Journal of Food Measurement and Characterization (2024) 18:5673–5682
https://doi.org/10.1007/s11694-024-02597-5

ORIGINAL PAPER

Distinguishing Turkish pine honey from multi-floral honey through MALDI-MS-based *N*-glycomics and machine learning

Saad Masri[1] · Sena Aksoy[1] · Hatice Duman[2] · Sercan Karav[2] · Haci Mehmet Kayili[1] · Bekir Salih[3]

Received: 1 February 2024 / Accepted: 24 April 2024 / Published online: 21 May 2024

Abstract

Honey, a multifaceted blend of sugars, amino acids, vitamins, proteins, and minerals, exhibits compositional variability dependent upon the floral source. While previous studies have attempted to categorize honey, the use of glycomic profiles for honey classification remains an unexplored avenue. This investigation seeks to establish a methodology for distinguishing honey types, specifically multi-floral and pine honey, employing mass spectrometry-based glycomic analysis in tandem with machine learning. In this search, seven samples of pine honey and eight samples of multi-floral honey were obtained from diverse regions of Turkey. Subsequently, the proteins within these honey samples were extracted, and glycans were enzymatically released. The released glycans were labeled with 2-aminobenzoic acid (2-AA) and subjected to analysis via matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS). The glycan profiles of pine and multi-floral honey were determined through these analytical procedures, revealing a total of 76 distinct *N*-glycan structures. Among these, 13 *N*-glycan profiles consistently established at high levels across experimental replicates and were incorporated in subsequent analyses. Following the quantification of individual glycan abundances, statistically significant differences in glycan profiles were determined. Notably, *N*-glycans Hex5HexNAc2, Hex4HexNAc3, and Hex5HexNAc3 displayed considerable differences. Using the 13 *N*-glycan profiles, an accuracy rate of 93.5% was obtained from machine learning analysis, which increased to 100% when incorporating the identified significantly changed glycans. The most productive models were identified as "subspace and fine k-nearest neighbors (KNN)." The findings underscore the potential of mass spectrometry-based glycomics in conjunction with machine learning as a robust tool for precise honey type classification and its prospective utility in quality control and honey product authentication.

Keywords Honey classification · Pine honey · Multi-floral honey · Machine learning · Glycomics

Introduction

Honey, a natural sweetener formed by bees from the nectar of various types of flowers, has a composition that varies depending on factors such as botanical and geographical origin, as well as processing and storage conditions [1, 2]. Therefore, it is essential to develop reliable analytical methods for evaluating the quality and authenticity of honey. The components found in honey, including carbohydrates, amino acids, vitamins, and minerals, are crucial factors that determine its nutritional and therapeutic properties [2]. Honey sourced from different types of flowers possesses unique chemical constituents, including antimicrobial, anti-inflammatory, and antioxidant properties [2–8]. However, concerns have arisen about the quality and authenticity of honey due to common issues in the market, such as impurity and mislabeling [9].

In honey authentication research, conventional approaches are used to determine its botanical origin. Sensory and physicochemical assessments are employed to learn the provenance of monofloral honey [10]. At the same time, the customary practice involves the utilization of

✉ Haci Mehmet Kayili
h.mehmetkayili@karabuk.edu.tr

✉ Bekir Salih
bekir@hacettepe.edu.tr

[1] Department of Medical Engineering, Faculty of Engineering, Karabük University, 78000 Karabük, Türkiye

[2] Department of Molecular Biology and Genetics, Faculty of Science, Çanakkale Onsekiz Mart University, 17100 Çanakkale, Türkiye

[3] Department of Chemistry, Faculty of Science, Hacettepe University, 06800 Ankara, Türkiye

melissopalynological analysis for the microscopic scrutiny and identification of floral pollen grains within honey [11]. Nevertheless, it should be noted that melissopalynological methodologies may not be as pertinent when applied to other honey varieties due to the inherent variability in pollen content, which typically exhibits lower levels [12]. In addition, the melissopalynological analysis is an intricate process that requires a limited number of highly trained experts, possibly owing to the rigorous training needed for this specialized field [13].

The limitations of conventional methods for verifying the botanical and geographical origins of honey emphasize the necessity for more dependable and contemporary analytical approaches. Cutting-edge analytical instruments and sensor arrays, including chromatography [14], mass spectrometry (MS) [15] techniques, vibrational spectroscopy such as infrared (IR) [16] and Raman spectroscopy methods [17], nuclear magnetic resonance (NMR) [18], are joined in research to assess sugar profiles, mineral content, phenolic and flavonoid compositions, aroma characteristics, and amino acid compositions[19]. The application of these tools is crucial in ensuring the precision of honey origin authentication.

Among many analytical techniques, mass spectrometry-based glycomic methods have emerged as practical tools for the detailed analysis of carbohydrate components in glycoproteins [20, 21]. This method analyzes mass-to-charge ratios and fragmentation patterns to identify complex carbohydrate structures located in honey glycoproteins for the honey samples and determine their quantities [22, 23]. This allows for obtaining a detailed profile of the glycans in honey glycoproteins, including information about glycan compositions, linkage types, and branching patterns. However, interpreting this complex data can be challenging due to the limitations of traditional statistical methods. In this regard, the use of advanced data analysis with machine learning approaches may be a good choice [24]. Machine learning algorithms can analyze large datasets to uncover situations or relationships that may go unnoticed by observers. In this context, machine learning is commonly used to classify honey types and determine their essential characteristics[17].

In this study, honey samples from different origins were collected and categorized as multi-floral and pine honey based on botanical type. The *N*-glycan compositions of these samples were analyzed using mass spectrometry-based glycomic methods, and the classification potential of machine learning was explored for distinguishing between multi-floral and pine honey. The performance of the machine learning models was evaluated. The outcomes of this study will introduce a technique that can be used to classify two widely consumed honey types with different botanical origins and contribute to understanding the chemical diversity of these honey types.

Methods

Materials

All chemicals used within the scope of the study were sourced from Sigma-Aldrich (St Louis, MO, USA) unless otherwise specified. PNGase F enzyme was procured from Promega. Multi-floral (n = 8) and pine honey (n = 8) were obtained from various regions in Turkey where honey production is carried out.

Extraction of honey glycoproteins

A 100 µL sample was taken from each honey type (from pine and flower sources). To each sample, 100 µL of chloroform was added and mixed. Then, 300 µL of deionized water and 400 µL of methanol were added. The prepared samples were centrifuged at 14,000 relative centrifugal force (rcf), and the resulting liquid phase was separated without disturbing the pellet. An additional 400 µL of methanol was added, and the samples were centrifuged again at 14,000 rcf for 5 min. Once again, the upper phase was removed, and in the final step, the samples were dried under rapid vacuum at 45 °C.

Release and labeling of *N*-glycans from honey glycoproteins

After dissolving protein extracts in 50 µL of 1% SDS, they were incubated at 90 °C for 10 min for denaturation. Subsequently, 25 µL of 2% Igepal-CA630 and 25 µL of 5×PBS were added to the solution and mixed. To perform enzymatic deglycosylation, 1 U of PNGase F enzyme was added, and the samples were left overnight at 37 °C. Following the removal of *N*-glycans from glycoproteins, the samples were labeled with 2-aminobenzoic acid (2-AA). For this process, solutions of 48 mg/mL 2-AA (in dimethyl sulfoxide/acetic acid, 10:3, v/v) and 63 mg/mL sodium cyanoborohydride ($NaCNBH_3$) (in DMSO) were prepared. Then, 50 µL of 2-AA and 50 µL of $NaCNBH_3$ were sequentially added to the honey samples subjected to glycan release. The labeling process was carried out by incubating the samples at 65 °C for 2 h.

Purification of *N*-glycans via HILIC and PGC-containing solid phase extraction cartridges

2-AA labeled *N*-glycans were first purified using solid-phase extraction with cellulose. Samples were supplemented with 100% ACN to reach 85% ACN in their final volumes. Approximately 20 mg of cellulose solid-phase extraction

cartridges were used for this purpose. The cartridges were initially washed twice with 1 mL of 100% water and then 85% ACN. The samples were added to the cartridges and left to interact with the material for 5 min. Subsequently, the cellulose-containing cartridges were washed three times with 85% ACN containing 1% TFA and 85% ACN solutions, effectively removing excess labels and other chemicals. Elution of 2-AA labeled *N*-glycans was carried out using 0.75 mL of water. Later, 0.75 µL of TFA was added to the samples, followed by purification with porous graphitized carbon. Approximately 20 mg of porous graphitized carbon-containing material was placed inside the cartridges. Initially, carbon-containing cartridges were washed twice, with 1 mL of 80% ACN containing 0.1% TFA and 0.1% TFA in water, respectively. Subsequently, the samples previously purified with cellulose were added to the carbon-containing cartridges and incubated for 5 min to allow the samples to interact with the material. Then, the PGC-containing cartridges were washed five times with water containing 0.1% TFA. Elution of 2-AA labeled glycans was performed using 80% ACN containing 0.1% TFA. Finally, the samples were dried using a speed vacuum concentrator at 45 °C.

MALDI-MS analysis

One microliter (1 µL) of the elution solution was spotted onto an MTP 384 Anchor sample holder and allowed to dry. Subsequently, 1 µL of a solution containing 5 mg of 2,5-dihydroxybenzoic acid (DHB) matrix dissolved in acetonitrile/H_2O (1/1, v/v) with 0.1% orthophosphoric acid was added to the dried sample. The analysis was performed using MALDI mass spectrometry with a Bruker rapifleXTM MALDI TissuetyperTM instrument (Germany, Bremen). The analysis was conducted in the reflectron mode with at least 8000 laser shots for negative ionization. Prior to each analysis, mass calibration of the Bruker rapifleXTM MALDI TissuetyperTM mass spectrometry was performed using a peptide mixture. Mass spectra in the range of 1000–4000 m/z were obtained with an applied acceleration voltage of 25 kV.

Data analysis of *N*-glycans

The data obtained for the analysis of *N*-glycans were transferred to ProteinScape software for identification. The Glycoquest algorithm was used to determine which glycan components each peak detected. The CarbBank database was utilized for the searches, with a mass error tolerance of 100 ppm. The areas of the obtained glycan components were determined using FlexAnalysis software, and the relative area of each *N*-glycan was obtained using a total area normalization approach. Each sample was analyzed with two experimental replicates. Spectrum quality guided the selection of 17 pine honey and 14 multi-floral honey spectra for the analysis.

Statistical and machine learning analyses

Perseus software was employed for statistical analyses. First, the normalized data was imported into the software. After necessary groupings were made, the data was classified to include relative areas of at least 80% for the defined group. Values without defined relative areas were set to 0. Classic t-tests were used for statistical analyses, and a change rate with a p-value less than 0.05 was considered significant. Volcano plots were created to identify components showing substantial changes.

In machine learning analyses, two different datasets were used. The first dataset included 13 different *N*-glycan compositions, while the second dataset comprised three different *N*-glycans found to have significant statistical changes. Both datasets were loaded into the classification learner application in MatLab software. The cross-validation rate was set at "5," and the classification performance of all algorithms available in the application was tested. Subsequently, machine learning models were evaluated for their performance using metrics such as accuracy and ROC analysis.

Results and discussion

In this study, samples were obtained from two different honey sources (pine honey and multi-floral honey) from various regions. Initially, honey proteins were extracted from the samples, and then glycan release was performed, followed by labeling with 2-AA. Subsequently, purification of 2-AA labeled glycans was carried out. Glycans were analyzed using MALDI-MS. The peak areas for the identified glycans were extracted, and the relative abundance of each glycan was calculated. Finally, statistical and machine learning analyses were conducted.

As a result of the analyses, a total of 76 different *N*-glycan compositions were identified (Table 1). In Fig. 1, MALDI mass spectra for multi-floral and pine kinds of honey are presented. Out of these, 13 *N*-glycan profiles that were consistently present in experimental replicates were utilized in the statistical analyses. After determining the relative abundance of each glycan, statistically significant variations in glycan profiles were identified. A classical t-test was applied to detect statistically significant changes in glycan abundance between the two honey types. Significantly regulated *N*-glycans that contributed to the differentiation between pine and multi-floral honey are shown in Fig. 2. The volcano plot seen in Fig. 2A reveals that the glycans Hex5HexNAc3, Hex4HexNAc3, and Hex5HexNAc2 exhibited statistically significant differences between multi-floral

Table 1 The list of the identified *N*-glycans belonging to pine and multi-floral honeys

Row	Proposed Composition	m/z meas	z	m/z calc	Δ MH + [Da]
1	Hex3HexNAc2-AA	1030.469	− 1	1030.373	0.096
2	Hex3HexNAc2Pen1-AA	1162.465	− 1	1162.416	0.049
3	**Hex3HexNAc2dHex1-AA**	1176.473	− 1	1176.431	0.041
4	Hex3HexNAc2Me2Pen1-AA	1190.443	− 1	1190.447	− 0.004
5	**Hex4HexNAc2-AA**	1192.466	− 1	1192.426	0.040
6	**Hex3HexNAc3-AA**	1233.476	− 1	1233.453	0.024
7	Hex3HexNAc3S1-AA	1313.443	− 1	1313.409	0.034
8	Hex4HexNAc2dHex1-AA	1338.476	− 1	1338.484	− 0.008
9	*Hex5HexNAc2-AA*	1354.471	− 1	1354.479	− 0.008
10	Hex3HexNAc3Pen1-AA	1365.477	− 1	1365.495	− 0.018
11	**Hex3HexNAc3dHex1-AA**	1379.496	− 1	1379.511	− 0.015
12	*Hex4HexNAc3-AA*	1395.493	− 1	1395.505	− 0.013
13	**Hex3HexNAc4-AA**	1436.511	− 1	1436.532	− 0.021
14	Hex3HexNAc4S1-AA	1516.503	− 1	1516.489	0.014
15	Hex6HexNAc2-AA	1516.530	− 1	1516.532	− 0.002
16	Hex4HexNAc3Pen1-AA	1527.456	− 1	1527.548	− 0.092
17	Hex4HexNAc3dHex1-AA	1541.594	− 1	1541.563	0.030
18	*Hex5HexNAc3-AA*	1557.532	− 1	1557.558	− 0.027
19	Hex3HexNAc4Pen1-AA	1568.566	− 1	1568.574	− 0.008
20	**Hex3HexNAc4dHex1-AA**	1582.557	− 1	1582.590	− 0.033
21	Hex4HexNAc4-AA	1598.554	− 1	1598.585	− 0.031
22	**Hex3HexNAc5-AA**	1639.586	− 1	1639.611	− 0.026
23	Hex6HexNAc2dHex1-AA	1662.645	− 1	1662.590	0.055
24	Hex4HexNAc4S1-AA	1678.560	− 1	1678.542	0.018
25	Hex7HexNAc2-AA	1678.588	− 1	1678.585	0.004
26	Hex3HexNAc5S1-AA	1719.580	− 1	1719.568	0.012
27	Hex6HexNAc3-AA	1719.609	− 1	1719.611	− 0.002
28	Hex4HexNAc4dHex1-AA	1744.610	− 1	1744.643	− 0.033
29	Hex4HexNAc4Me2Pen1-AA	1758.565	− 1	1758.658	− 0.094
30	Hex5HexNAc4-AA	1760.617	− 1	1760.638	− 0.021
31	Hex3HexNAc5dHex1-AA	1785.643	− 1	1785.669	− 0.026
32	**Hex4HexNAc5-AA**	1801.651	− 1	1801.664	− 0.013
33	Hex5HexNAc4S1-AA	1840.606	− 1	1840.594	0.011
34	Hex8HexNAc2-AA	1840.629	− 1	1840.637	− 0.008
35	Hex3HexNAc6-AA	1842.776	− 1	1842.691	0.085
36	Hex4HexNAc5S1-AA	1881.634	− 1	1881.621	0.013
37	Hex5HexNAc4dHex1-AA	1906.760	− 1	1906.696	0.064
38	Hex4HexNAc3NeuAc1S1dHex1-AA	1912.624	− 1	1912.616	0.008
39	Hex8HexNAc2P1-AA	1920.638	− 1	1920.604	0.035
40	Hex5HexNAc4Me2Pen1-AA	1920.672	− 1	1920.711	− 0.039
41	Hex3HexNAc6S1-AA	1922.661	− 1	1922.648	0.014
42	Hex6HexNAc4-AA	1922.683	− 1	1922.690	− 0.007
43	Hex4HexNAc5dHex1-AA	1947.717	− 1	1947.722	− 0.005
44	Hex5HexNAc5-AA	1963.729	− 1	1963.717	0.012
45	Hex6HexNAc4Me3-AA	1964.692	− 1	1964.737	− 0.045
46	Hex5HexNAc4S1dHex1-AA	1986.670	− 1	1986.652	0.018
47	Hex3HexNAc6dHex1-AA	1988.805	− 1	1988.749	0.057
48	Hex3HexNAc6S2-AA	2002.634	− 1	2002.604	0.030
49	**Hex9HexNAc2-AA**	2002.705	− 1	2002.690	0.015
50	Hex4HexNAc6-AA	2004.866	− 1	2004.744	0.123
51	Hex8HexNAc3-AA	2043.577	− 1	2043.717	− 0.139

Table 1 (continued)

Row	Proposed Composition	m/z meas	z	m/z calc	Δ MH + [Da]
52	Hex3HexNAc7-AA	2045.738	− 1	2045.770	− 0.032
53	Hex5HexNAc4NeuAc1-AA	2051.847	− 1	2051.733	0.114
54	Hex6HexNAc4dHex1-AA	2068.814	− 1	2068.748	0.066
55	**Hex9HexNAc2P1-AA**	2082.722	− 1	2082.657	0.065
56	Hex7HexNAc4-AA	2084.737	− 1	2084.743	− 0.007
57	Hex6HexNAc5-AA	2125.888	− 1	2125.770	0.118
58	Hex4HexNAc6dHex1-AA	2150.895	− 1	2150.801	0.094
59	Hex10HexNAc2-AA	2164.713	− 1	2164.743	− 0.030
60	Hex5HexNAc6-AA	2166.825	− 1	2166.796	0.029
61	Hex4HexNAc4NeuAc2-AA	2180.802	− 1	2180.776	0.026
62	Hex4HexNAc7-AA	2207.962	− 1	2207.823	0.139
63	Hex7HexNAc4Me4Pen1-AA	2272.727	− 1	2272.848	− 0.121
64	Hex7HexNAc5-AA	2287.746	− 1	2287.823	− 0.077
65	Hex4HexNAc5NeuAc1S1dHex1-AA	2318.785	− 1	2318.774	0.011
66	Hex5HexNAc6Me2Pen1-AA	2326.899	− 1	2326.870	0.029
67	Hex5HexNAc4NeuAc2-AA	2342.871	− 1	2342.828	0.043
68	Hex3HexNAc8dHex1-AA	2394.943	− 1	2394.907	0.035
69	Hex5HexNAc4Neu2dHex1-AA	2404.769	− 1	2404.865	− 0.097
70	Hex3HexNAc9-AA	2451.780	− 1	2451.929	− 0.149
71	Hex5HexNAc4NeuAc2S1dHex1-AA	2568.899	− 1	2568.843	0.055
72	Hex5HexNAc6Pen1dHex2-AA	2590.802	− 1	2590.954	− 0.153
73	Hex16HexNAc2-AA	3137.288	− 1	3137.060	0.229
74	Hex7HexNAc8S4-AA	3216.897	− 1	3216.888	0.009
75	Hex7HexNAc8S6-AA	3376.759	− 1	3376.802	− 0.043
76	Hex9HexNAc6NeuAc2dHex1-AA	3543.094	− 1	3543.256	− 0.162

The glycans highlighted in italic font indicate statistical significance, while those highlighted in bold font are selected for machine learning analysis aimed at distinguishing between pine honey and multi-floral honey

and pine honeys. These glycans displayed notable differences in abundance and were found to be significantly higher in one honey type compared to the other. These specific glycans emerge as key discriminators facilitating the differentiation between pine and multi-floral honey. Specifically, the glycan Hex5HexNAc3 exhibited substantial up-regulation in multi-floral honey compared to pine honey. Additionally, Hex5HexNAc2 demonstrated the most pronounced variation between multi-floral and pine honey types.

The principal component analysis in Fig. 2B demonstrates the ability to distinguish between pine and multi-floral honey using the significant change of three *N*-glycans in the honey samples. Red dots represent glycans from pine honey samples, while blue dots represent multi-floral honey. In the two-dimensional analysis, component 1 explains 85.4% of the variance, and component 2 explains 14.1% of it. This analysis indicates a high differentiation between pine and multi-floral honeys. The differentiation between pine and multi-floral kinds of honey can be attributed to significant quantitative differences in *N*-glycan profiles for significantly changed profiles. The clustering of the samples observed in the chart demonstrates that the glycan compositions of the two honey types are pretty different. Component 1, capturing the most significant variance in the data, likely reflects the fundamental differences in glycan abundance between the two honey types. Similarly, Component 2 highlights additional differences contributing to the differentiation between pine and multi-floral honey.

In the machine learning analysis, the relative areas of the glycans were determined based on the mass spectrometric data. Subsequently, machine learning models were developed using MATLAB, employing a 5% cross-validation. The performance of all available models within the MATLAB classification application was comprehensively assessed. To evaluate the machine learning models' effectiveness, various criteria were considered, including accuracy and the calculation of the area under the curve (AUC). Figure 3A in the study provides a detailed insight into the machine learning classification results based on 13 *N*-glycans obtained from pine and multi-floral glycoproteins. The findings from this figure indicate that the true positive score for predicting pine honey is an impressive 100%, signifying that all pine honey samples were accurately classified. On the other hand, there is a false positive rate of 12.5% in the classification of

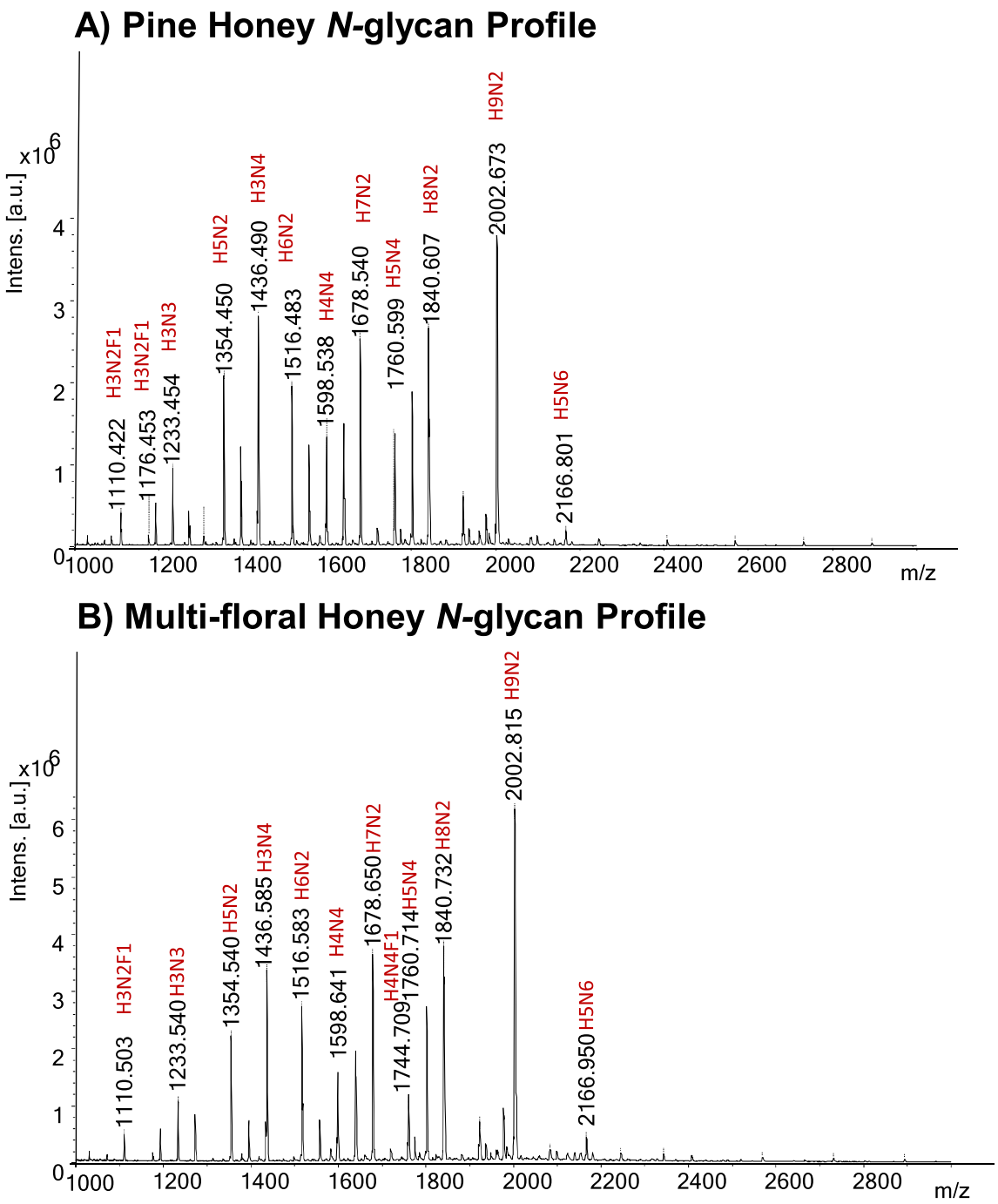

Fig. 1 MALDI-MS spectra of identified *N*-glycans derived from **A** pine honey and **B** multi-floral honey

multi-floral honey. This indicates that while the model excels at classifying pine honey, there is room for improvement in distinguishing multi-floral honey samples. Overall, the classification accuracy of the model stands at 93.5%. Figure 3B presents the ROC curve analysis for this model, displaying the area under the curve (AUC). The obtained AUC value is 0.93, indicating a high level of classification performance (Fig. 3B). The fact that the ROC curve predominantly lies above the diagonal line signifies the model's success in distinguishing between pine and multi-floral honey. The high AUC value further substantiates the effectiveness of accurately classifying samples based on the selected glycans.

In Fig. 3C, a confusion matrix is presented for the machine learning classification results obtained using three statistically significant glycans (Hex5HexNAc3, Hex4HexNAc3, and Hex5HexNAc2). Remarkably, in this

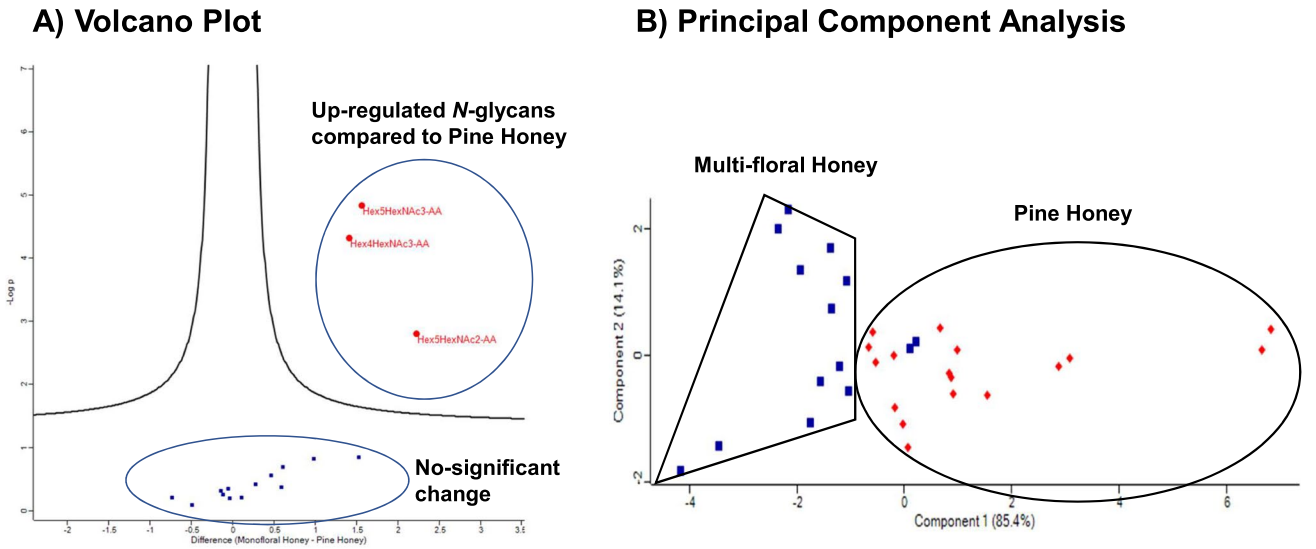

Fig. 2 The figure elucidates the graphical representations employed in the comparative analysis of *N*-glycans between pine honey and multi-floral honey. Firstly, **A** encompasses a volcano plot, visually depicting the distinctively regulated *N*-glycans between the two honey types. And, **B** displays a diagram derived from principal component analysis, providing a visual representation of the multivariate variations among the *N*-glycan profiles of the honey varieties (Color figure online)

scenario, all pine honey and multi-floral honey samples were perfectly and precisely distinguished from one another, resulting in a 100% accuracy. This remarkable achievement in accurately discerning the two honey types underlines the discriminative capability of the three selected glycans. Figure 3D showcases the ROC curve analysis for the classification model employing these three statistically significant glycans. The ROC curve consistently remains positioned above the diagonal line, with an AUC value of 1.00 reported. This AUC value signifies an exceptionally high level of classification performance. This analysis unequivocally demonstrates the algorithm and model's exceptional effectiveness in the precise classification of samples based on the selected glycans.

The literature extensively covers the proteomic characterization of various honey types. A recent study thoroughly examined the protein composition of 13 honeys, identifying a total of 130 proteins [25]. Utilizing a proteomic approach, honey from distinct geographical and botanical sources was clearly differentiated [26]. It has been understood that the proteomic profiles of honey types differ among honey types. However, there is currently a lack of literature on the *N*-glycans of honey glycoproteins. This study is the pioneer in utilizing MALDI-MS-based *N*-glycomics and machine learning for the classification of honey types.

The variation in glycosylation machinery across organisms is known. Unlike proteomes, there is no genetic template for the glycosylation machinery within cells [27]. It is emphasized that multiple genes play a role in the formation of glycans attached to proteins, rendering glycans highly intricate and site-, tissue- and species-specific [28]. In contrast to comprehensive proteomics experiments, *N*-glycan profiling can be accomplished through MALDI-MS. In this method, *N*-glycans extracted from pine and multi-floral honey were swiftly profiled and utilized for classification purposes.

The categorization of different types of honey is crucial to protect consumers from deceptive practices [29]. The existing literature has explored various methodologies for classifying different kinds of honey, with a predominant focus on categorizing them according to botanical and regional origins. The primary objective is to establish a bioanalytical method that exhibits high accuracy and specificity in discerning between various honey types. Mass spectrometry stands out as a highly suitable approach owing to its rapidity, accuracy, and precision. An illustrative instance is the utilization of gas chromatography-mass spectrometry (GC–MS) analysis to assess volatile organic compounds in honey, serving as an effective means for classification [30]. Applying this approach alongside Kohonen's self-organizing map has enabled the discrimination between Turkish pine honey and Greek pine honey [31].

Conclusion

In conclusion, a bioanalytical technique integrating MALDI-MS-based *N*-glycomics and machine learning was created for the classification of Turkish pine honey in comparison to multi-floral honey. This method demonstrated

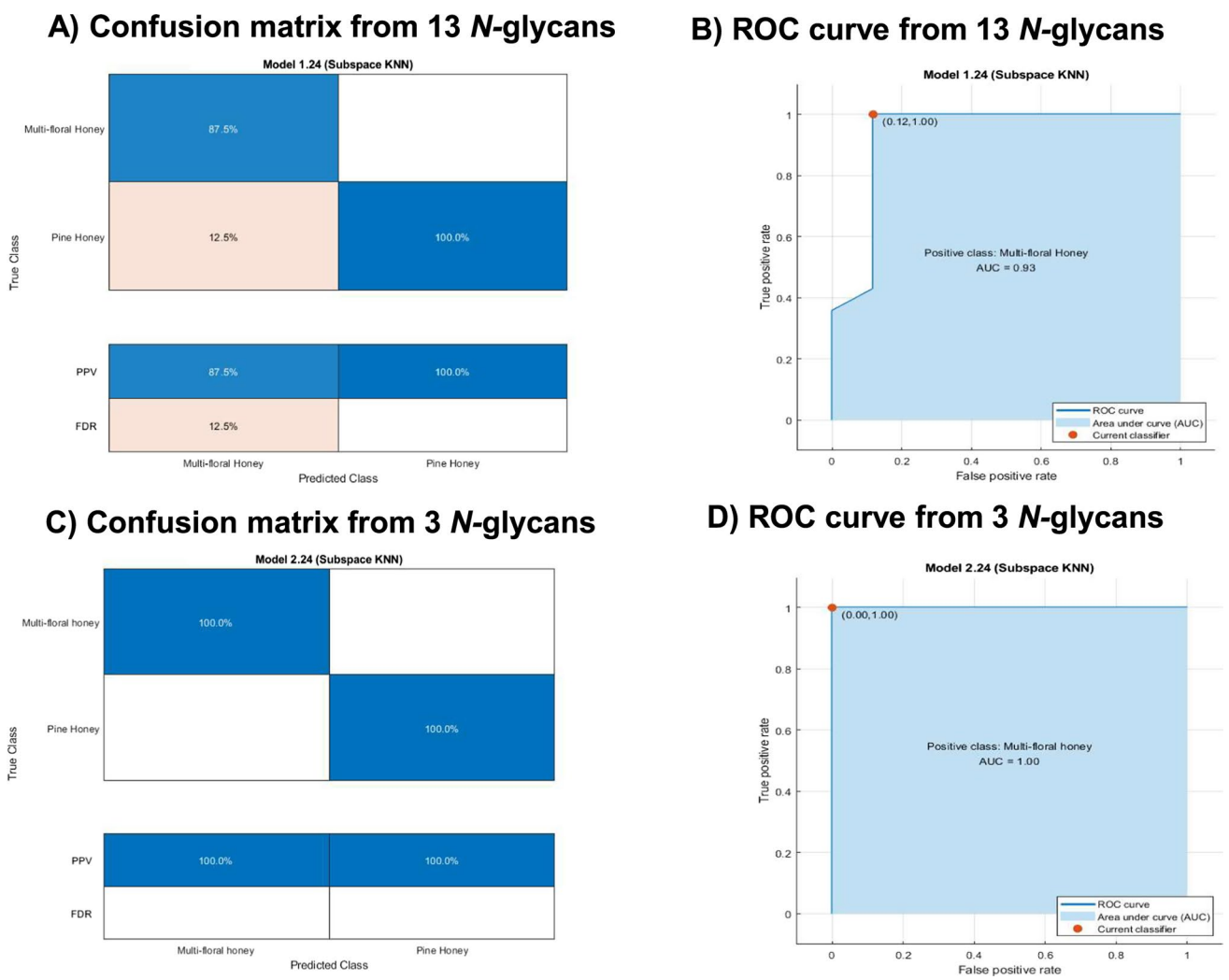

Fig. 3 The confusion matrix presented herein pertains to a machine-learning model employed in the discriminatory classification of pine honey versus multi-floral honey. The model's training involved two distinct feature sets: **A** comprised 13 *N*-glycans, and **C** restricted to three *N*-glycans demonstrating significant alterations. Additionally, the receiver operating characteristic (ROC) curves, delineating the models' performance characteristics, are illustrated in (**B**) for the 13 *N*-glycans dataset and **D** specifically for the subset of three significantly changed *N*-glycans

high accuracy and specificity in distinguishing pine honey from multi-floral honey. Utilizing statistically significant glycans for discrimination resulted in a 100% accuracy rate. This approach enables swift differentiation between Turkish pine honey and multi-floral honey. The application of this approach could extend to distinguishing between pine and multi-floral honey originating from diverse botanical sources. Additionally, this method has the potential to differentiate between various types of multi-floral honey and mono-flower honey. This suggests that MALDI-MS-based glycomics coupled with machine learning hold promise as a reliable candidate for identifying honey-type adulteration and mitigating variations.

Acknowledgements Partial funding for the research was received from the Ministry of Development-Republic of Türkiye, with Project Number 2016 K121230. Bekir Salih extends his gratitude to the Turkish Academy of Science (TUBA) for their financial support.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

Data Availability Data will be made available on request.

Declarations

Conflict of interest The authors have no conflicts of interest to disclose.

References

1. P.M. da Silva, C. Gauche, L.V. Gonzaga, A.C.O. Costa, R. Fett, Honey: chemical composition, stability and authenticity. Food Chem. **196**, 309–323 (2016). https://doi.org/10.1016/j.foodchem.2015.09.051
2. T. Eteraf-Oskouei, M. Najafi, Traditional and modern uses of natural honey in human diseases: a review. Iran. J. Basic Med. Sci. **16**(6), 731–742 (2013)
3. V.R. Pasupuleti, L. Sammugam, N. Ramesh, S.H. Gan, Honey, propolis, and royal jelly: a comprehensive review of their biological actions and health benefits. Oxid. Med. Cell. Longev. (2017). https://doi.org/10.1155/2017/1259510
4. S. Samarghandian, T. Farkhondeh, F. Samini, Honey and health: a review of recent clinical research. Pharmacogn. Res. **9**(2), 121–127 (2017). https://doi.org/10.4103/0974-8490.204647
5. J.M. Alvarez-Suarez, S. Tulipani, S. Romandini, E. Bertoli, M. Battino, Contribution of honey in nutrition and human health: a review. Mediterr. J. Nutr. Metab. **3**(1), 15–23 (2010). https://doi.org/10.1007/s12349-009-0051-6
6. S. Bogdanov, T. Jurendic, R. Sieber, P. Gallmann, Honey for nutrition and health: a review. J. Am. Coll. Nutr. **27**(6), 677–689 (2008). https://doi.org/10.1080/07315724.2008.10719745
7. M. Küçük, S. Kolayli, S. Karaoglu, E. Ulusoy, C. Baltaci, F. Candan, Biological activities and chemical composition of three honeys of different types from Anatolia. Food Chem. **100**(2), 526–534 (2007). https://doi.org/10.1016/j.foodchem.2005.10.010
8. G. Beretta, P. Granata, M. Ferrero, M. Orioli, R.M. Facino, Standardization of antioxidant properties of honey by a combination of spectrophotometric/fluorimetric assays and chemometrics. Anal. Chim. Acta **533**(2), 185–191 (2005). https://doi.org/10.1016/j.aca.2004.11.010
9. R. Fakhlaei, J. Selamat, A. Khatib, A.F.A. Razis, R. Sukor, S. Ahmad, A.A. Babadi, The toxic impact of honey adulteration: a review. Foods (2020). https://doi.org/10.3390/foods9111538
10. J.M.B. de Sousa, E.L. de Souza, G. Marques, M. de Toledo Benassi, B. Gullón, M.M. Pintado, M. Magnani, Sugar profile, physicochemical and sensory aspects of monofloral honeys produced by different stingless bee species in Brazilian semi-arid region. LWT **65**, 645–651 (2016). https://doi.org/10.1016/j.lwt.2015.08.058
11. M.A. Rodopoulou, C. Tananaki, M. Dimou, V. Liolios, D. Kanelis, G. Goras, A. Thrasyvoulou, The determination of the botanical origin in honeys with over-represented pollen: combination of melissopalynological, sensory and physicochemical analysis. J. Sci. Food Agric. **98**(7), 2705–2712 (2018). https://doi.org/10.1002/jsfa.8764
12. P. Molan, The limitations of the methods of identifying the floral source of honeys. Bee World **79**(2), 59–68 (1998). https://doi.org/10.1080/0005772x.1998.11099381
13. I. Escriche, M. Juan-Borrás, M. Visquert, J.M. Valiente, An overview of the challenges when analysing pollen for monofloral honey classification. Food Control **143**, 109305 (2023). https://doi.org/10.1016/j.foodcont.2022.109305
14. E. Guzelmeric, I. Ciftci, P.I. Yuksel, E. Yesilada, Importance of chromatographic and spectrophotometric methods in determining authenticity, classification and bioactivity of honey. LWT **132**, 109921 (2020). https://doi.org/10.1016/j.lwt.2020.109921
15. H.E. Tahir, Z. Xiaobo, H. Xiaowei, S. Jiyong, A.A. Mariod, Discrimination of honeys using colorimetric sensor arrays, sensory analysis and gas chromatography techniques. Food Chem. **206**, 37–43 (2016). https://doi.org/10.1016/j.foodchem.2016.03.032
16. A. Noviyanto, W.H. Abdulla, Honey botanical origin classification using hyperspectral imaging and machine learning. J. Food Eng. **265**, 109684 (2020). https://doi.org/10.1016/j.jfoodeng.2019.109684
17. D.A. Magdas, F. Guyon, C. Berghian-Grosan, M.C. Muller, Challenges and a step forward in honey classification based on Raman spectroscopy. Food Control **123**, 107769 (2021). https://doi.org/10.1016/j.foodcont.2020.107769
18. K. Rachineni, V.M. Rao Kakita, N.P. Awasthi, V.S. Shirke, R.V. Hosur, S.S. Chandra, Identifying type of sugar adulterants in honey: combined application of NMR spectroscopy and supervised machine learning classification. Curr. Res. Food Sci. **5**, 272–277 (2022). https://doi.org/10.1016/j.crfs.2022.01.008
19. S.P. Kek, N.L. Chin, S.W. Tan, Y.A. Yusof, L.S. Chua, Classification of honey from its bee origin via chemical profiles and mineral content. Food Anal. Methods **10**(1), 19–30 (2017). https://doi.org/10.1007/s12161-016-0544-0
20. H.M. Kayili, M. Atakay, A. Hayatu, B. Salih, Sample preparation methods for N-glycomics. Adv. Sample Prep. **4**, 100042 (2022). https://doi.org/10.1016/j.sampre.2022.100042
21. H.M. Kayili, R. Sakhta, B. Salih, Comparison of denaturing agent effects in enzymatic N-glycan release for human plasma N-glycan analysis. Turk. J. Chem. **46**(5), 1524 (2022). https://doi.org/10.55730/1300-0527.3457
22. D.B. Demirhan, H. Yılmaz, H. Erol, H.M. Kayili, B. Salih, Prediction of gastric cancer by machine learning integrated with mass spectrometry-based N-glycomics. Analyst **148**(9), 2073–2080 (2023). https://doi.org/10.1039/D2AN02057B
23. H.M. Kayili, B. Salih, Site-specific N-glycosylation analysis of human thyroid thyroglobulin by mass spectrometry-based glycoanalytical strategies. J. Proteomics (2022). https://doi.org/10.1016/j.jprot.2022.104700
24. P.S. Reel, S. Reel, E. Pearson, E. Trucco, E. Jefferson, Using machine learning approaches for multi-omics data analysis: a review. Biotechnol. Adv. **49**, 107739 (2021). https://doi.org/10.1016/j.biotechadv.2021.107739
25. T. Erban, E. Shcherbachenko, P. Talacko, K. Harant, The unique protein composition of honey revealed by comprehensive proteomic analysis: allergens, venom-like proteins, antibacterial properties, royal jelly proteins, serine proteases, and their inhibitors. J. Nat. Prod. **82**(5), 1217–1226 (2019). https://doi.org/10.1021/acs.jnatprod.8b00968
26. Y.-F. Zheng, M.-C. Wu, H.-J. Chien, W.-C. Wang, C.-Y. Kuo, C.-C. Lai, Honey proteomic signatures for the identification of honey adulterated with syrup, producing country, and nectar source using SWATH-MS approach. Food Chem. **354**, 129590 (2021). https://doi.org/10.1016/j.foodchem.2021.129590
27 H.M. Kayili, M. Atakay, A. Hayatu, B. Salih, Sample preparation methods for N-glycomics. Adva. Sample Prep. **4**, 100042 (2022)
28. C.M. West, D. Malzl, A. Hykollari, I.B.H. Wilson, Glycomics, glycoproteomics, and glycogenomics: an inter-taxa evolutionary perspective. Mol. Cell. Proteomics **20**, 100024 (2021). https://doi.org/10.1074/mcp.R120.002263

29. A. Noviyanto, W.H. Abdulla, Honey botanical origin classification using hyperspectral imaging and machine learning. J. Food Eng. (2020). https://doi.org/10.1016/j.jfoodeng.2019.109684
30. A.C. Soria, I. Martínez-Castro, J. Sanz, Study of the precision in the purge-and-trap–gas chromatography–mass spectrometry analysis of volatile compounds in honey. J. Chromatogr. A **1216**(15), 3300–3304 (2009). https://doi.org/10.1016/j.chroma.2009.01.065
31. C. Tananaki, A. Thrasyvoulou, J.L. Giraudel, M. Montury, Determination of volatile characteristics of Greek and Turkish pine honey samples and their classification by using Kohonen self organising maps. Food Chem. **101**(4), 1687–1693 (2007). https://doi.org/10.1016/j.foodchem.2006.04.042