

Analysis of volatile compounds in New Zealand unifloral honeys by SPME–GC–MS and chemometric-based classification of floral source

Laura E. Revell · Bruce Morris · Merilyn Manley-Harris

Received: 5 September 2013 / Accepted: 29 November 2013 / Published online: 8 December 2013
© Springer Science+Business Media New York 2013

Abstract New Zealand unifloral honeys have a higher commercial value than polyfloral honeys; however identification of floral source can be difficult and time-consuming. In this study, we aimed to establish a rapid and semi-automated method for identifying the floral source of New Zealand honeys. Volatile compounds from ten types of New Zealand unifloral honeys (a total of 234 samples) were analyzed by solid-phase microextraction (SPME) and gas chromatography coupled to mass spectrometry (GC–MS). For 37 compounds, probability plots of \log_{10} [GC–MS peak area] versus cumulative probability enabled visual identification of those that could be possible markers used to discriminate floral source. GC–MS peak areas were also analyzed by hierarchical cluster analysis and principal component analysis. Results showed data falling into groups based on floral source, indicating that supervised pattern recognition could be used to build a model with which to classify honeys based on floral source. A model was built using WEKA (Waikato Environment for Knowledge Analysis) machine-learning software. The logistic model tree algorithm in WEKA produced a model that classified 89.8 % of samples correctly. Overall, results show that the methods employed here have the potential to

be used as a basis for routine testing and classification of New Zealand unifloral honeys.

Keywords Honey · SPME · GC–MS · Botanical origin · Chemometrics

Introduction

Unifloral honeys (derived mostly from one floral source), as opposed to polyfloral honeys (derived from multiple floral sources), have characteristic sensory properties including odor and aroma, which make them of higher commercial value. Traditionally, identification of floral source has involved pollen analysis, together with determination of sensory and physico-chemical properties (color, moisture, pH, conductivity and sugar profiles), however such methods can be difficult and time-consuming [1, 2].

Volatile compounds are useful in linking honeys to their floral source, as the volatile fraction of honey can originate from the plant from which it was produced [1]. In previous studies, characteristic volatile compounds from European unifloral honeys were identified as markers, using a range of extraction techniques including ultrasound-assisted solvent extraction [3], simultaneous distillation–extraction [4], dynamic head-space extraction [5] and solid-phase microextraction, SPME [6]. In contrast to other extraction techniques, SPME requires minimal sample preparation and combines sample preparation and instrument introduction steps when it is coupled to GC–MS (gas chromatography–mass spectrometry), therefore making it promising as a convenient and semi-automated method for the extraction of honey volatiles from numerous samples.

Previous studies have applied multivariate statistical and chemometric methods to SPME–GC–MS data from

L. E. Revell · M. Manley-Harris
Department of Chemistry, University of Waikato,
Private Bag 3105, Hamilton, New Zealand

Present Address:

L. E. Revell
Institute for Atmospheric and Climate Science, ETH Zurich,
Zurich, Switzerland

B. Morris (✉)
Food and Bioanalytical Division, R. J. Hill Laboratories,
Private Bag 3205, Hamilton, New Zealand
e-mail: bruce.morris@hill-labs.co.nz

European and Argentinean honeys in order to build classification models based on floral source [7–12]. Such approaches, which involve analyzing honey volatile profiles as a whole, have been shown to yield superior classification rates compared with searching for individual marker compounds [8]. Based on the success of previous studies on European and Argentinean honeys, we have investigated whether SPME–GC–MS coupled with chemometric methods could form a suitable approach to classify New Zealand unifloral honey samples on a routine basis. New Zealand unifloral honeys have been subject to a number of composition analyses in the past, for example the recent studies using LC–MS (liquid chromatography–mass spectrometry) [13] and SIFT-MS (selected ion flow tube–mass spectrometry) [14]. To our knowledge, ours is the first study to analyze New Zealand unifloral honeys by SPME–GC–MS, with the exception of two studies focusing solely on mānuka honey [15, 16]. Here, we describe analysis of ten New Zealand unifloral honey types by SPME–GC–MS, identification of compounds that could serve as markers of floral source, and chemometric analysis of the volatile profiles.

Materials and methods

Standards

3-Methylpentanoic acid, *o*-methoxyacetophenone, 2-methylbenzofuran, myrtenal, hexanoic acid, 3-methyl-2-butenal, 2-ethylhexanoic acid, linalool and 1-phenylethanol were purchased from Sigma-Aldrich (St. Louis, MO). Thymol was obtained from BDH (United Kingdom), phenol from Univar (Seattle, WA), (*E*)-cinnamaldehyde from May and Baker (United Kingdom), 1-(2-methoxyphenyl)-ethanol from Acros Organics (Geel, Belgium), and dimethyl sulfide from Scharlau.

Samples

234 New Zealand honey samples, identified by the suppliers as unifloral via pollen counting and physico-chemical analysis, were obtained from Airborne Honey Ltd., Haddrell's of Cambridge, Waitemata Honey Co. Ltd., and the New Zealand Honey Food and Ingredient Advisory Service. These honeys were of the following floral types: pohutukawa (*Metrosideros excelsa*) 17 samples, thyme (*Thymus vulgaris*) 12 samples, mānuka (*Leptospermum scoparium*) 28 samples, southern beech honeydew (*Nothofagus* spp.) 26 samples, southern rata (*Metrosideros umbellata*) 23 samples, kamahi (*Weinmannia racemosa*) 18 samples, viper's bugloss (*Echium vulgare*) 23 samples, clover (*Trifolium repens*) 43 samples, tawari (*Ixerba*

brexioides) 28 samples, and rewarewa (*Knightia excelsa*) 16 samples. These are common unifloral honey types in New Zealand and originate from different regions. Pohutukawa, tawari and rewarewa are predominantly produced in the North Island; thyme, beech honeydew, southern rata and viper's bugloss are predominantly produced in the South Island. Mānuka, clover and kamahi are collected in both islands, with clover being the most common honey type in New Zealand. Samples were stored below 4 °C prior to analysis.

SPME

Fibers with the following coatings were obtained from Supelco (Bellafonte, PA): 85 μm carboxen/polydimethylsiloxane (CAR/PDMS), 50/30 μm divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS), 100 μm polydimethylsiloxane (PDMS), 70 μm carbowax/divinylbenzene (CAR/DVB), 65 μm polydimethylsiloxane/divinylbenzene (PDMS/DVB) and 85 μm polyacrylate (PA), and were used in an automated SPME holder. Fibers were conditioned in the GC injection port for the time and at the temperature suggested by the manufacturer, before being used for analyses. For each analysis the sample was weighed into an amber 2-mL vial which was capped using a magnetic crimp-cap with silicone/Teflon septum (Microliter Analytical). Caps and septa were heated (150 °C, 16 h) before use, to reduce siloxane levels. The SPME was automated, using a CTC CombiPal auto injector (CTC Analytics, Switzerland) mounted on the GC–MS, with a heated agitator unit and a fiber conditioning station. Before SPME, the sample was placed in the agitator to thermally equilibrate. During this time, the SPME fiber was heated (250 °C) in the conditioning station under a flow of N₂ gas to desorb contamination arising from volatile solvents present in the laboratory.

GC–MS

An Agilent 6890 GC with a split-splitless inlet and CTC Combi-pal robotic autosampler was interfaced with an Agilent 5793 MS (Agilent Technologies), with an electron impact source (set at 70 eV). GC used a DB-VRX capillary column (Agilent 122-1534; 30 m × 250 μm × 1.4 μm film thickness), with helium carrier gas (>99.995 %) at 1.2 mL/min (constant flow mode). SPME fiber desorption was carried out in splitless mode, with an injector temperature of 250 °C. A GC oven temperature program of 35 °C (4 min hold) to 225 °C (5 min hold) at 15 °C/min was used. The MS was operated in full scan mode, using a mass range of *m/z* 33–350. Compounds were identified by comparison of their mass spectra with the NIST (National Institute of Standards and Technology) 98 spectral library,

or by comparison of mass spectra and retention times with purchased standards.

Optimization of method parameters

The recovery of volatiles was optimized by varying parameters such as fiber type, temperature and extraction time. Good sensitivity at a reasonable analysis time was achieved by introducing honey (0.5 g) into a 2 mL vial, pre-incubating (70 °C, 5 min), then extracting (70 °C, 20 min) using a CAR/PDMS fiber (similar to the method used by Pérez et al. [17]). Some researchers have used water to decrease the density of the honey matrix [18, 19] and added salt to decrease the solubility of hydrophilic compounds in the aqueous matrix [6, 20]. However if water is not used, salt is not required to counteract the increased retention of hydrophilic compounds; hence was decided that neither water nor salt would be used.

Selection of marker compounds

The ten New Zealand unifloral honeys were analyzed by the chosen SPME method. Two chromatograms from each floral source were examined in detail and peaks were identified by matching their spectra with those in the NIST 98 mass spectral library, resulting in a list of 240 compounds. Ions and retention times for these compounds were entered into an Agilent Enviroquant method [21], which was used to examine the chromatograms of a further five samples from each floral source. The following compounds were eliminated from further analyses: acetone and ethyl acetate (common to all honey samples); compounds in the first 5 min of the chromatograms; siloxanes or peaks with siloxane interferences in the last 5 min of the chromatograms; compounds that were in a single sample of a unifloral type. 37 compounds were selected for analysis based on their presence in the majority of samples for one honey type, and confidence in their identification. Peak areas of these 37 compounds were integrated for all 234 honey samples and converted to their logarithm (base ten) to produce the data set that was subjected to chemometric and statistical analyses. Probability plots were used to further narrow the choice of marker compounds, as described below.

Data processing

The peak areas of the 37 identified compounds in all 234 honey samples were integrated using Agilent Enviroquant software from the SPME–GC–MS chromatogram of each honey sample. Due to the large spread of values, peak areas were converted to their logarithms (base ten) to give more normal probability distributions (as seen in the

linearity of probability plots with the y-axis set as a normal distribution). This formed the data set used in statistical analyses.

Statistical and chemometric analyses

Probability plots

Probability plots for each of the 37 compounds analyzed in the honey samples were generated using Minitab 15 [22]. Comparing the resulting data distributions for different honey types allowed possible markers for floral source to be identified. Estimated cumulative probability values were calculated using the Median Rank formula in Minitab, and the Normal probability distribution was used to transform the y-axis. This yielded linear plots for most compounds, allowing easy visual comparison. For compounds identified as potential markers by probability plots (those listed in boldface in Table 1), the independent two-sample Student's *t*-test was applied to ascertain whether there was a significant difference ($p < 0.05$) between abundances of those compounds in different honey types.

While some compounds are suitable as marker compounds for particular honey types (such as thymol for thyme honey and *p*-benzoquinone for viper's bugloss honey), it is difficult to find marker compounds for all honey types due to the complexity of the honey matrix. We therefore aimed to use supervised methods to create a chemometrics-based model capable of classifying floral source. Prior to this however, it was necessary to establish whether the underlying data structure was suitable for the application of supervised methods, by undertaking unsupervised multivariate statistical analyses.

Unsupervised learning techniques

These are multivariate statistical techniques in which class membership is unknown, and are helpful in revealing natural groupings within data. Two such techniques; hierarchical cluster analysis (HCA) and principal component analysis (PCA) were applied to the honey volatiles data set (all compounds listed in Table 1) using Matlab R2012b [23] (PCA) and Minitab 15 (HCA). Euclidean distances and Ward linkage were used for the HCA.

Chemometrics/supervised learning techniques

These were used to build a model which could discriminate between honey samples from different unifloral sources. Supervised learning algorithms contained within the data mining software WEKA (Waikato Environment for Knowledge Analysis) [24, 25] were applied to the honey

Table 1 Retention times, characteristic ions and relative abundances of volatile organic compounds in New Zealand unifloral honeys used in statistical analyses

	RT ^a	Ions ^b , <i>m/z</i>	Clover ^c	Honeydew	Kamahi	Mānuka	Pohutukawa	Southern rata	Rewarewa	Tawari	Thyme	Viper's bugloss
Dimethyl sulfide	2.69	62, 61, 47			2.5 (4.6)		40.0 (18.7)	20.4 (15.5)	5.7 (3.8)	5.7 (6.9)		
3-Methyl-2-butenal ^d	7.73	84, 55, 50				1.7 (1.2)						
Dimethyl sulfoxide ^d	9.43	96, 78, 63				2.5 (1.6)	3.2 (2.6)					
1,3,5,7-Cyclooctatetraene	10.49	104, 103, 78	2.6 (1.7)	1.6 (0.8)	0.3 (0.1)	0.6 (0.4)	1.2 (0.8)	1.1 (0.6)	1.4 (1.2)	2.9 (1.6)	1.3 (3.1)	2.2 (1.1)
<i>p</i> -benzoquinone	10.63	108, 82, 80										2.0 (0.9)
3-Methylpentanoic acid ^d	10.72	87, 60, 57	15.3 (13.7)	14.6 (12.6)	3.0 (3.3)	0.9 (0.8)	1.3 (0.5)	0.7 (0.2)	1.5 (0.8)	3.0 (2.6)	3.4 (2.7)	2.4 (1.3)
Hexanoic acid ^d	11.3	87, 73, 60									18.8 (7.5)	
5-Methyloxolan-2-one	11.36	100, 56, 42	1.9 (2.3)	2.2 (1.5)	5.3 (5.0)	1.6 (2.7)	2.0 (0.8)	2.9 (1.5)	3.4 (3.2)	1.8 (1.0)		1.3 (0.4)
Phenol ^d	11.40	94, 66, 65	3.3 (2.3)	3.3 (2.3)		0.4 (0.4)						
Benzaldehyde	11.47	106, 105, 77	6.8 (4.4)	6.1 (3.5)	1.0 (0.8)	1.6 (1.3)	11.5 (8.6)	1.7 (0.7)	2.7 (0.9)	5.7 (2.7)		11.8 (5.3)
1,1'-Bicyclopentyl	11.97	138, 95, 68	3.0 (2.2)		1.6 (0.7)	1.1 (1.0)	1.4 (0.4)	3.9 (2.4)	1.3 (0.5)	1.3 (0.7)		7.6 (0.1)
2,2,4,4-Tetramethylcyclobutane-1,3-dione	12.19	140, 70, 42	7.6 (10.1)	0.9 (0.3)	5.6 (4.1)	3.5 (4.5)	7.5 (5.4)	12.6 (6.5)	7.7 (3.9)	7.8 (6.3)	4.4 (3.6)	7.2 (9.4)
Benzyl alcohol	12.24	108, 107, 79	3.5 (3.0)	2.5 (1.7)	0.8 (0.6)	1.0 (0.8)	8.5 (5.0)	0.7 (0.5)	5.8 (2.6)	4.4 (3.1)	2.2 (2.0)	7.0 (6.7)
Pantoyl lactone	12.26	71, 57, 43	3.9 (2.5)	2.7 (1.6)	1.0 (0.4)	0.7 (0.6)	1.1 (0.7)	0.8 (0.4)	1.7 (0.7)	3.2 (2.3)	1.0 (1.3)	3.9 (1.8)
Benzeneacetaldehyde	12.46	120, 91, 65	8.1 (7.4)	2.9 (1.8)	1.2 (1.4)	3.2 (4.2)	2.0 (2.0)	1.8 (0.7)	3.2 (1.7)	5.2 (6.3)	5.1 (3.8)	37.8 (17.3)
1-Phenylethanol ^d	12.47	122, 107, 79				0.6 (0.4)						
4-Methyl-5 <i>H</i> -furan-2-one	12.51	98, 69, 41	1.9 (1.1)		7.9 (2.8)	2.1 (2.2)	2.1 (0.6)	2.2 (1.5)	3.0 (1.9)	3.8 (1.1)		
<i>Trans</i> -linalool oxide	12.68	111, 94, 59	19.6 (11.2)	1.9 (1.3)	2.4 (3.2)	2.6 (3.3)	1.8 (2.8)	5.9 (4.2)	4.5 (2.3)	3.7 (2.2)	2.2 (2.6)	
Acetophenone	12.79	120, 105, 77	0.5 (0.1)			0.7 (0.3)	0.8 (0.3)					
2-Ethylhexanoic acid ^d	12.89	116, 101, 88	6.7 (3.9)	2.0 (0.5)			1.9 (0.9)			8.6 (3.5)		
<i>Cis</i> -linalool oxide	12.90	124, 111, 94	8.6 (3.4)		2.0 (1.8)	2.0 (1.5)	4.2 (3.1)	2.7 (1.6)	3.1 (1.0)	8.8 (5.0)	2.2 (2.3)	6.6 (4.1)
Hotrienol	12.91	82, 71, 67	8.1 (5.0)	3.6 (0.9)	17.5 (3.5)	3.8 (2.5)	6.5 (4.1)	15.4 (3.7)	11.3 (3.3)	12.8 (4.1)	28.8 (9.4)	6.8 (3.0)
Linalool ^d	12.92	136, 121, 93	3.7 (1.2)	14.6 (10.1)	0.9 (0.5)	2.1 (1.5)	1.6 (1.5)		11.9 (1.9)	0.5 (0.2)		
Nonanal	13.03	98, 82, 70	4.4 (0.9)			4.8 (2.3)						
2-Phenylethanol	13.23	122, 92, 91	5.8 (4.2)	6.7 (2.8)	1.3 (1.4)	2.6 (2.3)	5.5 (2.8)	2.7 (4.6)	6.9 (4.0)	8.0 (5.7)	6.4 (15.2)	8.1 (6.5)
2-Methylbenzofuran ^d	13.37	132, 131, 103	0.5 (0.2)		0.2 (0.1)	0.5 (0.3)						
2,6-Dimethyl-1,3,5,7-octatetraene	13.44	134, 119, 91	0.3 (0.1)		0.3 (0.1)		0.3 (0.1)	0.3 (0.1)	0.8 (1.2)	0.4 (0.1)	0.4 (0.1)	
Benzoic acid	13.58	122, 105, 77	7.0 (5.3)	7.1 (4.2)	1.4 (1.3)	1.2 (0.7)	8.4 (8.5)	1.8 (1.2)	3.5 (2.4)	4.2 (3.2)	3.2 (2.0)	9.6 (6.2)
2,6,6-Trimethylcyclohexane-1,4-dione	13.87	154, 139, 69	2.4 (3.8)	2.5 (2.7)	4.5 (3.0)	3.5 (7.6)	0.4 (0.2)	5.2 (4.0)	2.9 (2.5)	1.6 (1.1)	3.0 (1.3)	1.4 (0.9)
2,6-Dimethyl-3,7-octadiene-2,6-diol	13.94	82, 71, 43	3.3 (3.0)	1.3 (0.6)	35.0 (13.8)	2.5 (3.1)	1.4 (2.5)	13.0 (4.8)	11.4 (4.7)	12.0 (8.6)	21.1 (7.5)	1.8 (1.1)
Terpineol	14.11	136, 121, 93	1.4 (0.7)	1.5 (0.7)		1.0 (0.9)	1.5 (0.9)	0.8 (0.3)	0.6 (0.2)	1.4 (0.3)		1.2 (0.4)
Myrtenal ^d	14.24	150, 135, 107	1.4 (0.4)			2.1 (1.6)						
(<i>E</i>)-cinnamaldehyde ^d	14.86	132, 131, 103				1.4 (1.2)						
Thymol ^d	14.96	150, 135, 115									2.4 (1.9)	
1-(2-Methoxy-phenyl)ethanol ^d	15.00	152, 137, 121	1.5 (1.2)	4.7 (3.0)	1.4 (0.8)	10.0 (15.6)	0.8 (0.0)		3.5 (3.3)	2.5 (1.3)		

Table 1 continued

	RT ^a	Ions ^b , <i>m/z</i>	Clover ^c	Honeydew	Kamahai	Mānuka	Pohutukawa	Southern rata	Rewarewa	Tawari	Thyme	Viper's bugloss
<i>o</i> -Methoxyacetophenone ^d	15.23	150, 135, 77	4.0 (5.2)	31.8 (24.4)	8.4 (9.0)	60.8 (24.6)	5.8 (8.6)	2.5 (2.2)	21.6 (11.4)	9.7 (13.0)	1.6 (1.5)	2.0 (2.2)
2,6-Dimethyl-2,7-octadiene-1,6-diol	15.43	137, 119, 71	3.3 (3.1)		5.2 (3.8)	0.9 (0.4)	1.1 (0.6)	2.1 (2.0)	3.0 (2.3)	3.1 (2.2)	5.4 (2.0)	

^a Retention time (min)

^b Base peak *m/z* selected ion chromatograms used to calculate peak areas

^c Mean relative abundances in % and standard deviation (in brackets). The relative abundance was calculated by expressing the peak area as a percentage of the combined peak areas for all compounds. Data is displayed only for compounds present in two or more samples

^d Compound identification was performed with reference to purchased standards. All other compounds were tentatively identified with reference to the NIST 98 MS library

volatiles data set (all compounds listed in Table 1). The J48, logistic model tree (LMT), logistic, IBk (*K*-nearest neighbor) and classification via regression algorithms were studied. Ten-fold cross-validation was used in building and testing models; this procedure enabled as much data as possible to be used for training, while still allowing a reasonable error estimate to be obtained [25]. Results from WEKA were in the form of confusion matrices showing the number of each class (honey type) assigned to that class or to other classes, giving an overall classification rate (percent of samples correctly classified).

Results

Selection of marker compounds

The total ion chromatograms obtained by SPME of the ten New Zealand unifloral honeys are shown in Fig. 1. Elimination of common, spurious and dubious peaks and confirmation of the appearance of the selected peaks in all samples of the specified unifloral honey resulted in a list of 37 possible marker compounds (Table 1), identified by comparison of their mass spectrum with the NIST 98 MS library. This list was then refined further by use of probability plots (Fig. 2) to identify a list of potential marker compounds (compounds listed in boldface in Table 1), and the identity of each of the chosen compounds was confirmed by comparison of their mass spectra and retention times with purchased standards.

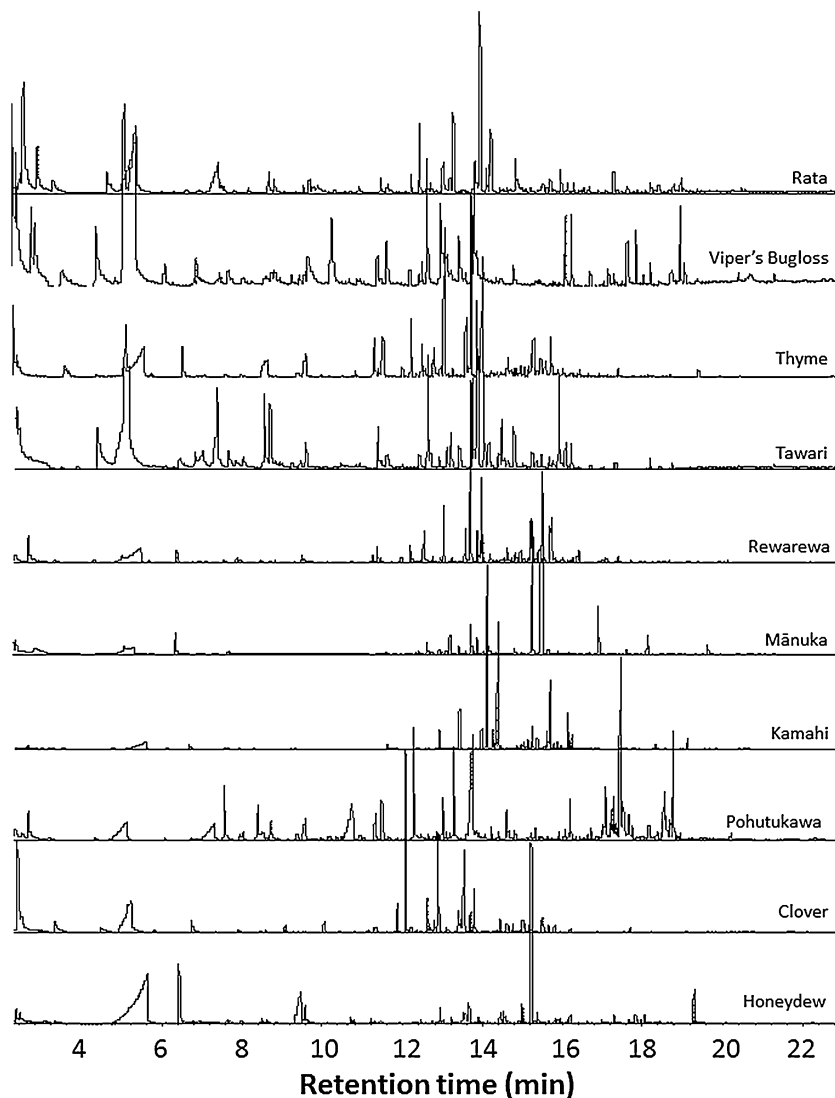
For each honey type, the independent two-sample *t*-test was used to identify statistically significant differences between the abundances of the chosen marker compounds compared with other honey types. Statistically significant differences ($p < 0.05$) were found for most compounds and unifloral types with a few exceptions; there was no statistically significant difference between relative amounts of dimethyl sulfoxide in pohutukawa and southern rata honeys, or between relative amounts of *o*-methoxyacetophenone in tawari and rewarewa honeys. In addition, 1-(2-methoxyphenyl)-ethanol was not present in statistically significantly different relative amounts between honeydew and rewarewa honeys, and honeydew and tawari honeys.

Unsupervised learning

HCA

Hierarchical cluster analysis revealed clusters of honey samples, with each cluster corresponding to predominantly one type of honey. Table 2 displays information on how many samples were correctly classified per

Fig. 1 Representative total ion chromatograms obtained by SPME of the ten New Zealand unifloral honey types analyzed



cluster, and how many samples were correctly classified per honey type. Three of the clusters contained 100 % of one honey type (thyme, pohutukawa and tawari), and two or fewer samples were misclassified for thyme, honeydew, mānuka, pohutukawa and rewarewa honeys.

PCA

The first two components obtained by PCA accounted for 40 % of variation. The score plot for the first two components (Fig. 3) indicates distinction between honey types, although with only 40 % of variance explained by these first two components, there was some overlap between clusters (perhaps with the exception of mānuka honey); 10 components were needed to explain 80 % of the variance.

1,1'-bicyclopentyl and 4-methyl-5*H*-furan-2-one gave large contributions to the first component, and 1-(2-methoxyphenyl)ethanol and 2-methylbenzofuran gave large contributions to the second component.

Chemometrics/supervised learning

Of the five algorithms tested on the honey volatile data set, the LMT algorithm yielded the most favorable results in terms of large true classification rates and small false classification rates. This algorithm classifies data by building decision trees with linear logistic regression models at nodes [25]. The confusion matrix resulting from the LMT model is shown in Table 3 and the evaluation statistics are in Table 4.

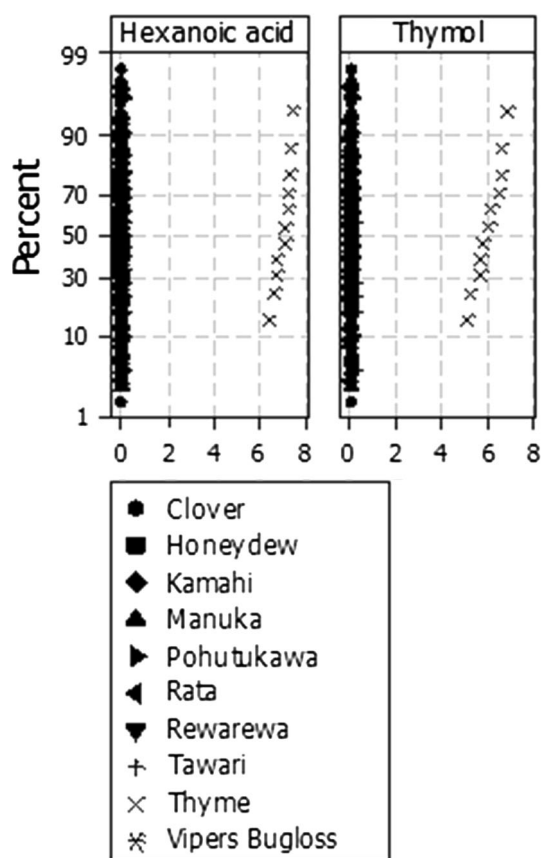


Fig. 2 Examples of probability plots for hexanoic acid and thymol; the clear separation of thyme from other honey types indicates that these compounds are unique to thyme honey

Discussion

Markers specific to unifloral honeys

Pohutukawa and southern rata honey

These species are both in the genus *Metrosideros*, and both contained significantly greater relative amounts of dimethyl sulfoxide and/or dimethyl sulfide than other honeys analyzed. As a percentage of the total peak area, the peak area of dimethyl sulfide averaged 40 % in the pohutukawa honeys and 20 % in the southern rata honeys. The relative peak area of dimethyl sulfoxide averaged 2.5 % in the pohutukawa honeys and 3.2 % in the southern rata honeys. This suggests that volatile sulfur compounds are characteristic of the *Metrosideros* genus (Table 1). These compounds have a low odor threshold, and are therefore likely to contribute significantly to the aroma of these honeys. Dimethyl sulfoxide has previously been found in Spanish rosemary honey, and could result from oxidation of dimethyl sulfide [26]. Dimethyl sulfide was found in lesser relative amounts in other “bush” honeys (from wild forest areas and possibly containing contributions from

Table 2 Classification rates obtained from HCA for clusters of honey samples

a	b	c
Clover	78	33
Viper's bugloss	75	22
Thyme	100	17
Honeydew	79	0
Mānuka	96	7
Pohutukawa	100	12
Rewarewa	50	13
Tawari	100	39
Kamahi	65	17
Southern rata	89	26

Column a Honey type predominantly associated with each cluster

Column b Percentage of observations of the dominant honey type within the cluster

Column c Percentage of samples of the type listed in *column a* assigned to a cluster other than the dominant cluster for that honey type

Metrosideros spp.); rewarewa, kamahi and tawari, although with a smaller average peak area, of 5.7 % for rewarewa and tawari honeys and 2.5 % for kamahi. It has also been found by SPME–GC–MS in Spanish orange, eucalyptus, rosemary, lavender and thyme honeys [17], and in Cuban black mangrove honey [27]. In the present study the relatively high level and presence of dimethyl sulfide in all southern rata samples made it a useful marker, while dimethyl sulfoxide was only found in pohutukawa and southern rata, plus one viper's bugloss sample.

Other possible markers for pohutukawa were 3-methyl-1-2-butenal and (*E*)-cinnamaldehyde. The former has been found in Spanish citrus, rosemary and polyfloral honeys [28] and in a mixture of honeys from Europe and Brazil [26]; the latter in Greek cotton honey [29]. However, of the ten New Zealand unifloral honeys analyzed in this study, 3-methyl-2-butenal and (*E*)-cinnamaldehyde were found only in pohutukawa honey and may be useful markers for this honey type in New Zealand.

Thyme honey

Thymol was only found in thyme honey samples (Table 1), and has also been found in Greek and Palestinian thyme honeys, along with Spanish honeys from a large variety of floral sources [30–33]. Thymol was considered a marker in rosemary and *Dorycnium pentaphyllum* honeys [33] and in lime tree (*Tilia* spp.) honey from France [4]. Since none of these honey types apart from thyme are produced in New Zealand it appears that thymol could be a useful marker compound for New Zealand thyme honey, although this could be compromised if thymol is used for control of

Fig. 3 Score plot showing the first two principal components obtained from PCA; these explained 40 % of the variance and show clustering of the honey types due to their volatile composition

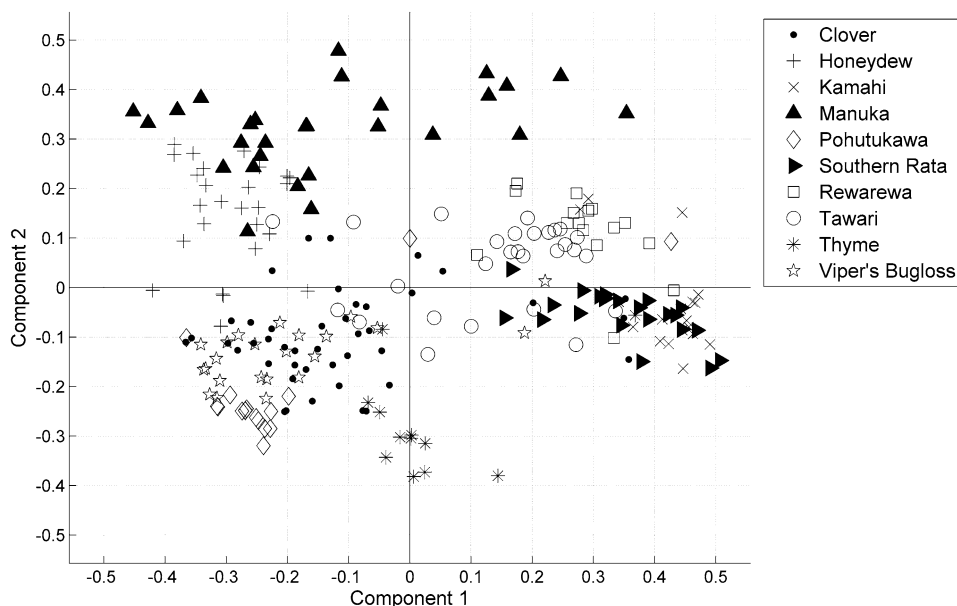


Table 3 Confusion matrix showing results of the LMT algorithm applied to honey data

Classification	Clover	Honeydew	Kamahī	mānuka	Pohutukawa	Southern rata	Rewarewa	Tawari	Thyme	Viper's bugloss
Clover	37	1	0	0	0	0	0	4	0	1
Honeydew	1	25	0	0	0	0	0	0	0	0
Kamahī	0	0	17	0	0	0	1	0	0	0
Mānuka	1	0	1	26	0	0	0	0	0	0
Pohutukawa	0	0	0	0	15	2	0	0	0	0
Southern rata	0	0	1	0	0	21	0	1	0	0
Rewarewa	1	0	0	0	0	0	15	0	0	0
Tawari	1	1	0	1	0	0	1	24	0	0
Thyme	0	0	0	0	0	0	0	1	11	0
Viper's bugloss	4	0	0	0	0	0	0	0	0	19

Rows show how the samples of a particular honey were classified. Columns show, for a particular honey, how many of the other honeys were misclassified as that honey type. Correct classifications are shown on the left to right diagonal

varroa mites, *Varroa destructor* [34]. Hexanoic acid was also only found in the thyme honeys, and has previously been identified in European [6, 32, 35] and Cuban [10] honeys from a variety of sources, but again none of these honey types are produced in New Zealand. Neither hexanoic acid nor thymol were identified in a previous study on New Zealand thyme honey [36], possibly due to the liquid-liquid extraction or derivatisation procedures used, although 3-hexenoic acid was reported. The degraded carotenoid marker 1-(3-oxo-*trans*-1-butenyl)-2,6,6-trimethylcyclohexane-*trans*, *cis*-1,2,4-triol previously isolated from New Zealand thyme honey [36] was not detected in SPME-GC-MS chromatograms, possibly due to its relatively high molecular weight and high polarity.

Mānuka honey

Myrtenal ((1R)-6,6-dimethylbicyclo[3.1.1]hept-2-en-2-carboxaldehyde), 1-phenylethanol, 2-methylbenzofuran and 1-(2-methoxyphenyl)ethanol appeared to be useful markers for mānuka honey. There was a significant difference between the relative amounts of these compounds in mānuka and other honeys (independent two-sample *t*-test [$p < 0.05$]), with the exception of 1-(2-methoxyphenyl)ethanol. Myrtenal was also extracted at lower levels from two clover samples and one pohutukawa sample, and 2-methylbenzofuran from six honeydew, one rewarewa and two kamahi honeys.

1-(2-Methoxyphenyl)ethanol was present in relatively high levels in all mānuka honeys tested (Table 1), however

Table 4 Evaluation statistics for the model created with the LMT algorithm and ten-fold cross-validation

TPR ^a	FPR ^b	P ^c	R ^d	ROC ^e	Class
0.860	0.042	0.822	0.860	0.957	Clover
0.962	0.010	0.926	0.962	0.998	Honeydew
0.944	0.009	0.895	0.944	0.971	Kamahi
0.929	0.005	0.963	0.929	0.988	Mānuka
0.882	0	1	0.882	0.998	Pohutukawa
0.913	0.009	0.913	0.913	0.996	Southern rata
0.938	0.009	0.882	0.938	0.942	Rewarewa
0.857	0.029	0.800	0.857	0.965	Tawari
0.917	0	1	0.917	0.982	Thyme
0.826	0.005	0.95	0.826	0.958	Viper's bugloss

^a True positive rate

^b False positive rate

^c Precision = (number of samples of a particular class correctly classified)/(number of samples classified as that class)

^d Recall = (number of samples of a particular class correctly classified)/(number of samples in that class)

^e ROC area = Receiver Operating Characteristic. This characterizes the trade-off between the true positive rate and false negative rate (should be close to one but not one exactly as this is an indication of data over-fitting)

it was also found in most other honey types tested at levels that were not statistically significantly different by Student's *t*-test ($p < 0.05$) to those in mānuka honey. The aromatic acids 2-hydroxy-3-phenylpropanoic acid (phenyllactic acid), 2-methoxybenzoic acid and 4-hydroxy-3,5-methoxybenzoic acid (syringic acid), previously found in mānuka honey by liquid–liquid extraction [37], were not detected in SPME–GC–MS chromatograms, likely due to their high polarity and poor extractability by SPME. However, these compounds were found in mānuka honey by liquid chromatography–tandem mass spectrometry [13]. Similarly 2-methoxybenzoic acid, 4-methoxyphenyllactic acid, phenylacetic acid, phenyllactic acid and methyl syringate (methyl 3,5-dimethoxy-4-hydroxybenzoate), which were found to be useful markers for mānuka honey by HPLC (high performance liquid chromatography) following solid phase extraction [38] were not detected here.

Myrtenal has not been identified in honey before. 1-(2-Methoxyphenyl)ethanol is also a new honey volatile, although its oxidation product *o*-methoxyacetophenone was identified in SPME–GC–MS chromatograms in the current study; it was also found in mānuka honey as a major component following liquid–liquid extraction [37] and by SPME–GC–MS [15]. *o*-Methoxyacetophenone and the related hydroxy species have been identified previously as marker compounds by head-space SPME–GC–MS of mānuka honeys together with benzofuran and dimethylbenzofuran [16];

differences between this result and the current study could be attributed to the different solid phase employed for the extraction in the previous study or to variations incurred in storage or trans-shipment overseas.

2-Methylbenzofuran has previously been identified in a mixture of European and Brazilian honeys by SPME [18], and in polyfloral honeys from Portugal, together with 1-phenylethanol [39]. These compounds were not detected in previous studies of mānuka honeys [37, 40, 41]. 1-Phenylethanol has been extracted as a characteristic volatile from chestnut (*Castanea sativa*) honey [4, 33], however this honey is not produced in New Zealand and few chestnut trees are present in locations where mānuka honey is collected.

Southern beech honeydew honey

There were significantly greater relative amounts of phenol and linalool (Student's *t*-test [$p < 0.05$]) in honeydew honey, compared with the other honey types. Phenol was present in only seven other samples (six mānuka and one pohutukawa), while linalool was also found in 50 % of mānuka and pohutukawa samples, together with four clover, three kamahi, two thyme, one southern rata and two tawari samples. Linalool may therefore be a poor marker for southern beech honeydew honey. Phenol has previously been found in several different unifloral honeys, including lavender (*Lavandula stoechas*) [33], chestnut [4, 33, 35], heather (*Calluna vulgaris*) [42–44] and lime-honeydew [44], although was not considered characteristic in these honeys due to the relatively low levels. Of these honey types only heather honey is collected in New Zealand, and was not available for inclusion in this study (heather is designated a pest species and efforts are underway to eradicate it), therefore the relative amounts of phenol extracted by SPME from heather and southern beech honeydew honeys could not be compared. However southern beech honeydew honey has distinctive physico-chemical properties distinguishing it from heather honey.

Kamahi, tawari and rewarewa honeys: "bush honeys"

These three honey types have >45 % pollen from the tree species kamahi (*W. racemosa*), tawari (*I. brexioides*), and rewarewa (*K. excelsa*), that grow in New Zealand native forests ("bush"); therefore they may have common nectar sources among other forest trees apart from these. Significantly greater levels of 4-methyl-5*H*-furan-2-one were extracted from kamahi than from other honey samples in this study (Student's *t*-test [$p < 0.05$]), however this compound was also present in lower levels in southern rata, rewarewa and tawari honeys (all originating from forest areas), therefore either these honeys contain kamahi nectar,

or 4-methyl-5H-furan-2-one originates from other bush sources.

70 % of tawari samples contained significantly more 2-ethylhexanoic acid (Student's *t*-test [$p < 0.05$]) than other honeys, but this compound was also found in lesser relative amounts in 80 % of pohutukawa honeys. No possible volatile marker compounds were revealed for rewarewa honey, with the relative amounts of 4-methyl-5H-furan-2-one not significantly different (Student's *t*-test [$p < 0.05$]) from those in southern rata and tawari honeys, although the levels in southern rata, kamahi, tawari and rewarewa were all significantly greater (Student's *t*-test [$p < 0.05$]) than those in the other six honeys tested. This means that 4-methyl-5H-furan-2-one could be used as a marker for honey from New Zealand lowland native forests, although it has not been previously reported as a honey constituent. Other furan-2-ones have been found in Japanese haze honey (3-methyl-3H-furan-2-one) [45] and Spanish citrus honey (5-methyl-3H-furan-2-one) [32].

Viper's bugloss honey

The presence of *p*-benzoquinone was characteristic of viper's bugloss samples and was absent from all other honey types, therefore it could be used as a marker. *p*-Benzoquinone, together with hydroquinone, was found in viper's bugloss honey by liquid-liquid extraction in a previous study [46], in which hydroquinone was proposed as a marker. These compounds appear to be unique to viper's bugloss honey, although hydroquinone was not found in the current study.

Chemometric analysis

The results from the HCA and PCA indicate that the underlying data structure was suitable for supervised learning techniques to be applied; HCA sorted samples into clusters of predominantly one honey type even through class membership was unknown. Clusters of honeys produced by PCA (in which class membership was also unknown) are shown in Fig. 3, although because this figure represents only the first two components of the PCA (and ten components were needed to explain 80 % of the variance), the clusters are therefore not clearly defined. The results from the unsupervised learning techniques are sufficiently promising to justify the application of supervised learning algorithms, as was done with WEKA software. The LMT algorithm was relatively successful in classifying tawari and viper's bugloss honeys, which were problematic honey types when other algorithms were used. Furthermore, the rate of true positives for all honey types was high and the rate of false positives low. High rates of correct classification were reflected by the high precision, recall

and receiver operating characteristic area values (Table 4). Overall, the classification rate was 89.8 %.

The results presented here show that a method coupling SPME-GC-MS and chemometric analysis has the potential to form the basis of a routine testing method for the New Zealand honey industry. The probability plots and LMT algorithm provided by WEKA software were both able to classify the floral source of New Zealand honeys with approximately 90 % success. Future research should focus on testing larger numbers of samples to improve model performance.

Acknowledgments The authors would like to thank Ray Littler and Lyn Hunt of the Statistics Department at The University of Waikato for their help with the statistical analyses involved in this work. LER was supported by a Technology New Zealand Education Fellowship.

References

1. E. Anklam, *Food Chem.* **63**, 549 (1998)
2. S. Bogdanov, K. Ruoff, L. Persano Oddo, *Apidologie* **35**, 4 (2004)
3. E. Alissandrakis, D. Daferera, P.A. Tarantilis, M. Polissiou, P.C. Harizanis, *Food Chem.* **82**, 575 (2003)
4. C. Guyot, A. Bouseta, V. Scheirman, S. Collin, *J. Agric. Food Chem.* **46**, 625 (1998)
5. F. Bianchi, M. Careri, M. Musci, *Food Chem.* **89**, 527 (2005)
6. A. Verzera, S. Campisi, M. Zappalà, I. Bonaccorsi, *Am. Lab.* **33**, 18 (2001)
7. E. de la Fuente, I. Martínez-Castro, J. Sanz, *J. Sep. Sci.* **28**, 1093 (2005)
8. M.V. Baroni et al., *J. Agric. Food Chem.* **54**, 7235 (2006)
9. K.A. Aliferis, P.A. Tarantilis, P.C. Harizanis, E. Alissandrakis, *Food Chem.* **121**, 856 (2010)
10. L. Ceballos, J.A. Pino, C.E. Quijano-Celis, A. Dago, *J. Food Qual.* **33**, 507 (2010)
11. S. Daher, F. Gülaçar, *J. Agric. Food Chem.* **56**, 5775 (2008)
12. T. Čajka, J. Hajslova, F. Pudil, K. Ridellova, *J. Chrom. A* **1216**, 1458 (2009)
13. J.M. Stephens et al., *Food Chem.* **120**, 78 (2010)
14. V. Langford, J. Gray, B. Foulkes, P. Bray, M.J. McEwan, *J. Agric. Food Chem.* **60**, 6806 (2012)
15. S. Daher, F. Gülaçar, *E-J. Chem.* **7**(S1), S7 (2010)
16. S. Oelschlaegel, *Characterisation of unifloral honeys using HS-SPME/GC/MS, 42nd Apimondia* (Buenos Aires, Argentina, 2011)
17. R.A. Pérez, C. Sánchez-Brunete, R.M. Calvo, J.L. Tadeo, *J. Agric. Food Chem.* **50**, 2633 (2002)
18. T. Čajka, J. Hajslova, J. Cochran, K. Holadová, E. Klimánková, *J. Sep. Sci.* **30**, 534 (2007)
19. A.C. Soria, I. Martínez-Castro, J. Sanz, *J. Sep. Sci.* **26**, 793 (2003)
20. L. Piasenzotto, L. Gracco, L. Conte, *J. Sci. Food Agric.* **83**, 1037 (2003)
21. Enviroquant Chemstation, G1701BA Version B.01.00 (Agilent Technologies)
22. Minitab Inc., Minitab 15, www.minitab.com
23. MathWorks, Matlab R2012b, www.mathworks.com
24. WEKA (Waikato Environment for Knowledge Analysis), version 3.5.4, <http://www.cs.waikato.ac.nz/ml/weka>
25. I.H. Witten, E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd edn. (Morgan Kaufmann, San Francisco, 2005), p. 560

26. L. Castro-Vázquez, M.S. Pérez-Coello, M.D. Cabezedo, *Chromatographia* **57**, 227 (2003)
27. J.A. Pino, *Int. J. Food Sci. Tech.* **47**, 1688 (2012)
28. I. Escriche, M. Visquert, M. Juan-Borrás, P. Fito, *Food Chem.* **112**, 329 (2009)
29. E. Alissandrakis, A.C. Kibaris, P.A. Tarantilis, P.C. Harizanis, M. Polissiou, *J. Sci. Food Agric.* **85**, 1444 (2005)
30. E. Alissandrakis, P.A. Tarantilis, P.C. Harizanis, M. Polissiou, *J. Agric. Food Chem.* **55**, 8152 (2007)
31. I. Odeh et al., *Food Chem.* **101**, 1393 (2007)
32. L. Castro-Vázquez, M.C. Díaz-Maroto, M.S. Pérez-Coello, *Food Chem.* **103**, 601 (2007)
33. J.S. Bonvehí, F.V. Coll, *J. Sci. Food Agric.* **83**, 275 (2003)
34. S. Adamczyk, R. Lázaro, C. Pérez-Arquillue, P. Conchello, A. Herrera, *J. Agric. Food Chem.* **53**, 10085 (2005)
35. I. Jerković, J. Mastelić, Ž. Marijanović, Z. Klein, M. Jelić, *Ultrason. Sonochem.* **14**, 750 (2007)
36. S.T. Tan, A.L. Wilkins, P.T. Holland, T.K. McGhie, *J. Agric. Food Chem.* **38**, 1833 (1990)
37. S.T. Tan, P.T. Holland, A.L. Wilkins, P.C. Molan, *J. Agric. Food Chem.* **36**, 453 (1988)
38. S. Oelschlaegel, M. Gruner, P.N. Wang, A. Boettcher, I. Koelling-Speer, K. Speer, *J. Agric. Food Chem.* **60**, 7729 (2012)
39. M. Pontes, J.C. Marques, J.S. Câmara, *Talanta* **74**, 91 (2007)
40. A.L. Wilkins, Y. Lu, S.T. Tan, *J. Agric. Food Chem.* **41**, 873 (1993)
41. R.J. Weston, L.K. Brocklebank, Y.R. Lu, *Food Chem.* **70**, 427 (2000)
42. G. Guyot, V. Scheirman, S. Collin, *Food Chem.* **64**, 3 (1999)
43. S.T. Tan, A.L. Wilkins, P.T. Holland, T.K. McGhie, *J. Agric. Food Chem.* **37**, 1217 (1989)
44. T. Wolski, K. Tambor, H. Rybak-Chmielewska, B. Kedzia, *J. Apic. Sci.* **50**, 115 (2006)
45. M. Shimoda, Y. Wu, Y. Osajima, *J. Agric. Food Chem.* **44**, 3913 (1996)
46. A.L. Wilkins, S.T. Tan, P.C. Molan, *J. Apic. Res.* **34**, 73 (1995)