

SNP Array for Small-Shrimp (Genus *Acetes*) Origin Determination Using Machine Learning

Eun Soo Noh *, Mi Nan Lee, Chun-Mae Dong, Jungwook Park, Hyo Sun Jung, Woo-Jin Kim and Young-Ok Kim

Biotechnology Research Division, National Institute of Fisheries Science, 216, Gijanghaean-ro, Gijang-eup, Gijang-gun, Busan 46083, Republic of Korea; minan012@naver.com (M.N.L.); ehdcnsao@naver.com (C.-M.D.); jjuwoogi@korea.kr (J.P.); jhs3010@korea.kr (H.S.J.); wj2464@korea.kr (W.-J.K.); yobest12@korea.kr (Y.-O.K.)

* Correspondence: laperm@korea.kr; Tel.: +82-51-720-2451

Abstract: Accurate origin determination of seafood is crucial for consumer trust and safety. This study was performed to develop a machine learning-based single-nucleotide polymorphism (SNP) analysis technique to determine the origin of *Acetes* species in salted small-shrimp products. Mitochondrial DNA (COI and 16S rRNA) analysis revealed genetic variations among species and origins. Eight candidate SNPs were identified, six of which were developed into markers for genotyping analysis. Using the developed markers, an SNP array was created and SNP data from salted small-shrimp samples were obtained. Machine learning analysis using a supervised learning algorithm achieved 100% accuracy in classifying the origin of *Acetes* based on SNP data. This method offers a reliable method for regulatory bodies to combat food fraud and ensure product integrity. The approach can be further improved by expanding the data set to encompass a wider range of species and origins. This study highlights the potential of SNP analysis and machine learning for ensuring seafood authenticity and promoting sustainable practices.

Keywords: small shrimp; machine learning; origin; single-nucleotide polymorphism



Citation: Noh, E.S.; Lee, M.N.; Dong, C.-M.; Park, J.; Jung, H.S.; Kim, W.-J.; Kim, Y.-O. SNP Array for Small-Shrimp (Genus *Acetes*) Origin Determination Using Machine Learning. *Foods* **2024**, *13*, 2087. <https://doi.org/10.3390/foods13132087>

Academic Editor: Efrén Pérez Santín

Received: 3 June 2024

Revised: 23 June 2024

Accepted: 27 June 2024

Published: 1 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Issues related to the safety and quality of seafood have emerged, with consumers increasingly paying more attention to the origin of products due to concerns about health and food safety [1]. Knowing the origin of seafood allows consumers to make informed choices and avoid potential health hazards related to the use of antibiotics in aquaculture, pollution in waterways, and mislabeling [2]. Clearly indicating the origin of food ensures trust in the product, reflecting the natural environment, production technologies, and conditions involved in food production [3]. For example, Norwegian salmon may be perceived as superior due to specific production methods, environmental factors, or regulations [4]. Knowing the origin of foods allows producers to charge a premium for these desired qualities. Unfortunately, the mislabeling of seafood is an ongoing problem. Unscrupulous sellers may pass off lower quality or less desirable species as more valuable alternatives, which not only deceives consumers but also undermines trust in the seafood industry. Accurate origin determination helps to combat food fraud and ensure consumers receive the products they have paid for. Furthermore, producers in each country or region strive to enhance the competitiveness of their seafood and food products by emphasizing their origin [5].

Consumers often place high trust in foods from certain countries, including domestic production [6]. The mislabeling of seafood of lower quality as more desirable species poses risks to consumer health due to unreliable production environments or processes [7]. To prevent such issues, ensure consumer safety, and establish order in distribution, research on origin determination of various seafood products is required.

In the context of Korean cuisine, where small shrimp (*Acetes* species) are consumed in large quantities and play a significant role in traditional dishes like kimchi, ensuring their

authenticity and origin is particularly crucial [8]. In particular, the rising demand for these shrimp has led to a decline in domestic catches and a subsequent price increase, creating an opportunity for fraudulent mislabeling of imported shrimp as domestic [9]. Accurate species and origin determination is essential to maintain consumer confidence and uphold the quality of domestic products.

There are two main approaches to determining the origin of seafood. The first approach leverages genetic differences within the same species resulting from geographical isolation among countries. Population genetic analysis plays a crucial role in such studies. By studying these genetic variations, populations from different regions can be distinguished; for example, studies can use microsatellite markers and genetic variations to determine the origin of small shrimp (*Acetes chinensis*) and Manila clams (*Ruditapes philippinarum*) [9,10]. The second approach focuses on identifying the species of seafood, which becomes particularly relevant when dealing with species that are not native to a particular country or that closely resemble others. Here, techniques such as multiplex PCR analysis are used to differentiate between domestic and imported seafood or to identify non-native species that may be present. For instance, studies have used multiplex PCR to analyze sole (*Cynoglossus*), croakers (*Larimichthys*, *Atrobucca*, and *Pseudotolithus*), and loaches (*Misgurnus* and *Paramisgurnus*) [11–13].

Species identification is fundamental to taxonomic research and is a crucial element in studies of ecology and evolution [14]. Accurate species identification underpins research on biodiversity, the monitoring of endangered populations, and aquaculture and fisheries research [15,16]. Morphological and molecular biological methods are generally used for identification, with molecular methods complementing the limitations of morphological classification and enhancing accuracy [17]. One commonly used molecular method involves the analysis of single-nucleotide polymorphisms (SNPs), genetic variations where a single nucleotide differs between members of the same or different species. These variations in genes serve as effective genetic markers in both animals and plants, providing insight into population compositions [18]. SNP analysis technology has applications in a number of fields, including forensic research and studies of biological evolution [18,19].

Research on SNPs can be significantly enhanced when combined with machine learning (ML), a rapidly growing field of computer science that can uncover complex correlations hidden in large data sets obtained by next-generation sequencing [20]. ML uses a range of algorithms to generate models from existing data, facilitating pattern recognition and classification, thus yielding predictions. These algorithms can be broadly divided into supervised learning, which involves training with well-classified objects to predict the classification of new objects, and unsupervised learning, which involves classifying provided objects without a training process [21]. Such ML analyses have applications in various biological fields, such as pattern recognition in coding sequences, the identification of biomarkers, and the exploration of disease genes [22,23].

In this study, an ML-based SNP analysis technique was developed and evaluated to accurately discriminate the origin of small shrimp. By leveraging the power of ML algorithms, complex correlations within mitochondrial DNA data sets were uncovered to facilitate efficient pattern recognition and classification, ultimately leading to robust predictive models for determining the geographical origin of small-shrimp samples. This integrated approach combines the high accuracy of SNP analysis with the computational efficiency of ML, offering a promising solution for addressing challenges related to seafood mislabeling and ensuring the authenticity of small-shrimp products.

2. Materials and Methods

2.1. Sample Preparation

Salted shrimp samples were purchased from a local market for genetic analysis. The origin of each sample was labeled on the product packaging (Table 1). Genomic DNA was extracted using a DNeasy Blood & Tissue Kit (Qiagen Inc., Toronto, ON, Canada) in accordance with the manufacturer's protocol. Briefly, 20 mg of muscle tissue was lysed

with 180 μ L of ATL buffer and 20 μ L of proteinase K at 56 °C for 1 h. Then, 180 μ L of AL buffer and 180 μ L of ethanol (99%) were added, the sample was mixed, and the supernatant was transferred to a column and centrifuged at 8000 rpm. The DNA was washed twice with 300 μ L of AW1 buffer and an equal volume of AW2 buffer, and then eluted with TE buffer. The extracted genomic DNA was visualized on a 1% (*w/v*) agarose gel and quantified by spectrophotometry (NanoVue; GE Healthcare, Fairfield, NJ, USA).

Table 1. The information of salted small-shrimp products used in this study.

Sample No.	Indicated Origin		Collection Date	Population Size
1	Korea	Sinan-gun	7 March 2023	48
2	Korea	-	7 March 2023	48
3	Korea	Ganghwa-gun	7 March 2023	47
4	Korea	-	8 March 2023	48
5	Korea	Incheon-si	24 March 2023	48
6	Korea	Yeonpyeong Island	24 March 2023	43
7	Korea	Boryeong-si	24 March 2023	48
8	Korea	Taeaeon-gun	24 March 2023	47
9	Korea	-	24 March 2023	48
10	Korea	Mokpo-si	24 March 2023	48
11	China	-	24 March 2023	46
12	Vietnam	-	7 March 2023	46
13	Vietnam	-	7 March 2023	48
14	Malaysia	-	26 June 2023	23

2.2. Sequence Analyzing

Mitochondrial DNA (mtDNA) barcode regions—the cytochrome oxidase subunit I (COI) gene and 16S ribosomal RNA (16S rRNA) gene—were amplified and sequenced to identify the species and compare interspecific genetic variation. The following universal primers were used: LCO1490 (5'-GGT CAA CAA ATC ATA AAG ATA TTG G-3') and HCO2198 (5'-TAA ACT TCA GGG TGA CCA AAA AAT CA-3') for COI; and 16Sar (5'-CGC CTG TTT ATC AAA AAC AT-3') and 16Sbr (5'-CCG GTC TGA ACT CAG ATC ATG T-3') for 16S rRNA [24,25]. PCR was performed in a mixture consisting of 1 μ L of genomic DNA, 2 μ L of 10 \times Ex Taq buffer, 1.6 μ L of dNTP mixture (2.5 mM), 1 μ L of forward primer (10 pmol), 1 μ L of reverse primer (10 pmol), and 5 units of Ex Taq polymerase (Takara Bio, Otsu, Japan) in a total volume of 20 μ L. PCR amplification was performed using an ABI Veriti Fast Thermal Cycler (Applied Biosystems, Foster City, CA, USA) with an initial denaturation step at 94 °C for 10 min followed by 32 cycles of denaturation at 94 °C for 30 s, annealing at 58 °C for 30 s, and extension at 72 °C for 1 min, with a final extension at 72 °C for 7 min. The PCR products were visualized on a 1% (*w/v*) agarose gel to confirm amplification, and then sequenced using a BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and an ABI3730XL sequencer.

2.3. SNP Discovery

Sequence data produced from the samples were assembled and aligned using SeqMan Pro v17.3.0 (DNA Star, Madison, WI, USA). Intraspecific genetic variation and haplotypes by group were analyzed using DnaSP v5.10.01 [26]. Single-nucleotide variants that showed differences between each group were then identified. Intraspecific genetic distances were analyzed using the Kimura-2-parameter model in MEGA 10 and summarized in a phylogenetic tree using the neighbor-joining (NJ) method [27].

2.4. Designing SNP Genotyping Array

Primers for the eight selected SNPs were designed using the Fluidigm D3 Assay Design Tool (<http://d3.standardbio.com/>; accessed on 11 October 2023) along with their flanking sequences. The primers and Integrated Fluidic Circuit (96.96 IFC) array were

ordered from Fluidigm (Fluidigm Corp., San Francisco, CA, USA). To test the performance of the manufactured SNP panel, 96 randomly selected small-shrimp samples that had been used for sequence analysis were analyzed.

2.5. Machine Learning Analysis

SNP information obtained from each salted shrimp product was used as training data for the ML analysis to enable rapid result interpretation and improve accuracy. The ML analysis was performed using WEKA 3.8.5 with the Random Forest algorithm, a supervised learning method chosen for its ability to handle high-dimensional data, such as SNP information, and its robustness against overfitting [28]. Where a well-trained model was used to analyze the provided data, the accuracy of the developed ML workflows was evaluated using SNP information from random samples obtained using the Fluidigm array as test data.

3. Results

3.1. Sequencing Analysis

Genetic information was obtained from up to 48 samples selected from 14 salted small-shrimp products labeled as originating from Korea, China, Vietnam, and Malaysia. The COI and 16S rRNA gene regions were obtained from the products labeled as coming from Korea. NCBI BLAST analysis revealed that both gene regions showed 100% identity to *A. chinensis* MK941905.1 in the COI region and 99.8% identity to MK928508.1 in the 16S rRNA region. For shrimp paste labeled as coming from China, the COI and 16S rRNA regions were also obtained, and both regions showed 100% identity to *A. chinensis* MK941908.1 and MK928507.1. Although both origins were identified as the same species, *A. chinensis* from Korea and China exhibited genetic differences of 3.95% in the COI region and 0.96% in the 16S rRNA region (Figure 1).

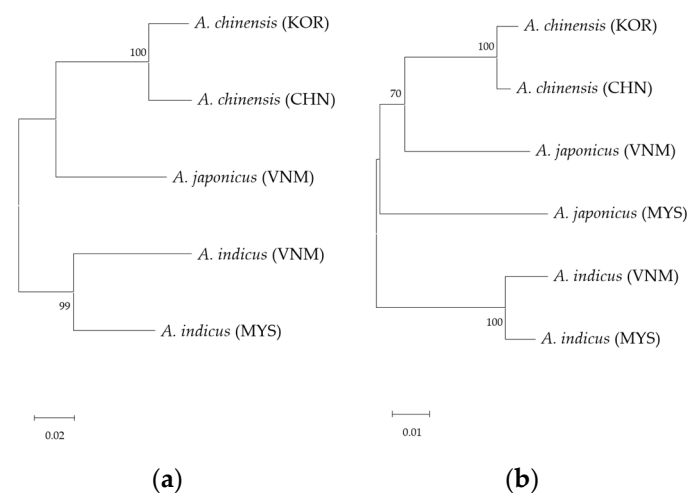


Figure 1. Phylogenetic analysis of small shrimp from Korea (KOR), China (CHN), Vietnam (VNM), and Malaysia (MYS) using neighbor-joining method. (a) Mitochondrial COI region; (b) mitochondrial 16S rRNA region.

One of the salted small-shrimp products labeled as coming from Vietnam was identified as *A. japonicus*, showing 100% identity to MZ412582.1 in the COI region and 99.42% identity to MZ412582.1 in the 16S rRNA region. Another product from the same origin was identified as *Acetes indicus*, exhibiting 100% identity to OP420230.1 in the COI region and 99.42% identity to MK245778.1 in the 16S rRNA region. The salted small-shrimp product labeled as coming from Malaysia contained a mixture of *A. japonicus* and *A. indicus*. *A. japonicus* showed 99.36% identity to MK412582.1 in the 16S rRNA region, but the COI gene could not be obtained. In addition, *A. indicus* exhibited 100% identity to HQ630436.1 in the COI region and 97.31% identity to MK245778.1 in the 16S rRNA region. However,

A. japonicus from Vietnam and Malaysia samples showed a genetic difference of 8.43% in the 16S rRNA gene region. Furthermore, *A. indicus* from Vietnam and Malaysia samples exhibited genetic differences of 9.81% in the COI region and 2.12% in the 16S rRNA region.

3.2. SNP Discovery

The haplotype diversity within groups of salted small-shrimp products classified by species and origin was confirmed by analyzing two mitochondrial gene regions. Subsequently, SNPs that can distinguish between the groups were identified. In the COI region, two candidate SNPs (sites 171 and 195) were identified (Figure 2), and in the 16S rRNA region, six candidate SNPs (sites 135, 206, 291, 322, 411, and 458) were identified (Figure 3). Using these candidate SNPs, the accuracy of distinguishing between groups was assessed. Within the *A. chinensis* group, only one SNP (COI-171) could distinguish between those of Korean and Chinese origin, whereas two or more SNPs were suitable for this purpose in all other groups (Table 2).

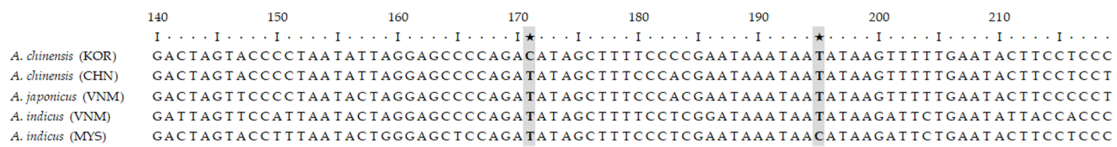


Figure 2. Candidate SNP marker sites selected from mitochondrial COI region. The asterisks in the figure indicate the polymorphic sites for origin determination.



Figure 3. Candidate SNP marker sites selected from mitochondrial 16S rRNA region. The asterisks in the figure indicate the polymorphic sites for origin determination.

Table 2. Candidate SNP marker information for identifying origin of small shrimp.

Species (Origin)	COI Region				16S rRNA Region			
	171	195	135	206	291	322	411	458
<i>A. chinensis</i> (KOR)	C	T	A	C	A	A	A	T
<i>A. chinensis</i> (CHN)	T	T	A	C	A	A	A	T
<i>A. japonicus</i> (VNM)	T	T	A	C	T	A	A	C
<i>A. japonicus</i> (MYS)	-	-	A	T	A	A	T	T
<i>A. indicus</i> (VNM)	T	T	A	C	A	G	A	C
<i>A. indicus</i> (MYS)	T	C	G	C	A	A	A	C

3.3. Designing SNP Genotyping Array

To facilitate genotyping using microarrays, marker development was performed for the selected SNP candidates using the Fluidigm D3 Assay Design Tool. Marker development was successfully completed for six candidates, excluding the 291 and 322 SNPs in the 16S rRNA region (Table 3). The two SNPs for which markers were not developed were found to be specific to *A. japonicus* and *A. indicus* from Vietnam, making it impossible to distinguish between them. Additionally, with the exclusion of the two markers, only one SNP (16S-458) was involved in discriminating between *A. chinensis* from China and *A. indicus* from Vietnam.

Table 3. A list of Fluidigm SNP-type assays developed in this study.

SNP ID	Target Gene	Position	SNP	Fluidigm Assay ID
COI-171	COI	171	..CCAGA[C/T]ATAGC..	GTA0343170
COI-195	COI	195	..AATAA[C/T]ATAAG..	GTA0343171
16S-135	16S rRNA	135	..TTTAA[G/A]AATTA..	GTA0343174
16S-206	16S rRNA	206	..CTTTA[T/C]TGTTT..	GTA0343420
16S-417	16S rRNA	411	..GTCCA[T/A]ATCGA..	GTA0343172
16S-464	16S rRNA	458	...CCTTT[T/C]TAATG..	GTA0343173

3.4. Analyzing Genotyping Data

Genotype information for six SNPs obtained from a total of 636 small-shrimp samples was used as training data for the ML analysis. The training data were composed of groups consisting of *A. chinensis* from Korea, *A. chinensis* from China, *A. japonicus* and *A. indicus* from Vietnam, and *A. japonicus* and *A. indicus* from Malaysia. The constructed training data were validated using the Random Forest algorithm with 80% as training data and 20% as test data, and achieved 90.65% accuracy. For Vietnamese *A. japonicus* and *A. indicus*, it was impossible to differentiate between the species due to them sharing the same SNP markers. However, the identification accuracy for other species and their origins was confirmed to be 100%.

Subsequently, salted small-shrimp samples were randomly collected for analysis, including 16 samples each from Korea, China, and Vietnam, and 32 samples from Malaysia. These samples were subjected to Fluidigm chip analysis. The results confirmed a 100% call rate for all samples, and the obtained genotype information was used as test data to perform species and origin identification through ML analysis. Korean salted small-shrimp samples were identified as 100% (16/16) Korean *A. chinensis*, and Chinese samples were also identified as 100% (16/16) Chinese *A. chinensis*. The Vietnamese samples identified as *A. japonicus* and *A. indicus* matched 100% (32/32) with the same origin. The Malaysian products matched 21.9% (7/32) with *A. japonicus* and 78.1% (25/32) with *A. indicus* (Table 4).

Table 4. The accuracy of small-shrimp origin identification using machine learning analysis with 6 SNP markers.

Species (Origin)	Sample 1 (Korea)	Sample 2 (China)	Sample 3 (Vietnam 1)	Sample 4 (Vietnam 2)	Sample 5 (Malaysia)
<i>A. chinensis</i> (KOR)	16	-	-	-	-
<i>A. chinensis</i> (CHN)	-	16	-	-	-
<i>A. japonicus</i> (VNM)	-	-	16	16	-
<i>A. japonicus</i> (MYS)	-	-	-	-	7
<i>A. indicus</i> (VNM)	-	-	-	-	-
<i>A. indicus</i> (MYS)	-	-	-	-	25
Species accuracy	100%	100%	100%	0%	100%
Origin accuracy	100%	100%	100%	100%	100%

4. Discussion

The integrity of the seafood industry is critically dependent on accurate origin determination, a concern that has become increasingly significant amid rising consumer awareness about food safety and quality [29]. This study was performed to develop a reliable method to discriminate the origin of small shrimp, specifically targeting the species *A. chinensis*, *A. japonicus*, and *A. indicus*, which are commonly used in Korean cuisine. ML-based SNP analysis was performed to address issues related to mislabeling and ensure that consumers receive accurate information.

The results of the mtDNA analysis of the COI region and 16S rRNA region provided genetic distinctions among small-shrimp species and their geographical origins. Notably, *A. chinensis* from Korea and China exhibited significant genetic differences, as did *A. japonicus* and *A. indicus* from Vietnam and Malaysia. These genetic markers are critical for accurately identifying the provenance of small-shrimp products. Through SNP discovery, crucial genetic variations capable of differentiating between species and origins were identified. Specifically, eight candidate SNPs were identified in the COI and 16S rRNA regions. Six of these were successfully developed into markers using the Fluidigm D3 Assay Design Tool. These markers were instrumental in the genotyping process, which involved the analysis of a large number of shrimp samples to establish a robust database for ML.

The application of ML, in particular, supervised learning, showed high accuracy in classifying the origin of shrimp samples. The training data, derived from 636 small-shrimp samples, showed 100% accuracy in identifying the origin of test samples. This high level of precision underscored the potential of combining SNP analysis with ML for effective seafood origin determination. However, there are still challenges in identifying interspecific differences in Vietnamese samples. This remains a task to be addressed in future research.

The findings of this study have important implications for the seafood industry. Accurate origin determination is essential for maintaining consumer trust. By ensuring that products are correctly labeled, food fraud can be prevented and consumers can be protected from potential health risks associated with mislabeled seafood [30]. The method developed here provides a robust means for regulatory bodies to enforce labeling laws and standards. This can help to establish order in the distribution of seafood products and deter unscrupulous practices in the industry. For producers, especially those in regions with a reputation for high-quality seafood, accurate labeling can justify premium pricing. Highlighting the origin of products, such as Korean *A. chinensis* used in kimchi, can enhance market competitiveness and consumer preference.

However, as the countries of origin for imports become more diverse, the range of target species requiring additional analysis may increase. To address this issue, it will be necessary to expand the data set by including additional samples to improve the accuracy and performance of the analysis through continuous monitoring.

Overall, this study underscores the importance of accurate seafood origin determination for ensuring food safety, maintaining consumer trust, and supporting regulatory compliance. The successful integration of SNP analysis with ML provides a powerful

means of determining the origins of small shrimp, addressing the challenges posed by mislabeling. As the seafood industry continues to evolve, such innovative approaches will be crucial for safeguarding the integrity of food products and promoting sustainable practices. Continued research and technological advances will be required to expand these methods to a broader range of species and regions, ultimately benefiting consumers, producers, and regulators.

5. Conclusions

This study provides insights into the species and origins of salted small-shrimp products from Korea, China, Vietnam, and Malaysia through genetic analysis. The results confirmed that Korean and Chinese small shrimp belonged to the same species (*A. chinensis*), but showed differences in some nucleotide variations. Similarly, Vietnamese and Malaysian small shrimp were identified as the same species (*A. japonicus* and *A. indicus*), with genetic differences based on their origins. These genetic variations are considered to result from genetic fixation due to geographic isolation. Based on these genetic differences, this study developed genetic markers and used ML analysis to establish a method for distinguishing the origins of small shrimp. The findings underscore the value of genetic analysis for accurately identifying small-shrimp species in commercial products and revealing potential variations based on origin. In addition, this method can be applied across various fields and will contribute to food distribution safety.

Author Contributions: Conceptualization, E.S.N.; methodology, E.S.N. and M.N.L.; software, J.P. and H.S.J.; validation, M.N.L. and C.-M.D.; formal analysis, E.S.N.; investigation, E.S.N., W.-J.K. and Y.-O.K.; writing—original draft preparation, E.S.N.; writing—review and editing, E.S.N.; supervision, E.S.N.; project administration, E.S.N.; funding acquisition, E.S.N. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by a grant from the National Institute of Fisheries Science (R2024018), Republic of Korea.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data sets generated and/or analyzed in the present study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hicks, D. Seafood safety and quality: The consumer's role. *Foods* **2016**, *5*, 71. [[CrossRef](#)] [[PubMed](#)]
2. Lulijwa, R.; Rupia, E.J.; Alfaro, A.C. Antibiotic use in aquaculture, policies and regulation, health and environmental risks: A review of the top 15 major producers. *Rev. Aquac.* **2020**, *12*, 640–663. [[CrossRef](#)]
3. El Sheikha, A.F.; Xu, J. Traceability as a key of seafood safety: Reassessment and possible applications. *Rev. Fish. Sci. Aquac.* **2017**, *25*, 158–170. [[CrossRef](#)]
4. Onozaka, Y.; Honkanen, P.; Altintzoglou, T. Sustainability, perceived quality and country of origin of farmed salmon: Impact on consumer choices in the USA, France and Japan. *Food Policy* **2023**, *117*, 102452. [[CrossRef](#)]
5. Mizik, T. Agri-food trade competitiveness: A review of the literature. *Sustainability* **2021**, *13*, 11235. [[CrossRef](#)]
6. Claret, A.; Guerrero, L.; Aguirre, E.; Rincón, L.; Hernández, M.D.; Martínez, I.; Peleteiro, J.B.; Grau, A.; Rodríguez-Rodríguez, C. Consumer preferences for sea fish using conjoint analysis: Exploratory study of the importance of country of origin, obtaining method, storage conditions and purchasing price. *Food Qual. Prefer.* **2012**, *26*, 259–266. [[CrossRef](#)]
7. Noh, E.S.; Park, Y.J.; Kim, E.M.; An, C.M.; Park, J.Y.; Kim, K.H.; Song, J.H.; Kang, J.H. Development of primer set for the identification of fish species in surimi products using denaturing gradient gel electrophoresis. *Food Control* **2017**, *79*, 74–79. [[CrossRef](#)]
8. Kang, J.-H.; Kang, H.-S.; Noh, E.-S.; Park, J.-Y.; An, C.-M. Isolation and characterization of novel microsatellite markers for the northern mauxia shrimp, *Acetes chinensis*, using pyrosequencing. *Mar. Genom.* **2014**, *18*, 67–69. [[CrossRef](#)] [[PubMed](#)]
9. Kang, J.-H.; Noh, E.-S.; Park, J.-Y.; An, C.-M.; Choi, J.-H.; Kim, J.-K. Rapid origin determination of the northern mauxia shrimp (*Acetes chinensis*) based on allele specific polymerase chain reaction of partial mitochondrial 16S rRNA gene. *Asian Australas. J. Anim. Sci.* **2015**, *28*, 568–572. [[CrossRef](#)] [[PubMed](#)]

10. Kim, E.M.; Byun, S.G.; An, C.M.; Kang, J.H.; An, H.S.; Dong, C.M.; Kim, M.J.; Hong, Y.K.; Park, J.Y. Development of single nucleotide polymorphism markers from ESTs for discrimination between domestic and imported manila clams, *Ruditapes philippinarum*. *Food Control* **2014**, *40*, 208–213. [[CrossRef](#)]
11. Noh, E.S.; Kang, H.S.; An, C.M.; Park, J.Y.; Kim, E.M.; Kang, J.H. Rapid and specific identification of genus *Cynoglossus* by multiplex PCR assays using species-specific derived from the COI region. *J. Life Sci.* **2016**, *26*, 1007–1014. [[CrossRef](#)]
12. Noh, E.S.; Lee, M.N.; Kim, E.M.; Park, J.Y.; Noh, J.K.; An, C.M.; Kang, J.H. Development of a multiplex PCR assay for rapid identification of *Larimichthys polyactis*, *L. crocea*, *Atrubucca nibe*, and *Pseudolithus elongates*. *J. Life Sci.* **2017**, *27*, 746–753. [[CrossRef](#)]
13. Kim, E.-M.; Dong, C.M.; Lee, M.N.; Noh, J.K.; Noh, E.S.; Nam, B.H.; Kim, Y.-O.; Jung, H.S. Development of multiplex species-specific PCR for the simultaneous identification of three closely related species in the genera *Misgurnus* and *Paramisgurnus*. *Aquac. Rep.* **2022**, *24*, 101144. [[CrossRef](#)]
14. Asaad, I.; Lundquist, C.J.; Erdmann, M.V.; Costello, M.J. Ecological criteria to identify areas for biodiversity conservation. *Biol. Conserv.* **2017**, *213*, 309–316. [[CrossRef](#)]
15. Schwartz, M.K.; Luikart, G.; Waples, R.S. Genetic monitoring as a promising tool for conservation and management. *Trends Ecol. Evol.* **2007**, *22*, 25–33. [[CrossRef](#)] [[PubMed](#)]
16. Breed, M.F.; Harrison, P.A.; Blyth, C.; Byrne, M.; Gaget, V.; Gellie, N.J.C.; Hodgson, R.; Mills, J.G.; Prowse, T.A.A.; Steane, D.A.; et al. The potential of genomics for restoring ecosystems and biodiversity. *Nat. Rev. Genet.* **2019**, *20*, 615–628. [[CrossRef](#)] [[PubMed](#)]
17. Čandek, K.; Kuntner, M. DNA barcoding gap: Reliable species identification over morphological and geographical scales. *Mol. Ecol. Resour.* **2015**, *15*, 268–277. [[CrossRef](#)] [[PubMed](#)]
18. Leaché, A.D.; Oaks, J.R. The utility of single nucleotide polymorphism (SNP) data in phylogenetics. *Annu. Rev. Ecol. Evol. Syst.* **2017**, *48*, 69–84. [[CrossRef](#)]
19. Canturk, K.M.; Emre, R.; Kinoglu, K.; Başpınar, B.; Sahin, F.; Ozen, M. Current status of the use of single-nucleotide polymorphisms in forensic practices. *Genet. Test Mol. Biomark.* **2014**, *18*, 455–460. [[CrossRef](#)] [[PubMed](#)]
20. Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255–260. [[CrossRef](#)] [[PubMed](#)]
21. Ang, J.C.; Mirzal, A.; Haron, H.; Hamed, H.N.A. Supervised, unsupervised, and semi-supervised feature selection: A review on gene selection. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2016**, *13*, 971–989. [[CrossRef](#)] [[PubMed](#)]
22. Swan, A.L.; Mobasheri, A.; Allaway, D.; Liddell, S.; Bacardit, J. Application of machine learning to proteomics data: Classification and biomarker identification in postgenomics biology. *OMICS* **2013**, *17*, 595–610. [[CrossRef](#)] [[PubMed](#)]
23. Han, S.; Liang, Y.; Li, Y.; Du, W. Long noncoding RNA identification: Comparing machine learning based tools for long noncoding transcripts discrimination. *Biomed. Res. Int.* **2016**, *2016*, 8496165. [[CrossRef](#)] [[PubMed](#)]
24. Folmer, O.; Black, M.; Hoeh, W.; Lutz, R.; Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **1994**, *3*, 294–299. [[PubMed](#)]
25. Palumbi, S.R.; Martin, A.; Romano, S.; McMillan, W.O.; Stice, L.; Grabowski, G.; University of Hawaii at Manoa; Kewalo Marine Laboratory (Eds.) *The Simple Fool's Guide to PCR, Version 2.0*; Department of Zoology and Kewalo Marine Laboratory, University of Hawaii: Honolulu, HI, USA, 2002.
26. Librado, P.; Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **2009**, *25*, 1451–1452. [[CrossRef](#)] [[PubMed](#)]
27. Kumar, S.; Stecher, G.; Li, M.; Niyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [[CrossRef](#)] [[PubMed](#)]
28. Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.H. The WEKA data mining software: An update. *ACM SIGKDD Explor. Newsl.* **2009**, *11*, 10–18. [[CrossRef](#)]
29. Manning, L.; Soon, J.M. Food safety, food fraud, and food defense: A fast evolving literature. *J. Food Sci.* **2016**, *81*, R823–R834. [[CrossRef](#)] [[PubMed](#)]
30. Jurica, K.; Brčić Karačonji, I.; Lasić, D.; Bursać Kovačević, D.; Putnik, P. Unauthorized food manipulation as a criminal offense: Food authenticity, legal frameworks, analytical tools and cases. *Foods* **2021**, *10*, 2570. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.