Detection of syrup adulterants in manuka and jarrah honey using HPTLC-multivariate data analysis

Md Khairul Islam[1,2], Kevin Vinsen[3], Tomislav Sostaric[1], Lee Yong Lim[1] and Cornelia Locher[1,2]

[1] Division of Pharmacy, School of Allied Health, University of Western Australia, Crawley, WA, Australia
[2] Cooperative Research Centre for Honey Bee Products Limited (CRC HBP), Perth, WA, Australia
[3] International Centre for Radio Astronomy Research (ICRAR), University of Western Australia, Crawley, WA, Australia

ABSTRACT

High-Performance Thin-Layer Chromatography (HPTLC) was used in a chemometric investigation of the derived sugar and organic extract profiles of two different honeys (Manuka and Jarrah) with adulterants. Each honey was adulterated with one of six different sugar syrups (rice, corn, golden, treacle, glucose and maple syrups) in five different concentrations (10%, 20%, 30%, 40%, and 50% w/w). The chemometric analysis was based on the combined sugar and organic extract profiles' datasets. To obtain the respective sugar profiles, the amount of fructose, glucose, maltose, and sucrose present in the honey was quantified and for the organic extract profile, the honey's dichloromethane extract was investigated at 254 and 366 nm, as well as at T (Transmittance) white light and at 366 nm after derivatisation. The presence of sugar syrups, even at a concentration of only 10%, significantly influenced the honeys' sugar and organic extract profiles and multivariate data analysis of these profiles, in particular cluster analysis (CA), principal component analysis (PCA), principal component regression (PCR), partial least-squares regression (PLSR) and Machine Learning using an artificial neural network (ANN), were able to detect post-harvest syrup adulterations and to discriminate between neat and adulterated honey samples. Cluster analysis and principal component analysis, for instance, could easily differentiate between neat and adulterated honeys through the use of CA or PCA plots. In particular the presence of excess amounts of maltose and sucrose allowed for the detection of sugar adulterants and adulterated honeys by HPTLC-multivariate data analysis. Partial least-squares regression and artificial neural networking were employed, with augmented datasets, to develop optimal calibration for the adulterated honeys and to predict those accurately, which suggests a good predictive capacity of the developed model.

Subjects Biochemistry, Food Science and Technology
Keywords High performance thin layer chromatography (HPTLC), Honey, Sugar syrups, Adulteration, Multivariate data analysis

INTRODUCTION

Honey has been regarded as a nutritional and medicinal food for thousands of years (*Kuropatnicki, Kłósek & Kucharzewski, 2018*; *Vorlová, 2015*). About 70–80% of the total solid content of honey is made of sugars (*Pita-Calvo & Vázquez, 2018*), in particular fructose, glucose, maltose and sucrose (*Arvanitoyannis et al., 2005*). Sugars play a vital role, not only in providing crucial nutrition to bee colonies, but also for maintaining the osmolality of the honey (*Olaitan, Adeleke & Ola, 2007*) and thus its self-preservation. As the bulk of the honey is made of sugars, sugar adulterations are common (*Amiry, Esmaiili & Alizadeh, 2017*), either by adding sugar syrups to the honey matrix to increase its bulk volume or by feeding bees with sugar syrups to increase honey production (*Du et al., 2015*; *Rios-Corripio, Rojas-López & Delgado-Macuil, 2012*). Glucose syrup, maple syrup, rice syrup, brown rice syrup, treacle syrup, golden syrup and high fructose corn syrups are commonly used in post-harvest honey adulteration (*Abdel-Aal, Ziena & Youssef, 1993*; *Amiry, Esmaiili & Alizadeh, 2017*; *Chen et al., 2014*; *Ferreiro-González et al., 2018*).

A wide range of methods exist for the detection of adulterants in honey, including $^{13}C/^{12}C$ Stable Carbon Isotope Ratio Analysis (*Bertelli et al., 2010*; *Jamróz et al., 2014*), High-Performance Liquid Chromatography-Mass Spectrometry (HPLC-MS) (*Wen et al., 2017*), Gas Chromatography–Mass Spectrometry (GC-MS) (*Tahir et al., 2016*), Nuclear Magnetic Resonance (NMR) Spectrometry (*Donarski, Jones & Charlton, 2008*; *He et al., 2020*; *Lolli et al., 2008*; *Spiteri et al., 2015*), Near-infrared (NIR) spectroscopy (*Başar & Özdemir, 2018*; *Mishra et al., 2010*), Raman spectroscopy (*Corvucci et al., 2015*), Laser-Induced Breakdown Spectroscopy (*Nespeca et al., 2020*; *Zhao et al., 2020*), Fourier Transform Infrared (FT-IR) Spectroscopy (*Hennessy, Downey & O'Donnell, 2010*; *Riswahyuli et al., 2020*; *Wang et al., 2010*), Plasma Mass Spectrometry (*Voica, Iordache & Ionete, 2020*) and X-ray fluorescence (*Fiamegos et al., 2020*). Chemometric analyses to confirm the geographic origin and authenticity of honey are also commonly used, based on routine chemical quality parameters (*Sant'Ana et al., 2012*), physicochemical characteristics (*Bentabol Manzanares et al., 2017*; *Rodríguez-Flores et al., 2019*; *Scholz et al., 2020*; *Truzzi et al., 2014*), mineral composition data (*Grembecka & Szefer, 2013*; *Voica, Iordache & Ionete, 2020*), amino acid and protein profiles (*Kečkeš et al., 2013*), phenolic components (*Ciucure & Geană, 2019*; *Dżugan et al., 2018*; *Rodríguez-Flores et al., 2019*; *Tonello et al., 2016*; *Wen et al., 2017*) as well as volatile content evaluation (*Tahir et al., 2016*).

All these methods have their limitations; NMR, GC, plasma mass spectrometry and X-ray fluorescence, for instance, require expensive instrumentation and might thus not be suitable for regular quality control purposes. Furthermore, NMR analysis is based on a spatial reference database and also requires expert personnel for data interpretation (*Ferreiro-González et al., 2018*). $^{13}C/^{12}C$ stable carbon isotope ratio analysis, NIR and FT-IR on the other hand are limited in their analytical capacity to the detection and/or quantification of sugars only (*Se et al., 2019*), which is not always suitable for the discrimination between different honeys. High-Performance Thin-Layer Chromatography

(HPTLC) analysis presents an attractive analytical alternative, in particular for day-to-day quality control purposes, as it is less expensive and can provide rich data (*e.g.*, quantification of sugars and organic extract fingerprint for floral source identification) (*Islam et al., 2020a*; *Islam et al., 2020b*; *Locher, Neumann & Sostaric, 2017*; *Locher et al., 2018*). HPTLC is a sophisticated and increasingly popular tool for the analysis of complex matrices. It facilitates semi-to fully automated analysis, it is very efficient as it supports the parallel analysis of multiple samples in a single run, it provides options for a wide range of chemical derivatisations (untargeted and targeted analysis) and has the ability of hyphenation with other analytical platforms and multivariate data analysis (*Islam et al., 2020b*; *Li et al., 2019*; *Stanek & Jasicka-Misiak, 2018*; *Stanek, Kafarski & Jasicka-Misiak, 2019*). Multivariate Data Analysis (MVDA), including cluster analysis (CA), hierarchical clustering analysis (HCA), principal component analysis (PCA), principal component regression (PCR), partial least-squares regression (PLSR) and artificial neural networks (ANN), have been specifically designed for the analysis and visualisation of complex sets of samples like those presented by honey and honey adulterant mixtures (*Amiry, Esmaiili & Alizadeh, 2017*; *Arvanitoyannis et al., 2005*; *Başar & Özdemir, 2018*; *Bertelli et al., 2010*).

The aim of this study was to determine the feasibility of HPTLC-based profiling in combination with MVDA for the quality control of Manuka and Jarrah honeys, specifically the detection of post-harvest adulterations with a range of sugar syrups. It was found that the combination of HPTLC and multivariate data analysis can not only assist in confirming and quantifying these adulterations, but also in identifying some of the adulterants used.

MATERIALS AND METHODS

Chemicals and reagents

All chemicals and reagents were of analytical grade. Commercial sugar syrups and honeys (Table 1) were obtained from supermarkets and other commercial suppliers in Western Australia.

Sample preparation

Adulterated honey samples were prepared by mixing the respective syrups (rice, corn, golden, treacle, glucose and maple) with honey (Manuka and Jarrah) in a final concentration of 10%, 20%, 30%, 40% and 50% (w/w) respectively. The samples were warmed in a water bath at 36 °C for about 30 min to assist with their mixing into homogenous blends. As described in *Islam et al. (2020b)* in more detail, for sugar analysis one mg/mL samples of the two honeys, six syrups and 60 adulterated honeys were prepared in 50% aqueous methanol. The organic honey extracts were obtained as previously described in *Locher, Neumann & Sostaric (2017)* and *Locher et al. (2018)*. In brief, approximately one g of each sample was mixed with two mL of deionised water and extracted three times with five mL of dichloromethane. After drying the combined organic extracts with anhydrous $MgSO_4$ followed by filtration, the solvent was evaporated at ambient temperature and the resulting extracts stored at 4 °C. Prior to HPTLC analysis

Table 1 Commercial syrups and honey samples.

	Sample name	Label and packaging information	Reference ID
Honeys	Manuka	Australian Manuka (MGO 514+) Barnes Naturals Pty Ltd.	MAN
	Jarrah	Boyanup Jarrah Sweet As Apiary, WA	JAR
Syrups	Rice	Organic Rice Syrup Pureharvest, VIC	RIC
	Corn	Corn Malt Syrup Korea Connections Ptl Ltd.	COR
	Golden	CSR Golden Syrup Sugar Australia Pty Ltd.	GOL
	Treacle	CSR Treacle Syrup Sugar Australia Pty Ltd.	TRE
	Glucose	Queen Glucose Syrup Queen Fine Foods Pty Ltd, QLD	GLU
	Maple	Queen Maple Syrup Dr. Oetker Queen Australia, QLD	MAP

they were reconstituted in 100 μL of dichloromethane. A solution of 4,5,7-trihydroxyflavanone (0.5 mg/mL) in methanol was prepared as a reference solution.

Instrumentation and sample analysis

The HPTLC sugar and organic honey extract profiles of all samples were obtained using a CAMAG HPTLC instrumentation (Muttenz, Switzerland) following previously established methods (*Islam et al., 2020b*; *Locher, Neumann & Sostaric, 2017*; *Locher et al., 2018*).

In brief, for sugar analysis, three μL of each sample were applied as eight mm wide bands eight mm from the lower edge and 20 mm from the side edge of the HPTLC plate (20 cm × 10 cm glass-backed silica gel 60 F_{254} plates) using a semi-automated HPTLC application device (Linomat 5, CAMAG, Muttenz, Switzerland). Track distances were 11.4 mm with 15 tracks on each 20 × 10 HPTLC plate, which represents the default plate setting. The development chamber was saturated for 60 min (33% relative humidity) and the plates developed to a migration distance of 85 mm in 1-butanol: 2-propanol: aqueous boric acid (5.0 mg/mL, 3:5:1 *v/v/v*). After drying for 5 min, the plates were derivatised with two ml of aniline-diphenylamine-phosphoric acid reagent (CAMAG Nozzle Spraying Derivatiser, Yellow Nozzle, Level 5) and heated for 10 min at 115 °C (CAMAG TLC Plate Heater III). After cooling to room temperature for two min, the plates were analysed at white light with the HPTLC imaging device (TLC Visualizer 2, CAMAG, Muttenz, Switzerland).

To obtain the organic extract profile, the reference solution (four μL) and the reconstituted extracts (five μL) were applied as eight mm bands at eight mm from the lower edge and 20 mm from the side edge of the HPTLC plates (20 cm × 10 cm glass-backed silica gel 60 F_{254} plates) with a semi-automated HPTLC application device (Linomat 5, CAMAG, Muttenz, Switzerland). Track distances were 11.4 mm with 15 tracks

on each 20×10 HPTLC plate, which represents the default plate setting. The development chamber was saturated for 20 min (33% relative humidity) and the plates developed to a final distance of 70 mm in toluene: ethyl acetate: formic acid (6:5:1 *v/v/v*). After drying for 5 min, the chromatographic results were documented at 254 and 366 nm using the HPTLC imaging device (TLC Visualizer 2, CAMAG, Muttenz, Switzerland) before being derivatised with three ml vanillin with sulfuric acid reagent (CAMAG Nozzle Spraying Derivatiser, Yellow Nozzle, Level 3) and being heated for 3 min at 115 °C (CAMAG TLC Plate Heater III). After cooling to room temperature for 2 min, the obtained images were again recorded at white light and 366 nm. The chromatographic analysis was performed using VisionCATS software (Version 2.5, CAMAG, Muttenz, Switzerland), which was also used to control the individual instrumentation modules.

Data acquisition and chemometric analysis

All multivariate data analysis were performed on a $1,253 \times 68$ matrix, which included data derived from the HPTLC sugar and phenolic organic honey extract profiles of the two pure honeys (Manuka and Jarrah), the six syrup adulterants (rice, corn, golden, treacle, glucose and maple) and the resulting 60 adulterated honeys. Sugar data included quantities of fructose, glucose, sucrose and maltose. From the organic honey extract HPTLC profiles, four sets of images (R 245 and R 366, and T white and R 366 after derivatisation with vanillin reagent) were converted into their respective chromatograms. The intensity (AU) of bands expressed as absorption peak height were extracted for data calculation along with their corresponding Rf values. To reduce complexity, only bands between Rf 0.05 and Rf 0.60 were considered, as this captured the majority of all detected bands (*Islam et al., 2021*). The obtained data sets were standardised and the multivariate data analysis performed using R and R Studio (Version 1.3.959) (*Team RC, 2020a*; *Team RC, 2020b*), Python3.9 (*Van Rossum & Drake, 2009*) and PyTorch (*Paszke et al., 2019*).

Multivariate data analysis

Multivariate data analysis included non-supervised techniques, like cluster analysis and principal component analysis, as well as supervised techniques, including principal component regression, partial least-squares regression and machine learning (*Agatonovic-Kustrin & Beresford, 2000*; *Donarski, Jones & Charlton, 2008*; *Se et al., 2019*; *Vorlová, 2015*; *Zhao et al., 2020*).

Chemometric validation

The sugar contents data had a dynamic range of 0 to ~2,000. All the methods chosen for the analysis prefer standardised data in the number range −1 to 1 or 0 to 1; so, the data was standardised. With the ANN, the AU values of the organic extracts naturally have a range from 0 to 1 and thus did not need standardising. To ensure the supervised models would work with unseen data, and not just repeat the labels of the samples that they had just seen, cross validation and k-fold cross validation was used. This involved holding back 30% of the data as a test set to validate the training of the models. With the basic raw data only 68 samples were available. This meant standard cross validation was

inappropriate for PCR, PLSR and ANN, and k-fold cross validation was required. k-Fold cross validation divides all the samples into *k* groups of samples, called folds of equal sizes (if possible). The prediction function is learned using *k-1* folds, and the fold left out is used for testing. Each fold is used for testing once. This process was applied to supervised PCR, PLSR and ANN chemometric techniques. For PCR and PLSR Root Mean Square Error of Cross Validation (RMSECV), Root Mean Square Error (RMSE) and Root Mean Square Error of Prediction (RMSEP) were deemed useful to predict the validation parameters.

Data augmentation

This study aimed to predict the honey/syrup type and the adulteration level of all samples, so all 68 classes. The original data set was too small to be used directly as input to train an ANN, PCR, PLSR, and potentially, a PCA. To address this, the fact that an analysis run using HPTLC is not deterministic was used. As different runs will produce small various in AU and the Rf value can slightly drift, this was used to augment the original dataset by repeating the following process 50 times:

- For the bands obtained in four different analysis conditions (R 245 and R 366, and T white and R 366 after derivatisation)-a drift of ±0.0173 Rf and adding Gaussian noise with a standard deviation of 1.25 times the standard deviation of the original AU value. The data was separated into the four light bands and the standard deviation of each band was calculated.
- For the sugar content values–Gaussian noise with a standard deviation of 5% of the value was add to the samples.

This increased the dataset from 68 to 71,468 ($50 \times 21 \times 68$ + the original 68) as shown in Fig. 1 (and Materials-S1) and Fig. 2 (and Materials-S2).

RESULTS AND DISCUSSIONS

HPTLC fingerprints and corresponding chromatograms

The obtained HPTLC sugars profiles of honey were very simple as they only captured glucose (Rf 0.30) and fructose (Rf 0.14) within the limit of quantification, as can be seen in Fig. 3. Apart from those two sugars, a very faint signal for maltose (Rf 0.20) was also noted, but it was well below the limit of quantification. Honeys adulterated with corn, rice or glucose syrup contained significant amounts of maltose, which was easily quantifiable, whereas honeys adulterate with golden, treacle and maple syrup were characterised by significant amounts of sucrose, which was also easily quantifiable (*Islam et al., 2020a*). In terms of their respective organic extracts most sugar syrups did not present any major bands. Thus, their incorporation into a honey only decreased the intensity of the major bands present in the organic extract of the pure honey. Only maple syrup extract was characterised by a major band at Rf 0.41 (data not presented), which could not be detected in the two honey extracts and therefore acted as signifying band for this particular syrup. Increasing concentrations of adulterants decreased the intensity of the bands in MAN and JAR accordingly (Figs. 4 and 5; only MAN shown for illustrative purposes).
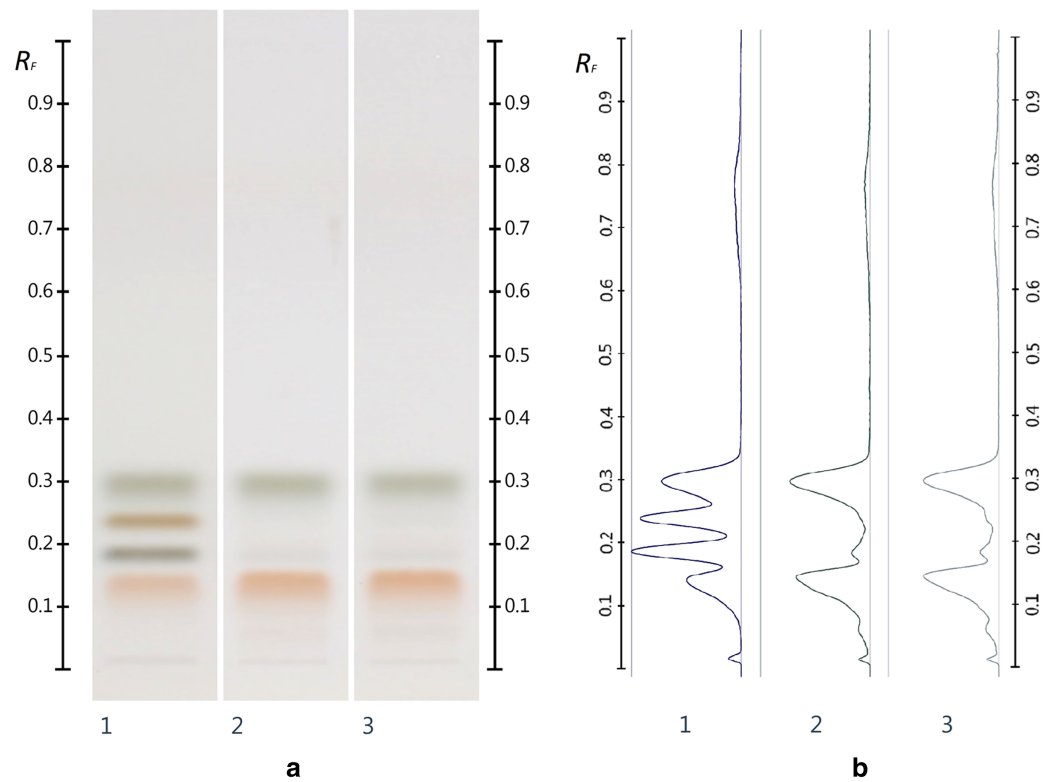
Class: MAN
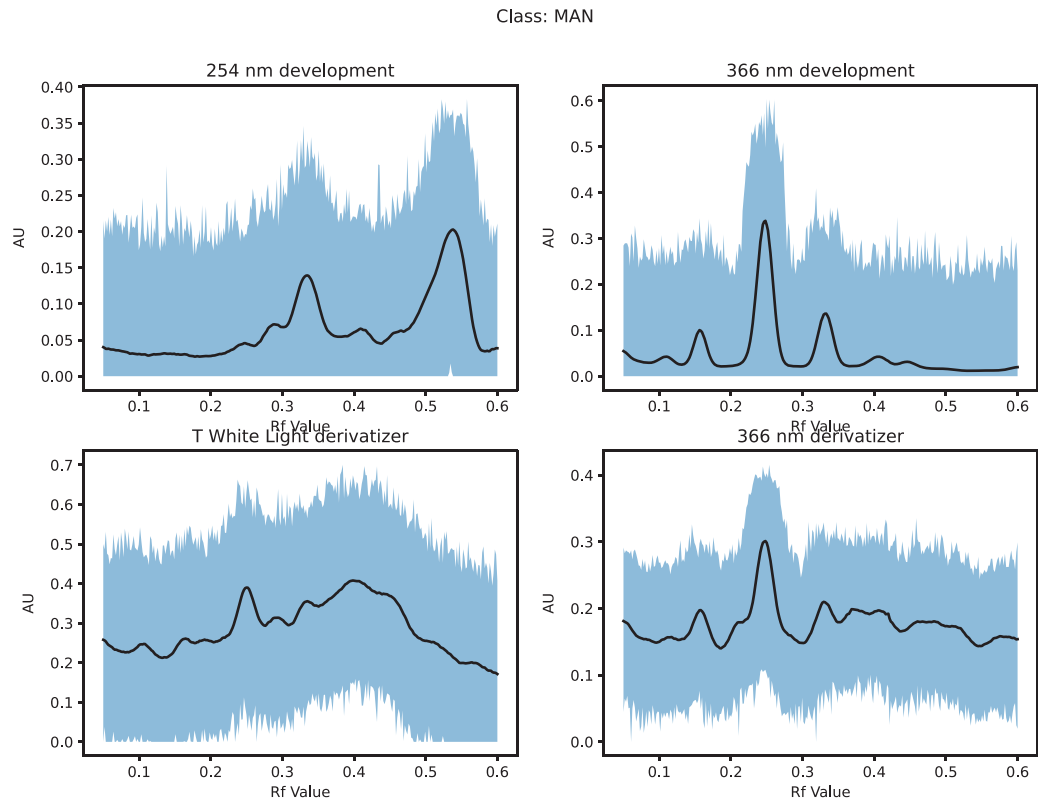
Figure 1 Augmented dataset for organic extracts Rf *vs* intensity values.

Full-size DOI: 10.7717/peerj.12186/fig-1
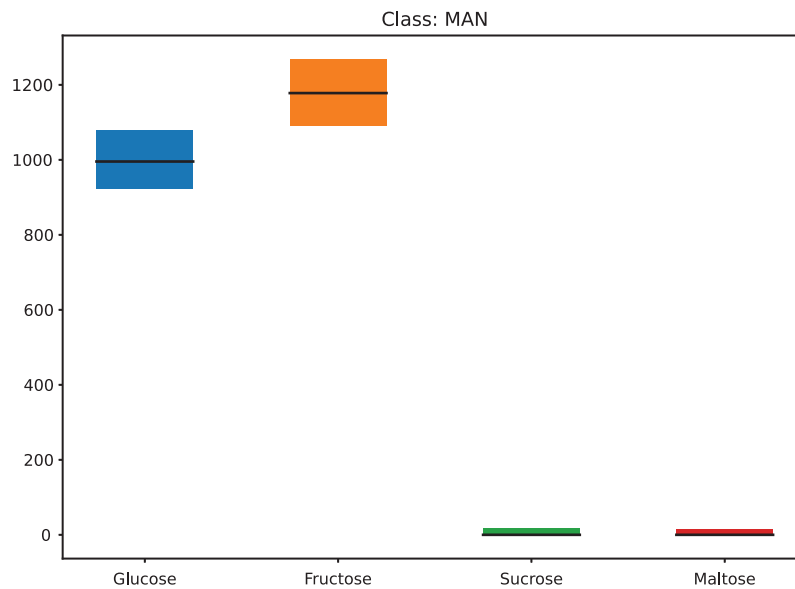
**Figure 2 Augmented dataset for sugar content values.** Full-size DOI: 10.7717/peerj.12186/fig-2

Figure 3 Images taken at T White light after derivatisation with aniline-diphenylamine-phosphoric acid reagent. Track 1–Standards (fructose, maltose, sucrose and glucose in increasing Rf values), Track 2–MAN, Track 3–JAR; three µL of each 50% aqueous methanolic solution (A) and their respective chromatograms (B). Full-size ⬀ DOI: 10.7717/peerj.12186/fig-3
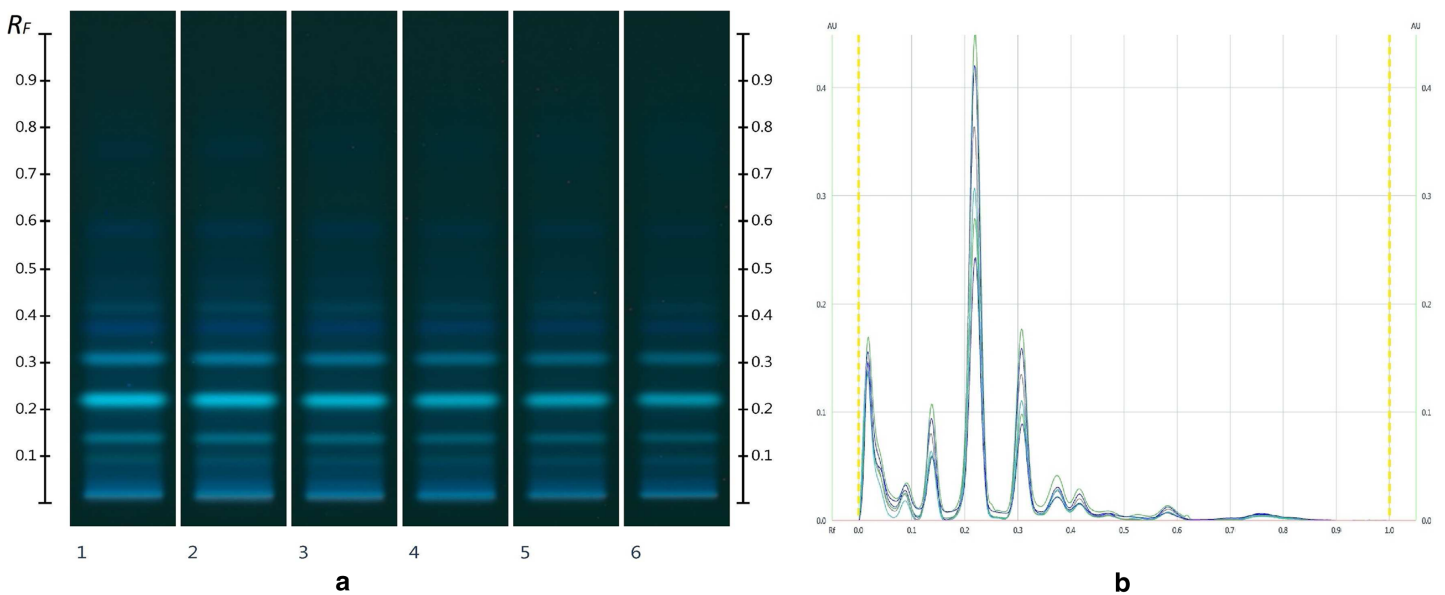
Figure 4 Images taken at R 366 after development (A) and their respective chromatograms (B). Track 1–MAN, Track 2–MAN-RIC 10%, Track 3–MAN-RIC 20%, Track 4–MAN-RIC 30%, Track 5–MAN-RIC 40% and Track 6–MAN-RIC 50%; five µL of each extract. Full-size ⬀ DOI: 10.7717/peerj.12186/fig-4
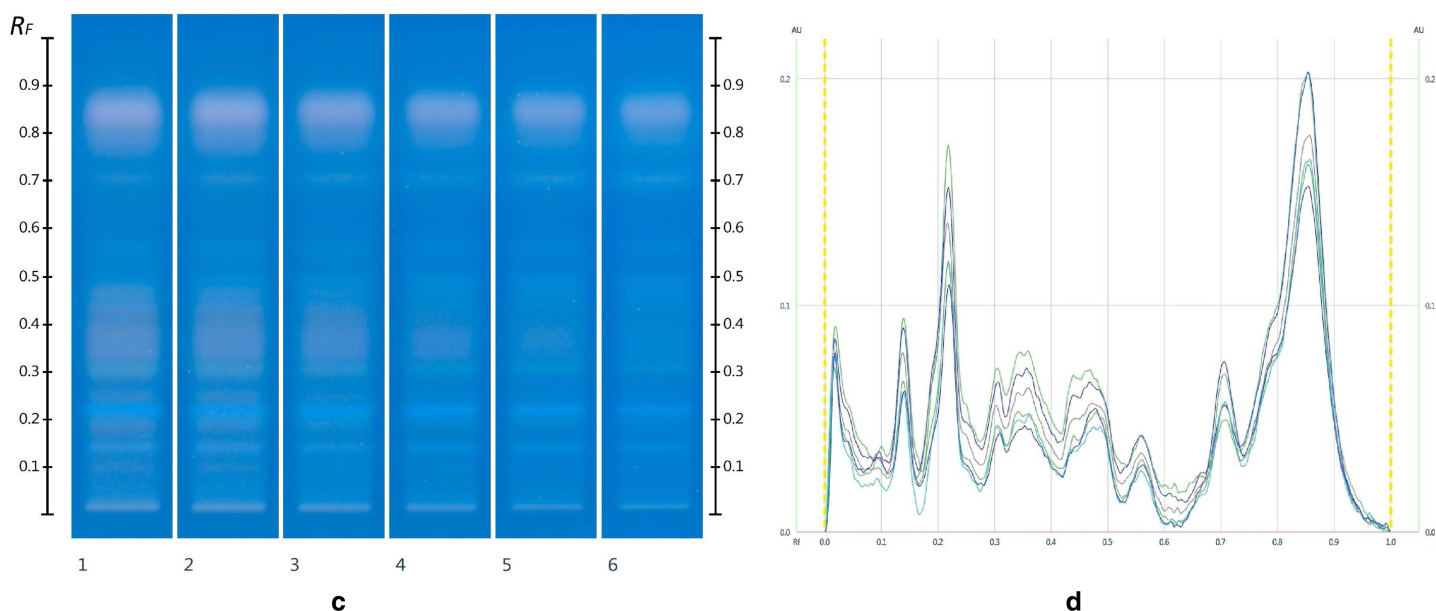
Figure 5 Images taken at R 366 after derivatisation with vanillin reagent (C) and their respective chromatograms (D). Track 1–MAN, Track 2–MAN-RIC 10%, Track 3–MAN-RIC 20%, Track 4–MAN-RIC 30%, Track 5–MAN-RIC 40% and Track 6–MAN-RIC 50%; five µL of each extract.

Full-size ☐ DOI: 10.7717/peerj.12186/fig-5

Cluster analysis

Cluster analysis was applied in this study in order to identify interrelationships between different honey groups. Clustering allows the grouping of similar data points in such a way that points in the same group, known as cluster, are more similar to each other than points in other groups. There are several types of cluster analysis, in this paper Hierarchical Clustering, K-Means Clustering and Density Based Clustering were investigated.

Hierarchical clustering

Hierarchical clustering (HC), an unsupervised clustering algorithm, is one of the most popular and easy to understand clustering techniques. It starts with placing each observation in its own cluster and then stepwise merging clusters by analysing the distances between adjacent points. HC gives an indication not only to what extent individual points are similar to each other but also how similar different clusters are to each other. The number of significant clusters found and displayed in the respective dendrogram is based on the Euclidian distance of the normalised data (*Dżugan et al., 2018*). The hierarchical clustering (average method) for honeys, syrups and adulterated honeys is shown in Fig. 6.

The dendrogram derived from the study's dataset displays three main clusters. The first contains sugar syrups (with the exception of TRE), the other two contain the pure and adulterated honeys. Within the other clusters two major sub-cluster can be distinguished; one contains pure Manuka and syrup adulterated Manuka honeys, the other contains pure Jarrah and syrup adulterated Jarrah honeys. These findings demonstrate the usefulness of hierarchical clustering in distinguishing syrups from pure honeys and syrup adulterated honeys.
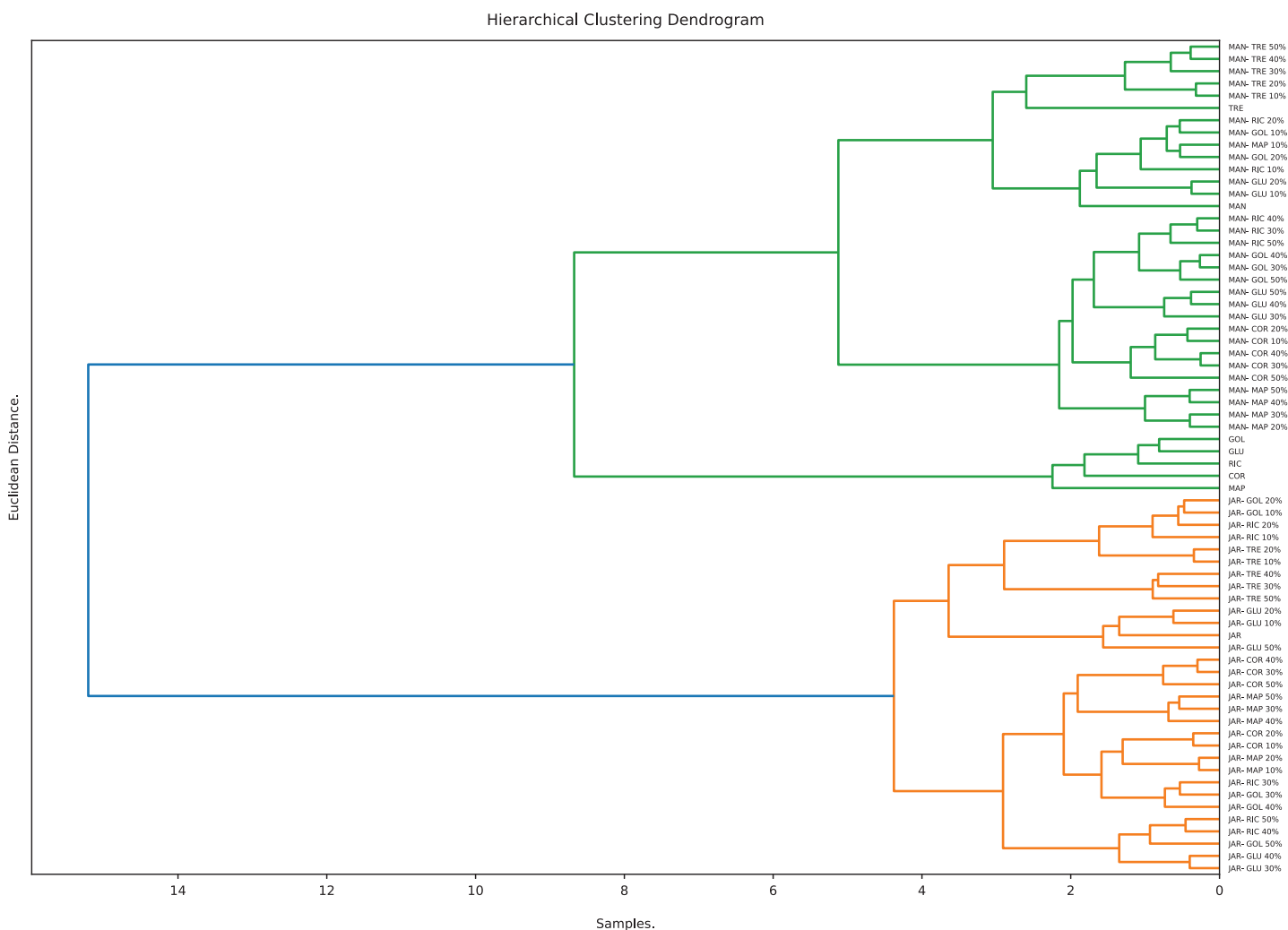
Figure 6 **Hierarchical clustering analysis.**

Full-size ▣ DOI: 10.7717/peerj.12186/fig-6

K-means clustering

K-means clustering is also a very simple and popular unsupervised algorithm. Typically, unsupervised algorithms use information from datasets as input vectors. A target number k is set first, which refers to the number of centroids needed in the dataset (*Ghosh & Dubey, 2013*). A centroid is the imaginary or real location representing the centre of the cluster (*Fränti, Rezaei & Zhao, 2014*). Every data point is allocated to each of the clusters through reducing the in-cluster sum of squares distance between data points and all centroids. In this analysis we used the elbow method to select the number of clusters k to be four and the clusters sizes (K-means clustering with four clusters) were 22, 5, 26, 15 (between_SS/total_SS = 59.5%) with higher "between Sum of Squares (between_SS)/total Sum of Squares (total_SS)" values signifying better cluster differences values (*Janrao, Mishra & Bharadi, 2019*). The scattered plot of the four clusters of Glucose + Fructose *vs* Sucrose + Maltose are shown in Fig. 7. In the cluster scatter plot, Manuka and Jarrah are clearly separated from the syrup and adulterated honeys (Fig. 7).
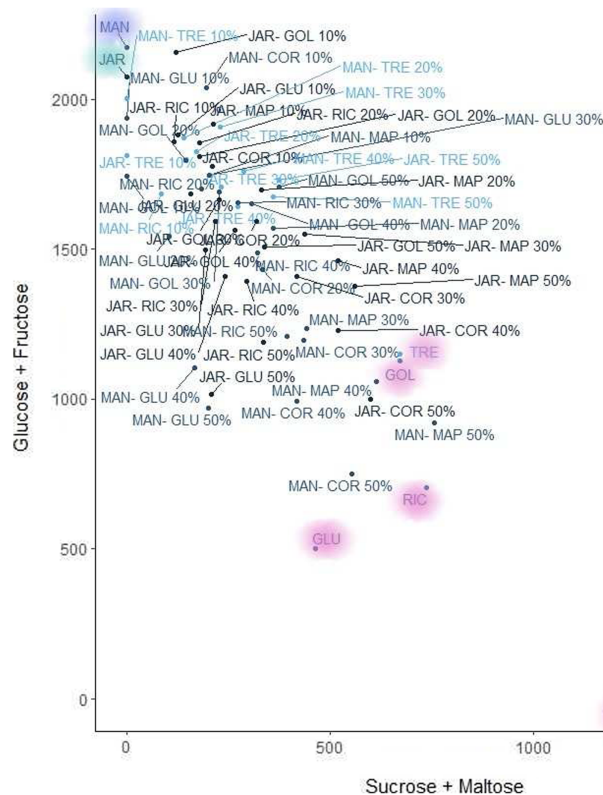
**Figure 7** K-means clustering.   Full-size 🖼 DOI: 10.7717/peerj.12186/fig-7

Density based clustering

DBSCAN is a density based clustering algorithm and is based on the concept of grouping samples into clusters if they are connected to one another by density populated areas in order for samples to be clustered into various shapes and sizes, in which the clusters are not sensitive to noise. In this analysis, the eps (*epsilon* or ϵ-neighbourhood) value, the radius of neighbourhood around a point (*Birant & Kut, 2007*), was set to 21 and minPts as the minimum number of neighbours within the "eps" radius (*Ruiz, Spiliopoulou & Menasalvas, 2007*) was set to three to optimise the clustering results. At eps (21) and minPts (3), the analysis resulted in five clusters (Fig. 8) and seven noise points. As the density-based clusters are not sensitive to noise, the analysis assigns samples to different clusters even within cluster data points and the noise points can be within the clusters but marked as different colours (Fig. 9). The Manuka honey and Manuka adulterated samples were assigned to two different clusters based on their density population and so were Jarrah honey and Jarrah adulterations. The four sugar syrups formed a cluster away from both the Manuka adulterated and Jarrah adulterated clusters. Examination of the seven noise points revealed that Point 1 was Manuka, Point 2 was Jarrah, Point 6 was maple syrup and Point 7 was treacle syrup. Thus, both pure Manuka and Jarrah honey were clearly separated from all other clusters.
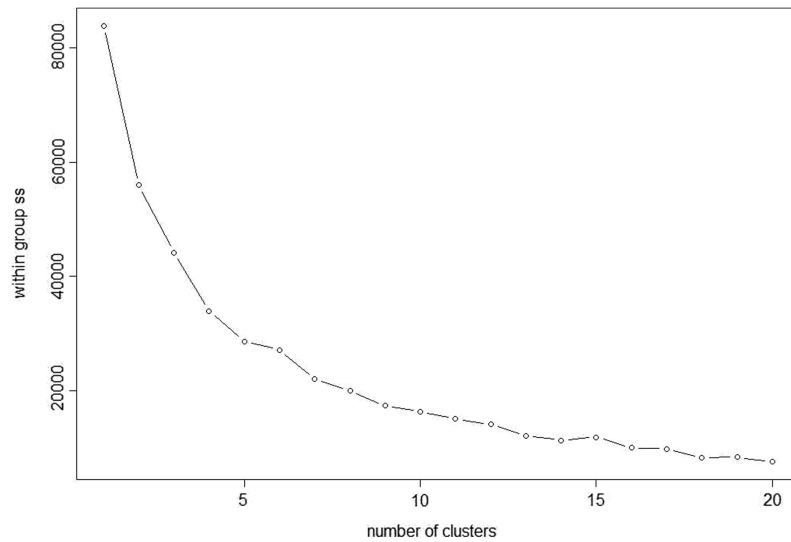
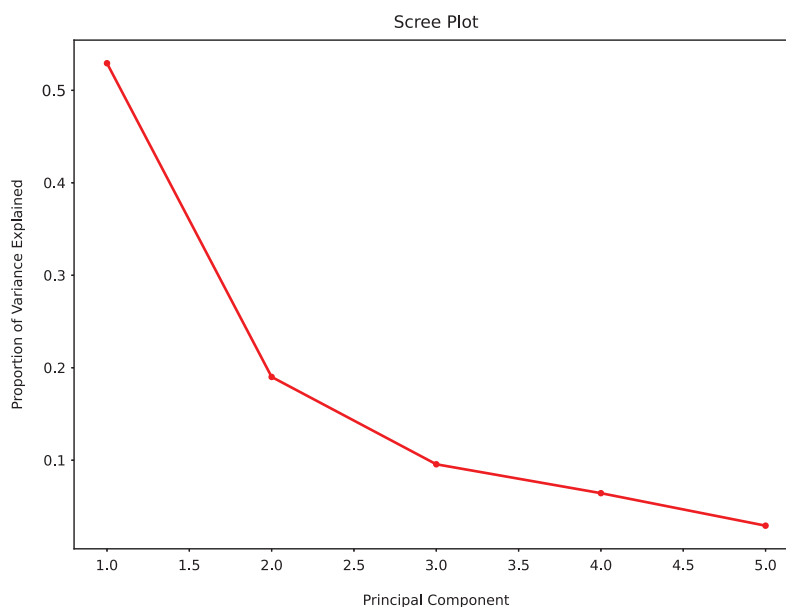**Figure 8 Density based scree plot.** Full-size 🖼 DOI: 10.7717/peerj.12186/fig-8
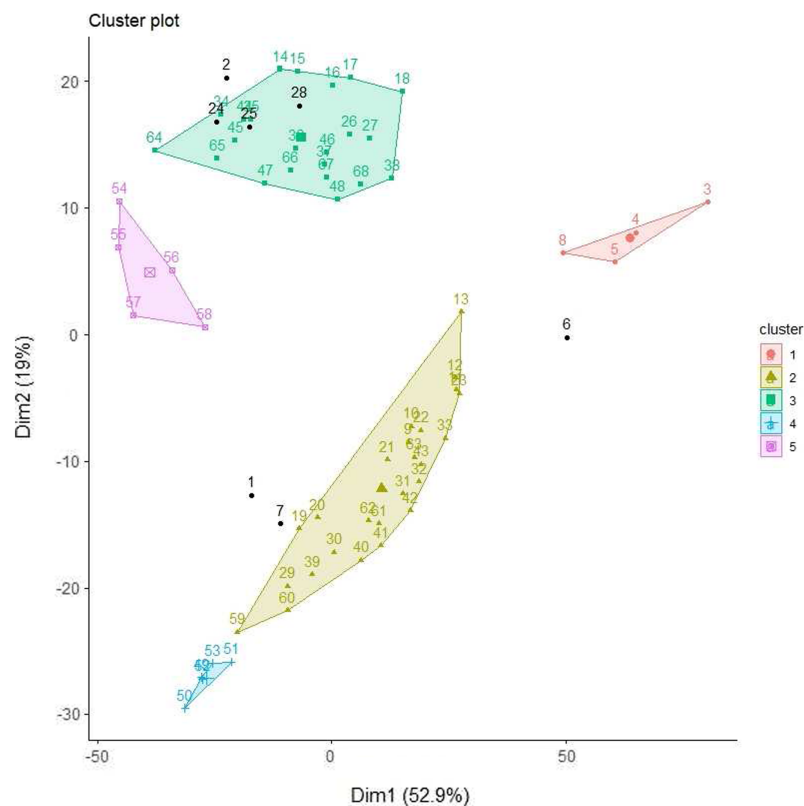
Figure 9 Density based scattered plot. Full-size 🖼 DOI: 10.7717/peerj.12186/fig-9

Principal component analysis

Unsupervised non targeted PCA was conducted to reduce the dimension of the original data to a smaller number of variables or components by examining the relationship between measured parameters. This allowed to explore and model the experimentally

Figure 10 Principal component analysis summary scree plot.
Full-size ☑ DOI: 10.7717/peerj.12186/fig-10

derived data, evaluating their correlation and variability; and to represent data visually into a linear transformation (*Hennessy, Downey & O'Donnell, 2010*; *Voica, Iordache & Ionete, 2020*) such as a scree plot and a loading plot.

The PCA scree plot allows to explore the possible number of clusters and variability among the clusters. Principal Component 1 (PC1) and Principal Component 2 (PC2) had the greatest variability, PC2, PC3 and PC4 had marginal variability. Little variability could be detected for PC5 and beyond (Fig. 10).

Although similar chromatographic profiles were obtained for all honey samples (Figs. 3–5), PCA can easily categorise the obtained data into three different clusters based on principal component 1 (PC1) and principal component 2 (PC2) (Fig. 11). The two main PCs, PC1 and PC2, accounted for 52.9% and 19%, thus taken together for 71.9% of total variability.

A clear differentiation between two clusters can be noted. Cluster 1 contains Manuka and Manuka-syrup adulterated honeys and Cluster 2 contains Jarrah and Jarrah-syrup adulterated honeys. Further exploring the PCA data reveals that in Cluster 1 pure Manuka honey is clearly separated from the other data points within this cluster, the same can be noted for Cluster 2, where pure Jarrah honey is also clearly separated from the adulterated samples (Fig. 11). Five of the sugar syrups (except TRE) formed a separate cluster/group. These findings demonstrate that the PCA model can discriminate not only between pure honeys and sugar syrups but also between different syrup adulterated honeys.

When PCA was applied to the augmented dataset, the general pattern of PC1 *vs* PC2 was retained, but the importance of PC1 and PC2 was significantly reduced (Figs. 12 and 13) due to the added noise.
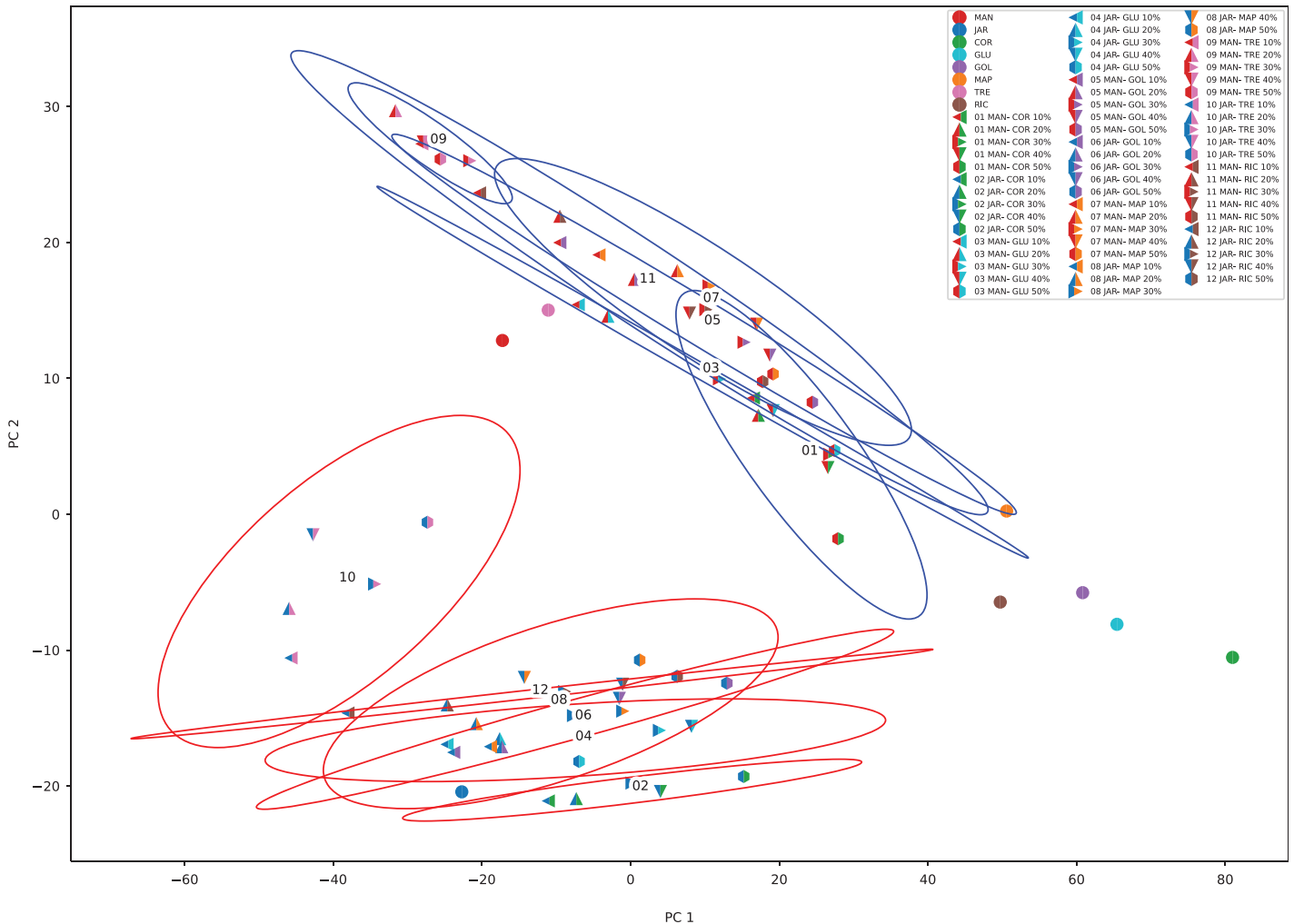
Figure 11 PCA biplot with clustering. Full-size ☒ DOI: 10.7717/peerj.12186/fig-11

Partial least squares regression and principal component regression

PLS is a classical linear multivariate regression tool, whilst PCR is a regression analysis technique that is based on PCA. These techniques were used to derive the levels of the sugars. The important values are those for Sucrose and Maltose, which if the levels are too high is an indication that the honey is adulterated. For completeness we also derived the Glucose and Fructose values. Supervised PLS and PCR were performed on the standardised data set and the augmented data set to independently calculate the values of the Glucose, Fructose, Sucrose and Maltose content from all the other variables (including the AU values). Generally, the number of PLS factors and the characteristic variables of the data can affect the performance of the PLS model. Too few PLS factors tend to decrease the reliability of the model, too many in turn might increase the model's complexity and weaken its stability (*Ferreiro-González et al., 2018*; *He et al., 2020*). In this work, the number of PCs was optimised by cross-validation in a model calibration process
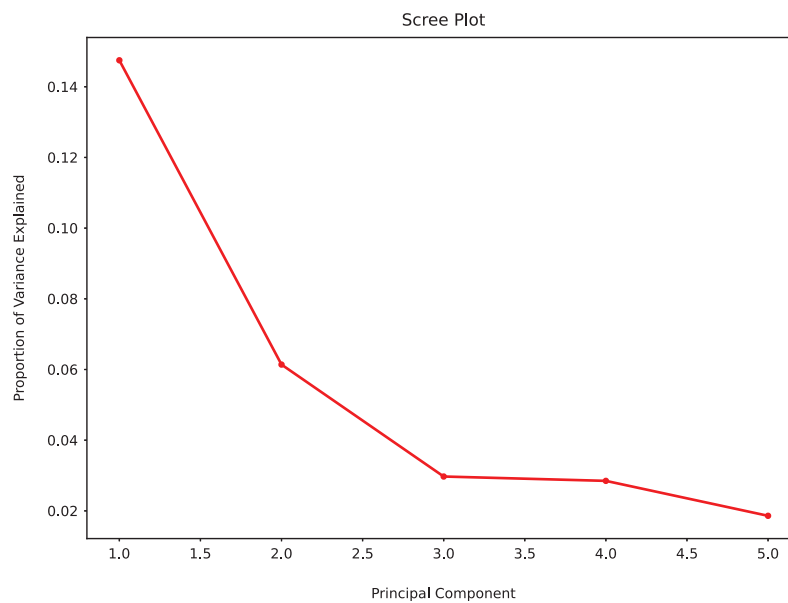
Figure 12 Principal component analysis (augmented data) summary scree plot.
Full-size ⊡ DOI: 10.7717/peerj.12186/fig-12

where the optimal number of PCs was corresponding to the lowest Root Mean Square Error of Cross Validation (RMSECV) value (Fig. 14). For cross-validation, the testing data were accounted (ten folds). To build the supervised model, 70% of the data was chosen to derive a training data set and the remaining 30% of the data were used for testing the model. This corresponded to 45 training set samples and 23 testing set samples in total for the raw standardised dataset and 50,025 training and 21,443 test samples for the augmented dataset. The lowest value of RMSE corresponded to the component number, that was best suited to describe the highest variable. The number of PLS factors or components for Glucose, for instance, was four (Fig. 14C) for the standardised data set and 10 for the augmented dataset. The different sugars required different numbers of components as shown in Table 2.

The main practical difference between PCR and PLSR is that PCR often needs more components than PLSR to achieve the same prediction error (*Mevik, 2007*; *Mevik & Wehrens, 2015*). On this data set, PCR for Glucose would need seven components (Fig. 14A) to achieve the same Root Mean Square Error of Prediction (RMSEP) for the standardised dataset and 14 for the augmented dataset.

Artificial neural networks

An ANN is a supervised mathematical model inspired by biological systems. It simulates the way mammalian brains work in processing information and learning from data. The ANN consists of a number of layers that perform different functions, such as multiplying weights, randomly dropping out nodes and performing non-linear activations. Training is performed iteratively, so that as it progresses, the ANN learns by comparing the ground truth to the predictions, and back propagating the losses to adjust the
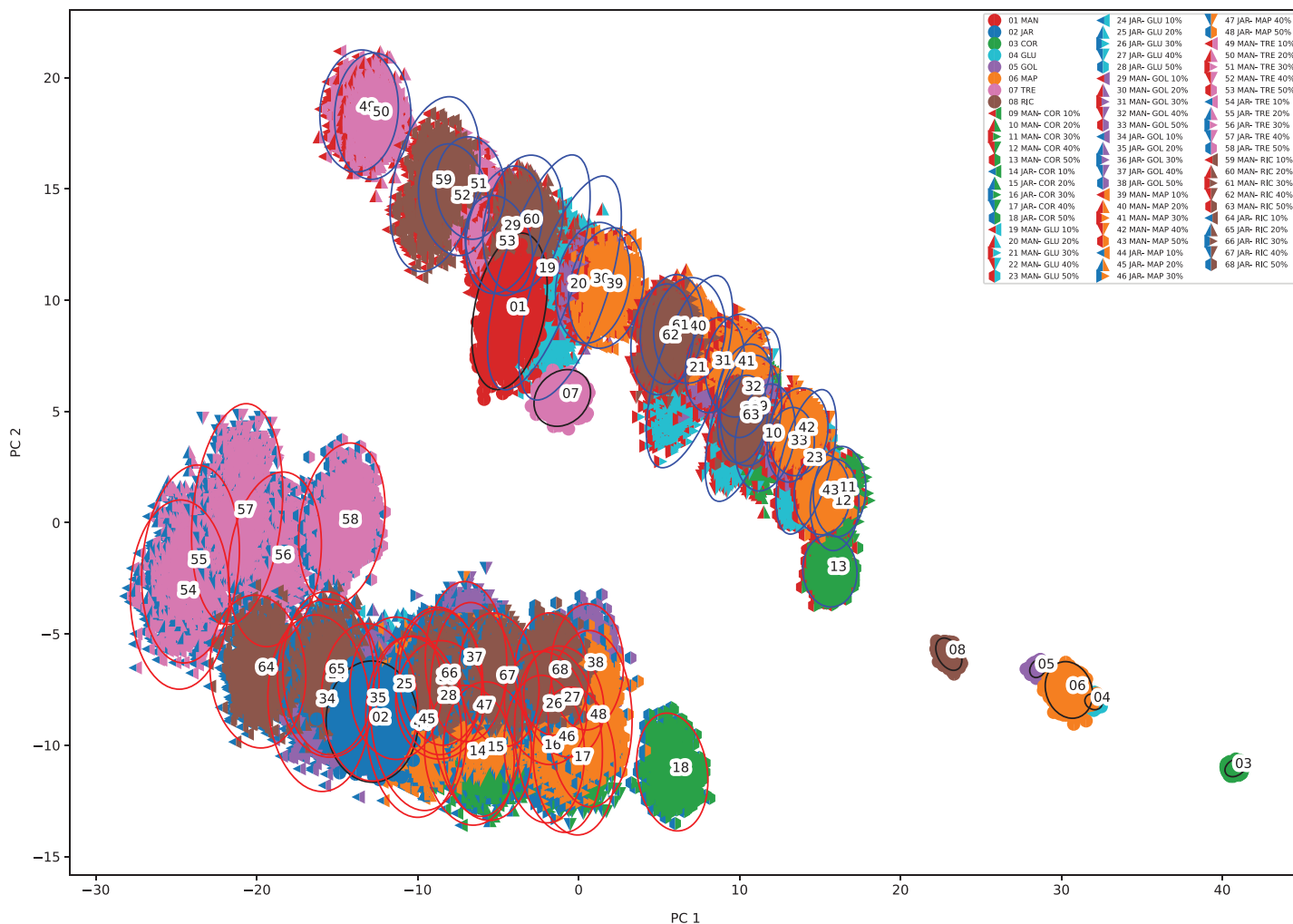
PCA 68 classes - 71468 inputs

Legend:
- 01 MAN
- 02 JAR
- 03 COR
- 04 GLU
- 05 GOL
- 06 MAP
- 07 TRE
- 08 RIC
- 09 MAN- COR 10%
- 10 MAN- COR 20%
- 11 MAN- COR 30%
- 12 MAN- COR 40%
- 13 MAN- COR 50%
- 14 JAR- COR 10%
- 15 JAR- COR 20%
- 16 JAR- COR 30%
- 17 JAR- COR 40%
- 18 JAR- COR 50%
- 19 MAN- GLU 10%
- 20 MAN- GLU 20%
- 21 MAN- GLU 30%
- 22 MAN- GLU 40%
- 23 MAN- GLU 50%
- 24 JAR- GLU 10%
- 25 JAR- GLU 20%
- 26 JAR- GLU 30%
- 27 JAR- GLU 40%
- 28 JAR- GLU 50%
- 29 MAN- GOL 10%
- 30 MAN- GOL 20%
- 31 MAN- GOL 30%
- 32 MAN- GOL 40%
- 33 MAN- GOL 50%
- 34 JAR- GOL 10%
- 35 JAR- GOL 20%
- 36 JAR- GOL 30%
- 37 JAR- GOL 40%
- 38 JAR- GOL 50%
- 39 MAN- MAP 10%
- 40 MAN- MAP 20%
- 41 MAN- MAP 30%
- 42 MAN- MAP 40%
- 43 MAN- MAP 50%
- 44 JAR- MAP 10%
- 45 JAR- MAP 20%
- 46 JAR- MAP 30%
- 47 JAR- MAP 40%
- 48 JAR- MAP 50%
- 49 MAN- TRE 10%
- 50 MAN- TRE 20%
- 51 MAN- TRE 30%
- 52 MAN- TRE 40%
- 53 MAN- TRE 50%
- 54 JAR- TRE 10%
- 55 JAR- TRE 20%
- 56 JAR- TRE 30%
- 57 JAR- TRE 40%
- 58 JAR- TRE 50%
- 59 MAN- RIC 10%
- 60 MAN- RIC 20%
- 61 MAN- RIC 30%
- 62 MAN- RIC 40%
- 63 MAN- RIC 50%
- 64 JAR- RIC 10%
- 65 JAR- RIC 20%
- 66 JAR- RIC 30%
- 67 JAR- RIC 40%
- 68 JAR- RIC 50%

**Figure 13 PCA biplot (augmented data) with clustering.** Full-size DOI: 10.7717/peerj.12186/fig-13

Table 2 Comparison of number of components and RMSE and Rsquare for PCR and PLS method for standardised and augmented dataset.

| | Standardised dataset | | Augmented dataset | | Standardised dataset | | | | Augmented dataset | | | |
| | No. of components | | No. of components | | PCR method | | PLS method | | PCR method | | PLS method | |
	PCR	PLS	PCR	PLS	RMSE	Rsquare	RMSE	Rsquare	RMSE	Rsquare	RMSE	Rsquare
Glucose	7	4	14	10	0.7063	0.6504	0.6328	0.7164	93.7469	0.7553	90.7861	0.7707
Fructose	6	7	12	13	0.5484	0.6629	0.5799	0.5810	147.623	0.7828	128.596	0.8352
Maltose	1	7	14	14	1.0016	0.0012	0.6541	0.2248	86.3064	0.8537	80.254	0.8735
Sucrose	10	7	15	14	0.7484	0.8605	0.6639	0.8960	84.4529	0.8997	74.1253	0.9227

weights of the varies layers. Upon completion the model will have established (linear or non-linear) relationships between input and output data (*Agatonovic-Kustrin & Beresford, 2000*) through the various layers.
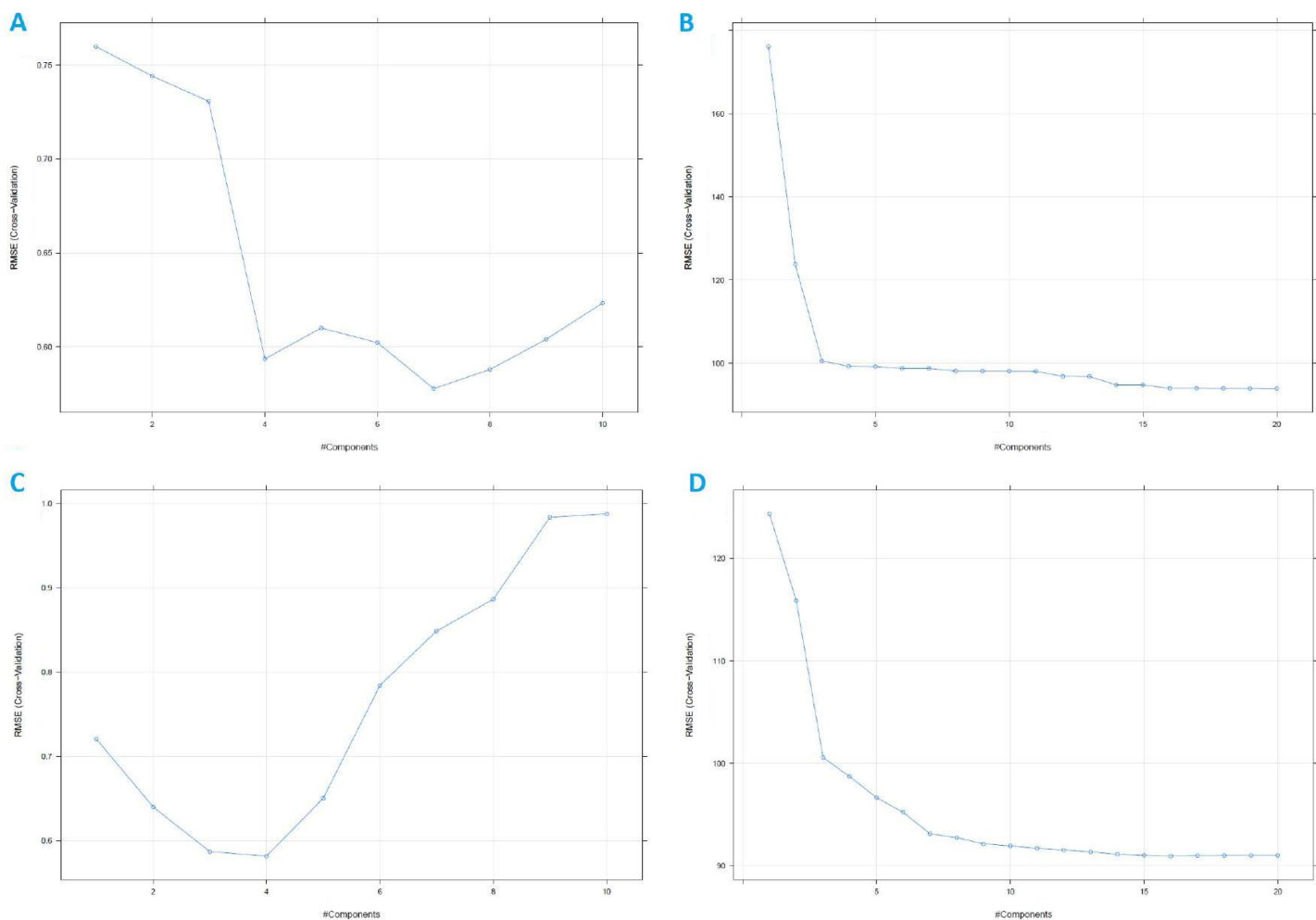
Figure 14 Partial least squares (PLS) regression RMSE cross-validation *vs* component (A) using PCR method on standardized dataset, (B) using PCR method on augmented dataset, (C) using PLS method on standardized dataset, and (D) using PLS method on augmented dataset (for Fructose, Maltose and Sucrose; see Materials–S3). Full-size ⬛ DOI: 10.7717/peerj.12186/fig-14

Table 3 Comparison of accuracy.

Fold	Sugars (%)	Organics extracts (%)	Sugars and organics extracts combined (%)
1	92.80	94.22	99.52
2	93.80	95.04	99.54
3	93.05	94.51	99.69
4	92.99	94.56	99.57
5	93.44	94.53	99.55
6	93.54	94.39	99.64
Mean	93.27	94.54	99.59

Confusion Matrix
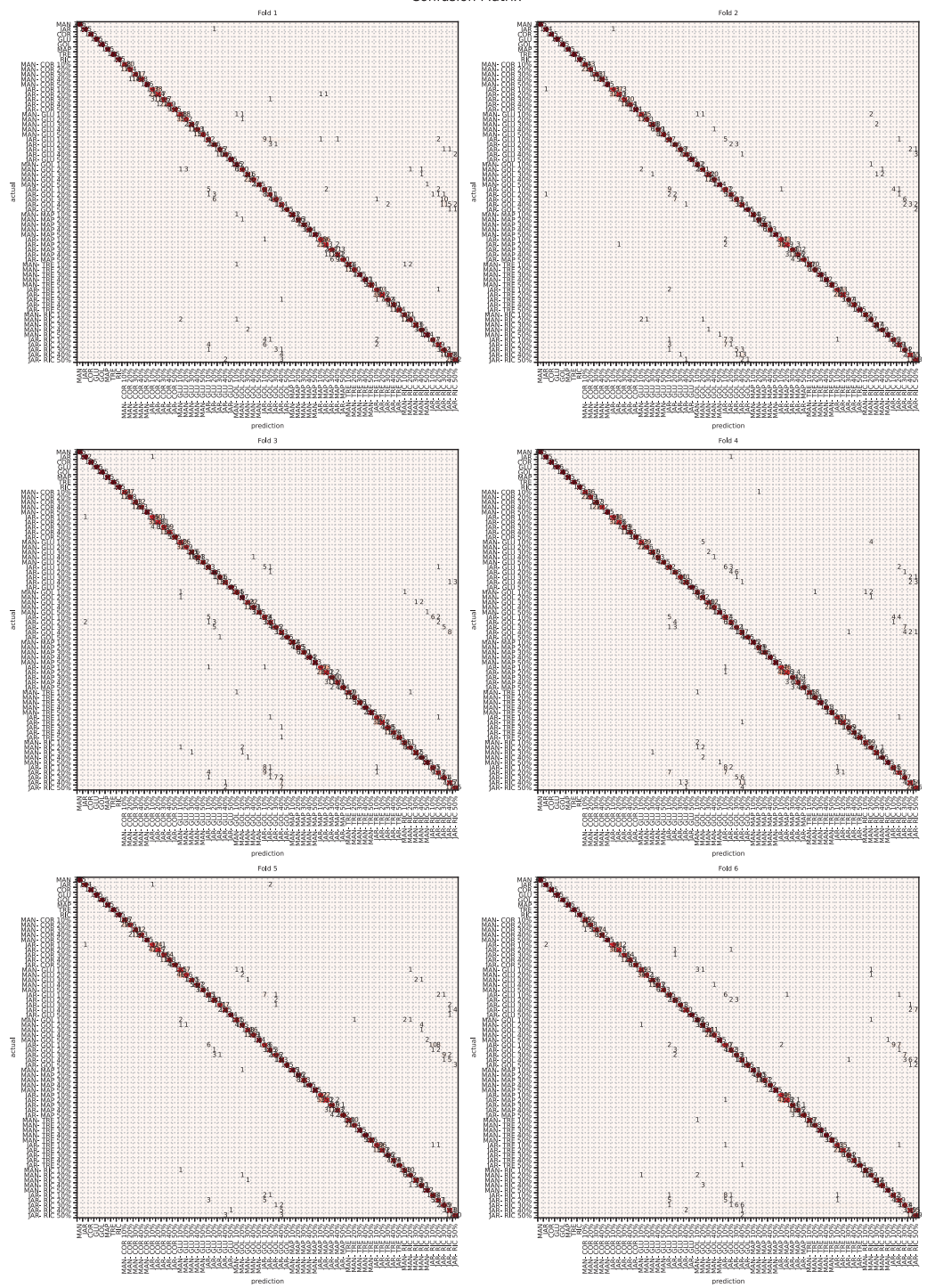
Figure 15 Confusion matrix for organic extracts Rf *vs* intensity values.
Full-size ☑ DOI: 10.7717/peerj.12186/fig-15
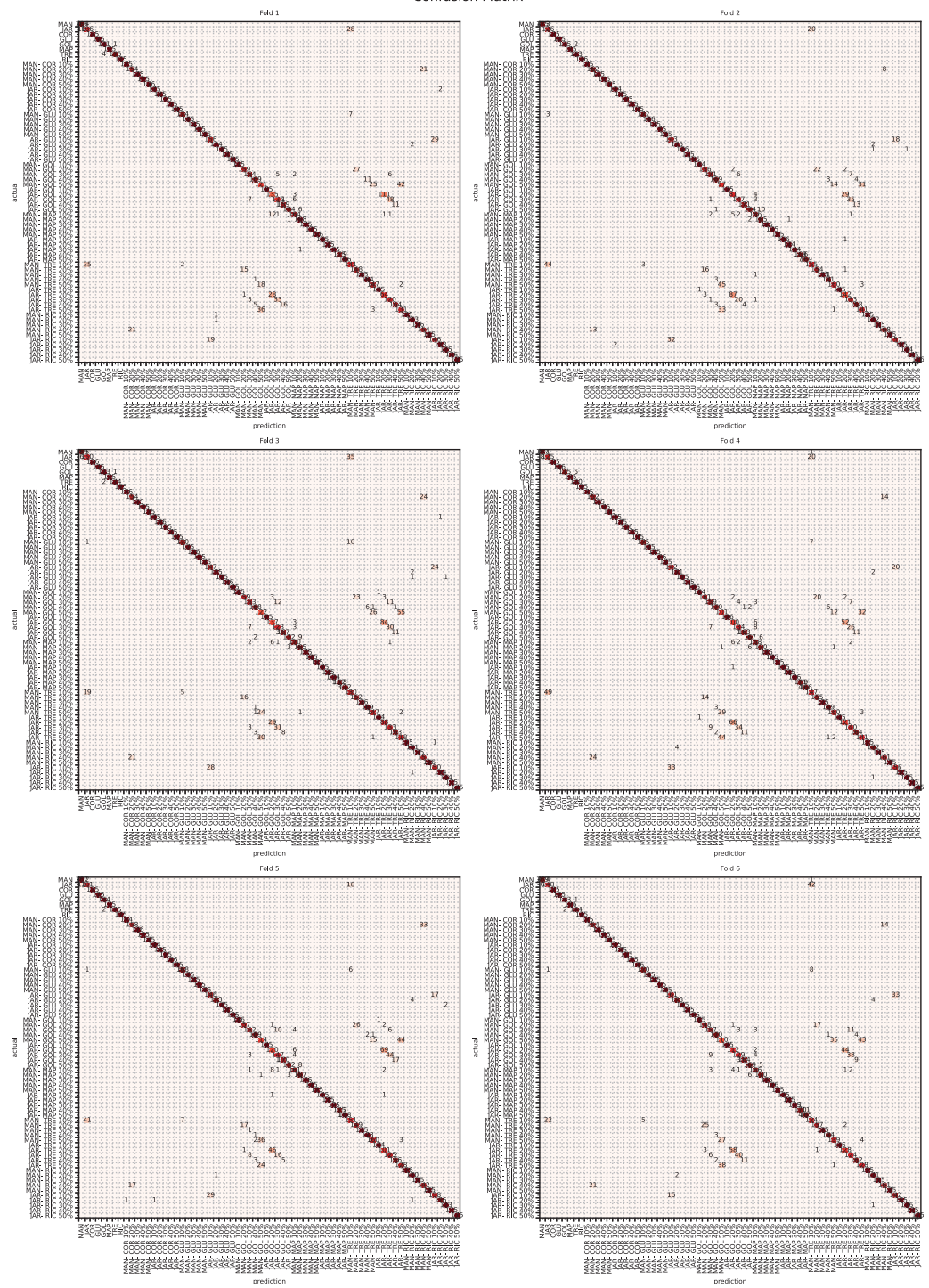
Figure 16 Confusion matrix for sugar content values. Full-size ☒ DOI: 10.7717/peerj.12186/fig-16
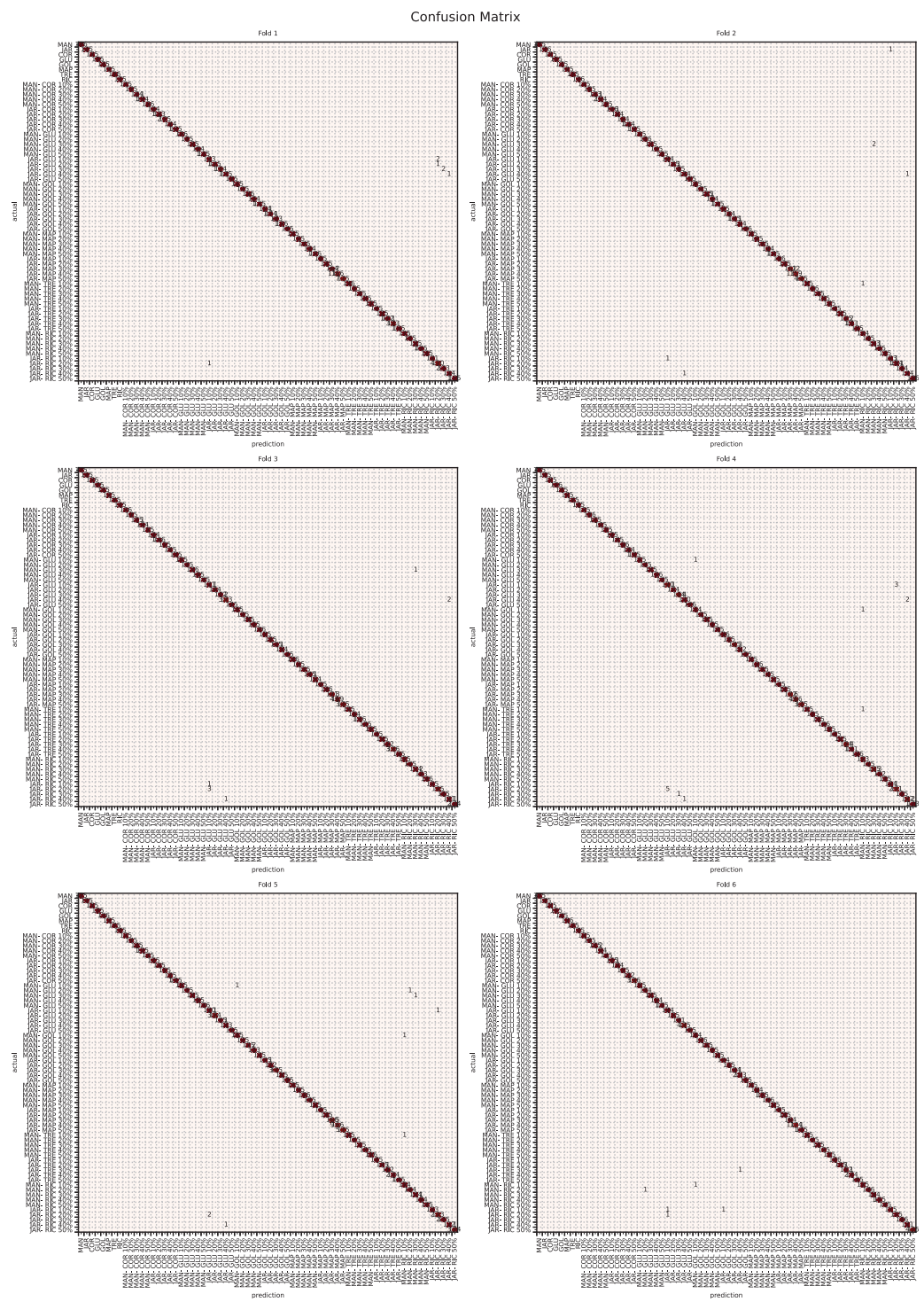
Figure 17 Confusion matrix for sugar content and organic extracts Rf *vs* intensity values learnt together. Full-size ☑ DOI: 10.7717/peerj.12186/fig-17

The augmented data was split into six groups to perform K-fold learning with six folds. For each of the six runs a different fold of the data was retained for testing as discussed in "Chemometric validation". Table 3 shows the final accuracy for each fold for sugars only, organic extracts only and sugars and organic extract combined. Individually the sugars and organic extracts are 93.27% and 94.54% accurate. When combined they reach 99.59% accuracy. The confusion matrixes for the Rf are given in Figs. 15–17. These show that main areas of confusion are the adulterants. A small number of MAN-TRE 10% are misclassified as either JAR or MAN when working with the sugars alone. The organic extracts and sugars plus organic extracts only misclassify the adulterants.

CONCLUSION

HPTLC derived organic extract profiles allow a reliable and reproducible authentication of a honey's floral source (*Locher, Neumann & Sostaric, 2017*; *Locher et al., 2018*; *Stanek & Jasicka-Misiak, 2018*; *Stanek, Kafarski & Jasicka-Misiak, 2019*), but any accidental or deliberate post-harvest adulteration with sugar syrups cannot be detected with this method alone. It requires an additional layer of analysis, the detection and quantification of simple sugars (glucose, fructose, sucrose, maltose) present in the sample. As has been demonstrated in this study, on the basis of both, HPTLC-derived honey organic extract and sugar profiles, multivariate data analysis allows for a definitive discrimination between pure and sugar syrup adulterated honeys. Cluster Analysis and Principal Component Analysis can easily cluster pure honeys, adulterated honeys and syrup adulterants into separate groups. The use of Partial Least Squares regression and the Artificial Neural Network model can successfully predict the outcome and correctly identify adulterated honeys and the percentage of adulteration based on the available data set of pure honeys and adulterated samples. Specifically, it was found that predictions based on the analysis of the samples' sugar profiles and the samples' organic extract profiles were 93.27% and 94.54% accurate respectively. When the combined data sets (sugar and organic extract profiles) were taken into account, the predictive capacity of ANN exceed 99% accuracy even for samples with post-harvest sugar adulterations of as low as 10% (w/w). This novel approach of combining HPTLC derived organic extract and sugar profile data and subjecting them to multivariate data analysis might therefore offer a powerful tool for the detection of post-harvest sugar syrup adulterants in the quality control of honeys.

ACKNOWLEDGEMENTS

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

Author Contributions

- Md Khairul Islam conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Kevin Vinsen conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Tomislav Sostaric performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Lee Yong Lim conceived and designed the experiments, authored or reviewed drafts of the paper, supervision, and approved the final draft.
- Cornelia Locher conceived and designed the experiments, authored or reviewed drafts of the paper, supervision, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:
   The raw data are available in a Supplementary File.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.12186#supplemental-information.

REFERENCES

Abdel-Aal ESM, Ziena HM, Youssef MM. 1993. Adulteration of honey with high-fructose corn syrup: detection by different methods. *Food Chemistry* **48(2)**:209–212 DOI 10.1016/0308-8146(93)90061-J.

Agatonovic-Kustrin S, Beresford R. 2000. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis* **22(5)**:717–727 DOI 10.1016/S0731-7085(99)00272-1.

Amiry S, Esmaiili M, Alizadeh M. 2017. Classification of adulterated honeys by multivariate analysis. *Food Chemistry* **224(15)**:390–397 DOI 10.1016/j.foodchem.2016.12.025.

Arvanitoyannis IS, Chalhoub C, Gotsiou P, Lydakis-Simantiris N, Kefalas P. 2005. Novel quality control methods in conjunction with chemometrics (multivariate analysis) for detecting honey

authenticity. *Critical Reviews in Food Science and Nutrition* **45(3)**:193–203
DOI 10.1080/10408690590956369.

Başar B, Özdemir D. 2018. Determination of honey adulteration with beet sugar and corn syrup using infrared spectroscopy and genetic-algorithm-based multivariate calibration. *Journal of the Science of Food and Agriculture* **98(15)**:5616–5624 DOI 10.1002/jsfa.9105.

Bentabol Manzanares A, Hernández García Z, Rodríguez Galdón B, Rodríguez-Rodríguez EM, Díaz Romero C. 2017. Physicochemical characteristics and pollen spectrum of monofloral honeys from Tenerife. *Spain Food Chemistry* **228(2)**:441–446
DOI 10.1016/j.foodchem.2017.01.150.

Bertelli D, Lolli M, Papotti G, Bortolotti L, Serra G, Plessi M. 2010. Detection of honey adulteration by sugar syrups using one-dimensional and two-dimensional high-resolution nuclear magnetic resonance. *Journal of Agricultural and Food Chemistry* **58(15)**:8495–8501 DOI 10.1021/jf101460t.

Birant D, Kut A. 2007. ST-DBSCAN: an algorithm for clustering spatial-temporal data. *Data & Knowledge Engineering* **60(1)**:208–221 DOI 10.1016/j.datak.2006.01.013.

Chen Q, Qi S, Li H, Han X, Ouyang Q, Zhao J. 2014. Determination of rice syrup adulterant concentration in honey using three-dimensional fluorescence spectra and multivariate calibrations. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **131**:177–182 DOI 10.1016/j.saa.2014.04.071.

Ciucure CT, Geană EI. 2019. Phenolic compounds profile and biochemical properties of honeys in relationship to the honey floral sources. *Phytochemical Analysis* **30(4)**:481–492
DOI 10.1002/pca.2831.

Corvucci F, Nobili L, Melucci D, Grillenzoni FV. 2015. The discrimination of honey origin using melissopalynology and Raman spectroscopy techniques coupled with multivariate analysis. *Food Chemistry* **169(2)**:297–304 DOI 10.1016/j.foodchem.2014.07.122.

Donarski JA, Jones SA, Charlton AJ. 2008. Application of cryoprobe 1H nuclear magnetic resonance spectroscopy and multivariate analysis for the verification of corsican honey. *Journal of Agricultural and Food Chemistry* **56(14)**:5451–5456 DOI 10.1021/jf072402x.

Du B, Wu L, Xue X, Chen L, Li Y, Zhao J, Cao W. 2015. Rapid screening of multiclass syrup adulterants in honey by ultrahigh-performance liquid chromatography/quadrupole time of flight mass spectrometry. *Journal of Agricultural and Food Chemistry* **63(29)**:6614–6623 DOI 10.1021/acs.jafc.5b01410.

Dżugan M, Tomczyk M, Sowa P, Grabek-Lejko D. 2018. Antioxidant activity as biomarker of honey variety. *Molecules* **23(8)**:2069 DOI 10.3390/molecules23082069.

Ferreiro-González M, Espada-Bellido E, Guillén-Cueto L, Palma M, Barroso CG, Barbero GF. 2018. Rapid quantification of honey adulteration by visible-near infrared spectroscopy combined with chemometrics. *Talanta* **188**:288–292 DOI 10.1016/j.talanta.2018.05.095.

Fiamegos Y, Dumitrascu C, Ghidotti M, de la Calle Guntiñas MB. 2020. Use of energy-dispersive X-ray fluorescence combined with chemometric modelling to classify honey according to botanical variety and geographical origin. *Analytical and Bioanalytical Chemistry* **412(2)**:463–472 DOI 10.1007/s00216-019-02255-6.

Fränti P, Rezaei M, Zhao Q. 2014. Centroid index: cluster level similarity measure. *Pattern Recognition* **47(9)**:3034–3045 DOI 10.1016/j.patcog.2014.03.017.

Ghosh S, Dubey SK. 2013. Comparative analysis of k-means and fuzzy c-means algorithms. *International Journal of Advanced Computer Science and Applications* **4**:35–39
DOI 10.14569/IJACSA.2013.040406.

Grembecka M, Szefer P. 2013. Evaluation of honeys and bee products quality based on their mineral composition using multivariate techniques. *Environmental Monitoring and Assessment* **185(5)**:4033–4047 DOI 10.1007/s10661-012-2847-y.

He C, Liu Y, Liu H, Zheng X, Shen G, Feng J. 2020. Compositional identification and authentication of Chinese honeys by (1)H NMR combined with multivariate analysis. *Food Research International (Ottawa, Ont)* **130(1)**:108936 DOI 10.1016/j.foodres.2019.108936.

Hennessy S, Downey G, O'Donnell CP. 2010. Attempted confirmation of the provenance of Corsican PDO honey using FT-IR spectroscopy and multivariate data analysis. *Journal of Agricultural and Food Chemistry* **58**:9401–9406 DOI 10.1021/jf101500n.

Islam MK, Sostaric T, Lim LY, Hammer K, Locher C. 2020a. Sugar profiling of honeys for authentication and detection of adulterants using high-performance thin layer chromatography. *Molecules* **25**:5289 DOI 10.3390/molecules25225289.

Islam MK, Sostaric T, Lim LY, Hammer K, Locher C. 2020b. A validated method for the quantitative determination of sugars in honey using high-performance thin-layer chromatography. *JPC–Journal of Planar Chromatography–Modern TLC* **33(5)**:489–499 DOI 10.1007/s00764-020-00054-9.

Islam MK, Sostaric T, Lim LY, Hammer K, Locher C. 2021. Development of an HPTLC-based dynamic reference standard for the analysis of complex natural products using Jarrah honey as test sample. *PLOS ONE* **16(7)**:e0254857 DOI 10.1371/journal.pone.0254857.

Jamróz MK, Paradowska K, Zawada K, Makarova K, Kaźmierski S, Wawer I. 2014. ^{1}H and ^{13}C NMR-based sugar profiling with chemometric analysis and antioxidant activity of herbhoneys and honeys. *Journal of the Science of Food and Agriculture* **94**:246–255 DOI 10.1002/jsfa.6241.

Janrao P, Mishra D, Bharadi V. 2019. Performance evaluation of principal component analysis for clustering on sugarcane dataset. In: *2019 International Conference on Advances in Computing, Communication and Control (ICAC3)*. 1–5.

Kečkeš J, Trifković J, Andrić F, Jovetić M, Tešić Z, Milojković-Opsenica D. 2013. Amino acids profile of Serbian unifloral honeys. *Journal of the Science of Food and Agriculture* **93(13)**:3368–3376 DOI 10.1002/jsfa.6187.

Kuropatnicki AK, Kłósek M, Kucharzewski M. 2018. Honey as medicine: historical perspectives. *Journal of Apicultural Research* **57(1)**:113–118 DOI 10.1080/00218839.2017.1411182.

Li T, Tian R, Yu X, Sun L, He Y, Xie P, Ma S. 2019. Application of chemometric algorithms in the high-performance thin-layer chromatography fingerprint of traditional chinese medicines. *Journal of AOAC International* **102(3)**:720–725 DOI 10.5740/jaoacint.18-0306.

Locher C, Neumann J, Sostaric T. 2017. Authentication of honeys of different floral origins via high-performance thin-layer chromatographic fingerprinting. *JPC-Journal of Planar Chromatography-Modern TLC* **30(1)**:57–62 DOI 10.1556/1006.2017.30.1.8.

Locher C, Tang E, Neumann J, Sostaric T. 2018. High-performance thin-layer chromatography profiling of Jarrah and Manuka honeys. *JPC-Journal of Planar Chromatography-Modern TLC* **31(3)**:181–189 DOI 10.1556/1006.2018.31.3.1.

Lolli M, Bertelli D, Plessi M, Sabatini AG, Restani C. 2008. Classification of Italian honeys by 2D HR-NMR. *Journal of Agricultural and Food Chemistry* **56(4)**:1298–1304 DOI 10.1021/jf072763c.

Mevik B-H. 2007. The pls package: principal component and partial least squares regression in R. *Journal of Statistical Software* **18(2)**:1–23 DOI 10.18637/jss.v018.i02.

Mevik B-H, Wehrens R. 2015. Introduction to the pls package. In: *Help Section of The Pls Package of R Studio Software*. Vienna, Austria: R Foundation for Statistical Computing, 1–23.

Mishra S, Kamboj U, Kaur H, Kapur P. 2010. Detection of jaggery syrup in honey using near-infrared spectroscopy. *International Journal of Food Sciences and Nutrition* **61(3)**:306–315 DOI 10.3109/09637480903476415.

Nespeca MG, Vieira AL, Júnior DS, Neto JAG, Ferreira EC. 2020. Detection and quantification of adulterants in honey by LIBS. *Food Chemistry* **311(2)**:125886 DOI 10.1016/j.foodchem.2019.125886.

Olaitan PB, Adeleke OE, Ola IO. 2007. Honey: a reservoir for microorganisms and an inhibitory agent for microbes. *African Health Sciences* **7**:159–165 DOI 10.5555/afhs.2007.7.3.159.

Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L. 2019. Pytorch: an imperative style, high-performance deep learning library. *Available at* https://arxiv.org/abs/1912.01703.

Pita-Calvo C, Vázquez M. 2018. Honeydew honeys: a review on the characterization and authentication of botanical and geographical origins. *Journal of Agricultural and Food Chemistry* **66(11)**:2523–2537 DOI 10.1021/acs.jafc.7b05807.

Rios-Corripio MA, Rojas-López* M, Delgado-Macuil R. 2012. Analysis of adulteration in honey with standard sugar solutions and syrups using attenuated total reflectance-Fourier transform infrared spectroscopy and multivariate methods. *CyTA-Journal of Food* **10(2)**:119–122 DOI 10.1080/19476337.2011.596576.

Riswahyuli Y, Rohman A, Setyabudi F, Raharjo S. 2020. Indonesian wild honey authenticity analysis using attenuated total reflectance-fourier transform infrared (ATR-FTIR) spectroscopy combined with multivariate statistical techniques. *Heliyon* **6(4)**:e03662 DOI 10.1016/j.heliyon.2020.e03662.

Rodríguez-Flores MS, Escuredo O, Míguez M, Seijo MC. 2019. Differentiation of oak honeydew and chestnut honeys from the same geographical origin using chemometric methods. *Food Chemistry* **297(2)**:124979 DOI 10.1016/j.foodchem.2019.124979.

Ruiz C, Spiliopoulou M, Menasalvas E. 2007. *C-DBSCAN: density-based clustering with constraints*. Berlin, Heidelberg: Springer Berlin Heidelberg, 216–223.

Sant'Ana LD, Sousa JP, Salgueiro FB, Lorenzon MC, Castro RN. 2012. Characterization of monofloral honeys with multivariate analysis of their chemical profile and antioxidant activity. *Journal of Food Science* **77(1)**:135–140 DOI 10.1111/j.1750-3841.2011.02490.x.

Scholz M, Quinhone Júnior A, Delamuta BH, Nakamura JM, Baudraz MC, Reis MO, Kato T, Pedrão MR, Dias LF, Dos Santos DTR, Kitzberger CSG, Bianchini FP. 2020. Indication of the geographical origin of honey using its physicochemical characteristics and multivariate analysis. *Journal of Food Science and Technology* **57(5)**:1896–1903 DOI 10.1007/s13197-019-04225-3.

Se KW, Wahab RA, Syed Yaacob SN, Ghoshal SK. 2019. Detection techniques for adulterants in honey: challenges and recent trends. *Journal of Food Composition and Analysis* **80(3)**:16–32 DOI 10.1016/j.jfca.2019.04.001.

Spiteri M, Jamin E, Thomas F, Rebours A, Lees M, Rogers KM, Rutledge DN. 2015. Fast and global authenticity screening of honey using 1H-NMR profiling. *Food Chemistry* **189(8)**:60–66 DOI 10.1016/j.foodchem.2014.11.099.

Stanek N, Jasicka-Misiak I. 2018. HPTLC phenolic profiles as useful tools for the authentication of honey. (Report). *Food Analytical Methods* **11(11)**:2979–2989 DOI 10.1007/s12161-018-1281-3.

Stanek N, Kafarski P, Jasicka-Misiak I. 2019. Development of a high performance thin layer chromatography method for the rapid qualification and quantification of phenolic compounds and abscisic acid in honeys. *Journal of Chromatography A* **1598**:209–215 DOI 10.1016/j.chroma.2019.04.052.

Tahir HE, Xiaobo Z, Xiaowei H, Jiyong S, Mariod AA. 2016. Discrimination of honeys using colorimetric sensor arrays, sensory analysis and gas chromatography techniques. *Food Chemistry* **206(2)**:37–43 DOI 10.1016/j.foodchem.2016.03.032.

Team RC. 2020a. *RStudio: integrated development environment for R.* Boston, MA: RStudio, PBC.

Team RC. 2020b. *R: a language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Tonello N, Moressi MB, Robledo SN, D'Eramo F, Marioli JM. 2016. Square wave voltammetry with multivariate calibration tools for determination of eugenol, carvacrol and thymol in honey. *Talanta* **158(11)**:306–314 DOI 10.1016/j.talanta.2016.05.071.

Truzzi C, Illuminati S, Annibaldia A, Finale C, Rossetti M, Scarponi G. 2014. Physicochemical properties of honey from Marche, Central Italy: classification of unifloral and multifloral honeys by multivariate analysis. *Natural Product Communications* **9(11)**:1595–1602 DOI 10.1177/1934578X1400901117.

Van Rossum G, Drake FL. 2009. *Python 3 reference manual.* Scotts Valley: CreateSpace.

Voica C, Iordache AM, Ionete RE. 2020. Multielemental characterization of honey using inductively coupled plasma mass spectrometry fused with chemometrics. *Journal of Mass Spectrometry* **55(7)**:e4512 DOI 10.1002/jms.4512.

Vorlová L. 2015. Adulteration of honey and available methods for detection–a review. *Acta Veterinaria Brno* **83(10)**:S85–S102 DOI 10.2754/avb201483S10S85.

Wang J, Kliks MM, Jun S, Jackson M, Li QX. 2010. Rapid analysis of glucose, fructose, sucrose, and maltose in honeys from different geographic regions using fourier transform infrared spectroscopy and multivariate analysis. *Journal of Food Science* **75(2)**:C208–C214 DOI 10.1111/j.1750-3841.2009.01504.x.

Wen YQ, Zhang J, Li Y, Chen L, Zhao W, Zhou J, Jin Y. 2017. Characterization of chinese unifloral honeys based on proline and phenolic content as markers of botanical origin, using multivariate analysis. *Molecules* **22(5)**:735 DOI 10.3390/molecules22050735.

Zhao Z, Chen L, Liu F, Zhou F, Peng J, Sun M. 2020. Fast classification of geographical origins of honey based on laser-induced breakdown spectroscopy and multivariate analysis. *Sensors (Basel)* **20**:1878 DOI 10.3390/s20071878.