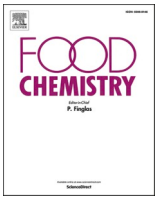
Food Chemistry 432 (2024) 137279

Contents lists available at ScienceDirect

Food Chemistry

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



MALDI-ToF MS and chemometric analysis as a tool for identifying wild and farmed salmon

Yunhe Hong [a], Nicholas Birse [a,*], Brian Quinn [a], Yicong Li [a], Wenyang Jia [a], Saskia van Ruth [b,d], Christopher T. Elliott [a,c]

[a] National Measurement Laboratory, Centre of Excellence in Agriculture and Food Integrity, Institute for Global Food Security, School of Biological Sciences, Queen's University Belfast, United Kingdom
[b] Food Quality and Design Group, Wageningen University and Research, Wageningen, Netherlands
[c] School of Food Science and Technology, Faculty of Science and Technology, Thammasat University, 99 Mhu 18, Pahonyothin Road, Khong Luang, Pathum Thani 12120, Thailand
[d] School of Agriculture and Food Science, University College Dublin, Dublin 4, Ireland

ARTICLE INFO

Keywords:
Food authenticity
Chemometric
Mass spectrometry
Salmon
MALDI-ToF

ABSTRACT

In this study, the difference between wild and farmed salmon production was successfully profiled and differentiated by matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-ToF MS) combined with chemometric analysis. The established method based on multivariate analysis mainly involved principal component analysis (PCA), partial least squares-discriminant analysis (PLS-DA), and orthogonal partial least squares-discriminant analysis (OPLS-DA) as the screening and verifying tools to provide insights into the distinctive features found in wild and farmed salmon products, respectively. The discrimination between farmed and wild salmon was accomplished with 100% classification accuracy using chemometric models, 100% identification accuracy was also achieved in distinguishing wild *Salmo salar* and *Oncorhynchus nerka* samples. The results of the present work suggest that the proposed method could serve as a reference for detecting salmon fraud relating to wild or farmed production and expand the application of MALDI-ToF technology further into food authenticity applications.

1. Introduction

In recent years the world has experienced significant population growth, and the demand for increased amounts of protein, including novel protein sources, has shown a significant increase. The use of seafood in general, specifically salmon, is one of the viable options to increase protein-rich foods available to consumers. Salmon is well-suited to commercial farming and also is readily available from traditional fisheries (wild-caught) (Nøstbakken et al., 2015). Salmon is a rich source of essential nutrients and is particularly sought after due to high levels of omega-3 fatty acids and low levels of saturated fats. It is also relatively abundant in essential elements such as iodine and selenium, and other nutrients such as vitamin D (Khalili Tilami & Sampels, 2018; Jakobsen et al., 2019; Colombo & Mazal, 2020).

The demand for salmon is also influenced by its cost-effectiveness and predictability in production and commercial harvest, making it an affordable and dependable source of protein for many consumers (Henriques et al., 2014).

Atlantic salmon (*Salmo salar*) and Pacific Sockeye salmon (*Oncorhynchus nerka*) are the two main species of salmon consumed in European countries (González et al., 2020; Ziegler & Hilborn, 2023). *Salmo salar* is both farmed commercially and caught from the wild, whereas all *Oncorhynchus nerka* are wild-caught (Mowi, 2021). Studies have demonstrated that wild salmon contains higher levels of omega-3 fatty acids and essential metals in comparison to farmed salmon (Thorstad et al., 2021; Lundebye et al., 2017). However, because of the high demand and price (Rickertsen et al., 2017), there have been instances of food fraud where farmed salmon has been deceptively labelled and sold as wild salmon (Kappel & Schröder, 2016). While it is possible to identify whole fish based on their morphology, smaller cuts of salmon can be challenging to authenticate, and once processed, they can be nearly impossible to visually differentiate (Cline, 2012).

The European Union recognised the growing concerns around mislabelling fraud in the seafood sector and implemented new labelling

* Corresponding author.
E-mail address: N.Birse@qub.ac.uk (N. Birse).

https://doi.org/10.1016/j.foodchem.2023.137279
Received 9 May 2023; Received in revised form 10 August 2023; Accepted 23 August 2023
Available online 26 August 2023

regulations for fishery and aquaculture products (Council Regulation (EC) No 104/2000 of 17 December 1999 on the Common Organisation of the Markets in Fishery and Aquaculture Products (Repealed), 2000). These regulations emphasise the commercial and scientific name, production modality (wild or farmed), and geographic origin information to help customers better understand the quality, origin, and production system of the products they are consuming (Council Regulation (EC) No 104/2000 of 17 December 1999 on the Common Organisation of the Markets in Fishery and Aquaculture Products (Repealed), 2000). However, the absence of an official analytical methodology to monitor mislabelling is a significant problem, despite the implementation of labelling regulations for fishery products. To address this issue, it is crucial to develop and validate a reliable and accurate analytical method for determining the authenticity of salmon production systems and the geographical origins of production. Accordingly, the objective of this project was to develop a rapid and reliable method capable of confirming the authenticity of both wild and farmed salmon from various geographic locations, thereby protecting both the salmon industry and consumers from the consequences of mislabelling fraud.

Given the increasing concerns about salmon fraud, several studies have been conducted to detect mislabelling fraud and confirm the authenticity of salmon products (Grazina et al., 2020). Thomas et al. (Thomas et al., 2008) developed a multi-probe and multi-element isotopic analysis combined with fatty acid analysis to determine the origin of salmon as wild or farmed. Fiorino et al. (Fiorino et al., 2019) more recently demonstrated that a Direct Analysis in Real Time-High Resolution Mass Spectrometry (DART-HRMS) platform can be a promising tool to distinguish between wild and farmed salmon. However, to enhance the applicability of the investigation method, a larger sample size is required for modelling purposes. Any newly developed assay would also require a high sample throughput to meet the demands of industry. In this context, a novel method using Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-ToF) has been developed to address these challenges. This method allows for the rapid and accurate identification of wild and farmed salmon production, thereby making it a valuable tool for ensuring the authenticity of salmon products.

MALDI-ToF works by ionising samples in the presence of a matrix substance and subjecting them to a laser pulse, causing the analyte molecules to become ionized (Zambonin et al., 2021). The resulting mass spectrum can then be analysed to identify the molecular components of the sample, providing accurate and sensitive detection of biomolecules such as proteins (Kritikou et al., 2022), peptides (Zhong et al., 2021), as well as small molecule compounds (Wang & Giese, 2017; Arroyo Negrete et al., 2019). MALDI-ToF has been deemed as an ideal tool for conducting high-throughput analyses, because it has relatively minimal sample preparation requirements (Murphy et al., 2021). Additionally, some degree of synergy exists with other applications in the seafood sector with Chai et al. (Chai et al., 2022) demonstrating a microchip-MALDI-TOF MS approach to investigate bacterial contamination in fish muscle commodities. Piras et al. (Piras et al., 2021) presented a MALDI-TOF method coupled with machine learning to discriminate milk from different mammalian species, and the method demonstrated 100% classification accuracy. Freitas et al. (Freitas et al., 2022) employed MALDI-ToF to achieve geographic discrimination of *Sparus aurata*, while Bi et al. (Bi et al., 2019) conducted species-level differentiation using a MALDI-ToF analysis of fish skin. The current applications of MALDI-ToF in food authenticity analysis represents a significant step forward in enhancing food safety measures and combatting fraudulent practices in the food industry. To date, MALDI-ToF has not yet been used in differentiating between wild-caught and farmed salmon.

Oils found in salmon are composed primarily of triglycerides (TAGs), which account for approximately 75% of the total lipid content (Cascant et al., 2018), whilst the number of fatty acids (FA) present in salmon results in a large variety of biomolecular species that make analysis difficult (Cascant et al., 2017). Thus, chemometric analysis was utilised for data analysis, which provides significant advantages by reducing data dimensions (Hong et al., 2023). Previous studies have demonstrated the utility of combining mass spectrometry with chemometric modelling or machine learning approaches for food authenticity applications (Birse et al., 2022). These studies have include the use of principal component analysis (PCA) (Wang et al., 2020), partial least squares-discriminant analysis (PLS-DA) (Bi et al., 2019), and orthogonal partial least squares-discriminant analysis (OPLS-DA) (Cajka et al., 2013). PCA is one of the most frequently used unsupervised multivariate data analysis methods in exploratory data analysis and data mining in food research (Cozzolino et al., 2019). PCA can also be used as a dimension-reduction technique to transform a large MS dataset into a smaller and more refined dataset whilst retaining much of the unique information found in the original dataset (Khulal et al., 2016). Thus, PCA modelling was used for an exploratory statistical analysis in this research to provide visualisations that show separation trends of MS data. PLS-DA is a linear classification tool that employs a partial least squares regression algorithm to calculate predictive models (Sampaio et al., 2020). OPLS-DA is a closely related algorithm, which is gaining increasing attention as a useful feature selector and sample classifier. OPLS-DA has recently been used for food fraud, authenticity, and quality relevant applications (Kang et al., 2022; Birse et al., 2020). OPLS-DA results from Trigg et al. (Trygg & Wold, 2002), who applied an orthogonal projection to the PLS-DA algorithm to construct the variations of the predictors correlated but orthogonal to the response. Accordingly, PLS-DA and OPLS-DA were evaluated to find the most suitable chemometric model for wild and farmed salmon identification.

In this study, a high-throughput, and accurate method combining MALDI-ToF MS with chemometric analyses was developed to profile and differentiate wild and farmed salmon production. This approach could serve as a reference for assessing fraud performed in commercially available wild and farmed salmon products.

2. Materials and methods

2.1. Reagents and chemicals

Acetonitrile (LC-MS grade) was supplied by Merck (Darmstadt, Germany). 1-Butanol (GPR grade) was supplied by VWR (Fontenay-sous-Bois, France). Trifluoroacetic acid (99%) was obtained from Sigma-Aldrich (St. Louis, USA). Sinapinic acid was supplied by Bruker Daltonics (Bremen, Germany). 18.2 MΩ/cm deionised water was obtained from a Millipore Milli-Q water-purification system (Billerica, MA, USA).

2.2. Sample collection and preparation

A total of 426 salmon muscle samples were obtained from a number of trusted sources and frozen at −20 °C prior to despatch and during transit from the relevant production facilities. The samples were then labelled upon arrival at the laboratory and stored at −20 °C, prior to analysis.

The salmon samples used in this study were obtained from different countries and production systems. Specifically, the Scottish samples (n = 78) and the Norwegian samples (n = 102) comprised exclusively of farmed *Salmo salar*. The Alaskan samples (n = 102) consisted solely of wild-caught *Oncorhynchus nerka*, and the Icelandic samples (n = 144) consisted of both farmed (n = 54) and wild-caught (n = 90) *Salmo salar*.

The samples for analysis were defrosted and approximately 20 g of sample was placed in a 50-mL centrifuge tube, the material was then refrozen at −45 °C before freeze-drying using a Lablyo freeze drier (Frozen in Time, York, UK) for a minimum of 48 h. The freeze-dried samples were then stored at −20 °C prior to extraction.

Extraction was undertaken by weighing 10 mg of freeze-dried sample into a 2-mL centrifuge tube then extracting with 1 mL of acetonitrile/1-butanol (1/1,v/v), as described by Cascant et al (Cascant et al., 2017).

The samples were then sonicated for 10 min using a camSonix C1275 sonicator (Camlab, Cambridge, UK) at maximum frequency, before being centrifuged using a Rotina 380R centrifuge (Hettich, Tuttlingen, Germany) at 5000 rpm for 10 min.

A sinapinic acid matrix solution was prepared by dissolving 10 mg of sinapinic acid in 1 mL of 0.1% trifluoroacetic acid (TFA) solution/acetonitrile (7/3, v/v) and mixed (1/1, v/v) with 0.5 μL of extract solution. The sample-matrix solution was pipetted onto a Bruker MTP Target Plate - Ground Steel (Bruker Daltonics, Bremen, Germany) and allowed to dry at room temperature.

2.3. MALDI-ToF MS analysis

Samples were analysed using a Bruker autoflex maX MALDI-ToF mass spectrometer (Bruker Daltonics, Bremen, Germany). The spectra were recorded in the LIFT positive mode (laser intensity 80%, ion source 1 = 19.00 kV, ion source 2 = 16.35 kV, lens = 8.35 kV, linear detector voltage = 2.6 kV). Mass spectrometric data were acquired over the range *m/z* 0–1500. Each spectrum corresponded to an ion accumulation of 2,000 laser shots randomly distributed on the desired spot (laser beam attenuation = 20, laser beam focus = 33, laser repetition rate = 2000 Hz).

The mass spectra from Alaskan wild salmon, Icelandic wild salmon, Icelandic farmed salmon, Scottish farmed salmon, and Norwegian farmed salmon, respectively, are shown in Fig. S1.

2.4. Data processing

MALDI-ToF data was acquired using Bruker FlexControl v3.4 (Build 169.5) (Bruker Daltonics, Bremen, Germany) and processed using Bruker FlexAnalysis v3.4 (Build 79) (Bruker Daltonics, Bremen, Germany). The spectra obtained were processed with default parameters for smoothing, variance stabilisation, baseline correction and peak detection, then exported in TXT, CSV and mzXML formats in readiness for further data preparation.

Mass binning at 0.2 Da was performed at this stage using R (v4.0.5, Vienna, Austria), and the mass range was set to 100 Da to 1500 Da, giving a total of 7000 mass bins for multivariate and statistical analysis. The MALDI-ToF data was normalised using the UV (Unit Variance) scaling method.

Multivariate analysis was undertaken in several stages using MetaboAnalyst 4.0 (Quinn et al., 2022). Unsupervised analysis was evaluated using PCA, while supervised analyses were undertaken using PLS-DA and OPLS-DA models. A receiver operating characteristic (ROC) analysis and an ANOVA were performed.

PCA was used to detect data trends and pattern analyses, and to perform an initial quality control check of the data generated. PLS-DA and OPLS-DA models were used to identify biomarker candidates that differentiated the wild and farmed salmon groups. The validity of the unsupervised and supervised models was verified using correlation coefficients (R^2) and cross-validation correlation coefficients (Q^2). The quality of the built models was assessed by the goodness of fit (R^2X), the proportion of the response matrix variance explained by the model (R^2Y) and the predictive performance of the model (Q^2Y). These three metrics have values between 0 and 1, and the higher they are, the better the performance of the model (Dinis et al., 2022). Generally, $Q^2(Y)$ values greater than 0.5 are regarded as good for biological models (Verplanken et al., 2017). The number of primary components utilised in modelling is represented by A, and N represents the number of samples analysed.

The five-fold cross-validation were used, designating 1/5 of the data as a hold-out test set in each iteration. The model was trained using 4/5 of the data, which was subsequently used to predict the classifications of the remaining 1/5. This process was executed five times, each instance using a different partition for prediction, with the model being trained on the complementary four partitions.

3. Results and discussion

3.1. Multivariate data analysis (MVA) of salmon specimen dataset

Unsupervised and supervised MVA were employed for the assessment of heterogeneity in the salmon metabolome for salmonid species differentiation, in addition to the country of origin in a targeted and untargeted form.

3.1.1. Principal components analysis of the whole-sample dataset

Mass spectra were generated from 426 samples representing five groups including Iceland farmed, Iceland wild, Alaska wild, Scotland farmed, and Norway farmed samples using the above-described method. The unsupervised PCA analysis was used to evaluate the sample differences. The PCA score plot showed clear separation amongst five salmon groups in PC1 and PC2 (Fig. 1a). The score plots across PC1-PC5 in differentiate five salmon groups are shown in Fig. S2(a). $R^2(X) = 0.99$ indicated the PCA model is well fitted, and $Q^2 = 0.98$, signifying the goodness of prediction ability of the model. The screen plot of the PCA variance explained were shown in Fig. 1(b). The variance explained by all PCs is shown by the green line at the top and the individual PC variance is shown by the blue line below. The first principal component (PC1) explained 59.9 cumulative percent (cum %) of the variance. The PC2 explained 26.7 cum %. PC3, PC4, and PC5 explained 6.7 cum %, 1.6 cum %, and 1.2 cum %, respectively. The first five principal components with 96.0 cum % of the total variance that could be considered adequate to demonstrate the variety of salmon samples.

A clear separation between wild (Iceland and Alaska) and farmed (Iceland, Norway, and Scotland) groups can be seen in Fig. 1(c). Therefore, the data were divided into wild and farmed groups for principal component analysis. The PCA plot (Fig. 1d) clearly demonstrates a distinct separation between the data from the wild and farmed groups, indicating the robustness and reproducibility of intra-group sample analysis. The first 3 principal components (PCs) can account for 93.3% of the observed group differences (Fig. 1e). The results indicated that a PCA based on MALDI-ToF datasets of all 426 salmon samples revealed a clear separation between wild salmon and farmed salmon and clearly showed separation amongst 5 different salmon groups.

3.1.2. Metabolic profiling analysis and univariate analysis

Fig. 2a presents the correlation heatmaps of 5 groups of salmon samples. The expression profile of MALDI-ToF MS results was sufficient to discriminate 5 groups of salmon samples. Interestingly, Icelandic wild salmon exhibited weak correlations with other sample groups, with all correlation coefficients registering values below 0.5. To better interpret the MS information from the salmon analysis datasets, heatmaps were generated from the top 20 features in all samples amongst 5 groups using an ANOVA test (Fig. 2b). The top 20 features represent the top 20 Variable Importance in the Projection (VIP) scores identified through ANOVA, which were selected as biomarkers. The VIP values corresponding to these 20 features are provided in Table S1.

In the heatmaps, blue indicates a lower intensity MS signal as opposed to red, which shows a higher intensity MALDI-ToF MS signal.

The box plots provide a detailed view of the distribution of these 20 features within each group (Fig. 2c). The differences between the wild and farmed groups are clearly to be seen in the box plot, constructed using mass bins *m/z* 120.8, *m/z* 146.8, *m/z* 148.8, *m/z* 152.4, *m/z* 162.8, *m/z* 181.2, *m/z* 191.2, *m/z* 206.8, *m/z* 210.4, and *m/z* 211.4. Compared to the other three farmed groups, the ion abundance of these features was found to be lower in the Alaska and Iceland wild groups. The mass intensity of mass bins with *m/z* 377.0, *m/z* 377.2, *m/z* 378.2, *m/z* 378.4, and *m/z* 379.2 was found to be higher in Scottish farmed salmon, compared to the other four groups. Additionally, a smaller mass bin value (*m/z* 104.2, *m/z* 105.2, *m/z* 116.2, and *m/z* 124.2) was found to have higher mass intensity in the Alaskan salmon group.

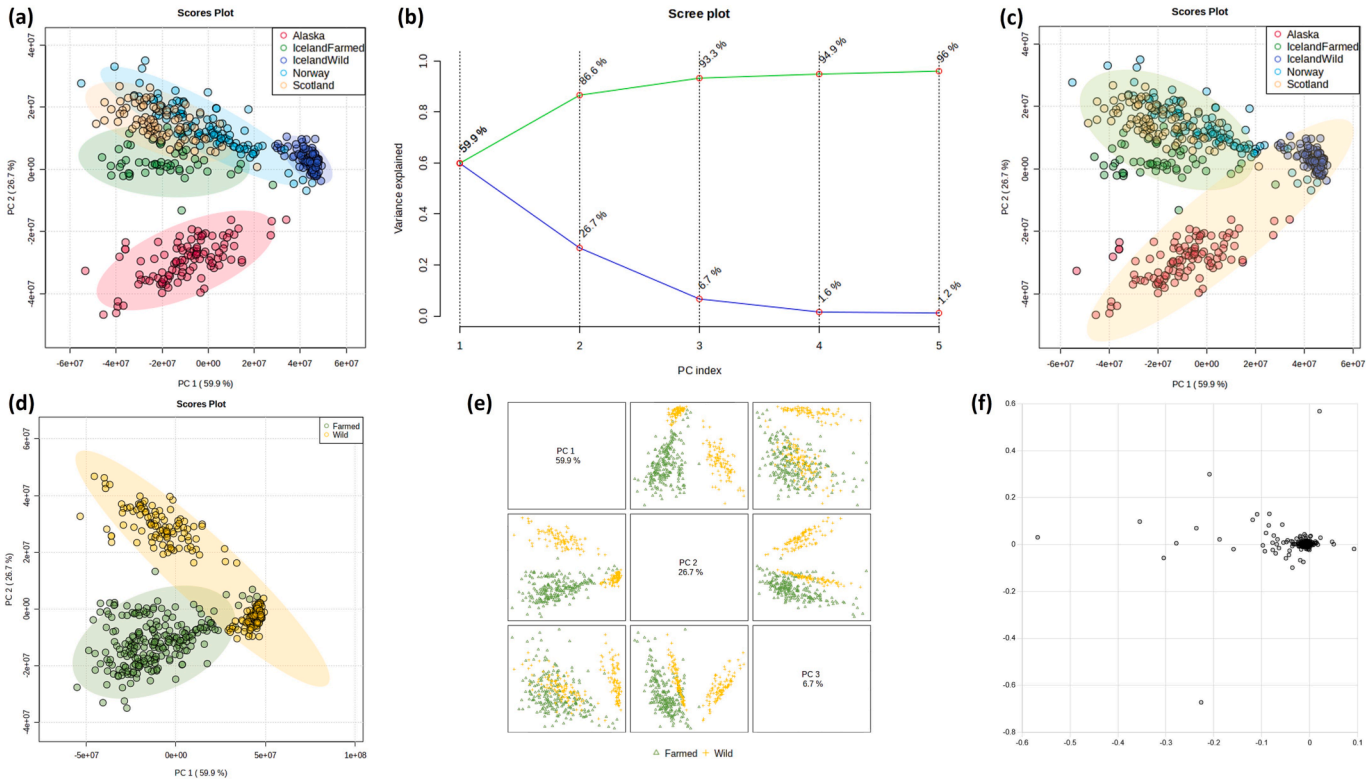


Fig. 1. PCA results. (a) PCA score plot between PC1 and PC2 for 5 groups of salmon samples (b) Scree plot. (c) PCA Plot of PC1 and PC2 for five salmon groups differentiated by wild (yellow) and farmed (green) production type. (d) PCA plot of PC1 and PC2 between wild and farmed salmon groups. (e) Data distribution between wild-type and farmed salmon group in PC1, PC2, and PC3. (f) Loading plot of PC1 and PC2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Two PLS-DA models were built using the whole dataset (Fig. 3a) and 20 biomarkers (Fig. 3b), respectively, populated by the mass spectra of 5 salmon groups, followed by an assessment of the modelling using a five-fold cross-validation. The differences between Alaska wild salmon and the other four groups can be found from PLS-DA loading plot by using the whole dataset (Fig. 3a). This likely resulted from Alaskan salmon being from a completely different genus/species *Oncorhynchus nerka*, while the other four groups are all *Salmo salar*, suggesting the variance is between *Salmo salar* and *Oncorhynchus nerka*. The PLS-DA model showed good performance with high R^2X, R^2Y, and Q^2 values of 0.99, 0.94, and 0.91, respectively. The first five components of the model accounted for 95.6% of the total variance (Fig. S2b), indicating its strong validity and predictive capability. The 5-fold cross-validation assessment demonstrated a 100% accuracy in identifying the 5 different salmon groups (Fig. 3c).

Additionally, the 20-feature PLS-DA model (Fig. 3b) shows slightly different results compared to Fig. 3(a). These top 20 metabolite markers exhibited a better separation amongst wild (Alaska and Iceland) and farmed (Scotland, Norway, and Iceland) groups in the PLS-DA loading plot. This PLS-DA model explained 99.9% of the total variance with components 1 to 3 (Fig. S2c). The R^2X, R^2Y, and Q^2 values of 0.99, 0.86, and 0.85, respectively, indicate that the PLS-DA model was highly reliable and possessed strong predictive capabilities. The cross-validation outcomes demonstrated that the 20-feature PLS-DA model achieved a classification accuracy of 97.4% in distinguishing among five groups of salmon samples.

3.1.3. *Wild and farmed salmon authenticity validation*

A supervised model was developed to authenticate global salmon production. Both the PLS-DA and OPLS-DA models demonstrate good classification results in wild (n = 192) and farmed (n = 234) salmon authenticity validation. The PLS-DA model (Fig. 4a) was a good fit for classifying wild and farmed salmon. The R^2X, R^2Y, and Q^2 values were 0.99, 0.98, and 0.93, respectively. Similarly, the OPLS-DA model (Fig. 4b) had R^2X, R^2Y, and Q^2 values of 0.99, 0.98, and 0.96, indicating that PLS-DA model and OPLS-DA model were not overfit and had strong predictive abilities. Cross-validation of both the PLS-DA and OPLS-DA models showed a 100% accuracy rate.

Fig. 4(c) illustrates the receiver operating characteristic (ROC) curve for the salmon modality production predictive models between wild salmon and farmed salmon groups (whole-sample dataset). The separation and discriminatory ability of the predictive models is demonstrated in ROC curves analysis. The area under the ROC curve (AUC) is 0.99 indicating that ROC analysis was acceptable based on a PLS-DA model with good sensitivity (99%) and specificity (99%), and the confidence intervals are greater than 97.5%. Two thirds of the samples were used to evaluate importance of the marker elements and to build classification models which were then validated using one third of the samples (Quinn et al., 2022). To assess the performance and confidence interval of the model, the process was repeated multiple times, and the predicted class probabilities (average of the cross-validation) for each sample using the PLS-DA classifier was shown in Fig. 4(d). All samples were correctly identified with a 99% accuracy.

The top 20 features in VIP analysis, combined with high S-plot (Fig. 4e) reliability (correlation) p[1] > |5E6| and p(corr) > |0.8| (Fig. 4f), were then selected as biomarkers (Table S1). PLS-DA model (Fig. 4g) and OPLS-DA (Fig. 4h) model were developed using these 20 features. Clear separation was obtained between wild and farmed salmon group. Fig. 4(g) displays the PLS-DA scores plot revealing the identification of two distinct clusters. The first two components accounted for a cumulative 98.8% of the total variation, with the first component explaining 98.4% of the variation and the second component explaining the variation between wild and farmed salmon samples (0.4%), while the cumulative variation of first 5 components was shown

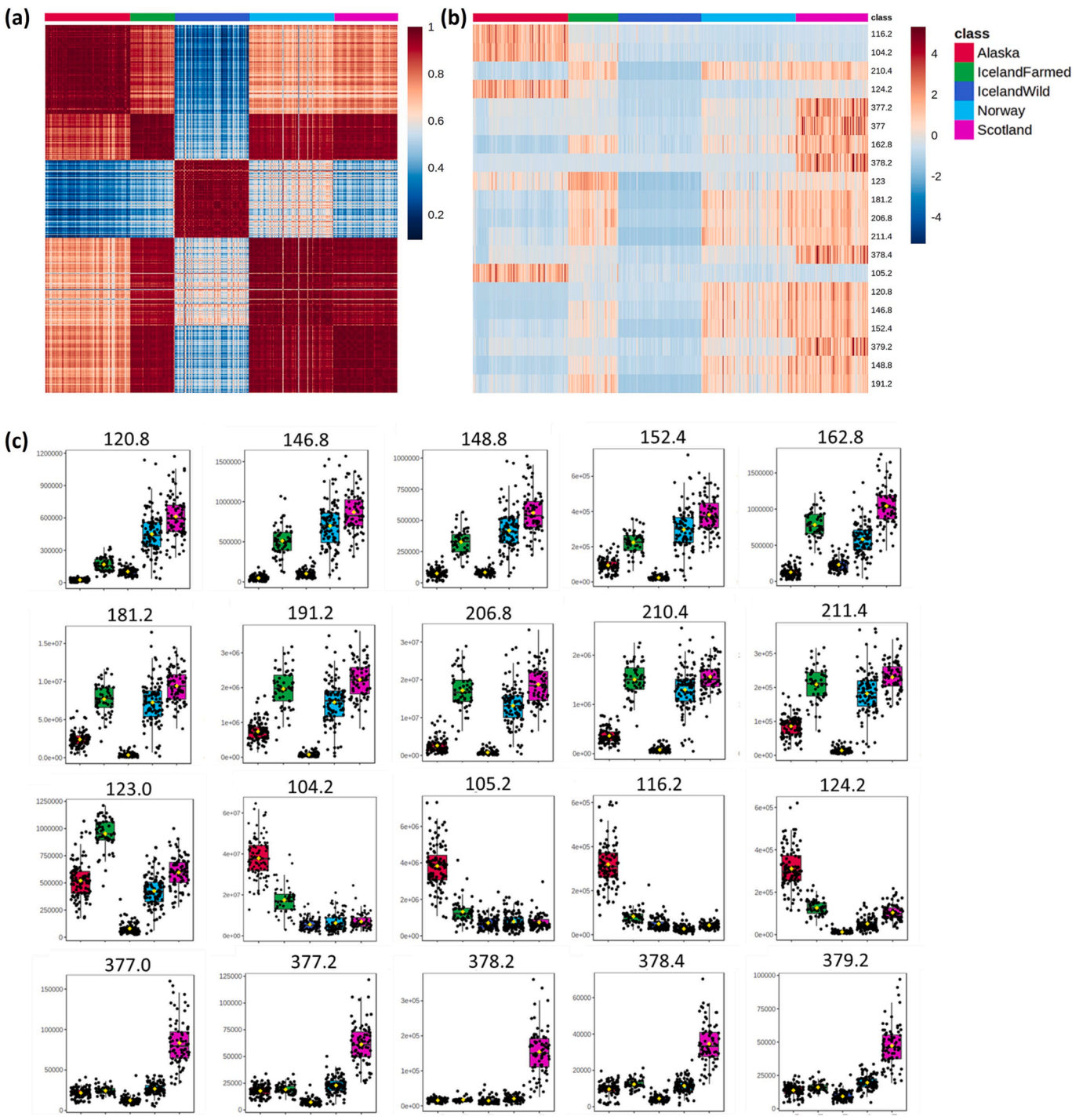


Fig. 2. Metabolic profiling and univariate analysis results: (a) the correlation heatmaps of these 426 salmon samples. (b) Heatmap of top 20 features in salmon samples among 5 groups. (c) Box plots of top 20 significant metabolites according to ANOVA.

in Fig. S2(e). $R^2X = 0.99$, $R^2Y = 0.92$, $Q^2 = 0.89$, and accuracy = 99.3% shows good fit of the PLS-DA model. The OPLS-DA resulted in 20 components with $R^2X = 0.99$, $R^2Y = 0.92$, and Q^2 (cum) = 0.91, evaluating using five-fold cross-validation. The classification accuracy of 99.3% was achieved for differentiation wild and farmed salmon samples.

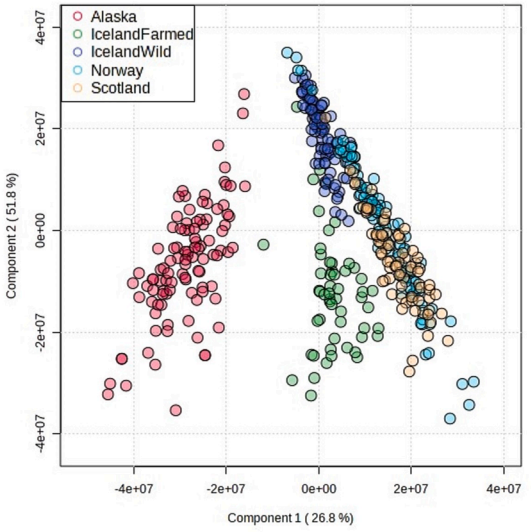
The results demonstrate that the combination of MALDI-ToF MS and chemometric analysis is competent in verifying the authenticity of wild-farmed salmon samples.

3.1.4. *Discriminant analysis of Salmo salar*

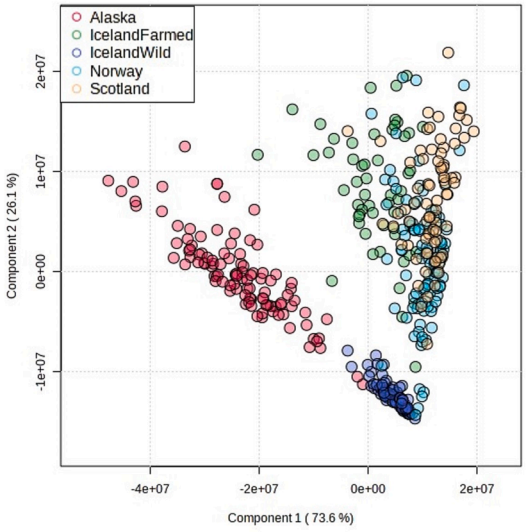
Salmo salar is the most commonly farmed salmon species in Europe and the most abundantly wild-caught salmon species in North Atlantic. Four groups of *Salmo salar* sample analysis results were examined to develop a model to identify the authenticity of *Salmo salar* production.

The unsupervised PCA model obtained from the MALDI-ToF MS analysis of all *Salmo salar* samples revealed the general structure of the complete dataset. The PCA plot (Fig. 5a) shows a clear two-part separation amongst farmed Norwegian salmon, farmed Scottish salmon, farmed Icelandic salmon, and wild Icelandic salmon. Thus, these 4 groups of *Salmo salar* were classified as the farmed group (farmed Norwegian salmon, farmed Scottish salmon, farmed Icelandic salmon) and the wild group (wild Icelandic salmon) for further analysis. The PLS-

(a)



(b)



(c)

Classification

Real Groups	Alaska Wild	Scotland Farmed	Norway Farmed	Iceland Farmed	Iceland Wild	Total
Alaska Wild	102 100%	0	0	0	0	102
Scotland Farmed	0	78 100%	0	0	0	78
Norway Farmed	0	0	102 100%	0	0	102
Iceland Farmed	0	0	0	54 100%	0	54
Iceland Wild	0	0	0	0	90 100%	90
Total	102	78	102	54	90	426

Group	Number of spectra	Number of passes	Number of failures	Number of outliers	Correct classification rate
Total	426	426	0	0	100%

(d)

Classification

Real Groups	Alaska Wild	Scotland Farmed	Norway Farmed	Iceland Farmed	Iceland Wild	Total
Alaska Wild	100 98.04%	0	0	0	2 1.96%	102
Scotland Farmed	0	74 94.87%	4 5.13%	0	0	78
Norway Farmed	0	0	97 95.10%	0	5 4.90%	102
Iceland Farmed	0	0	0	54 100%	0	54
Iceland Wild	0	0	0	0	90 100%	90
Total	100	74	97	54	90	426

Group	Number of spectra	Number of passes	Number of failures	Number of outliers	Correct classification rate
Total	426	415	0	0	97.42%

Fig. 3. Differential analysis among Alaskan wild salmon, Icelandic farmed salmon, Icelandic wild salmon, Norwegian farmed salmon, and Scottish farmed salmon. (a) PLS-DA for discrimination of 5 salmon groups using whole dataset. (b) PLS-DA model created by using top 20 metabolite markers. (c) Cross-validation classification table using whole dataset PLS-DA model. (d) Cross-validation classification table using the 20-biomarker PLS-DA model.

DA model and OPLS-DA model was used to investigate wild and farmed *Salmo salar*. Within a total of 426 salmon samples dataset, 324 of these were *Salmo salar*, with 90 of these being wild-caught salmon and 234 being farmed salmon. However, a significant difference in sample sizes among the various groups could potentially distort the model. To address this issue, three groups of farmed salmon were randomly sampled according to their proportional representation, thereby ensuring balance in the dataset. Specifically, in total of 180 *Salmo salar* samples were used for model building, 90 farmed salmon (21 samples from the Iceland-farmed group, 30 from the Scotland group, and 39 from the Norway group) and 90 wild salmon (Iceland wild). The remaining samples were held for use as external validation datasets.

The PCA between wild and farmed *Salmo salar* was performed (Fig. 5b), in which first three principal components cumulatively accounted for 95.4% of the total variation with PC1 explaining 86.3%, PC2 explaining 6.4%, and PC3 explaining 2.8% of the total variation respectively (Fig. 5c). The values of $R^2X = 0.98$ and $Q^2 = 0.93$ demonstrate both the goodness of fit and the predictive capability of the PCA model. The PLS-DA model generated from wild and farmed *Salmo salar* sample varieties revealed the overall structure of the complete data set, with components 1 to 5 explaining 86.2%, 5.7%, 3%, 1.2%, and 0.8% of the total variance, respectively (Fig. S2f). Clear clustering was observed between the wild and farmed sample groups, as evident from the PLS-DA scores plot presented in Fig. 5(d). To assess the predictive accuracy, five-fold cross-validation was performed. The cumulative values of the PLS-DA model, R^2X value of 0.89, R^2Y value of 0.98, and Q^2 value of 0.96, and a cross-validation correct classification rate of 100% suggest that the model has the potential to predict the classification of *Salmo salar*. The OPLS-DA model (Fig. 5e) also demonstrates good discriminatory ability in wild and farmed *Salmo salar* identification, $R^2X = 0.91$, $R^2Y = 0.98$, and $Q^2 = 0.97$ indicate that the model has a high capability to explain the wild and farmed *Salmo salar* differences. The results of 100% correct classification rate of cross-validation of OPLS-DA model indicating the model allows the authenticity analysis of the wild and farmed *Salmo salar* production patterns.

ROC curve analysis was performed to evaluate the performance of the model in predicting wild and farmed *Salmo salar* samples. The AUC = 1 indicating that ROC analysis was good based on PLS-DA model with good sensitivity (100%) and specificity (100%), and the confidence intervals are 100% in discriminate wild and farmed *Salmo salar* samples. Twenty biomarkers were found between the wild and farmed *Salmo salar* groups (Table S2) according to VIP (Fig. S3a) and S-plot (Fig. S3b) results. A PLS-DA model (Fig. S3c) and an OPLS-DA (Fig. S3d) model were developed using these 20 biomarkers. The classification accuracy of the PLS-DA model was 98.8% and 98.4% for the OPLS-DA model when evaluating the differentiation between wild and farmed *Salmo salar* samples.

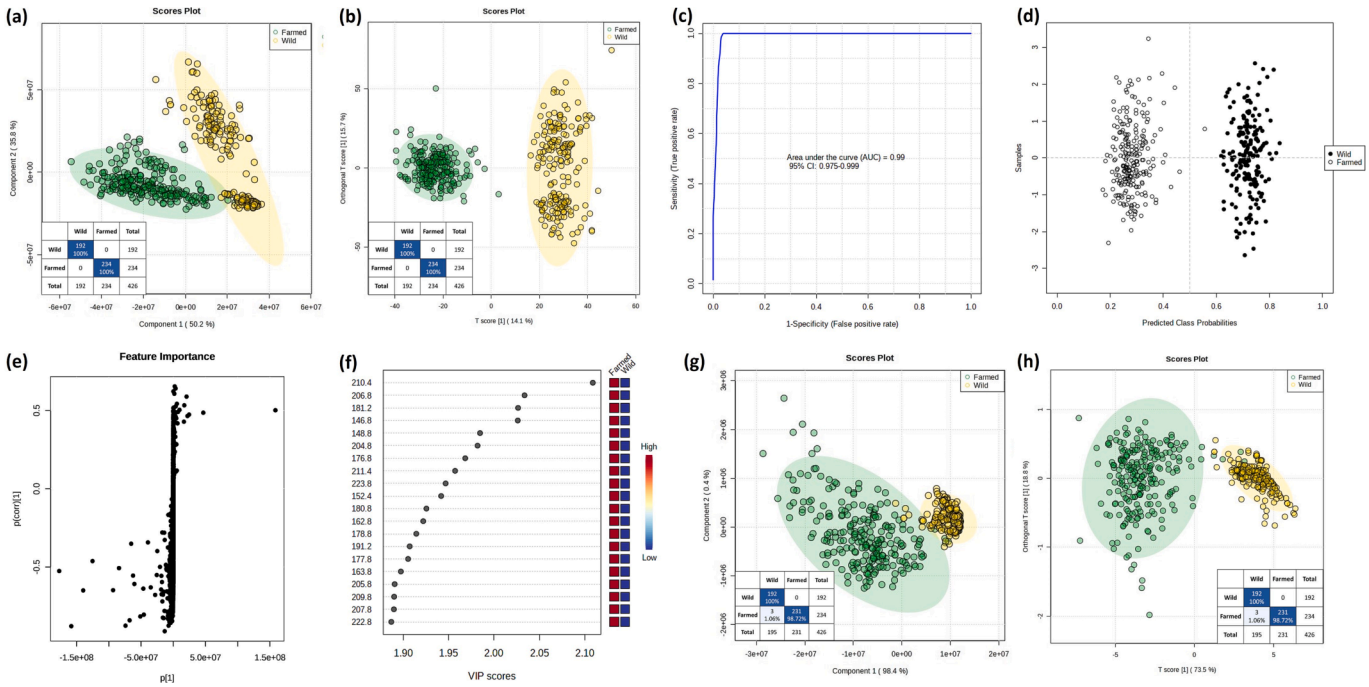


Fig. 4. Supervised model for authenticity validation of salmon modality production: (a) the PLS-DA scores plot between wild and farmed salmon group (component 1 and component 2). (b) OPLS-DA score plot of wild salmon and farmed salmon group. (c) ROC analysis between wild salmon and farmed salmon group. The specified number of latent variables (LV) for each model is set at 10. (d) The predicted class probabilities for each sample. (e) s-plot. (f) VIP scores plot with top 20 features (g) PLS-DA model created using 20 markers. (h) OPLS-DA model created using 20 markers.

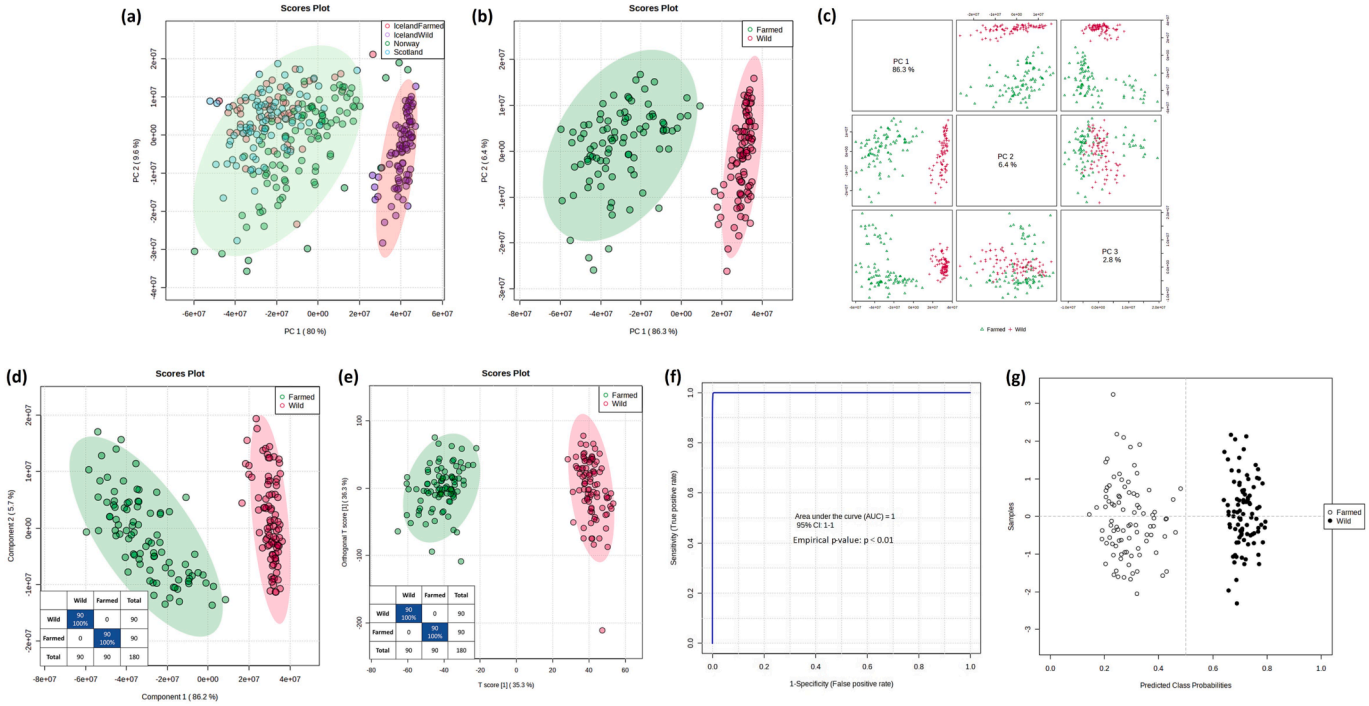


Fig. 5. Discriminant analysis results of *Salmo salar*: (a) the PCA scores plot amongst 4 groups *Salmo salar*. (b) The PCA scores plot between wild and farmed *Salmo salar*. (c) The first 5 PC scores plot between wild and farmed *Salmo salar*. (d) PLS-DA score plot and CV results of wild and farmed *Salmo salar*; (e) OPLS-DA score plot and CV results. (f) ROC analysis between wild salmon and farmed salmon group, AUC = 1. The specified number of LVs for each model is set at 10. (g) The predicted class probabilities for samples.

3.1.5. Discriminant analysis of wild Salmo salar and wild Oncorhynchus nerka

The use of advanced models for accurate classification between wild *Salmo salar* and *Oncorhynchus nerka* requires consideration in this research. Thus, the difference between two groups of wild salmon was further evaluated. The PCA score plot for wild *Salmo salar* and *Oncorhynchus nerka* was shown in Fig. 6(a), the first three principal components (PC1-PC3) contributed to 95.5% of the total explained variations

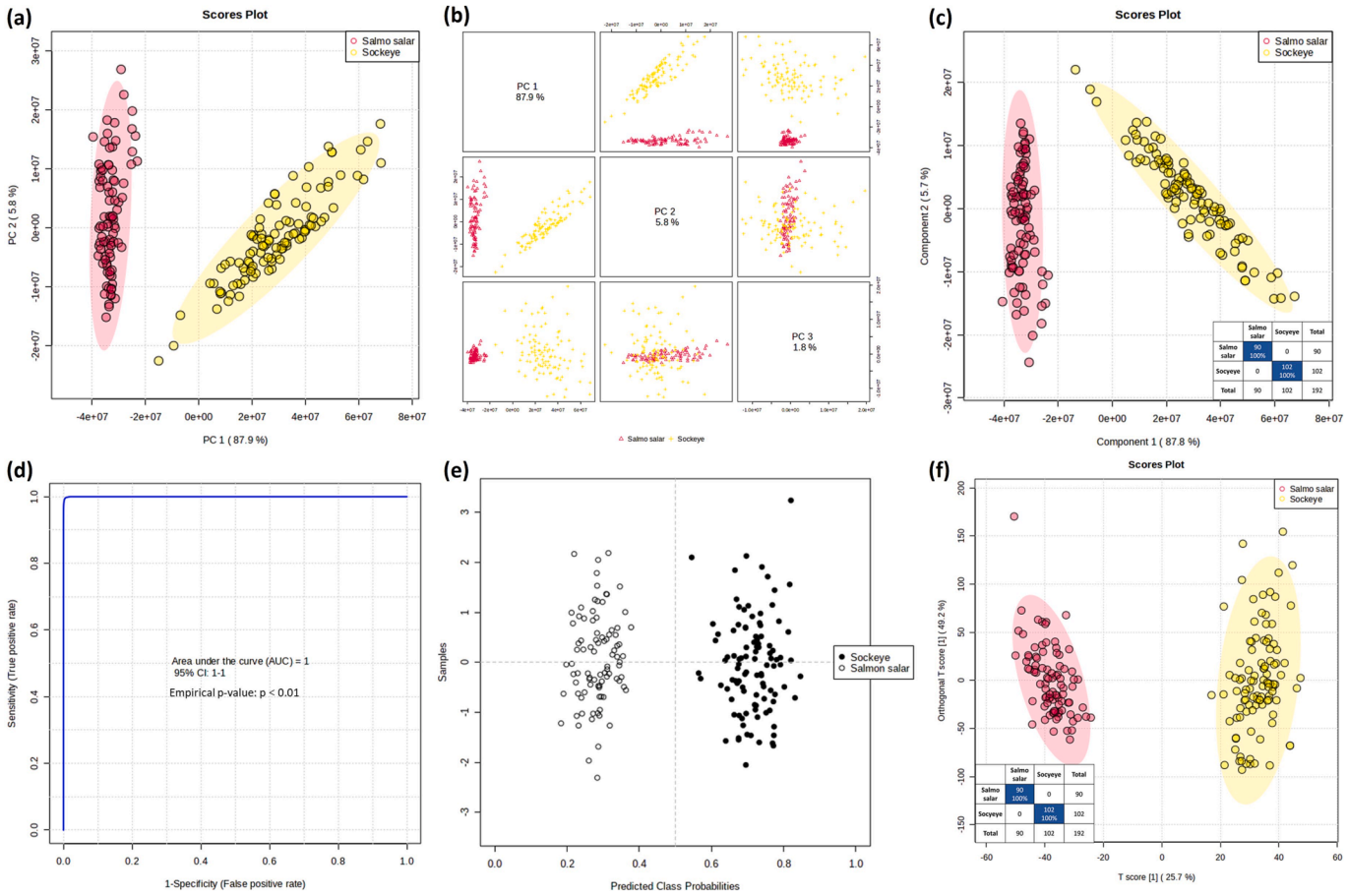


Fig. 6. Discriminant analysis results of wild *Salmo salar* and *Oncorhynchus nerka*; (a) the PCA scores plot between 2 wild groups. (b) The first 3 PC scores plot between 2 wild groups. (c) PLS-DA score plot. (d) ROC analysis for 2 datasets of wild *Salmo salar* and *Oncorhynchus nerka*. The specified number of LVs for each model is set at 10. (e) The predicted class probabilities for each sample. (f) OPLS-DA score plot and CV results.

(Fig. 6b). The model exhibited good performance when the initial 20 variables were employed to construct the PCA, resulting in high values for both R^2X (0.99) and Q^2 (0.99).

The PLS-DA model (Fig. 6c) demonstrated great applicability and predictive ability with the $R^2X = 0.99$, $R^2Y = 0.99$, $Q^2 = 0.93$, indicating that the model is in good fit and has a high level of predictive ability. The cross-validation result showed that the model achieved a 100% correct classification rate, indicating that it is highly accurate and reliable in predicting the samples. While The ROC curve between two wild groups is shown in Fig. 6(d). A value of AUC = 1 indicates that the prediction model can achieve a perfect prediction using at least one threshold. The predicted class probabilities using the PLS-DA classifier are shown in Fig. 6(e) with 100% accuracy. The OPLS-DA model achieved an R^2X value of 0.97, R^2Y value of 0.99, and Q^2 value of 0.98 and had a cross-validation correct classification rate of 100%, indicating that the model was not overfitted and had good prediction capabilities.

To distinguish between wild *Salmo salar* and *Oncorhynchus nerka* samples, the top twenty significant features were selected according to VIP analysis (Fig. S4a) and S-plot (Fig. S4b). The results are listed in Table S3. PLS-DA model (Fig. S4c) and OPLS-DA model (Fig. S4d) were developed using these 20 features. A classification accuracy of 100% was achieved in distinguishing between wild *Salmo salar* and *Oncorhynchus nerka* samples using both the PLS-DA and OPLS-DA models.

4. Conclusions

In this study, we demonstrated the potential of using MALDI-ToF MS coupled with chemometric analysis for the discrimination of wild and farmed salmon samples. The differences between wild and farmed *Salmo salar*, and the differences between wild *Salmo salar* and wild *Oncorhynchus nerka* were also explored. This technology provides the basis for the high-throughput identification of salmon resources in the market and unlocks the potential application areas of MALDI-ToF MS. The PCA, PLS-DA, and OPLS-DA models were successfully used in sample classification with 100% classification accuracy in wild and farmed salmon samples authentication. Owing to the high capacity to profile and distinguish amongst the different types of salmon samples, MALDI-ToF MS combined with chemometric analysis could be viewed as an effective approach for identifying salmon fraudulence related to wild/farmed production methods. Considering the important and growing role salmon currently plays in global protein production, the development of such technologies can effectively monitor incidences of fraud in the seafood sector and further safeguard the legal rights of consumers.

CRediT authorship contribution statement

Yunhe Hong: Conceptualization, Writing – review & editing. **Nicholas Birse:** Conceptualization, Writing – review & editing. **Brian Quinn:** Supervision, Project administration, Writing – review & editing. **Yicong Li:** . **Wenyang Jia:** Conceptualization. **Saskia van Ruth:** Supervision, Project administration. **Christopher T. Elliott:** Conceptualization, Supervision, Project administration, Writing – review & editing, Funding acquisition.

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.

Data availability

Data will be made available on request.

Acknowledgements

The authors would like to thank the Bruker Corporation for loaning us the MALDI-ToF instrument and funding support. This study was also supported by the Bualuang ASEAN Chair Professor Fund.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2023.137279.

References

Arroyo Negrete, M. A., Wrobel, K., Yanez Barrientos, E., Corrales Escobosa, A. R., Acevedo Aguilar, F. J., & Wrobel, K. (2019). Determination of sulfonated azo dyes in chili powders by MALDI-TOF MS. *Analytical and Bioanalytical Chemistry, 411*(22), 5833–5843. https://doi.org/10.1007/s00216-019-01965-1

Bi, H., Zhong, C., Shao, M., Wang, C., Yi, J., Qiao, L., & Zhang, J. (2019). Differentiation and authentication of fishes at the species level through analysis of fish skin by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry, 33*(16), 1336–1343. https://doi.org/10.1002/RCM.8474

Birse, N., Chevallier, O., Hrbek, V., Kosek, V., Hajšlová, J., & Elliott, C. (2020). Ambient mass spectrometry as a tool to determine poultry production system history: A comparison of rapid evaporative ionisation mass spectrometry (REIMS) and direct analysis in real time (DART) ambient mass spectrometry platforms. *Food Control, 107740*. https://doi.org/10.1016/j.foodcont.2020.107740

Birse, N., McCarron, P., Quinn, B., Fox, K., Chevallier, O., Hong, Y., … Elliott, C. (2022). Authentication of organically grown vegetables by the application of ambient mass spectrometry and inductively coupled plasma (ICP) mass spectrometry The leek case study. *Food Chemistry, 370*, Article 130851. https://doi.org/10.1016/J.FOODCHEM.2021.130851

Cajka, T., Danhelova, H., Vavrecka, A., Riddellova, K., Kocourek, V., Vacha, F., & Hajslova, J. (2013). Evaluation of direct analysis in real time ionization–mass spectrometry (DART–MS) in fish metabolomics aimed to assess the response to dietary supplementation. *Talanta, 115*, 263–270. https://doi.org/10.1016/J.TALANTA.2013.04.025

Cascant, M. M., Breil, C., Fabiano-Tixier, A. S., Chemat, F., Garrigues, S., & de la Guardia, M. (2018). Determination of fatty acids and lipid classes in salmon oil by near infrared spectroscopy. *Food Chemistry, 239*, 865–871. https://doi.org/10.1016/J.FOODCHEM.2017.06.158

Cascant, M. M., Breil, C., Garrigues, S., de la Guardia, M., Fabiano-Tixier, A. S., & Chemat, F. (2017). A green analytical chemistry approach for lipid extraction: computation methods in the selection of green solvents as alternative to hexane. *Analytical and Bioanalytical Chemistry 2017 409:14*, *409*(14), 3527–3539. https://doi.org/10.1007/S00216-017-0323-9.

Chai, Z., Soko, W. C., Xie, J., & Bi, H. (2022). Microchip coupled with MALDI-TOF MS for the investigation of bacterial contamination of fish muscle products. *Food Chemistry, 396*, Article 133658. https://doi.org/10.1016/J.FOODCHEM.2022.133658

Cline, E. (2012). Marketplace substitution of Atlantic salmon for Pacific salmon in Washington State detected by DNA barcoding. *Food Research International, 45*(1), 388–393. https://doi.org/10.1016/J.FOODRES.2011.10.043

Colombo, S. M., & Mazal, X. (2020). Investigation of the nutritional composition of different types of salmon available to Canadian consumers. *Journal of Agriculture and Food Research, 2*, Article 100056. https://doi.org/10.1016/j.jafr.2020.100056

Council Regulation (EC) No 104/2000 of 17 December 1999 on the common organisation of the markets in fishery and aquaculture products (repealed), https://webarchive.nationalarchives.gov.uk/eu-exit/https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02000R0104-20141213 (2000).

Cozzolino, D., Power, A., & Chapman, J. (2019). Interpreting and Reporting Principal Component Analysis in Food Science Analysis and Beyond. *Food Analytical Methods, 12*(11), 2469–2473. https://doi.org/10.1007/S12161-019-01605-5/TABLES/1

Dinis, K., Tsamba, L., Thomas, F., Jamin, E., & Camel, V. (2022). Preliminary authentication of apple juices using untargeted UHPLC-HRMS analysis combined to chemometrics. *Food Control, 139*, Article 109098. https://doi.org/10.1016/J.FOODCONT.2022.109098

Fiorino, G. M., Losito, I., De Angelis, E., Arlorio, M., Logrieco, A. F., & Monaci, L. (2019). Assessing fish authenticity by direct analysis in real time-high resolution mass spectrometry and multivariate analysis: Discrimination between wild-type and farmed salmon. *Food Research International, 116*, 1258–1265. https://doi.org/10.1016/J.FOODRES.2018.10.013

Freitas, J., Silva, P., Perestrelo, R., Vaz-Pires, P., & Câmara, J. S. (2022). Improved approach based on MALDI-TOF MS for establishment of the fish mucus protein pattern for geographic discrimination of Sparus aurata. *Food Chemistry, 372*, Article 131237. https://doi.org/10.1016/J.FOODCHEM.2021.131237

González, M.Á. P., Cavazza, G., Gustinelli, A., Caffara, M., & Fioravanti, M. (2020). Absence of anisakis nematodes in smoked farmed Atlantic salmon (Salmo salar) products on sale in European countries. *Italian Journal of Food Safety, 9*(4), 216–219. https://doi.org/10.4081/IJFS.2020.8615

Grazina, L., Rodrigues, P. J., Igrejas, G., Nunes, M. A., Mafra, I., Arlorio, M., Oliveira, M. B., & Amaral, J. S. (2020). Machine Learning Approaches Applied to GC-FID Fatty Acid Profiles to Discriminate Wild from Farmed Salmon. *Foods 2020, Vol. 9, Page 1622*, *9*(11), 1622. https://doi.org/10.3390/FOODS9111622.

Henriques, J., Dick, J. R., Tocher, D. R., & Bell, J. G. (2014). Nutritional quality of salmon products available from major retailers in the UK: Content and composition of n-3 long-chain PUFA. *British Journal of Nutrition, 112*(6), 964–975. https://doi.org/10.1017/S0007114514001603

Hong, Y., Birse, N., Quinn, B., Li, Y., Jia, W., McCarron, P., … Elliott, C. T. (2023). Data fusion and multivariate analysis for food authenticity analysis. *Nature Communications, 14*(1). https://doi.org/10.1038/S41467-023-38382-Z

Jakobsen, J., Smith, C., Bysted, A., & Cashman, K. D. (2019). Vitamin D in Wild and Farmed Atlantic Salmon (Salmo Salar)—What Do We Know? *Nutrients 2019, Vol. 11, Page 982*, *11*(5), 982. https://doi.org/10.3390/NU11050982.

Kang, C., Zhang, Y., Zhang, M., Qi, J., Zhao, W., Gu, J., … Li, Y. (2022). Screening of specific quantitative peptides of beef by LC–MS/MS coupled with OPLS-DA. *Food Chemistry, 387*, Article 132932. https://doi.org/10.1016/J.FOODCHEM.2022.132932

Kappel, K., & Schröder, U. (2016). Substitution of high-priced fish with low-priced species: Adulteration of common sole in German restaurants. *Food Control, 59*, 478–486. https://doi.org/10.1016/J.FOODCONT.2015.06.024

Khalili Tilami, S., & Sampels, S. (2018). Nutritional Value of Fish: Lipids, Proteins, Vitamins, and Minerals. In *Reviews in Fisheries Science and Aquaculture* (Vol. 26, Issue 2, pp. 243–253). Taylor & Francis. https://doi.org/10.1080/23308249.2017.1399104.

Khulal, U., Zhao, J., Hu, W., & Chen, Q. (2016). Nondestructive quantifying total volatile basic nitrogen (TVB-N) content in chicken using hyperspectral imaging (HSI) technique combined with different data dimension reduction algorithms. *Food Chemistry, 197*, 1191–1199. https://doi.org/10.1016/J.FOODCHEM.2015.11.084

Kritikou, A. S., Aalizadeh, R., Damalas, D. E., Barla, I. V., Baessmann, C., & Thomaidis, N. S. (2022). MALDI-TOF-MS integrated workflow for food authenticity investigations: An untargeted protein-based approach for rapid detection of PDO feta cheese adulteration. *Food Chemistry, 370*, Article 131057. https://doi.org/10.1016/J.FOODCHEM.2021.131057

Lundebye, A. K., Lock, E. J., Rasinger, J. D., Nøstbakken, O. J., Hannisdal, R., Karlsbakk, E., … Ørnsrud, R. (2017). Lower levels of Persistent Organic Pollutants, metals and the marine omega 3-fatty acid DHA in farmed compared to wild Atlantic salmon (Salmo salar). *Environmental Research, 155*, 49–59. https://doi.org/10.1016/J.ENVRES.2017.01.026

Mowi. (2021). *Salmon Farming Industry Handbook 2021*.

Murphy, F., Gathercole, J., Lee, E., Homewood, I., Ross, A. B., Clerens, S., & Maes, E. (2021). Discrimination of milk fermented with different starter cultures by MALDI-TOF MS and REIMS fingerprinting. *International Dairy Journal, 122*, Article 105143. https://doi.org/10.1016/J.IDAIRYJ.2021.105143

Nøstbakken, O. J., Hove, H. T., Duinker, A., Lundebye, A. K., Berntssen, M. H. G., Hannisdal, R., … Julshamn, K. (2015). Contaminant levels in Norwegian farmed Atlantic salmon (Salmo salar) in the 13-year period from 1999 to 2011. *Environment International, 74*, 274–280. https://doi.org/10.1016/J.ENVINT.2014.10.008

Piras, C., Hale, O. J., Reynolds, C. K., Jones, A. K., Taylor, N., Morris, M., & Cramer, R. (2021). Speciation and milk adulteration analysis by rapid ambient liquid MALDI mass spectrometry profiling using machine learning. *Scientific Reports, 11*(1), 3305. https://doi.org/10.1038/s41598-021-82846-5

Quinn, B., McCarron, P., Hong, Y., Birse, N., Wu, D., Elliott, C. T., & Ch, R. (2022). Elementomics combined with dd-SIMCA and K-NN to identify the geographical origin of rice samples from China, India, and Vietnam. *Food Chemistry, 386*, Article 132738. https://doi.org/10.1016/J.FOODCHEM.2022.132738

Rickertsen, K., Alfnes, F., Combris, P., Enderli, G., Issanchou, S., & Shogren, J. F. (2017). French Consumers' Attitudes and Preferences Toward Wild and Farmed Fish. *Marine Resource Economics, 32*(1), 59–81. https://doi.org/10.1086/689202/ASSET/IMAGES/LARGE/FG2.JPEG

Sampaio, P. S., Castanho, A., Almeida, A. S., Oliveira, J., & Brites, C. (2020). Identification of rice flour types with near-infrared spectroscopy associated with PLS-DA and SVM methods. *European Food Research and Technology, 246*(3), 527–537. https://doi.org/10.1007/S00217-019-03419-5/FIGURES/6

Thomas, F., Jamin, E., Wietzerbin, K., Guérin, R., Lees, M., Morvan, E., … Robin, R. J. (2008). Determination of origin of atlantic salmon (Salmo salar): The use of multiprobe and multielement isotopic analyses in combination with fatty acid composition to assess wild or farmed origin. *Journal of Agricultural and Food Chemistry, 56*(3), 989–997. https://doi.org/10.1021/JF072370D/ASSET/IMAGES/LARGE/JF-2007-02370D_0011.JPEG

Thorstad, E. B., Bliss, D., Breau, C., Damon-Randall, K., Sundt-Hansen, L. E., Hatfield, E. M. C., … Sutton, S. G. (2021). Atlantic salmon in a rapidly changing environment—Facing the challenges of reduced marine survival and climate change. *Aquatic Conservation: Marine and Freshwater Ecosystems, 31*(9), 2654–2665. https://doi.org/10.1002/AQC.3624

Trygg, J., & Wold, S. (2002). Orthogonal projections to latent structures (O-PLS). *Journal of Chemometrics, 16*(3), 119–128. https://doi.org/10.1002/CEM.695

Verplanken, K., Stead, S., Jandova, R., Poucke, C. V., Claereboudt, J., Bussche, J. V., … Vanhaecke, L. (2017). Rapid evaporative ionization mass spectrometry for high-throughput screening in food analysis: The case of boar taint. *Talanta, 169*, 30–36. https://doi.org/10.1016/j.talanta.2017.03.056

Wang, C., Bi, H., & Xie, J. (2020). Visualization of the Distance among Fishes by MALDI MS for Rapid Determination of the Taxonomic Status of Fish Fillets. *Journal of Agricultural and Food Chemistry, 68*(31), 8438–8446. https://doi.org/10.1021/ACS.JAFC.0C01291/SUPPL_FILE/JF0C01291_SI_001.PDF

Wang, P., & Giese, R. W. (2017). Recommendations for quantitative analysis of small molecules by matrix-assisted laser desorption ionization mass spectrometry. *Journal of Chromatography A, 1486*, 35–41. https://doi.org/10.1016/J.CHROMA.2017.01.040

Zambonin, C., Varzakas, T., & Panagiotis, T. (2021). MALDI-TOF Mass Spectrometry Applications for Food Fraud Detection. *Applied Sciences 2021, Vol. 11, Page 3374*, *11*(8), 3374. https://doi.org/10.3390/APP11083374.

Zhong, H., Abdullah, Zhang, Y., Deng, L., Zhao, M., Tang, J., Zhang, H., Feng, F., & Wang, J. (2021). Exploring the potential of novel xanthine oxidase inhibitory peptide (ACECD) derived from Skipjack tuna hydrolysates using affinity-ultrafiltration coupled with HPLC–MALDI-TOF/TOF-MS. *Food Chemistry, 347*, Article 129068. https://doi.org/10.1016/J.FOODCHEM.2021.129068

Ziegler, F., & Hilborn, R. (2023). Fished or farmed: Life cycle impacts of salmon consumer decisions and opportunities for reducing impacts. *Science of The Total Environment, 854*, Article 158591. https://doi.org/10.1016/J.SCITOTENV.2022.158591