Comparison of two commercial methods for smooth-shelled mussels (Mytilus spp.) species identification

Cynthia M. Asorey\textsuperscript{a,b}, Felipe Jilberto\textsuperscript{a,c,d}, Ilka Haase\textsuperscript{e,1}, Rainer Schubbert\textsuperscript{e}, María Angélica Larraín\textsuperscript{a,d}, Cristián Araneda\textsuperscript{a,c,7}

\textsuperscript{a} Food Quality Research Center. Universidad de Chile, Santiago, Chile
\textsuperscript{b} Universidad Católica del Norte, Facultad de Ciencias del Mar, Sala de Colecciones Biológicas, Departamento de Biología Marina, Larrondo 1281, Coquimbo, Chile
\textsuperscript{c} Universidad de Chile, Facultad de Ciencias Agronómicas, Departamento de Producción Animal, Avenida Santa Rosa 11315, La Pintana, Santiago, Chile
\textsuperscript{d} Universidad de Chile, Facultad de Ciencias Químicas y Farmacéuticas, Departamento de Ciencia de los Alimentos y Tecnología Química, Santiago, Chile
\textsuperscript{e} Eurofins Genomics, Anzinger Str. 7a, 85560 Ebersberg, Germany

ARTICLE INFO

Keywords: Seafood traceability, H1C gene, HRM, PAPM, FINS, DNA sequence analysis

ABSTRACT

Seaweed international trade has increased the labeling requirements in standards and regulations to include product information that enable traders and consumers to make informed choices. The European Union (EU) Regulation No. 1379/2013 imposes the declaration of an official commercial designation and scientific names for all the fishery and aquaculture products to be offered for sale to the final consumers. DNA analyses are used to enforce this regulation and to test authenticity in processed foods. We compared the performance of two monolocus approaches for species identification (SI) in 61 Mytilus mussels: the high-resolution melting analysis of the polyphenolic adhesive protein gene and the partial sequencing of the histone H1C gene. The H1C sequences were analyzed with five different methods. Both approaches show discrepancies in the identification of putative hybrids (0.0 < \(\kappa\) < 0.687 and 0.0 < MCC < 0.724). Excluding putative hybrids, methods show substantial to perfect agreement (0.772 < \(\kappa\) < 1.0 and 0.783 < MCC < 1.0). This study highlights the need to use standardized molecular tools, as well as to use multi-locus methods for SI of Mytilus mussels in testing laboratories.

1. Introduction

Smooth-shelled mussels (Mytilus spp.) includes, at least, seven species with statistics in FAO 2019 database. Among these edible mussels, 41.5%, 35.7% and 14.6% of the production came from Mytilus chilensis, M. edulis and M. galloprovincialis, respectively (FAO, 2021). It is well known that where more than one Mytilus species coexist in the same geographic area, they hybridize posing challenges to species identification (SI) (Coustau, Renaud, & Delay, 1991; Michalek, Ventura, & Sanders, 2016). Seafood international trade increased requirements regarding quality, safety, authenticity and sustainability production in food standards and regulations to give confidence to regulators, food traders and consumers, and to enable them to make informed choices (Codex Alimentarius, 2020; EU No. 1379/2013; EU No. 1380/2013; EU No. 1224/2009). These regulations require traceability “from ocean to fork”, that means among others, to identify the species to which the individuals belong (Ogden et al, 2008). For example, regulation EU No.1379/2013 establishes labelling requirements for marketed seafood and foodstuffs, such as declaring the commercial designation and scientific name of the species in the product labels (D’Amico, Armani, Gianfaldoni, & Guidi, 2016; Tinacci, Giusti, Guardone, Luisi, & Armani, 2019). To enforce this regulation, methods based on DNA analysis have been employed to investigate intentional and unintentional species substitution in seafood (Fernández-Tajes et al., 2011; Santaclera et al., 2006; Verrez-Bagnis et al., 2018).

The most popular approach to specimen identification is sequencing analysis, which can be performed by different methods. One of them is the Forensically Informative Nucleotide Sequencing (FINS), which involves the estimation of sequence similarity among specimens by phylogenetic methods (i.e. Kimura-2, Tamura-Nei or Jukes-Cantor) based on genetic distances and drawing a phylogenetic tree using UPGMA or Neighbor-Joining (NJ) algorithms (Bartlett & Davidson, 2006; Verrez-Bagnis et al., 2018).
sequences, and it is within the 95th percentile of all intraspecific distances (Meier, Shiyan, Vaidya, & Ng, 2006). This strategy requires determining the threshold similarity value that shows how similar a barcode match needs to be before it can be identified (Meier, Shiyan, Vaidya, & Ng, 2006). Finally, the “all barcodes” (AB) method is a more rigorous application of the best close match strategy (Meier et al., 2006). It uses information from all conspecific barcodes in the database, instead of just focusing on the most similar ones. The barcodes are sorted by similarity to the unknown sample using the same threshold as for the best close match. The identification is achieved if at least two sequences of the query species are available in the database and when all conspecific barcodes topped the list of the best matches. This method is more confident about assigning a species name to a query in cases where multiple species names are found on the list of best matches (Meier et al., 2006). Specifically, for the genetic identification of the three Mytilus species found in Europe (M. edulis, M. trossulus and M. galloprovincialis), Inoue, Waite, Matsuoka, Odo, and Harayama (1995) developed a PCR-length polymorphism analysis from the nonrepetitive region in the polyadenyllic adhesive protein gene (PAP). This region is an attractive target as it contains SNP and size polymorphisms between species within the Mytilus genus. Later, to differentiate between M. chilensis and M. galloprovincialis, Santacalora et al. (2006) added a Restriction Fragment Length Polymorphism analysis (PCR-RFLP) of the same amplicon. Always targeting this gene, Jilberto, Araneda, and Larraín (2017) developed a High-Resolution Melting (HRM) analysis (HRM-PAPM) to identify M. chilensis, M. galloprovincialis, M. edulis and their F_1 hybrids. By the HRM analysis is possible to obtain the genotypes of single nucleotide polymorphisms (SNPs) and detects length polymorphisms in small amplicons, showing distinguishable melting curves, permitting discrimination among species (Verrez-Bagnis et al., 2018). The HRM-PAPM analysis allows for distinguish, M. galloprovincialis (genotype GG, amplicon size 116 bp), M. edulis (genotype GG, 170 bp), and M. chilensis (genotype TT, 116 bp) (Jilberto et al., 2017). HRM is ideal for laboratory analysis because it is fast, accurate and less expensive in comparison with other DNA based methodologies (Jilberto et al., 2017). Another method used in the identification of Mytilus species is the comparison of the partial sequence of the histone H1C gene (Eirín-López, González-Tizon, Martínez, & Méndez, 2002). Specifically, H1C gene is used to separate M. chilensis from other Mytilus species (M. galloprovincialis, M. edulis, and M. trossulus). Both, the HRM-PAPM method and the HIC sequencing are easy to apply in routine testing to identify Mytilus species in traded seafood. However, basing SI on the analysis of a single gene, known as the mono-locus approach, can give contradictory results due to differences in the evolutionary rate of the analyzed genes (Larraín, González, Pérez, & Araneda, 2019; Vainió & Strelkov, 2011). Also, not all muscles are species-differentiable among all species (Larraín et al., 2019). To overcome these problems related to the analysis of a single marker, the multi-locus analysis integrating genomic information from SNP panels to identify M. trossulus, M. galloprovincialis, M. edulis, and M. chilensis has been used by Larraín, Zabwicka, Araneda, Gardner, and Wenne (2018) and Wenne et al. (2022) with 49 and 54 SNPs, respectively.

In this work, we tested and compared the performance of the HRM-PAPM method and the HIC sequence analysis for species identification in smooth-shelled mussels Mytilus galloprovincialis and M. chilensis using individuals previously identified with a panel of 49 SNPs.

2. Material and methods

2.1. Mussel samples, DNA extraction and species identification

Fifty-two samples of the Mytilus mussels had obtained from six growing centers in the Reloncavi Sound area (−41.700, −72.833), where the Chilean mussel aquaculture industry is located and one in Peel Island (−50.842, −74.011). Nine fresh individuals were obtained from a growing center in the Dichato Bay (−36.538, −72.957), 570 km away from the aquaculture area. Reloncavi Sound and Peel Island samples were collected in 2009 and 2013 (n = 26 per year), and Dichato samples were collected in 2009 (Table S1). All samples were processed fresh up to 24 h after collection, dissected and approximately 200 mg of mantle edge tissue was fixed with 95 % ethanol and stored at −20 °C until DNA extraction. DNA was obtained by the phenol-chloroform method, adapted for mussels (Larraín, Díaz, Lamas, Vargas, & Araneda, 2012). The DNA concentration was estimated with a NanoDrop ND-2000 spectrophotometer (Thermo Fischer Scientific). DNA integrity was assessed using agarose gel electrophoresis (0.7 % w/v). Mussel species were previously determined with a 49 SNPs panel by Larraín et al. (2018), resulting in nine as M. galloprovincialis (Mg) and 52 as M. chilensis (Mch).

2.2. High resolution melting (HRM-PAPM) analysis

The HRM-PAPM analysis was performed in an Eco Real-Time PCR System 4.0 (Illumina®) and MiC qPCR Cycler (Bio Molecular Systems). In all HRM analyses, M. galloprovincialis, M. edulis, and M. chilensis reference samples were included as controls. Also, negative control without DNA was included in all runs. Species identification was performed through clearly distinguishable melting curves, for extended protocol see Jilberto et al. (2017). The validation of the HRM-PAPM analysis was published in Quintrel et al. (2021), and a validation summary is included in Table S2.

2.3. Histone H1C gene sequencing

A partial sequence of approximately 400 bp of the HIC gene was amplified by PCR using the primer set HICF (5’-CATCATGGAACATT-CAACG-3’) and HICR (5’-GGCGTATAAGCTTCTGGCCAGA-3’) (Pérez-García, Morán, & Pasantes, 2014). The final product length was checked by electrophoresis on a 2 % agarose gel. PCR and sequencing were performed in the facilities of Eurofins Genomics GmbH (Ebersberg, Germany). All reactions were carried out in a 25-µL volume containing 1.5 U AmpliTaq Gold DNA Polymerase (Applied Biosystems®), 2.5 µL 10 × Gold Star Buffer (Promega Corporation®), 1 to 5 ng DNA and 0.2 µM of each primer). The PCR reaction was performed on a GeneAmp 9700 thermocycler (Applied Biosystems®). Initial denaturation was performed at 95 °C for 12 min; followed by 15 cycles of denaturation at 96 °C for 25 s, annealing at 55 °C for 20 s, and extension at 72 °C for 30 s; followed by 25 cycles of denaturation at 96 °C for 20 s, annealing at 58 °C for 20 s, and extension at 72 °C for 30 s with a final elongation step at 72 °C for 3 min. Double-stranded PCR amplicons sequencing was performed in an ABI 3130xl genetic analyzer (Applied Biosystems®).

2.3.1. Forensically informative nucleotide sequencing (FINS)

The 61 obtained sequences (GenBank accession numbers MT949777 to MT949837) were aligned together with 14 other Mytilus spp. HIC partial sequences recovered from GenBank (Drabant et al., 1999; Eirín-López et al., 2002) and two unpublished sequences used by Eurofins (Table 1), with the MAFFT 7.388 plugin (Katoh & Standley, 2013) in Geneious 11.5 (Biomatters) and manually edited. Phylogenetic analysis was carried out with Geneious tree builder module. A neighbour-joining phylogenetic tree based on the Tamura-Nei model genetic distances
Table 1

| ID           | GenBank     | Declared species | N' of sequences | Reference               |
|--------------|-------------|------------------|-----------------|-------------------------|
| AJ416421     |             | M. californianus  | 1               | Eirín-López et al. (2002). |
| AJ416422     |             | M. chilensis     | 1               | Eirín-López et al. (2002). |
| MT949777     | MT949820    | M. chilensis     | 44              | EUROFINS Genomics, Ebersberg, Germany |
| AJ416423     |             | M. edulis        | 1               | Eirín-López et al. (2002). |
| AJ224069     | AJ224077    | M. edulis        | 9               | Drabant et al. (1999). |
| AJ416424     |             | M. galloprovincialis | 1           | Eirín-López et al. (2002). |
| MT949821     | MT949828    | M. galloprovincialis | 9            | Current work |
| AJ416425     |             | M. trossulus     | 1               | Eirín-López et al. (2002). |

Total 77

Table 2

| Individual | GenBank ID | N' of individuals | Putative hybrid Mch × Mg | Putative hybrid Mch × Mg | Putative hybrid Mch × Mg |
|------------|------------|-------------------|--------------------------|--------------------------|--------------------------|
| M. chilensis | AJ49835 | 115                | Y                       | T                       | A                        |
| M. trossulus | MT949831 | 151                | Y                       | R                       | Y                        |
| M. chilensis | MT949833 | 193                | Y                       | G                       | R                        |
| M. galloprovincialis | MT949935 | 259                | Y                       | T                       | R                        |
| M. galloprovincialis | MT949408 | 322                | Y                       | R                       | T                        |

2.4. Agreement between the HRM-PAPM and histone H1C sequence analysis

The classification agreement between the HRM-PAPM and each of the Histone H1C sequence analyses described in 2.3 (DSC, BCM, AB and ABGD methods), was evaluated using individuals from dataset1 and dataset2. We calculate Cohen’s Kappa coefficient (κ) and the Matthews correlation coefficient (MCC) to evaluate the quality of SI obtained by the different molecular assays. κ statistic measures the agreement between two methods that classify items into mutually exclusive categories (Rotondi, 2018), and it was estimated with the 95% confidence interval (Cohen, 1960), using the “fmsb” R package version 0.6.3 (Nakasawa, 2018). FDR correction was applied to κ p-values to avoid the alpha-inflation produced by multiple testing (García, 2004). MCC is a measure of the quality of classification agreement between two categorical variables (Matthews, 1975) that was estimated using the “mltools” R package version 0.3.5.

3. Results

3.1. HRM-PAPM analysis

This method based on the melting curves from the 61 samples, classified nine individuals as M. galloprovincialis and the remaining 52 as M. chilensis, without any evidence of hybridization (Figure S1).

3.2. Histone H1C gene sequence analysis

3.2.1. Fins

The phylogenetic tree shows two principal clades (Fig. 1), the first one containing only two individuals from data uploaded by Eirín-López et al. (2002) on GenBank as M. californianus (AJ416421), and M. chilensis (AJ416422). The second clade grouped the remaining 75 individuals, among them, only M. trossulus was separated significantly with HIC gene analysis. The separations within this clade did not reach the minimum threshold confidence to be considered significant.

3.2.2. DSC

After aligning the HIC 397 bp amionc from all sequences available from GenBank and this work, 49 polymorphic sites were found. Despite

Verrez-Bagnis et al., 2018 was constructed with 10,000 bootstrap replicas as a reliability test (Terol, Mascarell, Fernández-Pedrosa, & Pérez-Alonso, 2002) using M. californianus as outgroup (AJ416421) because this species was the most distant taxa within Mytilus genus as revealed by these sequences. Species were identified when the individual clusters with conspecific barcodes (Meier et al., 2006).
Fig. 1. Neighbour-joining tree of the partial sequence of the \textit{HC1} gene of the 71 Mytilus sequences analyzed. GenBank downloaded sequences are identified by their accession number and the two unpublished sequences used by Eurofins. The twelve putative hybrids are indicated by (*).
these polymorphic sites, all sequences showed at least 91.44 % of similarity (Tables S3 and S4). The lowest similarity value between *M. galloprovincialis* and either *M. edulis* (download from GenBank) or *M. chilensis* was 97.73 %. (Table S4). Besides, in some cases, sequences corresponding to *M. edulis* and *M. galloprovincialis* were identical (Table S3).

Twelve individuals (GenBank accession numbers MT949820, MT949828 to MT949837 and MT949795) showed chromatograms with double peaks at five sites. These sites were: 115, 151, 193, 259, 322 (Table 2). In *M. galloprovincialis* these sites corresponded to T, A, C, G and A respectively, whereas in *M. chilensis* were C, G, T, A and C, respectively. The 12 individuals that presented double peaks in these five sites, showed the two nucleotides corresponding to the polymorphisms found between *M. chilensis* and *M. galloprovincialis* mentioned above, and were considered heterozygous for these sites and putative hybrids.

The sequence matrix showed a 97.73 % to 98.74 % of similarity between *M. chilensis* and *M. galloprovincialis* (Table S3 and S4). Considering the twelve sequences of putative hybrids carrying double picks in the five sites, the similarity with *M. chilensis* sequences ranged from 98.36 % to 99.87 % and with *M. galloprovincialis* ranged from 97.86 % to 99.37 % (Tables S3 and S4). Therefore, pairwise similarities among individuals from both species and their putative hybrids were higher than 97.73 % (Table S3 and S4).

Using dataset2 (excluding individuals whose sequences showed double picks) DSC analysis successfully assigned all samples to the correct species when a sequence similarity threshold over 99 % was considered.

### 3.2.3. Other sequence analysis methods

The ABGD method was not able to separate the individuals from dataset1 into groups according to the nominal species classification, the prior intraspecific distance ranged between 0.0028 and 0.0077. The BCM and the AB methods calculated a threshold similarity value of 1.28 %. BCM analysis correctly recognized the species from 55 sequences (90.16 %), four (6.56 %) were ambiguous (MT949829, MT949830, MT949832 and MT949837) and two (3.28 %) were incorrectly classified (MT949828 and MT949836). The ambiguous or incorrectly classified sequences corresponded to putative hybrids found in the DSC analysis. However, six sequences of those individuals who presented double picks (MT949795, MT949820, MT949831, MT949833, MT949834 and MT949835) were classified as *M. chilensis*. The AB method classified all the 61 individuals from dataset1 as ambiguous (Table 3).

Using dataset2, the ABGD method detected a gap in the intraspecific and interspecific distances in H1C sequences between *M. chilensis* and *M. galloprovincialis*. The BCM and the AB methods calculated a threshold similarity value of 0.25 %, to identify barcodes as conspecific. The BCM analysis successfully assigned to the species, all the 49 individuals.

### 4. Discussion

Molecular taxonomic identification or specimen assignment, i.e., to assign an individual to a species, is important for food safety and authenticity, wildlife forensic, conservation, property and consumers protection. In the food sector including, regulators, processors, retailers and consumers, there is special concern about species identification (Armani et al., 2015; Verrez-Bagnis et al., 2018). It is widely accepted the need for proper methods to support law enforcement, that could be easily implemented by private or government laboratories. However, when different analytical methods are used for the same purpose (in this case, SI) the results may not be concordant among them. Therefore, harmonization is necessary to avoid stakeholders being affected by the discrepancies among the different analytical methods (Coleman & Fontana, 2010). The harmonization among commercial methods currently used for mussel SI begins with a comparison of their performance. Discrepancies among methods could arise from the different molecular markers analyzed, but also from the sequence analysis methodology employed. In this study, we compare two mono-locus commercial methods used to identify the most traded *Mytilus* mussel species: the HRM-PAPM and the sequence analysis of the H1C barcode gene.

The high genetic similarity among *Mytilus* spp. has been revealed by many markers used for DNA barcoding (COI, SS rDNA, ITS-1, ITS-2 and NTS) (Giusti et al., 2020; Santana-Clar et al., 2006; Tinacci et al., 2018), which were not able to separate the different species. However, the polyphenolic adhesive protein (PAPM) and the histone H1C genes, have been shown to successfully differentiate *Mytilus* mussels (e.g., Santana-Clar et al., 2006, Fernández-Tajes et al., 2011, Eirín-López et al., 2002). SI results obtained by the HRM-PAPM and H1C gene sequence analysis showed a slight to a substantial agreement among them, using the complete dataset (dataset 1). Discrepancies between mono-locus methods are expected because they target genomic regions with different evolutionary histories (Larraín et al., 2012; Vainiola & Strelkov, 2011). The PAPM is a single copy gene, while the H1 genes (including H1C) have evolved by gene duplication with close to one hundred copies per haploid genome in *M. galloprovincialis* (Eirín-López et al., 2002) and *M. edulis* (Drabent et al., 1999). Besides, in contrast to the highly conserved family of core histones (H2A, H2B, H3, and H4), the H1 histones are less conserved during evolution (Drabent et al., 1999).

H1C sequences showed very high similarity among the *Mytilus* taxa analyzed, therefore, there was no barcoding gap in the H1C gene to separate intra- from inter-specific sequence variation when considering all the individuals analyzed. Probably, this is the reason why FINS methodology based on the H1C gene was not useful for SI in smooth-
shelled mussels (Fig. 1). Also, considering all the problems of tree-based identification techniques discussed by Meier et al. (2006) and Collins and Cruickshank (2013), FINS is not a recommendable approach to perform traceability in Mytilus species.

On the other hand, the DSC method uses a similarity criterion relying on an arbitrary threshold to consider two sequences belonging to the same species. Usually, a 1 or 2 % divergence is a reasonable rule-of-thumb in most cases, but it is prone to produce different rates of false-positive and false-negative (Collins & Cruickshank, 2013). In this study, a 1 % divergence threshold in pairwise similarities between M. chilensis and M. galloprovincialis H1C sequences, was suitable to correctly identify all samples, excluding putative hybrids (dataset2). The ABGD method was not able to identify a barcoding gap between M. chilensis and M. galloprovincialis sequences when all individuals were considered, as is in a real scenario in mussel traceability. This sequence analysis method only worked when putatively hybrids were previously excluded from the analysis. As expected, the BCM and AB methods showed a broader threshold similarity value (1.28 %) when all individuals were considered, compared with the 0.25 % threshold obtained when putatively hybrids were excluded. When these two methods were applied to identify the species in all individuals, the performance was lower. Moreover, the strictest criteria of the AB method, classified all individuals as ambiguous. It is important to consider that using the same dataset, sequence analysis methods could give different results.

Therefore, the election of the sequence analysis method for SI must be considered during the standardization and validation process.

The sequences for the HIC gen in Mytilus spp. available in the GenBank, were published in two papers. Eirín-López et al. (2002) uploaded sequences from M. edulis, M. galloprovincialis, M. trossulus, M. chilensis and M. californianus, whereas Drabent et al. (1999) contributed with nine M. edulis sequences for this gene. Until now, the NCBI Reference Sequence Database contains no sequences for the HIC gen from Mytilus spp. The reference sequences have their provenance and validity reviewed and checked and their GenBank ID begins with NC. This quality check is extremely important to avoid taxonomic misidentifications. Reviewing the HIC sequences published in GenBank, we realized that the sequence AJ416422 uploaded as M. chilensis by Eirín-López et al. (2002) grouped with the M. californianus sequence and not with M. chilensis (Fig. 1). Besides, 27 polymorphic sites were present between our sequences of M. chilensis and the sequence AJ416422, indicating a possible error in the species assigned to the last sequence in GenBank. Mistakes uploading sequences to GenBank, especially at the species level (but not at the genus level) are common, affecting the taxonomic reliability of this database (Leray, Knowlton, Ho, Nguyen, & Machida, 2019). Both species are very different in shell morphology, M. californianus is a ribbed mussel, while M. chilensis is a smooth shelled mussel. Unfortunately, the specimen used by Erin-Lopez do not have morphological data linked to the GenBank record. The availability of

---

### Table 4

Kappa concordance, confidence interval of 95%, agreement and Matthews correlation coefficient (MCC) between Species Identification methods based on the HRM-PAPM / Me15-16 and H1C barcode gene Direct Sequence Comparison (DSC), Best close Match (BCM) all barcode (AB) and Automatic Barcode Gap Discovery (ABGD) methods for a) dataset1(all sequences) and b) dataset2 (without putative hybrids).

|      | DSC       | BCM       | AB        | ABGD      |
|------|-----------|-----------|-----------|-----------|
| **a) Dataset1** |           |           |           |           |
| HRM PAPM | 0.5034   | 0.6676    | 0         | –         |
|          | (p = 0.00145) | (p = 0.00109) | (p = 0.5) | –         |
| Moderate agreement | 0.2516–0.7552 | 0.4150–0.9201 |
| MCC = 0.5775 | –         | Substantial agreement |
| DSC | 0.6870    | 0         | –         | –         |
| BCM | –         | –         | –         | –         |
| AB | –         | –         | –         | –         |

| **b) Dataset2** |           |           |           |           |
| HRM PAPM | 1.00     | 1.00      | 0.7717    | 1.00      |
|          | (p = 0.00006) | (p = 0.00006) | (p = 0.00054) | (p = 0.00006) |
| DSC | –         | Perfect agreement | MCC = 1.00 | –         |
| BCM | –         | –         | –         | –         |
| AB | –         | –         | –         | –         |
reference sequences obtained from voucher specimens in cured databases is necessary to standardize sequence-based methods as traceability tools and to avoid taxonomical uncertainties. The species from all the individuals whose HIC sequences were obtained in this work and uploaded to GenBank were previously checked by the 49 SNP panel (Larrain et al., 2018) and corresponded to *M. chilensis* and *M. galloprovincialis*.

Food traceability and authenticity in *Mytilus* mussels is challenging due to the natural hybridization observed in many areas where these species are cultured (Michael et al., 2016). In the case of *M. chilensis* and *M. galloprovincialis* the presence of hybrids has been highlighted in molecular identification studies conducted on cultured specimens in Arauco and Reloncavi gulfs (Larrain et al., 2012, 2019) and on exported Chilean mussel products collected from retail (Giusti et al., 2022). To overcome this challenge, new molecular techniques are necessary to separate hybrids from the parental taxa in an efficient a cost-effective way. However, according to our results, mono-locus assays are not efficient for this purpose, because they gave contradictory conclusions. Methods based on a multi-locus approach (Larrain et al., 2019) or genomic technology must be developed and validated to be used by regulatory agencies and the food industry. The higher the complexity of a method, the more time it consumes, affecting the practicability required in routine analysis. Consequently, using a small panel of informative SNPs is a good option to address the presence of hybrids (Quintrel et al., 2021). For example, all individuals used in this work were previously genotyped with a 49 SNP panel without any evidence of being hybrids. However, HIC sequence analysis showed strong evidence of hybridization in twelve individuals, reflecting the complexities of hybrids identification.

The SI discrepancies between these HIC and PAP are mainly because the putative hybrids detected by the HIC gene sequence analysis were not classified as such by the HRM-PAPM method, although this latter can also distinguish hybrids between these species (Jilberto et al., 2017).

The HRM-PAPM method showed total agreement with the 49 SNPs panel to identify *M. chilensis* from *M. galloprovincialis*. The HIC sequence analysis is helpful to differentiate between both species when the appropriate sequence analysis method is used. Our results show that HIC can reveal some hybrid not detected by the HRM-PAPM, finding that is highly possible because *Mytilus* mussel hybridizes, their genome has a size of ~ 1.6 Gb and of course we are screening only a very large fraction of it with our SI methods.

5. Conclusions

A high similarity of HIC gene within *Mytilus* spp. is evident and it hampers the differentiation between *M. edulis* and *M. galloprovincialis*. Nevertheless, the study confirms that the application of the BMC method allows both differentiation between the two species and the detection of hybrids.

Based on the HIC gene, the sequence analysis method affects SI outcomes. The FINS, ABGD and AB methods were not useful to identify *M. chilensis*, *M. edulis* and *M. galloprovincialis* specimens.

The different levels of agreement between the 49 SNP panel and both SI methods (HRM-PAPM and HIC gene sequence analysis) highlight the need for the standardization of molecular tools. The presence of hybrids is a realistic scenario in *Mytilus* mussel aquaculture that makes SI more complex. Our results indicates that analytical tools based on a multi-locus approach are needed to enface the challenge of the traceability of smooth-shelled mussels.

Each assay laboratory must decide what is the best method for mussel SI depending on the equipment available and if it can easily be combined with other methods used in the laboratory. If sanger sequencing is a routine procedure, we recommend the sequencing of the HIC gene followed by analysis with the BCM method. On the other hand, if the qPCR analysis is the routine procedure HRM-PAPM is the recommended method.

Our results highlight the need for standardized molecular tools to perform SI in smooth-shelled mussels, as well as the use of a multi-locus approach.

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.

Acknowledgement

This research was funded by the Chilean Agencia Nacional de Investigación y Desarrollo (ANID) through grants and ANID/CONICYT FONDECYT Regular 1191765 and FONDEF ID1610013/20013. Sanger sequencing was kindly provided by Eurofins Genomics Europe Applied Genomics GmbH.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jfochs.2022.100121.

References

Armani, A., Guardone, L., Castiglione, L., D’Amico, P., Menina, A., Malandra, R., ... Guidi, A. (2015). DNA and Mini-DNA barcoding for the identification of Porgies species (family Sparidae) of commercial interest on the international market. Food Control, 50, 589-596. https://doi.org/10.1016/j.foodcont.2014.09.025

Barbuto, M., Galimberti, A., Ferri, E., Labra, M., Malandra, R., Galli, P., & Casiraghi, M. (2010). DNA barcoding reveals fraudulent substitutions in shark seafood products: The Italian case of “palombo” (Mustelus spp.). Food Research International, 43(1), 376-381, https://doi.org/10.1016/j.foodres.2009.10.009

Bartlett, S. E., & Davidson, W. S. (1992). FINS (Forensically Informed Nucleotide Sequencing): A procedure for identifying the animal origin of biological specimens. BioTechniques, 12(3), 408–411. https://europepmc.org/article/med/1571152.

Codex Alimentarius. (2020). General Principles of Food Hygiene CXC 1-1969 (p. 35). https://www.fao.org/3/a-i7555e.pdf

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37–46.

Collins, R. A., & Cruickshank, R. H. (2013). The seven deadly sins of DNA barcoding. Molecular Ecology Resources, 13(6), 969-975. https://doi.org/10.1111/j.1755-099X.2013.03246.9

Cousten, C., Renaud, F., & Delay, B. (1991). Genetic characterization of the hybridization between Mytilus edulis and *M. galloprovincialis* on the Atlantic coast of France. Marine Biology, 111(1), 87–93. https://doi.org/10.1007/BF01986350

D’Amico, P., Armani, A., Gianfaldoni, D., & Guidi, A. (2016). New provisions for the labelling of fishery and aquaculture products: Difficulties in the implementation of Regulation (EU) n. 1379/2013. Marine Policy, 71, 147–156. https://doi.org/10.1016/j.marpol.2016.05.026

Drahent, B., Kim, J. S., Ahlbg, W., Prats, E., Cornudella, L., & Doncnek, D. (1999). Mytilus edulis histone gene clusters containing only H1 genes. Journal of Molecular Evolution, 49(5), 645–655. https://doi.org/10.1007/PL00006585

Erín-López, J. M., González-Tizón, A. M., Martínez, A., & Méndez, J. (2002). Molecular and evolutionary analysis of muscul histone genes (Mytilus spp.); Possible evidence of an “orphan origin” for H1 histone genes. Journal of Molecular Evolution, 55(3), 272–283. https://doi.org/10.1007/s00239-002-2235-1

EU No 1224/2009. (n.d.) Establishing a Community control system for ensuring compliance with the rules of the common fisheries policy. Official Journal of the European Union. EU.N.1380/2013. (2013). Regulation (EU) No 1380/2013 of the European Parliament and of the Council of 11 December 2013 on the Common Fisheries Policy, amending Council Regulations (EC) No 1954/2003 and (EC) No 1224/2009 and repealing Council Regulations (EC) No 2171/2002 and (EC. In Official Journal of the European Union. EU.N.1379/2013. (2013). Regulation (EU) No 1379/2013 of the European Parliament and of the Council of 11 December 2013 on the common organisation of the markets in fishery and aquaculture products, amending Council Regulations (EC) No 1184/2006 and (EC) No 1224/2009 and repealing. Official Journal of the European Union.

FAO. (2021). Fishery and Aquaculture Statistics. Global capture production 1950-2019 (Fishstat). In: FAO Fisheries and Aquaculture Department [online]. (Updated 2019). www.fao.org/fishery/statistics/software/fishstatj/en.
Fernández-Tajes, J., Longa, A., García-Gil, J., Chiu, Y. W., Huang, Y. S., Méndez, J., & Lee, R. S. (2011). Alternative PCR-RFLP methods for mussel Mytilus species identification. European Food Research and Technology, 233(5), 791–796. https://doi.org/10.1007/s00217-011-1574-x

García, L. V. (2004). Escaping the Bonferroni iron claw in ecological studies. Issue 3. In Giusti, A., Malloggi, C., Tosi, F., Boldini, P., Larraín Barth, M. A., Araneda, C., ... Araneda, C. (2019). Comparison between single and multi-locus approaches for specimen identification in Mytilus mussels. European Food Research and Technology, 233, 1084. https://doi.org/10.1007/s00217-019-10645-8

García, L. V. (2022). Escaping the Bonferroni iron claw in ecological studies. Issue 3. In Giusti, A., Malloggi, C., Tosi, F., Boldini, P., Larraín Barth, M. A., Araneda, C., ... Araneda, C. (2019). Comparison between single and multi-locus approaches for specimen identification in Mytilus mussels. European Food Research and Technology, 233, 1084. https://doi.org/10.1007/s00217-019-10645-8