

Article

# Discriminative Analysis of Different Grades of Gaharu (*Aquilaria malaccensis* Lamk.) via <sup>1</sup>H-NMR-Based Metabolomics Using PLS-DA and Random Forests Classification Models

Siti Nazirah Ismail <sup>1</sup>, M. Maulidiani <sup>1</sup>, Muhammad Tayyab Akhtar <sup>1</sup>, Faridah Abas <sup>1,2</sup>,  
Intan Safinar Ismail <sup>1,3</sup> , Alfi Khatib <sup>4</sup>, Nor Azah Mohamad Ali <sup>5</sup> and Khozirah Shaari <sup>1,3,\*</sup>

<sup>1</sup> Laboratory of Natural Products, Institute of Bioscience, Universiti Putra Malaysia, 43400 Serdang, Malaysia; sitinazirahismail12@yahoo.com (S.N.I.); dieni\_mauludiana@yahoo.com (M.M.); tayyabakhtar@hotmail.com (M.T.A.); faridah\_abas@upm.edu.my (F.A.); safinar@upm.edu.my (I.S.I.)

<sup>2</sup> Department of Food Science, Faculty of Food Science and Technology, Universiti Putra Malaysia, 43400 Serdang, Malaysia

<sup>3</sup> Department of Chemistry, Faculty of Science, Universiti Putra Malaysia, 43400 Serdang, Malaysia

<sup>4</sup> Department of Pharmaceutical Chemistry, Kulliyah of Pharmacy, International Islamic University Malaysia (Kuantan Campus), Bandar Indera Mahkota, 25200 Kuantan, Malaysia; alfikhatib@iiu.edu.my

<sup>5</sup> Forest Research Institute Malaysia, 52109 Kepong, Malaysia; norazah@frim.gov.my

\* Correspondence: khozirah@yahoo.com.my; Tel.: +60-3-8942148

Received: 17 August 2017; Accepted: 21 September 2017; Published: 25 September 2017

**Abstract:** Gaharu (agarwood, *Aquilaria malaccensis* Lamk.) is a valuable tropical rainforest product traded internationally for its distinctive fragrance. It is not only popular as incense and in perfumery, but also favored in traditional medicine due to its sedative, carminative, cardioprotective and analgesic effects. The current study addresses the chemical differences and similarities between gaharu samples of different grades, obtained commercially, using <sup>1</sup>H-NMR-based metabolomics. Two classification models: partial least squares-discriminant analysis (PLS-DA) and Random Forests were developed to classify the gaharu samples on the basis of their chemical constituents. The gaharu samples could be reclassified into a ‘high grade’ group (samples A, B and D), characterized by high contents of kusunol, jinkohol, and 10-epi- $\gamma$ -eudesmol; an ‘intermediate grade’ group (samples C, F and G), dominated by fatty acid and vanillic acid; and a ‘low grade’ group (sample E and H), which had higher contents of aquilarone derivatives and phenylethyl chromones. The results showed that <sup>1</sup>H-NMR-based metabolomics can be a potential method to grade the quality of gaharu samples on the basis of their chemical constituents.

**Keywords:** gaharu; *Aquilaria malaccensis*; quality; NMR-based metabolomics; PLS-DA; Random Forests classifier

## 1. Introduction

*Aquilaria* Lam. (family Thymelaeaceae), is a genus with 22 accepted species [1], distributed mainly in the tropical forest of South-East Asian countries, including Malaysia [2,3]. Species of this genus are the principal source of gaharu (also known as agarwood, aloeswood or eaglewood) which is one of the most valuable forest products, traded internationally for centuries. Gaharu refers to the fragrant, dark and dense, resinous heartwood, pathologically formed in response to injury and microbial infection [4–6]. This forest product is not only popular as incense and in perfumery, but it is also favored in traditional medicine, with sedative, carminative, cardioprotective and analgesic effects reported to be among the ethnopharmacological properties associated with it [7,8]. *Aquilaria* species that have been reported

to produce valuable gaharu include *A. malaccensis* Lamk., *A. crassna* Pierre ex Lecomte, *A. beccariana* Tiegh., *A. hirta* Ridl., *A. rostrata* Ridl., *A. sinensis* (Lour.) Spreng., *A. microcarpa* Baill., *A. filaria* (Oken) Merr. and *A. khasiana* Hallier f. [9]. By far, *A. malaccensis* or kekaras (in Malay), is the most common gaharu-producing species in Malaysia. The species has been well recognized for its commercial value, representing an important export commodity to the Middle East countries, Japan and China [4,10,11].

Gaharu is traded in the form of wood chips, sawn wood, and resin or distilled oil. In trade, the selling price varies greatly, depending on the 'quality' [12,13]. The quality of gaharu is usually graded based on its physical properties such as wood color, weight or density, and aroma upon burning, all of which, indirectly reflect the resin content of the gaharu samples. High quality gaharu would have higher content of resin, would be of higher density, and would have darker color and stronger aroma [3,14,15]. In the Malaysian gaharu market, the ABC Agarwood Grading System [11,16], which is highly based on physical characteristics, still remains as the most common method of grading gaharu. The system is widely used by traders, despite various efforts to develop a more viable and scientific method of grading [17–20]. However, relying on physical properties as a means of grading has many drawbacks since it is highly dependent on individual human perceptions, possibly resulting in bias and poor reproducibility. The subjective nature of the existing grading systems and lack of standard method of grading have contributed to a high incidence of adulteration and substitution in the gaharu trade [4,10]. Thus, an analysis based on the inherent chemical characteristics of the different types or class of gaharu would provide a more accurate method of their quality assessment. High grade gaharu samples are reported to have high contents of sesquiterpenes [12,14,21–23]. Other studies also showed that high grade gaharu contained high levels of 10-epi- $\gamma$ -eudesmol, aromadendrane,  $\beta$ -agarofuran,  $\alpha$ -agarofuran,  $\gamma$ -eudesmol, epoxybulnesene and  $\alpha$ -guaiene [24–27]. Meanwhile, Hung et al. [28] found that  $\alpha$ -copaene, *trans*-caryophyllene, and  $\delta$ -guaiene were present in the most expensive agarwood powders/extracts. Although these studies have provided useful insights into the chemical profile of gaharu, more information is needed in order to make meaningful correlations between chemical constituents and the quality of gaharu.

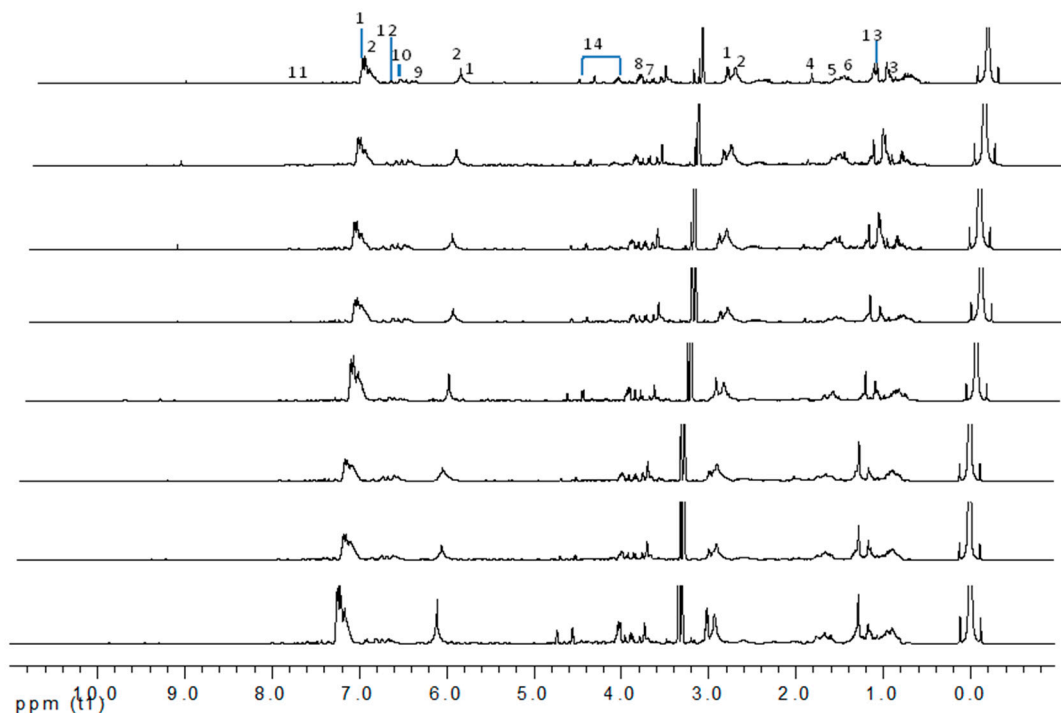
Metabolomics is an approach used to study the global profile of chemical constituents of an organism. It combines the use of analytical measurements (e.g., FTIR,  $^1\text{H-NMR}$ , GC-MS, LC-MS) and multivariate data analysis to classify and identify metabolites in biological samples [29–32]. Previously, metabolomics has been successfully used to classify different grades of agarwood powder [28,33,34] and oils [34–36]. The analytical tools used in these studies were mainly GC-MS [33–35], GC-MS coupled with solid-phase microextraction (SPME) [22,24,25,27,28] and two-dimensional GC coupled to accurate mass time-of-flight mass spectrometry (TOFMS) [36]. Although GC-MS can identify more metabolites in comparison to  $^1\text{H-NMR}$  spectroscopy, the identification of sesquiterpenoids (major component in agarwood essential oil) in gaharu samples using GC-MS alone may lead to errors [14]. It is important that these results be supported and substantiated with other complementary data using other spectroscopic techniques. Although less sensitive,  $^1\text{H-NMR}$  spectroscopy is fast and highly reproducible as well as requiring simple and non-destructive sample preparation. The technique can measure a wide range of metabolites [37,38] and is a method of choice for many metabolomics studies [30]. To date, the application of  $^1\text{H-NMR}$ -based metabolomics to differentiate various grades of gaharu based on their metabolite profiles has not been attempted. It is the objective of the present study to obtain a better insight into the chemical constituents of the various grades of gaharu. The study aims to ascertain if the different grades can be differentiated based on their metabolite profiles and whether the chemical information can be used as a means to grade the quality of gaharu.

## 2. Results and Discussion

### 2.1. Identification of Gaharu Metabolites

Figure 1 shows representative  $^1\text{H-NMR}$  spectra of the methanol extracts of gaharu samples obtained from Malaysia. The samples were of varying qualities and were grouped according to their

selling price (Supplementary Table S1). Visual inspection of the spectra indicated that the samples have very similar chemical profiles, with the expected differences in the concentration of the individual constituents. The main chemical shifts identified from the  $^1\text{H-NMR}$  spectra were in the regions of aromatics ( $\delta$  6.00–7.50), sugars and glycosides ( $\delta$  2.50–5.00), and fatty acids/aliphatics ( $\delta$  0.50–2.00). The complete  $^1\text{H-NMR}$  assignments for the identified metabolites are tabulated in Table 1.



**Figure 1.**  $^1\text{H-NMR}$  spectra of eight groups (A-H) of *A. malaccensis* gaharu samples of different grades. Metabolites labeled with numbers were tentatively identified. The spectra are arranged according to the selling price of the gaharu samples, from grades A (the most expensive) through B, C, D, E, F, G to H (the cheapest), respectively.

**Table 1.** Identified metabolites in the  $^1\text{H-NMR}$  spectra of (*A. malaccensis*) gaharu samples.

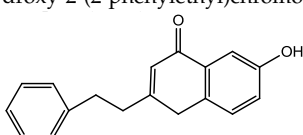
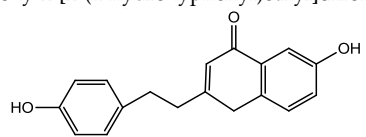
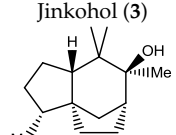
Tentative Compound	Chemical Shifts
6-Hydroxy-2-(2-phenylethyl)chromone (1) 	8.09 (d, $J = 2.5$ Hz, H-5); 7.26–7.19 (m, H-2'-H-6'); 7.14 (d, $J = 8.5$ Hz, H-8); 6.10 (s, H-3); 3.02–2.99 (m, H-7'); 2.92–2.90 (m, H-8')
6-Hydroxy-2-[2-(4-hydroxyphenyl)ethyl]chromone (2) 	7.98 (d, $J = 2.5$ Hz, H-5); 7.21 (d, $J = 8.0$ Hz, H-2'); 7.17 (d, $J = 8.0$ Hz, H-6'); 6.73 (dd, $J = 8.0, 2.5$ Hz, H-7); 6.11 (s, H-3); 2.95 (m, H-7'); 2.87 (m, H-8')
Jinkohol (3) 	2.01 (dd, $J = 4.9, 4.4$ Hz, H-8); 1.86 (m, H-3); 1.80 (m, H-2); 1.69 (dd, $J = 9.8$ Hz, H-5); 1.56 (ddd, $J = 10.6, 1.5$ Hz, H-11); 1.38 (dd, $J = 10.6, 4.4$ Hz, H-11); 0.90 (s, 6-Me); 0.84 (d, $J = 6.5$ Hz, 2-Me)

Table 1. Cont.

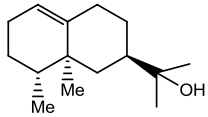
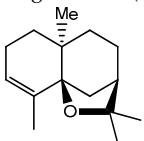
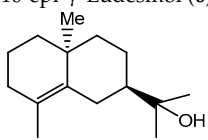
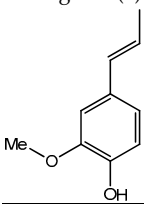
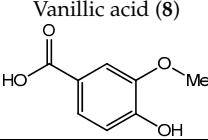
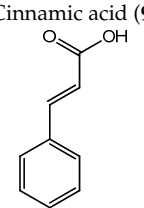
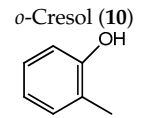
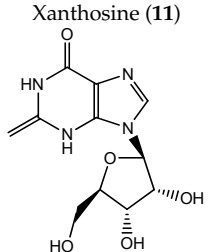
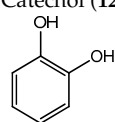
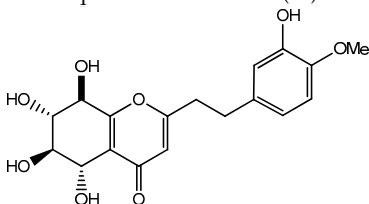
Tentative Compound	Chemical Shifts
<p>usunol (4)</p> 	<p>5.32 (ddd, <math>J = 5.7, 2.2</math> Hz, H-1);            2.27 (dddd, <math>J = 13.8, 12.4, 3.3</math> Hz, H-9);            1.62 (dddd, <math>J = 12.4, 3.3</math> Hz, H-7); 1.41 (m, H-4)</p>
<p><math>\alpha</math>-Agarofuran (5)</p> 	<p>5.59 (s, H-3); 2.22 (dd, <math>J = 12.5, 4.0</math> Hz, H-9); 1.72 (s, H-12);            1.23 (s, H-14); 0.91 (s, H-13)</p>
<p>10-epi-<math>\gamma</math>-Eudesmol (6)</p> 	<p>2.12 (d, <math>J = 15.8</math> Hz, H-3); 1.68 (s, H-12); 1.19 (s, H-13);            1.09 (s, H-11)</p>
<p>Isoeugenol (7)</p> 	<p>7.09 (dd, <math>J = 1.9, 0.5</math> Hz, H-5); 3.79 (s, H-10);            1.55 (d, <math>J = 7.3</math> Hz, H-9); 6.32 (d, <math>J = 16.9</math> Hz, H-7);            6.29 (dq, <math>J = 16.9, 6.9</math> Hz, H-8); 7.40 (dd, <math>J = 8.6, 1.9</math> Hz, H-3)</p>
<p>Vanillic acid (8)</p> 	<p>3.94 (s, H-8); 6.92 (d, <math>J = 8.2</math> Hz, H-3);            7.43 (dd, <math>J = 8.2, 1.7</math> Hz, H-4); 7.52 (d, <math>J = 1.7</math> Hz, H-6)</p>
<p>Cinnamic acid (9)</p> 	<p>7.60 (dd, <math>J = 7.9, 1.1</math> Hz, H-6); 7.45 (m, H-5);            7.40 (d, trans, <math>J = 16.0</math> Hz, H-7);            6.54 (d, trans, <math>J = 16.0</math> Hz, H-8)</p>
<p><i>o</i>-Cresol (10)</p> 	<p>2.29 (s, H-8); 6.82 (m, H-4, H-6); 7.14 (m, H-5); 7.20 (m, H-3)</p>
<p>Xanthosine (11)</p> 	<p>7.88 (s, H-7); 5.85 (d, <math>J = 6.4</math> Hz, H-2);            4.69 (t, <math>J = 5.7</math> Hz, H-3); 4.25 (q, <math>J = 2.7</math> Hz, H-5);            3.89 (m, H-17)</p>

Table 1. Cont.

Tentative Compound	Chemical Shifts
Catechol (12) 	6.87 (m, H-4, H-5); 6.94 (m, H-3, H-6)
Fatty acid: (13)	1.28 (m)
Aquilarone derivatives (14) 	4.72 (d, $J = 2.7$ Hz, H-5); 4.55 (d, $J = 7.3$ Hz, H-8); 4.29 (m, H-6); 3.99 (dd, $J = 6.2, 2.4$ Hz, H-7); 2.70–2.80 (m, 2H)

The representative signals of two phenylethyl chromones i.e., 6-hydroxy-2-(2-phenylethyl)-chromone (6HC, **1**) and 6-hydroxy-2-[2-(4-hydroxyphenyl)ethyl]chromone (6DHC, **2**) were observed in the  $^1\text{H-NMR}$  spectra of the gaharu extracts. Compound **1** was identified based on the  $^1\text{H-NMR}$  signals for an ABX coupling system at  $\delta$  7.14 (1H, d,  $J = 8.5$  Hz, H-8), 6.83 (1H, dd,  $J = 8.5, 2.5$  Hz, H-7), and 8.09 (1H, d,  $J = 2.5$  Hz, H-5) and a singlet proton at  $\delta$  6.10 (1H, s, H-3), characteristic of the chromone benzopyran moiety. In addition, signals for the two sets of methylene protons for the phenylethyl moiety were also observed at  $\delta$  3.02 (2H, m) and 2.92 (2H, m). The identification of compound **1** was further supported by 2D-NMR experiments ( $J$ -resolved, COSY, and HMBC spectra), LC-MS/MS analysis (data shown in Supplementary Table S2) and comparison with literature values [39]. Compound **2** was similarly assigned based on signals observed at 6.62 (1H, d,  $J = 8.0$  Hz, H-8), 6.73 (1H, dd,  $J = 8.0, 2.5$  Hz, H-7) and  $\delta$  7.98 (1H, d,  $J = 2.5$  Hz, H-5) for the ABX coupled protons, singlet proton at  $\delta$  6.11 (1H, s, H-3) and the two sets of methylene protons at  $\delta$  2.95 (2H, m) and 2.89 (2H, m). The presence of an  $\text{A}_2\text{B}_2$  coupled system at  $\delta$  7.21 (2H, d,  $J = 8.0$  Hz) and 7.17 (2H, d,  $J = 8.0$  Hz) supported the *para*-hydroxyphenyl ring of the phenylethylchromone structure.

Besides the signals for the phenylethylchromones, signals attributable to 5,6,7,8-tetrahydro-chromone (**14**) were also observed at  $\delta$  4.72 (d), 4.55 (d), 4.29 (m), 3.99 (dd) and two sets of methylene protons at  $\delta$  2.70–2.80 (m). Comparison with literature values [40] and a molecular ion peak observed at  $m/z$  363  $[\text{M} - \text{H}]^-$  in the LC-MS spectrum (Supplementary Table S2), supported the tentative identification of the compound as an aquilarone derivative. Minor constituents from the phenolic class of compounds were also identified in the gaharu samples, tentatively assigned as vanillic acid (**8**), cinnamic acid (**9**), *o*-cresol (**10**), xanthosine (**11**) and catechol (**12**) (Table 1).

Sesquiterpenes have been reported to be a major class of compounds present in the resin of *A. malaccensis* [13]. Yoneda et al. [41] also reported the presence of jinkohol, kusunol,  $\alpha$ -agarofuran, 10-epi- $\gamma$ -eudesmol, and agarospirol in the agarwood oil obtained from *A. malaccensis*. After detailed analysis and a comparison with literature data and online databases, the sesquiterpenoids jinkohol (**3**) and kusunol (**4**) [42], agarofuran (**5**) and epi-eudesmol (**6**) [43], and isoeugenol (**7**) (HMDB <http://www.hmdb.ca/>) were also identified in the gaharu extracts in the present study. Further examination of the  $^1\text{H-NMR}$  spectra also showed the presence of very small amounts of aldehydic compounds ( $\delta$  9.32 ppm) as can be seen in Figure 1. However, due to technical limitations of the present study, the structures of these aldehydes could not be identified.

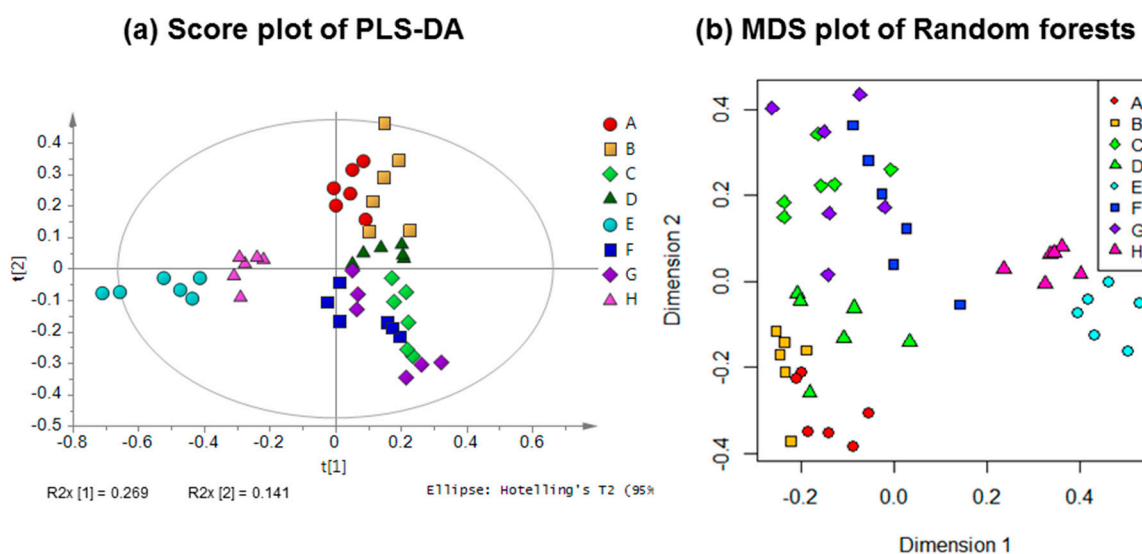
## 2.2. Discriminative Analysis of Gaharu Samples

The processed  $^1\text{H-NMR}$  data was initially subjected to principal component analysis (PCA) in order to see the differences between the eight groups of gaharu. However, PCA did not show any

clear clustering or differences among the gaharu samples. This could be due to the high variability of the different gaharu samples. Partial least squares-discriminant analysis (PLS-DA) was then used to model the relationships between the eight groups of gaharu samples. A permutation test was applied to evaluate the reliability of the model (Supplementary Figure S1). Overall, the PLS-DA model was found to be a reliable and good model for the classification. The model did not show over-fitting, based on the  $Y$ -axis intercept values of  $R^2 = 0.07$  and  $Q^2 = -0.14$ , and the fact that the  $R^2$  line was far from being horizontal.

The PLS-DA score plot showed that the eight groups of gaharu samples were differentiated into three distinct clusters (Figure 2a). The samples H and E were well separated from the rest of the samples (A, B, C, D, F, G), and were clustered together on the negative side of PLS component 1. Meanwhile, samples A, B, C, D, F, G could be further differentiated on the basis of PLS component 2 scores, where samples A, B and D were clustered on the positive side, while samples C, F and G were clustered on the negative side. The corresponding loading plot (Figure 3) showed the discriminant metabolites responsible for the separation of the three clusters in the score plot. From the loadings, it could be deduced that samples H and E contained higher levels of 6-hydroxy-2-(2-phenylethyl)-chromone (6HC, 1), 6-hydroxy-2-[2-(4-hydroxyphenyl)ethyl]chromone (6DHC, 2), cinnamic acid (9) and aquilarone derivatives (14). The samples A, B and D were marked by higher levels of jinkohol (3), kusunol (4), agarofuran (5), and 10-epi- $\gamma$ -eudesmol (6), whereas C, F and G were characterized by higher levels of isoeugenol (7), vanillic acid (8), xanthosine (11), catechol (12) and fatty acids (13). A blind test was carried out to evaluate the performance of the PLS-DA model (Supplementary Figure S2). A new batch of gaharu samples (test samples) belonging to low, medium and high grades were analyzed by  $^1\text{H-NMR}$  and subjected to PLS-DA together with the training set (previous NMR data of the different grades). In the PLS-DA score plot, the new gaharu samples were clustered well within the corresponding grades.

To further validate the results obtained from the PLS-DA, the Random Forests classifier was applied to the same  $^1\text{H-NMR}$  data. In contrast to PLS-DA, the application of Random Forests as a classification model [44] is relatively rare in metabolomics data analysis. Although it is available in freeware softwares such as the Random Forest package in the R software [45] and MetaboAnalyst [46], its applicability in metabolomics studies still needs to be explored. Figure 2b shows the Random Forests multi-dimensional scaling (MDS) plot of proximity matrix. The MDS plot showed the same clustering as was found in the score plot of PLS-DA. The accuracy of the models was evaluated using confusion matrices as shown in Table 2.



**Figure 2.** (a) PLS-DA score plot and (b) Random Forests multidimensional scaling (MDS) plot for *A. malaccensis* gaharu samples of different grades.

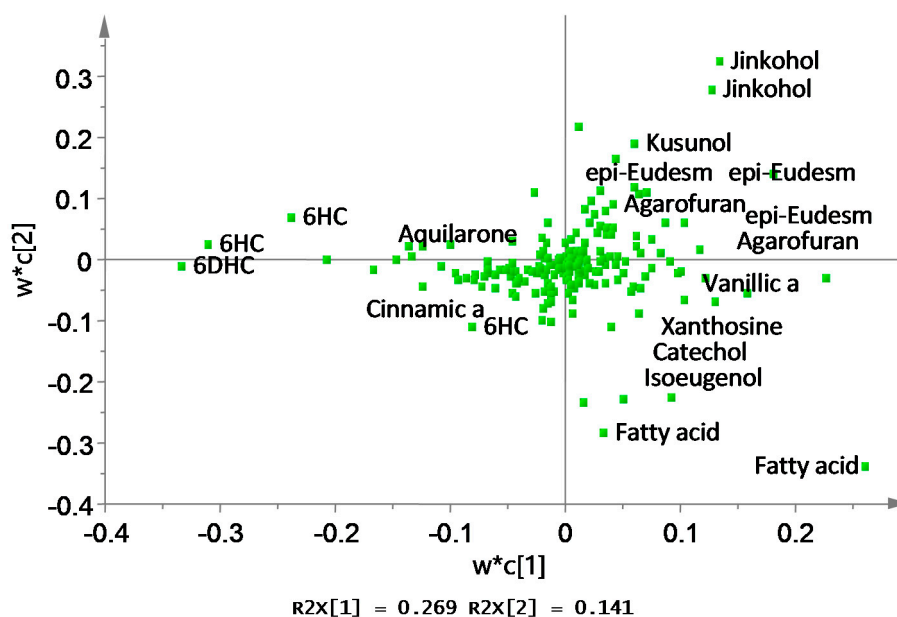


Figure 3. Loading plots corresponding to the PLS-DA score plot.

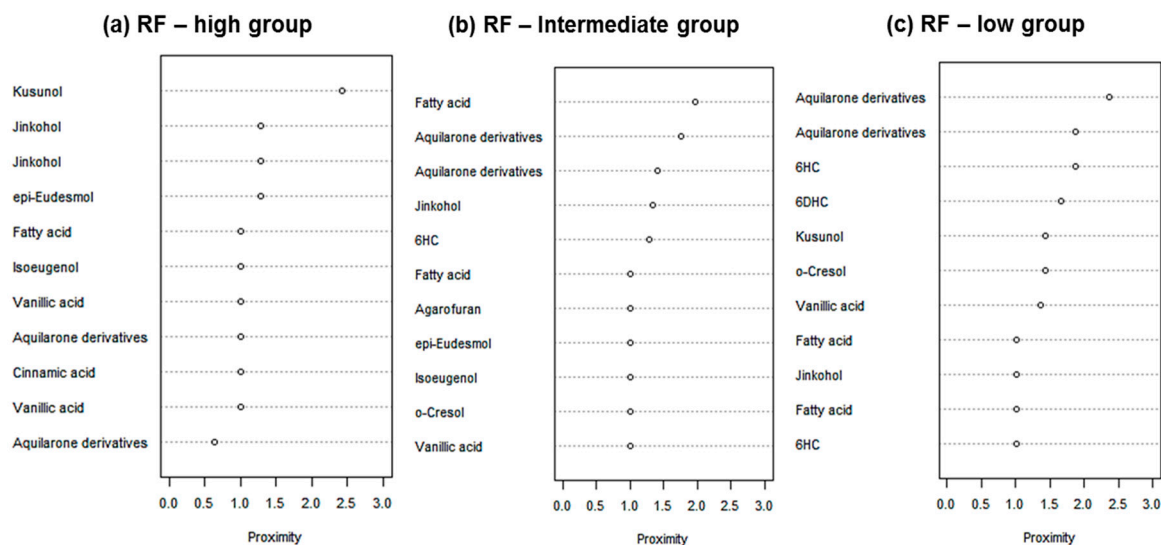
Table 2. Confusion matrix for Random Forests in the classification of eight groups of *A. malaccensis* gaharu samples of different grades.

	Random Forests Class									Producer Accuracy	
	A	B	C	D	E	F	G	H	Total	Percent Correct	Omission Error (%)
Reference class	A	4	2	0	0	0	0	0	6	67	33
	B	2	3	0	1	0	0	0	6	50	50
	C	0	0	5	0	0	0	1	6	83	17
	D	0	2	0	3	0	1	0	6	50	50
	E	0	0	0	0	6	0	0	6	100	0
	F	0	0	0	1	0	5	0	6	83	17
	G	0	0	1	1	0	1	3	6	50	50
	H	0	0	0	0	0	0	0	6	100	0
	Total	6	7	6	6	6	7	4	6	48	
Users accuracy											
Percent correct	67	43	83	50	100	71	75	100		72.92	
Commission error (%)	33	57	17	50	0	29	25	0			
Agreement											
By chance	4	3	5	3	6	5	3	6	35		
	0.75	0.88	0.75	0.75	0.75	0.875	0.5	0.75	6.00		
Kappa 0.69											

The percentage of overall agreement (given by  $100\sum X_{ii}/N$ ) and Kappa coefficient or  $\kappa$  (given by  $[(\sum X_{ii}) - (\sum X_i + X + i)/N]/[N - (\sum X_i + X + i)/N]$ ) values were calculated to be 72.9% and 0.69, respectively.

### 2.3. Identification of Discriminating Metabolites

The discriminating metabolites were identified from the chemical shifts in the PLS-DA loading plot (Figure 3), and from the VIP (variable importance) values in the Random Forests for each cluster of gaharu samples (Figure 4). In the latter, metabolites having high value of VIP are deemed to have high contribution to the clustering.



**Figure 4.** Variable importance (VIP) values for (a) high; (b) intermediate and (c) low grade clusters based on Random Forests.

For the PLS-DA model, the high grade cluster (groups A, B and D) was characterized by higher levels of jinkohol (3), kusunol (4), and 10-epi- $\gamma$ -eudesmol (6), whereas the intermediate grade cluster (groups C, F and G) contained higher levels of isoeugenol (7), vanillic acid (8), xanthosine (11), catechol (12) and fatty acid (13). The low grade cluster (groups H and E) was distinguished from the other groups by having higher levels of aquilarone derivatives (14) and phenylethylchromones 1 and 2. The Random Forests analysis basically resulted in the identification of the same discriminant metabolites in the established clusters as in the PLS-DA model.

Identification of the discriminant metabolites from the PLS-DA and Random Forests models were confirmed by analysis of the variable importance (VIP) values for all clusters as shown in the Supplementary Figure S3a,b, respectively. Furthermore, the relative quantification of the discriminant metabolites in the three clusters (high, intermediate and low grade clusters) was carried out using Tukey posthoc analysis, based on the average peak area of the corresponding  $^1\text{H-NMR}$  signals (Supplementary Figure S4).

The sesquiterpenoids jinkohol (3), kusunol (4),  $\alpha$ -agarofuran (5) and 10-epi- $\gamma$ -eudesmol (6) are well known volatile constituents that are associated with the fragrance of gaharu [14,26]. In the present study, the high grade gaharu samples were indeed characterized by high levels of jinkohol (3), kusunol (4) and 10-epi- $\gamma$ -eudesmol (6), and thus, were in agreement with previous findings [14,15,21]. The higher grade gaharu samples were also observably darker in color in comparison to the low grade gaharu samples which clearly reflected the higher contents of the resinous constituents. According to the literature, non-infected *A. malaccensis* wood is brighter in colour and almost odourless, whereas the infected wood is heavier and dark brown to black in colour [14,15]. On the other hand, chromones have been reported to be the metabolites responsible for the warm, sweet, balsamic and long-lasting odor when gaharu wood is burned or heated [14]. Therefore, the lower grade gaharu samples which were richer in these chromones are more suitable for use as incense. Jinkohol (3), the proposed chemical marker for high grade gaharu, also has a distinctive and extremely strong woody smell which contributed to the suitability of the high grade gaharu extract/resin as a perfume ingredient. Several studies have also reported high levels of agarofuran (5) in gaharu samples of high grade [24–27]. In the present study, however, both the PLS-DA and Random Forests models showed that the levels of the constituent in the different groups were not significantly different.

Interestingly, although PLS-DA and Random Forests are based on different concepts, both models yielded similar results in terms of class plots (score and MDS plots) and the chemical constituents for the new group clusters, as well as the VIP values. We noted that since Random Forests was developed



based on a random subset in both variables and individual data, repeating the Random Forests analysis may not always yield exactly the same results, albeit it was similar when we rerun the analysis. However, the Random Forests result may still explain about uncertainties in the biological system.

<sup>1</sup>H-NMR-based metabolomics was shown to be effective in classifying *A. malaccensis* gaharu samples of varying quality, as sampled from the market place. Using this approach, it was also possible to propose a new group of classification based on their chemical constituents. Random Forests and PLS-DA were found to be reliable chemometric methods to assess the differences and similarities among the different gaharu samples. Using the two methods, the gaharu samples analysed in the present study, could be reclassified into three groups based on their chemical characteristics. From the identified gaharu constituents, eight metabolites could be proposed as differentiating chemical constituents between the high (jinkohol (3), kusunol (4), and 10-epi- $\gamma$ -eudesmol (6)), intermediate (fatty acid (13) and vanillic acid (8)) and low (phenylethyl chromones (1 and 2) and aquilarone derivatives (14)) grades of *A. malaccensis* gaharu. However, the results were based on relative quantification of the metabolites. Further confirmatory analyses are required to determine the absolute quantification and identification of these chemical constituents.

### 3. Experimental Section

#### 3.1. Samples and Chemicals

Samples of *A. malaccensis* gaharu in the form of wood chips were purchased from an experienced collector and trader. The samples consisted of varying quality of gaharu samples (Supplementary Table S1), graded and valued (RM/kg) by the collector according to the ABC Agarwood Grading System. The gaharu samples were inspected for authenticity and the grading was double checked and confirmed by an in-house expert. For the purpose of the present study, the samples were grouped according to their selling price and labeled A to H. Each group of gaharu samples consisted of six replicates. Lab grade methanol (redistilled prior to use), methanol-d<sub>4</sub> (CD<sub>3</sub>OD, 99.8%), KH<sub>2</sub>PO<sub>4</sub>, sodium deuterium oxide (NaOD) and deuterium oxide (D<sub>2</sub>O) (99.8%) were purchased from Merck (Darmstadt, Germany).

#### 3.2. <sup>1</sup>H-NMR Sample Preparation

Each sample of gaharu wood chips was pulverized into fine powder using mortar pestle and grinder. To ensure that wood particles are of uniform size, the ground samples were sieved using sieve shaker (Retsch) to collect  $\leq 140$   $\mu$ m particles. The sieved samples (1 g) were then extracted with 10 mL lab grade methanol (sample:solvent ratio of 1:6 (*w/v*)) by sonication (Ultrasonic LC 60H, Elma, Singen, Germany), for 1 h at ambient temperature. The extracts were filtered and the collected filtrates taken to dryness under vacuum (MiVac, Genevac, Ipswich, UK), followed by lyophilization. All extracts were kept at  $-80$  °C until further analysis.

Samples for NMR measurements were prepared by resuspending 20 mg of each sample extract in 700  $\mu$ L CD<sub>3</sub>OD to which 0.5% tetramethylsilane (TMS) had been added as reference standard. The sample-solvent mixtures contained in 1.5 mL Eppendorf tubes were then sonicated for 15 min at room temperature to facilitate resolubilization of the extract in the NMR solvent. After centrifuging for a further 15 min at a speed of 13,000 rpm, the clear supernatant solutions of each sample (700  $\mu$ L) were transferred into 5 mm NMR tubes for NMR data acquisition.

#### 3.3. <sup>1</sup>H-NMR Data Acquisition and Data Preprocessing

In total, 48 samples were analyzed (8 groups  $\times$  6 replicates). <sup>1</sup>H-NMR spectra were recorded at 25 °C on a Unity Inova 500 MHz NMR spectrometer (Varian, Palo Alto, CA, USA) using 128 scans over a proton frequency range of 15 ppm. The PRESAT program was used to suppress undesirable signals caused by residual water. *J*-resolved and 2D NMR analysis were also performed for structural elucidation of chemical constituents present in the extract. The generated <sup>1</sup>H-NMR FIDs were

manually processed, phase corrected and referenced to the internal standard, TMS ( $\delta$  0.00 ppm). Baseline correction was applied to all spectra before converting to ASCII file and binned using Chenomx software after which the processed raw data was saved in an Excel spreadsheet (Microsoft, Washington, DC, USA). The raw data were binned into individual widths of  $\delta$  0.04 starting from chemical shift region  $\delta$  0.5 to  $\delta$  10.00 ppm. Water and solvent peaks in the region  $\delta$  4.70–4.96 and  $\delta$  3.28–3.33 ppm, respectively, were excluded in the multivariate analysis.

### 3.4. Metabolite Assignment

The metabolites were identified by comparing the characteristic peak signals in the  $^1\text{H-NMR}$  spectra of samples with published data and existing literature databases ([www.hmdb.ca](http://www.hmdb.ca); Chenomx NMR Suit Ver.7.1, company Edmonton, AB, Canada). Identification of compounds was also supported by 2D NMR and LC-MS analysis. Table 1 and Supplementary Table S2 show the identified metabolites in the gaharu samples.

### 3.5. Development of PLS-DA and Random Forests Models

PLS-DA is a supervised classification technique. This technique optimizes separation between different groups of samples and develops link between two data matrices X (i.e., data, binned spectra) and Y (i.e., groups, class membership etc.) by maximizing the covariance between these X and Y matrices and finding a linear subspace of the explanatory variables [30,47]. The Y-variables are represented with a special binary 'dummy' [30,48]. Data were Pareto-scaled and PLS-DA was carried out using SIMCA-P software (version 12.0, Umetrics, Umea, Sweden). Random Forests is a tree-based ensemble method where two subsets are operated in independent variables at each node and in individual observation data by bootstrapping technique. Random Forests can be used for unsupervised and supervised classification as well as regression. In the current study, random Forests was used for a supervised classification, performed using the Random Forest R package [45].

### 3.6. Statistical Analysis

The relative quantification of chemical constituents was based on the mean binned peak height of the related  $^1\text{H-NMR}$  signals. ANOVA and Tukey's honest multiple comparison tests were conducted to evaluate the significant difference ( $p < 0.05$ ) between the differentiating metabolites. The statistical analysis was performed using SPSS version 16.0 (SPSS Inc., Chicago, IL, USA) software.

## 4. Conclusions

The study showed that, using  $^1\text{H-NMR}$ -based metabolomics, it is possible to discriminate between *A. malaccensis* gaharu samples of different quality. The results provide an insight into the chemical characteristics of gaharu, categorizing the samples into high, intermediate and low grades. Although more extensive work needs to be done, such as applying the analysis to other gaharu-producing species, the information obtained in this study is of importance and contributes towards development of a 'chemical assay' that could make the process of grading gaharu or agarwood samples more accurate, practical and efficient.

**Supplementary Materials:** Supplementary Materials are available online.

**Acknowledgments:** This project is funded by Universiti Putra Malaysia grant (RUGS-2) No.: 01-02-10-0917RU. We thank Mohd Zin Jusoh from the Faculty of Forestry, UPM, for expert help in confirming the authenticity and grades of the gaharu samples.

**Author Contributions:** K.S and A.K conceived and designed the experiments. S.N.I. performed the experiment. S.N.I., M.M. and M.T.A analysed the data. S.N.I., M.M., M.T.A and K.S wrote the paper. F.A., N.A.M.A. and I.S.I. provided intellectual inputs.

**Conflicts of Interest:** The authors have no competing interest to declare.

## References

1. The Plant List. Version 1.1. 2013. Available online: <http://www.theplantlist.org/> (accessed on 2 February 2017).
2. Mabberley, D.J. *Mabberley's Plant-Book: A Portable Dictionary of Plants, Their Classifications, and Uses*; Cambridge University Press: Cambridge, UK, 2008.
3. Chen, D.; Xu, Z.; Chai, X.; Zeng, K.; Jia, Y.; Bi, D.; Ma, Z.; Tu, P. Nine 2-(2-phenylethyl)chromone derivatives from the resinous wood of *Aquilaria sinensis* and their inhibition of LPS-induced NO production in RAW 264.7 cells. *Eur. J. Org. Chem.* **2012**, *27*, 5389–5397. [[CrossRef](#)]
4. Barden, A.; Anak, N.A.; Mulliken, T.; Song, M. *Heart of the Matter: Agarwood Use and Trade and CITES Implementation for Aquilaria Malaccensis*; TRAFFIC International: Cambridge, UK, 2000.
5. Bhore, S.J.; Preveena, J.; Kandasamy, K.I. Isolation and identification of bacterial endophytes from pharmaceutical agarwood-producing *Aquilaria* species. *Pharmacogn. Res.* **2013**, *5*, 134–137. [[CrossRef](#)] [[PubMed](#)]
6. Xu, Y.; Zhang, Z.; Wang, M.; Wei, J.; Chen, H.; Gao, Z.; Sui, C.; Luo, H.; Zhang, X.; Yang, Y.; et al. Identification of genes related to agarwood formation: Transcriptome analysis of healthy and wounded tissues of *Aquilaria sinensis*. *BMC Genom.* **2013**, *14*, 227. [[CrossRef](#)] [[PubMed](#)]
7. Burkill, I.H. *A Dictionary of the Economic Products of the Malay Peninsula*; Ministry of Agriculture and Cooperatives: Kuala Lumpur, Malaysia, 1935.
8. Kim, Y.C.; Lee, E.H.; Lee, Y.M.; Kim, H.K.; Song, B.K.; Lee, E.J.; Kim, H.M. The effect of the aqueous extract of *Aquilaria agallocha* stems on the immediate hypersensitivity reactions. *J. Ethnopharmacol.* **1997**, *58*, 31–38. [[CrossRef](#)]
9. Ng, L.T.; Chang, Y.S.; Kadir, A.A. A review on agar (gaharu) producing *Aquilaria* species. *J. Trop. For. Prod.* **1997**, *2*, 272–285.
10. Antonopoulou, M.; Compton, J.; Perry, L.S.; Al-Mubarak, R. *The Trade and Use of Agarwood (Oudh) in the United Arab Emirates*; TRAFFIC Southeast Asia: Selangor, Malaysia, 2010.
11. Lim, T.W.; Anak, N.A. *Wood for Trees: A Review of the Agarwood (Gaharu) Trade in Malaysia*; TRAFFIC Southeast Asia: Petaling Jaya, Malaysia, 2010.
12. Ismail, N.; Ali, N.A.M.; Jamil, M.; Rahiman, M.H.F.; Tajuddin, S.N.; Tai, M.N. A Review study of agarwood oil and its quality analysis. *J. Teknol.* **2014**, *1*, 37–42. [[CrossRef](#)]
13. Ismail, N.; Rahiman, M.H.F.; Taib, M.N.; Ibrahim, M.; Zareen, S.; Tajuddin, S. A review on agarwood and its quality determination. In Proceedings of the IEEE 6th Control and System Graduate Research Colloquium (ICSGRC), Shah Alam, Malaysia, 10–11 August 2015; pp. 103–108.
14. Naef, R. The volatile and semi-volatile constituents of agarwood, the infected heartwood of *Aquilaria* species: A review. *Flavour Fragr. J.* **2011**, *26*, 73–87. [[CrossRef](#)]
15. Azah, M.A.N.; Husni, S.S.; Mailina, J.; Sahrim, L.; Majid, J.A.; Faridz, Z.M. Classification of agarwood (gaharu) by resin content. *J. Trop. For. Sci.* **2013**, *25*, 213–219.
16. Lee, S.Y.; Ng, W.L.; Mahat, M.N.; Nazre, M.; Mohamed, R. DNA barcoding of the endangered *Aquilaria* (Thymelaeaceae) and its application in species authentication of agarwood products traded in the market. *PLoS ONE* **2016**, *11*, e0154631. [[CrossRef](#)] [[PubMed](#)]
17. Hidayat, W.; Shakaff, A.Y.M.; Ahmad, M.N.; Adom, A.H. Classification of agarwood oil using an electronic nose. *Sensors* **2010**, *10*, 4675–4685. [[CrossRef](#)] [[PubMed](#)]
18. Najib, M.S.; Ali, N.A.M.; Arip, M.N.M.; Jalil, A.M.; Taib, M.N. Classification of agarwood region using ANN. In Proceedings of the 2010 IEEE Control and System Graduate Research Colloquium (ICSGRC 2010), Shah Alam, Malaysia, 22 June 2010; pp. 7–13.
19. Najib, M.S.; Ahmad, M.U.; Funk, P.; Taib, M.N.; Ali, N.A.M. Agarwood classification: A case-based reasoning approach based on E-nose. In Proceedings of the 2012 IEEE 8th International Colloquium on Signal Processing and Its Applications, Melaka, Malaysia, 23–25 March 2012; pp. 120–126.
20. Ismail, N.; Rahiman, M.H.F.; Taib, M.N.; Ali, N.A.M.; Jamil, M.; Tajuddin, S.N. Application of ANN in agarwood oil grade classification. In Proceedings of the 2014 IEEE 10th International Colloquium on Signal Processing and Its Applications, Kuala Lumpur, Malaysia, 7–9 March 2014; pp. 216–220.
21. Ishihara, M.; Tsuneya, T.; Uneyama, K. Components of the agarwood smoke on heating. *J. Essent. Oil Res.* **1993**, *5*, 419–423. [[CrossRef](#)]

22. Pripdeevech, P.; Khummueng, W.; Park, S.K. Identification of odor-active components of agarwood essential oils from Thailand by solid phase microextraction-GC/MS and GC-O. *J. Essent. Oil Res.* **2011**, *23*, 46–53. [[CrossRef](#)]
23. Zhang, X.L.; Liu, Y.Y.; Wei, J.H.; Yang, Y.; Zhang, Z.; Huang, J.Q.; Chen, H.Q.; Liu, Y.J. Production of high-quality agarwood in *Aquilaria sinensis* trees via whole-tree agarwood-induction technology. *Chin. Chem. Lett.* **2012**, *23*, 727–730. [[CrossRef](#)]
24. Ismail, N.; Azah, M.A.N.; Jamil, M.; Rahiman, M.H.F.; Tajuddin, S.N.; Taib, M.N. Analysis of high quality agarwood oil chemical compounds by means of SPME/GC-MS and Z-score technique. *Malays. J. Anal. Sci.* **2013**, *17*, 403–413.
25. Ismail, N.; Rahiman, M.H.F.; Taib, M.N.; Ali, N.A.M.; Jamil, M.; Tajuddin, S.N. Analysis of chemical compounds of agarwood oil based on headspace-solid phase microextraction combined with gas chromatography mass-spectrometry. In Proceedings of the IEEE 9th International Colloquium on Signal Processing and Its Applications, Kuala Lumpur, Malaysia, 8–10 March 2013; pp. 215–218.
26. Azah, M.A.N.; Ismail, N.; Mailina, J.; Taib, M.N.; Rahiman, M.H.F.; Hafizi, Z.M. Chemometric study of selected agarwood oils by gas chromatography-mass spectrometry. *J. Trop. For. Sci.* **2014**, *26*, 382–388.
27. Ali, N.A.M.; Ismail, N.; Jamil, M.; Aziz, A.; Lias, S.; Rahiman, M.H.F.; Taib, M.N. Identification of odor components of agarwood. *J. Teknol.* **2015**, *77*, 51–55.
28. Hung, C.H.; Lee, C.Y.; Yang, C.L.; Lee, M.R. Classification and differentiation of agarwoods by using non-targeted HS-SPME-GC/MS and multivariate analysis. *Anal. Methods* **2014**, *6*, 7449–7456. [[CrossRef](#)]
29. Gromski, P.S.; Correa, E.; Vaughan, A.A.; Wedge, D.C.; Turner, M.L.; Goodacre, R. A comparison of different chemometrics approaches for the robust classification of electronic nose data. *Anal. Bioanal. Chem.* **2014**, *406*, 7581–7590. [[CrossRef](#)] [[PubMed](#)]
30. Gromski, P.S.; Muhamadali, H.; Ellis, D.I.; Xu, Y.; Correa, E.; Turner, M.L.; Goodacre, R. A tutorial review: Metabolomics and partial least squares-discriminant analysis—A marriage of convenience or a shotgun wedding. *Anal. Chim. Acta* **2015**, *879*, 10–23. [[CrossRef](#)] [[PubMed](#)]
31. Gromski, P.S.; Xu, Y.; Correa, E.; Ellis, D.I.; Turner, M.L.; Goodacre, R. A comparative investigation of modern feature selection and classification approaches for the analysis of mass spectrometry data. *Anal. Chim. Acta* **2014**, *829*, 1–8. [[CrossRef](#)] [[PubMed](#)]
32. Gromski, P.S.; Xu, Y.; Hollywood, K.A.; Turner, M.L.; Goodacre, R. The influence of scaling metabolomics data on model classification accuracy. *Metabolomics* **2014**, *11*, 684–695. [[CrossRef](#)]
33. Gao, X.; Xie, M.; Liu, S.; Guo, X.; Chen, X.; Zhong, Z.; Wang, L.; Zhang, W. Chromatographic fingerprint analysis of metabolites in natural and artificial agarwood using gas chromatography-mass spectrometry combined with chemometric methods. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2014**, *967*, 264–273. [[CrossRef](#)] [[PubMed](#)]
34. Sen, S.; Talukdar, N.C.; Khan, M. A simple metabolite profiling approach reveals critical biomolecular linkages in fragrant agarwood oil production from *Aquilaria malaccensis*—A traditional agro-based industry in North East India. *Curr. Sci.* **2015**, *108*, 63–71.
35. Jayachandran, K.; Sekar, I.; Parthiban, K.T.; Amirtham, D.; Suresh, K.K. Analysis of different grades of agarwood (*Aquilaria malaccensis* Lamk.) oil through GC-MS. *Indian J. Nat. Prod. Resour.* **2014**, *5*, 44–47.
36. Wong, Y.F.; Chin, S.T.; Perlmutter, P.; Marriott, P.J. Evaluation of comprehensive two-dimensional gas chromatography with accurate mass time-of-flight mass spectrometry for the metabolic profiling of plant-fungus interaction in *Aquilaria malaccensis*. *J. Chromatogr. A* **2015**, *1387*, 104–115. [[CrossRef](#)] [[PubMed](#)]
37. Kooy, F.V.D.; Maltese, F.; Choi, Y.H.; Kim, H.K.; Verpoorte, R. Quality control of herbal material and phytopharmaceuticals with MS- and NMR-based metabolic fingerprinting. *Planta Med.* **2009**, *75*, 763–775. [[CrossRef](#)] [[PubMed](#)]
38. Kim, H.K.; Choi, Y.H.; Verpoorte, R. NMR-based metabolomic analysis of plants. *Nat. Protoc.* **2010**, *5*, 536–549. [[CrossRef](#)] [[PubMed](#)]
39. Konishi, T.; Konoshima, T.; Shimada, Y.; Kiyosawa, S. Six new 2-(2-phenylethyl)chromones from agarwood. *Chem. Pharm. Bull.* **2002**, *50*, 419–422. [[CrossRef](#)] [[PubMed](#)]
40. Chen, H.Q.; Wei, J.H.; Yang, J.S.; Zhang, Z.; Yang, Y.; Gao, Z.H.; Sui, C.; Gong, B. Chemical constituents of agarwood originating from the endemic genus *Aquilaria* plants. *Chem. Biodivers.* **2012**, *9*, 236–250. [[CrossRef](#)] [[PubMed](#)]

41. Yoneda, K.; Yamagata, E.; Nakanishi, T.; Nagashima, T.; Kawasaki, I.; Yoshida, T.; Mori, H.; Miura, I. Sesquiterpenoids in two different kinds of agarwood. *Phytochemistry* **1984**, *23*, 2068–2069. [CrossRef]
42. Nakanishi, T.; Yamagata, E.; Yoneda, K.; Miura, I.; Mori, H. Jinkoh-eremol and jinkohol II, two new sesquiterpene alcohols from agarwood. *J. Chem. Soc. Perkin Trans.* **1983**, *1*, 601–604. [CrossRef]
43. Nakanishi, T.; Yamagata, E.; Yoneda, K.; Nagashima, T.; Kawasaki, I.; Yoshida, T.; Mori, H.; Miura, I. Three fragrant sesquiterpenes of agarwood. *Phytochemistry* **1984**, *23*, 2066–2067. [CrossRef]
44. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
45. Liaw, A.; Wiener, M. Package ‘randomForests’. Breiman and Cutler’s Random Forests for Classification and Regression. Available online: <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf> (accessed on 25 September 2017).
46. Xia, J.; Psychogios, N.; Young, N.; Wishart, D.S. MetaboAnalyst: A web server for metabolomic data analysis and interpretation. *Nucleic Acids Res.* **2009**, *37*, 652–660. [CrossRef] [PubMed]
47. Kalivodová, A.; Hron, K.; Filzmoser, P.; Najdekr, L.; Janečková, H.; Adam, T. PLS-DA for compositional data with application to metabolomics. *J. Chemom.* **2015**, *29*, 21–28. [CrossRef]
48. Szymańska, E.; Saccenti, E.; Smilde, A.K.; Westerhuis, J.A. Double-check: Validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* **2012**, *8*, 3–16. [CrossRef] [PubMed]

**Sample Availability:** Samples of the compounds are not available from the authors.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).