trossulus* and cross-species amplification in North Sea *Mytilus edulis*. *Conservation Genetics*, *9*(4), 1003–1005.

Gardner, J. P. A., & Westfall, K. M. (2012). Geographic distribution and molecular identification of a metapopulation of blue mussels (genus *Mytilus*) in northeastern New Zealand. *Journal of Molluscan Studies*, *78*(1), 66–73.

Gérard, K., Bierne, N., Borsa, P., Chenuil, A., & Féral, J. -P. (2008). Pleistocene separation of mitochondrial lineages of *Mytilus* spp. mussels from Northern and Southern Hemispheres and strong genetic differentiation among southern populations. *Molecular Phylogenetics and Evolution*, *49*(1), 84–91.

Ghabooli, S., Zhan, A.B., Sardina, P., Paolucci, E., Sylvester, F., Perepelizin, P. V., et al. (2013). Genetic diversity in introduced golden mussel populations corresponds to vector activity. *PLoS ONE*, *8*(3).

Glover, K. A. (2008). Genetic characterisation of farmed rainbow trout in Norway: Intra- and inter-strain variation reveals potential for identification of escapees. *BMC Genetics*, *9*, 10.

Glover, K. A., Dahle, G., Westgaard, J. I., Johansen, T., Knutsen, H., & Jorstad, K. E. (2010). Genetic diversity within and among Atlantic cod (*Gadus morhua*) farmed in marine cages: A proof-of-concept study for the identification of escapees. *Animal Genetics*, *41*(5), 515–522.

Glover, K. A., Hansen, M. M., & Skaala, O. (2009). Identifying the source of farmed escaped Atlantic salmon (Salmo salar): Bayesian clustering analysis increases accuracy of assignment. *Aquaculture*, *290*(1–2), 37–46.

Glover, K. A., Skilbrei, O. T., & Skaala, O. (2008). Genetic assignment identifies farm of origin for Atlantic salmon Salmo salar escapees in a Norwegian fjord. *ICES Journal of Marine Science*, *65*(6), 912–920.

Groenenberg, D. S. J., Wesselingh, F. P., Rajagopal, S., Jansen, J. M., Bos, M. M., van der Velde, G., et al. (2011). On the identity of broad-shelled mussels (Mollusca, Bivalvia, *Mytilus*) from the Dutch delta region. *Contributions to Zoology*, *80*(2), 95–106.

Hauser, L., Seamons, T. R., Dauer, M., Naish, K. A., & Quinn, T. P. (2006). An empirical verification of population assignment methods by marking and parentage data: Hatchery and wild steelhead (*Oncorhynchus mykiss*) in Forks Creek, Washington, USA. *Molecular Ecology*, *15*(11), 3157–3173.

Hayes, B., Sonesson, A. K., & Gjerde, B. (2005). Evaluation of three strategies using DNA markers for traceability in aquaculture species. *Aquaculture*, *250*(1–2), 70–81.

Holsinger, K. E., & Weir, B.S. (2009). Genetics in geographically structured populations: Defining, estimating and interpreting FST. *Nature Reviews Genetics*, *10*(9), 639–650.

Inoue, K., Waite, J. H., Matsuoka, M., Odo, S., & Harayama, S. (1995). Interspecific variations in adhesive protein sequences of *Mytilus edulis*, *M. galloprovincialis*, and *M. trossulus*. *Biological Bulletin*, *189*(3), 370–375.

Jones, O. R., & Wang, J. L. (2012). A comparison of four methods for detecting weak genetic structure from marker data. *Ecology and Evolution*, *2*(5), 1048–1055.

Kijewski, T., Wijsman, J. W. M., Hummel, H., & Wenne, R. (2009). Genetic composition of cultured and wild mussels *Mytilus* from The Netherlands and transfers from Ireland and Great Britain. *Aquaculture*, *287*(3–4), 292–296.

Koehn, R. K., Milkman, R., & Mitton, J. B. (1976). Population genetics of marine pelecypods. IV. Selection, migration and genetic differentiation in the blue mussel *Mytilus edulis*. *Evolution*, *30*(1), 2–32.

Lallias, D., Stockdale, R., Boudry, P., Lapegue, S., & Beaumont, A.R. (2009). Characterization of ten microsatellite loci in the blue mussel *Mytilus edulis*. *Journal of Shellfish Research*, *28*(3), 547–551.

Larraín, M.A., Díaz, N. F., Lamas, C., Vargas, C., & Araneda, C. (2012). Genetic composition of *Mytilus* species in mussel populations from southern Chile. *Latin American Journal of Aquatic Research*, *40*(4), 1077–1084.

Li, J., Wang, G., & Bai, Z. (2009). Genetic variability in four wild and two farmed stocks of the Chinese freshwater pearl mussel (*Hyriopsis cumingii*) estimated by microsatellite DNA markers. *Aquaculture*, *287*(3–4), 286–291.

Lo Presti, R., Kohlmann, K., Kersten, P., Gasco, L., & Di Stasio, L. (2010). Tinca Gobba Dorata del Pianalto di Poirino: Genetic characterization by microsatellite markers. *Italian Journal of Animal Science*, *9*(4).

Loong, T. -W. (2003). Understanding sensitivity and specificity with the right side of the brain. *BMJ*, *327*(7417), 716–719.

Manel, S., Gaggiotti, O. E., & Waples, R. S. (2005). Assignment methods: Matching biological questions techniques with appropriate techniques. *Trends in Ecology & Evolution*, *20*(3), 136–142.

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*(1), 294–296.

Meirmans, P. G., & Hedrick, P. W. (2011). Assessing population structure: F_{ST} and related measures. *Molecular Ecology Resources*, *11*(1), 5–18.
Moretti, V. M., Turchini, G. M., Bellagamba, F., & Caprino, F. (2003). Traceability issues in fishery and aquaculture products. *Veterinary Research Communications*, *27*(1), 497–505.
Negrini, R., Milanesi, E., Filippini, F., Dunner, S., Leveziel, H., Williams, J., et al. (2007). Traceability of the PGI product "Vitellone Bianco dell' Appennino Centrale" by SNP markers. *Italian Journal of Animal Science*, *6*, 174.
Negrini, R., Nicoloso, L., Crepaldi, P., Milanesi, E., Marino, R., Perini, D., et al. (2008). Traceability of four European Protected Geographic Indication (PGI) beef products using Single Nucleotide Polymorphisms (SNP) and Bayesian statistics. *Meat Science*, *80*(4), 1212–1217.
Nei, M., Tajima, F., & Tateno, Y. (1983). Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution*, *19*(2), 153–170.
Nielsen, E. E., Cariani, A., Mac Aoidh, E., Maes, G. E., Milano, I., Ogden, R., et al. (2012). Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. *Nature Communications*, *3*, 851.
Ogden, R. (2008). Fisheries forensics: The use of DNA tools for improving compliance, traceability and enforcement in the fishing industry. *Fish and Fisheries*, *9*(4), 462–472.
Paetkau, D., Calvert, W., Stirling, I., & Strobeck, C. (1995). Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology*, *4*(3), 347–354.
Piry, S. (2004). GENECLASS2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, *95*(6), 536–539.
Presa, P., Pérez, M., & Diz, A. (2002). Polymorphic microsatellite markers for blue mussels (*Mytilus* spp). *Conservation Genetics*, *3*, 441–443.
Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.
Ramos, A.M., Megens, H. J., Crooijmans, R. P.M.A., Schook, L. B., & Groenen, M.A.M. (2011). Identification of high utility SNPs for population assignment and traceability purposes in the pig using high-throughput sequencing. *Animal Genetics*, *42*(6), 613–620.
Raymond, M., & Rousset, F. (1995). GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, *86*(3), 248–249.
Rego, I., Martínez, A., González-Tizón, A., Vieites, J., Leira, F., & Méndez, J. (2002). PCR technique for identification of mussel species. *Journal of Agricultural and Food Chemistry*, *50*(7), 1780–1784.
Rousset, F. (2008). genepop'007: A complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, *8*(1), 103–106.
Santaclara, F. J., Espiñeira, M., Cabado, A. G., Aldasoro, A., Gonzalez-Lavín, N., & Vieites, J. M. (2006). Development of a method for the genetic identification of mussel species belonging to *Mytilus*, *Perna*, *Aulacomya*, and other genera. *Journal of Agricultural and Food Chemistry*, *54*(22), 8461–8470.
Schwägele, F. (2005). Traceability from a European perspective. *Meat Science*, *71*(1), 164–173.
Selkoe, K. A., Henzler, C. M., & Gaines, S. D. (2008). Seascape genetics and the spatial ecology of marine populations. *Fish and Fisheries*, *9*(4), 363–377.
Shields, J. L., Heath, J. W., & Heath, D.D. (2010). Marine landscape shapes hybrid zone in a broadcast spawning bivalve: Introgression and genetic structure in Canadian west coast *Mytilus*. *Marine Ecology Progress Series*, *399*, 211–223.
Skibinski, D. O. F., Beardmore, J. A., & Cross, T. F. (1983). Aspects of the population genetics of *Mytilus* (Mytilidae; Mollusca) in the British Isles. *Biological Journal of the Linnean Society*, *19*(2), 137–183.
Sorenson, L., McDowell, J. R., Knott, T., & Graves, J. E. (2013). Assignment test method using hypervariable markers for blue marlin (*Makaira nigricans*) stock identification. *Conservation Genetics Resources*, *5*(1), 293–297.
Stewart, K. R., James, M. C., Roden, S., & Dutton, P. H. (2013). Assignment tests, telemetry and tag-recapture data converge to identify natal origins of leatherback turtles foraging in Atlantic Canadian waters. *Journal of Animal Ecology*, *82*(4), 791–803.
Toro, M.A., Fernandez, J., & Caballero, A. (2009). Molecular characterization of breeds and its use in conservation. *Livestock Science*, *120*(3), 174–195.
Van Oosterhout, C., Hutchinson, W. F., Wills, D. P.M., & Shipley, P. (2004). Micro-checker: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, *4*(3), 535–538.
Varela, M.A., González-Tizón, A., Mariñas, L., & Martínez-Lage, A. (2007). Genetic divergence detected by ISSR markers and characterization of microsatellite regions in *Mytilus* mussels. *Biochemical Genetics*, *45*(7–8), 565–578.
Vera, M., Pardo, B. G., Pino-Querido, A., Álvarez-Dios, J. A., Fuentes, J., & Martínez, P. (2010). Characterization of single-nucleotide polymorphism markers in the Mediterranean mussel, *Mytilus galloprovincialis*. *Aquaculture Research*, *41*(10), e568–e575.
Wei, K., Wood, A.R., & Gardner, J. P. A. (2013). Seascape genetics of the New Zealand greenshell mussel: Sea surface temperature explains macrogeographic scale genetic variation. *Marine Ecology Progress Series*, *477*, 107–121.
Wei, K. J., Wood, A.R., & Gardner, J. P. A. (2013). Population genetic variation in the New Zealand greenshell mussel: Locus-dependent conflicting signals of weak structure and high gene flow balanced against pronounced structure and high self-recruitment. *Marine Biology*, *160*(4), 931–949.
Weir, B.S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*, 1358–1370.
Westfall, K., & Gardner, J. A. (2013). Interlineage *Mytilus galloprovincialis* Lmk. 1819 hybridization yields inconsistent genetic outcomes in the Southern hemisphere. *Biological Invasions*, *15*(7), 1493–1506.
Wonham, M. J. (2004). Mini review distribution of the Mediterranean mussel *Mytilus galloprovincialis* and hybrids in the northeast pacific. *Journal of Shellfish Research*, *23*(2), 535–543.
Woolaver, L., Nichols, R., Morton, E., & Stutchbury, B.M. (2013). Population genetics and relatedness in a critically endangered island raptor, Ridgway's Hawk Buteo ridgwayi. *Conservation Genetics*, *14*(3), 559–571.
Yue, G. H., Xia, J. H., Liu, P., Liu, F., Sun, F., & Lin, G. (2012). Tracing Asian seabass individuals to single fish farms using microsatellites. *PLoS ONE*, *7*(12).
Zbawicka, M., Drywa, A., Śmietanka, B., & Wenne, R. (2012). Identification and validation of novel SNP markers in European populations of marine *Mytilus* mussels. *Marine Biology*, *159*(6), 1347–1362.