Classification and identification of Vietnamese honey using chemometrics based on ^{1}H-NMR data

The Anh Nguyen, Truong Giang Vu, Thi Lan Nguyen, Quang Trung Pham, Thi Thao Ta[1*]

Faculty of Chemistry, University of Science, Vietnam National University

Received 12 April 2017; accepted 2 June 2017

Abstract:

Honey is a natural, sweet, and syrupy fluid which has been used in Vietnam in a variety of ways; as a food supplement, beauty product, and natural drug. However, quality control and characterization of honey are blind problems. Consumers, and even market management committees, must believe in the producer's quality standards without using any special techniques to evaluate the botanical origins of honey in the Vietnamese market. The chemical composition and physical properties of natural honey vary per plant species of which the honey bees scrounged. Longan flower-honey has a high price and is commercially produced in Yen Bai, Bac Giang, and the more well-known Hung Yen Province in Vietnam, and now is being confused with original flower honey. In this work, a total of 57 honey samples (longan and non-longan) from different geographic and botanical origins have been analysed in terms of ^{1}H-NMR spectroscopy, coupled also with multivariate statistical analysis methods. Principal component analysis followed by icoshift algorithm analysis comes about as a proficient device in recognising ^{1}H-NMR spectra of longan honey samples.

Keywords: ***botanical origin, chemometrics, classification, ^{1}H-NMR, identification, Vietnamese honey.***

Classification number: ***2.2***

Introduction

As stated in the Codex Alimentarius Commission, honey is defined as a characteristic substance created by honey bees and is comprised of water and sugars, primarily fructose and glucose [1]. Other minor compounds include proteins, amino acids, flavours and aromatic molecules, pigments, vitamins, and numerous unpredictable parts establishing nutritious and organoleptic qualities. Honey is a global product due to its promptly accessible source of vitality, and its antibacterial and antioxidant capacities [2, 3]. Bees are considered to produce honey to serve as their main source of food during times of scarcity or harsh weather conditions. Bees transform pollen from flowers and trees of various kinds to produce honey, including both in-house trees and forest trees.

Currently, Vietnam ranks sixth in the world in regards to honey export. According to the Vietnamese Beekeepers Association, in 2013, the total domestic production of honey was more than 48,000 tonnes, with 37,000 tonnes exported. Recently, the honey export growth rate has steadily increased at a high rate (14%) [4]. Due to high market demand for forest honey, which often demands a much higher price, local producers often mix honey from various original and botanical sources, including lychee, coffee, *Melaleuca leucadendron L.,* and especially longan; they have been known to also mix money with sugar.

Longan (*Dimocarpus longan*) is an evergreen fruit crop grown in tropical and subtropical climates and is considered as a traditional fruit of Vietnam, having its main production areas in the south: Tien Giang, Ben Tre, Dong Thap, Vinh Long, Can Tho, and Ba Ria-Vung Tau; and in the north: Bac Giang, Lao Cai, Yen Bai, Thai Nguyen, Phu Tho, Son La, Hung Yen, and Thanh Hoa. In 1997, the total area planted with longan was 60,000 ha, and grew to reach to 75,200 ha in 2002 [5]. Because of a higher price of longan honey than synthesised honey, it is necessary to control the honey quality and authenticity in order to preserve the production areas, to develop quality standards, and to protect consumers from commercial speculation. Vietnamese officials are encouraging the development of new analytical methods to control and verify quality specification for honey with different botanical origins, quality controls, and original trademarks.

As of late, numerous different studies have been published to develop new

*Corresponding author: Email: thaott73@gmail.com

methodologies and diverse analytical techniques to evaluate either different or equal botanical and geographical origin of honey [6, 7]. Among the greater analytical methods applied in food characterization, nuclear magnetic resonance (NMR) is accepted as a powerful and trusted method [6-8] due to its non-destructible aspects, high reproducibility, and sensitivity as shown across a large range of utilizations. In contrast with chromatography, NMR requires a small amount of sample and simple sample preparation, so it can be used to perform metabolite characterization of honey for either geographical assessment or botanical assessment [9, 10].

The ^{1}H-NMR spectra of poly floral and honey samples were recorded and geographically characterised [11]. The classification of Brazilian honey's botanical origin by Principal Component Analysis and Hierarchical Cluster Analysis also using NMR analysis were investigated [12]. Independent component analysis had been also used to discriminate manuka honey from other floral honey types [13]. Factor analysis and general discriminant analysis was successfully applied to detect Honey Adulteration by Sugar Syrups [14]. The icoshift algorithm is based on the shift of spectral intervals, and is employed across all spectra simultaneously. The icoshift program is an open source and highly efficient program designed to solve signal alignment problems in metabolomic analysis [15]; however, it has not yet been applied to the ^{1}H-NMR spectra of honey.

In this work, the ^{1}H-NMR spectra of honey in water solvent was applied using a pulse sequence NOESYPR1D to saturate the signal of the solvent. The advantage of this method is that it has low cost and easy usage to prepare. A total of 57 honey samples coming from Vietnamese longan and other botanical origins were studied. By using Principal components analysis (PCA) combined with mean-centering calculation and icoshift tool, ^{1}H-NMR spectra have been used for building a model and identifying longan honey among different honey samples.

Materials and methods

Materials and sample collections

A total of 57 honey samples were collected on the trading market. The original and botanical information of the samples was recognised based on its packaging and onsite information. Among the samples, there were 18 longan honey samples, 10 non-longan honey samples (coming from other fruits) and 29 test samples recognised as non-identified samples.

NMR analysis

An NMR solvent was prepared from double distillated deionized water and deuterated water (9:1 in volume). A 0.1 ml of the sample was dissolved in 0.3 ml of the H_2O/D_2O solvent. The ^{1}H-NMR spectra were recorded at 300 K using a Bruker Advance 500 MHz (Bruker Biospin, Germany) operating at 11,7 T with a 5 mm BBFO probe. Solvent suppression was achieved by applying a presaturation scheme with low-power radiofrequency irradiation. The number of data points was 32 K, acquisition time was 2.04 s, the number of scans was 8 and spectral width was 8,012,820 Hz. An exponential function of LB 0.3 was applied before Fourier transformation, and the phase and baseline were automatically corrected using Topspin 3.2 (Bruker Biospin, Germany).

Statistical methods

NMR data was aligned, changed over into Excel 2016 (Microsoft) then transported into Matlab R2016a (The MathWorks, USA) for statistical analysis. Principal component analysis (PCA) was performed with mean-centering as a data pretreatment.

PCA is a chemometric standout method amongst unsupervised ones used in analysing NMR data. It is an essential statistical tool for introductory examinations of extensive data sets to investigate likely patterns, classifications and identification of outliers. The goal of the principal component analysis is to explain the maximum amount of variance with the fewest number of principal components [11]. This method includes a dimensional reduction of the data set using a smaller number of axes. These components (PCs) are shown graphically as a score plot, which is a summary of the relationship among the observations. Coefficients, by which the original variables are multiplied to obtain the PCs, are represented in loading plots that summarise the variables (chemical shift data points) - which is a means to interpret the patterns seen in the score plot [16]. Samples (or observations) that were similar, or highly correlated with one another, were closed in the same group, whereas samples that were dissimilar, or uncorrelated, were clustered in different groups. The higher eigenvalues, the more information of PCs contains the original data matrix [17].

One of the most common normalising methods is mean-centering, which calculates the mean of each column and subtracts this from the column itself. Another way of interpreting mean-centered data is that each row of the mean-centered data includes only the differences of each row from the average sample in the original data matrix. In other words, mean-centering involves the subtraction of the variable averages from the data.

Icoshift toolbox for Matlab is an open source tool provided by the University of Copenhagen. The icoshift algorithm represents a powerful and versatile tool used for dealing with all kinds of signal alignment problems. It allows the researcher to choose among a large variety of options, from fully automated corrections of the whole NMR spectrum

to supervised and targeted interventions covering only selected spectral regions [15].

Results and discussions

Chemical characterization of honey samples

Fig. 1 represents a complete spectrum of a longan honey sample from 0 to 9 ppm in chemical shift. The spectra of 57 honey samples were combined in Fig. 2. It can be seen that, the main compounds show their dominant resonances at regions of 1-2 ppm and from 3 to 5.5 ppm. The spectra were exported as text files from Topspin software into Matlab software to study the characteristics, identification and classification.

To specify compounds that characterise each part of the spectra, the whole spectra was divided into three main regions: 1-3 ppm, 3-5.5 ppm, and 6-8 ppm as shown in Fig. 3. The first region, 1-3 ppm, shows the appearance of two main peaks: lactic acid (1 ppm) and acetic acid (2.1 ppm). The region 3-5.5 ppm shows the percent of carbohydrates, and dominant resonances of main monosaccharides, like: (α - and β - glucopyranose, β - fructopyranose, α - and β - fructofuranose). For instance, α and β - anomeric hydrogen of glucopyranose could be recognised at 5.2 and 4.6 ppm [17]. The last region, 6-8 ppm, represents formic acid and some the aromatic amino acids including tyrosine, phenylalanine; in here, almost peaks have the too small intensity and are not convergent.

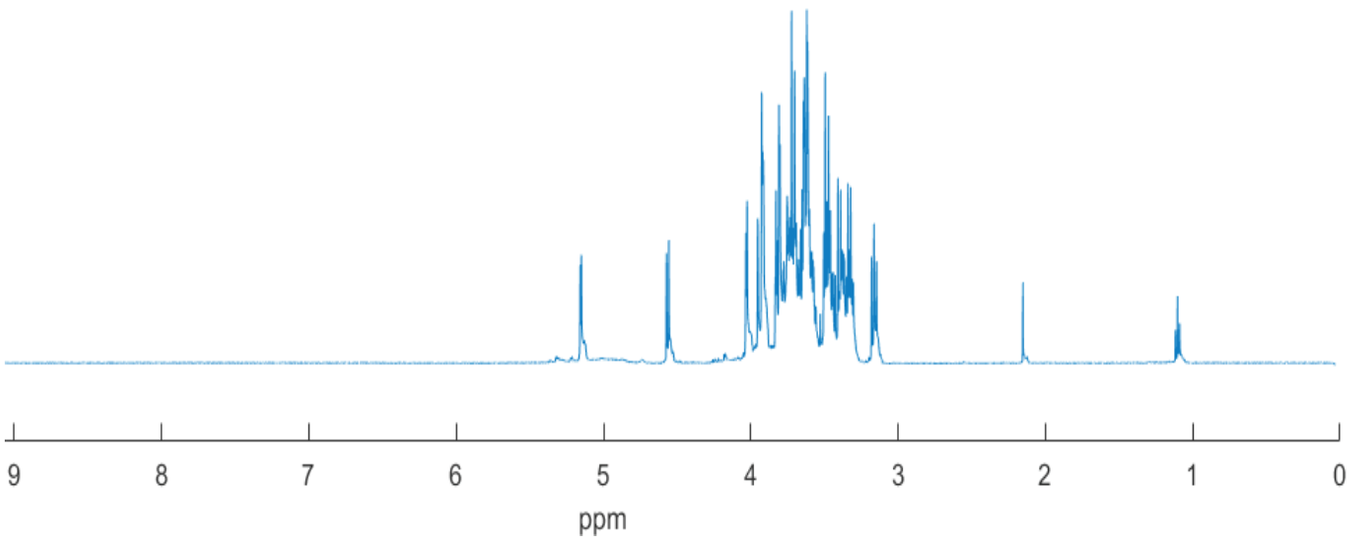

Fig. 1. A complete NMR spectrum of a longan honey sample.

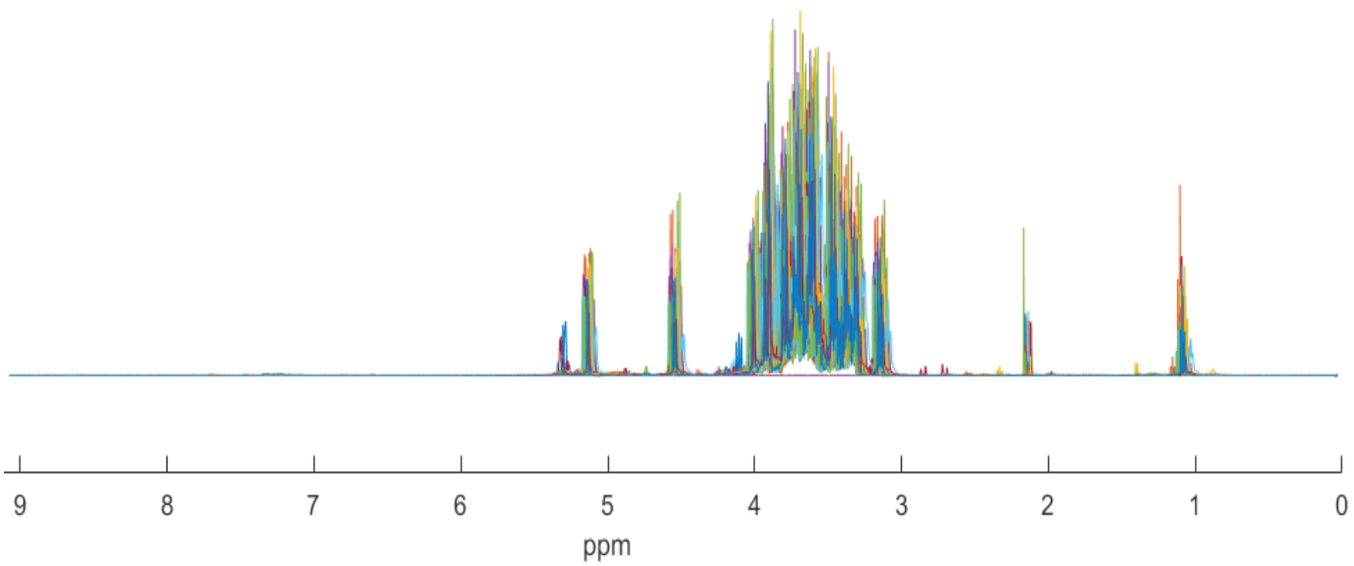

Fig. 2. The NMR combined spectra of 57 honey samples.

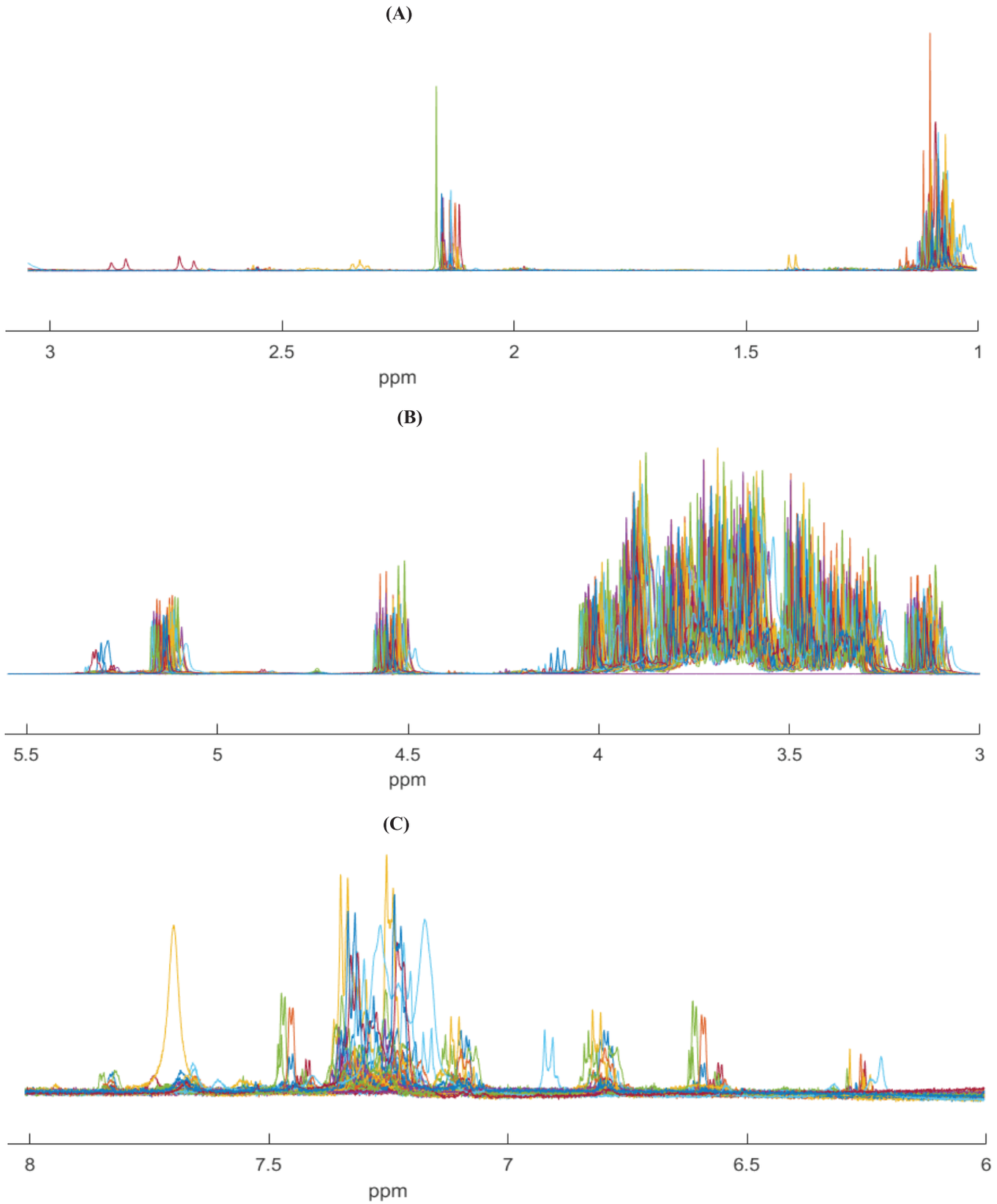

Fig. 3. ^{1}H-NMR spectra of honey samples in different regions corresponding to each group. (A) Acetic and lactic region; **(B)** Carbohydrate region; **(C)** Aromatic region.

It can be seen in the above figures, each honey sample gives us the differences in the concentrations of carbohydrate compounds, amino acids, lactic acids, and acetic acids. From this point of view, it can be proposed that the PCA method can be employed to discriminate longan flower honey and non-longan flower honey.

Chemometric application

Spectra transforming:

Figure. 2 shows the complexity and considerable deviation of each spectrum compared to others; it leads to the use of normalisation and mean-centering to standardise the data as well as subtract the variable averages.

Figure. 4A and 4B are the PCA score plots of the 1-3 ppm and 6-8 ppm regions, respectively. In these plots, it was impossible to distinguish the longan honey samples from others, due to the group-less distribution of them [18]. These results indicated that the concentrations of amino acids could not be used to discriminate honey samples with different origins because of their low quantities. Also, lactic and acetic acids are not able to distinguish the origins of honey because the amounts of these compounds vary due to the unprofessional collection, extracting and preserving techniques of farmers. However, lactic and acetic acids may consist of the information and the preservation time and conditions [19].

Data pretreatment method:

Figure. 5A represents the PCA score plot of the 3-5.5 ppm region. It can be seen that PC1 describes 26.23%, while PC2 describes 19.09% of the total variability; the samples are grouped into three clearly distinct clusters, and one of the figures has all 18 longan honey samples. It is highly possible that honey samples were botanically classified by their difference in carbohydrates ratio so that the glucose and fructose concentration can lead to a longan origin of honey. Compared to the PCA score plot without using data treatment by mean-centering (Fig. 4B), the increasing of the eigenvalue is almost two-fold.

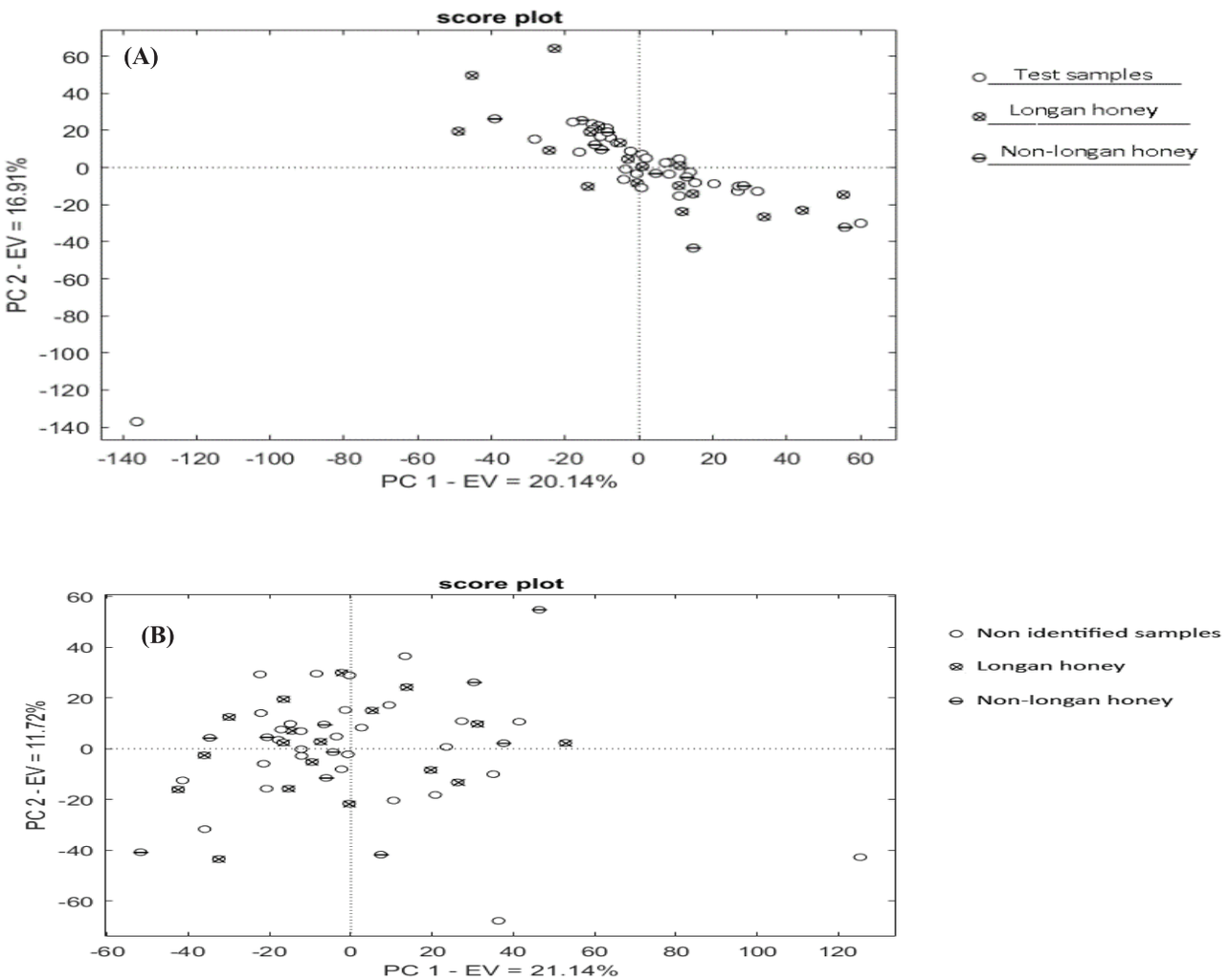

Fig. 4. PCA score plots of the ^{1}H-NMR spectra range: (A) 1-3 ppm and **(B)** 6-8 ppm regions.

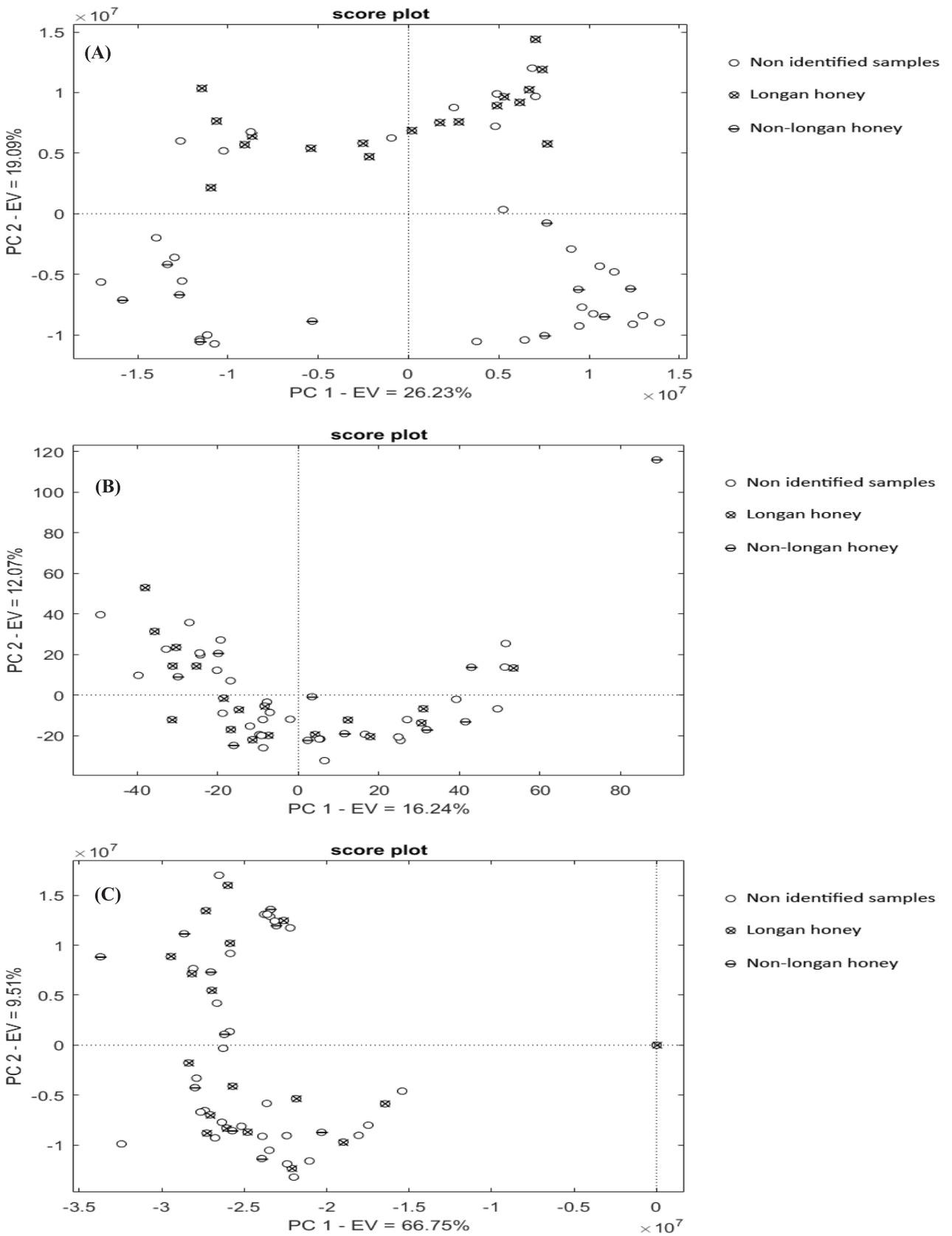

Fig. 5. PCA score plots of 3-5.5 ppm regions, with data pretreatment by: (A) mean-centering; **(B)** not treating and **(C)** range scaling.

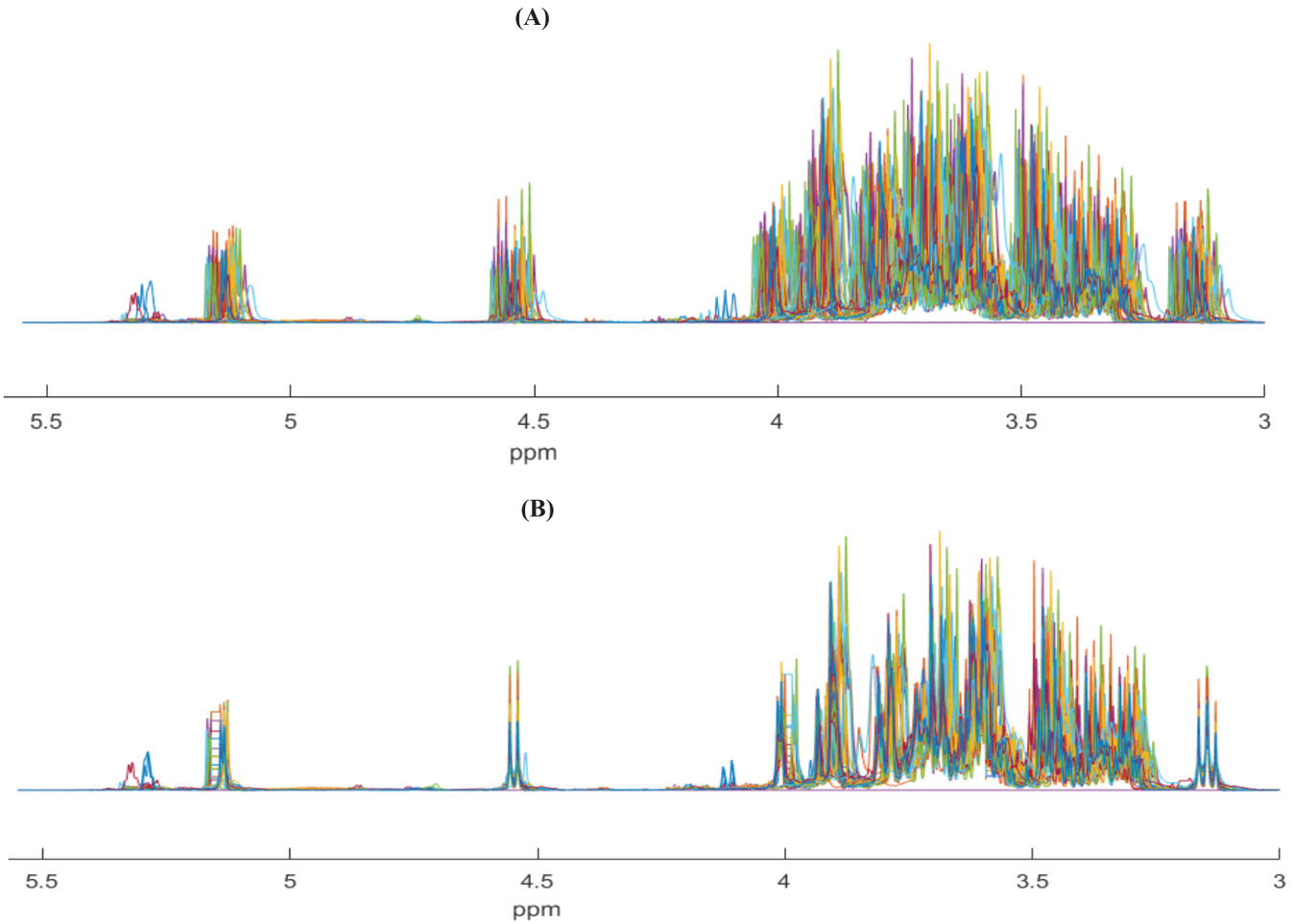

Fig. 6. ^{1}H-NMR spectra of 57 honey samples obtained with: (A) before using Icoshift algorithm; **(B)** after using Icoshift algorithm.

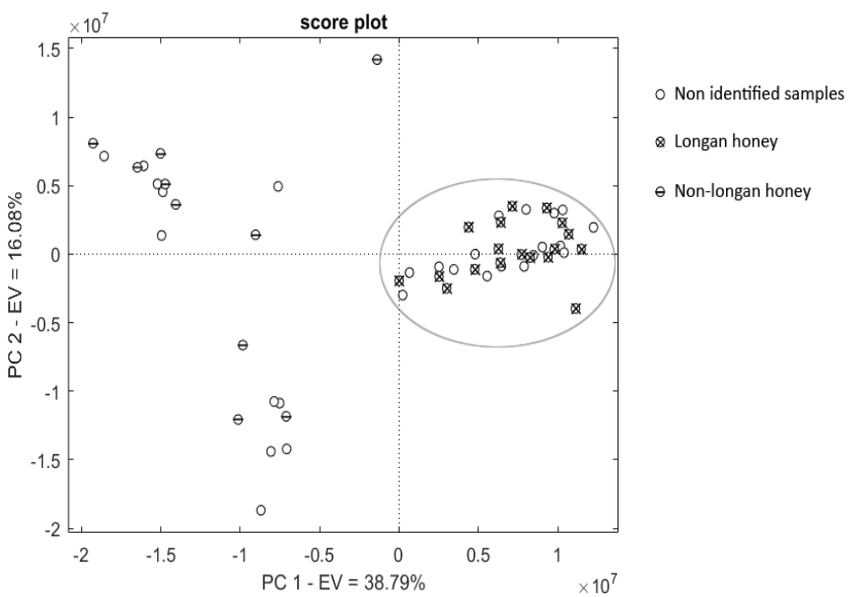

Fig. 7. PCA score plot of sample's spectra in the range of 3-5.5 ppm after icoshift.

Icoshift application:

The results obtained by using Icoshift toolbox to align the spectral data sets of all 57 honey samples were displayed in Fig. 6B. The results obtained by using PCA showed a better score plot with higher eigenvalue and clearer grouped samples.

The eigenvalues of the two first PCs were estimated at about 54%, which is an acceptable number. In the circled group, the predicted samples included 18 longan honey and 16 non-identified honey samples being present. Therefore, it was reasonable that this group contained longan honey.

It can also be recognised that all the longan honey samples belonged in the circle, whereas non-longan honey samples belonged on the outside (Fig. 7). It suggests

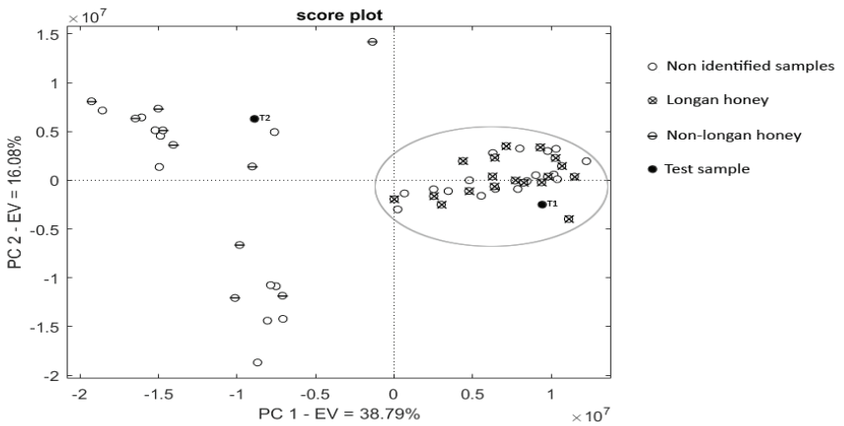

Fig. 8. PCA score plot of initial 57 honey samples and 2 test samples (T1 and T2).

that unknown samples' sources can be clearly identified based on their locations on the score plot.

Identification of unknown samples

For two unknown samples (T1 - a longan honey and T2 - coffee honey), the botanical sources can be classified as follows:

- Collect the data of ^{1}H-NMR spectrum of a honey sample in the range of 3-5.5 ppm.

- Extract the data into a spectrum of the data matrix and add the spectra (intensity vs. ppm) to the original data together with the 57 studied samples.

- Run the data pretreatment and PCA in the Matlab software.

The score plot obtained in Fig. 8 suggests that sample T1 is longan honey whereas T2 is not.

Conclusions

The ^{1}H-NMR spectra of honey samples in water solvent has been successfully applied for the classification of botanical origin of honey (longan flower honey or non-longan one). The ^{1}H-NMR data was pretreated by using the mean-centering algorithm. The PCA application was followed by icoshift algorithm which suggests good results in the classification of original longan honey based on the reference data of 57 honey samples in the range of carbohydrate 3-5.5 ppm. The longan honey (test sample) was grouped in its cluster showing suitable results to identify if an unknown sample is longan honey or not. The application of the data of 57 honey samples and PCA showed the appropriate results in the recognition of two test samples belonged to longan honey or non-longan honey. It can be seen that ^{1}H-NMR spectroscopy coupled with multivariate methods followed icoshift algorithm is a useful method of classifying the botanical of honey in the Vietnamese market. Therefore, it will be necessary to look after more reliable samples to develop a complete, quick and simple method for commercial application.

REFERENCES

[1] Codex Alimentarius Commission (2002), "Codex standard 12 Revised Codex Standard for Honey", *Stan. Methods,* **11,** pp.1-8.

[2] S. Bogdanov (1997), "Nature and origin of the antibacterial substances in honey", *Lebensm.-Wiss. u.- Technol,* **30(7),** pp.748-753.

[3] E. Perez, A.J. Rodriguez - Malaver, P. Vit (2006), "Antioxidant capacity of Venezuelan honey in wistar rat homogenates", *J. Med. Food,* **9(4),** pp.510-516.

[4] Cục Xúc tiến thương mại (2014), "Vietnam's honey export-opportunities and potentials", *VIETRADE.*

[5] Southern Horticultural Research Institute-SOFRI (2014), *Proceeding of the Workshop on Increasing Production and Market Access for Tropical Fruit in Southeast Asia,* pp.13-17.

[6] R. Karoui, J. De Baerdemaeker (2007), "A review of the analytical methods coupled with chemometrics tools for the determination of the quality and identity of dairy products", *Food Chem.,* **102(3),** pp.621-640.

[7] J.F.D. Kelly, G. Downey, V. Fouratier (2004), "Initial study of honey adulteration by sugar solutions using mid-infrared (MID) spectroscopy and chemometrics", *Journal of Agricultural and Food Chemistry,* **52(1),** pp.33-39.

[8] T.D. Alam, M.K. Alam (2004), "Chemometric analysis of NMR spectroscopy data: A review", *Annu. Rep. NMR Spectrosc.,* **54(1),** pp.41-80.

[9] R.A. Prestes, L.A. Colnago, L.A. Forato, L. Vizzotto, E.H. Novotny, E. Carrilho (2007), "A rapid and automated low resolution NMR method to analyze oil quality in intact oilseeds", *Analytica Chimica Acta,* **596(2),** pp.325-329.

[10] S. Bogdanov & M. Gfeller (2006), "Classification of honeydew and blossom honeys by discriminant analysis", *ALP Science,* **500,** pp.1-9.

[11] J. Riedl, S. Esslinger, C. Fauhl - Hassek (2015), "Review of validation and reporting of non-targeted fingerprinting approaches for food authentication", *Analytica Chimica Acta,* **885,** pp.17-32.

[12] R. Consonni, and L. Xagliani (2008), "Geographical Characterization of Polyfloral and Acacia Honeys by Nuclear Magnetic Resonance and Chemometrics", *Agricultural and Food Chemistry,* **56(16),** pp.6873-6880.

[13] F. Elisangela - Boffo, A. Leila - Tavares, C.T. Antonio - Tobias, M.C. Márcia - Ferreira, G. Antonio - Ferreira (2012), "Identification of components of Brazilian honey by ^{1}H-NMR and classification of its botanical origin by chemometric methods", *LWT - Food Science and Technology,* **49(1),** pp.55-63.

[14] M. Spiteri, K.M. Roger, E. Jamin, F. Thomas, S. Guyader, M. Lees, D.N. Rutledge (2017), "Combination of ^{1}H-NMR and chemometrics to discriminate manuk honey from other floral honey types from Oceania", *Food Chemistry,* **217,** pp.766-772.

[15] D. Bertelli, M. Lolli, G. Papotti, L. Bortolotti, J. Serra (2010), "Detection of Honey Adulteration by Sugar Syrups Using One-Dimensional and Two-Dimensional High - Resolution Nuclear Magnetic Resonance", *Agricultural and Food Chemistry,* **58(15),** pp.8495-8501.

[16] F. Savorani, G. Tomasi, S.B. Engelsen (2010), "icoshift: A versatile tool for the rapid alignment of 1D NMR spectra", *Journal of Magnetic Resonance,* **202(2),** pp.190-202.

[17] M. Spiteri, E. Dubin, J. Cotton, M. Poirel, B. Corman, E. Eric Jamin, M.L. Rutledge (2016), "Data fusion between high resolution ^{1}H-NMR and mass spectrometry: a synergetic approach to honey botanical origin characterization", *Analytical Bioanalytical Chemistry,* **408(16),** pp.4389-401.

[18] J. Devillers, M. Morlot, M.H. Pham-Delegue, J.C. Dore (2004), "Classification of monofloral honeys based on their quality control data", *Food Chemistry,* **86(2),** pp.305-312.

[19] R. Consonni, L.R. Cagliani, C. Cogliati (2012), "NMR characterization of saccharides in Italian honeys of different floral sources", *Agricultural and Food Chemistry,* **60(18),** pp.4526-34.