European Food Research and Technology (2025) 251:2903–2913
https://doi.org/10.1007/s00217-025-04814-x

ORIGINAL PAPER

Unveiling the putative porcine myosin-based peptide markers for non-halal meat through chemometrics-assisted MRM-based proteomics

Mohd Hafis Yuswan[1] · Norazlina Ali[2] · Syaiful Izwan Ismail[2] · Muhamad Shirwan Abdullah Sani[3] · Lai Kok Song[4]

Received: 18 February 2025 / Revised: 14 May 2025 / Accepted: 3 June 2025 / Published online: 16 June 2025

Abstract
Handling massive proteomics datasets poses challenges due to assessing dataset quality and dealing with multiple dimensions of the dataset when establishing putative peptide markers. Therefore, this study aims to confirm putative porcine peptide markers for precision halal proteomics through chemometrics-assisted MRM-based proteomics. Chemometric data mining was employed to access the dispersion characteristics and normality of 509 commercial processed meat samples (beef, chicken, fish, and pork). All the samples displayed normal distributions and showed significant differences in the median. By employing chemometric principal component analysis, two significant dimensions were identified to select the putative porcine peptide markers. Out of 1204 identified peptides, two putative porcine peptide markers were critically selected: P25 and P68, derived from myosin-1. MRM acquisition was developed to verify the P25 and P68 for precision halal proteomics. Notably, only the MRM chromatogram of P68 showed a modified peptide peak. Nonetheless, the process of confirming putative porcine peptide markers from massive proteomics datasets is robust and reliable through chemometrics-assisted MRM-based proteomics for halal authentication in the context of meat speciation. It is recommended utilizing P25 as the peptide marker due to its purity and unmatched sequence with bovine, chicken, and fish based on the UniProtKB search.

Keywords Chemometrics · MRM · Proteomics · Peptide · Halal · Meat authentication

Introduction

Meatballs, meatloaf, minced meat, frikadelle, frankfurter, salami, sausage, corned meat, canned meat, half-cooked meat, smoked meat, floss meat, and various other commercial meat products have become an integral part of our local markets. These products undergo intensive processing methods, including thermal and non-thermal processing, along with the addition of food additives [1]. However, ensuring the absence of pork in these products is crucial for the halal industry. The halal industry holds significant potential, sharing commonalities with kosher practices [2–6]. Recently, even non-Muslim consumers prefer halal food due to its perceived product integrity, brand trustworthiness, and low risk of foodborne zoonotic diseases [7].

The realm of halal meat speciation faces a major challenge with conventional DNA-based methods for pork identification in commercial processed meat products. Harsh industrial processing negatively affects DNA stability, rendering DNA-based methods less reliable for detecting pork [8]. This study explores an alternative protein-based approach, focusing on peptide identification through liquid chromatography-tandem mass spectrometry (LC-MS/MS) [9–13].

The LC-MS/MS is a powerful analytical tool, allowing precise identification and characterization of proteins and peptides in complex samples. However, most of the peptide

✉ Mohd Hafis Yuswan
hafisyuswan@upm.edu.my

1 Halal Products Research Institute, Universiti Putra Malaysia, Serdang 43400 UPM, Selangor, Malaysia

2 Malaysia Halal Analysis Centre (MyHAC), Department of Islamic Development Malaysia, No. 1 Persiaran Teknologi 1, Lebuh Enstek, Bandar Baru Enstek, Nilai 71760, Negeri Sembilan, Malaysia

3 International Institute for Halal Research and Training, International Islamic University Malaysia, Jalan Gombak, Kuala Lumpur 53100, Selangor, Malaysia

4 Health Sciences Division, Abu Dhabi Women's College, Higher Colleges of Technology, Abu Dhabi 41012, United Arab Emirates

markers were only established from a single database search engine of the LC-MS/MS technique [14], which might be prone to false positives. In our previous study [15], we analysed 135 raw LC-MS/MS data files from 15 types of commercial processed pork-based food products and discovered the prevalence of myofibrillar proteins, particularly myosin isoforms identified from four comparable database search engines. In contrast to previous studies that primarily focussed on raw or in-house cooked samples [16–21], this study analyses massive proteomic datasets from commercial processed meat products including beef, chicken, fish, and pork. However, analysing such datasets requires advanced chemometrics for data mining, extracting chemically relevant information [22].

Therefore, the objective of this study is to establish and verify putative porcine peptide markers for precision halal proteomics in the context of meat speciation. To achieve this, this study undertakes a meticulous analysis of additional samples, including beef ($n=14$), chicken ($n=14$), and fish ($n=14$). To handle the massive proteomics datasets, this study explains the chemometrics of the data mining process, using boxplots and Q–Q plots to ensure data quality and reliability. Subsequently, this study employs principal components analysis to select robust peptide markers. The resulting data is processed through multiple reaction monitoring (MRM) of LC-MS/MS acquisition to verify the presence of peptide markers for precision halal proteomics.

These findings have significant implications for the halal industry, providing insight into handling massive proteomics datasets and selecting precise peptide markers for precision halal proteomics. The innovative application of chemometric-assisted MRM-based proteomics represents a valuable resource for researchers exploring halal proteomics for meat speciation. Ultimately, this study enhances halal food authentication methodologies, strengthening consumer confidence, and meeting the stringent demands of the halal market.

Materials and methods

Materials

LC-MS grade formic acid and acetonitrile were purchased from Fisher Chemical, Fair Lawn, NJ, USA. Ultra-pure water with a conductivity of ≤18 MΩ cm was freshly prepared using an ultrapure water system (Arium®611VF, Sartorius Stedim, Goettingen, Germany). Mass spectrometry grade Trypsin Gold was obtained from Promega, Madison, WI, USA. The samples consist of commercial processed meat products including pork ($n=15$), beef ($n=14$), chicken ($n=14$), and fish ($n=14$) randomly purchased on the same day from local markets in the Serdang city area, Selangor, Malaysia. Details of samples are provided in Supplementary 1. All samples were stored at −20 °C until further analysis.

Sample preparation and digestion

All samples were cleaned by washing with ethanol/water (70:30% v/v) three times to remove potential salt and other soluble low molecular weight compounds, followed by three times rinsing with ultra-pure water. Subsequently, approximately 20–30 g of samples were transferred into 50-mL Falcon tubes for lyophilisation (1 L benchtop freeze dry system, FreeZone LabConco, Kansas City, MO, USA). Next, each sample was individually ground using a ProMix Handblender (HR2533/01, Philips, China) to produce a fine dry powder. The fine dry powder of each sample was then pooled accordingly to its sample type (pork, beef, chicken, and fish) for randomization. Adulterated samples were prepared according to previous study [23], by mixing the pooled pork sample (0.5%, 1%, and 5%, w/w) with the pooled beef, chicken, and fish samples, respectively. Next, 10 mg of each sample was rehydrated in 100 µL of ultra-pure water. Each sample was reduced and alkylated by adding 0.5 mM of dithiothreitol (DTT) and 0.5 mM of iodoacetamide (IAA) in a 1:20 ratio (v/v) at 37 °C for 6 h. Subsequently, each sample was digested by adding 1 g/L Trypsin Gold in 50 mM acetic acid in a 1:25 ratio (v/v), in accordance with the manufacturer's instructions as outlined in the certificate of analysis (Promega Trypsin Gold, Mass Spectrometry Grade, Part No. V528A, Revised 10/7). The digestion was carried out at 37 °C for a duration of 4 h. Tryptic peptides were collected by transferring the supernatant into a new, clean 1.5 mL Eppendorf tube after centrifugation for 15 min at 12,000 rcf and 5 °C. All samples were prepared in triplicate.

LC-MS/MS and scheduled MRM acquisition

Chromatographic separation and tandem mass spectrometry detection were conducted using high-performance liquid chromatography (Agilent 1200 Series), which included an autosampler (G1367D), a binary pump (G1312B), and a column oven (G1316A). This setup was coupled with the AB SCIEX 4000 QTrap mass spectrometer (Singapore). The initial concentration of mobile phase A began at 97% and gradually decreased to 71.6% over 22 min. Subsequently, the concentration of mobile phase A was set to 0% and maintained until 28 min. Finally, at 29 min, the concentration of mobile phase A was restored to its initial level of 97% and held steady until 35 min. The Turbo Spray ionization source for electrospray ionization was set at 400 °C, with a nebulizer voltage of 5200 V at 40 psi. The pressures for source gases 1 and 2 were set at 40 psi and 30 psi, respectively,

while the curtain gas pressure was maintained at 30 psi. The declustering potential was set to 50 V, and the scan rate was 1000 Da/s. Mass spectrometry (MS) and targeted MS/MS analyses were performed using enhanced mass spectrometry (EMS) with a scan range of 500–2800 m/z and enhanced product ion analysis with a scan range of 100–1800 m/z, respectively, in positive mode. Instrument control and data acquisition were conducted using Analyst® 1.5 Software (AB Sciex). For scheduled MRM scan mode, MRM parameters were established and optimized using Skyline 22.2.0.312 software (64-bit, MacCoss Lab, Department of Genome Sciences, University of Washington). Optimization of collision energy (CE) and declustering potential (DP) was performed through stepwise adjustment to maximize signal intensity for each transition. Transitions were selected based on previous spectral library [15] and manual evaluation to ensure both sensitivity and specificity. Key parameters, including precursor and product ion m/z values (transitions), optimized CE and DP, retention times, and dwell time, are reported in Table 1. Peptide markers were detected by monitoring the five most intense transitions per peptide to enhance reliability. Each sample was injected in triplicate, with an injection volume of 10 µL per run.

Database searching

Database searching was conducted according to a previous study [15] with some modifications. All LC-MS/MS chromatographic raw data were converted into centroided files (mzML) using the ProteoWizard MSConvertGUI (64-bit, Version: 3.0.1908-43e675997) with the peak picking filter type and vendor algorithm (MS level = 1–2) parameters. The centroided files were then subjected to Comet sequential database search engine (Version 2018.01 rev.4), through the Petunia Trans-Proteomic Pipeline (TPP, version 5.2.0). An in-house target-decoy database was constructed from reference proteomes of pig (*Sus scrofa*), bovine (*Bos taurus*), chicken (*Gallus gallus*), and Zebrafish (*Danio rerio*), downloaded from UniProtKB with proteome identifiers UP000314985, UP000009136, UP000000539, and UP000000437, respectively. For the Comet parameters, the precursor mass tolerance was 0.5 Da; the fragment mass tolerance was 0.7 Da; trypsin digestion was considered as semi-digestion; the maximum missed cleavage was set to one, and the variable modifications included carbamidomethylation of cysteine, oxidation of methionine, and reduction of asparagine and glutamines residues.

Data mining and chemometrics

The results of database searching were utilized for data mining and subsequent chemometric analysis, specifically by constructing a principal component analysis (PCA), as described previously with some modifications [23–25]. A data matrix X was generated for PCA with Pareto scaling (which reduces the influence of large variance features while preserving data structure), where the peak list of peptide-spectrum matches served as a variable (K), and each commercial processed pooled sample, including replicates, was treated as an observation (N) to identify the potential peptide markers. For data mining and chemometrics, the RStudio Team (Version 1.4.1717): Integrated Development Environment for R, 2015, was employed.

Results and discussions

Data analytics for massive proteomic dataset

The high-throughput nature of LC-MS/MS, particularly when applied to replicated samples, generates a substantial amount of complex data, including thousands of peptide-spectrum matches (PSMs) per sample. In this study, LC-MS/MS acquisition was performed on a total of 509

Table 1 MRM transition parameters for the detection of marker peptides

Variable number and peptide sequence	Protein (UniProtKB ID)	Precursor ion (m/z)	Product ion (m/z)	Retention time (min)	Declustering potential (V)	Collision energy (V)
P25 (DLEEATLQHEATAATLR)	Myosin-1 (Q9TV61)	934.971^{++}	1311.7015^{+} 1097.5698^{+} 969.5112^{+} 832.4523^{+} 703.4097^{+}	16.2	110	54
P68 (QLDEKDTLVSQLSR)	Myosin-1 (Q9TV61)	816.434^{++}	1018.5527^{+} 903.5258^{+} 802.4781^{+} 689.3941^{+} 590.3257^{+}	15.0	90	48

m/z = mass to charge ratio

processed samples, representing replicated analyses across four types of meat and yield a total of 36,504 PSM. For pork, a total of 135 samples were obtained from 15 types of samples, each analysed in triplicate for both biological and technical replicates ($n = 15 \times 3 \times 3$). Beef yielded 127 total samples from 14 types of samples, including one additional run performed unintentionally on a single sample ($n = 14 \times 3 \times 3 + 1$). Chicken yielded 123 total samples due to three failed runs ($n = 14 \times 3 \times 3 - 3$), while fish resulted in 124 total samples due to two failed runs ($n = 14 \times 3 \times 3 - 2$). Biological replicates refer to independently processed samples, capturing variability among different sources, while technical replicates represent repeated LC-MS/MS injections of each biological replicate to ensure analytical precision and reproducibility. To ensure robust interpretation of this data, statistical treatments were employed to assess the quality of the dataset, such as evaluating data distribution and normality across replicated samples [26]. These analyses provided critical insights into the reliability and reproducibility of the acquired spectra and enabled the identification of consistent and meaningful peptide signals. Rather than relying solely on a simple selection of peptide markers, this data-driven approach ensured that selected pork-specific markers were supported by statistical evidence, strengthening the validity and applicability of the method for routine halal authentication and food integrity analysis.

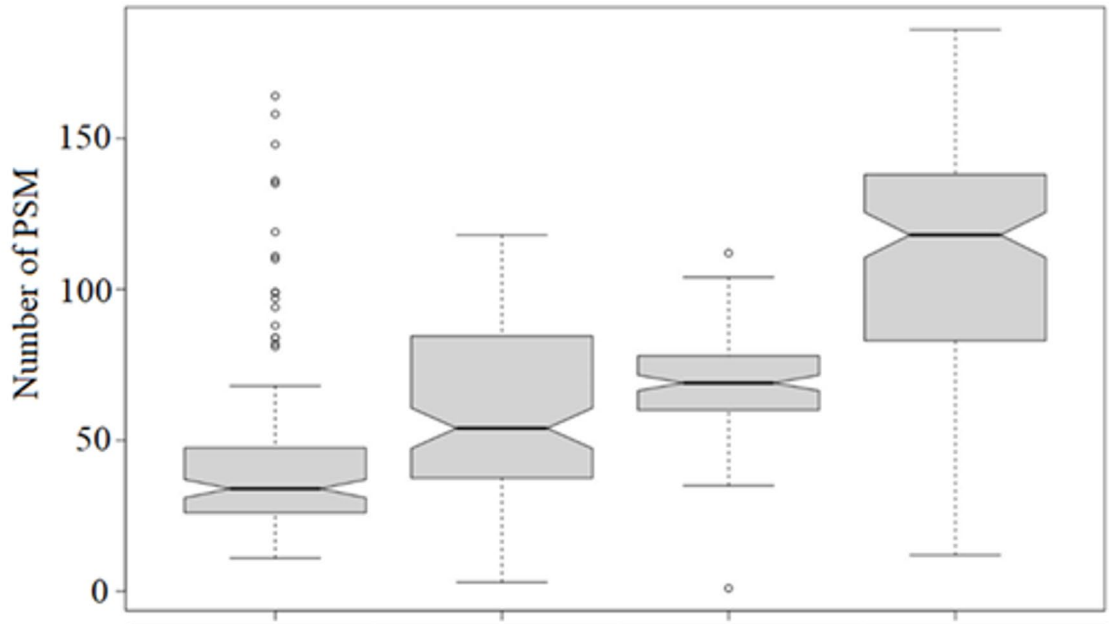
Subsequently, the dataset was subjected to data mining using a boxplot to examine its dispersion characteristics, such as range, quartiles, interquartile range, variance, and outliers. Figure 1 presents the boxplot comparison of the commercial processed samples based on the number of PSMs. Data points that differ by 1.5 times the interquartile range from the first or third quartiles (Q_1 and Q_3) are considered outliers [27]. In comparison to the fish samples, which had a single data point that was varied both $Q_1 = 60$ and $Q_3 = 78$, the commercial processed samples of beef had 18 data points that were above the $Q_3 = 48$ (Fig. 1). The presence of a high number of outliers in the commercial processed samples of beef may indicate false-positive data points resulting from the Comet search engine during our previous target-decoy strategy [15]. The non-normal distribution of the population may also be attributed to the complexity of the sample matrix after pooling [28]. A computer simulation showed that regardless of the sample size, approximately 30% of samples drawn from a normally distributed population will contain one or more outliers [28].

Additionally, the shape of the boxplot provided information about the spread of the massive proteomic data points for each commercial processed sample. The variance among the commercial processed samples was found to be distinct (p-value $= 2.2 \times 10^{-16}$) according to Bartlett's Test for homogeneity of variance [29]. Instead of conducting a one-way analysis of variance, the notches of each boxplot were used to determine significant differences in the median (Q_2) [27]. In this study, the medians for the commercial processed samples of beef ($Q_2 = 34$), chicken ($Q_2 = 54$), fish ($Q_2 = 69$), and pork ($Q_2 = 118$) were significantly different at a 95% confidence level, as none of the notches overlapped (Fig. 1). The median represents the value where 50% of the massive proteomic data points fall. Furthermore, all the massive proteomic data points were within the upper and lower whisker ends, except for the commercial processed samples of beef and fish (the outliers). This indicates that each boxplot covers most of the massive proteomic data points and suggests that the boxplots exhibit a normal distribution (non-skewed shape) [27].

To confirm the normality of the massive proteomic data points, a normality test was conducted using a Q–Q plot to analyse the distribution [30]. Figure 2 displays the Q–Q plot for each commercial processed sample along with its regression line. Adequate representation of the data points by the regression line increases the likelihood of the data being normally distributed [31]. The regression (r^2) values for the commercial processed samples of pork, beef, chicken, and fish were 0.988, 0.862, 0.979, and 0.983, respectively. Overall, the massive proteomic data points appeared to be normally distributed. Some Q–Q plots also revealed heavy- or light-tailed data points. The low regression value ($r^2 = 0.862$) and the presence of heavy-tailed data points in the commercial processed samples of beef can be attributed to the 18 outliers, as illustrated in Fig. 2 [31]. This result aligns with the findings from the boxplot analysis in Fig. 1. The Q–Q plots for the other samples indicated light-tailed behaviour, suggesting the absence of extreme data points [31].

Selection of putative porcine myosin isoforms-based peptide markers

The massive proteomic dataset consisted of sample description, PSM, mass-to-charge ratio (m/z) value, retention time, protein accession number, and protein mass. This massive dataset was subsequently subjected to Principal Component Analysis (PCA) as the initial clustering method. Generally, the data matrix for PCA consisted of 36 observations (4 types of grouping commercial processed pooled samples × 3 biological replicates × 3 technical replicates) and 1204 variables (identified peptides). The establishment of the 1204 identified peptides was explained clearly in our previous study [15]. The PCA was employed to assess the systematic variation present in the large data matrix containing a high number of dimensions. Figure 3a shows a scree plot for the first to tenth dimensions of PCA with a percentage of explained variance for each dimension. For 100% of explained variance, the PCA model required 35



	Beef	Chicken	Fish	Pork
Upper whisker	68	118	104	186
3rd quartile (Q_3)	48	85	78	138
Median (Q_2)	34	54	69	118
Notch*	31 – 37	47 – 61	67 – 72	111 – 126
1st quartile (Q_1)	26	38	60	83
Lower whisker	11	3	35	12
Numbers of outliers	18	0	2	0
Number of samples (n)	127	123	124	135
Total number of PSM	5547	7112	8578	15267

Fig. 1 Boxplot comparison of samples against the number of peptide-spectrum match (PSM). The bold line inside the boxplot is a median (Q_2), *notch represents a 95% confidence interval of Q_2, and dots above the upper whisker or below the lower whisker indicate outliers

dimensions. By applying dimension reduction, PCA effectively decreased the dimensionality of the highly correlated dataset while retaining crucial information [32]. However, in this study, only the first (13.2%) and second (8.7%) dimensions were retained, explaining a total variance of 21.9%. In a previous study that examined potential porcine peptide markers in 45 raw meat samples (beef, chicken, and pork), the PCA model accounted for 59.6% of the explained variance [23]. The difference in explained variance percentages between the two studies can be attributed to the complexity of the commercial processed pooled samples' matrix and the addition of commercial processed pooled fish samples in the PCA data matrix.

Since there were only 2 out of 35 significant dimensions in the PCA model, only variables (identified peptides) that surpassed the contribution's reference line were considered crucial for those dimensions. Figure 3b presents a bar chart illustrating the contribution of the variables to the first and second dimensions of PCA, with a threshold for the contribution's reference line. Out of the 1204 identified peptides, only 283 were deemed crucial for the first and second dimensions of the PCA model after sorting the contributions in descending order (Fig. 3b). However, to select highly

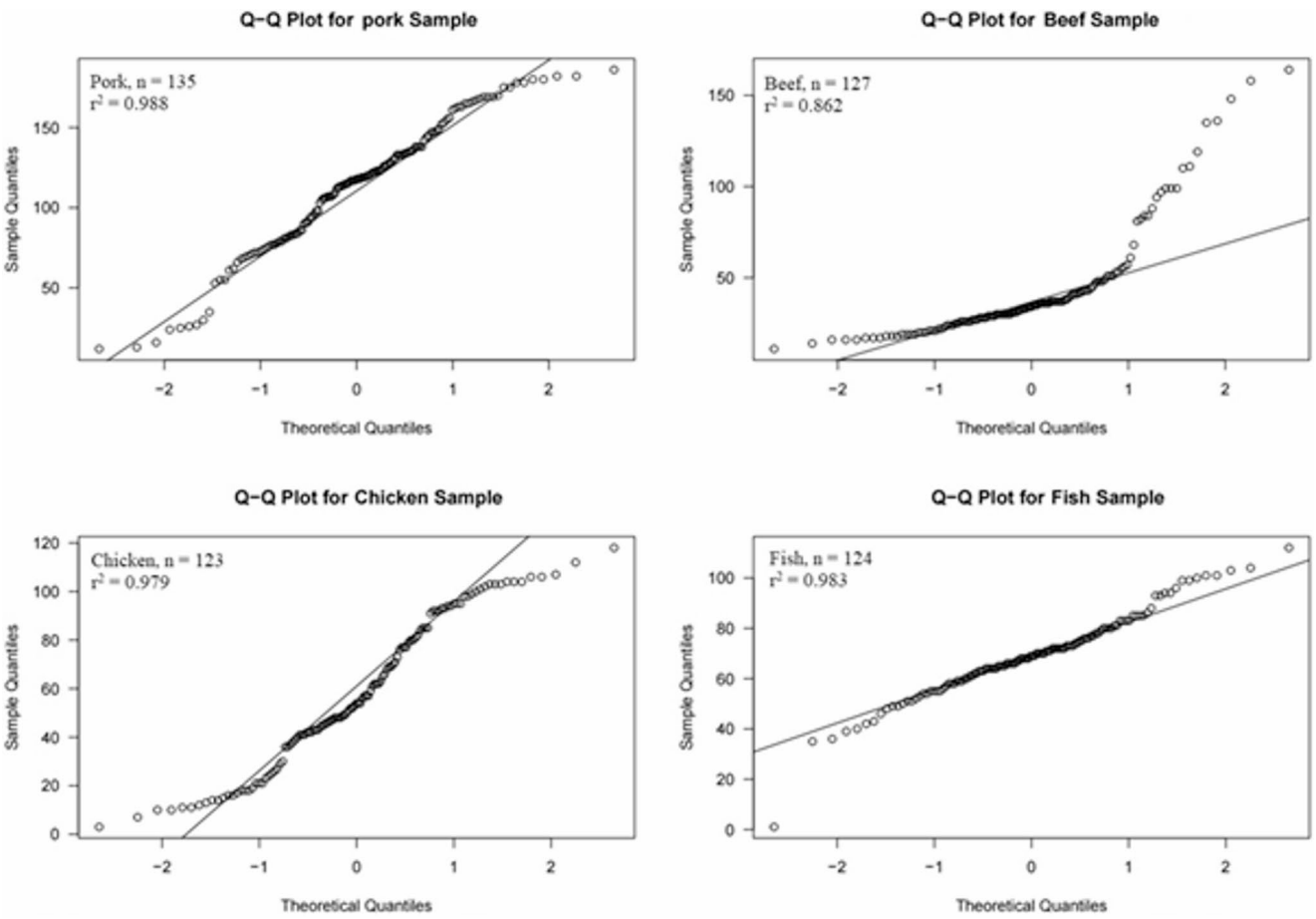


Fig. 2 Q–Q plot for commercial processed samples of pork, beef, chicken, and fish. Dots in each Q–Q plot represent data points. The straight line in each Q–Q plot indicates a regression line

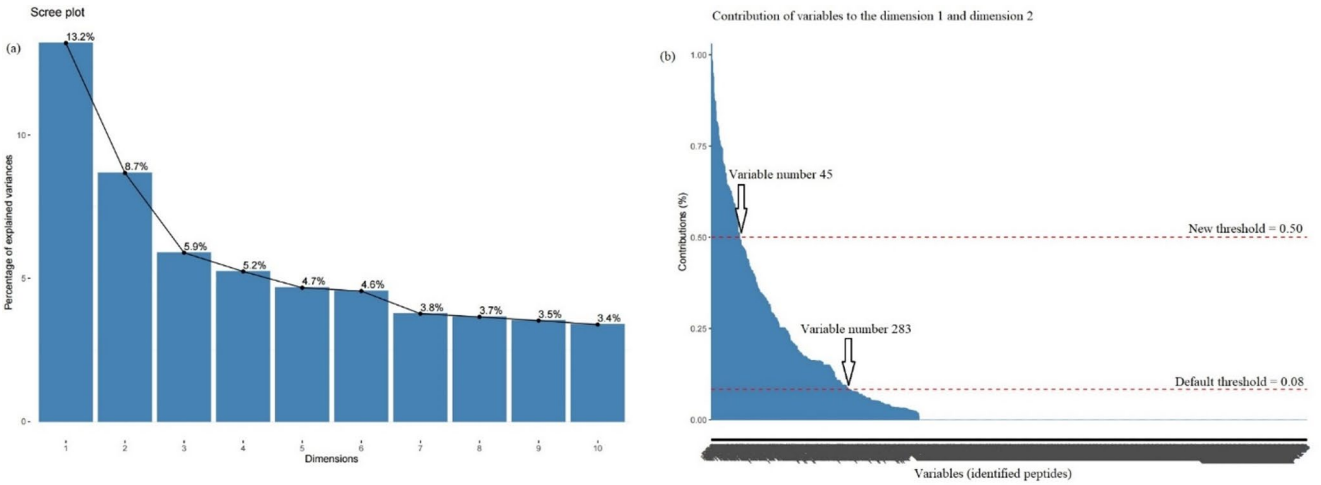


Fig. 3 Scree plot for the first to tenth dimensions of Principal Component Analysis (PCA) (**a**). The percentage value explains the variance for each dimension. The total dimension for the PCA model is 35 for 100% of explained variance. Bar chart for the contribution of variables (identified peptide) to the first and second dimensions of Principal Component Analysis (PCA) (**b**). A dashed line is a threshold of the reference line to determine the important variables contributing to the dimension. The total variables are 1204 in descending order of contribution

significant contributions of potential peptide markers, a reference line threshold of 50% was used instead of the default 8%. Consequently, only 45 identified peptides met the reference line threshold. Unfortunately, these 45 identified peptides were not specific to the commercial processed sample of pooled pork but were present across all commercial processed pooled samples.

To facilitate the selection of potential peptide markers for non-halal pork, the score and loading plots of PCA were utilized for the 45 identified peptides. Figure 4a displays the score plot, which describes the clustering of commercial processed pooled samples based on their similarity. Each commercial processed pooled sample is represented by an ellipse, encompassing 95% of the data points associated with that sample. The ellipse helps visualize the distribution of the commercial processed pooled samples and reveals their distinctive characteristics, correlations, and differences [33]. In this study, all the commercial processed pooled samples showed significant differences and possessed unique characteristics based on the ellipse regions. While all the commercial processed pooled samples exhibited a negative correlation relative to the first dimension of PCA, the commercial processed pooled samples of beef and fish displayed a positive correlation.

Furthermore, Fig. 4b illustrates the loading plot, indicating the contribution of each identified peptide to the first and second dimensions of PCA. The dark-coloured identified peptides had higher contributions compared to the light-coloured ones. Since the score and loading plots are correlated, the selection of porcine peptide markers for precision halal proteomics is feasible. Specifically, the selection of porcine peptide markers focused on the dark data points in the upper left quadrant of the loading plot. Therefore, the specific porcine peptide markers identified were variable numbers P25 and P68 with contributions of 88 and 63%, respectively.

Characteristics of putative porcine peptide markers

Considering the variable numbers P25 and P68 as putative porcine peptide markers, an optimized multiple reaction monitoring (MRM) of LC-MS/MS acquisition was developed before these peptide markers can be used for halal authentication in terms of meat speciation. Table 1 presents the characteristics and MRM transition parameters used to detect these peptide markers for halal authentication. Our Comet database search, based on a previous study [15], confirmed that all the putative porcine peptide markers belong to myosin (Table 1). Peptides P25 and P68 have the sequences DLEEATLQHEATAATLR and QLDEKDTLVSQLSR, respectively, and they originate from myosin-1 proteins. These peptides were frequently detected in a porcine study by LC-MS [9–13]. The peptides derived from myofibrillar protein, specifically myosin, account for nearly 55–60% of the total protein content in meat [34]. However, another study that employed MRM to authenticate meat species and cuts in single-cut meat products did not report the presence of these two putative porcine peptide markers [35], suggesting that these peptides may be alternative for the authentication. During the MRM of LC-MS/MS acquisition, P68 showed inadequate retention on the stationary phase, taking approximately 15 min, whereas P25's retention was 16.2 min.

For identification purposes, each peptide marker has a unique precursor ion and specific product ions. In this study, five product ions were established for each precursor ion for precision halal proteomics (Table 1). For precision halal proteomics using P25 as a peptide marker, this study applied a declustering potential of 110 volts and collision energy of 54 volts. This allows the detection of P25's precursor ion at 934.971 *m/z*, with products ions representing the amino acids TLQHEATAATLR (1311.7015 *m/z*), QHEATAATLR (1097.5698 *m/z*), HEATAATLR (969.5112 *m/z*), EATAATLR (832.4523 *m/z*), and ATAATLR (703.4097 *m/z*), respectively, as identified through the Comet database search. Likewise, for P68, this study used the declustering potential of 90 volts and the collision energy of 48 volts to detect its precursor ion at 816.434 *m/z*. The product ions corresponding to the amino acids DTLVSQLSR (1018.5527 *m/z*), TLVSQLSR (903.5258 *m/z*), LVSQLSR (802.4781 *m/z*), VSQLSR (689.3944 *m/z*), and SQLSR (590.3257 *m/z*) were identified through the Comet database search. These MRM parameters for precision halal proteomics align with The European Commission, 2017 [36], which requires a minimum of two m/z ions for MRM acquisition.

While the developed MRM of LC-MS/MS acquisition is reliable for precision halal proteomics, a modified peptide peak was observed in the MRM chromatogram for P68. Figure 5 displays the MRM of LC-MS/MS acquisition chromatograms of P25 and P68. The modified peptide in P68 may be attributed to three potential modifications in its peptide sequence (QLDEKDTLVSQLSR): acetylation of threonine (at T7), deamination of glutamine (at Q1), and pyroglutamate of glutamine (at any position). These modified peptides were supported by a previous study that characterizes the Maillard reaction in meat under different cooking treatments, revealing advanced glycation end products, protein crosslinks, or post-translational modifications [13]. In that study, the same peptide sequence of P68 was reported from the myosin of pork with the three modifications, while no peptide sequence of P25 was reported. The MRM chromatogram of P25 displayed no modified peptide.

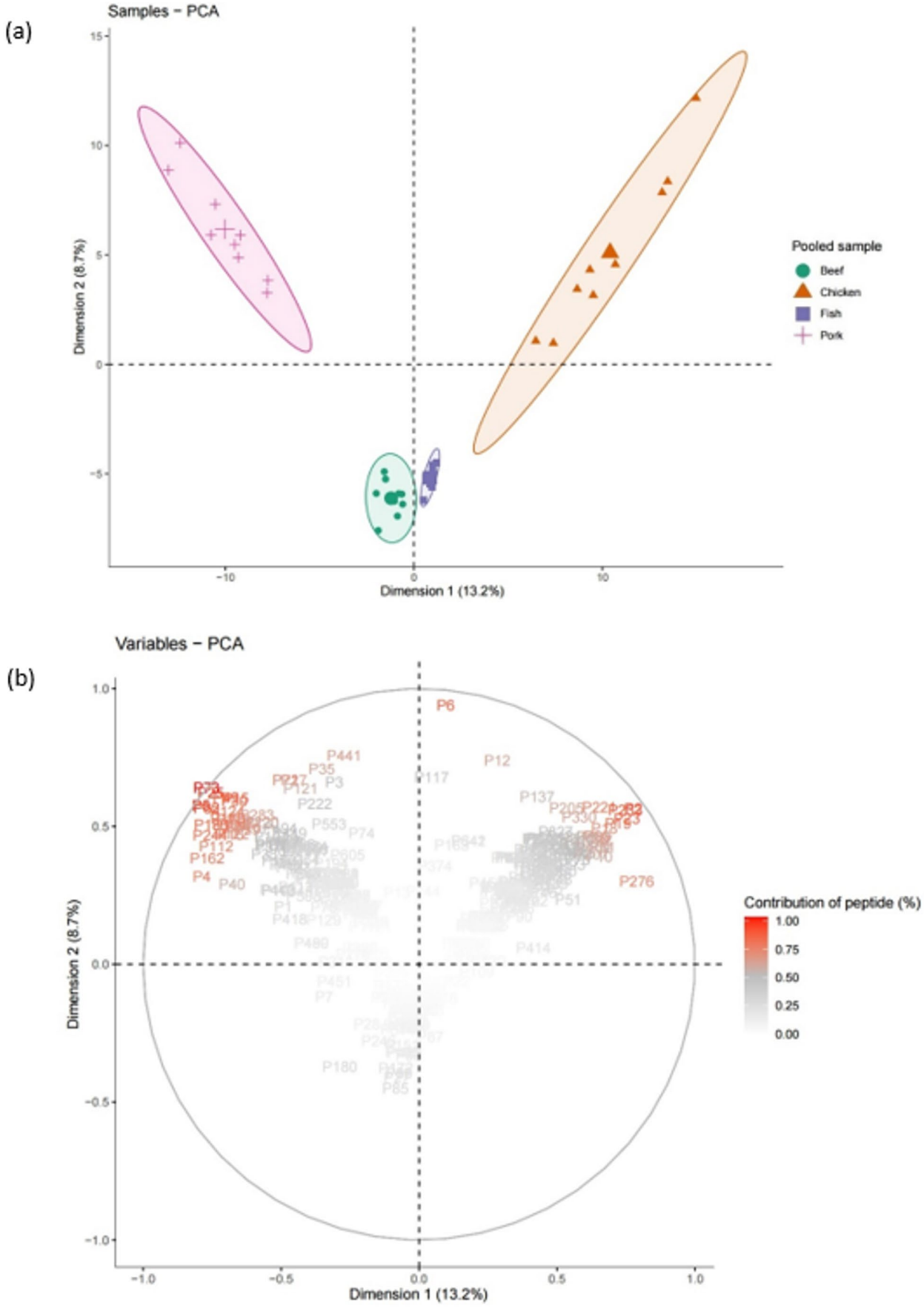

(a)
Samples – PCA
Dimension 2 (8.7%)
Dimension 1 (13.2%)
Pooled sample
Beef
Chicken
Fish
Pork
(b)
Variables – PCA
Dimension 2 (8.7%)
Dimension 1 (13.2%)
Contribution of peptide (%)
1.00
0.75
0.50
0.25
0.00
P6
P12
P276
P162
P4
P40
P112
P222
P441
P35
P121
P3
P137
P553
P74
P51
P414
P480
P451
P7
P180
P85

Fig. 4 Principal component analysis (PCA) for commercial processed samples of pooled beef, chicken, fish, and pork. The upper side is a score plot (**a**), and the lower side is a loading plot (**b**)

Conclusion

This study contributes to the emerging field of halal proteomics, which focuses on the targeted application of proteomic tools for ensuring halal authentication and integrity. This comprehensive analysis of chemometrics-assisted MRM-based proteomics allowed for the successful establishment and confirmation of putative porcine peptide markers from 509 commercial processed samples of beef, chicken, fish, and pork, enabling precision halal proteomics in the context of meat speciation. Chemometric data mining techniques, such as boxplots and Q–Q plots, were considered crucial in the initial stage for researchers to investigate the quality of massive proteomic datasets in terms of dispersion characteristics and normality. Subsequently, the chemometric principal component analysis managed to minimize the multiple dimensions of massive proteomics datasets, allowing the selection of putative porcine peptide markers across 36 groups of commercial processed pooled samples and 1204 identified peptides. Only two specific porcine peptide markers, P25 and P68 originating from myosin-1 proteins were critically selected for the precision halal proteomics. An optimized MRM of LC-MS/MS acquisition confirmed the reliability of P25 and P68 peptide markers. However, a modified peptide peak was observed in the MRM chromatogram of P68, possibly due to post-translational modifications. Instead, this study strongly recommends using the P25 peptide marker, as it did not show modified peptide and did not match sequences from bovine, fish, and chicken during the UniProtKB search. This study provides a valuable insight into analysing massive proteomic datasets and their application in ensuring the integrity and authenticity of halal meat products in the context of meat speciation. Further investigations are required to uncover the factors causing the modified peptide of P68 and validate P25 to fully harness the potential of these putative porcine peptide markers for practical applications in the food industry. Although this study focused on commercial processed meat products, future work should expand validation to a broader range of meat-based matrices to enhance general applicability.

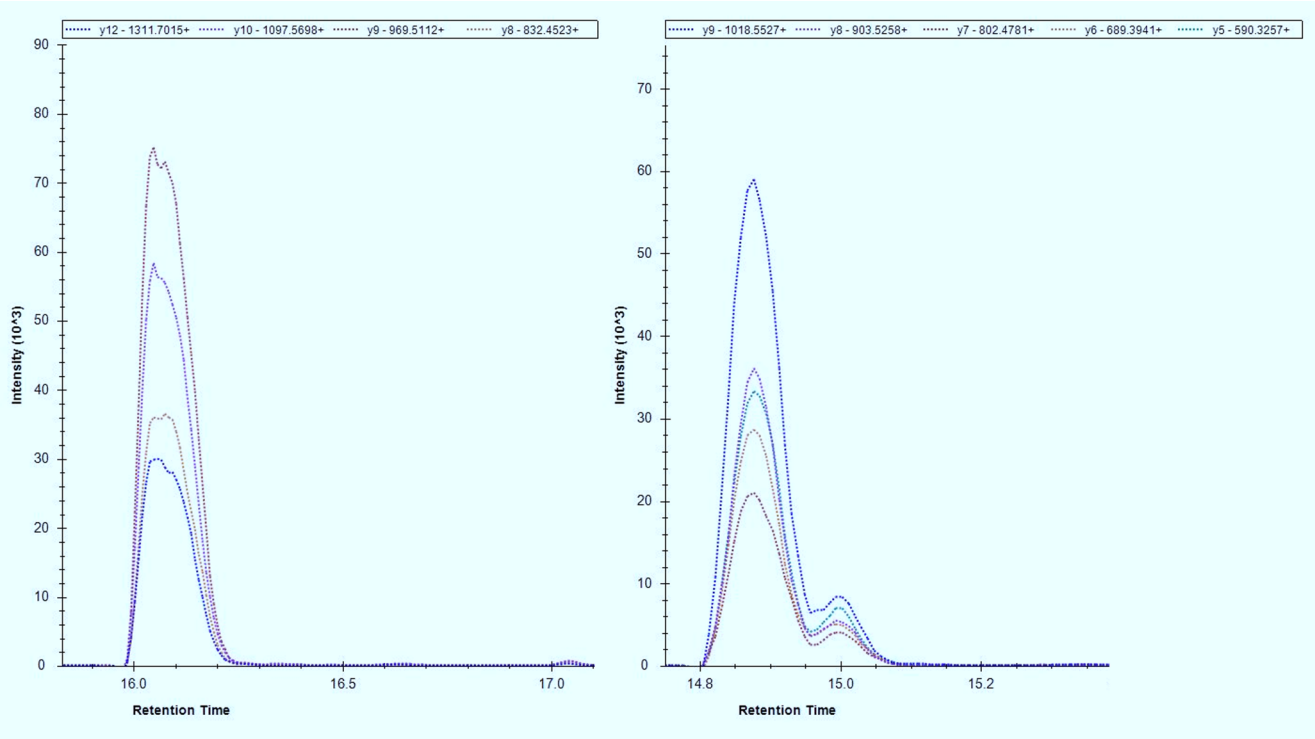


Fig. 5 MRM of LC-MS/MS chromatograms of P25 (DLEEATLQHEATAATLR) (left) and P68 (QLDEKDTLVSQLSR) (right)

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00217-025-04814-x.

Author contributions MHY: project administration, conceptualization, methodology, formal analysis, investigation, data curation, writing—original draft preparation, writing—reviewing and editing. NA: writing—reviewing and editing. SII: writing—reviewing and editing. MSAS: writing—reviewing and editing. LKS: writing—reviewing and editing.

Funding Open access funding provided by The Ministry of Higher Education Malaysia and Universiti Putra Malaysia. Open access funding provided by the Islamic Economics Research and Innovation Fund (IERIF) [I-RISE/IERIF/AWARD/2025/BATCH 2/71] with a Trust Account No. 6300564-10205.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

Ethical approval This study does not contain any studies with human participants or animal performed of any of the authors.

References

1. van Boekel M et al (2010) A review on the beneficial aspects of food processing. Mol Nutr Food Res 54(9):1215–1247
2. Fazryatul HAR, Nor Aini M, Faridah HH (2017) Halal and kosher marketing strategies. J Islam Manage Stud 1(1):104–116
3. Adam AR, Bahari CA, Jalil MNA (2020) Halal quality management system: a comparison between halal and kosher. Al-itqan J Islam Sci Comp Stud 4(1):89–115
4. Al-Teinaz YR, Regenstein JM, Lever J, Katme AM, Unsdorfer S (2020) The Halal and Kosher food experience in the UK. In: Al-Teinaz YR, Spear S, El-Rahim IHAA (eds) The Halal food handbook, 1st edn. John Wiley & Sons Ltd, pp 331–341. https://doi.org/10.1002/9781118823026.ch20
5. Regenstein JM (2020) Kosher and halal: how they affect muslim and jewish dietary practices. In: Meiselman HL (ed) Handbook of eating and drinking. Springer International Publishing, Ithaca, NY, USA, pp 593–614. https://doi.org/10.1007/978-3-030-14504-0_158
6. Fischer J (2016) Markets, religion, regulation: kosher, halal and hindu vegetarianism in global perspective. Geoforum 69:67–70. https://doi.org/10.1016/j.geoforum.2015.12.011

7. Chu DT et al (2019) The possible zoonotic diseases transferring from pig to human in Vietnam. Eur J Clin Microbiol Infect Dis 38(6):1003–1014. https://doi.org/10.1007/s10096-018-03466-y
8. Yusop MHM, Bakar MFA (2020) Review on halal forensic: a focus on DNA-based methods for pork authentication. Food Res 4(6):2347–2354. https://doi.org/10.26656/fr.2017.4(6).180
9. Kim GD, Yang HS, Jeong JY (2017) Comparison of myosin heavy chain content determined by label-free quantification between Porcine longissimus thoracis, Psoas major and semimembranosus muscles. Food Res Int 100(July):504–513. https://doi.org/10.1016/j.foodres.2017.07.050
10. Kim GD, Jeong TC, Cho KM, Jeong JY (2017) Identification and quantification of myosin heavy chain isoforms in bovine and Porcine longissimus muscles by LC-MS/MS analysis. Meat Sci 125(3):143–151. https://doi.org/10.1016/j.meatsci.2016.12.002
11. Jeong JY, Jung EY, Jeong TC, Yang HS, Kim GD (2016) Analysis of methionine oxidation in myosin isoforms in porcine skeletal muscle by LC-MS/MS analysis. Korean J Food Sci Anim Resour 36(2):254–261
12. Kim GD, Jeong JY, Yang HS, Joo ST (2013) Identification of myosin heavy chain isoforms in Porcine longissimus dorsi muscle by electrophoresis and mass spectrometry. Electrophoresis 34(9–10):1255–1261. https://doi.org/10.1002/elps.201300049
13. Mitra B, Lametsch R, Greco I, Ruiz-Carrascal J (2018) Advanced glycation end products, protein crosslinks and post translational modifications in pork subjected to different heat treatments. Meat Sci 145(July):415–424. https://doi.org/10.1016/j.meatsci.2018.07.026
14. Watson AD, Gunning Y, Rigby NM, Philo M, Kemsley EK (2015) Meat authentication via multiple reaction monitoring mass spectrometry of myoglobin peptides. Anal Chem 87(20):10315–10322. https://doi.org/10.1021/acs.analchem.5b02318
15. Amir SH et al (2021) Comparative database search engine analysis on massive tandem mass spectra of pork-based food products for halal proteomics. J Proteom 241:104240
16. Kim GD, Seo JK, Yum HW, Jeong JY, Yang HS (2017) Protein markers for discrimination of meat species in raw beef, pork and poultry and their mixtures. Food Chem 217:163–170. https://doi.org/10.1016/j.foodchem.2016.08.100
17. Pan XD, Chen J, Chen Q, Huang BF, Han JL (2018) Authentication of pork in meat mixtures using PRM mass spectrometry of myosin peptides. RSC Adv 8(20):11157–11162. https://doi.org/10.1039/c8ra00926k
18. Li Y, Zhang Y, Kang C, Zhao W, Li S, Wang S (2020) Assessment of carbonic anhydrase 3 as a marker for meat authenticity and performance of LC-MS/MS for pork content. Food Chem 342(October):128240. https://doi.org/10.1016/j.foodchem.2020.128240
19. Von Bargen C, Brockmeyer J, Humpf HU (2014) Meat authentication: a new HPLC-MS/MS based method for the fast and sensitive detection of horse and pork in highly processed food. J Agric Food Chem 62(39):9428–9435. https://doi.org/10.1021/jf503468t
20. von Bargen C, Dojahn J, Waidelich D, Humpf HU, Brockmeyer J (2013) New sensitive high-performance liquid chromatography–tandem mass spectrometry method for the detection of horse and pork in halal beef. J Agric Food Chem 61(49):11986–11994
21. Sarah SA, Faradalila WN, Salwani MS, Amin I, Karsani SA, Sazili AQ (2016) LC-QTOF-MS identification of porcine-specific peptide in heat treated pork identifies candidate markers for meat species determination. Food Chem 199:157–164. https://doi.org/10.1016/j.foodchem.2015.11.121
22. Wold S, Sjöström M (1998) Chemometrics, present and future success. Chemometr Intell Lab Syst 44(1–2):3–14
23. Yuswan MH et al (2018) Chemometrics-assisted shotgun proteomics for establishment of potential peptide markers of non-halal pork (Sus scrofa) among halal beef and chicken. Food Anal Methods 11(12):3505–3515
24. Yuswan MH et al (2019) Improved gel-enhanced liquid chromatography-mass spectrometry by chemometrics for halal proteomics. Chemometr Intell Lab Syst 192:103825
25. Yuswan MH et al (2021) Hydroxyproline determination for initial detection of halal-critical food ingredients (gelatin and collagen). Food Chem 337:127762. https://doi.org/10.1016/j.foodchem.2020.127762
26. Schork K, Podwojski K, Turewicz M, Stephan C, Eisenacher M (2021) Important issues in planning a proteomics experiment: statistical considerations of quantitative proteomic data. In: Marcus K, Eisenacher M, Sitek B (eds) Quantitative methods in proteomics, 2nd edn, vol 2228, ch 1. Humana Press Inc., pp 1–20. https://doi.org/10.1007/978-1-0716-1024-4_1
27. Nuzzo RL (2016) The box plots alternative for visualizing quantitative data. PM R 8(3):268–272. https://doi.org/10.1016/j.pmrj.2016.02.001
28. Dawson R (2011) How significant is a boxplot outlier? J Stat Educ 19(2). https://doi.org/10.1080/10691898.2011.11889610
29. Odoi B, Twumasi-Ankrah S, Samita S, Al-Hassan S (2022) The efficiency of bartlett's test using different forms of residuals for testing homogeneity of variance in single and factorial experiments-a simulation study. Sci Afr 17:e01323. https://doi.org/10.1016/j.sciaf.2022.e01323
30. Dubois E, Galindo AN, Dayon L, Cominetti O (2022) Assessing normalization methods in mass spectrometry-based proteome profiling of clinical samples. BioSystems 215–216. https://doi.org/10.1016/j.biosystems.2022.104661
31. Pleil JD (2016) QQ-plots for assessing distributions of biomarker measurements and generating defensible summary statistics. J Breath Res 10(3):35001. https://doi.org/10.1088/1752-7155/10/3/035001
32. Singh M, Domijan K (2019) Comparison of machine learning models in food authentication studies. In: 30th Irish signals and systems conference (ISSC). IEEE, Maynooth, Ireland, pp 1–6. https://doi.org/10.1109/ISSC.2019.8904924
33. Pang T et al (2021) Quantitative analysis of a weak correlation between complicated data on the basis of principal component analysis. J Anal Methods Chem. https://doi.org/10.1155/2021/8874827
34. Yu TY, Morton JD, Clerens S, Dyer JM (2017) Cooking-induced protein modifications in meat. Compr Rev Food Sci Food Saf 16(1):141–159. https://doi.org/10.1111/1541-4337.12243
35. Nalazek-Rudnicka K, Kłosowska-Chomiczewska I, Wasik A, Macierzanka A (2019) MRM–MS of marker peptides and their abundance as a tool for authentication of meat species and meat cuts in single-cut meat products. Food Chem 283:367–374. https://doi.org/10.1016/j.foodchem.2019.01.007
36. European Commission (2017) Guidance document on analytical quality control and method validation procedures for pesticides residues analysis in food and feed. SANTE/11813/2017. https://doi.org/10.13140/RG.2.2.33021.77283