Classification of Hoi-An and Sin-Chew Agarwood by Components Analysis of VOCs Released in Heat-Treated Agarwood using TD-GCMS and Chemometric Methods

Dongyu Jia and Songlin Yi*

Agarwood can be divided into resinous heartwood from the Hoi-An zone and Sin-Chew zone. Traditionally, an experienced human grader classifies agarwood by odor. However, sensory errors can follow from subjectivity, poor reproducibility, and time consumption during manual work. In this study, agarwood samples were heat-treated to release volatile organic compounds (VOCs), which were analyzed using the thermal desorption - gas chromatograph mass spectrograph (TD-GCMS) method and chemometrics analysis. The classification of agarwood was then identified. Sesquiterpenes and other aromatic compounds were the main compounds of heat-treated VOCs. Twenty-six characteristic compounds were screened *via* stepwise regression. Fisher discriminant analysis and Bayes discriminant analysis were conducted, based on the 26 compounds, to classify the agarwood samples. Discriminant functions of the two analysis methods were obtained.The results showed that it is feasible to use the TD-GCMS method combined with chemometrics analysis to analyze VOCs from heat-treated agarwood instead of experienced graders to classify the agarwood samples as being from either the Hoi-An zone and Sin-Chew zone. This study also provides a way to classify unknown samples by odor through 26 characteristic compound'srelative peak area and the discriminant equations, offering the possibility of testing an unknown sample's cultivation region.

Keywords: Agarwood; TD-GCMS; Heat-treated; Chemometrics methods

Contact information: College of Materials Science and Technology, Beijing Forestry University, Beijing, 100083, PR China;*Corresponding author: ysonglin@126.com

INTRODUCTION

Agarwood, the resinous heartwood of the *Aquilaria* species (Thymelaeaceae) (Naef 2011; Gao *et al.* 2014; Yang *et al.* 2016; Liu *et al.* 2017), is otherwise known as Chen-Xiang, eaglewood, gaharu, jinko, aloeswood, pokok karas, kalamabak, or oud in different cultures (Gao *et al.* 2014; Yang *et al.* 2016). The resin is widely used and well known for its fragrance and applications in aromatherapy, incense, perfume, religions, medicine, *etc*. (Ito 2008; Lancaster and Espinoza 2012; Yang *et al.*2016). The scent or aroma of agarwood is complex and pleasing, with few or no similar natural analogues (Liu *et al.* 2017).

Agarwood is widely distributed in South and Southeast Asia. The main agarwood-producing regions are China, India, Indonesia, Laos, Malaysia, Myanmar, Thailand, Vietnam, Bangladesh, Bhutan, Iran, the Philippines, and Singapore (Gao *et al.* 2014). The agarwood-producing region is roughly broken into two zones by agarwood collectors, the Hoi-An zone and Sin-Chew zone. It is generally understood that agarwood from China, Vietnam, Cambodia, India, Thailand, Laos, and Myanmar belong to the Hoi-An zone,

bioresources.com

and agarwood from Malaysia, Indonesia, and Brunei belong to the Sin-Chew zone. Agarwood from the Hoi-An zone generally sells at a much higher price because of its more pleasant scent (Nor Azah *et al.* 2013; Liu *et al.* 2017).

A common method of grading agarwood is for trained human graders to burn an agarwood piece or its oil and smell the scent leisurely (Ishihara *et al.* 1993; Ismail *et al.* 2014; Liu *et al.* 2017). When agarwood is heated or ignited, it releases volatile organic compounds (VOCs) (Ishihara *et al.* 1993). Chemical analyses of agarwood show that the most abundant compounds in the resin are sesquiterpenes (52%) and 2-(2-phenylethyl) chromone derivatives (41%) (Naef 2011; Lancaster and Espinoza 2012; Mei *et al.* 2013; Jong *et al.* 2014). Most sesquiterpenes possess unique odoriferous properties (Ishihara *et al.* 1993; Yang *et al.* 2016). In fact, the key components of VOCs are sesquiterpenes, which vary depending on geographic factors (Subasinghe and Hettiarachchi 2015), hence the unique scents of different cultivation areas and odor as the basis of agarwood classification.

The gas chromatograph mass spectrograph (GCMS) method has been used to successfully determine the oil composition of agarwood smoke from heating (Ishihara *et al.* 1993). Analyses of GC and GCMS have shown the existence of major sesquiterpenes and their chromone derivatives in agarwood oil (Tajuddin and Yusoff 2010; Ismail *et al.* 2014).

Explorations into sesquiterpenes, and other aromatic compounds of agarwood, detected *via* the GCMS method, have been attempted for years. Research on the classification of agarwood based on chemical composition has also attracted scholarly attention. Ishihara *et al.* (1993) used GC to identify the main components of agarwood and found sesquiterpenes to be the main components of its VOCs. When agarwood was burned, few pyrolyzed resin products were found, such as sesquiterpene hydrocarbons; the sesquiterpene composition of the smoke was quite similar to that of the extract. An analysis of the composition of oils obtained from healthy, naturally infected, and artificially inoculated eaglewood using GC and GCMS analyses found that a marked difference was shown between naturally infected and artificially inoculated eaglewood(Tamuli *et al.*2005). Resin content analysis can support the grading and quality assessment of gaharu in Malaysia and Southeast Asia (Nor Azah *et al.* 2013). The GCMS method has been used to distinguish essential oils obtained from agarwood simulated by chemical methods, wild agarwood, and healthy trees (Chen *et al.* 2011). Lancaster and Espinoza (2012) distinguished agarwood and other woods through DART-TOFMS analysis of 17 kinds of ions, and thoroughly discussed the principal ionization mechanisms (Ishihara *et al.* 1993). Gao *et al.* (2014) used gas GCMS combined with chemometric methods to identify natural and artificial agarwood; the authors also evaluated the quality of artificial agarwood. Nor Azah *et al.* (2013) used the GCMS method and principal components analysis to identify high-quality incense. Studies conducted on agarwood resin formation in three different locations using the GCMS method have revealed that the key components of sesquiterpenes variations in resin depend upon a trees geographic location and physical and biological damage. For example, Subasinghe and Hettiarachchi (2015) used GCMS to test and analyze 19 chemical components in Gyrinops Walla, especially typical sesquiterpene components. Studies have also shown that sesquiterpenes differ across agarwood species (Yang *et al.* 2016).

In this study, VOCs released from heat-treated agarwood were collected and tested *via* thermal desorption-gas chromatograph mass spectrometry (TD-GCMS). The

sesquiterpenes and other aromatic compounds in VOCs were the main target materials for research. Chemometrics analysis was used to facilitate classifications. Chemometrics is an analytical method of gathering information from multivariate chemical data through the application of statistical or mathematical techniques (Carbone *et al.* 2018). Discriminant analysis methods (kinds of chemometrics analysis) were evaluated in this study, namely stepwise regression (SR), Fisher discriminant analysis (FDA), and Bayes discriminant analysis (BDA). The SR method was used to pick up agarwood samples' characteristic chemical composition. Discriminant functions were gained through FDA and BDA. The aim of the study was to simulate artificial discrimination through TD-GCMS and chemometrics analysis. During this process, human sensory error caused by subjectivity, poor reproducibility, and time consumption may be effectively avoided (Liu *et al.*2017). In addition, the characteristic chemical compositions and the corresponding discriminant equationswere provided through chemometrics analysis, by which unknown growing region samples can be precisely judged. Moreover, the discriminant result was hoped to validate the work of the human grader. There are few studies that apply a TD-GCMS method combined with chemometrics analysis to the analysis of VOCs in heat-treated agarwood from the Hoi-An and Sin-Chew zones.

EXPERIMENTAL

Agarwood Samples

A total of 40 agarwood samples were tested: 26 from the Hoi-An zone and 14 from the Sin-Chew zone. The 26 Hoi-An zone samples were from Vietnam, China, Laos, Kampuchea, India, and Thailand; the 14 Sin-Chew zone samples were grown in Indonesia, Malaysia, Brunei, Papua New Guinea, and East Timor. All test samples were provided by the Beijing Agarwood Association (Beijing, China), and all were evaluated by the Association's assessment experts. The specific identification process was conducted by at least five experts with each sample, and the final evaluation result was based on the comprehensive opinions of all experts. Worthy of note is that the final result was adopted only when the consensus rate was no less than 80%.

The agarwood samples were dried at room temperature before being cut into small pieces and then filtered through 40-mesh sieves. The powdered samples (20 mg) were placed into a headspace bottle and washed with nitrogen before being heated to 160 °C and maintained at constant temperature for 1 h. Throughout the above process, VOCs released by the heat-treated agarwood were collected in the headspace bottle. All VOCs were used in GCMS analysis.

TD-GCMS Method

VOC analysis was conducted using a Markes Series 2 TD unit (Markes International Ltd., Llantrisant, England) installed on a gas chromatograph (Agilent 7890A, Agilent Technologies Inc., Beijing, China) coupled with a mass spectrometer (Agilent 5975C, Agilent Technologies Inc., Beijing, China) equipped with a column (DM 726641A:50 m × 320 μm × 1 μm; 325 °C, Dikma Technologies Inc., Beijing, China). Thermal desorption was measured with a TD unit using the following parameters: flow path temp: 180 °C; tube desorption: 280 °C; trap temperature: -10 °C; carrier gas (helium); flow rate: 80 mL/min. The ion sources were set at 230 °C. Mass spectra were scanned in the range m/z 35 to 400 amu. The chromatographic temperature program

proceeded as follows: 40 C at 1 min → 150 °C at 5° C/min →170 °C (3 min) at 1 °C/min →180 °C (3 min) at 1 °C/min → 200 °C at 10 °C/min → 280 °C (1 min) at 20 °C/min. The carrier gas (helium) had a splitless flow rate of 3 mL/min. VOC identification was achieved by comparing mass spectra. The minimum peak area of the output was 1% of the maximum peak area. The NIST 14 Mass Spec Library and Search Programs helped to identify the compounds (http://www.sisweb.com/manuals/nist.htm).

Stepwise Regression

Stepwise regression (SR) is one of the earliest statistical methods for analyzing and classifying a complicated dataset (Abdullah *et al.* 2001). A feature of SR is to eliminate one side and introduce data to the other side until no additional power of discrimination is obtained (Iskandrian *et al.* 1993; Lu *et al.* 2017). If the variable is inconspicuous for the model, it will be excluded until no discriminant variables are being entered into and excluded from the model (Zhu *et al.* 2016). The purpose of introducing SR in this study was to screen characteristic compounds.

There are two key parameters, Wilks' Λ and F,

$$\Lambda = |E| \,/\, |H+E| \tag{1}$$

where Λ is the ratio between the intergroup deviation cross-product matrix and the total deviation cross-product matrix, E is the intergroup deviation cross-product matrix, and H isthe between-groups deviation cross-product matrix. Therefore, both the numerator and denominator are values of the determinant of the matrix. The range of Λ values is between 0 and 1. A smaller value indicates a larger difference between groups. A value close to 1 indicates no between-group difference.

The F value is the ratio of the mean square to the mean square in the group,

$$F = [SS_{reg}(X_j)/1]/[SS_{reg}(n\text{-}p\text{-}1)] \tag{2}$$

where n is the total sample number, p is the number of independent variables in the equation, $SS_{reg}$is the sum of partial regression equations in X_j, and $SS_{res}$is the sum of squares of residuals. If the significance level is very small (*i.e.*, < 0.05), there is a significant difference between groups; if the value is larger (*e.g.*, > 0.10), the difference between groups is not significant.

Fisher Discriminant Analysis

Fisher discriminant analysis (FDA) (Fisher1936) utilizes a dimension reduction method to find the best (D-1)-dimensional hyperplane(s) that can divide a D-dimensional space into two or more subspaces. FDA is used to find a linear combination of continuous independent variables that characterize or separate two or more classes of objects or events (Fisher 1936; Lu *et al.* 2017). It is a classic and popular supervised learning method commonly used in face recognition, data rating recognition, and animal and plant recognition through some main features (Liu and Wechsler 2002; Alexandre-Cortizo *et al.* 2005; Witten and Tibshirani 2011).The Fisher criterion is defined as the ratio of the between-class variance to the within-class variance (Mahmoudi and Duman2015). Taking the standard deviation of data in both classes leads to a proper weight vector estimate to prevent overlap between the projected data in each class. To do so, by calculating within-class scatters (in matrix form), the weight vector onto which the data are projected can optimally split two classes. Its mathematical model is expressed as follows,

bioresources.com

$$\Delta(a) = SSR/SSE = (a^TBa)/(a^TAa) \rightarrow max \qquad (3)$$

where a is a p-dimension vector and a normal projection. SSE denotes within-class variance, while SSR denotes between-class variance. A represents within-class and B represents between-class scatter matrices. The mathematical model is the objective function in FDA; maximizing it results in better-separated classes with as little within-class overlap as possible.

Differentiating FDA with regard to a and equating to zero results in Eq. 4,

$$Ba = \lambda Aa \qquad (4)$$
$$a'Aa = 1$$

where λ and a, respectively, are the eigenvalue and eigenvector of $A^{-1}B$. The value of a is calculated using the above formula.

Bayes Discriminant Analysis

Bayes discriminant analysis (BDA), a common method based on probability and statistics theory, is characterized by high efficiency and accuracy. It allows for rapid and precise online testing and classification of observation objects.

The main mechanism of BDA is prior probability. Posterior probability is obtained by adjusting the prior probability according to a discriminant function; then, the probability of categories of predicted samples can be obtained based on posterior probability (Zhu *et al.* 2016). The maximized posterior probability is the foundation of the Bayes discriminant model. In this study, it is divided into two collectivities: Sin-Chew zone samples and Hoi-An zone samples. To determine the categories of new samples, the posterior probability of the totality was calculated as follows:

$$P(G_i/X) = q_i|\textstyle\sum^{(i)}|^{-1/2}exp[-1/2D^2(X,G_i)] / \textstyle\sum^2_{j=1}q_j|\textstyle\sum^{(i)}|^{-1/2}exp(-1/2D^2(X,G_j)) \qquad (5)$$

where,

$$q_i = n_i / [\textstyle\sum^2_{j=1} n_j]$$
$$D^2(X/G_i) = (x-u^{(i)})'(\textstyle\sum^{(i)})^{-1}(x-u^{(i)})$$

G_i (i=1,2) denotes the totality; $u^{(i)}$ and $\sum^{(i)}$, respectively, indicate the mean and covariance of the totality; and q_1, q_2 is the prior probability of the totality. The discrimination principle is,

$$P(G_i/X) = max_{1 \leq i \leq 2}P(G_j/X) \qquad (6)$$

where $X \in G_i$.

All data were processed using SPSS19.0 software for Windows (SPSS Inc., Chicago, IL, USA), and all figures were drawn in OriginPro 9.0 software (OriginLab Corporation, Northampton, MA, USA).

RESULTS AND DISSCUSION

Target Masses and Search Library

Diagnostic sesquiterpenes and other volatile aromatic compounds were used to construct a search library (Naef 2011). Due to isomeric configurations of the 67 sesquiterpenes and 37 simple aromatic compounds reported, there were 42 unique masses in total. Only 26 masses were detected in the 40 samples (Table 1). The average relative

peak area of the corresponding mass of two classification samples was calculated and recorded in the table. Because of the complex naming convention for the sesquiterpenes, the naming system was based on data reported by Naef (2011) and Chen *et al.* (2012), in which each sesquiterpene and simple volatile aromatic compound is assigned a number. The number appears in the upper right corner of the compound name in Table 1. The sesquiterpenes and volatile aromatics compounds were identified by their assigned numbers 1 through 66 and 106 through 142 (Naef 2011) and 68* (Chen *et al.* 2012).

General Chemical Composition

Table 1 displays the 26 unique target masses detected in 40 agarwood samples. The mass of sesquiterpenes was greater than 190, and the mass of the other volatile aromatic compounds was less than 190. Few pyrolyzed sesquiterpenes were found in the VOCs, as noted in previous studies (Ishihara *et al.* 1993). In the analysis of VOCs, many other volatile aromatic compounds were found in addition to sesquiterpene.

Table 1. Relative Peak Areas of Unique Target Masses Detected in 40 Agarwood Samples

m/z	Compound Name	Relative Peak Area on Average	
		Hoi-An	Sin-Chew
92.063	Toluene[107]	0.39	0.36
94.042	Phenol[119]	0.04	0.05
96.021	Furfural[134]	0.30	0.78
98.037	Furfuryl alcohol[135]	0.03	0.05
104.063	Styrene[108]	0.13	0.27
106.042	Benzaldehyde[115]	3.65	4.96
120.058	Acetophenone[127]	0.54	0.86
122.073	p-Methylanisol[112]	0.01	0.01
124.052	p-Guaiacol[123]	0.01	0.02
128.063	Naphthalene[109]	0.03	0.02
134.073	p-Vinylanisol[114]	0.61	0.52
136.052	Anisaldehyde[117]	1.38	0.89
136.089	p-Ethylanisol[113]	0.02	0.02
148.052	Dihydrocoumarin[137]	0.09	0.36
148.089	Benzylacetone[128]	3.30	5.56
150.068	p-Vinylguaiacol[125]	0.01	0.02
178.099	p-Methoxy-benzylacetone=Anisyl acetone[129]	0.79	0.64
190.172	(S)-4a-Methyl-2-(1-methylethylidene)-1,2,3,4,4a,5,6,7-octahydronaphthalene[27];(S)-4a-Methyl-2-(1-methylethyl)-3,4,4a,5,6,7-hexahydronaphthalene[28];(2R,4aS)-4a-Methyl-2-(1-methylethenyl)-1,2,3,4,4a,5,6,7-octahydronaphthalene[29]	0.14	0.02
202.172	ar-Curcumene[68*]	3.15	2.73
204.188	α-Guaiene[41];α-Bulnesene[42]	5.58	15.02
218.167	(5S,7S,10S)-(-)-Selina-3,11-dien-9-one[17];Selina-3,11-dien-14-al[20];Selina-4,11-dien-14-al[22];(+)-(4S,5R)-Dihydrokaranone[31];Eremophila-9,11-dien-8-one(Neopetasane)[33];rel-(3R,7R,9R,10S)-9,10-Dimethyl-6-methylene-4-oxatricyclo[7.4.0.0^{3,7}]tridec-1-ene,(8,12-Epoxyeremophila-9,11(13)-diene)[35];(-)-Guaia-1(10),11-dien-15-al[45];(+)-Guaia-1(10),11-dien-9-one[50];Rotundone[51];rel-(5R,10R)-2-Isopropylidene-10-methyl-spiro[4.5]dec-6-ene-6-	1.51	1.33

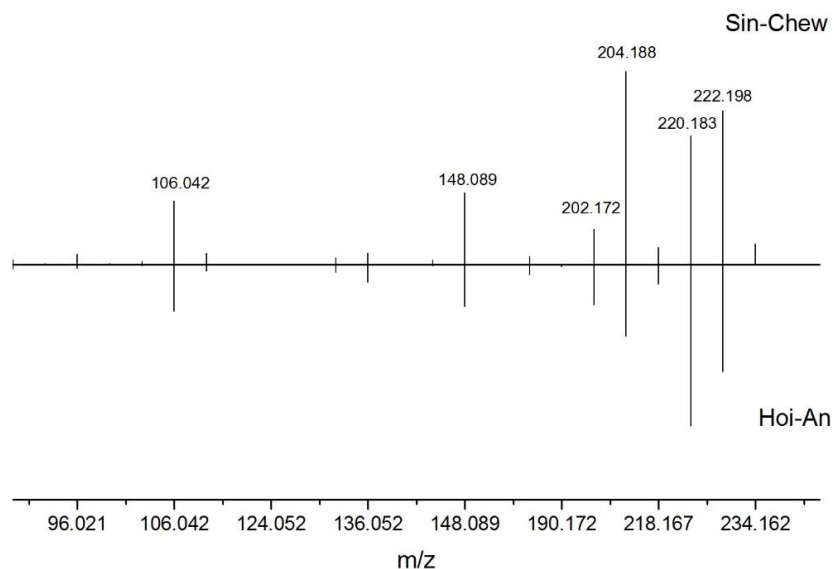| *m/z*	Compound Name	Relative Peak Area on Average	
		Hoi-An	Sin-Chew
	carbaldehyde(Vetispira-2(11),6-dien-14-al)[62];rel-(1R,2R)-9-Isopropyl-2-methyl-8-oxatricyclo[7.2.1.0[1,6]]dodeca-4,6-diene(2,14-Epoxy-vetispira-6(14),7-diene)[64]		
220.183	α-Agarofuran[1];β-Agarofuran[2];(5S,7S,9S,10S)-(+)-Selina-3,11-dien-9-ol[18];Selina-3,11-dien-14-ol[19];rel-(2R,8R,8aS)-2-(1,2,3,5,6,7,8,8a-Octahydro-8,8a-dimethyl-2-naphthyl)-prop-2-en-1-ol(Eremophila-9,11(13)-dien-12-ol[34];Dehydro-jinkoh-eremol[37];rel-(2R,8S,8aS)-2-(1,2,6,7,8,8a-Hexahydro-8,8a-dimethyl-2-naphthyl)-propan-2-ol(Valenca-1(10),8-dien-11-ol)[39](-)-Epoxyguai-11-ene(epoxybulnesene)[43];(-)-Guaia-1(10),11-dien-15-ol[44];(+)-1,5-Epoxy-nor-ketoguaiene[52];rel-(1R,2R)-9-Isopropyl-2-methyl-8-oxatricyclo[7.2.1.0[1,6]]dodec-5-ene(2,14-Epoxy-vetispir-6-ene)[63];rel-(5R,7S,10R)-2-Isopropylidene-10-methyl-6-methylene-spiro[4.5]decan-7-ol(Vetispira-2(11),6(14)-dien-7-ol)[65]	12.58	10.00
222.198	Dihydro-β-agarofuran[3];nor-Keto-agarofuran[14];Agarol (11(13)-Eudesmen-12-ol)[25];10-epi-γ-Eudesmol[26];(-)-(4R,5S,7R)-Jinkoh-eremol[36];(+)-(4R,5S,7R)-Kusunol (= Valerianol)[38];epi-Ligulyl oxide[53];Jinkohol[56];Jinkohol II[57];Agarospirol, (2R,5R,10R)- α, α,6,10-tetramethyl-spiro [4,5]dec-6-ene-2-methanol[58];Isoagarospirol[59]	8.35	11.93
234.162	Selina-3,11-dien-14-oic acid(as methyl ester)[21];Selina-4,11-dien-14-oic acid(as methyl ester)[23];(-)-Guaia-1(10),11-diene-15-carboxylic acid[46]	0.05	1.59
236.178	Dehydrobaimuxinol[9];(1S,2S,6S,9R)-6,10,10-Trimethyl-11-oxatricyclo[7.2.1.0[1,6]]dodecane-2-carbaldehyde[10];Sinenofuranal[55];Oxo-agarospirol = Baimuxinal[60];(-)-(4R,5R,7R)-11-Hydroxy-vetispir-6-en-8-one[66]	0.02	0.00
250.157	9-Hydroxy-selina-4,11-dien-14-oic acid(as methyl ester)[24];(-)-2α-Hydroxyguaia1(10),11-dien-15-oic acid(identified in acidic fraction as Me-ester)[49]	0.00	0.02

Fig. 1. GCMS spectrum of 26 target unique masses in Sin-Chew and Hoi-An agarwoods

Figure 1 displays the 26 unique masses of 40 samples and their relative peak areas. The major peaks observed were m/z = 204.188, 222.198, 220.183, 148.089, 106.042, 202.172, *etc*. Comparing the relative peak areas in Table 1 with the data in Fig. 1, many minor peaks were observed between m/z=90 and 260. All the peaks roughly showed common points and different points in the two sample areas. The relative peak areas of the components of the two area samples differed at the following points: m/z=204.188,222.198, 148.089, 250.157, and 148.052.

It is worth mentioning that the heating temperature was selected based on the few pyrolyzed sesquiterpenes that appeared when samples were heat-treated. The reference temperature used in previous studies was 180 to 210 °C. In combination with artificial identification and previous studiesby temperature, 160 °C was chosen as the heat-treating temperature in the study.Based on these experiments, a temperature range from 120 °C to 180 °C could be considered in future research.

Principal Component Analysis

Principal component analysis (PCA) was employed on the 40 samples and on the relative peak areas of 26 unique masses. The first 10 PCs accounted for 81.0% of the total variance. Figure 2 shows the sum of the first three functions, which explained 41.8% of the total variance. Figure 2 also displays the effects of the first three functions, which were not significant in terms of the classification of the two areas' samples.
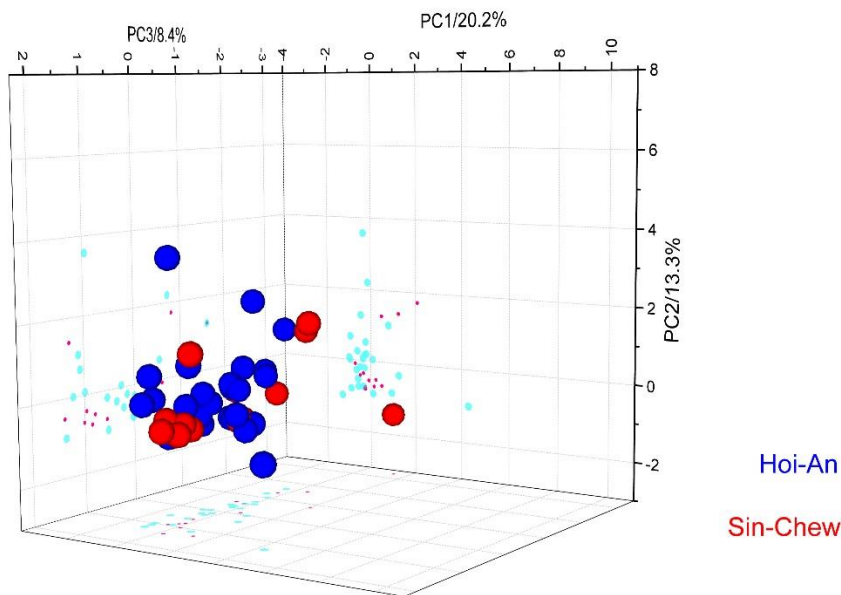

Fig. 2. PCA of 26 unique target masses detected in Sin-Chew and Hoi-An agarwoods

PCA was used to select a main component structure direction of the samples. Through the PCA calculation process, the loadings of variables from the top 10 principal components were derived. By comparing the variable loadings, eight of 26 unique target masses were identified as making the greatest contributions to the model. The masses of 104.063, 136.052, 222.198, 202.172, 96.021, 204.188, 250.157, and 128.063 represented the main composition structure of all agarwood samples.

Stepwise Regression

Among the multivariate methods available, PCA is one of the most common unsupervised techniques, while stepwise discriminant analysis (SDA) is frequently applied as a supervised technique for sample classification purposes. While PCA selects a direction that retains the maximum structure of data on a reduced dimension, SDA selects a direction that achieves maximum separation between given sample classes (Berrueta *et al.* 2007, Li *et al.* 2014; Figueiredo *et al.* 2016). Stepwise regression was introduced to screen key VOCs components as characteristic chemicals to make a classification.

During the subsequent statistical analysis, 26 unique masses were subdivided by their peak time. In total, 155 ions, identified by peak time, with 26 unique masses and the ions' relative peak areas were used. In this study, a matrix was established for 155 ions from 40 samples. The Λ, P, and F values were calculated, and the characteristic compounds were screened based on two principles. First, the compound with the minimum Λ value was selected to enter the model, after which the Λ values of the remaining components and the selected component were recalculated separately. The remaining compound with the minimum Λ value was entered into the model following the compound of the previous entry. Principle two stipulates that when $F>F_{\alpha in}$ or $P<\alpha_{in}$ is entered, $F\leq F_{\alpha out}$ or $P\geq \alpha_{out}$ is removed, such that $\alpha_{in} = 0.05$ and $\alpha_{out} = 0.10$.

The above process of entering and removing variables step-by-step to screen the characteristic compounds was repeated. Finally, 26 characteristic compounds were obtained, represented as X1 through X26 for use in the subsequent discriminant analysis (Table 2).

Table 2. 26 Characteristic Compounds and the Peak Time Detected in 40 Agarwood Sample

Characteristic Compounds	*m/z*	Peak Time	Characteristic Compounds	*m/z*	Peak Time
X1	222.198	49.663	X14	204.188	53.317
X2	204.188	37.31	X15	204.188	57.926
X3	122.073	19.501	X16	204.188	52.445
X4	92.063	10.901	X17	220.183	56.025
X5	222.198	49.161	X18	202.172	39.549
X6	204.188	53.44	X19	202.172	48.034
X7	204.188	41.038	X20	218.167	60.444
X8	134.073	23.781	X21	204.188	51.704
X9	220.183	62.979	X22	204.188	44.593
X10	222.198	52.017	X23	218.167	56.839
X11	148.089	26.71	X24	222.198	54.313
X12	220.183	45.129	X25	202.172	38.224
X13	204.188	41.31	X26	222.198	46.486

SR selects a direction that achieves maximum separation between given sample classes. In this study, the SR method was used to screen 26 characteristic compounds with 9 masses, as follows: 92.063, 122.073, 134.073, 148.089, 202.172, 204.188, 218.167, 220.183, and 222.198. Masses of 202.172, 204.188, 218.167, 220.183, and 222.198 represented sesquiterpenes; masses of 92.063, 122.073, 134.073, and 148.089 represented the other volatile aromatic compounds.

To compare the SR results with those from theabove analysis, the VOCs with

respective masses of 204.188, 222.198, and 148.089 were highlighted in the difference observed in Fig. 1. The results of the respective masses of 204.188, 222.198, and 202.172 aligned with the PCA results. Thus, the sesquiterpene compounds with respective masses of 204.188 and 222.198 were the key chemical substances, denoting not only the main components of the agarwood, but also the main substances to classify the difference. In previous studies of chemical compositions in agarwood, there have been many observations of 204.188 and 222.198 masses (CAS No.: 88-84-6, 3691-11-0, 20053-66-1, 5986-25-4, 1460-64-6, 15051-81-7, 86747-08-2, 20489-45-6, 18680-81-4, 66512-57-0, 86703-03-9, 1460-73-7, *etc*.).

Fisher Discriminate Analysis

In this study, the Fisher discriminant function is defined as follows,

$$F_{H\text{-}S} = a^T x + c = a^T x + a^T \bar{x} \tag{7}$$

The barycentric coordinates are defined as,

$$D_j = a^T (\bar{x}^{(j)} - \bar{x}) \quad (j=1,2) \tag{8}$$

where a is the coefficient and c is a constant. $\bar{x}^{(j)}$ are means of classes; $\bar{x}$ is the combined mean of all samples.

The agarwood samples were divided into two classes: the Hoi-An zone and Sin-Chew zone. When j=1, the formulate belongs to the Sin-Chew zone; when j=2, the formulate represents the Hoi-An zone. In this study, the model was established for 40 agarwood samples and their 26 characteristic compounds.

The calculated discriminant equation is as follows,

$$Y_{FH\text{-}S} = -35.696 \times X1 + 26.545 \times X2 - 381.429 \times X3 - 2.212 \times X4 + 11.166 \times X5 - 4.880 \times X6 + 5.354 \times X7 - 11.099 \times X8 + 51.132 \times X9 - 1.771 \times X10 + 0.479 \times X11 + 12.806 \times X12 + 19.542 \times X13 + 0.849 \times X14 - 34.822 \times X15 + 86.344 \times X16 - 10.656 \times X17 - 124.136 \times X18 - 61.055 \times X19 - 76.724 \times X20 - 2.085 \times X21 + 41.704 \times X22 + 87.366 \times X23 - 3.834 \times X24 - 12.010 \times X25 + 9.094 \times X26 - 33.817 \tag{9}$$

where $Y_{FH\text{-}S}$ is the value of Fisher discriminant function.

The barycentric coordinates of the Hoi-An zone and Sin-Chew zone were -35.795 and 66.477, respectively.

Table 3. The Eigenvalue of the Discriminant Function

Discriminant Function	Eigen value	% of Variance	Canonical Correlation
$F_{H\text{-}S}$	2504.774	100.0	1.000

The important parameters of the FDA model are displayed in Table 3. The eigenvalue of the discriminant function was 2504.774, and the discriminant function completely explained the variance. The canonical correlation was 1.000. The above parameters indicate that the function's discriminating effect was significant. Hence, FDA was applied to divide the agarwood samples into two classes: the Hoi-An zone and Sin-Chew zone.

For illustration, Fig. 3 depicts Fisher discriminant function for the 40 samples from the Hoi-An zone and Sin-Chew zone using their 26 characteristic compounds. The

discriminant function plot shows that the function discriminated the Hoi-An zone samples from the Sin-Chew zone samples. Compared with artificial discrimination, discriminant functional leviated human sensory error induced from subjectivity. Therefore, this function can be used to evaluate the accuracy of human grader'sjudgment result.
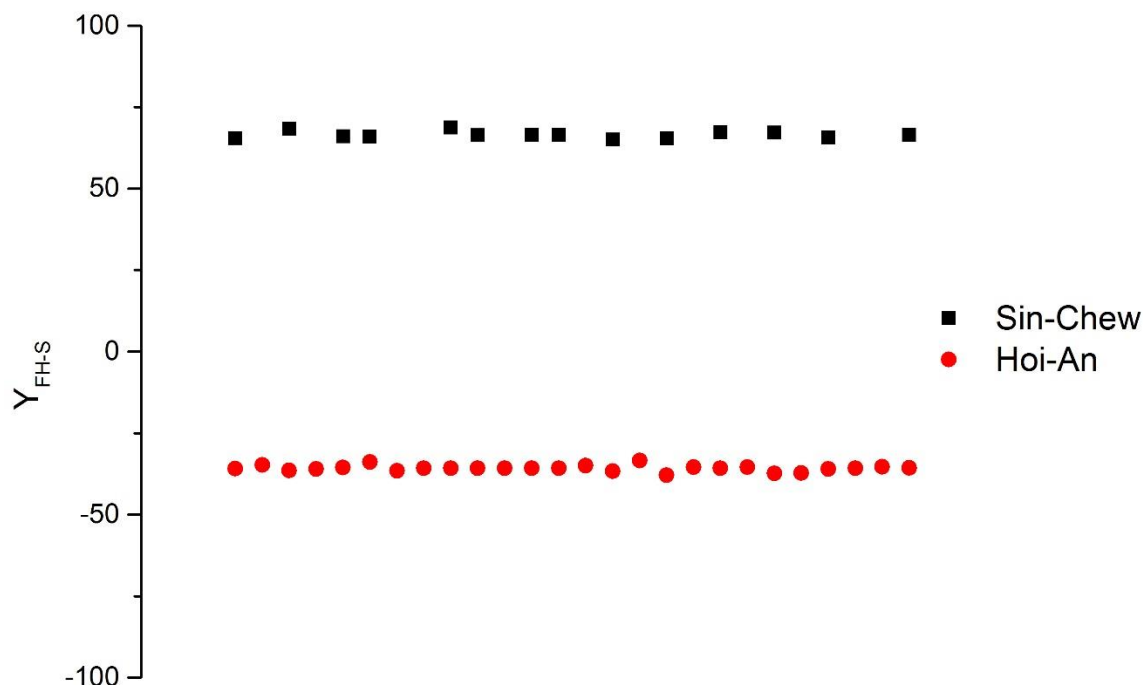

Fig. 3. Fisher discriminant function values for 40 samples

After analyzing theFisher discriminant function and its coefficients, the top three components with significant influence were 122.073 (X3), 202.172 (X18), and 218.167 (X23), with respective coefficients of -381.429, -124.136, and 87.366. The most influential substance appeared to be X3 (p-Methylaniso), and two other substances thatwere both sesquiterpenes.

Moreover, taking the mass of 204.188 as an example, the number of corresponding ions was 9, but the contribution coefficient of each ion in the formula was different. This indicates that it is necessary to subdivide components by peak time when the agarwood classification was determined to be in the Hoi-An zone or Sin-Chew zone, according to discriminate analysis.

It is worth mentioning that the Fisher discriminant model does not provide taxonomy. The unweighted methodwas used to calculate barycentric coordinates. When encountering a sample of unknown classification, after the relative peak areas of the characteristic chemicals are tested, the relative peak areas should be put into the FDA formula (Eq. 9) to calculate the unknown samples' results. Comparing distances between the results and the barycentric coordinates helped to classify samples from the Hoi-An zone and Sin-Chew zone. The classification of barycentric coordinates closer to the result helped to identify unknown samples.

Bayes Discriminant Analysis

The Bayes discriminant function is defined as follows,

$$y(j/x) = \hat{\beta}_{j0} + \hat{\beta}'_j x \tag{10}$$

where $\hat{\beta}_j = S^{-1}\bar{x}^{(j)}$ is the discriminant coefficient, and $\hat{\beta}_{j0} = \ln q_j - \frac{1}{2}\bar{x}'^{(j)}S^{-1}\bar{x}^{(j)}$ is a constant.

S is the total sample covariance matrix, $\bar{x}^{(j)}$ are means of classes, and q_j is the posterior probability for class j samples.

When j=1, the formulate represents the Sin-Chew zone; when j=2, the formulate represents the Hoi-An zone. In this study, the model was established for 40 agarwood samples and their 26 characteristics.

Bayes discriminant functions of the Sin-Chew and Hoi-An samples were as follows,

$$
\begin{aligned}
Y_{1BH\text{-}S} = &- 3580.050 \times X1 + 2664.676 \times X2 - 38255.655 \times X3 \\
&- 217.588 \times X4 + 1119.588 \times X5 - 488.123 \times X6 + 536.508 \\
&\times X7 - 1109.138 \times X8 + 5133.602 \times X9 - 177.312 \times X10 \\
&+ 48.427 \times X11 + 1284.653 \times X12 + 1960.714 \times X13 \\
&+ 85.107 \times X14 - 3489.342 \times X15 + 8658.033 \times X16 \\
&- 1067.824 \times X17 - 12446.075 \times X18 - 6125.367 \times X19 \\
&- 7697.598 \times X20 - 209.158 \times X21 + 4180.503 \times X22 \\
&+ 8760.754 \times X23 - 384.387 \times X24 - 1199.074 \times X25 \\
&+ 912.432 \times X26 - 5033.486
\end{aligned} \tag{11}
$$

$$
\begin{aligned}
Y_{2BH\text{-}S} = &\ 70.589 \times X1 - 50.126 \times X2 + 753.724 \times X3 + 8.588 \\
&\times X4 - 22.341 \times X5 + 10.965 \times X6 - 11.032 \times X7 + 25.951 \\
&\times X8 - 95.771 \times X9 + 3.846 \times X10 - 0.539 \quad \times X11 - 24.990 \\
&\times X12 - 37.844 \times X13 - 1.720 \times X14 + 71.937 \times X15 - 172.554 \\
&\times X16 + 21.948 \times X17 + 249.565 \times X18 + 118.803 \times X19 \\
&+ 149.102 \times X20 + 4.027 \times X21 - 84.626 \times X22 - 174.325 \\
&\times X23 + 7.760 \times X24 + 29.247 \times X25 - 17.641 \times X26 - 6.078
\end{aligned} \tag{12}
$$

where $Y_{1BH\text{-}S}$ and $Y_{2BH\text{-}S}$ are values of Bayes discriminant functions.

Table 4. Test of Wilks' Λ

Wilks' Λ	Chi-square	df	Sig
0.000	195.659	26	0.000

Table 5. Results of Bayes Discrimination Model

	Initial validation		Cross-validation	
Sample	Hoi-An zone	Sin-Chew zone	Hoi-An zone	Sin-Chew zone
Error number	0	0	1	0
Accuracy(%)	100	100	96.2	100
Total accuracy(%)	100		97.5	

*df indicates degree of freedom; Sig indicates significance.

Table 4 displays the Wilks' test results of the discriminant function, indicating whether the function is statistically significant. In this study, Wilks' test was an important measure for correctly discriminating between agarwood samples from the Hoi-An zone

bioresources.com

and those from the Sin-Chew zone. The discriminatory power of the discriminant function was significant (Table 4, sig. 0.000).

Table 5 presents the results of the Bayes stepwise discriminant model. Only 26 variables were included in this model. The initial-validation accuracy of the model was 100%, and the cross-validation accuracy was 97.5%. One Hoi-An Zone agarwood was wrongly discriminated; the accuracy reached the desired level.

The BDA model was applied to divide agarwood samples into two classes (*i.e.*, the Hoi-An zone and Sin-Chew zone). The discriminant functions plot (Fig. 4) showed that the functions discriminated samples from these two zones. It is similar to the Fisher discriminant function that the Bayes functions can be used to validate the work of the assessment experts.
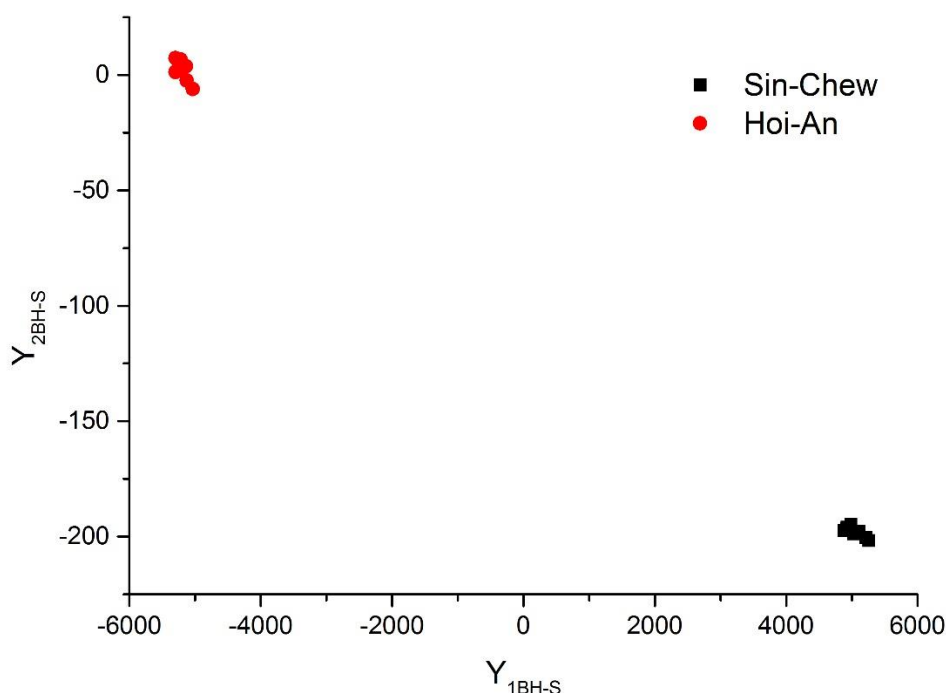

Fig. 4. Bayes discriminant functions values for 40 samples

Compared with the absolute values of the difference between the coefficients of the two Bayes formulas (Eqs. 11 and 12), the top three components with the absolute values were 122.073 (X3), 202.172 (X18), and 218.167 (X23); their respective coefficients were -38255.655/753.724, -12446.075/249.565, and 8760.754/-174.325. The most influential substances were determined to be X3 (p-Methylaniso), consistent with the FDA results. The other two substances were sesquiterpenes.

When encountering a sample of unknown classification, the relative peak areas of the corresponding characteristic compounds were substituted into the BDA formulas, Eqs. 11 and 12, and formula values were obtained. When comparing two values, the corresponding classification with a larger value was used to classify the unknown samples.

It is worth mentioning that the study only uses chemometrics analysis (SR, FDA, and BDA) to distinguish the different agarwood sample's cultivation region. However,

more specific researchwere hoped to be carried out by using the chemometrics analysisfor reference in future, such as the survey of the exact location from different countries and the identification of different agarwood species.

CONCLUSIONS

1. It is feasible to use the TD-GCMS method combined with chemometrics analysis to analyze the VOCs of heat-treated agarwood instead of using experienced graders to classify agarwood samples from the Hoi-An zone and Sin-Chew zone. Sesquiterpenes and other volatile aromatic compounds are the key components of VOCs. These key substances are not only the main components of agarwood samples but also an important means of discrimination.

2. Chemometrics analysis *via* SR, FDA, and BDA can be applied to agarwood classification. During this process, a stepwise regression model was found to simplify the characteristic compounds and improve predictive accuracy. Twenty-six characteristic compounds were screened *via*the stepwise regression.These characteristic components can be used as a basis for discriminant analysis (FDA and BDA).

3. FDA and BDA were conducted, based on the 26 compounds, to classify the agarwood samples. The results showed that FDA and BDA both classified agarwood effectively. The results of this study also helped to identify previously unknown growing districts of agarwood to make classifications. The specific method is to substitutethe 26 characteristic compound'srelative peak areaof unknown sample into FDA or BDA discriminant equations, and compared the formula values.The growing district of unknown agarwood was then identified.

ACKNOWLEDGMENTS

REFERENCES CITED

Abdullah, M. Z., Guan, L. C., and Mohd Azemi, B. M. N. (2001). "Stepwise discriminant analysis for colour grading of oil palm using machine vision system," *Food Bioproducts Processing* 79(4), 223-231. DOI: 10.1205/096030801753252298

Alexandre-Cortizo, E., Rosa-Zurera, M., and Lopez-Ferreras, F. (2005). "Application of Fisher linear discriminant analysis to speech/music classification," in: *EUROCON 2005-The International Conference on Computer as a Tool*, Belgrade, Serbia, pp.1666-1669. DOI: 10.1109/EURCON.2005.1630291

Berrueta, L. A., Alonso-Salces, R. M., and Héberger, K. (2007). "Supervised pattern recognition in food analysis," *Journal of Chromatography A* 1158(1), 196-214. DOI: 10.1016/j.chroma.2007.05.024

Carbone, K., Ciccoritti, R., Paliotta, M., Rosato, T., Terlizzi, M., and Cipriani, G. (2018). "Chemometric classification of early-ripening apricot (*Prunus armeniaca*, L.) germplasm based on quality traits, biochmical profiling and in virto biological

activity," *Scientia Horticulturae*227(2018), 187-195. DOI: 10.1016/j.scienta.2017.09.020

Chen, H. Q., Yan, Y., Xue, J., Wei, J. H., Zhang, Z., and Chen, H. J. (2011). "Comparison of compositions and antimicrobial activities of essential oils from chemically stimulated agarwood, wild agarwood and healthy *Aquilaria sinensis* (Lour.) gilg trees," *Molecules* 16(6), 4884-4896. DOI: 10.3390/molecules16064884

Figueiredo, A. B., Magina, S., Evtuguin, D. V., Cardoso, E. F., Ferra, J. M., and Cruz, P. (2016). "Factors affecting the dimensional stability of decorative papers under moistening,"*BioResources* 11(1), 2020-2029. DOI: 10.15376/biores.11.1.2020-2029

Fisher, R. A. (1936). "The use of multiple measurements in taxonomic problems," *Annals of Human Genetics* 7(2), 179-188. DOI: 10.1111/j.1469-1809.1936.tb02137.x

Gao, X., Xie, M., Liu, S., Guo, X., Chen, X., Zhong, Z.,Wang, L., and Zhang, W. (2014). "Chromatographic fingerprint analysis of metabolites in natural and artificial agarwood using gas chromatography-mass spectrometry combined with chemometric methods," *Journal of Chromatography B Analytical Technologies in the Biomedical & Life Sciences* 967, 264-273. DOI: 10.1016/j.jchromb.2014.07.039

Ismail, N., Mohd Alib, N. A., Jamil, M., Rahiman, M. H. F., Tajuddin, S. N., and Taib, M.N. (2014). "A review study of agarwood oil and its quality analysis," *JurnalTeknologi(Sciences & Engineering)* 68(1), 37-42. DOI: 10.11113/jt.v68.2419

Ishihara, M., Tsuneya, T., and Uneyama, K.(1993). "Components of the agarwood smoke on heating," *Journal of Essential Oil Research* 5(4), 419-423. DOI: 10.1080/104129

Ito, M. (2008). "Studies on perilla, agarwood, and cinnamon through a combination of fieldwork and laboratory work," *Journal of Natural Medicines* 62(4), 387-395. DOI: 10.1007/s11418-008-0262-z

Jong, P. L., Tsan, P., and Mohamed, R. (2014). "Gas chromatography-mass spectrometry analysis of agarwood extracts from mature and juvenile *Aquilaria malaccensis*," *International Journal of Agriculture & Biology* 16(3), 644-648.

Lancaster, C., and Espinoza, E. (2012). "Evaluating agarwood products for 2-(2-phenylethyl) chromones using direct analysis in real time time-of-flight mass spectrometry," *Rapid Communications Mass Spectrometry* 26(23), 2649-2656. DOI: 10.1002/rcm.6388

Li, B., Liu, H., Xu, H., Pang, B., Mou, H.,Wang, H., and Mu, X. (2014). "Characterization of the detailed relationships of the key variables in the process of the alkaline sulfite pretreatment of corn strover by multivariate analysis," *BioResources* 9(2), 2757-2771. DOI: 10.15376/biores.9.2.2757-2771

Liu, C., and Wechsler, H. (2002). "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society* 11(4), 467. DOI: 10.1109/TIP.2002.999679

Liu, Y., Wei, J., Gao, Z., Zhang,Z., and Lyu, J. (2017). "A review of quality assessment and gradingfor agarwood," *Chinese Herbal Medicines* 9(1),22-30. DOI: 10.1016/s1674-6384(17)60072-8

Lu, J., Ehsani, R., Shi, Y., Abdulridha, J., de Castro, A. I., and Xu, Y. (2017). "Field detection of anthracnose crown rot in strawberry using spectroscopy technology," *Computers & Electronics in Agriculture* 135, 289-299. DOI: 10.1016/j.compag.2017.01.017

Mahmoudi, N., and Duman, E. (2015). "Detecting credit card fraud by modified Fisher discriminant analysis," *Expert Systems with Applications* 42(5), 2510-2516. DOI:

10.1016/j.eswa.2014.10.037

Mei, W., Yang, D., Wang, H., Yang, J., Zeng, Y., Guo, Z., Dong, W., Li, W., and Dai, H. (2013). "Characterization and determination of 2-(2-phenylethyl)chromones in agarwood by GC-MS," *Molecules* 18(10), 12324-12345. DOI: 10.3390/molecules181012324

Naef, R.(2011). "The volatile and semi-volatile constituents ofagarwood, the infected heartwood of *Aquilaria* species: A review," *Flavour and Fragrance Journal* 26, 73-89. DOI: 10.1002/ffj.2034

Nor Azah, M.A., Saidatul Husni, S., Mailina, J., Sahrim, L., Abdul Majid, J., and Mohd Faridz, Z. (2013). "Classification of agarwood (gaharu) by resin content," *Journal of Tropical Forest Science* 25(2), 213-219.

Subasinghe, S. M. C. U. P., and Hettiarachchi, D. S. (2015). "Characterisation of agarwood type resin of *Gyrinops walla* Gaertn growing in selected populations in Sri Lanka," *Industrial Crops & Products* 69, 76-79. DOI: 10.1016/j.indcrop.2015.01.060

Tajuddin, S. N., and Yusoff, M. M. (2010). "Chemical composition of volatile oils of *Aquilaria malaccensis* (Thymelaeaceae) from Malaysia," *Natural Product Communications* 5(12), 1965-1968.

Tamuli, P., Boruah, P., Nath, S. C., and Leclercq, P. (2005). "Essential oil of eaglewood tree: a product of pathogenesis," *Journal of Essential Oil Research* 17(6), 601-604. DOI: 10.1080/10412905.2005.9699008

Witten, D. M., and Tibshirani, R. (2011). "Penalized classification using Fisher's linear discriminant," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73(5), 753-772. DOI: 10.1111/j.1467-9868.2011.00783.x

Yang, D., Li, W., Dong, W., Wang, J., Mei, W., and Dai, H. (2016). "Five new 5,11-epoxyguaiane sesquiterpenes in agarwood "Qi-Nan" from *Aquilaria sinensis*," *Fitoterapia* 112, 191-196. DOI: 10.1016/j.fitote.2016.05.014

Zhu, Z., Li, W., Wang, Q., Tang, Y., Cao, F., and Ma, R. (2016). "Online discriminant model of blood spot eggs based on spectroscopy," *Journal of Food Process Engineering*, 40(3), 1-7 DOI: 10.1111/jfpe.12435