metabolites

MDPI

Article

Detection of Sugar Syrups in Honey Using Untargeted Liquid Chromatography–Mass Spectrometry and Chemometrics

Jule Hansen [1], Christof Kunert [2], Kurt-Peter Raezke [2] and Stephan Seifert [1,*]

[1] Hamburg School of Food Science, Institute of Food Chemistry, University of Hamburg, Grindelallee 117, 20146 Hamburg, Germany
[2] Eurofins Food Integrity Control Services GmbH, Berliner Str. 2, 27721 Ritterhude, Germany
* Correspondence: stephan.seifert@uni-hamburg.de; Tel.: +49-40-428-38-8818

Abstract: Background: Honey is one of the most adulterated foods worldwide, and several analytical methods have been developed over the last decade to detect syrup additions to honey. These include approaches based on stable isotopes and the specific detection of individual marker compounds or foreign enzymes. Proton nuclear magnetic resonance (^{1}H-NMR) spectroscopy is applied as a rapid and comprehensive screening method, which also enables the detection of quality parameters and the analysis of the geographical and botanical origin. However, especially for the detection of foreign sugars, ^{1}H-NMR has insufficient sensitivity. **Methods**: Since untargeted liquid chromatography–mass spectrometry (LC-MS) is more sensitive, we used this approach for the detection of positive and negative ions in combination with a recently developed data processing workflow for routine laboratories based on bucketing and random forest for the detection of rice, beet and high-fructose corn syrup in honey. **Results**: We show that the distinction between pure and adulterated honey is possible for all three syrups, with classification accuracies ranging from 98 to 100%, while the accuracy of the syrup content estimation depends on the respective syrup. For rice and beet syrup, the deviations from the true proportion were in the single-digit percentage range, while for high-fructose corn syrup they were much higher, in some cases exceeding 20%. **Conclusions**: The approach presented here is very promising for the robust and sensitive detection of syrup in honey applied in routine laboratories.

Keywords: syrup adulteration; honey; routine analysis; untargeted LC-MS; machine learning; random forest; classification; regression

check for updates

Citation: Hansen, J.; Kunert, C.; Raezke, K.-P.; Seifert, S. Detection of Sugar Syrups in Honey Using Untargeted Liquid Chromatography–Mass Spectrometry and Chemometrics. *Metabolites* **2024**, *14*, 633. https://doi.org/10.3390/metabo14110633

Academic Editors: Marina Kranjac, Carlo I.G. Tuberoso, Piotr Marek Kuś and Saša Prđun

Received: 28 October 2024
Revised: 11 November 2024
Accepted: 14 November 2024
Published: 16 November 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Honey is one of the most adulterated foods in the world, adulterated through false labelling of the geographical or botanical origin, illegal processing and the addition of sugar syrup [1,2]. The latter is economically motivated by the high price of honey compared to syrups made from cheaper sources such as rice or beets, as producers can increase their profits significantly even by adding only small amounts of syrup. However, in the European Union (EU), it is prohibited to add any food ingredient, including food additives or other additions to honey (Honey Directive, Council Directive 2001/110/EC relating to honey, Annex II), and the EU coordinated action From the Hives showed that a significant proportion of the honey imported into the EU is suspicious or has been found not to comply with the provisions of the Honey Directive [3]. Sensitive and reliable analytical methods are therefore required to detect syrup adulteration in honey to ensure compliance with regulations and the trustworthiness of beekeepers.

Currently, syrup in honey is detected by commercial and official laboratories using ^{13}C/^{12}C stable carbon isotope ratio mass spectrometer coupled with elemental analyzer liquid chromatography (^{13}C-EA/LC-IRMS) [4], untargeted proton nuclear magnetic resonance (^{1}H-NMR) [5] and targeted Liquid Chromatography–Mass Spectrometry (LC-MS).

In the latter, known marker molecules defined for specific syrups are detected, which is time-consuming because extensive sample preparation and multiple analyses are required [6]. Furthermore, it is not trivial to find suitable marker molecules due to the similarity of honeys and sugar syrups and the high variance within the honey samples [7]. This diversity is due to the fact that the diet of bees is less well defined than that of other animals or plants, for example because they collect pollen and nectar within a 5 km radius of their hive [8,9]. Furthermore, the suitable markers found, which are stored in marker databases, must constantly be updated, as fraudsters are aware of the substances being tested and adapt their fraud methods accordingly [5,10]. This has also been demonstrated in a comprehensive study of economically motivated adulteration based on a large number of specific cases [11].

^{1}H-NMR is often applied untargeted, which means that the entire ^{1}H-NMR data of the samples are analysed with multivariate approaches. These can be supervised or unsupervised, i.e., the membership of the samples to the targeted grouping can be included into the analysis or not. Principal Component Analysis (PCA) is an unsupervised approach that generates latent principal components used to analyse the main variances in the data [12–14]. Supervised approaches such as Random Forest (RF) and Support Vector Machine train models based on specific differences between predefined classes [15–18], e.g., regarding pure or adulterated honey [5,19,20]. RF is a non-parametric ensemble learning algorithm based on multiple binary decision trees, which has several advantages for this application. Because each of the decision trees is trained on a different subset of the samples, called bootstrap sample, the remaining samples, called out-of-bag samples, can be used to generate an independent error, called out-of-bag (OOB) error [21]. This means that no additional data is necessary to validate the performance of the obtained model. The other advantages are its easy adaptation to imbalanced training data and the minimal need for parameter tuning [22], which is why ^{1}H-NMR in combination with RF has been applied for the authentication of various foods [18,23–27].

Even though the use of ^{1}H-NMR as an untargeted method offers the advantages of relatively simple sample preparation, the possibility of quantifying metabolites and a high degree of experimental reproducibility, this approach has the disadvantage of a rather low sensitivity [28]. This means that syrup additions are often only detected in comparatively large quantities [5]. Therefore, it would be a great asset for honey monitoring laboratories to use the more sensitive LC-MS method as an untargeted approach.

Untargeted LC-MS is already used successfully for the authentication of various foods [29–33], including honey [7,34–36]. In these applications, however, usually only samples measured within a short period of time using one single instrument [37] can be analysed together due to instrumental shifts and the resulting need to define peaks during processing [38]. The lack of robustness is primarily due to the varying age and degree of contamination of the devices and the columns used [39]. This causes a slightly different interaction between the compounds and the columns and, hence, results in measurement-based differences in the spectra if they are not obtained at the same time with the same device [40]. These circumstances are the reason why current LC-MS data processing workflows such as vendor-specific software (e.g., Compound Discoverer^{TM} Software [41]) or the framework for processing and visualizing LC-MS data xcms [42–44] include a so-called retention time alignment step. Additionally, in common data processing strategies, a correspondence step is included to match the detected and aligned chromatographic peaks between the different spectra to generate a common peak list, which is unique for each analysis. Therefore, it is currently not possible to analyse untargeted LC-MS spectra that were processed separately. For long-term application, this means an immense computational effort, as the entire data set would have to be reprocessed each time a newly analysed sample is to be classified or new authentic data are added. In routine analysis, several hundred samples are analysed every week and it is therefore currently very difficult to apply untargeted LC-MS in this field. This is why untargeted LC-MS is often used to identify marker molecules for specific groups, which are then applied in

targeted analyses [6,38,45,46]. However, to use the full potential of the large amounts of data generated, the long-term use and separate processing of the LC-MS data are necessary. We have recently developed a data processing approach, Bucketing of Untargeted LC-MS Spectra (BOULS), which makes this possible, as demonstrated by the authentication of the geographical origin of honey [47].

In this paper we evaluate LC-MS-based approaches in combination with BOULS and multivariate analysis for the detection of different syrups used in a routine laboratory for the sensitive analysis of honey. For this, we analysed 34 North German honey samples mixed with rice, beet and high-fructose corn syrup with different LC approaches. For the detection of polar and non-polar compounds, Hydrophilic Interaction Liquid Chromatography (HILIC) in negative ionization mode and Reversed Phase (RP) chromatography in positive ionization mode are used, respectively [8,35], and we evaluate whether one of these approaches should be used individually, together or in addition with fragmentation data.

2. Materials and Methods

2.1. Sample Preparation

For authentication, mainly to ensure that the North German honey samples were actually free of foreign sugars, they were analyzed using pollen microscopy (pollen identification and determination of the relative pollen content, DIN 10760 mod. (2002–05)), sensory (ICS SOP 520–02 (2018–08)), ^{13}C-EA/LC-IRMS [4], untargeted proton nuclear magnetic resonance (^{1}H-NMR) [5] and targeted LC-HRMS [35]. As foreign sugar was detected in one of the 35 samples, it was excluded from the following analysis. For the analysis of the remaining 34 honey samples, 1 g of honey sample, as well as 1 g of rice, beet and high-fructose corn syrup was dissolved in 9.8 mL of deionised water and centrifuged. Subsequently, to produce adulterated honey samples, each syrup was added to each honey, creating mixtures with 5, 10, 20, 50 and 80% v/v syrup content. The 5 mixtures of each honey and syrup, as well as samples of the pure honey and syrup were stored at −20 °C until the measurement.

2.2. LC-MS Analysis

The samples were analysed in a randomized order using three different LC-HRMS devices (Thermo Scientific™ UltiMate™ 3000 systems coupled to Thermo Scientific™ QExactive™ Hybrid Quadrupole-Orbitrap™ High-Resolution Mass Spectrometers, Thermo Scientific™, Bremen, Germany). Each sample was analysed using two different LC approaches. Due to preliminary tests, HILIC (Accucore-150-Amide-HILIC, 150 × 2.1 mm, Thermo Scientific™, Bremen, Germany) was applied in negative mode to analyse polar compounds, and non-polar compounds were analysed using RP (Hypersil Gold C18, 150 × 2.1 mm, Thermo Scientific™, Bremen, Germany) chromatography in positive mode. For both chromatographic methods, the mobile phases acetonitrile and water were used with acetic acid as modifier. The ionization was performed using electrospray ionization (ESI) and the data were acquired in profile mode. For the mass spectrometric analysis, the variable data-independent acquisition (vDIA) approach developed by Thermo Scientific™ Orbitrap™ was applied using MS/MS precursor isolation windows of differing mass ranges, covering the entire mass range of 100–1500 Da (MS1) of the preceding full scan [48]. Ions of the following masses were fragmented: 100–200, 200–300, 300–400, 400–500, 500–1000 and 1000–1500 Da and the fragments were detected in the ranges of 50–225, 50–330, 50–430, 50–535, 69–1045, and 104–1555 Da, respectively [35].

This experimental setup resulted in 4 different data sets: HILIC full-scan data, HILIC full-scan and fragment data, RP full-scan data, RP full-scan and fragment data.

2.3. Data Analysis

The data processing and analysis were carried out in R (version 4.4.1) [49] using the BOULS approach [47] (https://github.com/AGSeifert/BOULS, accessed on 11 November 2024), requires Linux OS), which is based on the xcms workflow and uses the same functions

for data import and chromatographic peak detection [50]. For data import, the thermo-specific raw files of the HILIC and RP data of each analysed sample were converted to open-format mzML files using MSConvert, which is part of the ProteoWizard software package (version 3.0.21078-7da1f1136 (developer build)), using the filter peakPicking to convert the profile data into centroided data [51]. The Bioconductor package mzR (version 2.26.0) was used for data import into R [51–54]. The package MSnbase (version 2.18.0) [55,56] was used to load and store the data in an object that was compatible with the xcms package (version 3.14.0) [42,44]. The centWave algorithm [44] was used for peak detection with the parameters peakwidth of 15 s and ppm of 5 Da. Retention time alignment was performed using the obiwarp method (bin size of 0.1 and the localAlignment set to TRUE). The data of one honey were used for retention time alignment and the 50% dilution with the respective syrup was used as center spectrum [57]. Subsequently, the data were processed using the BOULS approach with bucket sizes of 20 s in the retention time dimension and 2 Da in the mass dimension. The data were normalized by dividing the summarized intensities of each bucket by the sum of intensities of all signals, as previously determined during optimization of data processing with BOULS [47].

For multivariate data analysis, the processed HILIC and RP data were analyzed separately and after a low-level data fusion. This resulted in 6 analyses using HILIC full-scan data (42,000 variables), HILIC full-scan and fragment data (87,060 variables), RP full-scan data (67,200 variables), RP full-scan and fragment data (85,344 variables), the fused data set of HILIC and RP full-scan data (109,200 variables) and the fused data set of HILIC and RP full-scan and fragment data (172,404 variables). In this manuscript, the results of the data fusion of HILIC and RP full-scan data for the three syrups are shown. Due to the similarities, the results of the remaining data are shown in the Supplementary Materials.

For PCA, the stats package [58–60] (version 4.4.1) was used and the scatter plots were visualized using the package ggplot2 [61] (version 3.5.1). To visualize the main variance in the data, the first and second principal components were used. For RF analysis, nine honey samples were randomly selected and the respective data were used as test data for all analyses. The data of the other 25 honey samples were used for training of all RF classification and regression models using the ranger function of the R package ranger [62] (version 0.16.0). One sample (50% syrup proportion) was removed from the high-fructose corn syrup training data set, as the RP spectrum did not contain any signals, presumably due to injection problems during the measurement. For classification and regression, the number of trees (num.trees) was set to 5000 and default settings for mtry (205 for HILIC, 259 for RP and 330 for the fused data set) and min.node.size (1 for classification and 5 for regression) were chosen. As the data sets contained less data from pure honey samples than from adulterated samples, for classification, the case.weights parameter was chosen according to the size of the respective classes to compensate for class imbalance. The Root Mean Square Error (RMSE) was used to evaluate the precision of the predictions of the RF regression models (Metrics package, version 0.1.4 [63]).

3. Results and Discussion

In this study, untargeted LC-MS data in combination with multivariate data analysis were applied for analysis as in a routine laboratory to detect three different sugar syrups in honey. Various approaches for mass spectrometric analysis with several devices were used for this purpose and compared to verify the detection of syrups over a wide range of proportions and at comparatively low additions. In the following sections, the high-dimensional data were first analyzed by PCA to assess overall variances in the data. An RF classification model was then used to validate whether syrup additives can be detected qualitatively, and an RF regression model was used to estimate whether quantitative determination is feasible.

3.1. PCA

Figure 1 shows the scores plots of the first two principal components of the analysis of the fused data set of honey samples adulterated with rice (a), beet (b) and high-fructose corn syrup (c). Groupings based on the syrup content can be observed in all three plots. However, an extremely strong influence of the device used is evident, particularly in the scores of the first principal component for the samples adulterated with rice and high-fructose corn syrup and the second principal component for the beet syrup adulteration. Similar results were also obtained when the HILIC and RP data were analyzed individually or when fragment data were added (Figures S1–S5). From this it can be concluded that even after the application of the BOULS data processing approach, device-specific variations were still present in the data to a significant extent. The superposition of these variations with sample-specific variations, which result in a lack of robustness of LC-MS approaches, is generally a well-known challenge when implementing untargeted LC-MS methods in routine laboratories [39,64–66]. Therefore, as explained above, LC-MS is currently mainly used either targeted or on a single day by a single device. The processing approach used here nevertheless results in a robust data structure due to the three-dimensional bucketing, which makes it possible to analyze and compare samples over a longer period of time, independent of the processing batch [47].

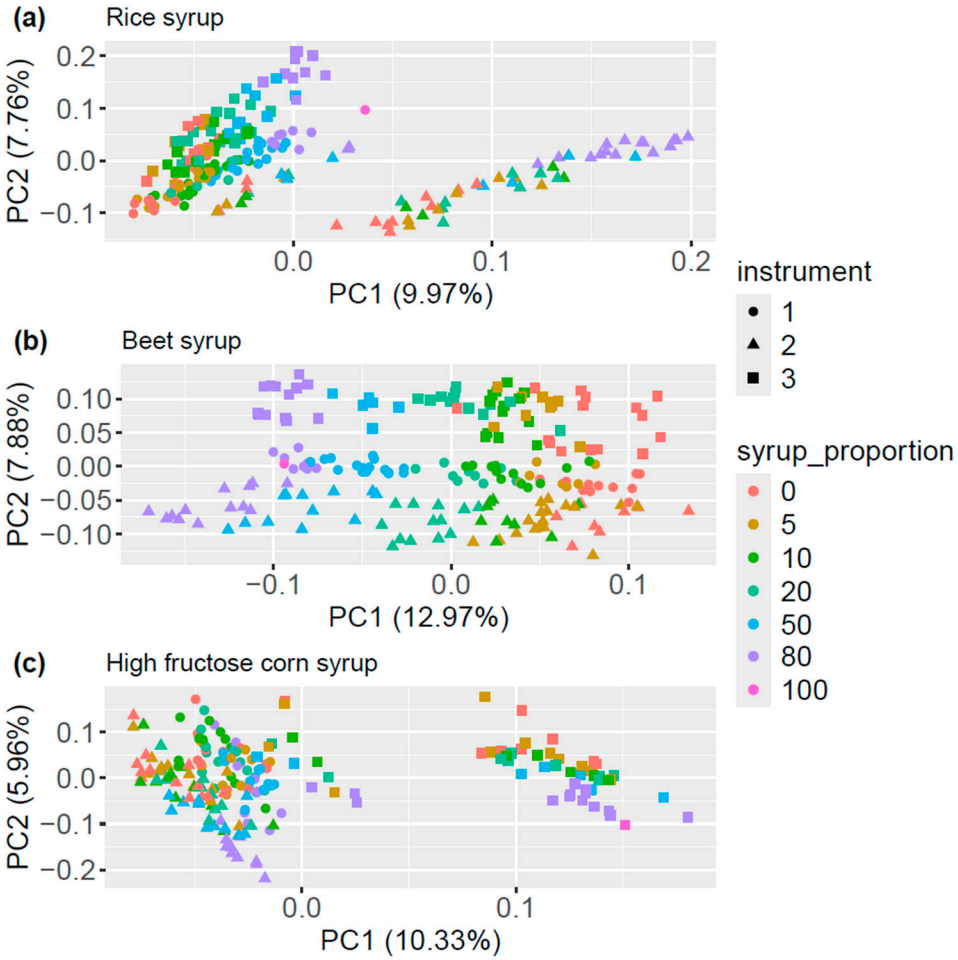


Figure 1. PCA of the fused data of the honey samples adulterated with rice (**a**), beet (**b**) and high-fructose corn syrup (**c**) labeled according to the proportion of syrup [%] (colors) and the device used (shapes).

In the following sections, we will investigate whether a supervised multivariate approach can be used to train models that are based only on the relevant variances and allow device-specific information to be ignored in the evaluation of the authenticity of honey.

3.2. Identification of Honey Samples Containing Syrup

To analyze whether honey samples adulterated with syrups can be identified using a supervised multivariate approach, three random forest models were trained to discriminate between the class of pure honey samples and an adulterated class, which included all samples with the addition of one of the syrups. The results for the fused training and test data are shown in Tables S1 and 1. It can be seen that all pure and beet syrup adulterated samples are correctly classified by the respective model, while one of the test samples adulterated with 10% beet syrup is classified incorrectly. The same classification results for the identification of beet and rice syrup are obtained when the HILIC and RP data are analyzed individually (Tables S2 and S3) and/or fragment data are added (Tables S4–S6). Since the same sample with 10% beet syrup adulteration is misclassified in all data sets but all samples with 5% adulteration are correctly classified, it is likely that the misclassification is due to a pipetting error. The results show that the addition of rice and beet syrup to honey can be clearly detected using LC-MS in combination with BOULS and the application of random forest, a supervised classification approach. For the analysis of high-fructose corn syrup, however, one pure honey sample of the test data (see Table 1) and various samples of the training data (see Table S4) are falsely classified as adulterated, which is also reflected in an increased OOB error of 9.7% for the respective model. This suggests that it is generally more difficult to distinguish between pure honey samples and those adulterated with high-fructose corn syrup than with the other two syrups. This is also supported by the fact that the results vary when analyzing the different data types, so that a larger number of samples are incorrectly assigned when using the pure HILIC data (Table S2), as well as when fragmented data are added (Table S5), than when analyzing the individual data or fused data (Tables S1, S3, S4 and S6). The reason for this is probably that high-fructose corn syrup is more similar to the sugars contained in honey. For example, glucose and fructose each make up 30–40% of high-fructose corn syrup and honey sugars [67], whereas these sugars make up only 1–10% of rice [68] and beet syrup [69]. The fact that almost all samples were correctly classified in the test data set indicates that adulterations with high-fructose corn syrup can also be detected if a sufficiently large data set is used for training. Thus, BOULS data processing combined with random forest classification results in focusing the analysis on the relevant syrup-specific, rather than device-specific, variances, as was already the case for the authentication of honey regarding the geographical origin [47]. The fact that an addition of 5% could be detected for all three syrups shows the great potential of this approach, as the analytical techniques currently used are mostly not capable of this [4,5,35]. This is particularly true for the detection of beet and rice syrups, which are currently particularly difficult to detect [4].

Table 1. Classification results of the fused data of the honey samples adulterated with rice, beet and high-fructose corn syrup. The adulterated class contained the honey samples with 5, 10, 20, 50 and 80% syrup addition. The RF classification model trained with the rice, beet and high-fructose corn syrup data set showed OOB errors of 0%, 0.7% and 9.7%, respectively, and the results of the training data are shown in Table S6.

True Syrup Portion [%]	Correct Predictions Rice Syrup [%]	Correct Predictions Beet Syrup [%]	Correct Predictions High-fructose Corn Syrup [%]
0	9/9	9/9	8/9
5	9/9	9/9	9/9
10	9/9	8/9	9/9
20	9/9	9/9	9/9
50	9/9	9/9	9/9
80	9/9	9/9	9/9

3.3. Estimation of the Syrup Content in Adulterated Honey Samples

In order to analyze whether the approach could also be used to predict the syrup content, regression models were trained for each syrup adulteration. Figure 2 shows the results for the analysis of the fused data set. The coefficients of determination show values of 0.95 and 0.97 for the analysis of adulteration with rice and beet syrup, respectively (Figure 2a,b). The fact that these values are quite close to 1 shows that there is a relevant correlation between the respective LC-MS data and the actual proportion of syrup present. This makes it possible to predict the syrup content in the test data with an error that is in the single digits, ranging from 2% to 7% for all analyzed proportions (see Figure 2 and Table S7). Again, the results are quite similar whether the HILIC and RP data are analyzed separately or whether fragment data are added (see Table S7). For the prediction of the high-fructose corn syrup proportion, models with coefficients of determination of around 0.8 are obtained (see Figures 2c and S6–S10). The correlation between the respective LC-MS data and the actual percentage of syrup is therefore much lower here than for the analysis of the other syrups. This is also reflected in the prediction of the test data, which shows larger errors ranging from 8% to 22% for the analysis of all data types (see Figure 2 and Table S7). As already surmised in the previous section, this is probably due to the greater similarity of this syrup to the sugars present in honey. In this case, the device-specific variance present in the data set is probably an obstacle to achieving a more precise determination of the syrup proportion.

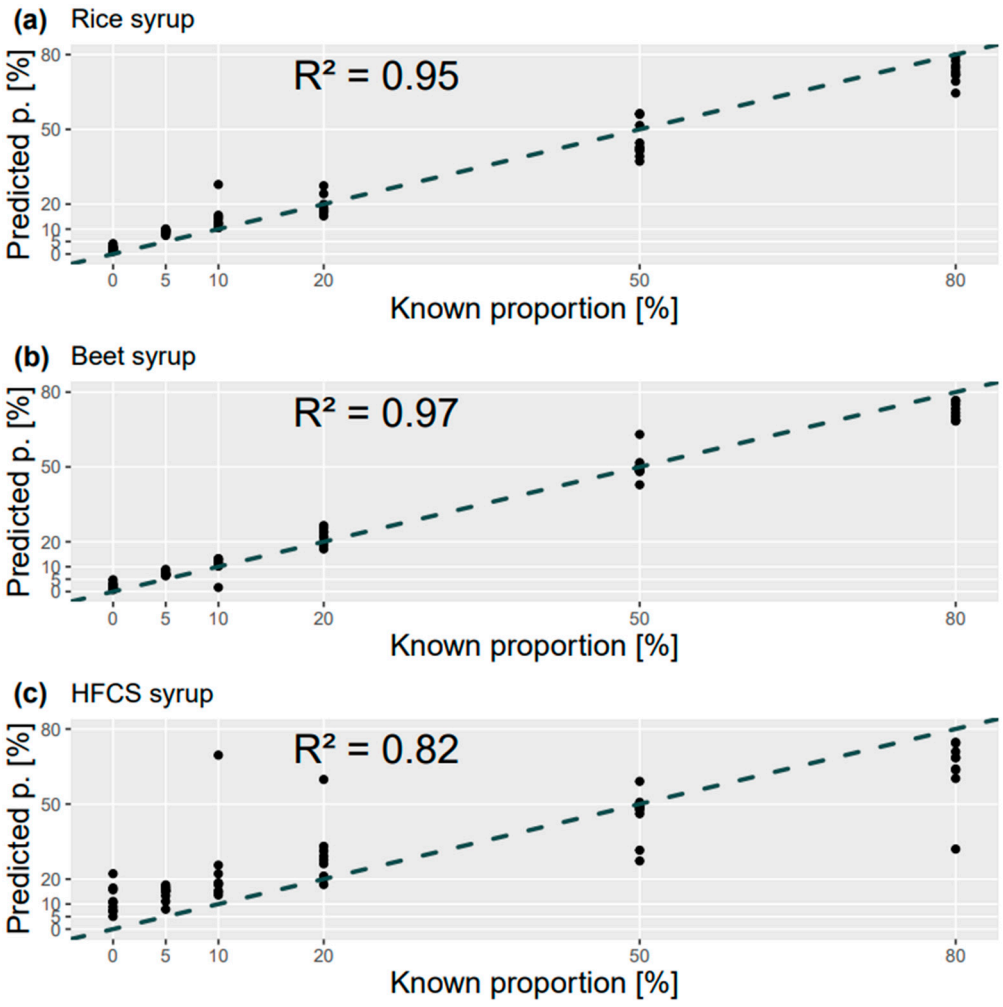


Figure 2. Results of random forest regression showing the predicted proportions of test samples with rice (**a**), beet (**b**) and high-fructose corn syrup (**c**) adulteration plotted against the known proportions.

Overall, the results show that when analysing LC-MS data of honey samples with BOULS, device-specific information is retained in the data, but its influence can be minimised by analysing the data with a supervised multivariate method. This enables the development of a simple and fast approach, as no re-evaluation of the entire data and no batch correction [70,71], feature identification [72] or feature matching [73] are required for the detection of syrup additives in honey. This can be based on HILIC or RP data or on a fused data set of both approaches for the detection of rice and beet syrup. An addition of fragment data was not necessary here, as no significant improvement in syrup detection was achieved. However, if the approach is used to analyse the addition of a syrup with a honey-like sugar composition, it is possible to identify the addition but it is much more difficult to determine the exact proportion, presumably due to instrumental variations in the data. In this case, it may be useful to combine different mass spectrometric data for a better prediction or even combine LC-MS data with other analytical techniques such as ^{13}C-EA/LC-IRMS, which is particularly suitable for C4 sugars like high-fructose corn syrup [4].

4. Conclusions and Outlook

We showed that the combination of data processing using BOULS and the application of a supervised multivariate method is very promising for robust and sensitive detection of syrup in honey based on LC-MS, which could be applied in private or state laboratories. In order to establish this approach for routine analysis, the application presented here, a proof-of-concept study using North German honeys, should be further developed for application of samples with higher variance and adulteration with other syrups, e.g., tailored syrups whose compositions are specifically adapted to the composition of the honey that is to be adulterated. In addition, other food samples should also be analyzed, as this approach promises to be robust and sensitive to a wide range of authentication issues.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/metabo14110633/s1: Figure S1. PCA of the fused data set of the HILIC and RP full-scan and fragment data of the honey samples adulterated with rice (a), beet (b) and high-fructose corn syrup (c) labeled according to the proportion of syrup [%] (colors) and the device used (shapes). Figure S2. PCA of the HILIC full-scan data of the honey samples adulterated with rice (a), beet (b) and high-fructose corn syrup (c) labeled according to the proportion of syrup [%] (colors) and the device used (shapes). Figure S3. PCA of the RP full-scan data of the honey samples adulterated with rice (a), beet (b) and high-fructose corn syrup (c) labeled according to the proportion of syrup [%] (colors) and the device used (shapes). Figure S4. PCA of the HILIC full-scan and fragment data of the honey samples adulterated with rice (a), beet (b) and high-fructose corn syrup (c) labeled according to the proportion of syrup [%] (colors) and the device used (shapes). Figure S5. PCA of the RP full-scan and fragment data of the honey samples adulterated with rice (a), beet (b) and high-fructose corn syrup (c) labeled according to the proportion of syrup [%] (colors) and the device used (shapes). Figure S6. Results of random forest regression for fused full-scan and fragment data showing the predicted proportions of test samples with rice (a), beet (b) and high-fructose corn syrup (c) adulteration plotted against the known proportions. Figure S7. Results of random forest regression for HILIC full-scan data showing the predicted proportions of test samples with rice (a), beet (b) and high-fructose corn syrup (c) adulteration plotted against the known proportions. Figure S8. Results of random forest regression for RP full-scan data showing the predicted proportions of test samples with rice (a), beet (b) and high-fructose corn syrup (c) adulteration plotted against the known proportions. Figure S9. Results of random forest regression for HILIC full-scan and fragment data showing the predicted proportions of test samples with rice (a), beet (b) and high-fructose corn syrup (c) adulteration plotted against the known proportions. Figure S10. Results of random forest regression for RP full-scan and fragment data showing the predicted proportions of test samples with rice (a), beet (b) and high-fructose corn syrup (c) adulteration plotted against the known proportions. Table S1. Classification and regression results of the analysis of the fused full-scan and fragment data of the honey samples adulterated with rice, beet and high-fructose corn syrup. For the classification, the adulterated class contained all of the samples with 5, 10, 20, 50 and 80% addition of the respective syrup (p = pure; a = adulterated; HFCS = high-fructose corn syrup). Table S2. Classification and

regression results of the analysis of the fused full-scan data of the honey samples adulterated with rice, beet and high-fructose corn syrup. For the classification, the adulterated class contained all of the samples with 5, 10, 20, 50 and 80% addition of the respective syrup (p = pure; a = adulterated; HFCS = high-fructose corn syrup). Table S3. Classification and regression results of the analysis of the HILIC full-scan data of the honey samples adulterated with rice, beet and high-fructose corn syrup. For the classification, the adulterated class contained all of the samples with 5, 10, 20, 50 and 80% addition of the respective syrup (p = pure; a = adulterated; HFCS = high-fructose corn syrup). Table S4. Classification and regression results of the analysis of the RP full-scan data of the honey samples adulterated with rice, beet and high-fructose corn syrup. For the classification, the adulterated class contained all of the samples with 5, 10, 20, 50 and 80% addition of the respective syrup (p = pure; a = adulterated; HFCS = high-fructose corn syrup). Table S5. Classification and regression results of the analysis of the HILIC full-scan and fragment data of the honey samples adulterated with rice, beet and high-fructose corn syrups. For the classification, the adulterated class contained all of the samples with 5, 10, 20, 50 and 80% addition of the respective syrup (p = pure; a = adulterated). Table S6. Classification and regression results of the analysis of the RP full-scan and fragment data of the honey samples adulterated with rice, beet and high-fructose corn syrup. For the classification, the adulterated class contained all of the samples with 5, 10, 20, 50 and 80% addition of the respective syrup (a = adulterated; p = pure; HFCS = high-fructose corn syrup). Table S7. RMSE for the prediction of the proportions of rice, beet and high-fructose corn syrup in honey samples by the random forest regression models.

Author Contributions: Conceptualization, J.H., C.K., K.-P.R. and S.S.; methodology, J.H. and S.S.; software, J.H.; validation, J.H. and S.S.; formal analysis, J.H.; investigation, J.H. and C.K.; resources, S.S., K.-P.R. and C.K.; data curation, J.H.; writing—original draft preparation, J.H.; writing—review and editing, S.S. and C.K.; visualization, J.H.; supervision, S.S. and C.K.; project administration, J.H.; funding acquisition, S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data sets analyzed during the current study are not publicly available, as they were generated from samples from customers of a commercial laboratory. However, the BOULS approach is published in an R package here: https://github.com/AGSeifert/BOULS (accessed on 11 November 2024) and example data are provided here: https://www.fdr.uni-hamburg.de/record/13583 (accessed on 11 November 2024).

Conflicts of Interest: Authors Christof Kunert and Kurt-Peter Raezke were employed by the company Eurofins Food Integrity Control Services GmbH. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Cozzolino, D.; Dayananda, B.; Chapman, J. Food Adulteration. In *Chemometrics*; Elsevier: Amsterdam, The Netherlands, 2024; pp. 273–291, ISBN 978-0-443-21493-6.
2. Moore, J.C.; Spink, J.; Lipp, M. Development and Application of a Database of Food Ingredient Fraud and Economically Motivated Adulteration from 1980 to 2010. *J. Food Sci.* **2012**, *77*, R118–R126. [CrossRef] [PubMed]
3. European Commission. *EU Coordinated Action "From the Hives" (Honey 2021–2022)*; Brussels, Belgium. 2022. Available online: https://food.ec.europa.eu/safety/eu-agri-food-fraud-network/eu-coordinated-actions/honey-2021-2022_en (accessed on 11 November 2024).
4. Elflein, L.; Raezke, K.-P. Improved Detection of Honey Adulteration by Measuring Differences between $^{13}C/^{12}C$ Stable Carbon Isotope Ratios of Protein and Sugar Compounds with a Combination of Elemental Analyzer—Isotope Ratio Mass Spectrometry and Liquid Chromatography—Isotope Ratio Mass Spectrometry ($\delta^{13}C$-EA/LC-IRMS). *Apidologie* **2008**, *39*, 574–587. [CrossRef]
5. Spiteri, M.; Jamin, E.; Thomas, F.; Rebours, A.; Lees, M.; Rogers, K.M.; Rutledge, D.N. Fast and Global Authenticity Screening of Honey Using ^{1}H-NMR Profiling. *Food Chem.* **2015**, *189*, 60–66. [CrossRef] [PubMed]
6. Esslinger, S.; Riedl, J.; Fauhl-Hassek, C. Potential and Limitations of Non-Targeted Fingerprinting for Authentication of Food in Official Control. *Food Res. Int.* **2014**, *60*, 189–204. [CrossRef]
7. Martinello, M.; Stella, R.; Baggio, A.; Biancotto, G.; Mutinelli, F. LC-HRMS-Based Non-Targeted Metabolomics for the Assessment of Honey Adulteration with Sugar Syrups: A Preliminary Study. *Metabolites* **2022**, *12*, 985. [CrossRef]

8. Tsagkaris, A.S.; Koulis, G.A.; Danezis, G.P.; Martakos, I.; Dasenaki, M.; Georgiou, C.A.; Thomaidis, N.S. Honey Authenticity: Analytical Techniques, State of the Art and Challenges. *RSC Adv.* **2021**, *11*, 11273–11294. [CrossRef]
9. Magdas, D.A.; Guyon, F.; Puscas, R.; Vigouroux, A.; Gaillard, L.; Dehelean, A.; Feher, I.; Cristea, G. Applications of Emerging Stable Isotopes and Elemental Markers for Geographical and Varietal Recognition of Romanian and French Honeys. *Food Chem.* **2021**, *334*, 127599. [CrossRef]
10. Gao, B.; Holroyd, S.E.; Moore, J.C.; Laurvick, K.; Gendel, S.M.; Xie, Z. Opportunities and Challenges Using Non-Targeted Methods for Food Fraud Detection. *J. Agric. Food Chem.* **2019**, *67*, 8425–8430. [CrossRef]
11. Everstine, K.; Spink, J.; Kennedy, S. Economically Motivated Adulteration (EMA) of Food: Common Characteristics of EMA Incidents. *J. Food Prot.* **2013**, *76*, 723–735. [CrossRef]
12. Abdi, H.; Williams, L.J. Principal Component Analysis. *WIREs Comput. Stats* **2010**, *2*, 433–459. [CrossRef]
13. Ehlers, M.; Horn, B.; Raeke, J.; Fauhl-Hassek, C.; Hermann, A.; Brockmeyer, J.; Riedl, J. Towards Harmonization of Non-Targeted ^{1}H NMR Spectroscopy-Based Wine Authentication: Instrument Comparison. *Food Control* **2022**, *132*, 108508. [CrossRef]
14. Sobolev, A.P.; Thomas, F.; Donarski, J.; Ingallina, C.; Circi, S.; Cesare Marincola, F.; Capitani, D.; Mannina, L. Use of NMR Applications to Tackle Future Food Fraud Issues. *Trends Food Sci. Technol.* **2019**, *91*, 347–353. [CrossRef]
15. Truzzi, E.; Marchetti, L.; Piazza, D.V.; Bertelli, D. Multivariate Statistical Models for the Authentication of Traditional Balsamic Vinegar of Modena and Balsamic Vinegar of Modena on ^{1}H-NMR Data: Comparison of Targeted and Untargeted Approaches. *Foods* **2023**, *12*, 1467. [CrossRef]
16. Huo, Y.; Kamal, G.M.; Wang, J.; Liu, H.; Zhang, G.; Hu, Z.; Anwar, F.; Du, H. 1 H NMR-Based Metabolomics for Discrimination of Rice from Different Geographical Origins of China. *J. Cereal Sci.* **2017**, *76*, 243–252. [CrossRef]
17. Jiménez-Carvelo, A.M.; González-Casado, A.; Bagur-González, M.G.; Cuadros-Rodríguez, L. Alternative Data Mining/Machine Learning Methods for the Analytical Evaluation of Food Quality and Authenticity—A Review. *Food Res. Int.* **2019**, *122*, 25–39. [CrossRef]
18. Wenck, S.; Mix, T.; Fischer, M.; Hackl, T.; Seifert, S. Opening the Random Forest Black Box of ^{1}H NMR Metabolomics Data by the Exploitation of Surrogate Variables. *Metabolites* **2023**, *13*, 1075. [CrossRef]
19. Siddiqui, A.J.; Musharraf, S.G.; Choudhary, M.I.; Rahman, A.-u. Application of Analytical Methods in Authentication and Adulteration of Honey. *Food Chem.* **2017**, *217*, 687–698. [CrossRef]
20. He, C.; Liu, Y.; Liu, H.; Zheng, X.; Shen, G.; Feng, J. Compositional Identification and Authentication of Chinese Honeys by ^{1}H NMR Combined with Multivariate Analysis. *Food Res. Int.* **2020**, *130*, 108936. [CrossRef]
21. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
22. Probst, P.; Bischl, B.; Boulesteix, A.-L. Tunability: Importance of Hyperparameters of Machine Learning Algorithms. *J. Mach. Learn. Res.* **2019**, *20*, 1–32.
23. Shakiba, N.; Gerdes, A.; Holz, N.; Wenck, S.; Bachmann, R.; Schneider, T.; Seifert, S.; Fischer, M.; Hackl, T. Determination of the Geographical Origin of Hazelnuts (*Corylus avellana* L.) by Near-Infrared Spectroscopy (NIR) and a Low-Level Fusion with Nuclear Magnetic Resonance (NMR). *Microchem. J.* **2021**, *174*, 107066. [CrossRef]
24. Shakiba, N.; Lösel, H.; Wenck, S.; Kumpmann, L.; Bachmann, R.; Creydt, M.; Seifert, S.; Fischer, M.; Hackl, T. Analysis of Hazelnuts (*Corylus avellana* L.) Stored for Extended Periods by ^{1}H NMR Spectroscopy Monitoring Storage-Induced Changes in the Polar and Nonpolar Metabolome. *J. Agric. Food Chem.* **2023**, *71*, 3093–3101. [CrossRef] [PubMed]
25. Wenck, S.; Bachmann, R.; Barmbold, S.-M.; Horns, A.L.; Paasch, N.; Seifert, S. Authentication of Apples (*Malus* × *domestica* Borkh.) According to Geographical Origin, Variety and Production Method Using ^{1}H NMR Spectroscopy and Random Forest. *Food Control* **2025**, *167*, 110817. [CrossRef]
26. Cui, C.; Xu, Y.; Jin, G.; Zong, J.; Peng, C.; Cai, H.; Hou, R. Machine Learning Applications for Identify the Geographical Origin, Variety and Processing of Black Tea Using ^{1}H NMR Chemical Fingerprinting. *Food Control* **2023**, *148*, 109686. [CrossRef]
27. Mascellani, A.; Hoca, G.; Babisz, M.; Krska, P.; Kloucek, P.; Havlik, J. ^{1}H NMR Chemometric Models for Classification of Czech Wine Type and Variety. *Food Chem.* **2021**, *339*, 127852. [CrossRef]
28. Emwas, A.-H.; Roy, R.; McKay, R.T.; Tenori, L.; Saccenti, E.; Gowda, G.A.N.; Raftery, D.; Alahmari, F.; Jaremko, L.; Jaremko, M.; et al. NMR Spectroscopy for Metabolomics Research. *Metabolites* **2019**, *9*, 123. [CrossRef]
29. Schütz, D.; Achten, E.; Creydt, M.; Riedl, J.; Fischer, M. Non-Targeted LC-MS Metabolomics Approach towards an Authentication of the Geographical Origin of Grain Maize (*Zea mays* L.) Samples. *Foods* **2021**, *10*, 2160. [CrossRef]
30. Creydt, M.; Hudzik, D.; Rurik, M.; Kohlbacher, O.; Fischer, M. Food Authentication: Small-Molecule Profiling as a Tool for the Geographic Discrimination of German White Asparagus. *J. Agric. Food Chem.* **2018**, *66*, 13328–13339. [CrossRef]
31. Klockmann, S.; Reiner, E.; Bachmann, R.; Hackl, T.; Fischer, M. Food Fingerprinting: Metabolomic Approaches for Geographical Origin Discrimination of Hazelnuts (*Corylus avellana*) by UPLC-QTOF-MS. *J. Agric. Food Chem.* **2016**, *64*, 9253–9262. [CrossRef]
32. Lösel, H.; Brockelt, J.; Gärber, F.; Teipel, J.; Kuballa, T.; Seifert, S.; Fischer, M. Comparative Analysis of LC-ESI-IM-qToF-MS and FT-NIR Spectroscopy Approaches for the Authentication of Organic and Conventional Eggs. *Metabolites* **2023**, *13*, 882. [CrossRef]
33. Wenck, S.; Creydt, M.; Hansen, J.; Gärber, F.; Fischer, M.; Seifert, S. Opening the Random Forest Black Box of the Metabolome by the Application of Surrogate Minimal Depth. *Metabolites* **2021**, *12*, 5. [CrossRef] [PubMed]
34. García-Seval, V.; Saurina, J.; Sentellas, S.; Núñez, O. Characterization and Classification of Spanish Honey by Non-Targeted LC–HRMS (Orbitrap) Fingerprinting and Multivariate Chemometric Methods. *Molecules* **2022**, *27*, 8357. [CrossRef] [PubMed]
35. Kunert, C. *Honey Authenticity Testing by LC-Orbitrap-HRMS [White Paper]*; Eurofins Whitepaper: Ritterhude, Germany, 2021.

36. Du, B.; Wu, L.; Xue, X.; Chen, L.; Li, Y.; Zhao, J.; Cao, W. Rapid Screening of Multiclass Syrup Adulterants in Honey by Ultrahigh-Performance Liquid Chromatography/Quadrupole Time of Flight Mass Spectrometry. *J. Agric. Food Chem.* **2015**, *63*, 6614–6623. [CrossRef] [PubMed]
37. Riedl, J.; Esslinger, S.; Fauhl-Hassek, C. Review of Validation and Reporting of Non-Targeted Fingerprinting Approaches for Food Authentication. *Anal. Chim. Acta* **2015**, *885*, 17–32. [CrossRef]
38. Lacalle-Bergeron, L.; Izquierdo-Sandoval, D.; Sancho, J.V.; López, F.J.; Hernández, F.; Portolés, T. Chromatography Hyphenated to High Resolution Mass Spectrometry in Untargeted Metabolomics for Investigation of Food (Bio)Markers. *TrAC Trends Anal. Chem.* **2021**, *135*, 116161. [CrossRef]
39. Zelena, E.; Dunn, W.B.; Broadhurst, D.; Francis-McIntyre, S.; Carroll, K.M.; Begley, P.; O'Hagan, S.; Knowles, J.D.; Halsall, A.; HUSERMET Consortium; et al. Development of a Robust and Repeatable UPLC−MS Method for the Long-Term Metabolomic Study of Human Serum. *Anal. Chem.* **2009**, *81*, 1357–1364. [CrossRef]
40. Zhou, B.; Xiao, J.F.; Tuli, L.; Ressom, H.W. LC-MS-Based Metabolomics. *Mol. BioSyst.* **2012**, *8*, 470–481. [CrossRef]
41. Thermo Scientific. *Compound Discoverer User Guide Software, Version 3.1*; Thermo Fisher Scientific Inc.: Waltham, MA, USA, 2019.
42. Benton, H.P.; Want, E.J.; Ebbels, T.M.D. Correction of Mass Calibration Gaps in Liquid Chromatography–Mass Spectrometry Metabolomics Data. *Bioinformatics* **2010**, *26*, 2488–2489. [CrossRef]
43. Smith, C.A.; Want, E.J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **2006**, *78*, 779–787. [CrossRef]
44. Tautenhahn, R.; Böttcher, C.; Neumann, S. Highly Sensitive Feature Detection for High Resolution LC/MS. *BMC Bioinform.* **2008**, *9*, 504. [CrossRef]
45. Zhong, P.; Wei, X.; Li, X.; Wei, X.; Wu, S.; Huang, W.; Koidis, A.; Xu, Z.; Lei, H. Untargeted Metabolomics by Liquid Chromatography-mass Spectrometry for Food Authentication: A Review. *Comp. Rev. Food. Sci. Food. Safe.* **2022**, *21*, 2455–2488. [CrossRef] [PubMed]
46. Creydt, M.; Fischer, M. Food Phenotyping: Recording and Processing of Non-Targeted Liquid Chromatography Mass Spectrometry Data for Verifying Food Authenticity. *Molecules* **2020**, *25*, 3972. [CrossRef] [PubMed]
47. Hansen, J.; Kunert, C.; Münstermann, H.; Raezke, K.-P.; Seifert, S. Application of Untargeted Liquid Chromatography-Mass Spectrometry to Routine Analysis of Food Using Three-Dimensional Bucketing and Machine Learning. *Sci. Rep.* **2024**, *14*, 16594. [CrossRef] [PubMed]
48. Scheibner, O.; Kellmann, M.; Yang, C.; Bromirski, M. Variable Data-Independent Acquisition (vDIA) Delivers High Selectivity and Sensitivity in Combined Targeted and Untargeted Analyses for Small Molecules. Available online: https://assets.thermofisher.com/TFS-Assets/CMD/Application-Notes/TN-64283-LC-MS-Veterinary-Drugs-Bio-Matrices-TN64283-EN.pdf (accessed on 8 February 2023).
49. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2023.
50. Rainer, J. *Metabolomics Data Pre-Processing Using Xcms, Version 1.0.0*; Zenodo: Geneve, Switzerland, 2020. [CrossRef]
51. Chambers, M.C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D.L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; et al. A Cross-Platform Toolkit for Mass Spectrometry and Proteomics. *Nat. Biotechnol.* **2012**, *30*, 918–920. [CrossRef] [PubMed]
52. Martens, L.; Chambers, M.; Sturm, M.; Kessner, D.; Levander, F.; Shofstahl, J.; Tang, W.H.; Römpp, A.; Neumann, S.; Pizarro, A.D.; et al. mzML—A Community Standard for Mass Spectrometry Data. *Mol. Cell. Proteom.* **2011**, *10*, R110.000133. [CrossRef]
53. Pedrioli, P.G.A.; Eng, J.K.; Hubley, R.; Vogelzang, M.; Deutsch, E.W.; Raught, B.; Pratt, B.; Nilsson, E.; Angeletti, R.H.; Apweiler, R.; et al. A Common Open Representation of Mass Spectrometry Data and Its Application to Proteomics Research. *Nat. Biotechnol.* **2004**, *22*, 1459–1466. [CrossRef]
54. Keller, A.; Eng, J.; Zhang, N.; Li, X.; Aebersold, R. A Uniform Proteomics MS/MS Analysis Platform Utilizing Open XML File Formats. *Mol Syst Biol* **2005**, *1*, 2005.0017. [CrossRef]
55. Gatto, L.; Lilley, K.S. MSnbase-an R/Bioconductor Package for Isobaric Tagged Mass Spectrometry Data Visualization, Processing and Quantitation. *Bioinformatics* **2012**, *28*, 288–289. [CrossRef]
56. Gatto, L.; Gibb, S.; Rainer, J. MSnbase, Efficient and Elegant R-Based Processing and Visualization of Raw Mass Spectrometry Data. *J. Proteome Res.* **2021**, *20*, 1063–1069. [CrossRef]
57. Prince, J.T.; Marcotte, E.M. Chromatographic Alignment of ESI-LC-MS Proteomics Data Sets by Ordered Bijective Interpolated Warping. *Anal. Chem.* **2006**, *78*, 6140–6152. [CrossRef]
58. Sánchez, J. MARDIA, K. V., J. T. KENT, J. M. BIBBY: Multivariate Analysis. Academic Press, London-New York-Toronto-Sydney-San Francisco 1979. Xv, 518 Pp., $ 61.00. *Biom. J.* **1982**, *24*, 502-502. [CrossRef]
59. Becker, R.M.; Chambers, J.M.; Wilks, A.R. *The New S Language Data Analysis: A Programming Environment for Data Analysis and Graphics*; The Wadsworth & Brooks/Cole statistics, probability series; Wadsworth & Brooks/Cole: Pacific Grove, CA, USA, 1988; ISBN 978-0-534-09193-4.
60. Venables, W.N.; Ripley, B.D. *Modern Applied Statistics with S*, 4th ed.; Springer: New York, NY, USA, 2002.
61. Wickham, H. *Ggplot2*; Use R! Springer International Publishing: Cham, Switzerland, 2016; ISBN 978-3-319-24275-0.
62. Wright, M.N.; Ziegler, A. Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *J. Stat. Soft.* **2017**, *77*, 1–17. [CrossRef]

63. Hamner, B.; Frasco, M. *Metrics: Evaluation Metrics for Machine Learning*; The Comprehensice R Archive Network. 2018. Available online: https://metadatacatalogue.lifewatch.eu/srv/api/records/a5f2540d-bb75-4c25-b9b5-68fee6870b86 (accessed on 11 November 2024).
64. Broadhurst, D.; Goodacre, R.; Reinke, S.N.; Kuligowski, J.; Wilson, I.D.; Lewis, M.R.; Dunn, W.B. Guidelines and Considerations for the Use of System Suitability and Quality Control Samples in Mass Spectrometry Assays Applied in Untargeted Clinical Metabolomic Studies. *Metabolomics* **2018**, *14*, 72. [CrossRef]
65. Märtens, A.; Holle, J.; Mollenhauer, B.; Wegner, A.; Kirwan, J.; Hiller, K. Instrumental Drift in Untargeted Metabolomics: Optimizing Data Quality with Intrastudy QC Samples. *Metabolites* **2023**, *13*, 665. [CrossRef]
66. The Human Serum Metabolome (HUSERMET) Consortium. Dunn, W.B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J.D.; Halsall, A.; et al. Procedures for Large-Scale Metabolic Profiling of Serum and Plasma Using Gas Chromatography and Liquid Chromatography Coupled to Mass Spectrometry. *Nat. Protoc.* **2011**, *6*, 1060–1083. [CrossRef]
67. Doner, L.W.; Kushnir, I.; White, J.W. Assuring the Quality of Honey. Is It Honey or Syrup? *Anal. Chem.* **1979**, *51*, 224A–232A. [CrossRef]
68. Ofoedu, C.E.; Osuji, C.M.; Ojukwu, M. Sugar Profile of Syrups from Malted and Unmalted Rice of Different Varieties. *J. Food Res.* **2019**, *8*, 52. [CrossRef]
69. Ghazi, I.; Fernandez-Arrojo, L.; Gomez De Segura, A.; Alcalde, M.; Plou, F.J.; Ballesteros, A. Beet Sugar Syrup and Molasses as Low-Cost Feedstock for the Enzymatic Production of Fructo-Oligosaccharides. *J. Agric. Food Chem.* **2006**, *54*, 2964–2968. [CrossRef]
70. Han, W.; Li, L. Evaluating and Minimizing Batch Effects in Metabolomics. *Mass Spectrom. Rev.* **2022**, *41*, 421–442. [CrossRef]
71. Malinka, F.; Zareie, A.; Prochazka, J.; Sedlacek, R.; Novosadova, V. Batch Alignment via Retention Orders for Preprocessing Large-Scale Multi-Batch LC-MS Experiments. *Bioinformatics* **2022**, *38*, 3759–3767. [CrossRef]
72. Yang, Q.; Li, B.; Wang, P.; Xie, J.; Feng, Y.; Liu, Z.; Zhu, F. LargeMetabo: An out-of-the-Box Tool for Processing and Analyzing Large-Scale Metabolomic Data. *Brief. Bioinform.* **2022**, *23*, bbac455. [CrossRef] [PubMed]
73. Habra, H.; Kachman, M.; Bullock, K.; Clish, C.; Evans, C.R.; Karnovsky, A. *metabCombiner*: Paired Untargeted LC-HRMS Metabolomics Feature Matching and Concatenation of Disparately Acquired Data Sets. *Anal. Chem.* **2021**, *93*, 5028–5036. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.