Research Article

Food forensics on gelatine source via ultra-high-performance liquid chromatography diode-array detector and principal component analysis

Azilawati Mohd Ismail[1] · Muhamad Shirwan Abdullah Sani[2,3] [iD] · Azman Azid[4] · Nor Nadiha Mohd Zaki[3,5] · Syariena Arshad[3,5] · Nur Azira Tukiran[2,3] · Siti Aimi Sarah Zainal Abidin[6,7] · Mohd Saiful Samsudin[8] · Amin Ismail[5,9]

Abstract

This study provided a step-by-step procedure to investigate the distribution of 17 amino acids (AAs) in 50 fish, 50 bovine and 54 porcine gelatines using Ultra-High-Performance Liquid Chromatography Diode-Array Detector (UHPLC–DAD) with the incorporation of principal component analysis (PCA). Dataset pre-processing step, including outlier removal, analysis of variance (ANOVA), dataset adequacy test, dataset transformation and correlation test was performed before the PCA. The method rendered linearity range of 37.5–1000 pmol/μL and accuracy of 85–111% recovery. The bovine and porcine gelatines showed a similar ranking while the L-Alanine (Ala), L-Arginine (Arg) and L-Glutamic acid (Glu) concentrations had differed the fish gelatine from the bovine and porcine gelatines. The PCA, which explained 77.013% cumulative variability at eigenvalue of 5.436, showed AAs with strong FL in PC1 had polar and nonpolar side chains while AAs with strong FL in PC2 had polar side chain. The AAs with moderate and weak FL in PC1 had a nonpolar side chain. The AAs with strong FL of in PC1 were also the same AAs with 7, 6 and 5 strong CMs as determined in the correlation test. The second PCA showed that the L-Serine (Ser), Arg, Glycine (Gly), L-Threonine (Thr), L-Methionine (Met), L-Histidine (His) and L-Hydroxyproline (Hyp) were significant in fish gelatine; Hyp, Met, Thr, Ser, His, Gly, and Arg in bovine gelatine; and L-Proline (Pro), L-Tyrosine (Tyr), L-Valine (Val), L-Leucine (Leu), and L-Phenylalanine (Phe) in porcine gelatine. The 100% fish, bovine and porcine gelatines accommodated grouping 1, 2 and 3, respectively, which proved that AAs with strong FL (Hyp, His, Ser, Arg, Gly, Thr, Pro, Tyr, Met, Val, Leu and Phe) were the significant AAs and becomes the biomarkers to identify the gelatine source. From this study, the PCA was a useful tool to analyse a multivariate dataset that could provide an in-depth understanding of AA distributions as compared to ANOVA and correlation test.

✉ Muhamad Shirwan Abdullah Sani, shirwansany@iium.edu.my | [1]Malaysia Halal Analysis Centre (MyHAC), No.1, Persiaran Teknologi 1, Lebuh Enstek, 71760 Bandar Baru Enstek, Negeri Sembilan, Malaysia. [2]International Institute for Halal Research and Training, Level 3, KICT Building, International Islamic University Malaysia, Jalan Gombak, 53100 Kuala Lumpur, Malaysia. [3]Konsortium Institut Halal IPT Malaysia, Ministry of Higher Education, Block E8, Complex E, Federal Government Administrative Centre, 62604 Putrajaya, Malaysia. [4]Faculty of Bioresources and Food Industry, Universiti Sultan Zainal Abidin (UniSZA), 22200 Besut, Terengganu, Malaysia. [5]Halal Product Research Institute, Universiti Putra Malaysia, Serdang 43400, Selangor, Malaysia. [6]Faculty of Applied Sciences, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia. [7]Malaysia Institute of Transport, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia. [8]Faculty of Entrepreneurship and Business, Universiti Malaysia Kelantan, UMK City Campus, Pengkalan Chepa, Kelantan, Malaysia. [9]Department of Nutrition and Dietetics, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang 43400, Selangor, Malaysia.

1 Introduction

Gelatine is a mixture of insoluble collagenous polypeptides which is derived from acidic and alkaline hydrolyses of animal bones, cartilage, hides, tendons, skins, and sinews [1]. The gelatine production worth of USD1.34 billion consists of 80% porcine skin, 15% bovine hide and 5% porcine bone, bovine bone, and fish skin. The manufacturing of gelatine is aiming for various applications entailing dairy products, sausages, candies, gummies and marshmallows, capsules [2], excipients and beauty products. Gelatine contains 92% polypeptides, including 18 AAs and 2% salt and 6% moisture [3]. Although the human body can synthesize protein from naturally developed AAs, some AAs could not be produced by the human body; thus, it has to digest the AAs from food [4]. Such AAs known as essential AAs are His, Ile, Leu, Lys, Met, Phe, Thr, tryptophan, and Val [5]. On the other hand, the human body naturally synthesizes nonessential AAs that includes Ala, Arg, asparagine, Asp, cysteine, Glu, glutamine, Gly, Pro, Ser, and Tyr. Also, there is conditional AAs which are nonessential but important for the human body during stress and illness entailing Arg, cysteine, glutamine, Tyr, Gly, ornithine, Pro and Ser. Due to this reason, gelatine receives interest from food manufacturers to cater to human needs. However, some manufacturers make a false claim of the gelatine source for their business purposes which then has raised concern from certain groups of consumers such as vegetarians [6], Jews [7] and Muslims consumers [8]. This false claim has tainted the food integrity when the food does not correctly represent its claim [9] and may expose the consumer to diseases such as bovine spongiform encephalopathy or mad cow disease from the consumption of bovine-sourced food [10].

Various testing methods have been utilized to address the issue of a false claim and food integrity—most gelatine testing use polymerase-chain-reaction (PCR) method to identify the source of the gelatine. Beside of the PCR, the immunological method using polyclonal anti-peptide antibodies in indirect and competitive indirect enzyme-linked immunosorbent assay (ELISA) has also been reported as another successful method; however, these methods are sophisticated, very costly and prone to contamination [2]. Liquid chromatography methods, e.g. liquid chromatography time-of-flight mass spectrometer (LC-QTOF/MS) and liquid chromatography-mass spectrometer (LC/MS) have rendered capability to differentiate the animal sources. Still, most of their applications went to meat speciation and highly expensive [11] for the maintenance in the testing laboratories, especially for the new ones. However, a much affordable liquid chromatography method such as high-performance liquid chromatography has proven to a successful application to differentiate gelatine sources [10]. To improve the sensitivity of the AA analysis in gelatine, this study employed Ultra-High-Performance Liquid Chromatography Diode-Array Detector (UHPLC-DAD), which has not been reported previously.

Although UHPLC-DAD may render lower sensitivity of AA detection, it could not differentiate the gelatine source via comparison of the individual AA from the gelatine against the AA standard since all gelatines of animal source possess similar distribution of AAs. This claim was evident from a comparison study between AAs from porcine and bovine skins without indication of the significant difference between the two gelatines [12]. Previous research also did not identify which AAs were selected as the biomarkers to differentiate the gelatine source. Thus, our study performed the analysis of variance (ANOVA) to identify which AAs were significantly different in fish, bovine and porcine gelatines. Nevertheless, the ANOVA was not enough to discriminate the gelatine source unless the dataset was subjected to multivariate data analysis such as principal component analysis [14].

Principal component analysis (PCA) is an unsupervised technique to explore the dataset of multivariate data which requires prerequisite analyses prior to the analysis including (1) removal of outliers [15], (2) ensuring the dataset adequacy [16], and (3) data transformation for normal distribution [17]. However, most dataset exploratory via PCA for food analysis did not undergo these prerequisite steps, including AA analysis prior to PCA [18]. Other studies performed dataset transformation based on previous reports without exhaustively investigate the suitable dataset transformation method for a specific matrix. For instance, although dataset transformation of standardize (n − 1) has been tested on gelatine matrix [10], the same gelatine dataset should also be tested on other transformation methods such as the log transformation before concluding the most suitable dataset transformation for the gelatine matrix. Without the fulfilment of these prerequisite analyses, the PCA may lead to erroneous result and interpretation. Previous researches have neglected the analysis of correlation test between AAs to support the PCA result and explained the effect of factor loading (FL) of each AA. The FL can be divided into strong, moderate and weak FL where the FL play an essential role in determining the apportionment of AAs in fish, bovine and porcine gelatines. Hence, this study provided step-by-step PCA procedure to explore the gelatine dataset and identify the apportionment of the AAs in the fish, bovine and gelatine sources. This study also anticipated the certification or regulatory bodies at the governmental level to adopt

this step-by-step PCA procedure in their development of guideline or standard for testing laboratory.

2 Methodology

2.1 Preparation of calibration standard solution of amino acids

A mixture of 17 standard stock solutions (SSS) of AA hydrolysate was purchased from Waters, USA containing 2.5 µmol/mL of L-Histidine (His), L-Serine (Ser), L-Arginine (Arg), Glycine (Gly), L-Aspartic acid (Asp), L-Glutamic acid (Glu), L-Threonine (Thr), L-Alanine (Ala), L-Proline (Pro), L-Lysine (Lys), L-Tyrosine (Tyr), L-Methionine (Met), L-Valine (Val), L-Isoleucine (Ile), L-Leucine (Leu) and L-Phenylalanine (Phe), and 1.25 µmol/mL L-Cystine (Cys). An SSS of L-Hydroxyproline (Hyp) and internal standard solution (ISS) of L-Aminobutyric acid (AABA) were prepared at 2500 µmol/mL each. A series of calibration standard solution (CSS) consisted of 37.5 pmol/µL, 100 pmol/µL, 250 pmol/µL, 500 pmol/µL and 1000 pmol/µL were prepared from the SSS while 100 pmol/µL of AABA was prepared from the ISS for each CSS.

2.2 Sample preparation

Before the sample preparation, gelatines from cold-water fish skin (G7041), bovine skin (G9382) and porcine skin (G6144) were freeze-dried to ensure their moisture content was less than 10%. An amount of 0.1–0.2 g of 54 G7041 gelatines, 50 G9382 gelatines and 54 G6144 gelatines were acid hydrolysed with 5 mL of 6 N hydrochloric acid and incubated at 110 °C for 24 h. The hydrolysate solution was mixed with 100 pmol/µL AABA, diluted to 100 mL with distilled water, filtered by 0.45 µm cellulose acetate membrane to produce mixture A.

2.3 Sample derivatization

A volume of 10 µL mixture A was derivatized with 70 µL AccQ.Fluor borate buffer (Waters, Massachusetts, USA) to generate pH 8.2–10.0 condition, derivatized with 20 µL of prepared AccQ.Fluor reagent (Waters, Massachusetts, USA) to produce mixture B. The prepared AccQ.Fluor reagent consists of a mixture of 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate (AQC) derivatizing reagent and AccQ. Fluor reagent diluent. The mixture B was heated at 55 °C for 10 min to ensure complete derivatization of the AQC with primary and secondary AAs prior injection to Ultra-High-Performance Liquid Chromatography Diode-Array Detector (UHPLC-DAD).

2.4 Amino acids analysis of gelatine samples

Prior to injection, a binary solvent system was used consisting of (A) aqueous solution of AccQ.Tag™ Eluent A, concentrate (WAT052890) (1:10) and (B) acetonitrile in deionised water (60:40) that were filtered using 0.45 µm membrane filters. A volume of 1 µL mixture B was injected into UHPLC-DAD (Agilent, USA) and eluted by the mobile phases according to gradient elution set-up: 0–0.5 min: 99.9% A, 0.1% B, 0.5 min: 99.9% A, 0.1% B, 5.7 min: 90.9% A, 9.1% B, 6.4 min: 87.4% A, 12.6% B, 6.6 min: 87% A, 13% B, 7.7 min: 78.8% A 21.2% B, 8.0 min: 40.4% A, 59.6% B, 8.6 min: 40.4% A, 59.6% B, 8.73 min: 99.9% A, 0.1% B and 9.5 min: 99.9% A, 0.1% B. The AAs in the mixture B were separated by Waters AccQ.Tag column (3.9 mm × 150 mm) at 1 mL/min and 36 °C and detected at 260 nm. The AA peaks are recorded in a chromatogram.

2.5 Method linearity and accuracy

The method linearity was established by injecting the 37.5 pmol/µL, 100 pmol/µL, 250 pmol/µL, 500 pmol/µL and 1000 pmol/µL CSS into the HPLC-DAD. Each concentration was measured in triplicate, and the ratio of the peak area of CSS over the peak area of AABA was computed and plotted against the ratio of CSS concentration over the AABA concentration. This plot constructed 17 linear regression lines with their respective slope, intercept and determination coefficients (R^2). The linear regression line with R^2 nearest to 1.00 indicates method linearity [19].

The method accuracy was carried out under repeatable analysis on fish, bovine and porcine gelatine which were spiked with 50 pmol/µL, 250 pmol/µL and 1000 pmol/µL amino acid standards that covered the working range. Each spiked concentration was analyzed separately in ten replication, and the percentage recovery of the spiked concentration was determined via this formulae:

$$\text{Recovery } (\%), \ R = \frac{\bar{x} - x_{spike}}{x_{spike}} \times 100$$

where $\bar{x}$ is the mean concentration of amino acid in the gelatines and x_{spike} is the spiking concentration of amino acids standards.

2.6 Dataset pre-processing

Peak area of AAs from each gelatine was imported to the dataset table in XLSTAT 2016 software. An about 54 fish, 50 bovine and 54 porcine gelatines with 17 AAs identified in the dataset was pre-processed to facilitate the process of differentiation among samples. This pre-processing

step is applied to reduce the variation of the amino acids in the dataset. Dataset pre-processing tests consisted of the analysis of variance (ANOVA), box and whisker plot, dataset transformation, Kaiser–Meyer–Olkin (KMO) test, correlation test and principal component analysis (PCA) were conducted using XLSTAT-Pro (2017) statistical software (Addinsoft, Paris, France).

2.6.1 Outlier removal by box and whisker plot

Prior to ANOVA and principal component analysis (PCA), the individual dataset of fish, bovine and porcine gelatines was subjected to outlier removal by using box and whisker plot (BWP) method from a standardized dataset. The confidence interval of the BWP was set at 95%. The skewness of the BWP was examined to confirm the need for dataset transformation. The dataset, which showed different pattern within the individual AA, was discriminated and shown in the BWP as an outlier. Outlier value which exceeded three times of the box's height, was signed with a dot, star or asterisk [20] and subjected to removal. After removing the outliers, the new dataset consisting 41, 40 and 45 fish, bovine and porcine gelatines, respectively was subjected to KMO test.

$$R_{ab} = n \sum a_i b_i - \sum a_i \sum b_i \Big/ \left(\sqrt{n \sum a_i^2 - (\sum a_i)^2} \times \sqrt{n \sum b_i^2 - (\sum b_i)^2} \right)$$

2.6.2 Analysis of variance

Results were expressed as mean ± standard deviation of triplicate for distribution of AAs in the gelatines. The analysis of variance (ANOVA) test of Tukey's test was applied to determine the significant difference between the means of AAs at a 95% confidence level ($p < 0.05$).

2.6.3 Kaiser–Meyer–Olkin test

The dataset was analysed for dataset adequacy by KMO test. Adequate dataset determines the ability of a generated model to extract latent variables from the dataset. In this study, the KMO test was employed at a significant level (α) of 0.05. The calculated KMO was ranked as: KMO < 0.5 = inadequate, 0.5 < KMO < 0.7 = mediocre, 0.7 < KMO < 0.8 = good, 0.8 < KMO < 0.9 = very good and KMO > 0.9 excellent to indicate the dataset adequacy; and only KMO > 0.5 was acceptable for PCA [16, 21].

2.6.4 Dataset transformation

To ensure dataset follow normal distribution before the PCA, the dataset normality was tested using Shapiro–Wilk, Anderson–Darling and Lilliefors at α of 0.05. The dataset was transformed using standardize n-1, standardize (n), centre, standard deviation^{-1} (n-1), standard deviation^{-1} (n), rescale from 0 to 1, rescale from 0 to 100, Pareto and log transformation methods. For log transformation, the AA with 0 pmol/L was removed before the transformation. The transformation of each AA was employed to ensure the transformed dataset remained closer to the original AA dataset [22]. The normal distribution of the transformed dataset was evaluated by normality test of Shapiro–Wilk, Anderson–Darling and Lilliefors at α of 0.05. The best transformation method and normality test were selected from the result.

2.6.5 Correlation test

Study of correlation between was carried out using Pearson correlation to measure the strength (weak, moderate or strong) and direction (positive and negative) of the linear relationship between two AAs. The formula of the Pearson correlation matrix (R) of AA a and b as follows:

where n was the number of gelatines, ai was the value of AA a for ith gelatine and bi was value of AA b for ith gelatine. In this study, the strong, weak and moderate correlation were determined correlation matrix (CM) value; $|0.000| < R < |0.300|$ for weak, $|0.300| < R < |0.700|$ for moderate and $|0.700| < R < |1.000|$ for strong CMs.

2.7 Dataset exploratory by principal component analysis

The PCA of Pearson correlation was applied to recognize the dataset pattern, explore the contribution of each AA to the gelatines, find and explain the variance of intercorrelated AAs and transform the dataset into smaller sets of new independent variables which are called as principal components (PCs). The principle of PCA was to reduce significantly ($p < 0.05$) the dataset dimensionality to achieve these aims. The PCs can be expressed as:

$$C_{xy} = a_{x1} b_{1y} + a_{x2} b_{2y} + a_{x3} b_{3y} + \cdots + a_{xn} b_{ny}$$

where C is the component score, a is the factor loading (FL), b is the AA concentration, x is the PC number, y is the sample number and n is the total number of the AA.

In this study, the number of PCs selected was based on cumulative variability of ≥ 75% and eigenvalue criteria ≥ 1. The AA with strong FL (FL ≥ |0.750|) was considered as an AA with strong factor loading while AAs with VF coefficient of |0.500| < FL < |0.749| and FL ≤ |0.499| were considered as having moderate and weak factor loadings, respectively [23]. To determine the contribution of AAs to the gelatine sources, the correlation, symmetric and distance biplots were plotted, and the biplot that clearly showed the gelatine groupings was selected and the AAs which contributed to each grouping were examined and further explained.

2.8 Apportionment of amino acids in gelatine groupings by principal component analysis

The AAs with moderate and weak FL and AAs which did not contribute to the gelatine grouping were removed from the gelatine dataset (Ile, Glu, Ala, Lys and Asp). By using 12 AAs with strong FL, a new PCA was executed to produce a new AA plot and correlation biplot. The FL and correlations of the AAs were assessed, the apportionment of AAs to the gelatine groupings was examined, and the improvement of the gelatines that made the gelatine groupings was observed and compared to the previous PCA of 17 AAs.

3 Result and discussion

3.1 Method linearity and accuracy

Table 1 exhibits the linearity of the CSS by providing the linear regression line for each amino acid. The relationships between amino acid concentration and peak area were linear over the 37.5–1000 pmol/µL range with R^2 between 0.96 and 1.00. Since $R^2 > 0.95$ denotes a strong linear relationship between the amino acid concentration and the peak area [19], thus the established R^2 in this study indicated method linearity.

Table 1 explains the method accuracy, where the overall amino acids had method recovery between 85 and 111%. This is evident when 50 pmol/µL, 250 pmol/µL and 1000 pmol/µL amino acid standards rendered 85.3–110.7%, 92.4–111.4%, 87.6–110.8% recoveries, respectively. Among the standards, the 50 pmol/µL Leu had the lowest recovery of 85.3% while 1000 pmol/µL Gly had the highest recovery. These results indicated that the

Table 1 Linearity and accuracy of method of analysis for amino acid

Amino acid[1,2]	Retention time, min	Linearity			Accuracy (Recovery, %)		
		Linear range, pmol/µL	Calibration linear equation	Determination coefficients (R^2)	50 pmol/µL	250 pmol/µL	1000 pmol/µL
Hyp	1.773	37.5–1000	Y = 0.0109x − 0.1242	0.99	92.5	97.2	110.5
His	1.887	37.5–1000	Y = 0.0123X − 0.2404	0.98	102.0	103.5	88.5
Ser	2.596	37.5–1000	Y = 0.0104x − 0.0948	0.98	92.6	100.5	97.3
Arg	2.692	37.5–1000	Y = 0.0122x + 0.3025	0.97	110.7	100.2	102.7
Gly	2.858	37.5–1000	Y = 0.0107x + 0.3503	0.97	101.8	107.3	110.8
Asp	3.122	37.5–1000	Y = 0.0075x + 0.5745	0.98	103.4	95.9	98.1
Glu	3.576	37.5–1000	Y = 0.0079x + 0.2700	0.96	94.7	100.5	94.9
Thr	3.997	37.5–1000	Y = 0.0101x + 0.4367	0.96	107.7	95.9	93.4
Ala	4.419	37.5–1000	Y = 0.0087x + 0.5774	0.97	101.0	92.4	102.4
Pro	5.158	37.5–1000	Y = 0.0095x + 0.2339	0.96	85.2	112.4	106.7
Cys	6.786	37.5–1000	Y = 0.0102x + 0.1753	0.97	99.5	94.8	98.5
Lys	6.927	37.5–1000	Y = 0.0121x + 0.3505	0.97	109.4	103.0	99.9
Tyr	7.142	37.5–1000	Y = 0.0126x + 0.2992	0.98	104.3	100.5	87.6
Met	7.371	37.5–1000	Y = 0.0112x + 0.1467	0.97	101.3	102.1	102.9
Val	7.983	37.5–1000	Y = 0.0103x + 0.4938	0.97	97.2	107.2	88.0
Ile	8.272	37.5–1000	Y = 0.0104x − 0.0750	1.00	104.9	100.2	98.9
Leu	8.361	37.5–1000	Y = 0.0100x + 0.0813	0.98	85.3	111.4	99.3
Phe	8.473	37.5–1000	Y = 0.0123x − 0.2459	0.98	100.4	99.1	100.5

[1]Hyp = L-Hydroxyproline, His = L-Histidine, Ser = L-Serine, Arg = L-Arginine, Gly = Glycine, Asp = L-Aspartic acid, Glu = L-Glutamic acid, Thr = L-Threonine, Ala = L-Alanine, Pro = L-Proline, Lys = L-Lysine, Tyr = L-Tyrosine, Met = L-Methionine, Val = L-Valine, Ile = L-Isoleucine, Leu = L-Leucine and Phe = L-Phenylalanine

[2]Total of 54, 50 and 50 fish, bovine and porcine gelatines were tested

method bias was minute, which was in accordance with the recovery result produced by Azilawati [19] and fulfilled the requirement of the Eurachem guidelines [24].

3.2 Amino acids concentration in fish, bovine and porcine gelatines

The concentrations of individual AA in fish, bovine and porcine gelatines are presented in Table 2, and the chromatogram of the fish, bovine and porcine gelatines are depicted in Fig. 1. For fish gelatine, the lowest and highest concentration of AAs were Tyr at the mean value of 113 pmol/L and Gly at the mean value of 12,268 pmol/L, respectively. The ranking of AA concentration in fish gelatine as follows: Gly > Ala > Pro > Arg > Glu > Hyp > Ser > Lys > Asp > Thr > Val > Leu > Met > Phe > Ile > His > Tyr. This ranking supported the finding of AAs in *Chitala ornata* gelatines with Gly > Ala > Pro ranking [25] although other researches findings have slight differences; Gly, Pro and Ala were the highest identified AAs with Gly > Pro > Ala ranking in *Oreochromis nilotica*, *Clarias batrachus* and *Pangasius sutchi* [26]. To date, many research of fish-sourced gelatine found the highest concentration of Gly in *Epinephelus sexfasciatus*, *Lutjianus argentimaculatus*, *Rastrelliger kanagurta*, and *Pristipomodes typus* [27],

Oreochromis mossambicus [28], *Probarbus Jullieni* [29] as well as fish skin in our study. Jamilah et al. [26] also identified Leu > Lys > Tyr as the AAs in *Oreochromis nilotica* with the lowest concentrations which were absent of His and Phe as obtained in our finding. These different findings were due to our fish gelatine was of cold-water fish skin while Jamilah et al. [26] analysed freshwater fish skin.

The ranking of AA concentration in bovine gelatine as follows: Gly > Pro > Hyp > Ala > Glu > Arg > Lys > Asp > Ser > Leu > Val > Thr > Phe > Ile > Met > Tyr > His (Table 2). The presence of Gly, Pro and Hyp was anticipated since these AAs were the building block of collagen; thus, bovine gelatine was preferable in producing collagen products [30]. However, Gly and Pro were of the highest concentration of AAs in the same type of bovine skin [12] that was following our finding while Hyp was not analysed in Hafidz's study.

The ranking of AA concentration in porcine gelatine as follows: Gly > Pro > Hyp > Ala > Glu > Arg > Lys > Asp > Ser > Leu > Val > Thr > Phe > Ile > Met > Tyr > His. Azilawati, et al. [10] identified Gly, Pro and Hyp and Tyr, Met and Ile as the highest and lowest concentrations of AAs in porcine, respectively. This finding of the Tyr, Met and Ile differed to our finding might be due to lower UHPLC-DAD employed in our study.

Table 2 The concentration of amino acids in fish, bovine and porcine gelatines

Amino acid[1,2]	Fish gelatine[3], pmol/L			Bovine gelatine[3], pmol/L			Porcine gelatine[3], pmol/L		
	Minimum	Maximum	Mean	Minimum	Maximum	Mean	Minimum	Maximum	Mean
Hyp	1987	2250	2144 ± 57^c	2850	3789	3538 ± 165^b	3366	3824	3688 ± 122^a
His	300	367	331 ± 19^a	0	142	4 ± 23^b	0	212	36 ± 74^b
Ser	1807	2245	2076 ± 87^a	881	1105	1015 ± 49^c	1006	1193	1082 ± 43^b
Arg	2130	2471	2330 ± 83^a	1607	2080	1937 ± 89^c	2015	2352	2204 ± 93^b
Gly	11,496	13,216	$12,268 \pm 376^a$	9592	11,613	$11,023 \pm 462^c$	10,708	12,495	$11,811 \pm 445^b$
Asp	1122	1630	1329 ± 137^a	956	1500	1200 ± 162^b	818	1549	1211 ± 172^b
Glu	1912	2580	2182 ± 183^a	1770	2679	2204 ± 248^a	1677	2739	2267 ± 244^a
Thr	808	1025	949 ± 42^a	497	628	584 ± 27^c	585	663	626 ± 20^b
Ala	3064	3877	3393 ± 229^a	2605	3861	3265 ± 339^a	2609	3925	3395 ± 309^a
Pro	3000	3608	3282 ± 162^c	3297	4402	3880 ± 292^b	3522	4750	4228 ± 291^a
Lys	1162	1598	1334 ± 118^{ab}	912	1563	1262 ± 186^b	929	1641	1355 ± 161^a
Tyr	85	130	113 ± 11^b	72	111	92 ± 10^c	128	180	157 ± 13^a
Met	549	673	621 ± 29^a	93	318	214 ± 64^c	212	300	268 ± 17^b
Val	637	773	703 ± 34^c	624	808	727 ± 50^b	714	894	824 ± 44^a
Ile	374	458	415 ± 23^a	110	454	403 ± 54^a	321	403	368 ± 17^b
Leu	627	782	701 ± 39^c	695	894	802 ± 54^b	738	927	854 ± 45^a
Phe	441	535	495 ± 20^b	387	646	481 ± 34^c	509	576	544 ± 17^a

[1]Hyp=L-Hydroxyproline, His=L-Histidine, Ser=L-Serine, Arg=L-Arginine, Gly=Glycine, Asp=L-Aspartic acid, Glu=L-Glutamic acid, Thr=L-Threonine, Ala=L-Alanine, Pro=L-Proline, Lys=L-Lysine, Tyr=L-Tyrosine, Met=L-Methionine, Val=L-Valine, Ile=L-Isoleucine, Leu=L-Leucine and Phe=L-Phenylalanine

[2]Total of 54, 50 and 50 fish, bovine and porcine gelatines were tested

[3]Means with the different superscript letter are significantly different ($p < 0.05$)
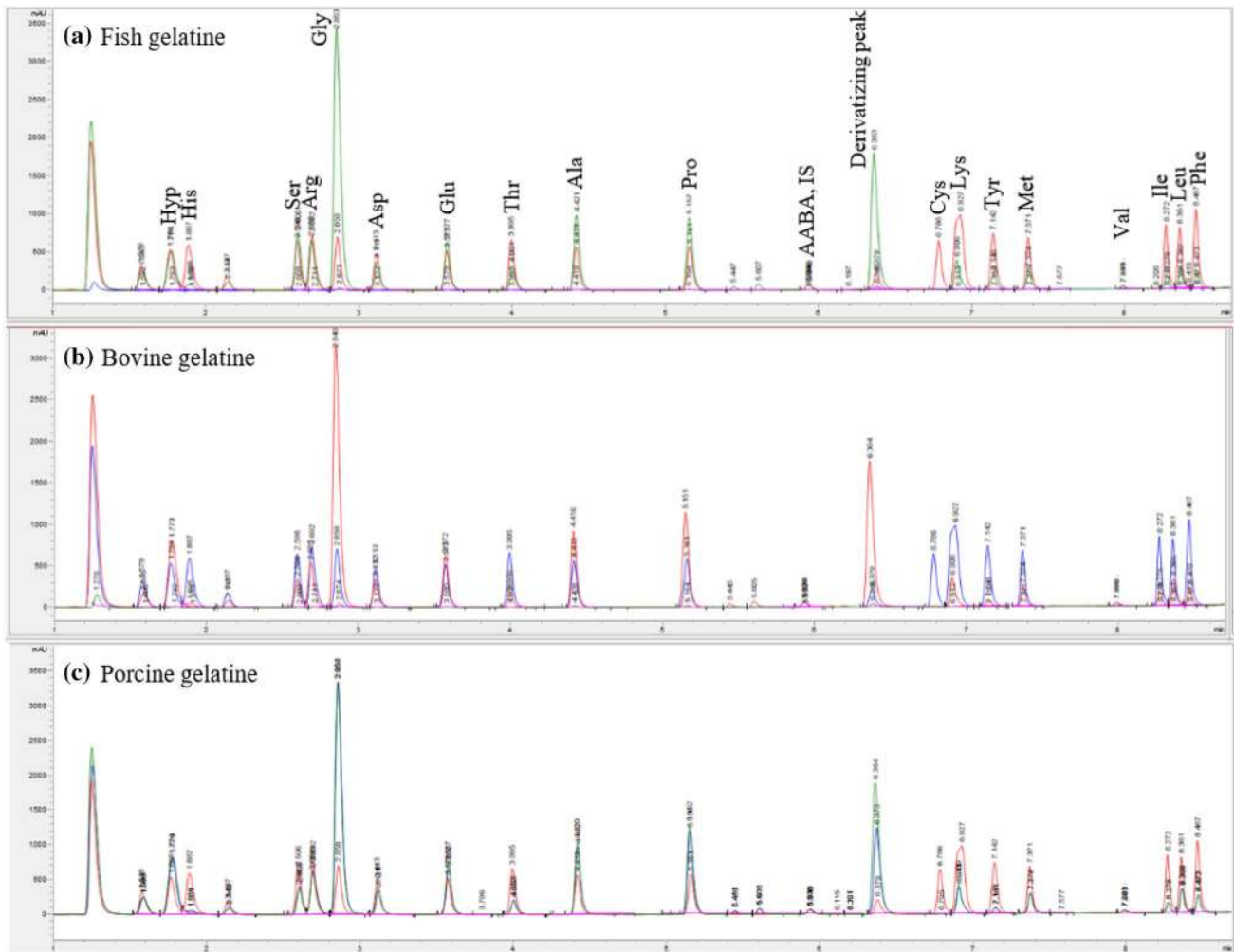
Fig. 1 Chromatogram of amino acids of **a** fish, **b** bovine and **c** porcine gelatines

By comparing the acid amino rankings, bovine and porcine gelatines showed a similar ranking while the Ala, Arg and Glu concentrations had differed the fish gelatine form the bovine and porcine gelatines. These findings indicated that (1) fish gelatine could be clearly discriminated from bovine and porcine gelatines and (2) bovine and porcine gelatines were hardly differentiated via UHPLC-DAD alone. Some studies suggested an application of multivariate dataset analysis to discriminate observations with similar distributions [13, 14]. Nevertheless, the discrimination of these gelatines could be achieved by assessing the significantly different ($p < 0.05$) AAs among the gelatines in Table 2. The table exhibited the highest concentration of Ser, Arg, Gly, Thr and Met in fish gelatine and Hyp, Pro, Tyr, Val, Leu and Phe in porcine gelatine. These AAs may characterize the gelatine sources based on the concentration of Hyp, Ser, Thr, Pro, Met as discovered by Azilawati et al. [10] and Gly, Pro and Hyp by Gómez-Guillén et al. [31]. Azilawati et al. [10] also recommended applying

multivariate dataset analysis, such as principal component analysis (PCA) [18] to assist the exploratory of gelatine dataset. Although only limited studies have employed PCA in gelatine study, the exploration of gelatine dataset may contribute to upholding the food integrity since gelatine which was not represent the actual source as it was claimed may deteriorate the confidence on a gelatine product especially from the vegetarians [6], Jews [7] and Muslims consumers [8]. False claim of the gelatine source may expose consumers to potential infection of bovine spongiform encephalopathy or 'mad cow disease' [10].

3.3 Outlier removal

Figure 2 shows the BWP of fish, bovine and porcine gelatines. The BWP of fish gelatine in Fig. 2a demonstrated outliers in the Hyp (1987 pmol/L), Ser (1807 pmol/L), Arg (2130 pmol/L), Gly (13,215 pmol/L), Thr (13,215 pmol/L) and Met (549 pmol/L) where observations appeared

Fig. 2 Box plots of standardized amino acids of **a** fish, **b** bovine and **c** porcine gelatines

(a) Box-whisker plots of standardized amino acids of fish gelatine

(b) Box-whisker plots of standardized amino acids of bovine gelatine

(c) Box-whisker plots of standardized amino acids of porcine gelatine
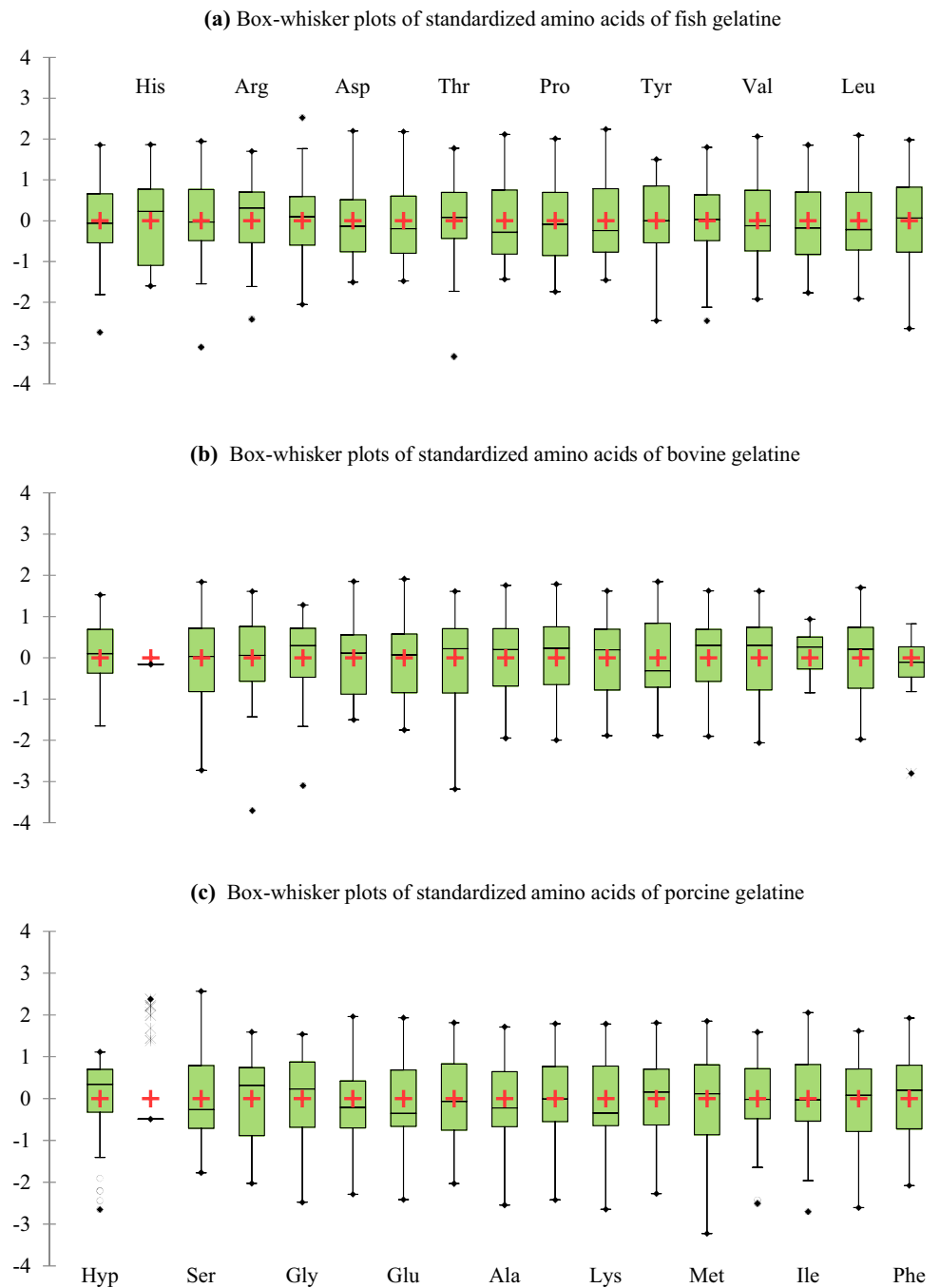
beyond the box's whisker [32]. Of these outliers, all outliers are of AAs with low concentrations than the mean except the Gly. The dataset of fish gelatine in Fig. 2a exhibited five AAs (Hyp, Ser, Pro, Tyr and Met) which had the median value that equal to the mean value where the mean value was assigned as a red plus. The His, Arg, Gly, Thr and Phe demonstrated right-skewed distribution because of the median value > mean value. Besides, the Asp, Glu, Ala, Lys, Val, Ile and Leu showed left-skewed distribution since the median value < mean value [15].

The dataset of bovine gelatine in Fig. 2b exhibited the Ser and Arg had median value = mean value. Other AAs demonstrated right-skewed distribution except for Tyr and Phe. The BWP of bovine gelatine in Fig. 2b also showed outliers in Arg, Gly and Phy where observations appeared beyond the box's whisker. Of these outliers, all outliers are of AAs with low concentration.

The dataset of porcine gelatine in Fig. 2c exhibited three AAs (Pro, Val and Ile), which had a median value = mean value. The Hyp, Arg, Gly, Tyr, Met, Leu and

Phe demonstrated right-skewed distribution because of the median value > mean value. Besides, the Ser, Asp, Glu, Thr, Ala, and Lys confirmed left-skewed distribution. The BWP of porcine gelatine in Fig. 2c demonstrated outliers in Hyp, His, Val and Ile where observations appeared beyond the box's whisker. Of these outliers, outliers of the Hyp, Val, Ile are of AAs with low concentrations except His. The Hyp, His and Val exhibited more than one outlier.

The removal of outliers had reduced the dataset distortion of the PCA as achieved in AAs study by Azilawati et al. [10]. Post outlier removal via BWP, the new dataset comprised 41, 40 and 45 fish, bovine and porcine gelatines, respectively. The AAs with the skewed distribution of BWP were also subjected to dataset normalization before the PCA.

3.4 Kaiser–Meyer–Olkin test

The KMO test calculated the value of 0.7542 for the new dataset. Researches from different fields have set appropriate KMO values prior to KMO test; KMO > 0.6 for study on sleep quality [33], customer attitude [34] and seafood spoilage [35]; and KMO > 0.7 for water quality assessment [36]. To the authors' knowledge, very limited researches have tested their dataset adequacy with KMO test prior to dataset transformation and PCA. This absent was evident in gelatine studies by Azilawati et al. [10] and Widyaninggar et al. [18] which may render erroneous findings due to small dataset for PCA [16]. Moreover, the preliminary study of AAs in gelatine by Widyaninggar et al. [18] did not explain total samples in their research. Based on these researches, KMO > 0.6 had been accepted as the appropriate KMO before performing PCA. Since our KMO value was between 0.7 and 0.8, the dataset was deemed as suitable for PCA and subjected to dataset transformation and PCA.

3.5 Dataset transformation

This study investigated issues of (1) which dataset transformation was suitable for AAs analysis in gelatine and (2) which normality test was the best to examine dataset normality.

The most common dataset transformation was the standardize (n-1) and followed by other dataset transformations deemed suitable according to sample type. After testing the dataset with all dataset transformations, the AAs of transformed dataset still demonstrated non-normal distribution except for log transformation.

Prior to dataset transformation, the normality test of Shapiro–Wilk exhibited three AAs involving Asp, Ala and Lys that followed a normal distribution (Table 3). The normality test for Glu demonstrated 0.1642 p value after log

transformation, which rendered normal distribution. However, the log transformation also changed Ala and Lys into non-normal distribution via p-value changes from 0.0940 to 0.0145 and from 0.0794 to 0.0045, respectively. Since the premise of normal distribution applied on the individual AA, hence Asp, Ala, Lys and Glu were confirming to the normal distribution after dataset transformation.

The normality test of Anderson–Darling exhibited that Ala, Lys and Val followed a normal distribution (Table 4) before the dataset transformation. The Asp and Glu also showed the normal distribution after log transformation with 0.1703 and 0.1023 p-values, respectively. Nonetheless, Lys exhibited opposite result after the same transformation as the p-value changed from 0.1274 to 0.0380. In this given view, the Ala, Lys, Val, Asp and Glu followed the normal distribution after the dataset transformation.

The normality test of Lilliefors showed the Asp, Ala, Lys, Val and Ile followed the normal distribution prior to the dataset transformation (Table 5). After the log transformation, His and Glu exhibited normal distribution with 0.0723 and 0.0717, respectively, while no other AAs exhibited non-normal distribution after the log transformation. To note, the AAs of porcine gelatine contained 0 pmol/L of His, which the log transformation of 0 value yielded undefined value. This study removed the 0 pmol/L before the log transformation and normality test; hence, we can omit the p-value of His because not all His concentrations were included in this dataset transformation. Subsequently, the Asp, Ala, Lys, Val, Ile and Glu had achieved this normal distribution after the dataset transformation. To note, Glu was the only AA followed the normal distribution after log transformation by the three normality tests.

Other studies have investigated the suitability of different dataset transformations for specific matrix before confirming the most suitable one. For instance, log, Box-Cox Standardize (n), 0 to 100 rescaling and Pareto transformations were confirmed suitable for microbiological, freshwater, sugarcane spirits [37], water quality [38] plant volatiles [39] matrices, respectively. For gelatine matrix, Azilawati et al. [10] has transformed the gelatine dataset via standardize (n-1) but has never tested the same dataset on other transformations, which our study had proven that log transformation was also suitable for gelatine matrix provided the dataset only has an integer value. On the contrary, a study of medical dataset recommended omitting any dataset transformation because the dataset distribution will never achieve normal distribution [40]. Our study suggested testing the gelatine dataset with different transformations before confirming the most suitable transformation for the gelatine matrix. Hence, for further analysis in this study, the 16 AAs were transformed using the standardize (n-1) while Glu was converted using log transformation.

Table 3 Normality test of Shapiro–Wilk after dataset transformations

Amino acid[1]	p-value of Shapiro-Wilk[2,3,4]									
	No transformation	Standardize (n-1)	Standardize (n)	Centre	Standard deviation (n-1)	Standard deviation (n)	Rescale from 0 to 1	Rescale from 0 to 100	Pareto	Log[5]
Hyp	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
His	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.0129
Ser	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Arg	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Gly	0.0060	0.0060	0.0060	0.0060	0.0060	0.0060	0.0060	0.0060	0.0060	0.0017
Asp	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	0.1615
Glu	0.0438	0.0438	0.0438	0.0438	0.0438	0.0438	0.0438	0.0438	0.0438	0.1642
Thr	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Ala	0.0940	0.0940	0.0940	0.0940	0.0940	0.0940	0.0940	0.0940	0.0940	0.0145
Pro	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0001
Lys	0.0794	0.0794	0.0794	0.0794	0.0794	0.0794	0.0794	0.0794	0.0794	0.0045
Tyr	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.0003
Met	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Val	0.0200	0.0200	0.0200	0.0200	0.0200	0.0200	0.0200	0.0200	0.0200	0.0296
Ile	0.0234	0.0234	0.0234	0.0234	0.0234	0.0234	0.0234	0.0234	0.0234	0.0301
Leu	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0001
Phe	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0004

[1]Hyp=L-Hydroxyproline, His=L-Histidine, Ser=L-Serine, Arg=L-Arginine, Gly=Glycine, Asp=L-Aspartic acid, Glu=L-Glutamic acid, Thr=L-Threonine, Ala=L-Alanine, Pro=L-Proline, Lys=L-Lysine, Tyr=L-Tyrosine, Met=L-Methionine, Val=L-Valine, Ile=L-Isoleucine, Leu=L-Leucine and Phe=L-Phenylalanine

[2]Null hypothesis (H$_0$)=The amino acids of the dataset followed a normal distribution and alternative hypothesis (H$_a$)=The amino acids from the dataset did not follow a normal distribution

[3]Bold p-value indicated p-value < 0.05, thus this study accepted the H$_a$, indicated that the amino acids from the dataset did not follow a normal distribution

[4]Dataset entailed 41, 40 and 45 fish, bovine and porcine gelatines, respectively

[5]The p-value of log transformation determined the p-value after removing 0 pmol/L amino acid

Table 4 Normality test of Anderson−Darling after dataset transformations

Amino acid[1]	p-value of Anderson-Darling[2,3,4]									
	No transformation	Standardize (n-1)	Standardize (n)	Centre	Standard deviation (n-1)	Standard deviation (n)	Rescale from 0 to 1	Rescale from 0 to 100	Pareto	Log[5]
Hyp	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
His	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.0064
Ser	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Arg	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Gly	0.0084	0.0084	0.0084	0.0084	0.0084	0.0084	0.0084	0.0084	0.0084	0.0029
Asp	0.0383	0.0383	0.0383	0.0383	0.0383	0.0383	0.0383	0.0383	0.0383	0.1703
Glu	0.0141	0.0141	0.0141	0.0141	0.0141	0.0141	0.0141	0.0141	0.0141	0.1023
Thr	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Ala	0.2873	0.2873	0.2873	0.2873	0.2873	0.2873	0.2873	0.2873	0.2873	0.1256
Pro	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Lys	0.1274	0.1274	0.1274	0.1274	0.1274	0.1274	0.1274	0.1274	0.1274	0.0380
Tyr	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.0002
Met	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Val	0.0652	0.0652	0.0652	0.0652	0.0652	0.0652	0.0652	0.0652	0.0652	0.1052
Ile	0.0263	0.0263	0.0263	0.0263	0.0263	0.0263	0.0263	0.0263	0.0263	0.0369
Leu	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Phe	0.0008	0.0008	0.0008	0.0008	0.0008	0.0008	0.0008	0.0008	0.0008	0.0027

[1] Hyp = L-Hydroxyproline, His = L-Histidine, Ser = L-Serine, Arg = L-Arginine, Gly = Glycine, Asp = L-Aspartic acid, Glu = L-Glutamic acid, Thr = L-Threonine, Ala = L-Alanine, Pro = L-Proline, Lys = L-Lysine, Tyr = L-Tyrosine, Met = L-Methionine, Val = L-Valine, Ile = L-Isoleucine, Leu = L-Leucine and Phe = L-Phenylalanine

[2] Null hypothesis (H_0) = The amino acids of the dataset followed a normal distribution and alternative hypothesis (H_a) = The amino acids from the dataset did not follow a normal distribution

[3] Bold p-value indicated p-value < 0.05, thus this study accepted the H_a, indicated that the amino acids from the dataset did not follow a normal distribution

[4] Dataset entailed 41, 40 and 45 fish, bovine and porcine gelatines, respectively

[5] The p-value of log transformation determined the p-value after removing 0 pmol/L amino acid

Table 5 Normality test of Lilliefors after dataset transformation

Amino acid[1]	p-value of Lilliefors[2,3,4]									
	No transformation	Standardize n-1	Standardize (n)	Centre	Standard deviation (n-1)	Standard deviation (n)	Rescale from 0 to 1	Rescale from 0 to 100	Pareto	Log[5]
Hyp	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**
His	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	0.0723
Ser	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**
Arg	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**
Gly	**0.0246**	**0.0246**	**0.0246**	**0.0246**	**0.0246**	**0.0246**	**0.0246**	**0.0246**	**0.0246**	**0.0207**
Asp	0.0783	0.0783	0.0783	0.0783	0.0783	0.0783	0.0783	0.0783	0.0783	0.4760
Glu	**0.0134**	**0.0134**	**0.0134**	**0.0134**	**0.0134**	**0.0134**	**0.0134**	**0.0134**	**0.0134**	0.0717
Thr	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**
Ala	0.3938	0.3938	0.3938	0.3938	0.3938	0.3938	0.3938	0.3938	0.3938	0.7157
Pro	**0.0021**	**0.0021**	**0.0021**	**0.0021**	**0.0021**	**0.0021**	**0.0021**	**0.0021**	**0.0021**	**0.0002**
Lys	0.4542	0.4542	0.4542	0.4542	0.4542	0.4542	0.4542	0.4542	0.4542	0.1497
Tyr	**0.0009**	**0.0009**	**0.0009**	**0.0009**	**0.0009**	**0.0009**	**0.0009**	**0.0009**	**0.0009**	**0.0038**
Met	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**
Val	0.3101	0.3101	0.3101	0.3101	0.3101	0.3101	0.3101	0.3101	0.3101	0.3366
Ile	0.1620	0.1620	0.1620	0.1620	0.1620	0.1620	0.1620	0.1620	0.1620	0.1944
Leu	**0.0063**	**0.0063**	**0.0063**	**0.0063**	**0.0063**	**0.0063**	**0.0063**	**0.0063**	**0.0063**	**0.0004**
Phe	**0.0179**	**0.0179**	**0.0179**	**0.0179**	**0.0179**	**0.0179**	**0.0179**	**0.0179**	**0.0179**	**0.0246**

[1]Hyp=L-Hydroxyproline, His=L-Histidine, Ser=L-Serine, Arg=L-Arginine, Gly=Glycine, Asp=L-Aspartic acid, Glu=L-Glutamic acid, Thr=L-Threonine, Ala=L-Alanine, Pro=L-Proline, Lys=L-Lysine, Tyr=L-Tyrosine, Met=L-Methionine, Val=L-Valine, Ile=L-Isoleucine, Leu=L-Leucine and Phe=L-Phenylalanine

[2]Null hypothesis (H$_0$)=The amino acids of the dataset followed a normal distribution and alternative hypothesis (H$_a$)=The amino acids from the dataset did not follow a normal distribution

[3]Bold p-value indicated p-value < 0.05, thus this study accepted the H$_a$, indicated that the amino acids from the dataset did not follow a normal distribution

[4]Dataset entailed 41, 40 and 45 fish, bovine and porcine gelatines, respectively

[5]The p-value of log transformation determined the p-value after removing 0 pmol/L amino acid

This study also recommended the Lilliefors as the best normality test to investigate dataset normality of AAs in gelatine, which was contrary to the study of Razali et al. [41] that found that the Shapiro–Wilk was the best normality test. However, Razali et al. [41] also found that normality test of Shapiro–Wilk was ineffective in low sample size (n < 10,000) which in line with our result due to low sample size (n = 125). Nonetheless, Bower [42] accepted the non-normal distribution prior to further dataset analysis because (1) the instrumental measurement of food sample was deemed as a continuous dataset that follows a normal distribution and (2) sample size (n > 100) has already followed the central limit theorem of the normal distribution [40, 43, 44].

3.6 Correlation test

The correlation test was sought to select the suitable correlation types to determine correlations between two AAs. The XLSTAT2017 software provided two correlation tests: Pearson or Spearman correlations. This study employed the Pearson correlation to explore the dataset in investigating the strength (weak, moderate, or strong) and direction (positive and negative) of the linear relationship between two AAs. Our study also emphasized the application of the Pearson correlation than the Spearman correlation since the former investigated the linear relationship between two continuous or parametric variables, i.e. AAs. On the contrary, the Spearman correlation is more suitable for non-parametric dataset such as ordinal or ranking dataset [45]. Thus, our study used the Pearson correlation to explore gelatine dataset.

Table 6 shows the correlation matrix (CM) of the AAs in fish, bovine, and porcine gelatines. The CM measures the strength and direction of a linear relationship between the two AAs. The CM indicated that positive and negative correlations among the AAs at $|0.700| < R < |1.000|$ for strong CM, $|0.300| < R < |0.700|$ for moderate CM and $^3|0.000| < R < |0.300|$ for weak CM as in generic guideline [46] although various research fields provided lower ranges for strong CM; $|0.500| < R < |1.000|$ in household food security [47] and in phenolic compounds of mungbean [48]. Study on AAs in *Limonia acidissima* fruit adopted the generic guideline of $|0.700| < R < |1.000|$ for strong CM [49]; therefore, our study referred this range to study the strong CM of the AAs. Table 6 exhibited the CM of these 17 AAs; Ser had 7 strong correlation matrices (CMs); Hyp and His had 6 strong CMs; Thr, Pro, Met and Leu had 5 strong CMs; Arg and Val had 4 strong CMs; Asp, Ala, Glu and Lys had 3 strong CMs; Tyr and Phe had 2 strong CMs; Gly had 1 CM; and Ile had no strong CM. Although these AAs rendered strong CM, the CM value

only explained the correlation between two AAs [46] while the CMs among more than two AAs brought more meaningful information, which the PCA can cater for this purpose [1].

3.7 Dataset exploratory by principal component analysis

The aim of using PCA in this study was to explore the gelatine dataset by (1) identifying AAs with different factor loadings, (2) differentiating type of biplots to study the correlation of the AAs and gelatine sources and (3) proposing the significant AAs that might contribute to the gelatine sources.

The exploratory of the dataset via PCA demonstrated PC1, PC2 and PC3 with eigenvalue (EV) > 1 [50] which explained 93.366% cumulative variability (CV) of the dataset (Table 6). The EV and percentage variability (PV) reflect the size and significant PC ($p < 0.05$) whereby PC1 has the largest EV than the next PC. The EV information supported or our result that the EVs decreased as the PC number increased, i.e. PC1 (EV = 7.656, PV = 45.036) > PC2 (EV = 5.436, PV = 31.977). Although there is no cut-off value of EV or PV, our study adopts the suggested EV > 1 as a principal guideline for food composition study [50] and further determined the PC selection based on the best PC numbers that explained the highest CV [51] via visual dimension. According to this principle, the PC1 and PC2 explained 77.013% CV at EV of 5.436 in this study. Furthermore, although the PCA generated three PCs, only PC1 and PC2 provided AAs with FL ≥ |0.75|: Hyp, His, Ser, Thr, Pro, Met, and Leu in PC1 and Asp, Glu, Ala, and Lys in PC2. The AAs with moderate FL ($|0.500| < FL < |0.749|$) were Arg, Gly, Val in PC1 and weak FL (FL ≤ |0.499|) were Tyr, Ile and Phe in PC1. Our study found the AAs with strong FL in PC1 had polar and nonpolar side chains. The Hyp, His, Ser and Thr had polar side chain while Pro, Met and Leu had nonpolar side chain. The AAs with strong FL in PC2 had polar side chain. AAs with moderate and weak FL in PC1 had nonpolar side chain [52]. Therefore, the PCA had identified the PC according to the polarity of the side chain of AA.

The strong FL of AAs in PC1 were also the same AAs with 7, 6 and 5 strong CMs as determined in the correlation test (Table 6). In contract, the AAs of strong FL in PC2 were the AAs with 3 strong CMs via correlation matrix study. Surprisingly, AAs with moderate FL were the AAs with 1 and 4 strong CMs and AAs with weak FL were the AAs with 0 and 2 strong CMs. This finding indicated that AAs with 0 and 2 strong CMs were the least contributing to dataset variability since dataset variability assists the determination of gelatine sources.

The positive and negative signs of the FL (Table 6) or AA plot in Fig. 3a explained the direction of the linear

Table 6 Correlation matrix and factor loading of amino acids in bovine, porcine and fish gelatines

Amino acid[1]	Correlation matrix, R[2,3,4,5,6,7]																	Factor loading (FL)[7,8,9]		
	Hyp (6)	His (6)	Ser (7)	Arg (4)	Gly (1)	Asp (3)	Glu (3)	Thr (5)	Ala (3)	Pro (5)	Lys (3)	Tyr (2)	Met (5)	Val (4)	Ile (0)	Leu (5)	Phe (2)	PC1	PC2	PC3
Hyp	**1.000**	−0.986	−0.971	−0.576	−0.519	−0.349	0.103	−0.953	−0.080	0.805	−0.055	0.309	−0.945	0.592	−0.447	0.789	0.346	−0.987[a]	−0.081[c]	0.036[c]
His	−0.986	**1.000**	0.992	0.675	0.600	0.340	−0.104	0.983	0.097	−0.777	0.071	−0.216	0.968	−0.528	0.417	−0.758	−0.239	**0.986**[a]	0.136[c]	0.067[c]
Ser	−0.971	0.992	**1.000**	0.710	0.641	0.360	−0.071	0.993	0.133	−0.738	0.109	−0.161	0.972	−0.464	0.422	−0.711	−0.173	**0.970**[a]	0.191[c]	0.101[c]
Arg	−0.576	0.675	0.710	**1.000**	0.906	0.226	0.029	0.772	0.278	−0.239	0.244	0.451	0.739	0.129	0.047	−0.229	0.440	0.582[b]	0.494[c]	0.610[b]
Gly	−0.519	0.600	0.641	0.906	**1.000**	0.367	0.205	0.695	0.463	−0.082	0.416	0.339	0.673	0.213	0.165	−0.096	0.339	0.514[b]	0.621[b]	0.426[c]
Asp	−0.349	0.340	0.360	0.226	0.367	**1.000**	0.874	0.343	0.883	0.142	0.891	0.010	0.411	0.205	0.536	0.151	−0.075	0.258[c]	**0.793**[a]	−0.460[c]
Glu	0.103	−0.104	−0.071	0.029	0.205	0.874	**1.000**	−0.067	0.944	0.578	0.953	0.223	0.011	0.568	0.377	0.589	0.131	−0.201[c]	**0.861**[a]	−0.424[c]
Thr	−0.953	0.983	0.993	0.772	0.695	0.343	−0.067	**1.000**	0.157	−0.698	0.125	−0.102	0.971	−0.405	0.415	−0.666	−0.105	**0.954**[a]	0.233[c]	0.161[c]
Ala	−0.080	0.097	0.133	0.278	0.463	0.883	0.944	0.157	**1.000**	0.488	0.978	0.229	0.213	0.556	0.456	0.498	0.182	−0.012[c]	**0.945**[a]	−0.303[c]
Pro	0.805	−0.777	−0.738	−0.239	−0.082	0.142	0.578	−0.698	0.488	**1.000**	0.490	0.492	−0.664	0.888	−0.211	0.977	0.506	−0.853[a]	0.494[c]	0.000[c]
Lys	−0.055	0.071	0.109	0.244	0.416	0.891	0.953	0.125	0.978	0.490	**1.000**	0.266	0.193	0.570	0.413	0.507	0.165	−0.040[c]	**0.937**[a]	−0.308[c]
Tyr	0.309	−0.216	−0.161	0.451	0.339	0.010	0.223	−0.102	0.229	0.492	0.266	**1.000**	−0.059	0.718	−0.546	0.479	0.842	−0.341[c]	0.485[c]	0.725[b]
Met	−0.945	0.968	0.972	0.739	0.673	0.411	0.011	0.971	0.213	−0.664	0.193	−0.059	**1.000**	−0.379	0.397	−0.645	−0.129	**0.934**[a]	0.280[c]	0.124[c]
Val	0.592	−0.528	−0.464	0.129	0.213	0.205	0.568	−0.405	0.556	0.888	0.570	0.718	−0.379	**1.000**	−0.211	0.889	0.716	−0.645[b]	0.673[b]	0.270[c]
Ile	−0.447	0.417	0.422	0.047	0.165	0.536	0.377	0.415	0.456	−0.211	0.413	−0.546	0.397	−0.211	**1.000**	−0.097	−0.448	0.439[c]	0.297[c]	−0.665[b]
Leu	0.789	−0.758	−0.711	−0.229	−0.096	0.151	0.589	−0.666	0.498	0.977	0.507	0.479	−0.645	0.889	−0.097	**1.000**	0.517	−0.832[a]	0.512[b]	−0.031[a]
Phe	0.346	−0.239	−0.173	0.440	0.339	−0.075	0.131	−0.105	0.182	0.506	0.165	0.842	−0.129	0.716	−0.448	0.517	**1.000**	−0.360[c]	0.434[c]	0.741[b]
Eigenvalue	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	7.656	5.436	2.780
Variability (%)	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	45.036	31.977	16.353
Cumulative explained variability (%)	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr	45.036	77.013	93.366

[1] Hyp = L-Hydroxyproline, His = L-Histidine, Ser = L-Serine, Arg = L-Arginine, Gly = Glycine, Asp = L-Aspartic acid, Glu = L-Glutamic acid, Thr = L-Threonine, Ala = L-Alanine, Pro = L-Proline, Lys = L-Lysine, Tyr = L-Tyrosine, Met = L-Methionine, Val = L-Valine, Ile = L-Isoleucine, Leu = L-Leucine, Phe = L-Phenylalanine

[2] Kaiser-Meyer-Olkin (KMO) value = 0.7542

[3] |0.000| < R < |0.300| = weak correlation, |0.300| < R < |0.700| = moderate correlation and |0.700| < R < |1.000| = strong correlation matrix

[4] Correlation matrix with the bold value indicated a strong correlation between amino acids

[5] Value in parenthesis of the amino acid indicated numbers of strong correlation matrix between amino acids

[6] nr = not related

[7] Dataset entailed 41, 40 and 45 fish, bovine and porcine gelatines, respectively

[8] FL ≥ |0.750| = strong factor loading, [b] |0.500| < FL ≤ |0.749| = moderate factor loading and [c] FL ≤ |0.499| = weak factor loading in the principal component

[9] Factor loading with bold superscript a indicated strong factor loading in the principal component

Fig. 3 **a** Amino acid plot and biplot of **b** correlation, **c** symmetric, and **d** distance of amino acids and gelatines
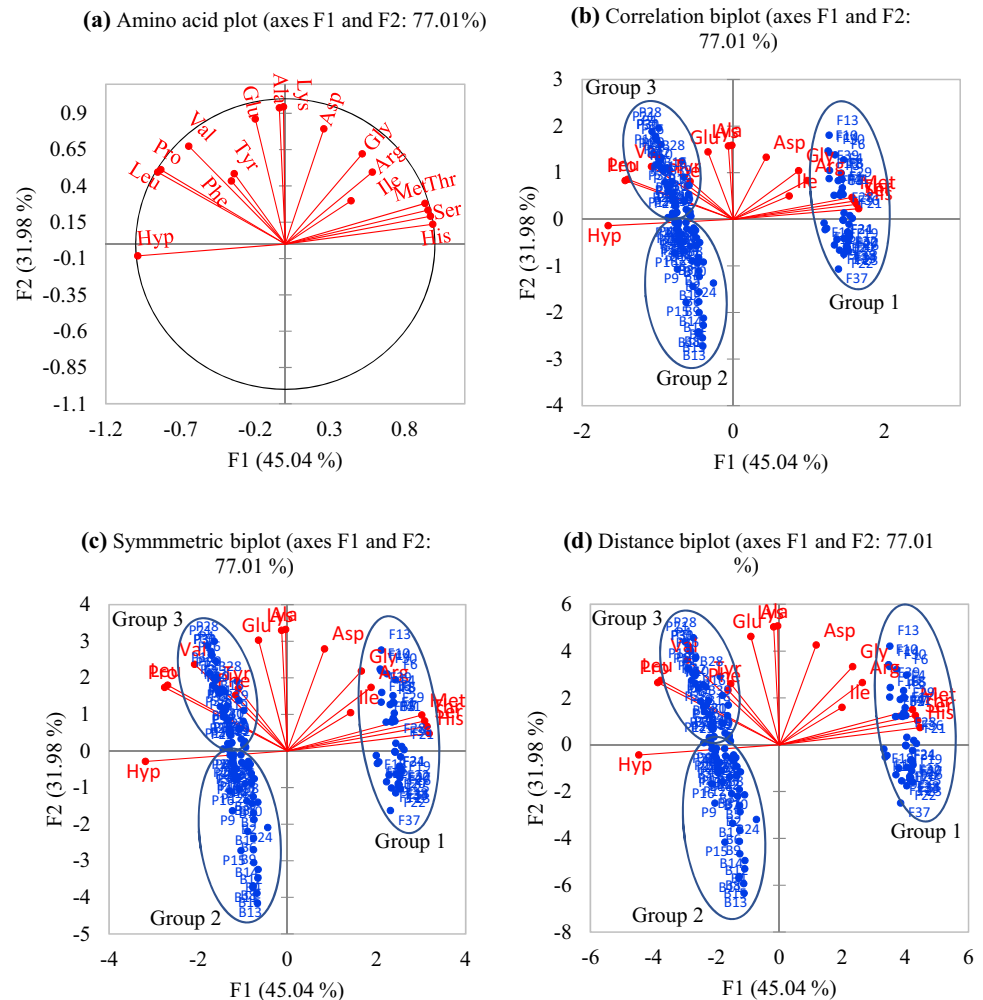
relationship between two AAs. The AAs with positively correlated FL indicated that positive change of AA would render positive change towards other AAs, while negatively correlated FL exhibited inverse change among them. The AAs with the same sign were positively correlated and vice versa. Based on this principle, the Hyp, Pro and Leu were positively correlated, and had negative correlations against His, Ser, Thr and Met which were aligned with Azilawati et al. [10] except our study (1) did not identify the positive correlation of Ile and Val and (2) provided with His as additional AA. For PC2, the Asp, Glu, Ala and Lys were positively correlated and was slightly different from Azilawati et al. [10], which the Ala and Lys were absent in the latter's result.

Figure 3a shows the AAs with weak or without correlations due to the directions of their FL were ~ 90° [53]; Met, Thr, Ser and His against Glu, Ala and Lys; Hyp against Glu, Ala and Lys; and Val, Pro and Leu against Asp, Gly and Arg. These results supported the CM in Table 6

but were not discussed in Azilawati et al. [10], and not harmonized with the result of Widyaninggar et al. [18]. The instrumentation set-up, low sample numbers or deficiency of dataset pre-processing before PCA might have caused these differences.

3.8 Apportionment of amino acids in gelatine groupings by principal component analysis

Biplot in Fig. 3 showed the AAs that significantly contributed to the gelatine sources. This study compared the correlation (CB), symmetric (SB) and distance biplots (DB) to determine which biplot had the best visual display of the AAs and gelatine sources. All biplots showed the same pattern of AA and gelatine distributions, whereby the strong FL and correlations in the biplots supported the FL results in Table 6. Generally, each biplot exhibited three-gelatine groupings, and the biplots were not rendered visual difference. However, the CB in Fig. 3b showed the gelatine

proximity in each grouping, while SB showed a little space among the gelatines in each grouping in Fig. 3c. The DB in Fig. 3d depicted the most significant space among the gelatines in each grouping than in the SB and CB, although the three-gelatine groupings in the DB remained separated. Hence, we proposed that the CB was the best visual display of biplot to explore the gelatine dataset. This is because the gelatines within each grouping were closer to each other than the gelatines in SB and DB. In the CB, 100% fish, 65% bovine and 71% porcine gelatines dominated the groupings of 1, 2 and 3, respectively; hence, we may conclude that the fish, bovine and porcine gelatines represented grouping 1, 2 and 3, respectively (Fig. 3b).

The Fig. 3b showed which AA significantly contributed ($p < 0.05$) to the gelatine grouping by examining which variable fall into the observation grouping [54]: (1) high concentration of Ser, Arg, Gly, Thr, Met, His and Ile and low concentration of Hyp for fish gelatine; (2) high concentration of Hyp and low concentration of Met, Thr, Ser, His, Gly, Arg and Ile for bovine gelatine; and (3) high concentration of Pro, Tyr, Val, Leu, and Phe for porcine gelatine. The Ser, Thr and Met, which have been associated with fish gelatine were in line with Azilawati et al. [10]. Nonetheless, the AAs-correlated bovine and porcine gelatines contradicted Azilawati et al. [10].

In Fig. 3b, the Glu, Ala, Lys, Asp did not contribute to any gelatine grouping while Ile had moderate FL; thus our study had removed these AAs from the dataset and performed another PCA to assess: (1) the FL and AAs correlation, (2) AAs contribution to the gelatine groupings and (3) gelatine type in the gelatine groupings. The AA plot in Fig. 4a showed a higher CV at 90.89% than the CV in Fig. 3a. The FLs were visually improved and had maintained the

strong and directive of FL (Fig. 4a). The apportionment of the AAs to the gelatine grouping was also maintained (Fig. 4b) as follows: (1) high concentration of Ser, Arg, Gly, Thr, Met and His and low concentration of Hyp for fish gelatine; (2) high concentration of Hyp and low concentration of Met, Thr, Ser, His, Gly, and Arg for bovine gelatine; and (3) high concentration of Pro, Tyr, Val, Leu, and Phe for porcine gelatine. The Fig. 4b also exhibited 100% fish, 100% bovine and 100% porcine made the grouping 1, 2 and 3, respectively, which had improved the gelatine grouping as compared to CB in Fig. 3b. These results proved that AAs with strong FL had improved the CB and the gelatine grouping. Hence, our study had selected the Hyp, His, Ser, Arg, Gly, Thr, Pro, Tyr, Met, Val, Leu and Phe as the biomarkers for identification of the gelatine source.

The AAs that significantly contributed ($p < 0.05$) to the fish gelatine supported our ANOVA (Table 2) except the CB rendered His and Hyp as additional AAs. For porcine gelatine, the CB (Fig. 4b) identified all significant AAs ($p < 0.05$) calculated from ANOVA (Table 2) except Hyp. The PCA identified Hyp, Met, Thr, Ser, His, Gly, and Arg as the biomarkers ($p < 0.05$) to the bovine gelatine, which ANOVA failed to identify. From these results, it was evident that PCA was a useful tool to analyse a multivariate dataset that could provide an in-depth understanding of AA distributions as compared to ANOVA and correlation test.

4 Conclusion

The demand to produce gelatines for food, pharmaceutical and cosmetic products has shown an exponential growth which requires gelatine authentication as a
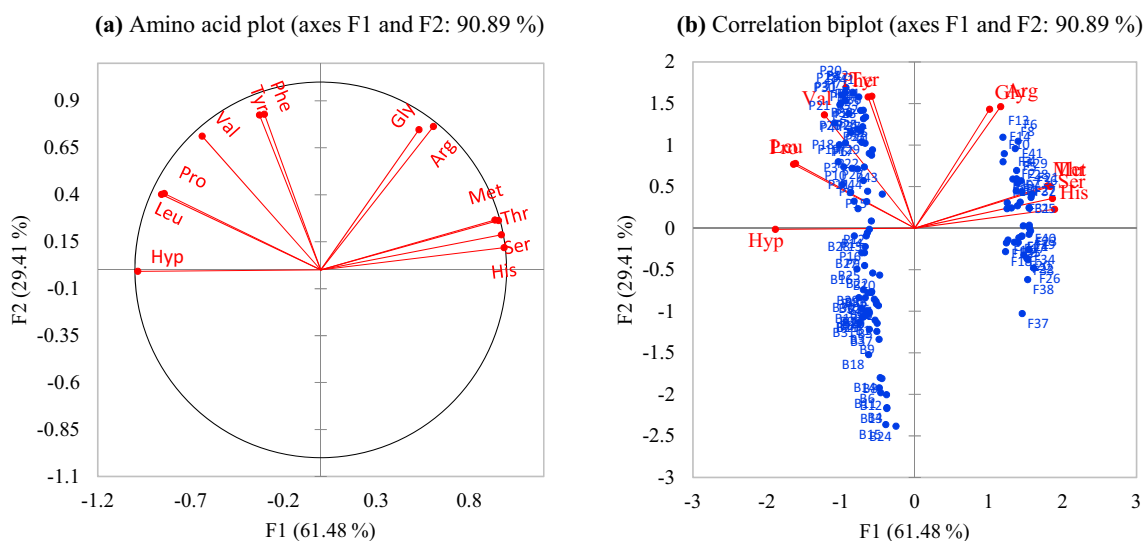

(a) Amino acid plot (axes F1 and F2: 90.89 %)   **(b)** Correlation biplot (axes F1 and F2: 90.89 %)

Fig. 4 a Amino acid plot and **b** correlation biplot of strong factor loading amino acids on fish, bovine and porcine gelatines

forensic food tool by the testing laboratory. The AA testing analysis by UHPLC-DAD incorporated with PCA is imperative to address the false claim and food integrity of gelatines by the manufactures. To explore the gelatine dataset, this manuscript only explained about the PCA as one of frequently utilized multivariate data analysis (MDA) for authentication while other MDA such as multiple linear regression, discriminant analysis and partial least square were not discussed. The analysis undergoes a step-by-step procedure of gelatine hydrolysis, AA derivatization, dataset pre-processing and PCA to avoid false-positive and false-negative results. Since only limited numbers of testing laboratories provide the AA testing, this manuscript may guide the testing laboratories to extend their scope of analysis and suggested the application of UHPLC-DAD-incorporated PCA as an alternative for the PCR method. The certification or regulatory bodies in the governmental level and testing laboratories could adopt this step-by-step procedure to develop a standard of AA analysis and achieve quality testing to authenticate gelatine products.

Compliance with ethical standards

Conflict of interest We declare no conflict of interest.

Research involving human participants and/or animals We declare no human and/or animals involved in this study.

Informed consent All authors have granted permission in full knowledge for the publication of this study.

References

1. Jannat B, Ghorbani K, Kouchaki S, Sadeghi N, Eslamifarsani E, Rabbani F et al (2020) Distinguishing tissue origin of bovine gelatin in processed products using LC/MS technique in combination with chemometrics tools. Food Chem 319:126302. https://doi.org/10.1016/J.Foodchem.2020.126302

2. Yap BK, Gam LH (2019) Differentiation of bovine from porcine gelatin capsules using gel electrophoresis method. Food Chem 274:16–19. https://doi.org/10.1016/J.Foodchem.2018.08.111

3. Gelatine Manufacturers of Europe. Gelatine manufacturing. Gelatine Manufacturers of Europe. 2020:1–4. https://www.gelatine.org/gelatine/manufacturing.html.Accessed April 4 2020

4. Lv LC, Huang QY, Ding W, Xiao XH, Zhang HY, Xiong LX (2019) Fish gelatin: the novel potential applications. J Funct Foods. https://doi.org/10.1016/j.jff.2019.103581

5. Liu Z, Rochfort S (2013) A fast liquid chromatography-mass spectrometry (LC-MS) method for quantification of major polar metabolites in plants. J Chromatogr B 912:8–15. https://doi.org/10.1016/J.Jchromb.2012.10.040

6. Mutalib SA, Muin NM, Abdullah A, Hassan O, Wan Mustapha WA, Abdullah Sani N et al (2015) sensitivity of polymerase chain reaction (PCR)-southern hybridization and conventional PCR analysis for Halal authentication of gelatin capsules. LWT: Food Sci Technol 63:714–719. https://doi.org/10.1016/J.Lwt.2015.03.006

7. Romi Mukherjee S (2014) Global Halal: meat, money, and religion. Religions. 5:22–75. https://doi.org/10.3390/Rel5010022

8. Department of Standards Malaysia. Malaysian Standard MS 1500: 2019 Halal food: General requirements (Third revision). 2019

9. Soon JM, Chandia M, Regenstein JM (2017) Halal integrity in the food supply chain. British Food Journal. 119:39–51. https://doi.org/10.1108/bfj-04-2016-0150

10. Azilawati MI, Hashim DM, Jamilah B, Amin I (2015) RP-HPLC method using 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate incorporated with normalization technique in principal component analysis to differentiate the bovine, porcine and fish gelatins. Food Chem 172:368–376. https://doi.org/10.1016/J.Foodchem.2014.09.093

11. Abbas O, Zadravec M, Baeten V, Mikuš T, Lešić T, Vulić A et al (2018) Analytical methods used for the authentication of food of animal origin. Food Chem 246:6–17. https://doi.org/10.1016/J.Foodchem.2017.11.007

12. Hafidz R, Yaakob C (2011) Chemical and functional properties of bovine and porcine skin gelatin. Int Food Res J 18:813–817

13. Hashim DM, Man YBC, Norakasha R, Shuhaimi M, Salmah Y, Syahariza ZA (2010) Potential use of Fourier transform infrared spectroscopy for differentiation of bovine and porcine gelatins. Food Chem 118:856–860. https://doi.org/10.1016/J.Foodchem.2009.05.049

14. Hameed AM, Asiyanbi-HT Idris M, Fadzillah N, Mirghani MES (2018) A review of gelatin source authentication methods. Trop Life Sci Res 29:213–227. https://doi.org/10.21315/tlsr2018.29.2.15

15. Bower JA. Descriptive statistics. In: Bower JA, (ed) Stat. Methods Food Sci. Introd. Proced. Food Pract. Second Edi, John Wiley & Sons, Ltd.; 2013, pp 44–90. https://doi.org/10.1007/978-3-642-25697-4_4

16. Williams B, Brown T (2012) Exploratory factor analysis : a five-step guide for novices. J Emerg Prim Health Care

17. Granato D, de Araújo Calado VÕM, Jarvis B (2014) Observations on the use of statistical methods in Food Science and Technology. Food Res Int 55:137–149. https://doi.org/10.1016/J.Foodres.2013.10.024

18. Widyaninggar A, Triwahyudi Triyana K, Rohman A (2012) Differentiation between porcine and bovine gelatin in commercial capsule shells based on amino acid profiles and principal component analysis. Indon J Pharm 23:104–1099

19. Azilawati MI, Hashim DM, Jamilah B, Amin I (2014) Validation of a reverse-phase high-performance liquid chromatography

method for the determination of amino acids in gelatins by application of 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate reagent. J Chromatogr A 1353:49–56. https://doi.org/10.1016/J.Chroma.2014.04.050

20. Saiful M, Azid A, Iskandar S, Shirwan M, Sani A, Lananan F (2019) Comparison of prediction model using spatial discriminant analysis for marine water quality index in mangrove estuarine zones. Mar Pollut Bull 141:472–481. https://doi.org/10.1016/J.Marpolbul.2019.02.045

21. Kaiser HF (1974) An index of factorial simplicity. Psychometrika 39:31–36

22. Van Den Berg RA, Hoefsloot HCJ, Westerhuis JA, Smilde AK, Van Der Werf MJ (2006) Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genom 7:1–15. https://doi.org/10.1186/1471-2164-7-142

23. Retnam A, Zakaria MP, Juahir H, Aris AZ, Zali MA, Kasim MF (2013) Chemometric techniques in distribution, characterisation and source apportionment of polycyclic aromatic hydrocarbons (PAHS) in aquaculture sediments in Malaysia. Mar Pollut Bull 69:55–66. https://doi.org/10.1016/J.Marpolbul.2013.01.009

24. Eurachem. Eurachem guide: The fitness for purpose of analytical methods:a laboratory guide to method validation and related topics. 2014. DOI:978-91-87461-59-0

25. Kittiphattanabawon P, Benjakul S, Sinthusamran S, Kishimura H (2016) Gelatin from clown featherback skin: extraction conditions. LWT: Food Sci Technol 66:186–192. https://doi.org/10.1016/J.Lwt.2015.10.029

26. Jamilah B, Tan KW, Umi Hartina MR, Azizah A (2011) Gelatins from three cultured freshwater fish skins obtained by liming process. Food Hydrocolloids 25:1256–1260. https://doi.org/10.1016/J.Foodhyd.2010.11.023

27. Irwandi J, Faridayanti S, Mohamed ESM, Hamzah MS, Torla HH, Che Man YB (2009) Extraction and characterization of gelatin from different marine fish species in Malaysia. Int Food Res J 16:381–389

28. Chen H, Shi P, Fan F, Chen H, Wu C, Xu X, et al. Hofmeister effect-assisted one step fabrication of fish gelatin hydrogels. vol 121. Elsevier Ltd; 2020. https://doi.org/10.1016/j.lwt.2019.108973

29. Ali AMM, Kishimura H, Benjakul S (2018) Physicochemical and molecular properties of gelatin from skin of golden carp (Probarbus Jullieni) as influenced by acid pretreatment and prior-ultrasonication. Food Hydrocolloids 82:164–172. https://doi.org/10.1016/J.Foodhyd.2018.03.052

30. Noorzai S, Verbeek CJR, Lay MC, Swan J (2019) Collagen extraction from various waste bovine hide sources. Waste Biomass Valorizat. https://doi.org/10.1007/S12649-019-00843-2

31. Gómez-Guillén MC, Giménez B, López-Caballero ME, Montero MP (2011) Functional and bioactive properties of collagen and gelatin from alternative sources: a review. Food Hydrocolloids 25:1813–1827. https://doi.org/10.1016/J.Foodhyd.2011.02.007

32. Huang R, Zeng J, Zhao D, Cook KV, Hambright KD, Yu Z (2020) Sediment microbiomes associated with the rhizosphere of emergent macrophytes in a shallow, subtropical lake. Limnol Oceanogr 65:S38–S48. https://doi.org/10.1002/Lno.11325

33. Begdache L, Kianmehr H, Sabounchi N, Marszalek A, Dolma N (2019) Principal component regression of academic performance, substance use and sleep quality in relation to risk of anxiety and depression in young adults. Trends Neurosci Edu 15:29–37. https://doi.org/10.1016/J.Tine.2019.03.002

34. Yener D (2015) Factors that affect the attitudes of consumers toward halal-certified products in Turkey. J Food Prod Market 21:160–178. https://doi.org/10.1080/10454446.2013.843483

35. Kuuliala L, Abatih E, Ioannidis AG, Vanderroost M, De Meulenaer B, Ragaert P et al (2017) Multivariate statistical analysis for the identification of potential seafood spoilage indicators. Food Control 84:49–60. https://doi.org/10.1016/J.Foodcont.2017.07.018

36. Zhang Y, Guo F, Meng W, Wang X-Q (2009) Water quality assessment and source identification of Daliao River Basin using multivariate statistical methods. Environ Monit Assess 152:105–121. https://doi.org/10.1007/S10661-008-0300-Z

37. Granato D, Cardoso C, Oliveira D, Solange M, Caruso F, Araújo L et al (2014) Feasibility of different chemometric techniques to differentiate commercial Brazilian sugarcane spirits based on chemical markers. FRIN. 60:212–217. https://doi.org/10.1016/J.Foodres.2013.09.044

38. Juahir H, Zain SM, Yusoff MK, Hanidza TIT, Armia SM, Toriman ME et al (2011) Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. Environ Monit Assess 173:625–641. https://doi.org/10.1007/s10661-010-1411-x

39. Gogna N, Hamid N, Dorai K (2015) Metabolomic profiling of the phytomedicinal constituents of *Carica papaya* L leaves and seeds by 1H NMR spectroscopy and multivariate statistical analysis. J Pharm Biomed Anal 115:74–85. https://doi.org/10.1016/j.jpba.2015.06.035

40. Rodriguez A (2020) Know your data: evaluation of normality for continuous variables and tools to deal with non-normal distributions. Medicina Clínica. 154:142–145. https://doi.org/10.1016/J.Medcle.2019.05.011

41. Razali NM, Wah YB, Sciences M (2011) Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. J Statist Model Anal 2:21–33

42. Bower JA (2013) The nature of data and their collection. In: Bower JA (ed) Stat. Methods Food Sci. Introd. Proced. Food Pract. Second Edi, John Wiley & Sons, Ltd. Published, pp 15–43. https://doi.org/10.1002/9781118541593.ch2

43. Sainani KL (2012) Dealing with non-normal data. Am Acad Phys Med Rehabilit 4:1001–1005. https://doi.org/10.1016/J.Pmrj.2012.10.013

44. Kwak SG, Kim JH (2017) Central limit theorem: the cornerstone of modern statistics. Korean J A Anesthesiol 70:144–156. https://doi.org/10.4097/Kjae.2017.70.2.144

45. Miri M, Rostami-Aghdam-Shendi M, Ghaffari HR, Ebrahimi-Aval H, Ahmadi E, Taban E et al (2016) Investigation of outdoor BTEX: concentration, variations, sources, spatial distribution, and risk assessment. Chemosphere 163:601–609. https://doi.org/10.1016/j.chemosphere.2016.07.088

46. Ratner B (2009) The correlation coefficient: its values range between +1/− 1, or do they? J Target Measure Anal Mark 17:139–142. https://doi.org/10.1057/Jt.2009.5

47. Vaitla B, Coates J, Glaeser L, Hillbruner C, Biswal P, Maxwell D (2017) The measurement of household food security: correlation and latent variable analysis of alternative indicators in a large multi-country dataset. Food Policy 68:193–205. https://doi.org/10.1016/J.Foodpol.2017.02.006

48. Kim J-K, Kim E-H, Lee O-K, Park S-Y, Lee B, Kim S-H et al (2013) Variation and correlation analysis of phenolic compounds in mungbean (*Vigna radiata* L.) varieties. Food Chem 141:2988–2997. https://doi.org/10.1016/j.foodchem.2013.05.060

49. Priya Darsini DT, Maheshu V, Vishnupriya M, Nishaa S, Sasikumar JM (2013) Antioxidant potential and amino acid analysis of underutilized tropical fruit *Limonia acidissima* L. Free Radic Antioxid 3:1–8. https://doi.org/10.1016/J.Fra.2013.08.001

50. Falcó I, Randazzo W, Sánchez G, López-Rubio A, Fabra MJ (2019) On the use of carrageenan matrices for the development of antiviral edible coatings of interest in berries. Food Hydrocolloids 92:74–85. https://doi.org/10.1016/J.Foodhyd.2019.01.039

51. Brereton RG (2003) Chemometrics: data analysis for the laboratory and chemical plant

52. Cooper GM, Hausman RE (2007) The cell: a molecular approach. ASM Press, Washington. https://doi.org/10.5860/choice.35-0264

53. Hair JF, Black WC, Babin BJ, Anderson RE (2014) Multivariate data analysis. https://doi.org/10.1007/978-3-319-01517-0_3

54. Gelman A, Rabe-Hesketh S, Long JS, Skrondal A (2008) Analysis of multivariate social science data. 2008

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.