Journal Pre-proof

Detection of adulteration activities in edible bird's nest using untargeted ^{1}H-NMR metabolomics with chemometrics

Chin-Hong Yong, Syahidah Akmal Muhammad, Fatimatuzzahra' Abd Aziz, Jing-Sheng Ng, Fatin Ilyani Nasir, M.N.H. Adenan, S. Moosa, Z. Othman, S.N.A. Abdullah, Z. Sharif, F. Ismail, Simon D. Kelly, Andrew Cannavan, Eng-Keng Seow

PII: S0956-7135(21)00680-0

DOI: https://doi.org/10.1016/j.foodcont.2021.108542

Reference: JFCO 108542

To appear in: *Food Control*

Received Date: 30 April 2021

Revised Date: 25 August 2021

Accepted Date: 5 September 2021

Please cite this article as: Yong C.-H., Muhammad S.A., Aziz Fatimatuzzahra'.Abd., Ng J.-S., Nasir F.I., Adenan M.N.H., Moosa S., Othman Z., Abdullah S.N.A., Sharif Z., Ismail F., Kelly S.D., Cannavan A. & Seow E.-K., Detection of adulteration activities in edible bird's nest using untargeted ^{1}H-NMR metabolomics with chemometrics, *Food Control* (2021), doi: https://doi.org/10.1016/j.foodcont.2021.108542.

Credit Author Statement

Chin-Hong Yong: Formal analysis, Investigation, Data Curation, Writing - Original Draft, Visualization

Syahidah Akmal Muhammad: Conceptualization, Methodology, Validation, Writing - Review & Editing, Supervision, Project administration, Funding acquisition

Fatimatuzzahra' Abd Aziz: Methodology, Software, Validation, Formal analysis, Writing - Review & Editing

Jing-Sheng Ng: Writing - Review & Editing

Fatin Ilyani Nasir: Data Curation

M.N.H. Adenan: Resources

S. Moosa: Resources

Z. Othman: Resources

S.N.A. Abdullah: Resources

Z. Sharif: Resources

F. Ismail: Resources

Simon D. Kelly: Writing - Review & Editing

Andrew Cannavan: Writing - Review & Editing

Eng-Keng Seow: Formal analysis, Writing - Review & Editing, Visualization

Detection of adulteration activities in edible bird's nest using untargeted ^{1}H-NMR metabolomics with chemometrics

Chin-Hong Yong[a], Syahidah Akmal Muhammad[a,b,*], Fatimatuzzahra' Abd Aziz[c], Jing-Sheng Ng[a], Fatin Ilyani Nasir[b], M.N.H. Adenan[d], S. Moosa[d], Z. Othman[d], S.N.A. Abdullah[d], Z. Sharif[e], F. Ismail[f], Simon D. Kelly[g], Andrew Cannavan[g], Eng-Keng Seow[h]

[a] Environmental Technology Division, School of Industrial Technology, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia
[b] Analytical Biochemistry Research Centre, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia
[c] Discipline of Clinical Pharmacy, School of Pharmaceutical Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia
[d] Malaysian Nuclear Agency, Bangi, 43000 Kajang, Selangor, Malaysia
[e] Surveillance Branch, Food Safety and Quality Division, Ministry of Health Malaysia, Presint 3, Federal Government Administrative Centre, Putrajaya, Malaysia
[f] Veterinary Public Health Laboratory, Department of Veterinary Services, Bandar Baru Salak Tinggi, 43900 Sepang, Selangor, Malaysia
[g] Food and Environmental Protection Laboratory, Joint FAO/IAEA Centre of Nuclear Techniques in Food and Agriculture, Department of Nuclear Sciences and Applications, International Atomic Energy Agency, Vienna International Centre, PO Box 100, 1400, Vienna, Austria
[h] Department of Food Science and Technology, Faculty of Applied Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
*Corresponding author: syahidah.muhammad@usm.my

Abstract

Edible bird's nests (EBNs) are the nests of swiftlets, made from the saliva of the male swiftlet (*Aerodramus fuciphagus*). Due to their nutritional value, EBNs are recognized as a premium food and highly in demand among the Chinese community. EBNs are commonly adulterated with cheaper ingredients and efforts are being made to combat these activities using different analytical techniques. In this study, nuclear magnetic resonance (^{1}H-NMR) metabolomic fingerprinting combined with chemometrics, particularly principal component analysis (PCA) and orthogonal partial least square-discriminant analysis (OPLS-DA), was employed to detect adulteration in EBNs. Authentic EBN samples from different locations in Malaysia were used and adulteration was simulated using nutrient agar, collagen, gelatine, karaya gum and melamine at 1, 5 and 10% w/w, respectively. Overall, unsupervised PCA was able to distinguish authentic EBNs from those adulterated with nutrient agar, collagen and gelatine down to 5% w/w adulteration level. As for EBN adulterated with karaya gum and melamine, a distinct peak can be observed at 1.91ppm and 6.10ppm, respectively. The supervised OPLS-DA predictive model was able to differentiate authentic EBNs from simulated adulterated EBNs with 100% accuracy. Conclusively, ^{1}H-NMR metabolomics combined with chemometrics could be a potential tool for the detection of adulteration in EBN.

Keywords: Edible bird's nest; adulteration; metabolomics; chemometrics; NMR

1. Introduction

Edible bird's nests (EBNs) are made of saliva produced by swiftlets (*Aerodramus* or *Collocalia* species). EBN is usually prepared for consumption as a delicacy known as 'bird's nest soup' (Kinnaird, 2003). It has been described as one of the premium foods among the Chinese community and is often referred as the 'Caviar of the East' (Kinnaird, 2003; Marcone, 2005). This is because EBN is believed to have anti-ageing properties and has been proven to be beneficial to health, showing potential for antiviral activity against influenza virus (Deng et al., 2006; Haghani et al., 2016, 2017; Norhayati et al., 2010). Thus, EBN can cost up to US$ 6000 for a kilogram in China (Lee et al., 2017). Being an expensive food product and having health benefits, EBN is commonly adulterated with cheaper foodstuffs for economic gain by fraudsters. Food products such as karaya gum, pork skin, *Tremella* fungus, agar and other cheaper items are among the common adulterants used to imitate the appearance of EBN but sold at the authentic retail price (Adenan et al., 2020; Lau et al., 1994; Lee et al., 2017; Marni et al., 2012; Tung et al., 2008).

Efforts have been made to ensure the authenticity of EBN using a number of analytical techniques, but each method has its own advantages and disadvantages. Spectroscopy is a good, precise screening method, but is mainly qualitative rather than quantitative, whereas hyphenated chromatographic analysis requires tedious sample preparation procedures, although it has highly sensitive qualitative and quantitative attributes. Concerted application of two or more major analytical methods such as combined DNA-based PCR and protein-based, gel electrophoresis and liquid chromatography two-dimensional gel electrophoresis (2D-GE) method as well as gas chromatography mass spectrometry and liquid chromatography mass spectrometry has been suggested for more accurate and promising results, yet this again requires tedious experimentation procedures and sample preparation (Lee et al., 2017). Likewise, metabolomics using chromatography technique with mass spectrometry has been used for metabolite profiling of edible bird's nest (Chua et al., 2014; Lee et al., 2017). Metabolomics is a scientific field which encompasses the comprehensive and simultaneous systematic profiling of multiple metabolite concentrations and their cellular and systemic fluctuations in response to external factors. Metabolomic analysis can be generally classified into targeted or untargeted methods. Targeted metabolomics measures a specified list of metabolites, focusing on one or more related pathways of interest, while untargeted metabolomics simultaneously measures as many metabolites as possible from biological samples, without bias (Alonso et al., 2015; Patti

et al., 2012; Roessner & Bowne, 2009). Metabolomics is an emerging approach for food analysis and authentication, but metabolomics using nuclear magnetic resonance (NMR) has not been used extensively to detect adulteration in EBN (Cubero-Leon et al., 2014; Wishart, 2008).

Metabolomics using proton nuclear magnetic resonance (^{1}H-NMR) is known for the minimal effort involved in sample preparation and its ability to measure as many metabolites as possible without tempering the metabolites present in the sample matrix in food authentication (Esslinger et al., 2014). Chemometrics is a technique that uses mathematical and statistical methods to interpret chemical data obtained from various analytical instruments. This type of analysis begins with principal component analysis (PCA), a commonly used unsupervised pattern recognition technique, to explore possible hidden patterns and relationships in data, and is followed by orthogonal partial least squares discriminant analysis (OPLS-DA), a supervised technique where classification is given to each sample and the data are projected into this new space, showing the relation between samples and variables (Abdi & Williams, 2010; Granato et al., 2018; Smolinska et al., 2012). Chemometrics has been incorporated into the analysis of both NMR data and mass spectroscopy (MS) data in various metabolomics applications including food authentication and the detection of adulteration (Cavanna et al., 2018; Dona et al., 2016; Erban et al., 2019; Madsen et al., 2010; Riedl et al., 2015; Trygg et al., 2007; Yi et al., 2016). The work done by Di Anibal et al., (2011), Spiteri et al. (2015), Jandrić et al., (2017) and Righetti et al. (2018) has clearly shown the application of chemometrics in metabolomics to determine the authenticity of culinary spices, floral honeys, fruits/fruit juices and wheats, respectively. Combined with chemometrics, the untargeted approach of NMR metabolomics takes into account the external factors that may affect the metabolites, and thus enables the analyst to potentially detect emerging fraudulent practices (Cubero-Leon et al., 2014; Kosmides et al., 2013).

The work done by Adenan et al. (2020) and Seow et al., (2016) has shown the possibility of discriminating the geographical origin of EBN using mid-infrared–attenuated total reflection spectroscopy and chemometric analysis of amino acid composition, respectively. In other words, any metabolites that are affected by external factors can contribute to the discrimination of authentic EBNs based on geographical origin alone. Since the metabolites can be affected by external factors, untargeted ^{1}H-NMR metabolomics with chemometrics was used in this study to first obtain the metabolic fingerprint of authentic

EBNs from different geographical regions, prior to the application of untargeted ^{1}H-NMR metabolomics with chemometrics to detect adulteration of the authentic EBN.

2. Materials and methods

2.1. Sampling

EBN samples from the swiftlet species, *Aerodramus fuciphagus*, were collected from the State Health Department, Ministry of Health, Malaysia; the authenticity of the EBN samples was guaranteed by the Ministry of Health as the samples were collected directly from processing plants registered with the Ministry of Health as part of its monitoring programme to ensure the safety and authenticity of Malaysian raw and cleaned EBNs. A total of 44 authentic EBN samples were collected from different locations across seven states in Malaysia, and the details are shown in Table 1.

2.2. Sample preparation

Cleaning, drying and grinding of all authentic EBN samples were done by officers from the Malaysia Nuclear Agency once they had obtained the raw samples from Malaysia Ministry of Health officials (Adenan et al., 2020). The EBN samples were placed in a 'Tefzel' (ethylene tetrafluoroethylene) container with stainless steel grinding balls (1.4034, hardness approx. 52 HRC) and ground to a fine powder. Other apparatus used in the handling of the EBNs was made of plastic. The ground samples were kept in airtight containers prior to analysis.

2.3. Simulated adulteration

The adulterants used in this study were pure melamine (Mel), karaya gum (KG), nutrient agar (Agar), collagen (Col) and gelatine (Gel). KG, Agar, Col and Gel are among the most commonly used adulterants to adulterate EBN, whereby Col and Gel represent the adulterants derived from animal skin such as those from porcine, bovine or fish (Adenan et al., 2020; Lee et al., 2017). Mel was also included as a potential adulterant used since there have been reports that Mel has been used as an apparent 'protein enhancer' (Adenan et al., 2020). Mel and KG were purchased from Sigma, USA, and Gel was purchased from

Lafleureshop.com, USA. Agar was obtained from Oxoid, UK, while Col was obtained from Lennox, Japan. Admixtures were prepared gravimetrically at the concentrations of 1, 5 and 10% w/w by combining appropriate quantities of adulterants with authentic EBN samples, as shown in Table 2. The resultant mixtures were thoroughly mixed by a hand action shaker for 30 min.

2.4. Buffered solution preparation

A 100 mL buffered solution (pH 7) containing 1mM of 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) internal standard was prepared as follows. Approximately 2.84 g of di-sodium hydrogen phosphate (Fluka, Switzerland) and 0.48 g of sodium dihydrogen phosphate (R&M Chemicals, UK) were dissolved in 50 mL of distilled water. In another vial, approximately 0.02154 g of DSS (Armar Chemicals, Switzerland) was dissolved in 20 mL of deuterated water (D_2O) (Armar Chemicals, Germany) before mixing it with the phosphate-buffered solution. The solution was then filled with distilled water up to 100 mL and kept for later usage.

2.5. Metabolite extraction

Approximately 35 mg of sample was dissolved in 1 mL of distilled water in a 1.5 mL microcentrifuge tube. The mixture was sonicated at 60 °C for 30 min. The mixture was cooled to room temperature and centrifuged at 16,000 rcf for 20 min. The supernatant was filtered through a 13 mm × 0.45 mm polytetrafluoroethylene (PTFE) syringe filter. Then, 400 µL of the filtrate was dissolved in 200 µL of buffered solution containing internal standard; and lastly, 550 µL of the solution was transferred into an NMR tube for analysis.

2.6. NMR spectroscopy

^{1}H-NMR metabolomic profiles of authentic EBNs, adulterated EBNs and adulterants were recorded on a Bruker 700 MHz ASCEND™ spectrometer (Bruker BioSpin, Germany), operating at 700.23 MHz and equipped with a cryoprobe, located at the Analytical Biochemistry Research Centre (ABrC), Universiti Sains Malaysia. The one-dimensional ^{1}H-nuclear Overhauser effect spectroscopy (NOESY) experiment method was used for water

suppression. The NMR experiment was performed using 128 scans with 4 prior dummy scans at a target temperature of 302.4 K. D_2O was used as an internal lock signal. Acquisition was done with a spectral width of 15 ppm with relaxation delay of 4 s, receiver gain of 28.5 and transmitter frequency offset at 3291.24 Hz. The data were acquired automatically under the control of ICON-NMR (Bruker BioSpin, Germany). Chemical shifts were calibrated with respect to DSS at 0.00 ppm.

2.7. Spectral pre-processing

All NMR spectra were manually phased and baseline-corrected using TOPSPIN 3.6.1 (Bruker BioSpin, Germany). AMIX Statistic Version 3.9.15 was used for the bucketing and normalization processes. Each NMR spectrum was bucketed by integrating regions having an equal bucket width of 0.04 ppm over a range of δ 0.0 to 15.0 ppm, with 0.02 to −0.02 ppm as the reference region. The spectrum was then converted to Microsoft Excel format (*.xls) for chemometrics.

2.8. Chemometrics

Chemometrics was performed to differentiate the adulterated EBNs from the authentic EBNs and to distinguish the geographical region of the EBNs. Prior to analysis, the water resonance between 4.28 and 5.88 ppm and the empty region between 8.52 and 15.00 ppm were eliminated to prevent these regions from contributing to the similarities between authentic and adulterated EBNs. The DSS internal standard peaks at 0.63 ppm (m), 1.75 ppm (m) and 2.91 ppm (m) were also excluded. All buckets were normalized using log 10, and Pareto scaling was used to build the soft independent modelling by class analogy (SIMCA) model. The unsupervised pattern recognition method using PCA, and the supervised method using (OPLS-DA) were performed using SIMCA® software, version 14.1 (MKS Umetrics AB, Umeå, Sweden).

PCA was first used to explore the possible patterns by extracting important information through dimension reduction of the dataset into a small number of factors, namely principal components (PCs), before projecting the variations present in the dataset (Abdi & Williams, 2010; Granato et al., 2018; Smolinska et al., 2012). The iteratively calculated PCs hold as much variation from the original data set as possible; the first PC

(denoted as PC1) explains the most data variation followed by the second PC (PC2), the third (PC3) and so on. In each subsequent PC, some part of the data was then taken out to perform cross-validation with a different number of PCs on the rest of the data. Dimension reduction of the dataset was repeated in PC2 to produce the third PC (PC3), and the calculation was repeated until all data had been reduced (Granato et al., 2018). Since PCA only projects or displays the dataset under investigation, it does not create a 'mathematical model' for classification and authentication purposes and thus the grouping of samples needs to be identified by the user (Granato et al., 2018).

The supervised method, OPLS-DA, was used as further analysis to the unsupervised PCA to develop a 'mathematical predictive model' for classification and authentication purposes (Alonso et al., 2015; Cubero-Leon et al., 2014). OPLS-DA factorizes the data variance into components that are correlated with the variable of interest and the other components, which are the uncorrelated components (Alonso et al., 2015). Each sample was given a specific classification in this supervised method and the data were projected into this new space, showing the relation between samples and variables (Smolinska et al., 2012). Specific indicators and information such as predictive ability of model, goodness of fit of data, and others can be obtained from this generated model.

3. Results and discussion

3.1. Authentic EBNs

The untargeted metabolomics method is known for its ability to take into account external factors, such as environmental factors and geographical origin, that may affect the metabolite composition (Cubero-Leon et al., 2014; Kosmides et al., 2013). Since the samples of EBN were from different locations in Malaysia, the ^{1}H-NMR metabolomic fingerprints of the authentic EBN samples were first explored to investigate whether there were any potential markers related to their origins in different geographical regions in Malaysia. Based on the NMR metabolomic fingerprints of the authentic EBN samples, there were differences in peak patterns and intensity depending on the geographical region as shown in Fig. 1. For instance, the peaks at 1.20–1.40 ppm (m), 2.60 ppm (q), 3.80–4.0 ppm (m) and 7.20–7.40 ppm (m) were absent in the samples from Kelantan but were present in the EBNs from other states. The metabolites that contributed at peaks, 1.20–1.40 ppm (m) and 2.60 ppm (q), are of the R-CH_2 or R-NH group whereby the postulated metabolites were alanine, lactic acid

derivatives or aspartic acid (Chua et al., 2014; Seow et al., 2016; Ulrich et al., 2008; Utomo et al., 2018). It is possible that the aromatic amino acid metabolites such as phenylalanine or tyrosine may contributed to the peaks at 7.20–7.40 ppm (m) but the exact metabolite was unable to be identified (Boffo et al., 2012; Ulrich et al., 2008). Based on the work done by Nakagawa et al., (2007), the possible metabolite that contributed to the chemical shift between 3.80 –4.0 ppm (m) was chondroitin proteoglycan or mainly N-Acetyl-D-glucosamine (GalNAc). In addition, the peak intensity at 2.60 ppm (q) and 7.95 ppm (dd) also varied possibly depending on the state where the EBN sample originates. Thus, the ^{1}H-NMR metabolomic fingerprints of authentic EBNs were subjected to PCA to explore possible patterns based on the geographical region in which the samples were collected, in Malaysia.

The PCA of the ^{1}H-NMR metabolic fingerprints of the authentic EBNs was performed to explore possible patterns and trends of clustering. However, from the PCA overview score plot shown in Fig. 2, it can be seen that there was no specific clustering of EBNs based on the geographical regions in Malaysia, as was potentially shown in the work of Seow et al. (2016) by chemometric analysis of amino acids and in the work of Adenan et al. (2020) using mid-infrared–attenuated total reflectance spectroscopy. The PCA score plot had a fitness of data (R^2) of 91.3%, which explains 70.7% of data variation based on two PCs, PC1 and PC2. Although most of the Sarawak EBN samples fell in the negative quadrant based on the first PC (denoted as t[1]), EBN samples from other regions did not form any specific cluster. In other words, it is possible that each EBN used in this study has its own unique ^{1}H-NMR metabolic fingerprint irrespective of the geographical region. Due to the indiscriminate PCA patterns for the determination of geographical origin, further analysis based on geographical origin was not pursued in this study.

3.2. Simulated adulteration

Prior to the simulated adulteration for the determination of EBN authenticity, the ^{1}H-NMR metabolic fingerprints of authentic EBNs were compared to those of possible adulterants used, and PCA of the ^{1}H-NMR metabolic fingerprints of these samples was performed. Since there is no particular EBN to be considered as an absolute standard, all of the authentic EBNs were used in the PCA as shown in Fig. 3. With just two PCs which explain 83.0% of the data variation, all the adulterants except KG are outside Hotelling's T2 ellipse at the 95% confidence interval. The adulterant KG falls closest to one of the EBN

samples from Samarahan and Bentong. However, based on the ^{1}H-NMR metabolomic fingerprints of 100% w/w KG, the fingerprint pattern was different than that of Samarahan and Bentong, as shown in Fig. A.1. Obvious differences include the peaks at the chemical shifts of 1.216–1.281 ppm (m), 1.905 ppm (s), 2.094–2.183 ppm (m) and 3.240–4.280 ppm (m) which may be polysaccharides such as rhamnose, polygalacturonic acid and galactose that made up the karaya gum (Ulrich et al., 2008; Wishart et al., 2018). The peaks could be crucial as potential markers for EBNs adulterated with KG. Hence, KG was used in the simulated adulteration to investigate the possibility of this method in determining the authenticity of EBNs.

To test the limitations of this method, simulated adulteration was conducted. The ^{1}H-NMR metabolomic fingerprints for the simulation were subjected to PCA to explore the possibility of discriminating the authentic EBNs from the adulterated EBNs down to 1% w/w adulteration. The rationale of adulterating the EBNs down to 1% w/w was that if this method was able to detect the adulteration for as low as 1% w/w, then it is very likely that this method can detect the adulteration at a much higher percentage. The PCA score plots show the separation of authentic EBNs from EBNs adulterated with Agar, Col and Gel down to 5% w/w adulteration using PC1 against PC3 as shown in Figures 4A – C. The differences in ^{1}H-NMR metabolomic fingerprints of authentic EBNs and those adulterated with 1% w/w Agar, Col and Gel were too small to be detected by PCA. The total variability explained by PC1 and PC3 in the PCA score plot as shown in Fig. 4 was 62.1% (Fig. 4A), 63.1% (Fig. 4B) and 66.2% (Fig. 4C). Based on the contribution plot of PC1 against PC3 (not shown) for EBNs adulterated with Agar, Col and Gel, the peaks that contributed to the differences in these adulterated EBNs were at the chemical shifts of 0.80–1.10 ppm (m), 1.40–1.56 ppm (m), 1.80–2.00 ppm (m), 2.30–2.38 ppm (m), 2.98–3.18 ppm (m) and 3.80–4.00 ppm (m). These were mainly polypeptides and amino acids such as proline, hydroxyproline, arginine and glutamic acid (Ulrich et al., 2008). The differences in peak patterns can be observed on the ^{1}H-NMR metabolomic fingerprint spectra in Fig. 5. The ^{1}H-NMR metabolomic fingerprint spectra of EBNs adulterated with Agar, Col and Gel at chemical shifts which were determined by the contribution plot were very different compared to those for authentic EBNs. In addition to that, the aromatic region at 6.80–8.40 ppm which showed additional peaks based on ^{1}H-NMR metabolomic fingerprint spectra may also be potential markers to detect EBNs adulterated with Agar, Col and Gel. For example, the high intensity of peaks at chemical shift region of 7.00 – 7.50 ppm (m) was most probably contributed by

phenylalanine and tyrosine whereby Agar, Col and Gel had higher percentage of these amino acids as compared to the EBN (Boffo et al., 2012; Seow et al., 2016; Ulrich et al., 2008). However, PCA failed to differentiate EBNs adulterated with 1% w/w Agar, Col and Gel. This was probably due to the content of the 1% w/w Agar which can be quite similar to that of authentic EBN, with protein constituents of approximately 58.31–63.88% w/w, made up of peptides and different types of amino acid, depending on the origin of the EBN (Adenan et al., 2020; Hamzah et al., 2013; Marcone, 2005). Agar mostly contains peptone, a water-soluble mixture of polypeptides and amino acids (Stephens, 2003). A similar observation can be made for the NMR metabolomic fingerprints of EBNs adulterated with Col and Gel, as both of these adulterants are made of protein composition as well.

On the other hand, PCA failed to differentiate authentic EBNs from those adulterated with KG even using up to three PCs (Fig. 4D). KG is a common adulterant in EBN and is made up of mostly polysaccharides (Adenan et al., 2020; BeMiller, 2019). Since EBNs contain approximately 11.3–26% w/w carbohydrates, including polysaccharides and glycoproteins, detection of EBN adulteration with KG is quite difficult based on the NMR spectrum as well (Adenan et al., 2020; Hamzah et al., 2013; Marcone, 2005; Shim et al., 2016; Tung et al., 2008). However, based on the contribution plot and the NMR spectra in Fig. 5A, there was only a single potential marker which could be used to distinguish authentic EBNs from those adulterated with KG, which is the singlet peak at the chemical shift at 1.91 ppm (s), possibly contributed by rhamnose (Wishart et al., 2018). Since KG is made up of mostly polysaccharides, the percentage of polysaccharide can be high as compared to the polysaccharide content of EBNs, including having a distinct singlet peak at the chemical shift of 1.91 ppm (s) with high intensity as shown in Fig A.1 (Marcone, 2005; Tung et al., 2008).

The PCA score plot failed to differentiate authentic EBNs from those adulterated with Mel as well, even using up to three PCs (Fig. 4E). Mel, the apparent ‘protein enhancer’ which was used as an adulterant, is a cyclic compound containing only N–H, N–C and N–C functional groups (Adenan et al., 2020; Daintith, 2008). Therefore, the only functional group that contains hydrogen is N–H, the only ^{1}H-NMR peak that can be used to differentiate authentic EBNs from those adulterated with Mel. This broad single peak can be observed at the chemical shift at 6.10 ppm (s) as shown in Fig. 5B, thus making it the only potential marker to differentiate authentic EBNs from those adulterated with Mel.

3.3. Predictive model

Since the unsupervised PCA method has demonstrated the possibility of discriminating authentic EBNs from adulterated ones with some limitations, a supervised method, OPLS-DA, was employed to construct the classification model to predict the authenticity of EBNs using ^{1}H-NMR metabolite fingerprints. The datasets of both the authentic and adulterated EBNs were divided into a training set and testing set at a ratio of 80:20. The 80% assigned to the training set were randomly selected from the dataset and put into two classes to build the predictive model. The first class was the authentic EBNs (denoted as Authentic Training Set) and the second class was adulterated EBNs (denoted as Adulterated Training Set). The remaining 20% of the testing set samples served as blind samples to evaluate the robustness of the predictive model via an external validation method. An internal validation method was performed with a total of seven cross-validation groups by default in SIMCA software. The outcome of the predictive model is shown in Fig. 6. The five-pointed red stars represent the authentic EBN training set while the four-pointed blue stars represent the adulterated EBN training set. This predictive model has a fitness of data (R2(X)) of 83.3% with moderate predictive ability (Q2) of 55.4%. The predictive model explained 91.4% of the total the sum of variation (R2(Y)). The P-value of the cross-validated analysis of variation (CV-ANOVA) was 0.01, which implies that the discrimination of authentic EBNs from adulterated EBNs was significant ($p < 0.05$).

To test the predictive ability of this model, the blind samples which consisted of the testing set were used for the external validation method. The blind samples of authentic EBNs were denoted as the Authentic Test Set (grey circles) and adulterated EBNs were denoted as the Adulterated Test Set (yellow squares). The blind samples in this test set were all correctly assigned to their respective classification as shown in Fig. 6 and no misclassification was observed in the misclassification table (Table 3). Although there was no misclassification using this predictive model, the predictive ability was only 55.4%, which could be further improved by increasing the sample size. Based on the Variables Importance in Projection (VIP) plot, important discriminatory markers should have a value higher than 1. Here, 55 out of 146 variable values were higher than 1 according to the VIP plot (not shown). According to the bucket from the VIP plot, the chemical shifts which could be potential markers in the aliphatic regions were 0.88–0.96 ppm, 1.22 ppm, 1.32–2.72 ppm, 2.96–3.04 ppm, 3.16–3.28 ppm, 3.42 ppm, 3.52–3.60 ppm and 3.76–4.00 ppm. In the aromatic region, chemical shifts at 7.24–7.40 ppm and 7.92–8.04 ppm also contributed to the separation of

authentic EBNs and adulterated EBNs and thus are potential markers as well. In short, the metabolites that contributed to the discrimination between authentic EBN and the adulterated EBN were mainly amino acids, with the exception of EBN adulterated with KG and Mel as explained in Section 3.2. The peak patterns of these potential markers to distinguish authentic EBNs from adulterated ones can be observed in the ^{1}H-NMR metabolomic fingerprint spectra shown in Fig. 5.

4. Conclusion

^{1}H-NMR metabolomics combined with chemometrics shows a lot of potential for the detection of adulteration activities in EBN. Although this method did not discriminate authentic EBNs according to geographical origin, it was able to differentiate authentic EBNs from adulterated ones. The PCA score plots were able to differentiate authentic EBNs from those adulterated with Agar, Col and Gel down to 5% w/w adulteration level. However, there were certain limitations as the PCA score plots were not able to differentiate authentic EBNs from those adulterated with KG and Mel. Nevertheless, based on the ^{1}H-NMR metabolomics fingerprint spectra, EBNs adulterated with KG and Mel have a distinct single peak at 1.91 ppm and a broad peak at 6.10 ppm, respectively, which could be potential markers for these types of adulterants. The OPLS-DA predictive model developed in this study showed promising results for distinguishing authentic EBN samples from adulterated EBN samples. The constructed predictive model may have high accuracy although it only has moderate predictive ability. The predictive ability of this model could be further improved by increasing the sample size. Overall, this work showed that the discrimination of authentic from adulterated EBNs using ^{1}H-NMR metabolomics was successful and the technique has promise as a tool for food authenticity work in the future.

Acknowledgements

This work was funded by the International Atomic Energy Agency (IAEA) under the Technical Cooperation Program (MAL5030) and the Coordinated Research Project (CRP) D52040: Contract 22215 (Grant number: USM 304/PTEKIND/650887). The authors would like to thank the Malaysian Nuclear Agency and Ministry of Health, Malaysia for providing the edible bird's nest and adulterants. The first author thanks the Kuok Foundation for the

MSM Fellowship Award provided. Lastly, the authors would also like to acknowledge the Analytical Biochemistry Research Centre, USM for use of the Bruker 700 MHz ASCEND™ nuclear magnetic resonance spectrometer in this study.

Appendices

Fig. A.1. ^{1}H-NMR metabolomic fingerprint spectra of 100% w/w KG, Samarahan and Bentong at 0.00–4.00 ppm.

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *WIREs Computational Statistics*, *2*(4), 433–459. https://doi.org/10.1002/wics.101

Adenan, M. N. H., Moosa, S., Muhammad, S. A., Abrahim, A., Jandric, Z., Islam, M., Rodionova, O., Pomerantsev, A., Perston, B., Cannavan, A., Kelly, S. D., Othman, Z., Abdullah Salim, N. A., Sharif, Z., & Ismail, F. (2020). Screening Malaysian edible bird's nests for structural adulterants and geographical origin using Mid-Infrared – Attenuated Total Reflectance (MIR-ATR) spectroscopy combined with chemometric analysis by Data-Driven – Soft Independent Modelling of Class Ana. *Forensic Chemistry*, *17*, 100197. https://doi.org/10.1016/j.forc.2019.100197

Alonso, A., Marsal, S., & Julià, A. (2015). Analytical methods in untargeted metabolomics: State of the art in 2015. *Frontiers in Bioengineering and Biotechnology*, *3*, 23. https://doi.org/10.3389/fbioe.2015.00023

BeMiller, J. N. (2019). 16 - Gum Arabic and Other Exudate Gums. In J. N. BeMiller (Ed.), *Carbohydrate Chemistry for Food Scientists (3rd Ed)* (pp. 313–321). AACC International Press. https://doi.org/10.1016/B978-0-12-812069-9.00016-9

Boffo, E. F., Tavares, L. A., Tobias, A. C. T., Ferreira, M. M. C., & Ferreira, A. G. (2012). Identification of components of Brazilian honey by 1H NMR and classification of its botanical origin by chemometric methods. *LWT - Food Science and Technology*, *49*(1), 55–63. https://doi.org/https://doi.org/10.1016/j.lwt.2012.04.024

Cavanna, D., Righetti, L., Elliott, C., & Suman, M. (2018). The scientific challenges in moving from targeted to non-targeted mass spectrometric methods for food fraud analysis: A proposed validation workflow to bring about a harmonized approach. *Trends in Food Science & Technology*, *80*, 223–241. https://doi.org/10.1016/j.tifs.2018.08.007

Chua, Y. G., Bloodworth, B. C., Leong, L. P., & Li, S. F. Y. (2014). Metabolite profiling of edible bird's nest using gas chromatography/mass spectrometry and liquid chromatography/mass spectrometry. *Rapid Communications in Mass Spectrometry*, *28*(12), 1387–1400. https://doi.org/doi:10.1002/rcm.6914

Cubero-Leon, E., Peñalver, R., & Maquet, A. (2014). Review on metabolomics for food

authentication. *Food Research International*, *60*, 95–107. https://doi.org/10.1016/j.foodres.2013.11.041

Daintith, J. (2008). *Melamine*. Oxford University Press. https://doi.org/10.1093/acref/9780199204632.013.2644

Deng, Y.-E., Sun, S.-Q., Zhou, Q., & Li, A. (2006). Analysis and discrimination of Collocalia esculenta L. via FTIR spectroscopy. *Guang pu xue yu guang pu fen xi = Guang pu*, *26*(7), 1242–1245. http://europepmc.org/abstract/MED/17020031

Di Anibal, C. V, Ruisánchez, I., & Callao, M. P. (2011). High-resolution 1H nuclear magnetic resonance spectrometry combined with chemometric treatment to identify adulteration of culinary spices with Sudan dyes. *Food Chemistry*, *124*(3), 1139–1145. https://doi.org/10.1016/j.foodchem.2010.07.025

Dona, A. C., Kyriakides, M., Scott, F., Shephard, E. A., Varshavi, D., Veselkov, K., & Everett, J. R. (2016). A guide to the identification of metabolites in NMR-based metabonomics/metabolomics experiments. *Computational and Structural Biotechnology Journal*, *14*, 135–153. https://doi.org/10.1016/j.csbj.2016.02.005

Erban, A., Fehrle, I., Martinez-Seidel, F., Brigante, F., Más, A. L., Baroni, V., Wunderlin, D., & Kopka, J. (2019). Discovery of food identity markers by metabolomics and machine learning technology. *Scientific Reports*, *9*(1), 9697. https://doi.org/10.1038/s41598-019-46113-y

Esslinger, S., Riedl, J., & Fauhl-Hassek, C. (2014). Potential and limitations of non-targeted fingerprinting for authentication of food in official control. *Food Research International*, *60*, 189–204. https://doi.org/10.1016/j.foodres.2013.10.015

Granato, D., Santos, J. S., Escher, G. B., Ferreira, B. L., & Maggio, R. M. (2018). Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective. *Trends in Food Science & Technology*, *72*, 83–90. https://doi.org/10.1016/j.tifs.2017.12.006

Haghani, A., Mehrbod, P., Safi, N., Aminuddin, N. A., Bahadoran, A., Omar, A. R., & Ideris, A. (2016). In vitro and in vivo mechanism of immunomodulatory and antiviral activity of Edible Bird's Nest (EBN) against influenza A virus (IAV) infection. *Journal of Ethnopharmacology*, *185*, 327–340. https://doi.org/10.1016/j.jep.2016.03.020

Haghani, A., Mehrbod, P., Safi, N., Kadir, F. A. A., Omar, A. R., & Ideris, A. (2017). Edible bird's nest modulate intracellular molecular pathways of influenza A virus infected cells. *BMC Complementary and Alternative Medicine*, *17*, 22. https://doi.org/10.1186/s12906-016-1498-x

Hamzah, Z., Jeyaraman, S., Ibrahim, N., Hashim, O., Lee, B. B., & Hussin, K. (2013). A rapid technique to determine purity of edible bird nest. *Advances in Environmental Biology*, *7*, 3758–3765.

Jandrić, Z., Islam, M., Singh, D. K., & Cannavan, A. (2017). Authentication of Indian citrus fruit/fruit juices by untargeted and targeted metabolomics. *Food Control*, *72*, 181–188. https://doi.org/10.1016/j.foodcont.2015.10.044

Kinnaird, M. (2003). Swiftlets of Borneo: Builders of Edible Nests by Lim Chan Koon and Earl of Cranbrook (2002), ix + 171 pp., Natural History Publications (Borneo) Sdn.

Bhd., Sabah, Malaysia. ISBN 983 812 060 X (hbk), price unmarked. *Oryx*, *37*(1), 119–121. https://doi.org/10.1017/S0030605303230210

Kosmides, A. K., Kamisoglu, K., Calvano, S. E., Corbett, S. A., & Androulakis, I. P. (2013). Metabolomic fingerprinting: Challenges and opportunities. *Critical Reviews in Biomedical Engineering*, *41*(3), 205–221. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4096240/

Lau, A., Melville, D., & International Traffic Network. (1994). *International trade in Swiftlet nests with special reference to Hong Kong*. Traffic International.

Lee, T. H., Wani, W. A., Koay, Y. S., Kavita, S., Tan, E. T. T., & Shreaz, S. (2017). Recent advances in the identification and authentication methods of edible bird's nest. *Food Research International*, *100*, 14–27. https://doi.org/10.1016/j.foodres.2017.07.036

Madsen, R., Lundstedt, T., & Trygg, J. (2010). Chemometrics in metabolomics—A review in human disease diagnosis. *Analytica Chimica Acta*, *659*(1), 23–33. https://doi.org/10.1016/j.aca.2009.11.042

Marcone, M. F. (2005). Characterization of the edible bird's nest the "Caviar of the East." *Food Research International*, *38*(10), 1125–1134. https://doi.org/10.1016/j.foodres.2005.02.008

Marni, S., Syahidah, H. W., Ketty, G. S. L., Roosnoor, F. H., Norakmar, I., Faridah, I., Norzela, A. M., Yahasmida, Y., & Bing, C. H. (2012). Current Application Techniques for Detection of Edible Bird Nest Adulteration and Fraud by the Department Of Veterinary Services Malaysia. *Proceedings of the 2nd Veterinary Research Convention*, 19–20.

Nakagawa, H., Hama, Y., Sumi, T., Li, S. C., Maskos, K., Kalayanamitra, K., Mizumoto, S., Sugahara, K., & Li, Y. T. (2007). Occurrence of a nonsulfated chondroitin proteoglycan in the dried saliva of Collocalia swiftlets (edible bird's-nest). *Glycobiology*, *17*(2), 157–164. https://doi.org/10.1093/glycob/cwl058

Norhayati, M. K., Azman, O., & Nazaimoon, W. M. W. (2010). Preliminary Study of the Nutritional Content of Malaysian Edible Bird's Nest. *Malaysian Journal of Nutrition*, *16*(3), 389–396.

Patti, G. J., Yanes, O., & Siuzdak, G. (2012). Metabolomics: the apogee of the omic trilogy. *Nature Reviews Molecular Cell Biology*, *13*(4), 263–269. https://doi.org/10.1038/nrm3314

Riedl, J., Esslinger, S., & Fauhl-Hassek, C. (2015). Review of validation and reporting of non-targeted fingerprinting approaches for food authentication. *Analytica Chimica Acta*, *885*, 17–32. https://doi.org/10.1016/j.aca.2015.06.003

Righetti, L., Rubert, J., Galaverna, G., Hurkova, K., Dall'Asta, C., Hajslova, J., & Stranska-Zachariasova, M. (2018). A novel approach based on untargeted lipidomics reveals differences in the lipid pattern among durum and common wheat. *Food Chemistry*, *240*, 775–783. https://doi.org/10.1016/j.foodchem.2017.08.020

Roessner, U., & Bowne, J. (2009). What is metabolomics all about? *Biotechniques*, *46*(5), 363.

Seow, E.-K., Ibrahim, B., Muhammad, S. A., Lee, L. H., & Cheng, L.-H. (2016). Differentiation between house and cave edible bird's nests by chemometric analysis of

amino acid composition data. *LWT - Food Science and Technology*, *65*, 428–435. https://doi.org/10.1016/j.lwt.2015.08.047

Shim, E. K. S., Chandra, G. F., Pedireddy, S., & Lee, S.-Y. (2016). Characterization of swiftlet edible bird nest, a mucin glycoprotein, and its adulterants by Raman microspectroscopy. *Journal of Food Science and Technology*, *53*(9), 3602–3608. https://doi.org/10.1007/s13197-016-2344-3

Smolinska, A., Blanchet, L., Buydens, L. M. C., & Wijmenga, S. S. (2012). NMR and pattern recognition methods in metabolomics: From data acquisition to biomarker discovery: A review. *Analytica Chimica Acta*, *750*, 82–97. https://doi.org/10.1016/j.aca.2012.05.049

Spiteri, M., Jamin, E., Thomas, F., Rebours, A., Lees, M. M., Rogers, K. M., & Rutledge, D. N. (2015). Fast and global authenticity screening of honey using 1H-NMR profiling. *Food Chemistry*, *189*, 60–66. https://doi.org/10.1016/j.foodchem.2014.11.099

Stephens, P. (2003). 6 - Culture methods. In T. A. McMeekin (Ed.), *Detecting Pathogens in Food* (pp. 123–146). Woodhead Publishing. https://doi.org/10.1533/9781855737044.2.123

Trygg, J., Holmes, E., & Lundstedt, T. (2007). Chemometrics in metabonomics. *Journal of Proteome Research*, *6*(2), 469–479.

Tung, C. H., Pan, J. Q., Chang, H. M., & Chou, S. S. (2008). Authentic determination of bird's nests by saccharides profile. *Journal of Food and Drug Analysis*, *16*(4), 86–91. https://doi.org/10.38212/2224-6614.2339

Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C. F., Tolmie, D. E., Kent Wenger, R., Yao, H., & Markley, J. L. (2008). BioMagResBank. *Nucleic Acids Research*, *36*(suppl_1), D402–D408. https://doi.org/10.1093/nar/gkm957

Utomo, B., Rosyidi, D., Radiati, L. E., Tri Puspaningsih, N. N., & Proborini, W. D. (2018). Use of keratinase to maintain pre-washing glycoprotein profiles of edible bird's nest. *Drug Invention Today*, *10*(Special Issue 2), 2986–2990.

Wishart, D. S. (2008). Metabolomics: Applications to food science and nutrition research. *Trends in Food Science & Technology*, *19*(9), 482–493. https://doi.org/10.1016/j.tifs.2008.03.003

Wishart, D. S., Feunang, Y. D., Marcu, A., Guo, A. C., Liang, K., Vázquez-Fresno, R., Sajed, T., Johnson, D., Li, C., Karu, N., Sayeeda, Z., Lo, E., Assempour, N., Berjanskii, M., Singhal, S., Arndt, D., Liang, Y., Badran, H., Grant, J., … Scalbert, A. (2018). HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Research*, *46*(D1), D608–D617. https://doi.org/10.1093/nar/gkx1089

Yi, L., Dong, N., Yun, Y., Deng, B., Ren, D., Liu, S., & Liang, Y. (2016). Chemometric methods in data processing of mass spectrometry-based metabolomics: A review. *Analytica Chimica Acta*, *914*, 17–34. https://doi.org/10.1016/j.aca.2016.02.001

559 **List of tables**

560

561 **Table 1.** Authentic EBNs according to production region.

562

563

564 **Table 2.** List of adulterants and adulteration percentage (% w/w) used.

565

566

567 **Table 3.** Misclassification table for the two-class predictive model.

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583 **List of figures**

584

585 **Fig. 1.** ^{1}H-NMR metabolomic fingerprint spectra of 44 authentic EBN samples according to
586 geographical region in Malaysia at chemical shifts of A) 0.00 to 4.0 ppm and B) 6.00 to
587 9.00 ppm.

588

589

590 **Fig. 2.** PCA of EBN samples according to geographical region in Malaysia.

591

592

593 **Fig. 3.** PCA of authentic EBNs and the five adulterants used in simulated adulteration.

594

595

596 **Fig. 4.** PCA of authentic EBNs against A: nutrient agar, B: collagen, C: gelatine, D: karaya
597 gum and E: melamine.

598

599

600 **Fig. 5.** ^{1}H-NMR metabolomic fingerprint spectra of authentic EBNs against those adulterated
601 with 10% w/w nutrient agar, collagen, gelatine, karaya gum and melamine, respectively, at A:
602 0.00–4.10 ppm and B: 5.90–8.50 ppm.

603

604

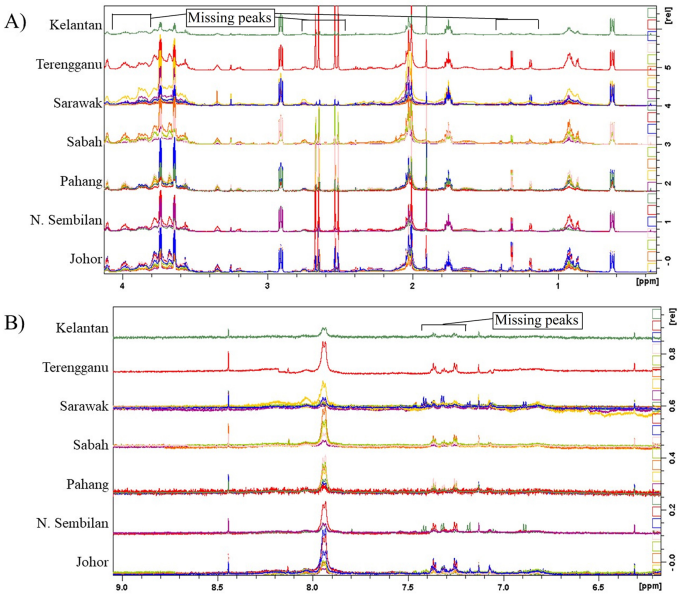
605 **Fig. 6.** OPLS-DA two-class predictive model of authentic EBNs and adulterated EBNs using
606 80% training set and 20% test set.

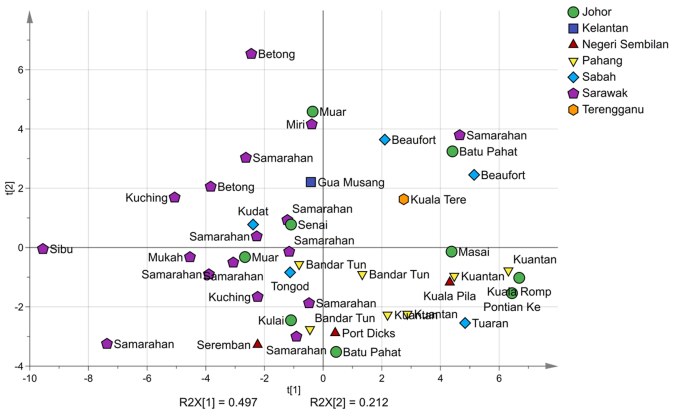
607

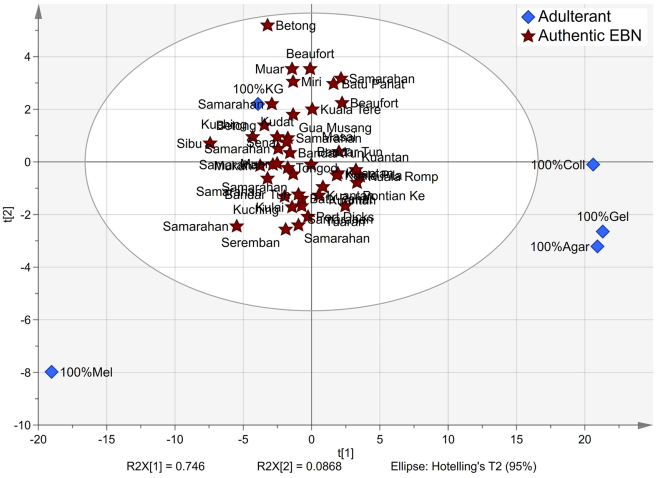
States	Production region (number of samples)
Johor	Batu Pahat (2), Kuala Rompin (1), Kulai (1), Masai (1), Muar (2), Pontian Kecil (1), Senai (1)
Kelantan	Gua Musang (1)
Negeri Sembilan	Kuala Pilah (1), Port Dickson (1), Seremban (1)
Pahang	Bandar Tun Razak (3), Kuantan (4)
Sabah	Beaufort (2), Kudat (1), Tongod (1), Tuaran (1)
Sarawak	Betong (2), Kuching (2), Miri (1), Mukah (1), Samarahan (10), Sibu (1)
Terengganu	Kuala Terengganu (1)

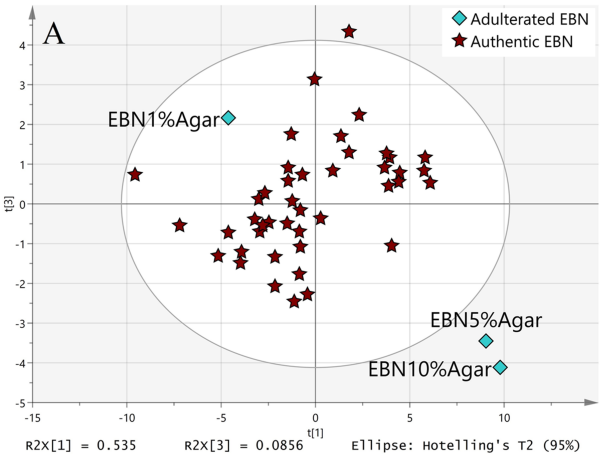
Adulterant	Adulteration percentage (% w/w)	Code
Nutrient Agar	100	100pAgar
	10	10pAgar
	5	5pAgar
	1	1pAgar
Collagen	100	100pColl
	10	10pColl
	5	5pColl
	1	1pColl
Gelatine	100	100pGel
	10	10pGel
	5	5pGel
	1	1pGel
Karaya Gum	100	100pKG
	10	10pKG
	5	5pKG
	1	1pKG
Melamine	100	100pMel
	10	10pMel
	5	5pMel
	1	1pMel

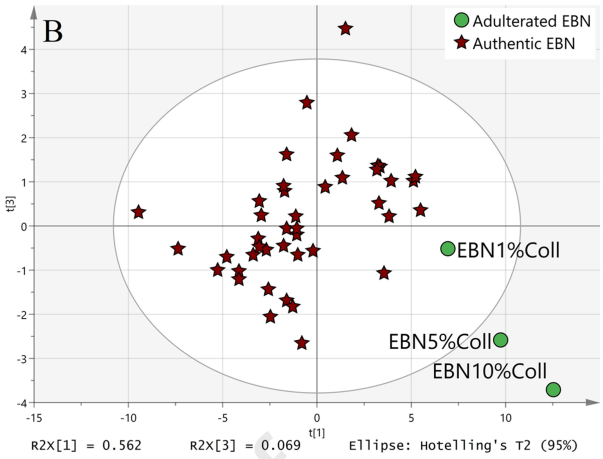
	Members	Correct	Authentic EBN	Adulterated EBN
Authentic EBN	9	100%	9	0
Adulterated EBN	3	100%	0	3
Total	12	100%	9	3

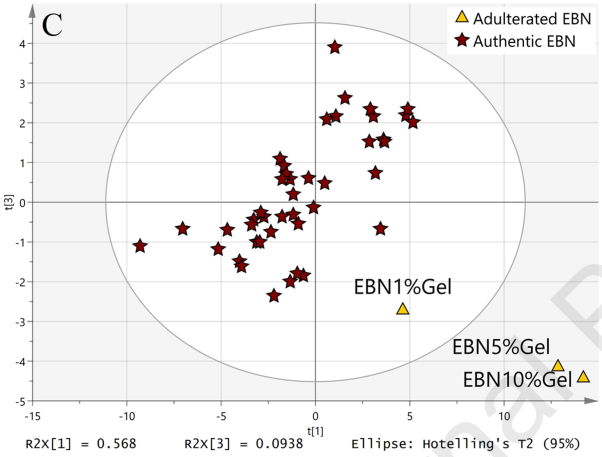


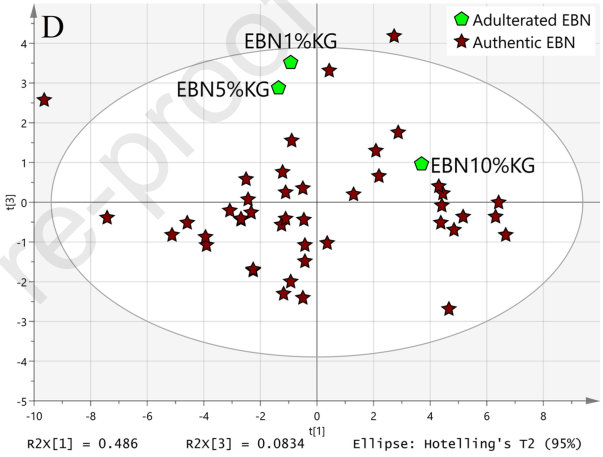


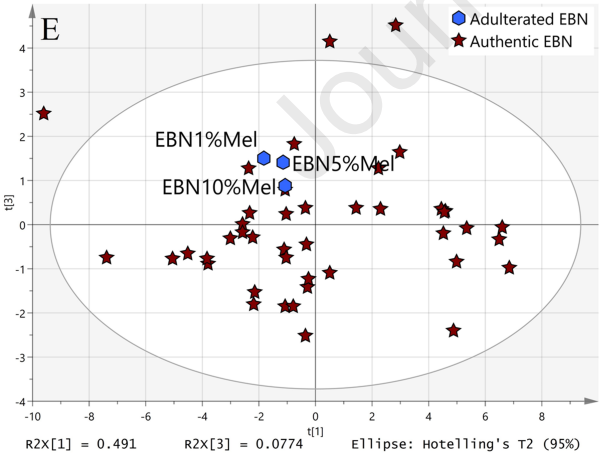




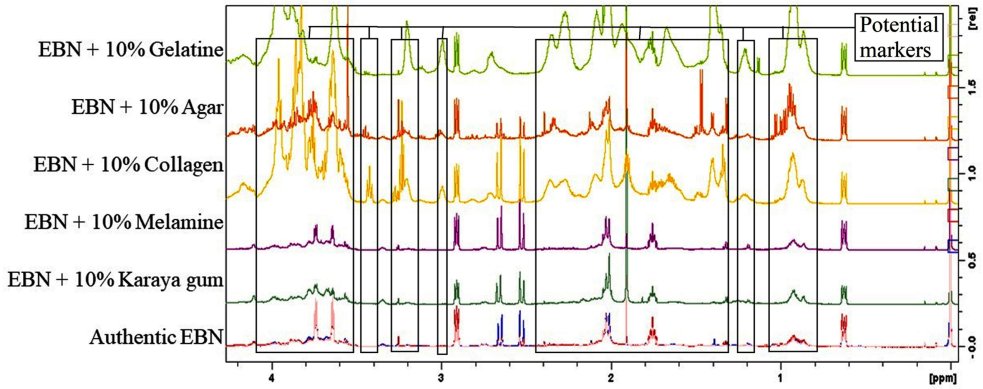




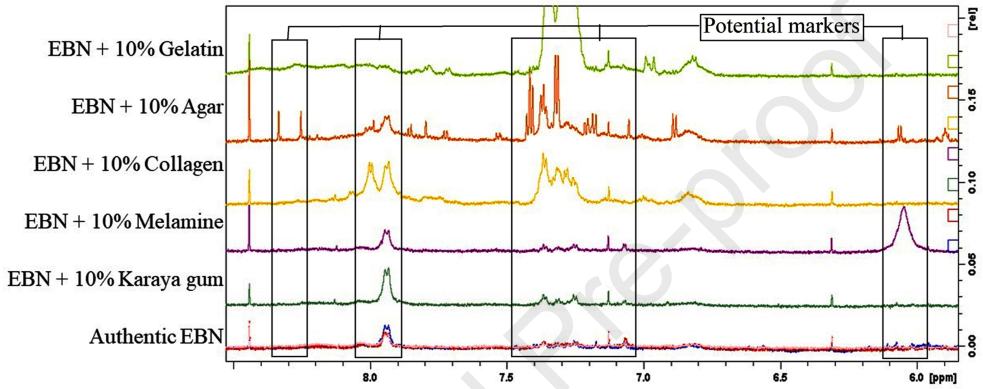


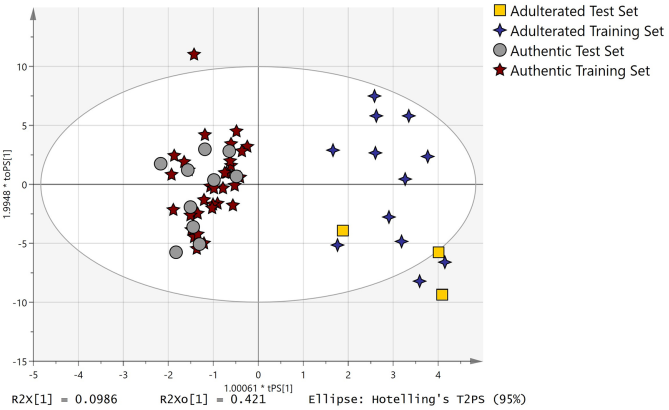


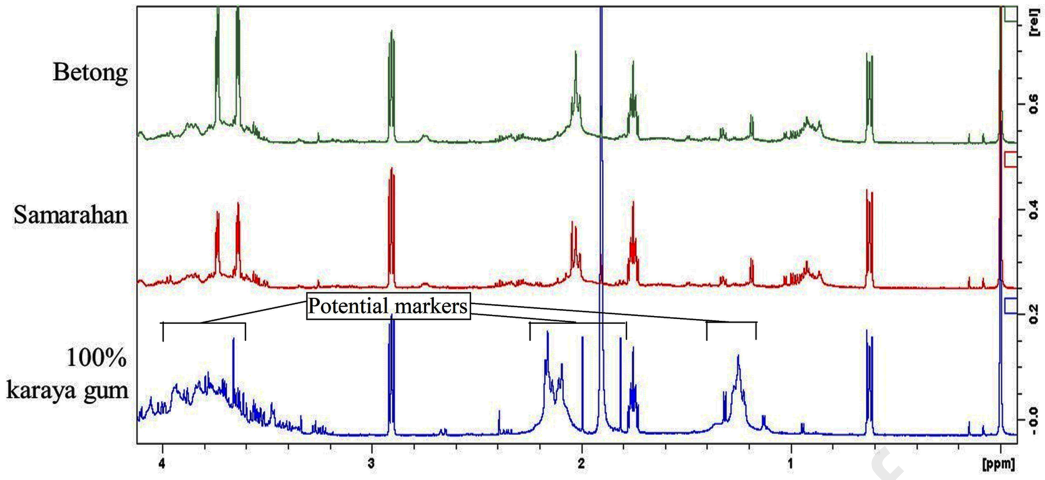
A)



B)







Highlights

- Detection of edible bird's nest adulteration using NMR metabolomic with chemometrics
- A singlet peak could potentially be a marker to detect karaya gum and melamine
- OPLS-DA predictive model differentiated authentic from that of adulterated EBN

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: