



OPEN Visualization OPLS class models of GC-MS-based metabolomics data for identifying agarwood essential oil extracted by hydro-distillation

Si-Zhu Qian¹, Yong-Mei Jiang¹, Qiao-Ling Yan¹, De-Huai Wu², Wen-Xian Zhang¹✉ & Jen-Ping Chung³✉

The composition of natural substances varies with plant species and cultivation environment factors, which is also a complex problem. A total of 127 substances of agarwood essential oils (AEOs) extracted by hydro-distillation were identified by GC-MS analysis. Among the components obtained from AEOs, sesquiterpenes and small molecule aromatic substances were the main components, and there were significantly fewer chromones. The aromatic compound 4-phenyl-2-butanone was the only common component. The VIP value and S-plot generated by the OPLS-DA model based on the comparison of regional groups or pairwise genotypes showed up to 26 potential markers at VIP > 1. The more common components of agarwood, such as sesquiterpenes α -guruene, agarospirol, guaioi, γ -eudesmol and chromone 2-phenylethyl-4H-chromen-4-one, contributed the most to the VIP value. Supervised OPLS-DA was better than that of PLS-DA, providing a reference for the quality evaluation of AEOs. This method emphasizes providing more information and obtaining additional information when combined with appropriate multivariate modeling and effective visualization of specific labeled metabolites for identification.

Keywords Agarwood, Essential oil, GC-MS, Chemometrics, OPLS-DA

The indigenous Chen-Xiang of China is a resinous part of *Aquilaria sinensis* (Lour.) Gilg in the Thymelaeaceae family. It is mainly produced in Guangdong, Guangxi, Hainan, and Fujian provinces¹. The genus *Aquilaria* consists of more than 21 species, of which all species of *Aquilaria* and *Gyrinops* appear in the Appendix II list of the *Convention on International Trade in Endangered Species of Wild Fauna and Flora*, since 2004 (Amendments to appendices I and II of CITES, 2004). *Aquilaria malaccensis* (*A. agallocha* Roxb.) is mainly produced in Malaysia, India, Laos, Cambodia, Thailand, and Taiwan island². In Vietnam, *Aquilaria crassna* (Kỳ Nam, Trâm Hương, Dó Bầu) is the most important variety and is also widely distributed in Cambodia and Thailand³. Agarwood is widely harvested extensively to obtain aromatic oils through a distillation process. The oils have traditionally been used in perfumes in the Middle East and have been widely used in advanced perfumes, toiletries, fragrance additives, and other biotechnology products⁴. The oleoresin component only exists in withered and dying *Aquilaria* trees. In recent years, due to the increasing demand and commercial value of agarwood, the trade in agarwood has intensified, leading to the destruction of natural agarwood forests.

As a crucial aspect of aromatherapy, the extraction methods of essential oils have a significant impact on their components. According to the definition of International Organization for Standardization (ISO) (ISO/DIS9235.2), essential oil is a product made by water or steam distillation, citrus peel machining or natural materials. In addition, the methods of CO₂ supercritical extraction, subcritical fluid extraction or organic solvent extraction are also called agarwood extracts or agarwood oils⁵. The main components are sesquiterpenes, phenylethyl chromone derivatives, and aromatic compounds^{6,7}. Chromone components can be used as important indicator components for the quality evaluation and identification of agarwood. It is generally believed that essential oils with a high content of chromone compounds have better quality^{8,9}. Among the methods of extracting AEOs, steam distillation is also common. The extraction device is simple, the product is natural and

¹ College of Life Sciences, Fujian Normal University, Qishan Campus, No.18 Middle Wulongjiang Avenue, Shangjie, Minhou, Fuzhou 350117, Fujian, China. ²National Quality Supervision and Inspection Center for Incense Products (Fujian), No. 1-1, West Liuan Shan Road, Taocheng Town, Yongchun County, Quanzhou City 362600, Fujian, China. ³Fujian Vocational College of Agriculture, No.116 Guishan, Hongxing Village, Jingyang Town, Fuqing City, Fuzhou 350119, Fujian, China. ✉email: wxzhang@fjnu.edu.cn; 3311312699@qq.com

pollution-free, but only low-boiling-point components can be extracted, and characteristic components such as chromones are lost, resulting in a low yield of essential oil¹⁰. Different extraction methods, plant species, and locational areas of agarwood oils lead to difference in market price performance, which makes the essential oils adulterate³. Research on AEOs mainly focuses on the identification of chemical components, but comparison of compounds is not enough to support the establishment of a quality evaluation system for agarwood¹¹.

Metabolomic data usually include a wide range of dynamic changes in metabolite concentrations, due to the geographical phenology and processing methods of plant varieties and production areas. Although the data obtained by GC-MS can be compared with the common components of AEOs through sampling from a wide range of sources, these common components of AEOs still cannot be used as indicators for statistical verification. Our work focused on seven regions of AEOs by hydro-distillation from different habitats and species. Different chemical constituents were identified by GC-MS fingerprint and multivariate statistical analysis, including partial least squares-discriminant analysis (PLS-DA), orthogonal partial least squares-discriminant analysis (OPLC-DA) and SPSS cluster analysis^{12,13}. This multivariate statistical method is used to handle the complex data generated by GC-MS. It helps in identifying characteristic chemical markers and distinguishing samples from different regions and species^{14–16}. The methods together provide a comprehensive approach to analyzing AEOs, ensuring accurate identification and differentiation of their chemical components. These components were identified by using NIST general database retrieval and literature review, providing reference for the overall quality evaluation of AEOs.

Results and discussion

GC-MS analysis of volatile components and common chemical compounds

According to the GC-MS standard mass spectrometry database NIST2020, the volatile components of AEO samples from different habitats and 3 species were analyzed, as shown in Table 1. The 2-(2-phenylethyl) chromone compounds were determined based on the mass spectrometry characteristics and fragmentation patterns summarized in literature¹⁷, combined with the characteristics of ion fragments in this study. A total of 127 compounds with more than 85% similarity were identified from the essential oils of 7 regions, accounting for 28.6–74.6% of the total volatile components (Table 1)(Supplement file Figure S1). According to different chemical structures, these volatile components are classified into sesquiterpenes (0.3–70.4%), aromatic compounds (0.6–24.9%), aliphatic compounds (0–8.4%), 2-(2-phenylethyl) chromones (0–12.0%) and others (0.3–7.6%) (Fig. 1). Sesquiterpenes had a total of 73 components, with the most abundant being sesquiterpenes. Aromatic groups followed, consisting of 12 components in total. Aliphatic groups and 2-(2-phenylethyl) chromones were the least, with 10 and 4 components, respectively.

In the AEOs extracted by hydro-distillation, due to the high temperature of water vapor in the extraction process, or the influence of the solvent used in the analysis, the 2-(2-phenylethyl) chromone component often did not appear¹¹. However, the AEOs obtained by supercritical fluid extraction and microwave-assisted extraction usually contain Flidersiachromone, 6-methoxy-2-(2-phenylethyl) chromone, 4 H-1-benzopyran-4-one chromone, and a few other semi-volatile 2-(2-phenylethyl) chromones^{11,17}. In this study, more aromatic components could be extracted from hydro-distillation AEO samples by pretreatment with 95% ethanol in GC analysis, and excellent chromone components could be detected¹⁸.

Low molecular weight aromatic compounds are important components of AEOs, and they are frequently regarded as the primary source of aroma in AEOs. More aromatic compounds were detected in the resinous agarwood, were absent from the non-resinous parts and confirmed as characteristic of the resinous parts¹⁹. AEOs contained abundant fatty acids, possibly affecting the complex process of resin accumulation, prolonging the accumulation time, and resulting in a longer formation time being required for agarwood oil yield⁴. Based on the aliphatic relative content of samples (Fig. 1), these were ranked as S3 Taiwan (8.38%) > S2 Hainan (2.34%) > S7 Cambodia (1.71%) > S1 Guangxi (1.64%) > S6 Vietnam B (0.68%) > S5 Vietnam A (0.15%) > S4 Malaysia (0%). We also considered in practical operations that the lower the fatty acid content, the more important it was to evaluate the quality of agarwood.

The phytocomplexity of the AEOs signifies the production of a multitude of plant–fungus mediated secondary metabolites as chemical signals for natural ecological communication. Table 1 shows an aromatic compound, 4-phenyl-2-butanone, as the only common component. A similar component, 3-phenyl-2-butanone, also appeared in hydro-distilled essential oil of *A. malaccensis* and *A. sub-integra* from Malaysia, Thailand, and Cambodia²⁰. This common 4-phenyl-2-butanone was presented in *A. malaccensis* represents an important basis for plant–fungal metabolic analysis chemistry in wild plants and in vitro plantlets^{7,21}. We consider that the S2 sample should contain a small amount or no agarwood, as most of its heartwood was directly extracted by hydro-distillation or might have been added with unknown chemical essences. Therefore, six common components could be detected in the other 6 AEOs except the S2 sample, namely, 1,1,4,5,6-pentamethyl-2,3-dihydro-1H-indene (aromatic compound), viridiflorol (sesquiterpenoid compound), bis(2-ethylhexyl)phthalate (aromatic compound, plasticizer), in addition to 5-(2-methylpropyl)-nonane and 2,6,10-trimethyl-dodecane (others). One terpenoid of particular interest is viridiflorol, a known common fragrance molecule of agarwood²². Viridiflorol has shown moderate antibacterial activity against *Mycobacterium tuberculosis*, the causative agent of tuberculosis, in an in vitro assay. It is also produced by the endophytic root fungus *Serendipita indica* and exhibits antifungal activity against *Colletotrichum truncatum*²³. It was particularly surprising that bis(2-ethylhexyl) phthalate, diethyl phthalate, and dibutyl phthalate were detected in this study (Table 1). These components are often used as plasticizers, condensing agents, anti-wear agents, and gas chromatographic stationary liquids for polyvinyl chloride resins. The plasticizers are mixed with some food oils to reduce product costs and should not be used as effective components of AEOs. Studies have shown that excessive intake of these plasticizers can have adverse effects on human reproduction, development, and the cardiovascular system²⁴. The total content of plasticizer added in the S4 sample accounted for about 23.9%. The quality of S4 essential oil was poor, and some samples

No.	Compounds	MW.	MF.	Relative content(%)							Notes
				S1	S2	S3	S4	S5	S6	S7	
1	Benzaldehyde	106	C ₇ H ₆ O	0.05	-	-	-	-	-	0.01	*
2	Decane, 4-methyl-	156	C ₁₁ H ₂₄	-	-	0.01	0.03	0.02	0.04	0.01	
3	3,3-dimethyloctane	142	C ₁₀ H ₂₂	0.02	-	-	-	-	-	-	
4	2-ethyl-1-hexanol	130	C ₈ H ₁₈ O	-	-	-	0.04	-	-	-	
5	5-(2-methylpropyl)-nonane,	184	C ₁₃ H ₂₈	0.05	-	0.03	0.06	0.07	0.16	0.03	
6	3-ethyl-3-methylheptane	142	C ₁₀ H ₂₂	-	-	-	-	-	-	0.03	
7	2,6,10-trimethyl dodecane	212	C ₁₅ H ₃₂	0.05	-	0.04	0.02	0.03	0.10	0.02	
8	3,7-dimethyldecane	170	C ₁₂ H ₂₆	-	-	-	0.05	-	-	-	
9	5,7-dimethylundecane	184	C ₁₃ H ₂₈	-	-	-	-	0.06	-	-	
10	Trans-linalool oxide (furanoid)	170	C ₁₀ H ₁₈ O ₂	-	-	0.01	-	0.02	-	-	
11	4-phenyl-2-butanone	148	C₁₀H₁₂O	0.27	0.21	2.27	0.30	0.13	1.10	0.16	*
12	β-selinene	204	C ₁₅ H ₂₄	-	0.12	-	-	-	-	-	△
13	β-dihydroagarofuran	222	C ₁₅ H ₂₆ O	-	0.43	0.20	-	-	-	-	△
14	α-cubebene	204	C ₁₅ H ₂₄	-	-	-	-	0.10	-	-	△
15	n-decanoic acid	172	C ₁₀ H ₂₀ O ₂	-	-	0.08	-	-	-	-	○
16	Copaene	204	C ₁₅ H ₂₄	-	-	-	-	0.49	-	-	△
17	β-elemen	204	C ₁₅ H ₂₄	-	-	-	-	0.48	-	0.05	△
18	Cyperene	204	C ₁₅ H ₂₄	-	-	-	-	1.46	-	-	△
19	α-gurjunene	204	C₁₅H₂₄	-	-	-	-	40.82	-	0.08	△
20	α-cedrene	204	C ₁₅ H ₂₄	-	-	-	-	-	-	0.27	△
21	β-caryophyllen	204	C ₁₅ H ₂₄	-	-	-	-	0.27	-	0.14	△
22	Valerena-4,7(11)-diene	204	C ₁₅ H ₂₄	-	-	-	-	0.06	-	-	△
23	(+)-calarene	204	C ₁₅ H ₂₄	-	-	-	-	-	-	0.35	△
24	cis-α-bergamotene	204	C ₁₅ H ₂₄	0.22	-	-	-	0.09	-	-	△
25	Thujopsene	204	C ₁₅ H ₂₄	0.22	-	-	-	-	-	-	△
26	α-guaiene	204	C ₁₅ H ₂₄	-	0.16	-	-	-	-	0.06	△
27	Alloaromadendrene	204	C ₁₅ H ₂₄	-	-	-	-	10.7	-	0.20	△
28	γ-gurjunene	204	C ₁₅ H ₂₄	-	-	-	-	3.65	-	-	△
29	(Z)-2-methoxy-4-(prop-1-enyl)phenol	164	C ₁₀ H ₁₂ O ₂	-	-	-	0.09	-	-	-	*
30	Humulene	204	C ₁₅ H ₂₄	0.06	-	-	-	0.46	-	0.07	△
31	Humulene epoxide II	220	C ₁₅ H ₂₄ O	-	0.08	-	-	-	-	-	△
32	(1R,4S,5S)-1,8-dimethyl-4-(prop-1-en-2-yl)spiro[4.5]dec-7-ene	204	C ₁₅ H ₂₄	-	-	-	-	-	-	0.17	△
33	(8R,8aS)-8,8a-dimethyl-3,4,6,7,8,8a-hexahydronaphthalen-2(1H)-one	178	C ₁₂ H ₁₈ O	-	-	0.06	-	-	-	-	
34	(4S,4aR,6R)-4,4a-dimethyl-6-(prop-1-en-2-yl)-1,2,3,4,4a,5,6,7-octahydronaphthalene	204	C ₁₅ H ₂₄	0.15	0.09	-	-	-	0.21	0.10	△
35	1H-cyclopropa[a]naphthalene, decahydro-1,1,3a-trimethyl-7-methylene-, [1aS-(1a.alpha.,3a.alpha.,7a.β.,7b.α.)]-	204	C ₁₅ H ₂₄	-	0.13	-	-	-	-	-	△
36	1,1,4,5,6-pentamethyl-2,3-dihydro-1H-indene	188	C₁₄H₂₀	0.06	-	0.09	0.12	0.81	0.07	0.33	*
37	α-curcumene	202	C ₁₅ H ₂₂	0.10	-	-	-	0.10	-	1.10	△
38	β-vetispirene	202	C ₁₅ H ₂₂	-	-	-	0.16	-	-	-	△
39	β-longipinene	204	C ₁₅ H ₂₄	-	-	-	-	-	-	0.33	△
40	α-bulnesene	204	C ₁₅ H ₂₄	0.13	-	-	-	1.24	-	0.22	△
41	(+)-ledene	204	C ₁₅ H ₂₄	-	1.15	-	-	1.06	-	-	△
42	β-chamigrene	204	C ₁₅ H ₂₄	-	-	-	-	-	-	0.24	△
43	2-butanone, 4-(4-methoxyphenyl)-	178	C ₁₁ H ₁₄ O ₂	-	-	0.84	0.18	-	-	-	*
44	α-murolene	204	C ₁₅ H ₂₄	-	-	-	-	0.21	-	0.83	△
45	γ-cadinene	204	C ₁₅ H ₂₄	-	-	-	-	0.11	-	0.78	△
46	Cadina-1(10),4-diene	204	C ₁₅ H ₂₄	-	-	-	-	-	-	0.91	△
47	δ-cadinene	204	C ₁₅ H ₂₄	-	-	-	-	0.29	-	-	△
48	trans-calamenene	202	C ₁₅ H ₂₂	-	-	-	-	-	-	0.11	△
49	Kessane	222	C ₁₅ H ₂₆ O	-	0.34	0.04	-	-	0.2	-	△
50	1H-cyclopropa[a]naphthalene, 1a,2,6,7,7a,7b-hexahydro-1,1,7,7a-tetramethyl-	202	C ₁₅ H ₂₂	-	-	-	-	0.10	-	-	△
51	α-cadinene	204	C ₁₅ H ₂₄	-	-	-	-	-	-	0.22	△
52	3,4-dihydro-8-hydroxy-3-methyl-(1H)-2-benzopyran-1-one	178	C ₁₀ H ₁₀ O ₃	0.09	-	-	-	-	-	-	
53	α-calacorene	200	C ₁₅ H ₂₀	-	-	-	-	-	-	0.36	△
54	α-agarofuran	220	C ₁₅ H ₂₄ O	-	-	0.94	-	-	-	-	△

Continued

No.	Compounds	MW.	MF.	Relative content(%)							Notes
				S1	S2	S3	S4	S5	S6	S7	
55	1-naphthalenol, 1,2,3,4,4a,5,6,7-octahydro-4a,5-dimethyl-3-(1-methylethenyl)-	220	C ₁₅ H ₂₄ O	-	-	-	-	-	-	0.23	△
56	Elemol	222	C ₁₅ H ₂₆ O	0.11	0.79	0.39	-	0.18	-	1.47	△
57	E-nerolidol	222	C ₁₅ H ₂₆ O	-	-	-	-	-	-	0.72	△
58	(4aS,8R)-4a,8-dimethyl-4,4a,5,6,7,8-hexahydronaphthalen-2(3 H)-one	178	C ₁₂ H ₁₈ O	0.10	-	-	-	-	-	-	△
59	(1aR,4 S,4aS,7R,7aS,7bR)-1,1,4,7-tetramethyldecahydro-1 H-cyclopropa[e]azulen-4a-ol	222	C ₁₅ H ₂₆ O	-	-	-	-	0.16	-	-	△
60	β-elemene	204	C ₁₅ H ₂₄	0.18	-	-	-	-	-	-	△
61	Espatulenol	220	C ₁₅ H ₂₄ O	-	-	-	-	0.30	-	-	△
62	Santalol, cis,α-	220	C ₁₅ H ₂₄ O	1.04	3.56	-	-	-	0.51	0.28	△
63	1-[(2E)-1,5-dimethyl-2-hexenyl]-7a-methyloctahydro-1 H-indene	248	C ₁₈ H ₃₂	-	-	-	0.69	-	-	-	
64	3,7,11,trimethyl-8,10- dodecadienylacetate	266	C ₁₇ H ₃₀ O ₂	-	-	-	3.31	-	-	-	
65	Diethyl phthalate (plasticizer)	222	C ₁₂ H ₁₄ O ₄	-	-	0.37	7.54	-	-	-	*
66	(-)-globulol	222	C ₁₅ H ₂₆ O	-	-	-	-	0.16	-	-	△
67	Cedrol	222	C ₁₅ H ₂₆ O	0.50	-	-	-	-	-	-	△
68	Bicyclo[5.3.0]decane, 2-methylene-5-(1-methylvinyl)-8-methyl-	204	C ₁₅ H ₂₄	-	-	0.49	-	-	-	-	△
69	2-naphthalenemethanol, 2,3,4,4a,5,6,7,8-octahydro-.alpha.,.alpha.,4a,8-tetramethyl-, [2R-(2.alpha.,4a.beta.,8.beta.)]-	222	C ₁₅ H ₂₆ O	-	-	-	-	0.66	-	-	△
70	1,10-di-epi-cubanol	222	C ₁₅ H ₂₆ O	-	-	-	-	-	-	0.32	△
71	γ-eudesmol	222	C ₁₅ H ₂₆ O	0.45	5.15	-	-	0.87	2.62	7.61	△
72	(-)- aristolene	204	C ₁₅ H ₂₄	1.51	4.76	4.94	-	1.33	3.87	-	△
73	Agarospinol	222	C ₁₅ H ₂₆ O	7.26	2.77	4.99	-	1.00	3.78	1.88	△
74	2-(6,10-dimethylspiro[4.5]dec-6-en-2-yl)-2-propanol	222	C ₁₅ H ₂₆ O	-	1.51	-	-	-	-	-	△
75	T-cadinol	222	C ₁₅ H ₂₆ O	-	-	-	-	-	-	1.89	△
76	α-cadinol	222	C ₁₅ H ₂₆ O	-	-	-	-	-	-	8.37	△
77	Guaiol	222	C ₁₅ H ₂₆ O	-	12.41	-	-	1.97	7.57	-	△
78	2-(4a,8-dimethyl-2,3,4,5,6,8a-hexahydro-1 H-naphthalen-2-yl)propan-2-ol	222	C ₁₅ H ₂₆ O	-	-	10.08	-	-	-	4.43	△
79	<i>trans</i> -guai-11-en-10-ol	222	C ₁₅ H ₂₆ O	-	-	-	-	-	1.72	-	△
80	Neointermedeol	222	C ₁₅ H ₂₆ O	1.48	3.69	-	-	-	-	-	△
81	α-bisabolene	222	C ₁₅ H ₂₆ O	0.50	-	-	-	-	-	-	△
82	Cycloisolongifolene, 8,9-dehydro-	202	C ₁₅ H ₂₂	0.16	-	-	-	-	-	-	△
83	β-vetivenene	202	C ₁₅ H ₂₂	-	0.90	0.70	-	-	-	-	△
84	Acorenone B	220	C ₁₅ H ₂₄ O	-	1.04	-	-	-	-	-	△
85	Eudesma-3,11(13)-dien-12-al	218	C ₁₅ H ₂₂ O	0.85	-	-	-	-	-	-	△
86	α-bisabolol	222	C ₁₅ H ₂₆ O	-	-	-	-	-	-	0.38	△
87	Humulene epoxide I	220	C ₁₅ H ₂₄ O	0.68	1.09	-	-	-	-	-	△
88	<i>trans</i> -valerenyl acetate	262	C ₁₇ H ₂₆ O ₂	-	-	-	-	-	-	0.25	○
89	1(2H)-naphthalenone, 3,4,4a,5,6,7-hexahydro-4a,5-dimethyl-3-(1-methylethenyl)-, [3 S-(3.alpha.,4a.alpha.,5.alpha.)]-	218	C ₁₅ H ₂₂ O	0.18	-	-	-	-	-	-	△
90	α-cyperonea	218	C ₁₅ H ₂₂ O	-	-	0.8	3.42	-	-	-	△
91	Ylangenal	218	C ₁₅ H ₂₂ O	-	-	-	-	-	-	0.34	△
92	Aristolone	218	C ₁₅ H ₂₂ O	-	-	0.49	-	0.57	-	-	△
93	(E)-5-((1 S,5R,8aR)-5-formyl-5,8a-dimethyl-2-methylenedecahydronaphthalen-1-yl)-3-methylpent-2-en-1-yl acetate	346	C ₂₂ H ₃₄ O ₃	0.89	-	-	-	-	-	-	
94	(E)-2-((8R,8aS)-8,8a-dimethyl-3,4,6,7,8,8a-hexahydronaphthalen-2(1 H)-ylidene)propyl formate	248	C ₁₆ H ₂₄ O ₂	-	-	-	-	0.15	-	-	○
95	α-Costol	220	C ₁₅ H ₂₄ O	0.34	0.88	-	-	-	-	-	△
96	2-naphthalenemethanol, decahydro-8-hydroxy-.alpha.,.alpha.,4a,8-tetramethyl	240	C ₁₅ H ₂₈ O ₂	1.73	-	-	-	-	-	-	△
97	2(3H)-naphthalenone, 4,4a,5,6,7,8-hexahydro-4a,5-dimethyl-3-(1-methylethylidene)-, (4aR-cis)-	218	C ₁₅ H ₂₂ O	3.20	5.11	-	-	1.23	4.00	2.59	△
98	(1R,4aR,5 S)-5-[(E)-5-hydroxy-3-methylpent-3-enyl]-1,4a-dimethyl-6-methylidene-3,4,5,7,8,8a-hexahydro-2 H-naphthalene-1-carbaldehyde	304	C ₂₀ H ₃₄ O ₂	-	1.06	-	-	-	-	-	
99	(1R,4aR,7R,8aR)-7-(2-Hydroxypropan-2-yl)-1,4a-dimethyldecahydronaphthalen-1-ol	240	C ₁₅ H ₂₈ O ₂	0.47	-	-	-	-	-	-	△
100	[7-(2-hydroxypropan-2-yl)-4a-methyl-1-methylidene-2,3,4,5,6,7,8,8a-octahydronaphthalen-2-yl] acetate	280	C ₁₇ H ₂₈ O ₃	-	-	0.47	-	-	-	-	
101	5,9-cyclopentacycloundecenediol, tetradecahydro-1-(1-hydroxy-1-methylethyl)-3a-methyl-6,10-bis(methylene)-	322	C ₂₀ H ₃₄ O ₃	-	-	0.87	-	-	-	-	
102	Arctioid	238	C ₁₅ H ₂₆ O ₂	0.73	-	-	-	-	-	-	△
103	Azuleno[6,5-b]furan-2(3 H)-one, 3a,4,4a,5,6,7,9a-octahydro-3,5,8-trimethyl-, (3R,3aR,4aS,5R,9aS)-	234	C ₁₅ H ₂₂ O ₂	0.88	-	-	0.14	-	-	0.46	△

Continued

No.	Compounds	MW.	MF.	Relative content(%)							Notes
				S1	S2	S3	S4	S5	S6	S7	
104	Viridiflorol	222	C₁₅H₂₆O	0.03	-	0.87	0.01	0.01	0.01	0.01	△
105	Dibutyl phthalate (plasticizer)	278	C ₁₆ H ₂₂ O ₄	0.38	0.13	0.2	-	-	-	-	*
106	n-hexadecanoic acid	256	C ₁₆ H ₃₂ O ₂	1.03	1.93	5.76	-	-	-	-	○
107	Hexadecanoic acid, ethyl ester	284	C ₁₈ H ₃₆ O ₂	-	0.16	-	-	-	-	0.62	○
108	Naphthalene, 1,2,3,4-tetrahydro-1-nonyl-	258	C ₁₉ H ₃₀	-	-	-	-	-	-	0.10	*
109	3-pentanone, 1,5-diphenyl-	238	C ₁₇ H ₁₈ O	0.56	0.25	1.07	0.32	-	-	0.37	*
110	9-octadecenoic acid, methyl ester	296	C ₁₉ H ₃₆ O ₂	-	-	-	-	-	-	0.09	○
111	Heneicosane	296	C ₂₁ H ₄₄	-	-	-	-	-	-	0.08	
112	9-octadecenoic acid	282	C ₁₈ H ₃₄ O ₂	0.61	-	2.00	-	-	-	-	○
113	Oleic acid	282	C ₁₈ H ₃₄ O ₂	-	0.25	-	-	-	-	0.75	○
114	1-penten-3-one, 1,5-diphenyl-	236	C ₁₇ H ₁₆ O	1.95	-	0.32	-	-	-	0.86	*
115	(E)-9-octadecenoic acid ethyl ester	310	C ₂₀ H ₃₈ O ₂	-	-	-	-	-	-	0.47	
116	Octadecanoic acid	284	C ₁₈ H ₃₆ O ₂	-	-	0.54	-	-	-	-	○
117	3-phenylbutylaldehyde	148	C ₁₀ H ₁₂ O	0.18	-	-	-	-	-	-	*
118	n-heptadecane	240	C ₁₇ H ₃₆	-	-	-	-	-	-	0.06	
119	n-eicosane	282	C ₂₀ H ₄₂	-	-	-	-	-	-	0.04	
120	2-phenethyl-4H-chromen-4-one	250	C₁₇H₁₄O₂	9.82	0.33	0.38	-	1.39	-	0.23	☆
121	2-monopalmitin	330	C ₁₉ H ₃₈ O ₄	-	-	-	-	-	0.28	-	○
122	Bis(2-ethylhexyl) phthalate (plasticizer)	390	C ₂₄ H ₃₈ O ₄	0.17	-	0.38	16.39	0.08	0.92	0.56	*
123	2 H-pyran-2-one, 5,6-dihydro-6-pentadecyl-	308	C ₂₀ H ₃₆ O ₂	-	0.07	0.03	-	-	-	-	
124	6-methoxy-2-phenethyl-4 H-chromen-4-one	280	C ₁₈ H ₁₆ O ₃	1.78	-	-	-	-	-	-	☆
125	2-(4-methoxyphenethyl)-4 H-chromen-4-one	280	C ₁₈ H ₁₆ O ₃	-	0.25	-	-	0.93	-	-	☆
126	Squalene	410	C ₃₀ H ₅₀	0.28	-	-	-	-	-	0.08	
127	6,7-dimethoxy-2-phenethyl-4 H-chromen-4-one	310	C ₁₉ H ₁₈ O ₄	0.39	-	0.29	-	-	0.12	-	☆

Table 1. Chemical compositions of agarwood essential oils analyzed by GC-MS. Sesquiterpenes Δ;aromatic compounds *;aliphatic compounds ○;2-(2-phenylethyl)chromones ☆.Noteworthy common components and potential markers in this study (**Black bold font**). Each batch of AEOs samples were analyzed by GC-MS every other day. Based on NIST 2020 database, substances with similarity greater than 85% were identified as chemical components. The value of relative content was the average of six samples from each region.

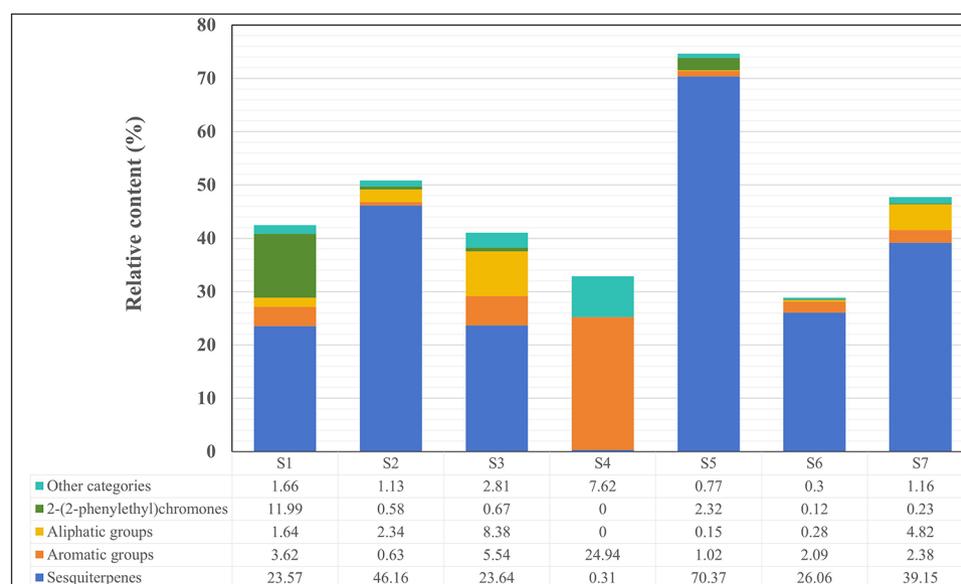


Fig. 1. Volatile constituents per each planting region of agarwood.

were also detected but low, which might be due to the accumulation of plants themselves, or pollution caused by GC analysis²⁵.

Sesquiterpenoids are natural terpenoids containing 15 carbon atoms in a molecule composed of three isoprene units¹¹. In addition to the viridiflorol, by comparing various samples, we have identified the following 6 common components that were worth noting: elemol (sesquiterpene), γ -eudesmol (sesquiterpene), (-)-aristolene (sesquiterpene), agarospirol (sesquiterpene), 2(3H)-naphthalenone, 4, 4a, 5,6,7,8-hexahydro-4a, 5-dimethyl-3- (1-methylidene)-, (4ar cis)- (sesquiterpene) and 2-phenyl-4 H-chromen-4-one (chromone derivative) (Table 1). According to modern pharmacology, sesquiterpene components of agarwood have good biological activity in the central nervous system, respiratory system, and digestive system, etc. Elemol is a natural product that sesquiterpenoid has a role as a fragrance, showing modest antioxidant, anti-inflammatory, and antiproliferative activities of the essential oil of *Cymbopogon nardus*^{26,27}. Plants with aromatic properties have multiple chemical components in their essential oils, such as the main component of *Blepharocalyx salicifolius*, which was the viridiflorol and eudesmane sesquiterpenes^{28,29}. In the fungus-mediated fermentation of resinous agarwood, the most significant finding was the appearance of key agarwood sesquiterpenes such as agarospirol, γ -eudesmol, (-)-aristolene²⁸. A sesquiterpenoid 2-(3H)-naphthalenone, 4, 4a, 5, 6, 7, 8-hexahydro-4a, 5-dimethyl-3- (1-methylidene)-, (4ar cis)-, which has not been found in other agarwood literatures. However, it showed higher performance in the comparison of common components in this study, and its relative content was higher than other components.

One of the main active components of agarwood, chromone, has been isolated and found to have 240 different subunits. It has anti-inflammatory and anti-tumor properties, neuroprotective effects, and inhibitory effects on acetylcholinesterase, tyrosinase, and glucosidase³⁰. It is worth noting that 2-(2-phenylethyl) chromones often do not appear in the essential oil extracted by hydro-distillation. However, among the four chromones analyzed in this study, the common component, 2-phenyl-4H-chromen-4-one, was presented in the 5 regions, and this chromone component has not been found in other literatures. There were significant differences in the relative content of each component, which might lead to differences in special flavors of agarwood from different habitats.

Multivariate statistical analysis of PLS-DA and OPLS-DA

PLS-DA is the deformation of PLS, used to establish classification and is suitable for supervised discriminant analysis methods with small intergroup differences¹⁴. It is applied to prediction and descriptive modeling, as well as selecting discriminative variables, determining the chemical compositions from different genotypes and product regions, and automatically generating more important principal components^{15,16}. The PLS-DA model displayed clear separation among the 7 regions and 3 genotypes of AEOs (Fig. 2a). The software automatically generated $R^2X(\text{cum}) = 0.848$, $R^2Y(\text{cum}) = 1$, and $Q^2(\text{cum}) = 0.854$ for predictive ability. In a previous agarwood study, HPLC chromatograms were used in combination with multivariate statistical screening to establish the identification methods for wild and cultivated agarwood. The Fisher linear recognition model and the PLS-DA recognition model were established³¹. This study established PLC-DA based on GC-MS data (Fig. 2a), $Q^2 > 0.5$ indicating a strong predictive ability. The result showed that there were significant differences in the volatile components of AEOs from different habitats. In addition, the discrimination of different genotypes was provided with a certain effect. The Permutation validation in SIMCA 14.1 software was used to verify the fitting of PLS-DA (Fig. 2b). Through 200 iterations of permutation testing, the model results showed that the Y-axis intercept was all less than 0, indicating that the PLS-DA model validation results were fitting and reliable. Alternative, the Hotelling's T2 analysis also verified that all samples were within the 95% confidence interval^{32,33}, validation results provided a more evaluation of model performance. (Supplementary file Fig. S2).

Supervised methods offer another approach to classification enhancing the discrimination between specimens by minimizing variance³⁴. In this study, PLS-DA was utilized to classify AEOs generated in China (CNA) and non-China (OCA). The model indicated that S4 - S7 were dispersed OCA, regardless of the first principal component or the second principal component. AEOs (S1-S3) produced in CNA showed a quadrant of aggregation (Fig. 2c). The software automatically generated $R^2X(\text{cum}) = 0.84$, $R^2Y(\text{cum}) = 1$, $Q^2(\text{cum}) = 0.981$, suggesting that the difference in volatile components of AEOs in OCA was significantly higher than that of CNA. Through 200 iterations of permutation testing, the model results showed that the Y-axis intercept was all less than 0, indicating that the PLS-DA model validation results were fitting and reliable (Fig. 2d). However, the model was not effective for screening differential volatile markers (Fig. 2c), so we conducted OPLS-DA to analyze the strategy of identifying these markers.

Supervised OPLS discriminant analysis (DA) was applied to identify the volatile markers for AEOs from different habitats. OPLS has excellent external prediction ability as well as a better visualization effect compared with PLS¹⁴. In the OPLS-DA scatter plot (Fig. 2e), the R^2X , R^2Y , and Q^2 of S1-S3 AEO samples from China and S4-S7 samples from outside China were 0.84, 1, and 0.997, respectively. The samples were located on both sides of the positive and negative axes of the first principal component with the X-axis was at 0, indicating that the volatile components of AEOs produced in CNA (S1-S3) could effectively distinguish the two quadrants from AEOs (except S5, S4-S7) in other regions, and the genotype and relative content of AEOs were different. The 200 permutation tests were conducted to verify the OPLS-DA model, and the Hotelling's T2 analysis showed that all samples were within the 95% confidence interval.

In the multivariate statistical analysis, S5 was an outlier phenomenon, with the difference components being the largest compared with the representativeness of other samples. The difference components of S5 could be used as the volatile markers of *A. crassna*. Although the PLS-DA genotype discrimination effect was good, the regional characteristics were not obvious in the quadrant, and the SIMCA software could not present the variable influence on projection (VIP). The difference between OPLS-DA groups was maximized, and the difference within the groups was weakened, which was more suitable for the separation of samples between groups³⁵. Therefore, the VIP diagram was presented for further analysis.

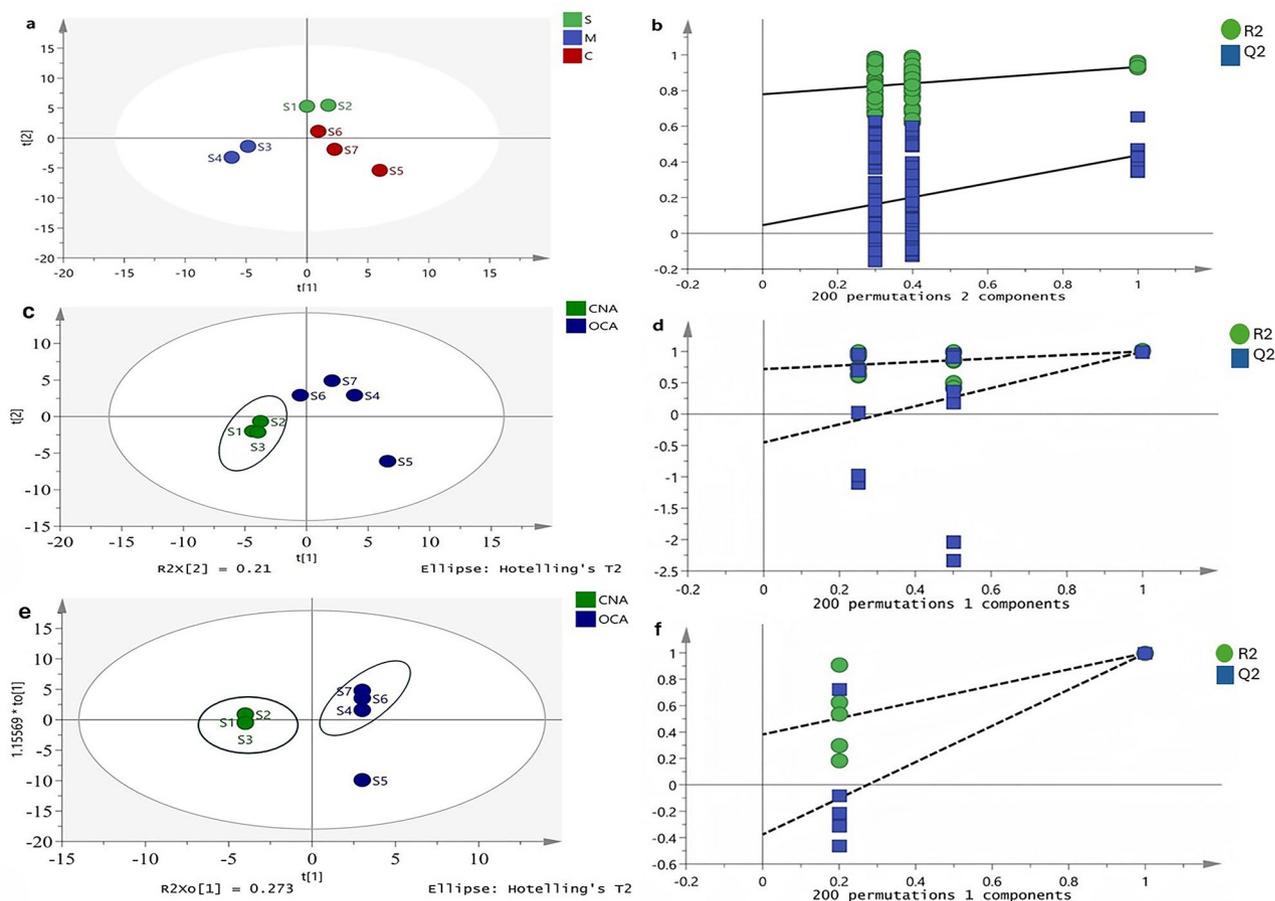


Fig. 2. The PLS-DA obtained scatter plot, (a) classification of AEOs in three genotypes; (c) classification of CNA and OCA areas, and 200 permutations of PLS-DA model validation results (b, d). The scatter plot of OPLS-DA, (e) classification of CNA and OCA areas, and the verification results of 200 permutation of OPLS-DA models (f). S: *A.sinensis*, M: *A.malaccensis*, C: *A.crassna*, CNA: AEOs from China (S1-S3 samples). OCA: AEOs from outside of China (S4-S7 samples). R2: Coefficient of determination, Q2: squared cross-validation. The Hotelling's T2 analysis showed that all samples were within the 95% confidence interval, and no samples beyond this confidence interval were found, so there was no significance.

OPLS-DA model screening the volatile markers of AEOs

The VIP value and S-plot evaluation method were employed to identify the key components contributing to the grouping of AEOs. The S-plot, a scatter plot combining covariance and correlation loading profiles resulting from an OPLS-DA model was utilized³⁶. Variables with a VIP greater than 1 were deemed statistically significant and served as important markers of the model^{37,38}. The VIP values (Fig. 3a) and S-plot (Fig. 3b) generated by the OPLS-DA model revealed 26 components with a VIP value > 1 (Table 2). These potential component values far away from the origin of S-plot represented variables that contributed a lot to the classification and were more reliable than the near origin components as potential markers to distinguish the AEOs from different producing regions. Statistical tests such as SPSS was carried out on significant variables to make the model acceptable. It is worth noting that sesquiterpenes and chromones are the index components of agarwood, sesquiterpenes α -gurjunene (VIP = 4.86), agarospirol (VIP = 2.86), alloaromadendrene (VIP = 2.49), (-)-aristolene (VIP = 2.37) and chromone 2-phenylethyl-4H-methylene-4-one (VIP = 2.63) were significantly VIP > 2 components between CNA and OCA. While α -gurjunene exhibited the highest VIP value, especially in “Hui-An” agarwoods. OPLS-DA analysis revealed, but this component did not appear in the CNA (S1-S3) group (Table 2, No. 19). The main factor affecting the VIP value was the detection of 40.82% content in S5, but this outstanding VIP value was also easy to distinguish among complex plant metabolites. Moreover, most aromatic components carry distinct aromas associated with AEOs³⁹. Although bis(2-ethylhexyl) phthalate raises concerns due to its toxicity, and it remains uncertain whether it originated from AEOs or pyrolysis during extraction^{24,25}. Due to the detection of 16.4% content of bis(2-ethylhexyl)phthalate in S4 sample, it also affected the VIP value of inter genomic comparison analysis. However, this could also clearly highlight the identification of this component.

AEOs from CNA contained more guaiol, and OCA (Malaysia, Vietnam, and Cambodia) contained more α -gurjunene. The total relative contents of differential components between the two producing regions were higher in OCA. Notably, two sesquiterpenes, α -gurjunene and agarospirol, stood out in the S-plot diagram, being distanced from the origin and the main compound groups (Fig. 3b). Specifically, α -gurjunene (VIP = 4.86)

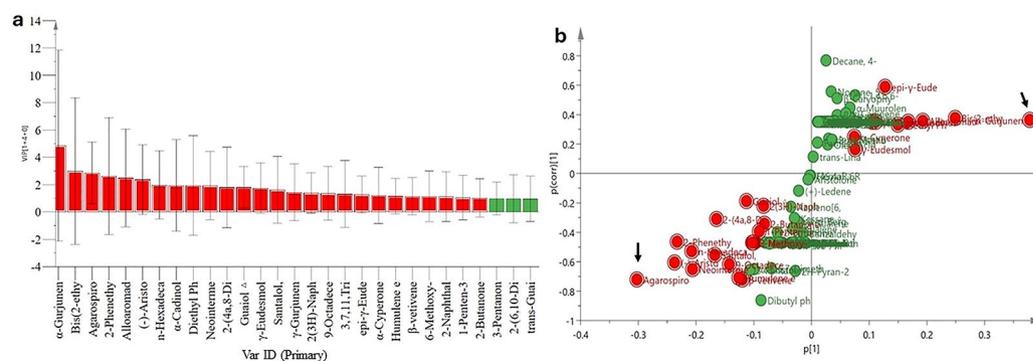


Fig. 3. The VIP value diagram (a) and S-plot diagram (b) of AEOs were obtained by OPLS-DA model. The OPLS-DA model showed that there were 26 components (red color) with a VIP value > 1. The two positions indicated by the arrow are agarospirol (left) and α -gurjunene (right).

Sort	Compounds	Average of total relative content (%)		VIP value	Notes
		CNA	OCA		
01	α -gurjunene	0.00	20.81	4.86	Δ
02	Bis(2-ethylhexyl) phthalate	0.28	4.49	2.99	*
03	Agarospirol	5.01	2.22	2.86	Δ
04	2-phenethyl-4 H-chromen-4-one	3.51	0.81	2.63	\star
05	Alloaromadendrene	0.00	5.45	2.49	Δ
06	(-)-Aristolene	3.74	2.60	2.37	Δ
07	n-hexadecanoic acid	2.91	3.11	1.98	\circ
08	α -cadinol	0.00	8.37	1.95	Δ
09	Diethyl phthalate	0.37	7.54	1.94	*
10	Neointermedeol	2.59	0.00	1.93	Δ
11	2-(4a,8-dimethyl-2,3,4,5,6,8a-hexahydro-1 H-naphthalen-2-yl)propan-2-ol	10.08	4.34	1.81	Δ
12	Guaiol	12.41	4.77	1.80	Δ
13	γ -eudesmole	2.80	3.70	1.75	Δ
14	Santalol, cis, α -	2.30	0.40	1.63	Δ
15	γ -gurjunene	0.00	3.65	1.44	Δ
16	2(3 H)-naphthalenone, 4,4a,5,6,7,8-hexahydro-4a,5-dimethyl-3-(1-methylethylidene)-, (4ar-cis)-	4.16	2.61	1.39	Δ
17	9-octadecenoic acid	1.31	0.00	1.35	\circ
18	3,7,11,Trimethyl-8,10- dodecedienylacetate	0.00	3.31	1.31	-
19	γ -eudesmol	0.59	1.22	1.28	Δ
20	α -cyperonea	0.80	3.42	1.21	Δ
21	Humulene epoxide I	0.89	0.00	1.17	Δ
22	β -vetivenene	0.10	0.00	1.16	Δ
23	6-methoxy-2-phenethyl-4 H-chromen-4-one	1.78	0.00	1.15	\star
24	2-naphthalenemethanol, decahydro-8-hydroxy-.alpha.,.alpha.,4a,8-tetramethyl	1.73	0.00	1.13	Δ
25	1-penten-3-one, 1,5-diphenyl-	1.14	0.85	1.06	*
26	4-phenyl-2-butanone	0.93	0.42	1.03	*

Table 2. Twenty-six distinguished violate compounds of AEOs between CNA and OCA. The VIP values and S-plot generated by the OPLS-DA model revealed 26 components with a VIP value > 1. CNA: AEOs from China (S1-S3 samples). OCA: AEOs from outside of China (S4-S7 samples). Sesquiterpenes Δ ;aromatic compounds *;aliphatic compounds \circ ;2-(2-phenylethyl)chromones \star .

significantly influenced the grouping of AEO samples and was positively correlated with the grouping of AEOs. Prior studies have employed OPLS-DA model to discriminate between the *A. sinensis* and its subspecies “Chin-Nan” and to identify potential distinguishing components. Notably, sesquiterpenes, particularly guaiane and eudesmane derivatives, were considered key markers contributing to their odoriferous properties³⁷. Similarly,

the sesquiterpenes in AEOs also exhibited significant differences, indicating their potential as characteristic components.

OPLS-DA model comparing the pairwise genotypes of AEOs

In this study, OPLS-DA effectively modeled two or more classes. In addition to the CNA and OCA analyses mentioned above, three agarwood genotypes, *A. sinensis*, *A. malaccensis* and *A. crassna*, were classified and compared according to the pairwise genotypes (Fig. 4). The OPLS-DA model analyzed different components among different producing regions. The model results indicated that when comparing *A. sinensis* and *A. malaccensis*, R^2X (cum) = 0.689, R^2Y (cum) = 1, Q^2 (cum) = 0.86; when comparing *A. sinensis* and *A. crassna*, R^2X (cum) = 1, R^2Y (cum) = 1, Q^2 (cum) = 1, when comparing *A. malaccensis* and *A. crassna*, R^2X (cum) = 1, R^2Y (cum) = 1, Q^2 (cum) = 1, indicating that the models could describe most of the GC-MS data and possessed good predictive ability. The volatile components of AEOs exhibited certain similarities within the same genotypes, but differences existed between different genotypes.

VIP values and S-plots were used to screen the differential chemical components contributing the most to the pairwise genotype group (Fig. 5). The VIP value results showed that there were 25 components with a VIP value > 1 between *A. sinensis* and *A. malaccensis*, with sesquiterpene guaiol (VIP = 2.55) being the largest contribution component between the two genotypes (excluding bis(2-ethylhexyl) phthalate) (Fig. 5a). Comparing *A. sinensis* and *A. crassna*, 25 components with VIP value > 1 were identified, with α -gurjunene (VIP = 5.28) being the largest contribution component (Fig. 5c). Between *A. malaccensis* and *A. crassna*, 22 components with VIP value > 1 were found, with α -gurjunene (VIP = 5.03) being the largest contribution component (Fig. 5e).

The results of the differential component analysis revealed that AEOs from *A. sinensis* contained relatively more guaiol and 2-phenethyl-4H-chromen-4-one, whereas those from *A. malaccensis* contained more sesquiterpene 2-(4a, 8-dimethyl-2,3,4,5,6,8a-hexahydro-1H-naphthalen-2-yl)propan-2-ol (Fig. 5a and b). Additionally, when comparing AEOs of *A. sinensis* and *A. crassna*, aside from the differences in guaiol, *A. sinensis* AEOs also contained more agarospirol and 2-phenethyl-4H-chromen-4-one, while the AEOs of *A. crassna* contained more sesquiterpenes α -gurjunene and alloaromadendrene (Fig. 5c and d). Compared with AEOs in pairwise genotypes, *A. malaccensis* and *A. crassna*, the α -gurjunene exhibited significant differences in composition contribution (Fig. 5e and f), with AEOs of *A. malaccensis* containing more sesquiterpene 2-(4a, 8-dimethyl-2,3,4,5,6,8a-hexahydro-1H-naphthalen-2-yl)propan-2-ol, while AEOs of *A. crassna* contained more sesquiterpene γ -eudesmol. These contribution classifications of pairwise genotypes could serve as potential markers to distinguish AEOs of different species. The results demonstrated that the production regions of AEOs could be better distinguished based on chemometrics.

Analysis of such multivariate data requires methodology capable of handling both the contribution to the OPLS model, i.e., concentration variant, and correlation to the OPLS model, i.e., concentration invariant³⁶. Statistical SPSS test ($P < 0.05$) was carried out on significant variables to make the model acceptable. Based on the above OPLS-DA results, the unique phytochemical characteristics of various species may be related to the genetic information of primitive plant germplasm or endophytic fungi. The current strategy focuses on this complex problem, emphasizing the strategy of obtaining additional information when appropriate multivariate modeling is combined with appropriate and effective visualization of specific marker metabolites to identification.

Conclusion

AEOs have a significant international market through hydro-extraction, particularly in Muslim regions. For the first time, we utilized GC-MS to delineate the chemical fingerprints of AEOs in three primary genotypes: *A. sinensis*, *A. malaccensis* and *A. crassna*, and analyzed the differences in aroma components across various production regions. Metabolomics data typically encompass vast dynamic ranges in metabolite concentration. Here, we reveal distinctive differences in sesquiterpenes, chromone and its derivatives, and low-molecular-weight aromatic compounds. A total of 127 compounds were identified from the AEOs, with sesquiterpenes comprising the majority, totaling 73 components. The aromatic compound 4-phenyl-2-butanone was the sole common component among the seven samples. Additionally, there were 7 common components with a higher occurrence of sesquiterpenes and chromone: viridiflorol; elemol; γ -eudesmol; (-)-aristolene; agarospirol; 2(3H)-naphthalenone, 4, 4a, 5,6,7,8-hexahydro-4a, 5-dimethyl-3-(1-methylidene)-, (4ar cis)- and 2-phenyl-4H-chromen-4-one. It was particularly surprising that plasticizers bis(2-ethylhexyl) phthalate, diethyl phthalate and dibutyl phthalate were detected in this study. The total content of plasticizers added in S4 sample accounted for about 23.9%, considering the poor quality of S4 essential oil. Other samples exhibited low levels of detection, likely due to contamination during GC analysis.

PLS-DA and OPLS-DA methods were employed for multivariate statistical analysis of the differential chemical components between different genotypes and habitats. The results demonstrated that the AEOs from different habitats could be effectively classified and identified based on GC-MS combined with chemometrics. In OPLS-DA, 26 differential markers including 17 sesquiterpenes, 2 chromones and 3 aromatics, were identified according to VIP value. The VIP value and S-plot generated by the comparison of regional groups (CNA and OCA) in the OPLS-DA model showed a total of 26 potential markers in VIP > 1, and a total of up to 25 potential markers were generated by comparison of two genotypes. The components of agarwood such as α -gurjunene, agarospirol, guaiol, γ -eudesmol and 2-phenethyl-4H-chromen-4-one were searched and summarized in the literature related to agarwood, which contributed the most to the VIP value. The unique phytochemical characteristics of agarwood may be related to the interactive information of original plant germplasm or invasive microorganisms. The current strategy focuses on this complex issue. By using multivariate statistical analysis, the indicator components can be scientifically highlighted, even if additional chemicals are added to reduce

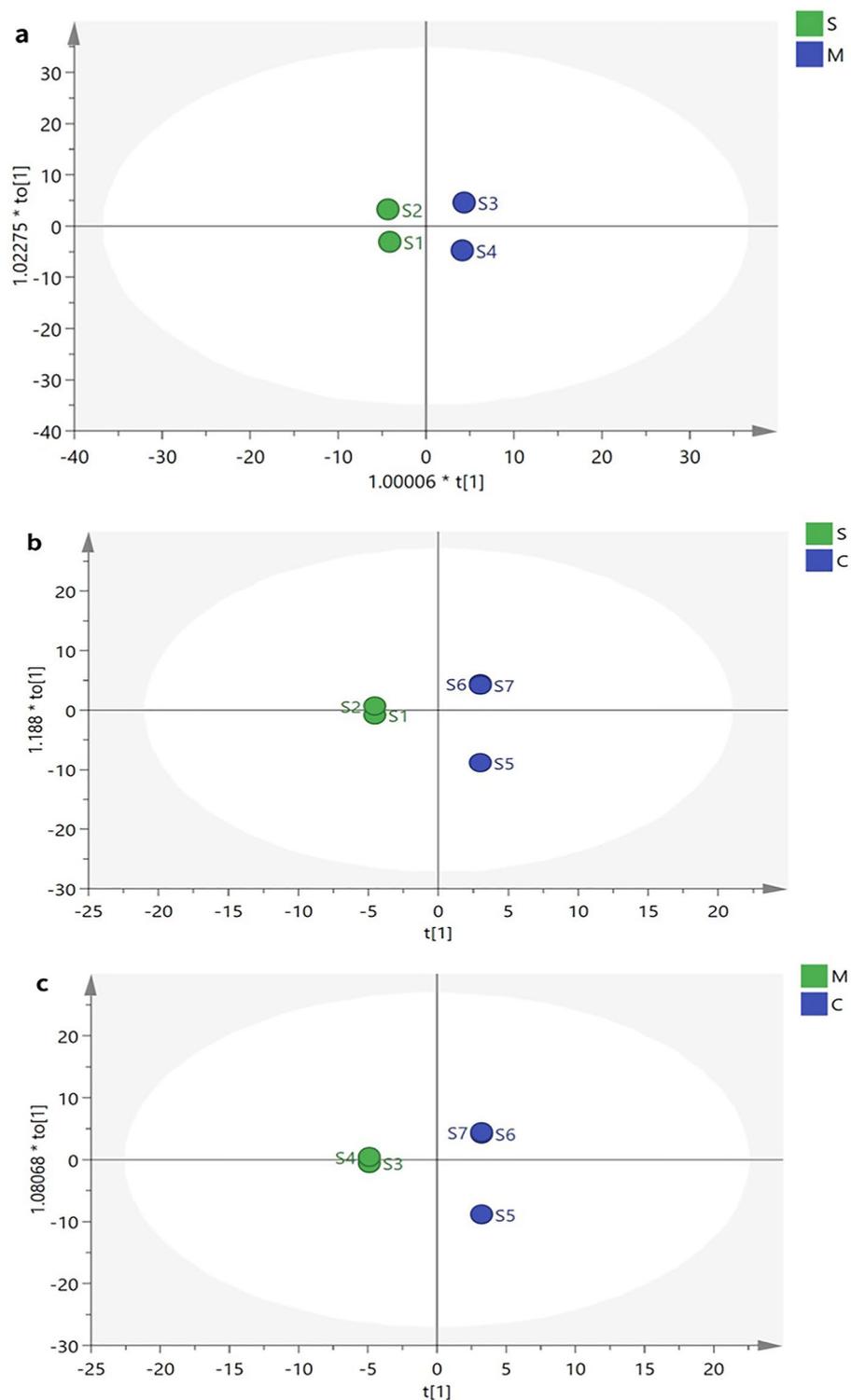


Fig. 4. OPLS-DA model was used to distinguish the producing areas of AEOs by comparison of pairwise genotypes. Compare S: *A.sinensis* with M: *A. malaccensis* (a), compare S: *A.sinensis* with C: *A.crassna* (b), compare M: *A.malaccensis* with C: *A.crassna* (c).

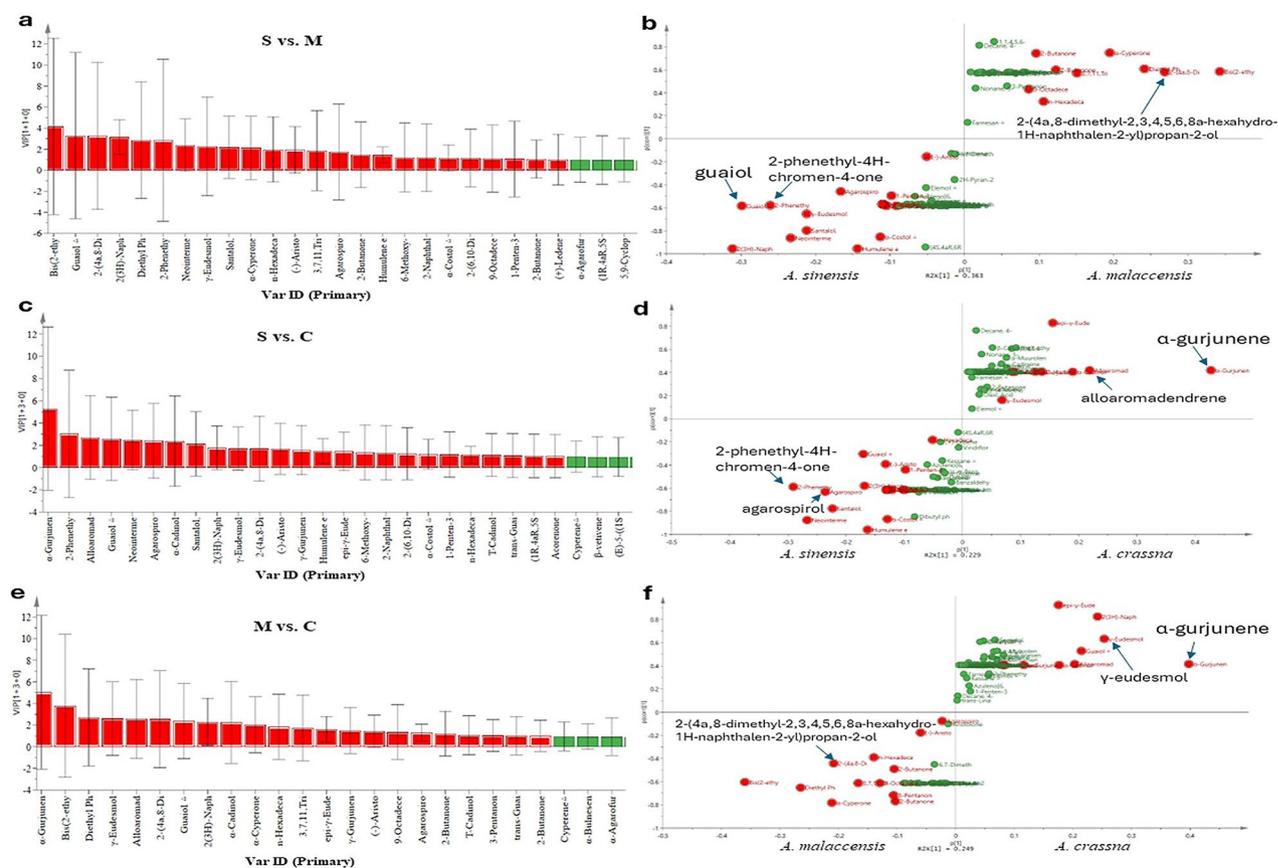


Fig. 5. The VIP value diagram (left) and S-plot diagram (right) of PLS-DA model were compared the pairwise genotypes of the AEOs. (a and b) *A.sinensis* and *A.malaccensis*, (c and d) *A.sinensis* and *A.crassna*, (e and f) *A.malaccensis* and *A.crassna*. The arrows exhibit significant differences in composition contribution between two-pair genotypes. Plasticizers should not be the component of AEOs, so the contributing component should be eliminated.

No.	Collecting regions	Species	Remarks
S1	Nanning, Guangxi	<i>A.sinensis</i>	Autonomous distillation
S2	Haikou, Hainan	<i>A.sinensis</i>	Shop purchase
S3	Changhua, Taiwan	<i>A.malaccensis</i>	Autonomous distillation
S4	Kuala Lumpur, Malaysia	<i>A.malaccensis</i>	Autonomous distillation
S5	Huian, Vietnam	<i>A.crassna</i>	Autonomous distillation
S6	Nha Trang, Vietnam	<i>A.crassna</i>	Autonomous distillation
S7	Phnom Penh, Cambodia	<i>A.crassna</i>	Autonomous distillation

Table 3. Source information of 7 regions of AEO samples. Autonomous distillation: autonomous distillation in local factory and extracts essential oils by solvent-free extraction method. Six samples from each region.

product costs such as plasticizers. Therefore, the strategy emphasizes providing more information and obtaining additional information when appropriate multivariate modeling is combined with appropriate and effective visualization of specific marker metabolites for identification.

Experimental section

Plant materials

Seven regions of AEOs were collected from Guangxi, Hainan and Taiwan for the China areas, and from Vietnam, Cambodia and Malaysia producing areas for Southeast Asia. Essential oils were obtained through water distillation or steam at the production regions and local shop purchase. Six samples were randomly selected from each planting region. See Table 3 for the source information.

GC–MS analyses of AEOs

Accurately weigh 30 mg of essential oil in a 5 mL EP tube, then add 2 mL of ethyl acetate solution (China National Pharmaceutical Group Chemical Reagent Co., Ltd., China) to dissolve. Shook well and let it stand for 2 h. Extract 1 mL of the essential oil solution and filter it through a 0.45 µm filter membrane, preparing it for gas chromatography-mass spectrometry analysis.

The compositions of the essential oils were analyzed by GCMS-QP2010 Plus (Shimadzu, Tokyo, Japan), equipped with an SH-Rxi-5Sil MS Cap. column (30 m × 0.25 mm i.d., 0.25 µm film thickness; Shimadzu, Japan). The temperature program was as follows: initial temperature 90 °C for 2 min, then increased by 2 °C min⁻¹ to 150 °C and held for 5 min, and then increased by 2 °C min⁻¹ to 280 °C and held for 5 min. The other parameters were as follows: injection temperature, 250 °C; ion source temperature, 230 °C; EI, 70 eV; carrier gas, He at 1 ml min⁻¹; injection volume, 1 ml; split ratio, 1:20; solvent delay of 2.5 min and mass range, *m/z* 50–550. Quantification was obtained from percentage peak areas from the gas chromatogram. Identification of individual compounds was carried out using the NIST2020 (National Institute of Standards and Technology, US. Department of Commerce) Registry of Mass Spectral Database to search the compounds of authentic references. Chromatographic results expressed as area percentages were calculated with a response factor of 1.0.

Methodological examination

Precision test

In S1–S7 regions of AEOs from different sources (Table 3), randomly selected one region, such as S1. Out of the six samples in each region, equal amounts were drawn and thoroughly mixed to form one sample. The test solution was prepared according to the above preprocessing description, and GC-MS analysis was conducted under the above chromatographic and mass spectrometric conditions. Following the same process, the analysis was repeated six times on the mixed S1 sample. The six data points were compared using the Similarity Evaluation System for Chromatographic Fingerprint of Traditional Chinese Medicine (Version 2012) (Chinese Pharmacopoeia Commission, China), and a similarity of no less than 0.99 indicated fine precision of the instrument.

Repeatability test

For the repeatability test, samples of AEOs from the same source (such as S1) were used. Six samples were made according to the steps description above. The weighing of each sample had to be precise. GC-MS analysis was conducted as described above. The six data points were compared using the Similarity Evaluation System for Chromatographic Fingerprint of Traditional Chinese Medicine, and a similarity of no less than 0.99 indicated good repeatability of the method.

Stability test

Any sample solution from S1–S7 was randomly selected, and the selected sample (such as S1) was dissolved into a tube of solution following the preprocessing steps. The solution was stored for different times: 2, 4, 6, 8, 12, and 24 h for GC-MS analysis. The six data points were compared using the Similarity Evaluation System for Chromatographic Fingerprint of Traditional Chinese Medicine, and a similarity of no less than 0.99 indicated that the test solution was stable within 24 h.

Data processing

Each experiment was repeated three times. Based on the NIST2020 database, the volatile components of the samples were qualitatively analyzed by mass spectrometry. Peak area normalization was used to calculate the relative percentage content. Substances with a similarity greater than 85% were identified as potential chemical components of AEOs. Using the software of Similarity Evaluation System for Chromatographic Fingerprint of Traditional Chinese Medicine (Version 2012) with a time window of 0.2, automatic matching was performed through multi-point correction using the median method. The similarity and common peaks between each sample and the reference map were calculated, and a GC-MS fingerprint map was constructed.

Multivariate analysis

The SIMICA14.1 software (Umetrics Co., Sweden) for multivariate data analysis was used. The compound data was normalized, and then the software performed diversified statistical analysis through PLS-DA analysis and OPLS-DA modules. PLS-DA and OPLS-DA were introduced for discrimination and derivation of potential markers (VIP score > 1)³⁶. Finally, the cluster analysis was carried out in combination with SPSS27.0 data processing software. The univariate statistical analysis was introduced to confirm those differentially expressed features (*p* < 0.05). The cluster analysis used between-cluster linkage, and the Euclidean distance was used as a sample measure to determine the difference between the producing regions and species of AEOs.

Data availability

Data availability All data generated or analyzed during this study are included in this published article and its supplementary information file.

Received: 19 August 2024; Accepted: 7 January 2025

Published online: 13 February 2025

References

- Hu, Z. K., Chen, Y., Yan, T. T. & Li, G. Y. Differences between Qi-Nan agarwood and ordinary agarwood in secretion properties and antibacterial activity. *Sci. Sil Sin.* **59**, 136–144. <https://doi.org/10.11707/j.1001-7488.LYKX20210536> (2023).
- Persoon, G. A. & Beek, H. H. Growing 'The wood of the gods': agarwood production in Southeast Asia. In: (eds Snelder, D. J. & Lasco, R. D.) *Smallholder Tree Growing for Rural Development and Environmental Services: Lessons from Asia*. Springer, Netherlands, Dordrecht. 245–262. (2008).
- Nguyen, H. T. et al. Multi-platform metabolomics and a genetic approach support the authentication of agarwood produced by *Aquilaria crassna* and *Aquilaria malaccensis*. *J. Pharmaceut Biomed.* **142**, 136–144. <https://doi.org/10.1016/j.jpba.2017.04.047> (2017).
- Chen, H. Q. et al. Comparison of compositions and antimicrobial activities of essential oils from chemically stimulated agarwood, wild agarwood and healthy *Aquilaria sinensis* (Lour.) Gilg trees. *Molecules* **16**, 4884–4896. <https://doi.org/10.3390/molecules16064884> (2011).
- Agarwood extract. Forestry Industry Standard of the People's Republic of China (LY/T 3361–2023). National Forestry and Grassland Administration. China Standards Press. (2023). <https://www.spc.net.cn>
- Li, W. et al. Five new eudesmane-type sesquiterpenoids from Chinese agarwood induced by artificial holing. *Fitoterapia* **100**, 44–49. <https://doi.org/10.1016/j.fitote.2014.11.010> (2015).
- 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*. **1387**, 104–115. <https://doi.org/10.1016/j.chroma.2015.01.096> (2015).
- Yang, D. L. et al. Fragrant agarofuran and eremophilane sesquiterpenes in agarwood 'Qi-Nan' from *Aquilaria sinensis*. *Phytochem Lett.* **8**, 121–125. <https://doi.org/10.1016/j.phytol.2014.03.003> (2014).
- Yan, T. T., Chen, Y., Wang, Q., Shang, L. L. & Li, G. Y. Quality analyses of domestic agarwood (*Aquilaria sinensis*) in China by LC-MS combining with Chemometrics. *China Wood Ind.* **33**, 14–18. <https://doi.org/10.19455/j.mcgy.20190604> (2019).
- Geng, T. Y., Luo, L. Y., Zeng, L., Wang, X. H. & Duan, Y. Z. On effect of different extraction methods on preparation of agarwood essential oil. *J. Southwest. China Normal Univ. (Natural Sci. Edition)*. **45**, 59–67. <https://doi.org/10.13718/j.cnki.xsxb.2020.02.009> (2020).
- Chen, X. Q. et al. Chemical composition and potential properties in mental illness (anxiety, depression and insomnia) of agarwood essential oil: a review. *Molecules* **27**, 4528–4545. <https://doi.org/10.3390/molecules27144528> (2022).
- Nguyen, H. T. et al. UPLC-QTOFMS based metabolomics followed by stepwise partial least square-discriminant analysis (PLS-DA) explore the possible relation between the variations in secondary metabolites and the phylogenetic divergences of the genus *Panax*. *J. Chromatogr. B*. 1012–1013. <https://doi.org/10.1016/j.jchromb.2016.01.002> (2016).
- Guo, D. Y. et al. Chemical composition analysis and discrimination of essential oils of *Artemisia Argyi* Folium from different germplasm resources based on electronic nose and GC/MS combined with chemometrics. *Chem. Biodivers.* **20**, e202200991. <https://doi.org/10.1002/cbdv.202200991> (2023).
- Bylesjo, M., Rantalainen, M., Cloarec, O. & Nicholson, J. K. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J. Chemometr.* **20**, 341–351. <https://doi.org/10.1002/cem.1006> (2006).
- Lee, L. C., Liong, C. Y. & Jemain, A. A. Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. *Analyst* **143**, 3526–3539. <https://doi.org/10.1039/C8AN00599K> (2018).
- Huang, Y. M. et al. Comparison of a novel PLS1-DA, traditional PLS2-DA and assigned PLS1-DA for classification by molecular spectroscopy. *Chemometr. Intell. Lab.* 209104225. <https://doi.org/10.1016/j.chemolab.2020.104225> (2021).
- Mei, W. L. et al. Characterization and determination of 2-(2-phenylethyl)chromones in agarwood by GC-MS. *Molecules* **18**, 12324–12345. <https://doi.org/10.3390/molecules181012324> (2013).
- Zhang, Y. et al. Temporal characteristics of agarwood formation in *Aquilaria sinensis* after applying whole-tree agarwood-inducing technique. *Chin. Herb. Med.* **15**, 37–44. <https://doi.org/10.1016/j.chmed.2022.07.003> (2023).
- Takamatsu, S. & Ito, M. Correction to: Agarotretol: a source compound for low molecular weight aromatic compounds from agarwood heating. *J. Nat. Med.* **73**, 685. <https://doi.org/10.1007/s11418-019-01310-x> (2019).
- Hashim, Y. Z. H. Y., Ismail, N. I. & Abbas, P. Analysis of chemical compounds of agarwood oil from different species by gas chromatography mass spectrometry (GCMS). *IJUM Eng. J.* **15**, 55–60. <https://doi.org/10.31436/ijumej.v15i1.469> (2014).
- Chung, J. P. & Chen, K. S. Developing an in vitro quasi-symbiotic culture system of *Aquilaria malaccensis* with potentially beneficial microbes on agarwood resinous formation. *Plant. Cell. Tiss. Org. Cult.* **133**, 193–202. <https://doi.org/10.1007/s11240-017-1372-7> (2018).
- Yan, T. T., Yang, S., Chen, Y., Wang, Q. & Li, G. Y. Chemical profiles of cultivated agarwood induced by different techniques. *Molecules*. **24**, (1990). <https://doi.org/10.3390/molecules24101990> (2019).
- Ntana, F. et al. Hamberger, B. A sesquiterpene synthase from the endophytic fungus *serendipita indica* catalyzes formation of viridiflorol. *Biomolecules* **11**, 898–914. <https://doi.org/10.3390/biom11060898> (2021).
- Luo, Q. et al. Global review of phthalates in edible oil: an emerging and nonnegligible exposure source to human. *Sci. Total Environ.* **704**, 135369. <https://doi.org/10.1016/j.scitotenv.2019.135369> (2019).
- Kdimy, A. et al. Isolation of two plasticizers, bis(2-ethylhexyl)terephthalate and bis(2-ethylhexyl) phthalate, from *Capparis spinosa* L. Leaves. *Chem. Biodivers.* **20**, e202300903. <https://doi.org/10.1002/cbdv.202300903> (2023).
- Pellegrini, M. C., Alonso-Salces, R. M., Umpierrez, M. L., Rossini, C. & Fuselli, S. R. Chemical composition, antimicrobial activity, and mode of action of essential oils against *Paenibacillus* larvae, etiological agent of American foulbrood on *Apis mellifera*. *Chem. Biodivers.* **14**, e1600382. <https://doi.org/10.1002/cbdv.201600382> (2017).
- Bayala, B. et al. Chemical composition, antioxidant, anti-inflammatory and antiproliferative activities of the essential oil of *Cymbopogon nardus*, a plant used in traditional medicine. *Biomol. Concepts*. **11**, 86–96. <https://doi.org/10.1515/bmc-2020-0007> (2020).
- Sen, S., Dehingia, M., Talukdar, N. C. & Khan, M. Chemometric analysis reveals links in the formation of fragrant biomolecules during agarwood (*Aquilaria malaccensis*) and fungal interactions. *Sci. Rep.* **7**, 44406. <https://doi.org/10.1038/srep44406> (2017).
- Furtado, F. B. et al. Chemical composition and bioactivity of essential oil from *Blepharocalyx Salicifolius*. *Int. J. Mol. Sci.* **19**, 33. <https://doi.org/10.3390/ijms19010033> (2018).
- Mi, C. N. et al. 2-(2-phenylethyl)chromone derivatives: promising α -glucosidase inhibitors in agarwood from *Aquilaria filaria*. *Phytochem* **181**, 112578. <https://doi.org/10.1016/j.phytochem.2020.112578> (2021).
- Shang, L. L., Chen, Y., Yan, T. T., Wang, Q. & Li, G. Y. Identification of wild agarwood and cultivated agarwood by HPLC coupled with multivariate data analysis. *Chem. Ind. For. Prod.* **38**, 33–41. <https://doi.org/10.3969/j.issn.0253-2417.2018.06.005> (2018).
- Hotelling, H. The generalization of Student's ratio. *Ann. Math. Stat.* **2**, 360–378 (1931).
- Srivastava, M. S. & Awan, H. M. On the robustness of Hotelling's T^2 -test and distribution of linear and quadratic forms III sampling from a mixture of two *mult1* variate normal populations. *Commun. Sta-Theor M.* **11**, 81–107. <https://doi.org/10.1080/036109282088282191982> (1982).
- Farag, M. A., Dokaalah, E. U., Eissa, T. F., Kamal, I. M. & Zayed, A. Chemometrics-based aroma discrimination of 14 Egyptian mango fruits of different cultivars and origins, and their response to probiotics analyzed via SPME coupled to GC–MS. *ACS Omega*. **7**, 2377–2390 (2022). <http://pubs.acs.org/journal/acsofd>

35. Trigg, J. & Wold, S. Orthogonal projections to latent structures (OPLS). *J. Chemometr.* **16**, 119–128. <https://doi.org/10.1002/cem.695> (2002).
36. Wiklund, S. et al. Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. *Anal. Chem.* **80**, 115–122. <https://doi.org/10.1021/ac0713510> (2008).
37. Yu, M. et al. Remarkable phytochemical characteristics of Chi-Nan agarwood induced from new-found Chi-Nan germplasm of *Aquilaria sinensis* compared with ordinary agarwood. *Int. J. Anal. Chem. Article ID.* **5593730**, 10pages. <https://doi.org/10.1155/2021/5593730> (2021).
38. Wang, W., Jin, S. & Guo, Y. L. Exploration of a method of distinguishing different Nongxiang Tieguanyin tea grades based on aroma determined by GC-MS combined with chemometrics. *Molecules* **24**, 1707. <https://doi.org/10.3390/molecules24091707> (2019).
39. Chen, H. Q. et al. Chemical constituents of agarwood originating from the endemic genus *Aquilaria* plants. *Chem. Biodivers.* **9**, 236–250. <https://doi.org/10.1002/cbdv.201100077> (2012).

Acknowledgements

This research received financial support from Fujian Provincial Department of Science and Technology, China under Contract No. 2022N5004 and 2023S0036.

Author contributions

SZ Qian, YM Jiang and WX Zhang conceived here search. SZ Qian and QL Yan performed the research. JP Chung wrote the paper. WX Zhang and JP Chung revised the paper. DH Wu provided the materials and resources for the research. WX Zhang and JP Chung performed the integrity of the work.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-85976-2>.

Correspondence and requests for materials should be addressed to W.-X.Z. or J.-P.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025