Differentiation of Gelsemium elegans-containing toxic honeys and non-toxic honeys by near infrared spectroscopy combine with aquaphotomics

Yuancui Su
Guangxi Normal University
Chuanmei Yang
Guangxi Normal University
Chengsen Tan
Guangxi Normal University
Yuanpeng Li (yuanpengli@gxnu.edu.cn)
Guangxi Normal University
Shan Tu
Guangxi Normal University
Siqi Zhu
Opto-electronic Department of Jinan University, Jinan University
Yongmei Wang
Guangxi Normal University
Lihu Wang
Guangxi Normal University
Junhui Hu
Guangxi Normal University
Yuxiang Mo
Guangxi Normal University
Hongxia Zhao
GuangDong Institute of Applied Biological Resources
Furong Huang
Opto-electronic Department of Jinan University, Jinan University

Research Article

Keywords: near-infrared spectroscopy, partial least squares, NIR aquaphotomics, toxic honeys, identification

Posted Date: April 14th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1544965/v1

Abstract

This work proposes a fast and accurate method based on near-infrared (NIR) spectroscopy with partial least-squares discriminant analysis (PLS-DA) and aquaphotomics to identify toxic honeys. PLS-DA was used to construct an optimal model for distinguishing toxic honey from non-toxic honey. The models based on preprocessed NIR spectra have an accuracy of 92.73% and were more accurate than the model based on raw NIR spectra. Based on the aquaphotomics analysis of the first overtone of water (1300–1600 nm), we found that the 1398 nm, 1440 nm, and 1472 nm bands can be used as markers to distinguish toxic honeys. Compared to non-toxic honeys, *G. elegans*-containing toxic honeys have a significantly smaller number of water molecules with multiple hydrogen bonds, due to the hydrogen bonding of the C-O-C, C = O, and NH_2 groups of gelsemine and koumine. These groups replace hydrogen bonds between glucose/polysaccharide molecules and water.

1 Introduction

Honey has health benefits as well as medicinal value and it is one of the most widely consumed nutritional products in China (Yang et al. 2020). China is a vast country with a wide variety of nectar plants. Common nectar source plants are *platycodon*, *rapeseed*, *lychee*, *loquat*, *osmanthus*, *safflower*. Toxic nectar plants like *Gelsemium elegans*, *Tripterygium wilfordii*, *Veratrum nigrum L*, *Golden Wisteria*, *Macleaya cordata*, and *Stellera chamaejasme Linn* are also found in China. At the end of the flowering period of non-toxic nectar plants or in excessively wet or dry conditions, toxic plants can become the dominant source of nectar, which results in the production of toxic honey (Kumaravelu and Gopal. 2015; Sun et al. 2019).

G. elegans is a poisonous plant that is mainly found in southern China (provinces like Zhejiang, Guangxi, Guangdong and Fujian), which is one of the poisonous nectar source plants. Indole alkaloids are the main toxic constituents of *G. elegans*, distributed throughout the plant, which has strong toxicity (Shen et al. 2020). In recent years, several cases of honey poisoning and illness have occurred in southern China due to the consumption of *G. elegans*-containing honeys; deaths have occurred in serious cases (Xu et al. 2015). Therefore, studies about methods for the identification of toxic honey are of great significance for preventing honey poisoning and ensuring the stability of the honey industry. Currently, the methods for identifying toxic honey include sensory analysis (Marcazzan et al. 2018), pollen identification (Kraaijeveld et al. 2015), gas chromatography (Kowalczyk et al. 2018), thin-layer chromatography(Islam et al. 2020), and liquid chromatography-tandem mass spectrometry (Koike et al. 2020). Mistakes can occur in sensory analysis and pollen identification due to interpretation bias. Chromatography is complex, time-consuming, and expensive. Therefore, there is an urgent need to develop a fast, simple, and accurate analytical method for the identification of toxic honey.

Near-infrared (NIR) spectroscopy is a fast, simple, and accurate analytical technique that is widely used in food analysis (Yakubu et al. 2020). In recent years, many studies around the world have demonstrated the feasibility of using NIR spectroscopy for honey quality inspections (Skaff et al. 2021). However, the

use of NIR to detect *G. elegans*-containing toxic honeys has yet to be reported. Honey consists of water and a diverse variety of carbohydrates. Due to its diverse components and water absorption, the NIR spectra of honey contains overlapping bands. Moreover, as the toxic components of *G. elegans* are present in low concentrations, their spectral absorptions are easily masked by those of other compounds. This makes it difficult to identify toxic honeys based on spectral differences alone. To address the aforementioned problems and extract useful information from NIR spectra, it is necessary to develop an algorithm that can extract the spectral features of toxic honey and use them to identify toxic honeys.

Aquaphotomics is a new scientific discipline introduced by Tsenkova (Tsenkova et al. 2018). In this novel approach to NIR analysis, water molecules are treated as the “matrix” of the system, and their absorptions are the primary source of information. Other substrates (solutes) are viewed as perturbing factors. By using the “extended water mirror approach” (EWMA), aquaphotomics allows changes in NIR water absorptions to reflect on changes in molecules in the aqueous system. Currently, aquaphotomics is widely used in food analysis (Muncan and Tsenkova. 2019; Yakubu et al. 2020) and disease diagnosis (Li et al. 2020; Tsenkova 2006). However, the use of NIR aquaphotomics to identify toxic honeys has not been reported.

The aim of this study was to provide a rapid and accurate method for identifying toxic honey. To this end, NIR spectroscopy was combined with partial least-squares discriminant analysis (PLS-DA) to develop a model for the discrimination of toxic and non-toxic honeys. NIR aquaphotomics was then used to elucidate changes in the NIR water absorptions of toxic and non-toxic honey and aquagrams were used to analyze the differences in their water spectral patterns (WASPs). A new approach for the rapid and accurate identification of toxic honeys was established.

2 Materials And Methods

2.1 Experimental apparatus

A FOSS XDS Rapid Content Analyzer with the rapid liquid analyzer accessory (FOSS, Denmark) was used for the acquisition of NIR spectra. The spectral acquisition range of the system was 400 nm – 2500 nm, and the detectors are Si (400 ~ 1100 nm) and PbS (1100 ~ 2500 nm). Gelsemine, gelsemicine, koumine, and gelsenicine concentrations were determined using an LC-20 AT HPLC system (Shimadzu, Japan), which was equipped with a UV detector, Welch Ultimate AQ-C18 column (250 mm × 4.6 mm, 5 μm), nitrogen blower, SK-1 vortex mixer, Oasis HLB solid phase extraction (SPE) cartridge (100 mg/3 mL), Cleanert PCX SPE cartridge (100 mg/3 mL), and C18 SPE cartridge (100 mg/3 mL).

2.2 Honey samples and spectral collection

The 218 honey samples used in this experiment were provided by the Guangdong Institute of Applied Biological Resources. The 95 toxic honey samples were collected from provinces in southern China including Guangdong, Guangxi, and Fujian, within the distribution range of *G. elegans*. The 123 non-toxic honey samples were pure natural honeys collected from the natural forests of Meizhou City, Zhaoqing

City, Yangjiang City, Puning City, Guangzhou City, Gaozhou City, Longmen County, Shaoguan City, Jieyang City, Conghua City, Heyuan City, Yunfu City in the Guangdong Province, and Beiliu City in the Guangxi Province. The concentrations of the four toxins of *G. elegans* in the toxic honeys is listed in Table 1.

Table 1
The concentrations of the four toxins of *G. elegans*

Content μg/kg	Gelsemine	Gelsemicine	Koumine	Gelsenicine
Min	0.04770	0.06690	0.06448	0.08560
Max	13.51000	59.78000	0.03580	10.70000
Avg	0.43630	1.37300	0.04859	0.26400
Bias	0.48690	1.56600	0.01711	0.20050

The glucose contained in honey crystallizes below 25 ℃, which leads to an uneven distribution of its constituents. This affects the consistency of the NIR spectra and thus the identification of toxic honey. Therefore, prior to NIR analysis, all honey samples were placed in a 50°C water bath for 1 h to dissolve all of the crystalline material or nuclei. The samples were then placed in a 30 ℃ water bath for 48 h in order to remove all air bubbles. Finally, the samples were loaded into sample tubes, labeled, and placed in a sample rack where they were stored at room temperature until used. During the NIR experiments, 2 mL of the honey sample was placed in a 2 mm cuvette. The absorbance spectrum of the sample was recorded from 400 nm to 2500 nm (in 2 nm intervals). Three measurements were taken and averaged for each honey sample. A total of 218 spectra were collected.

2.3 Aquaphotomics

Water is a natural biological matrix, as its molecules are bound to each other by strong hydrogen bonding interactions. Changes in the physical and chemical properties of a biological system produce changes in the absorption modes of their water molecules. Although all honeys primarily consist of water and sugars, there are significant differences in the chemical constitutions of non-toxic and toxic honeys. Therefore, the toxic components of *G. elegans* perturb the absorption modes of water in the honey. Hence, the water absorptions of honey could potentially be used to distinguish between non-toxic and toxic honeys.

Each water molecular structure has a distinct spectral range, known as the water matrix coordinates (WAMACS). There are 12 WAMACS that have been experimentally discovered in the first overtone of water (1300–1600 nm) and shown in Table 2 (Muncan and Tsenkova 2019). The combination of the activated water bands, at which the light absorbance is influenced by the perturbations, depicts a characteristic spectral pattern called Water Spectral Pattern (WASP), which reflects the condition of the whole water molecular system. It contains a huge amount of physical and chemical information for the

solution because the water hydrogen bonding network is easily influenced by any subtle perturbations including the solutes.

Table 2
Water absorption modes in the NIR region of toxic honeys and non-toxic honeys

WAMSCs	Range (nm)	Characteristic wavelengths (nm)	Assignment
C1	1336–1348		v_3
C2	1360–1366		OH stretch
C3	1370–1376		$v_1 + v_3$
C4	1380–1390	1386	OH stretch
C5	1398–1418	1398, 1412	S_0
C6	1420–1428	1422, 1428	Water hydration
C7	1434–1444	1440	S_1
C8	1448–1454	1452	$v_1 + v_3$
C9	1460–1468		S_2
C10	1472–1482	1472	S_3
C11	1482–1495		S_4
C12	1506–1516		Strongly bonded water or V_2
WAMACS - Water Matrix Coordinates.			
v_1 - symmetric stretching of first overtone of water.			
v_2 - bending of first overtone of water.			
v_3 - asymmetric stretching of first overtone of water.			
S_{0-4} - $(H_2O)_{1-5}$.			

This work proposes the use of NIR aquaphotomics to identify toxic honey. First, the characteristic bands that exhibit the largest difference between toxic and non-toxic honeys are selected based on difference spectra, derivative spectra, and principal component analysis (PCA) loading analysis. A radar chart (aquagram) is then plotted based on these features to perform a comprehensive aquaphotomics analysis

of non-toxic and toxic honeys. The procedures of this method are as follows: WASPs are represented as aquagrams, which are radar charts based on the normalized absorbance of several characteristic water bands. The value of the WASP on each axis is given by Eq. (1). First, the 1300–1600 nm band is preprocessed using the first-order derivative and multiple scattering correction (MSC) methods and the absorbances of the characteristic water bands are then extracted. Next, the absorbances of each sample are normalized and averaged and the absorbances of the selected bands are presented on their corresponding radial axes in the aquagram. The normalized absorbance of each band is given by the equation where $A_{q\lambda}$ is the normalized absorbance on the aquagram, A_λ is the absorbance obtained from first-order derivative and MSC preprocessing of the raw spectrum, μ_λ is the mean of all the spectra for the examined group of samples after transformation, σ_λ is the standard deviation of all spectra for the examined group after transformation, and λ is the wavelength of the selected band.

$$Aq_\lambda = \frac{A_\lambda - \mu_\lambda}{\sigma_\lambda} \quad (1)$$

2.4 Sample set partitioning and model evaluation

In this study, 218 honey samples were divided into training and validation sets in a 3:1 ratio using randomized smoothing. The training set consisted of 163 samples including 71 *G. elegans*-containing toxic honeys and 92 natural non-toxic honeys. The prediction set had 55 samples including 24 *G. elegans*-containing toxic honeys and 31 natural non-toxic honeys. The partitioning of the samples is shown in Table 3.

Table 3
Partitioning of honey samples

Category	Number of samples		Total
	Training set	Validation set	
Poisonous	71	24	95
Non-poisonous	92	31	123
Total samples	163	55	218

To maximize the performance of the classification model, it is necessary to evaluate its efficacy. Three indicators are used to evaluate prediction performance: accuracy, sensitivity, and specificity. High accuracy, specificity, and sensitivity values are indicative of a representative classification model(Liu et al. 2020). Accuracy is the percentage of samples out of the total number of samples that were correctly classified in the prediction set by the trained model. Sensitivity is the percentage of positive samples that were correctly classified and specificity is the percentage of negative samples that were correctly classified.

$$Accuracy = \frac{N_{correct}}{N_{total}} \quad (2)$$

$$Sensitivity = \frac{ET}{ET + MN} \quad (3)$$

$$Specificity = \frac{CT}{MF + CT} \quad (4)$$

$N_{correct}$ and N_{total} denote the number of correctly classified samples and the total number of samples in the prediction set, respectively. *ET*, *MN*, *MF*, and *CT* are the number of true positives, true negatives, false negatives, and false positives, respectively.

3 Results And Discussion

3.1 Spectral analysis

3.1.1 Analysis of raw spectra

The raw and averaged NIR spectra of the honey samples are shown in Figs. 1(a) and 1(b). In Fig. 1(a), the 1460 nm (O-H stretching vibration of first overtone) and 1938 nm (O-H stretching and bending vibration) bands are the two most significant absorption modes of water in the NIR region (Muncan et al. 2020). The 1690 nm mode is the characteristic peak of fructose in honey (the first overtone stretching vibration of C-H). The 1771 nm band is the characteristic peak of sugars and gelsemine in honey (second overtone of the C = O stretch). The 2098 nm band belongs to the O-H bending and C-O bond stretching vibrations of the C-O-C group in gelsemine. 2276 nm is the C-H stretching and bending mode. 2326 nm is the stretching and bending vibration of CH_2 (Zhang et al. 2013) and 2440 nm is a combination band between C-H and CH_2 (Grassi et al. 2018).

As shown in Fig. 1(b), the NIR spectra of natural non-toxic honey and toxic honey are identical, except for slight differences at 1460 nm and 2018 nm. These differences may be related to the second overtone of the C = O stretch in gelsemine or perturbations to the first overtone of water (1460 nm) due to the presence of gelsemine and koumine. Therefore, toxic and non-toxic honeys cannot be differentiated by inspecting the NIR spectra with the naked eye. To construct a classification model that distinguishes the NIR spectra of toxic and non-toxic honeys, spectral processing was used in tandem with machine learning to minimize the effects of noise and increase the information that is extracted from NIR spectra.

3.1.2 Spectral preprocessing

NIR signals carry a substantial amount of information about the chemical constitution of a sample. However, they also contain interfering factors such as noise, stray light, and the instrument response. These factors affect the stability of the spectra and cause spectral baseline drift, thus reducing the accuracy of the classification model. To obtain an accurate classification model, four preprocessing methods were applied to the raw NIR data: MSC, standard normal variate (SNV), first derivative, and second derivative processing.

Figure 1(c), 1(d), 1(e), and 1(f) show the preprocessed NIR spectra. Averaging the preprocessed spectra enabled the identification of spectral differences between toxic and non-toxic honeys. Although the trends in Figs. 1(c) and 1(d) are similar to those of the raw spectra, the MSC- and SNV- preprocessed spectra show that toxic and non-toxic honeys have different absorption intensities at 1460 nm, 1938 nm, 2018 nm, 2098 nm, and 2440 nm. The bands at approximately 2000 nm are O-H stretching vibrations and the 2098 nm band in particular is a combination band consisting of O-H bending and C-O stretching vibrations (Sakudo 2016).

In Fig. 1(e), the averaged first-derivative NIR spectra of toxic and non-toxic honeys differ at 1418 nm, 1904 nm, and 2050 nm. The 1389–1418 nm range corresponds to the absorption of free water molecules and the combination absorption band of water is located at approximately 1900 nm – 1950 nm. The differences between toxic and non-toxic honeys in the averaged second-derivative NIR spectra (Fig. 1(f)) occur at 1884 nm (the stretching-bending combination band of water), 1920 nm (the second-overtone stretching vibration of O-H from water and sugar and C = O from gelsemine), and 2090 nm (hydrogen bonding of O-H groups in the carbohydrate region) (Cakmak-Arslan et al. 2020; Yang et al. 2020). The second-derivative spectra of toxic honey contains an additional shoulder peak at 1944 nm (O-H overtone vibrations) that is absent in non-toxic honey. This peak may be attributed to gelsemine and koumine interfering with water-water and water-sugar hydrogen bonding interactions, which enhances the intensity of O-H overtone vibrations.

3.2 Toxic honey identification using NIR with PLS-DA

This work is the first report combining NIR spectroscopy with PLS-DA to construct a model in order to identify toxic honeys. Before constructing the PLS-DA model, the training and prediction sets were labeled and the non-toxic and toxic honey samples were assigned values of 1 and − 1, respectively. PLS-DA was then used to determine the accuracies of the principal components, select the optimal PLS factor, and construct the optimal model for toxic honey identification. Finally, the prediction set was used to evaluate the constructed model.

The accuracy, sensitivity, and specificity of the model based on raw NIR spectra are 90.9%, 100% and 83.87% (Table 4), respectively. The models based on MSC, SNV, first-derivative or second-derivative preprocessed NIR spectra have an accuracy, sensitivity, and specificity of 92.73%, 87.09%, and 100%, respectively. The optimal PLS factors were 8, 9, and 10. Although the classification model always has the same accuracy regardless of the type of preprocessing, the optimal PLS factor of the first-derivative

model is 8, which indicates that it has the lowest model complexity between all models. Hence, the identification model of toxic honey established after the first-derivative pretreatment has the best effect.

Table 4
Efficacies of PLS-DA models with different preprocessing methods

Pretreatment method	Optimal component score	Sensitivity	Specificity	Accuracy (%)
Untreated	8	100.00%	83.87%	90.90%
First Derivative	**8**	**100.00%**	**87.09%**	**92.73%**
Second Derivative	9	100.00%	87.09%	92.73%
SNV	10	100.00%	87.09%	92.73%
MSC	10	100.00%	87.09%	92.73%

Figure 2 compares toxic and non-toxic honey samples in the first three principal components. In Fig. 2(a), the toxic honey distribution profile is clustered and overlaps slightly with the distribution profile of non-toxic honey. Hence, the first and second principal components could be used to distinguish toxic honeys from non-toxic honeys. In Figs. 2(b) and (c), there are many overlaps between toxic and non-toxic honeys, which makes it more difficult to distinguish these honeys. In summary, it was found that the first three principal components cannot be used to accurately distinguish toxic honeys because they only extracted a small amount of feature information. However, after the principal components were optimized, an accuracy of 92.37% was achieved with a model using the first eight principal components.

3.3 Identification of toxic honeys using NIR aquaphotomics

3.3.1 Analysis of difference spectra and derivative spectra

The averaged spectra and difference in spectra from 1300 nm – 1600 nm of the honey samples are shown in Fig. 3(a) and (b). Non-toxic and toxic honeys have nearly identical spectra, but the 1460 nm absorption of toxic honey is slightly stronger than that of non-toxic honey. This shows that the water content differs slightly between toxic and non-toxic honeys. The difference spectrum shows that toxic and non-toxic honeys differ at 1380 nm and 1428 nm. The 1380 nm band originates from the solvent layer of water molecules while the 1428 nm band originates from crystallized waters (Bázár et al. 2016; Beć et al. 2020). The error bars represent the absorbance range of honeys from different batches and collection sites.

Figure 3(c) shows the second derivative and difference spectra. Second derivative processing effectively increased spectral resolution, which made it easier to identify differences between the NIR spectra of toxic and non-toxic honeys. In this figure, the second derivative transformation enhanced the peaks of the raw averaged spectrum. The largest positive peak is located at 1398 nm while the largest negative peak is located at 1440 nm. The 1398 nm band corresponds to free water molecules while the 1440 nm band comes from water species with one hydrogen bond. Furthermore, the difference spectra analysis shows

that toxic and non-toxic honeys mainly differ at 1386 nm (O-H stretch of water), 1422 nm (characteristic peak of crystallized waters), 1452 nm (symmetric stretching and doubly degenerate bending modes of water O-H), and 1472 nm (water species with three hydrogen bonds) (Breiman et al. 2017). The error bars indicate the absorbance range of honeys from different batches and collection sites.

3.3.2 Principal component analysis

NIR spectra are often highly correlated which can cause data redundancy. For this reason, PCA was applied to the 1300 nm – 1600 nm NIR spectra of all of the samples in this experiment to simplify the NIR data. The results of PCA on the NIR spectra of the honey samples are shown in Fig. 4. The 3D PCA score plot (Fig. 4(a)) shows the 3D projections of the first three principal components, whose cumulative explained variance is 93.23%. Therefore, the first three principal components reflect on the majority of the basic features of the raw data.

As one follows the PC3 axis from its negative end to the positive end, the honeys change from toxic to non-toxic (see Fig. 4(a)). However, there is a degree of overlap between the toxic and non-toxic honeys, which indicates that PCA alone cannot accurately distinguish toxic and non-toxic honeys. Figure 5(b) shows the loadings of the first three principal components. PC1 has the highest positive correlation at 1428 nm, while PC2 has almost no contribution. PC3 shows a degree of negative correlation at 1380 nm, 1412 nm, and 1434 nm. The peak at 1412 nm belongs to water molecules without hydrogen bonding while the 1434 nm band is the absorption peak of water species with one hydrogen bond.

3.3.3 Results of Aquaphotomics

We experimentally demonstrated that there are 12 WAMACS in the first overtone of water (1300 nm – 1600 nm). In this study, the 10 following characteristic wavelengths were selected based on difference spectra analysis, PCA, and derivative spectra analysis: 1386 nm, 1398 nm, 1412 nm, 1422 nm, 1428 nm, 1434 nm, 1440 nm, 1452 nm, 1470 nm, and 1522 nm. Based on these wavelengths and Eq. (1–8), aquagrams were plotted to analyze the differences between toxic and non-toxic honeys in terms of their WASPs.

The aquagrams of glucose solution and pure water are shown in Fig. 5(a), which shows that the most intense absorption modes of pure water are 1386 nm (solvent layer of water) and 1398 nm (free water molecules). This indicates that pure water contains a large number of free waters. As the concentration of glucose increased in the glucose solution, the 1472 nm and 1522 nm bands gradually intensified while the 1386 nm and 1398 nm bands gradually weakened. Therefore, increasing glucose concentration increases the number of water molecules with three hydrogen bonds and reduces the number of free waters. This is likely to be caused by glucose-water hydrogen bonds replacing water-water hydrogen bonds (as shown in Fig. 6), as a glucose molecule has six OH groups that can interact with water molecules via hydrogen bonds. Hence, increasing glucose concentration increases the number of water molecules with three hydrogen bonds and decreases the number of free water molecules.

The aquagrams of toxic and non-toxic honey are shown in Fig. 5(b) and there are significant differences between these aquagrams. *G. elegans*-containing toxic honeys have significantly higher absorptions than non-toxic honeys at five WAMACS: 1386 nm (solvent layer of water), 1398 nm and 1412 nm (free waters), 1422 nm (water molecules with no hydrogen bonds), and 1428 nm (C-H and O-H stretching vibration groups). Therefore, toxic honeys contain a larger number of free water molecules than non-toxic honeys. Non-toxic honey samples show significantly stronger absorbances than toxic honey samples at the 1440 nm (waters with one hydrogen bond), 1452 nm (bending and fundamental asymmetric stretching mode of water), 1472 nm (waters with three hydrogen bonds), and 1522 nm (unassigned) WAMACS.

The error bars represent the absorbance range of honeys from different batches and collection sites. Although the error bars in the aquagram indicate a significant degree of overlap between toxic and non-toxic honeys, there is no overlap at the 1386 nm (O-H stretch), 1440 nm (waters with one hydrogen bond), 1452 nm (bending and fundamental asymmetric stretching mode of water), 1472 nm (waters with three hydrogen bonds), and 1522 nm (unassigned) bands. Hence, there are significant differences between non-toxic and toxic honeys at these five WAMACS. This also shows that the number of free waters and the number of waters with three hydrogen bonds are important markers for the differentiation of toxic and non-toxic honeys.

The main components of honey are glucose and fructose. Hence, honey solutions are similar to glucose solutions, as both have a large number of waters with multiple hydrogen bonds and few free waters. In Fig. 5(b), we show that non-toxic honeys and glucose solutions have similar aquagrams that are significantly different than the aquagrams of *G. elegans*-containing toxic honeys. This is because each organic molecule has a distinct set of polar groups which have different hydrogen bonding affinities. As compared to the hydroxyl groups of glucose, the C-O-C, C = O and NH_2 groups of gelsemine and koumine have much stronger hydrogen bond energies and thus form stronger bonds with water (Chen et al. 2011). Hence, toxic honey has fewer hydrogen bonded waters than non-toxic honey due to the presence of these toxic substances. The strong binding between water and the C-O-C, C = O, and NH_2 groups of gelsemine and koumine effectively replaces water-water and glucose-water hydrogen bonds in honey (as shown in Fig. 6), which reduces the number of hydrogen bonds between water and glucose/polysaccharides. The chemical groups of gelsemine and koumine only form a hydrogen bond with one water molecule. As compared to non-toxic honeys, toxic honeys have a higher number of free waters (1398 nm) and small water clusters (1440 nm) and a smaller number of water species with multiple hydrogen bonds (1472 nm).

4 Conclusion

This study verified the feasibility of using NIR spectroscopy with PLS-DA and aquaphotomics in order to distinguish toxic honeys containing gelsemine and koumine from non-toxic honeys. To construct the NIR-based PLS-DA classification model, raw NIR spectra were first subjected to MSC, SNV, first-derivative, and second-derivative processing. We found that first-derivative preprocessing was optimal and that the sensitivity, specificity and accuracy of the optimal model are 100%, 87.09%, and 92.73%, respectively.

Based on NIR aquaphotomics analysis, the differences between non-toxic honeys and toxic honeys containing gelsemine and koumine are most significant at the 1386 nm, 1470 nm, and 1522 nm bands. The aquagrams show that toxic honeys have stronger absorptions in the shorter wavelengths (1386 nm, 1398 nm, 1412 nm, 1422 nm, and 1428 nm), which indicate a smaller number of hydrogen-bonded waters and a larger number of free waters. Non-toxic honeys have stronger absorptions in the longer wavelengths (1440 nm, 1452 nm, 1472 nm, and 1522 nm), which indicates that non-toxic honeys have a larger number of structural waters. In summary, we have developed a fast, simple, and accurate method for distinguishing toxic honeys from non-toxic honeys by combining NIR spectroscopy with PLS-DA and aquaphotomics. This is important for ensuring the stability and health of the honey industry and protecting consumer lives.

Declarations

Author Contribution Yuanpeng li ,Yuancui Su and Chengsen Tan designed experiments and wrote the manuscript. Chuanmei Yang collected data on common honey and honey with gelsemium. Yuanpeng li, Hongxia Zhao, and Furong Huang supervised this project. All other authors were involved in the study concept, design, analysis, and interpretation of data, drafting, and revision of the manuscript.

Funding This work was supported by the National Natural Science Foundation of China (62005056\61975069), GuangXi Natural Science Foundation of China (2021JJB110003), Guangzhou Science and Technology Project (202103000095), GuangXi Natural Science Foundation of China (No. 2018GXNSFBA050028),the Key-Area Research and Development Program of Guangdong Province (2020B090922006), Science and Technology Commissioner Project of Guangdong Province (GDKTP2020023100), Project of Guangzhou Academician Workstation (h2020-3-01), and Natural Science Foundation of Guangxi Province (2020GXNSFBA159008).

Data Availability The datasets analyzed during the current study areavailable from the corresponding author on reasonable request.

Ethics Approval This article does not contain any studies with humanor animal subjects.

Conflict of interest The authors of the manuscript declare we have no conflict of interest.

References

1. Breiman L, Friedman JH, Olshen RA, Stone CJ (2017) Classification and Regression Trees, Routledge. https://doi.org/10.1201/9781315139470
2. Beć KB, Grabska J, Huck CW (2020) Near-infrared spectroscopy in bio-applications. Molecules. 25(12): 2948.https://doi.org/10.3390/molecules251229 48
3. Bázár G, Romvári R, Szabó A, Somogyi T, Éles V, Tsenkova R (2016) NIR detection of honey adulteration reveals differences in water spectral pattern. Food chemistry.194: 873–880.

https://doi.org/10.1016/j.foodchem.2015.08.092

4. Chen LZ, Xue XF, Ye ZH, Zhou JH, Chen F, Zhao J (2011) Determination of Chinese honey adulterated with high fructose corn syrup by near infrared spectroscopy. Food Chemistry.128(4): 1110–1114.https://doi.org/10.1016/j.foodchem.2010.10.027
5. Cakmak-Arslan G, Haksoy H, Goc-Rasgele P, Kekecoglu M (2020) Determination of the dose-dependent toxic effects of mad honey on mouse liver using ATR-FTIR spectroscopy. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy. 228: 117719. https://doi.org/10.1016/j.saa.2019.117719
6. Grassi S, Alamprese C (2018) Advances in NIR spectroscopy applied to process analytical technology in food industries. Current Opinion in Food Science. 22: 17–21. https://do i.org/10.1016/j.cofs.2017.1 2.008
7. Islam MK, Sostaric T, Lim LY, Hammer K, Locher C (2020) Sugar profiling of honeys for authentication and detection of adulterants using high-performance thin layer chromatography. Molecules. 25(22): 5289. https://doi.org/10.3390/molecul es252252 89
8. Kumaravelu C, Gopal A (2015) Detection and quantification of adulteration in honey through near infrared spectroscopy. Int. J. International Journal of Food Properties. 18(9): 1930–1935. https://doi.org/10.108 0/10942912.2014.919320
9. Kraaijeveld K, De Weger LA, Ventayol García M, Buermans H, Frank J, Hiemstra PS, den Dunnen JT (2015) Efficient and sensitive identification and quantification of air borne pollen using next-generation DNA sequencing. Molecular ecology resources. 15(1): 8–16. https://doi.org/10.1111/1755-0998.12288
10. Kowalczyk E, Kwiatek K (2018) Pyrrolizidine alkaloids in honey: determination with liquid chromatography-mass spectrometry method. Journal of veterinary research. 62(2): 173–181. https://doi.org/10.2478/jvetres-2018-0027
11. Koike H, Kanda M, Hayashi H, Matsushima Y, Yoshikawa S, Ohba Y, Hayashi M, Nagano C, Sekimura M, Otsuka K, Kamiie J, Sasamoto T, Hashimoto T (2020) Development of an alternative approach for detecting botulinum neurotoxin type A in honey: Analysis of non-toxic peptides with a reference labelled protein via liquid chromatography-tandem mass spectrometry. Food Additives & Contaminants: Part A. 37(8): 1359–1373. https://doi.org/10.1080/19440049.2020.1766121
12. Li YP, Guo L, Li L, Yang CM, Guang PW, Huang FR, Chen ZQ, Wang LH, Hu JH (2020) Early diagnosis of Type 2 diabetes based on near-infrared spectroscopy combined with machine learning and aquaphotomics. Frontiers In Chemistry. 8: 580489. https://doi.org/10.3389/fchem.2020.580489
13. Liu Y, Li Y, Peng Y, Yang Y, Wang Q (2020) Detection of fraud in high-quality rice by near-infrared spectroscopy. *Journal of food science*. *85*(9): 2773–2782. https: //doi.o rg /10.1111/1750-3841.15314
14. Marcazzan GL, Mucignat-Caretta C, Marina Marchese C, Piana ML (2018) A review of methods for honey sensory analysis. J. Journal of Apicultural Research. 57(1): 75–87. https://doi.org/10.1080/0021883 9.2017.1357940

15. Muncan J, Tsenkova R (2019) Aquaphotomics—From innovative knowledge to integrative platform in science and technology. Molecules. 24(15): 2742. https:// doi.org/10.3390/molecules24152742
16. Muncan J, Tei K, Tsenkova R (2020) Real-time monitoring of yogurt fermentation process by aquaphotomics near-infrared spectroscopy. Sensors. 21(1): 177. https://doi.org/1 0.3390/s21010177
17. Sun MH, Zhao LG, Wang K, Han LD, Shan JH, Wu LM, Xue XF (2019) Rapid identification of “mad honey” from Tripterygium wilfordii Hook. f. and Macleaya cordata (Willd) R. Br using UHPLC/Q-TOF-MS. Food Chemistry. 294: 67–72. https://doi.org/10.1016/j.foodchem.2019.05.028
18. Shen XW, Ma JS, Wang XQ, Wen CC, Zhang ML (2020) Toxicokinetics of 11 Gelsemium Alkaloids in Rats by UPLC-MS/MS. BioMed Research International. 10: 8247270. https://doi.org/10.1155/2020/82472 70
19. Skaff W, El Hajj R, Hanna-Wakim L, Estephan N (2021) Detection of adulteration in honey by infrared spectroscopy and chemometrics: Effect on human health. Journal of Food Processing and Preservation. e15438. https://doi.org/10.11 11/jfpp.15438
20. Sakudo A (2016) Near-infrared spectroscopy for medical applications: Current status and future perspectives. Clinica Chimica Acta. 455: 181–188. https://doi.org/10.1016/ j.cca.2016.02.009
21. Tsenkova R, Munćan J, Pollner B, Kovacs (2018) Essentials of aquaphotomics and its chemometrics approaches. Frontiers in Chemistry. 6: 363. https://doi.org/10.3389/ fchem.2018.00363
22. Tsenkova R (2006) AquaPhotomics: water absorbance pattern as a biological marker. NIR news. *17*(7): 13–23. https://doi.org/10.1255/nirn.926
23. Xu YK, Yang L, Liao SG, Cao P, Wu B, Hu HB, Guo J, Zhang P (2015) Koumine, humantenine, and yohimbane alkaloids from Gelsemium elegans. Journal of Natural Products. 78(7): 1511–1517. https://doi.org/10.1021/np5009619
24. Yang XH, Guang PW, Xu GZ, Zhu SQ, Chen ZQ, Huang FR.(2020) Manuka honey adulteration detection based on near-infrared spectroscopy combined with aquaphotomics. LWT - Food Science and Technology. 132: 109837. https://doi. org/10.1016/j.lwt.2020.109837
25. Yakubu HG, Kovacs Z, Toth T, Bazar G (2020) The recent advances of near-infrared spectroscopy in dairy production—A review. Critical Reviews in Food Science and Nutrition. 12: 1–22. https://doi.org/10.10 80/10408398.2020.1829540
26. Yang SP, Liu Y, Sun FF, Zhang JZ, Jin Y, Li Y, Zhou JH, Yi L, Zhu K (2020) Gelsedine-type alkaloids: Discovery of natural neurotoxins presented in toxic honey. Journal of Hazardous Materials. 381: 120999. https://doi.org/10.1016/j.jhazmat.201 9.120999
27. Zhang N, Li WZ, Chen C, Zuo JG (2013) Evaluation of the application of hydrogen bonding criteria to DMSO aqueous solution. Acta Physico-Chimica Sinica. 29(9):1891–1899. https://doi.org/10.386 6/PKU.WHXB201307121

Figures

Figure 1

NIR spectra of honey samples and Averaged preprocessed NIR spectra of honey: **(a)** raw spectra; **(b)** averaged spectra; **(c)** averaged MSC spectra; **(d)** averaged SNV spectra; **(e)** averaged first-derivative spectra; **(f)** averaged second-derivative spectra.

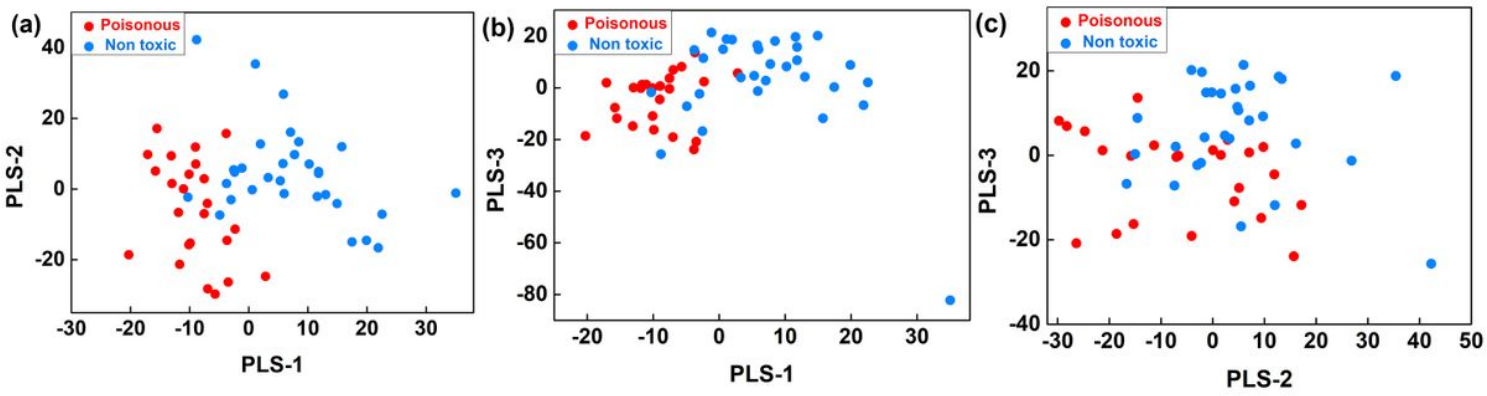

Figure 2

PLS-DA score distribution of the first three principal components extraction of sample.

Figure 3

1300 nm – 1600 nm spectra of toxic and non-toxic honeys: **(a)** averaged raw spectra; **(b)** averaged difference spectra; **(c)** Second-derivative spectra of honey (1300–1600 nm)

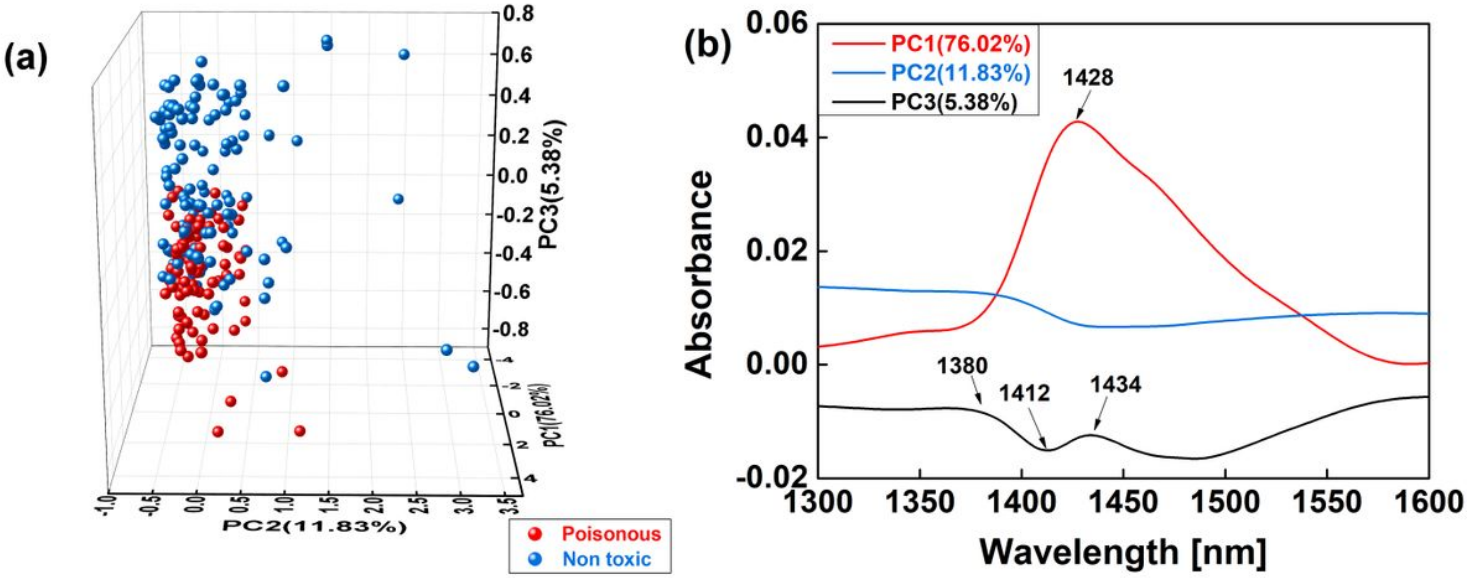

Figure 4

Result of PCA on the NIR spectra of honey samples: **(a)** PCA 3D score plot and **(b)** PCA loading plot

Figure 5

Aquagrams with 10 WAMACS for: **(a)** glucose solutions and pure water and **(b)** toxic honey and non-toxic honey

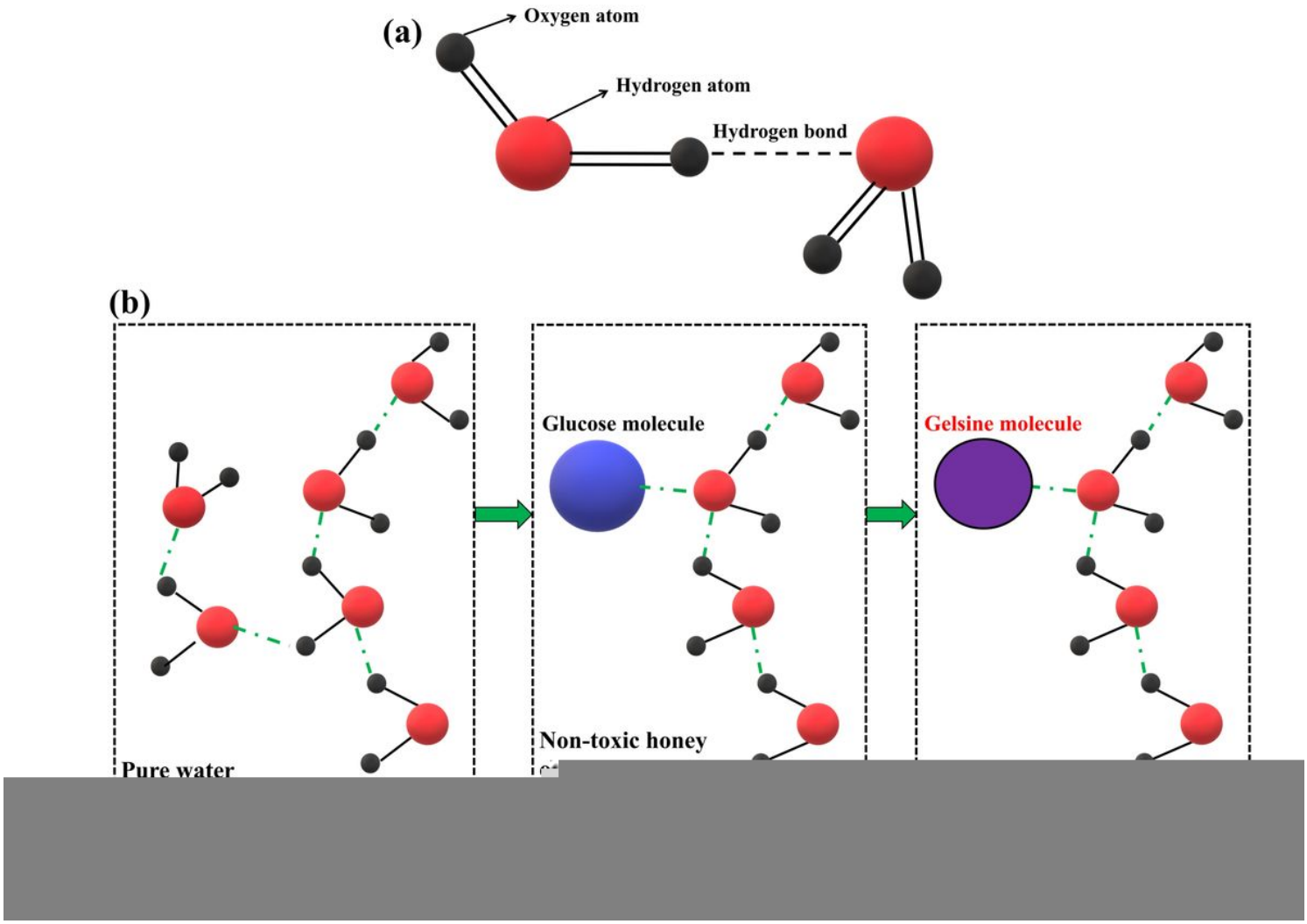

Figure 6

Hydrogen bonding interactions: **(a)** schematic diagram of hydrogen bonding between water molecules; **(b)** replacement of glucose by gelsemine in toxic honey.