Characterization of Botanical Parts of *Erythrina crista-galli* Using Pyrolysis-Gas Chromatography/Mass Spectrometry and Multivariate Analysis

Abd. Wahid Rizaldi Akili[1], Ari Hardianto[1], Jalifah binti Latip[2], Maya Ismayati[3], and Tati Herlina[1*]

[1]*Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Padjajaran, Jl. Raya Bandung Sumedang Km 21 Jatinangor, Sumedang 45363, Indonesia*

[2]*School of Chemical Sciences and Food Technology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Selangor 46300, Malaysia*

[3]*Research Center for Biomass and Bioproducts, National Research and Innovation Agency (BRIN), Jl. Raya Bogor Km 46, Cibinong 16911, Indonesia*

* ***Corresponding author:***

email: tati.herlina@unpad.ac.id

Received: August 27, 2022
Accepted: April 20, 2023

DOI: 10.22146/ijc.77325

Abstract: Erythrina crista-galli *is commonly used in folk medicines for its pharmacological properties which are associated with the bioactive compounds. Profiling botanical parts of* E. crista-galli *is an exciting topic and essential to uncover the similarity and clustering based on their chemical content. The botanical parts of* E. crista-galli, *including bark, flowers, leaves, roots, and twigs, were subjected to pyrolysis-gas chromatography/mass spectrometry. The samples were pyrolyzed using a multi-shot pyrolyzer. The relative abundance of the pyrolysate was subjected to multivariate analysis,* i.e., *principal component analysis (PCA) and hierarchical cluster analysis (HCA). The scree plot for PC.1, PC. 2, and PC. 3 accounted for 36.5%, 27.2%, and 20.3%, respectively. Together, the first three PCs explain 84% of the total variance. The PCA allows characterizing the roots of* E. crista-galli *by the highest relative abundance of lignin G, followed by the twigs, bark, and leaves, while the flowers had the least relative abundance of lignin G. The HCA allows to cluster the botanical parts of* E. crista-galli *into three different clusters based on their chemical component similarity,* i.e., *flowers-leaves, twigs, and roots-bark. In conclusion, Py-GC/MS analysis can be used in conjunction with multivariate data analysis to characterize the botanical parts of* E. crista-galli.

Keywords: E. crista-galli*; pyrolysis-GC/MS; multivariate analysis; principal component analysis; hierarchical clustering analysis*

■ INTRODUCTION

Erythrina (Fabaceae) is a large genus comprising around 200 species [1]. They are commonly used in folk medicines in Asian, African, and South American countries due to their pharmacological properties. One of the *Erythrina* species, *E. crista-galli*, was traditionally used as a wound healing and sedative. Meanwhile, people in Indonesia used *E. crista-galli* for malaria treatment by stewing the leaves and barks [2]. Additionally, *E. crista-galli* was also reported to have laxative, hypertensive, and diuretic activities. The botanical parts of *E. crista-galli* have various bioactivity; for example, the aerial parts of *E. crista-galli* have analgesic and anti-inflammatory activities; the root has antibacterial and antifungal activities, the bark has antibacterial, antimycobacterial, and antifungal activities; the leaves have antibacterial, antifungal, antivirus, animal repellent, and cytotoxic activities; while the flowers show antimutagenic activity [3]. These efficacies are associated with the metabolites constituents, which may unevenly spread within the botanical parts of *E. crista-galli*, as reported for some other species [4-5]. Thus, profiling the botanical parts of *E. crista-galli* is an exciting topic and essential to uncover

the similarity and the clustering of every botanical part based on their chemical content.

Various methods can be used for metabolite profiling, such as gas chromatography (GC) [6], high-performance liquid chromatography (HPLC) [7], gas chromatography coupled to mass spectrometry (GC-MS) [8], gas chromatography coupled to time-of-flight mass spectrometry (GC-TOF-MS) [9], ultra-performance liquid chromatography coupled to time-of-flight mass spectrometry (UPLC-TOFMS) [10], and pyrolysis-gas chromatography/mass spectrometry (Py-GC/MS) [5]. Among these methods, Py-GC/MS has the advantage that it is a fast analysis method, it requires simple sample preparation and a small amount of sample. Py-GC/MS can analyze diverse metabolite species, including high molecular weight metabolites, which in turn provides the opportunity to analyze the whole compound including primary and other metabolites [11].

Pyrolysis works by applying heat greater than the energy of specific bonds so that the molecule will fragment in a reproducible way. The fragments produced are then separated by the capillary column of the GC to produce the pyrogram. The interpretations of resulting pyrograms require detailed knowledge of the pyrolysis behavior of the desired compounds. This poses extreme difficulty for the global elucidation of metabolites, but since the Py-GC/MS of complex matrices results in a complex mixture of volatile fragments of the original sample, the resulting pyrogram can be used very effectively as a fingerprint of that particular sample. The analysis of the fingerprint pattern of these samples is often accomplished by the use of multivariate statistical techniques, which can be used to reveal relationships between samples and correlations between variables [12]. Two of the most used multivariate techniques to explore similarities and hidden patterns among samples are principal component analysis (PCA) and hierarchical cluster analysis (HCA) [13].

When the variables in a data set are highly correlated, which suggests data redundancy, PCA is extremely beneficial. PCA can be used to reduce the original variables into a smaller number of new variables called principal components that explain the majority of the variance in the original variable due to this redundancy [14]. PCA can also provide visualization to look for grouping in a data set. However, this method does not explicitly define clusters, and this is where the HCA method comes in [15]. HCA is a method to determine the underlying structure of observations by repeating a procedure that associates or dissociates each object until they are all processed wholly and equally. This method divides samples from a data set into groups that are related to one another [16]. Therefore, in our study, we use HCA in addition to PCA to explore similarities and hidden patterns among different parts of *E. crista-galli*.

In this study, Py-GC/MS was applied to characterize the botanical parts (bark, flowers, leaves, roots, and twigs) of *E. crista-galli.* The result obtained from Py-GC/MS was then subjected to PCA and HCA multivariate analysis to distinguish between parts of *E. crista-galli* based on their whole chemical component. The PCA and HCA analyses were performed in the R programming language. To the best of our knowledge, this is the first study that aimed to characterize five different parts of the *E. crista-galli* plant based on their whole metabolites using Py-GC/MS and to cluster these different parts of *E. crista-galli* based on their metabolite fingerprint similarity.

■ EXPERIMENTAL SECTION

Materials

Materials used were the botanical parts of *E. crista-galli*, including bark, flowers, leaves, roots, and twigs, that were collected from Bandung, West Java, Indonesia. These plant materials have been determined at the Laboratory of Agricultural Production Technology & Services, Agricultural Cultivation Department, Faculty of Agriculture, Universitas Padjajaran, under voucher specimen number 1020.

Instrumentation

The equipment used in this study was eco-cup SF PY1-EC50F, glass wool, multi-shot pyrolyzer (EGA/PY-3030D) interfaced with GC/MS system QP-2020 NX (Shimadzu, Japan) equipped with an SH-Rxi-5Sil MS column with electron impact of 70 eV.

Procedure

Pyrolysis-GC/MS measurement

Py-GC/MS was performed on several botanical parts of *E. crista-galli plants* (*i.e.,* bark, flowers, leaves, roots, and twigs). About 500 µg of samples were analyzed by Py-GC/MS. It was put in eco-cup SF PY1-EC50F and covered by glass wool. Furthermore, the eco-cup was pyrolyzed at 500 °C for 6 s using a multi-shot pyrolyzer (EGA/PY-3030D) which was interfaced (interface temperature 280 °C) with a GC/MS system QP-2020 NX (Shimadzu, Japan) equipped with an SH-Rxi-5Sil MS column (30 m × 0.25 mm i.d. film thickness 0.25 µm), with electron impact of 70 eV and helium as a carrier gas. The pressure was 20.0 kPa (15.9 mL/min, column flow 0.61 mL/min). The temperature profile for GC was as follows: 50 °C held for 1 min. Then the temperature increased until 280 °C (5 °C/min), and 13 min at 280 °C. Products resulting from the pyrolysis were identified by comparing their retention times and mass spectra data with NIST LIBRARY 2017.14. The identified pyrolysates were further compared with the literature [17].

Multivariate analysis

In this study, we performed two multivariate analyses, PCA which was followed by agglomerative hierarchical clustering or Hierarchical Clustering on Principal Components (HCPCs). Pyrograms of the botanical parts were assigned a matrix (row i, column k). The botanical parts were assigned as observations (i), whereas pyrolysis products were as descriptors (k). Mean centering and scaling were applied to the matrix during the preprocessing stage. The mean centering procedure was performed to maintain the important variation. The scaling step was employed due to the different scales of pyrolysis products.

An orthogonal linear transformation was applied to the matrix to produce principal components [18]. F_s (resp. G_s) indicates the coordinate vectors of the samples (resp. pyrolysis products), which can be expressed as Eq. (1) and (2):

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_k x_{ik} m_k G_s(k) \quad (1)$$

$$G_s(k) = \frac{1}{\sqrt{\lambda_s}} \sum_k x_{ik} p_i F_s(i) \quad (2)$$

whereas $F_s(i)$ and $G_s(k)$ represent the coordinates of the botanical part i and pyrolysis product k on the axis s. Notation λ_s is the eigenvalue corresponding to the axis s. Notations of m_k and p_i are the weights associated with pyrolysis product k and the botanical part i, respectively, whereas x_{ik} refers to the matrix (row i, column k). The first PCs responsible for at least 80% variance were retained and subjected to agglomerative hierarchical clustering. The most similar individual observations i were agglomerated iteratively based on the pairwise distance of Ward's criterion. The number of clusters was selected according to the hierarchical tree. PCA and HCPC were computed in the R programming language environment using FactoMineR [19]. The results were visualized using factoextra [20] or ggplot2 [21]. Leave-one-out cross-validation (LOOCV) computation for PCA was performed using chemometrics [22].

■ RESULTS AND DISCUSSION

Pyrolysis Products of the Botanical Parts of *E. crista-galli*

The chemical compositions of bark, flowers, leaves, roots, and twigs of *E. crista-galli* were analyzed by Py-GC/MS. This analysis method produces a pyrogram that plots retention time to its relative intensity. The resulting pyrograms from the analysis of botanical parts of *E. crista-galli* are given in Fig. 1.

According to the resulting pyrogram (Fig. 1), 93 pyrolysis products (pyrolysates) were identified by comparing their retention times with mass spectra data with NIST LIBRARY 2017.14. Table 1 shows pyrolysates and their relative intensities in each sample. The most abundant pyrolysates belong to polysaccharides, followed by lignins and extractives. This finding is unsurprising since polysaccharides and lignins are the main constituents of plant materials [23]. In softwood, polysaccharides such as cellulose and hemicellulose compose 41–50 and 11–33% of the biomass, respectively, while lignin constitutes 19–30%. The cellulose and

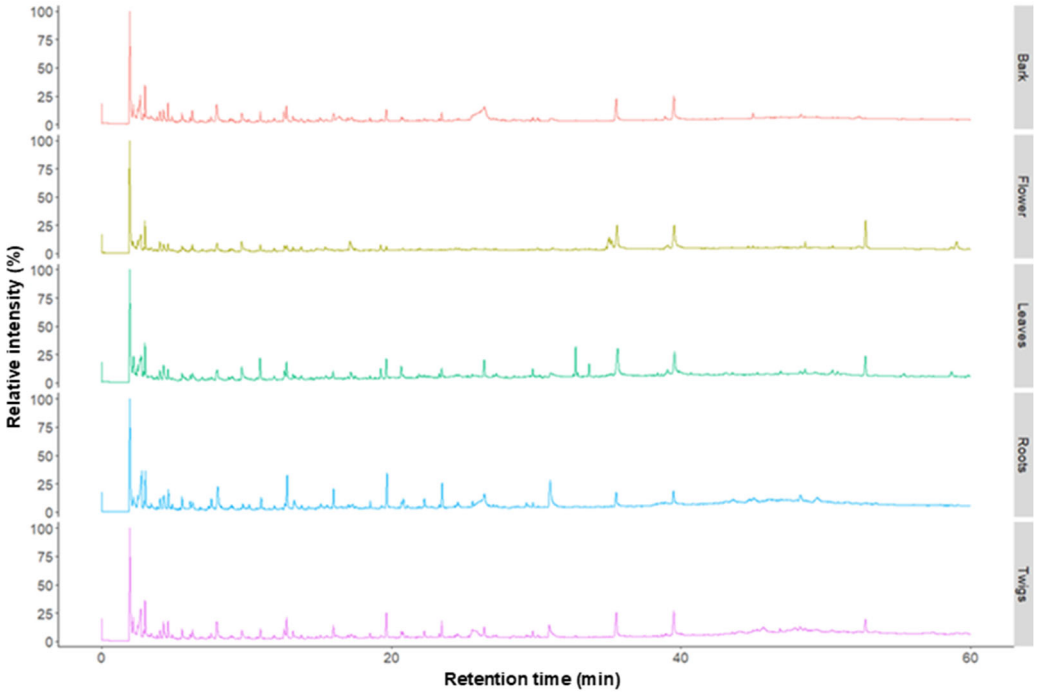

Fig 1. Pyrogram comparison of 5 botanical parts of *E. crista-galli*

Table 1. Pyrolysis products and their relative abundance

t_R (min) [a]	Pyrolysis product	SI (%) [b]	Molecular formula	Relative abundance (%)				
				Roots	Flowers	Leaves	Bark	Twigs
			Polysaccharide					
1.927	ammonium carbamate	98	$CH_6N_2O_2$	15.230	25.440	18.070	18.690	17.380
2.177	2-oxopropanal	88	$C_3H_4O_2$	3.690	1.590	4.830	4.610	3.030
2.343	2-methylpropanal	95	C_4H_8O	-	-	0.800	0.330	0.830
2.472	butane-2,3-dione	91	$C_4H_6O_2$	1.810	1.030	1.240	1.470	6.470
2.585	3-methylbutanoic acid	74	$C_5H_{10}O_2$	-	2.800	-	-	-
2.768	acetic acid	96	$C_2H_4O_2$	9.700	2.980	4.390	6.150	0.660
2.918	2,5-dihydrofuran	87	C_4H_6O	0.840	0.370	0.510	0.580	4.840
2.977	1-hydroxypropan-2-one	98	$C_3H_6O_2$	-	4.830	4.030	4.920	0.400
3.015	1-hydroxypropan-2-one	97	$C_3H_6O_2$	5.090	-	-	-	0.490
3.267	2-oxobutyl acetate	84	$C_6H_{10}O_3$	0.360	-	-	-	1.250
3.409	2,3-dihydro-1,4-dioxine	82	$C_4H_6O_2$	0.720	-	0.620	0.760	3.000
3.823	1-methylpyrrole	92	C_5H_7N	0.380	-	0.450	0.550	2.100
4.034	3-methylpenta-1,4-diene	84	C_6H_{10}	1.470	2.120	1.190	1.440	0.470
4.285	1-nitropropan-2-one	83	$C_3H_5NO_3$	2.430	1.830	2.500	2.090	1.160
4.603	methyl 2-oxopropanoate	96	$C_4H_6O_3$	2.330	0.890	0.920	2.260	0.700
4.861	5-(cyclohexylmethyl)pyrrolidin-2-one	83	$C_{11}H_{19}NO$	0.470	-	-	0.940	1.380
5.534	furan-2-carbaldehyde	90	$C_5H_4O_2$	1.320	0.570	0.570	1.200	0.820
6.100	2-hydroxycyclohexyl acetate	81	$C_8H_{14}O_3$	0.960	-	0.570	0.740	4.120
6.268	2-oxopropyl acetate	95	$C_5H_8O_3$	1.180	0.620	0.760	1.610	-
7.367	4,4-dimethyl-5-oxopentanenitrile	77	$C_7H_{11}NO$	0.250	-	-	-	-
7.564	2*H*-furan-5-one	91	$C_4H_4O_2$	1.170	-	-	0.720	-
7.964	cyclopentane-1,2-dione	90	$C_5H_6O_2$	-	2.630	2.510	4.840	-

t_R (min) [a]	Pyrolysis product	SI (%) [b]	Molecular formula	Relative abundance (%)				
				Roots	Flowers	Leaves	Bark	Twigs
8.009	4,5-dimethyloctane	83	$C_{10}H_{22}$	4.780	-	-	-	-
8.812	ethenyl propanoate	79	$C_5H_8O_2$	-	-	-	0.230	-
8.951	5-methylfuran-2-carbaldehyde	84	$C_6H_6O_2$	0.410	-	-	0.540	-
9.063	3-methylcyclopent-2-en-1-one	93	C_6H_8O	0.280	-	-	0.390	-
10.960	2-hydroxy-3-methylcyclopent-2-en-1-one	88	$C_6H_8O_2$	-	1.410	3.120	1.590	1.450
11.013	2-hydroxy-3-methylcyclopent-2-en-1-one	97	$C_6H_8O_2$	1.440	-	-	-	-
12.773	4-methoxyphenol (*p*-cresol)	96	$C_7H_8O_2$	-	1.530	-	-	-
13.277	cyclopropylmethanol	88	C_4H_8O	1.200	0.770	0.740	0.830	1.030
13.370	3-methylbutyl 2-methylpropanoate	80	$C_9H_{18}O_2$	0.410	0.480	-	-	-
13.690	3-hydroxy-2-methylpyran-4-one	89	$C_6H_6O_3$	-	-	-	0.320	-
13.802	3-ethyl-2-hydroxycyclopent-2-en-1-one	89	$C_7H_{10}O_2$	0.530	0.800	0.420	0.430	0.600
14.272	1,4-dioxaspiro[2.4]heptan-5-one	80	$C_5H_6O_3$	0.430	-	-	-	-
15.130	7-methyl-1,4-dioxaspiro[2.4]heptan-5-one	87	$C_6H_8O_3$	0.610	-	-	0.570	-
17.024	1,4:3,6-dianhydro-α-D-glucopyranose	91	$C_6H_8O_4$	0.590	-	-	0.660	-
17.185	2,3-dihydro-1-benzofuran	91	C_8H_8O	-	3.980	1.620	0.540	-
17.330	2,3-anhydro-D-mannosan	91	$C_6H_8O_4$	0.260	-	-	0.720	-
27.036	6,7-dimethoxy-1-[(*E*)-2-phenylethenyl]-1,2,3,4-tetrahydroisoquinoline	77	$C_{19}H_{21}NO_2$	0.220	-	-	-	0.260
30.117	3*H*-[1]benzofuro[3,2-d]pyrimidin-4-one	77	$C_{10}H_6N_2O_2$	-	-	-	0.520	0.370
31.110	tetradecanoic acid	91	$C_{14}H_{28}O_2$	-	-	-	-	0.920
32.728	7,11,15-trimethyl-3-methylidenehexadec-1-ene	94	$C_{20}H_{38}$	-	-	3.680	-	-
33.651	7,11,15-trimethyl-3-methylidenehexadec-1-ene	89	$C_{20}H_{38}$	-	-	1.680	-	-
35.045	(*E*)-octadec-6-enyl acetate	90	$C_{20}H_{38}O_2$	-	2.790	-	-	-
35.527	hexadecenoic acid	94	$C_{16}H_{32}O_2$	2.400	7.920	7.560	6.060	6.170
39.482	octadecanoic acid	93	$C_{18}H_{36}O_2$	1.850	9.290	6.250	5.860	5.930
46.895	dotriacontane	94	$C_{32}H_{66}$	-	-	0.510	-	0.620
48.570	tetracontane	88	$C_{40}H_{82}$	-	1.390	0.890	-	0.240
52.731	dotriacontane	95	$C_{32}H_{66}$	-	8.470	3.830	-	3.010
	Total			64.810	86.530	74.260	73.160	69.700
			Lignin G					
12.806	guaiacol	97	$C_7H_8O_2$	4.370	-	2.800	2.700	3.450
16.006	4-methylguaiacol	96	$C_8H_{10}O_2$	1.960	-	0.860	1.770	2.050
18.545	4-ethylguaiacol	94	$C_9H_{12}O_2$	0.740	-	-	0.400	0.570
19.697	4-vinylguaiacol	93	$C_9H_{10}O_2$	4.840	0.670	2.650	1.880	3.920
20.821	eugenol	94	$C_{10}H_{12}O_2$	1.070	-	0.330	0.470	0.860
21.092	4-propylguaiacol	89	$C_{10}H_{14}O_2$	0.240	-	-	0.480	-
22.280	*cis*-isoeugenol	83	$C_{10}H_{12}O_2$	1.190	-	-	-	1.030

t_R (min) [a]	Pyrolysis product	SI (%) [b]	Molecular formula	Relative abundance (%)				
				Roots	Flowers	Leaves	Bark	Twigs
23.508	*trans*-isoeugenol	95	$C_{10}H_{12}O_2$	3.120	-	1.100	1.330	2.480
24.584	acetoguaiacone	96	$C_9H_{10}O_3$	0.810	-	-	-	0.190
25.603	guaiacylacetone	92	$C_{10}H_{12}O_3$	0.670	-	-	-	0.880
29.322	(*E*)-4-(3-hydroxyprop-1-en-1-yl)-2-methoxyphenol	91	$C_{10}H_{12}O_3$	0.640	-	-	-	0.320
30.973	(*E*)-4-(3-hydroxyprop-1-en-1-yl)-2-methoxyphenol	92	$C_{10}H_{12}O_3$	6.940	-	-	1.210	3.800
	Total			26.590	0.670	7.740	10.240	19.550
			Lignin H					
9.731	phenol	98	C_6H_6O	0.650	2.770	2.260	1.760	1.460
11.901	2-methylphenol	94	C_7H_8O	-	0.660	0.450	0.650	11.893
12.642	*p*-cresol	96	C_7H_8O	0.820	1.550	1.690	2.220	12.579
	Total			1.470	4.980	4.400	4.630	25.932
			Lignin S					
20.707	syringol	94	$C_8H_{10}O_3$	0.810	-	1.710	0.630	0.950
23.320	4-methylsyringol	90	$C_9H_{12}O_3$	0.350	-	0.560	-	0.510
26.422	4-vinylsyringol	93	$C_{11}H_{14}O_4$	0.980	-	2.560	3.110	1.270
28.507	*cis*-4-propenylsyringol	83	$C_{11}H_{14}O_3$	0.170	-	-	0.330	-
28.720	syringaldehyde	92	$C_9H_{10}O_4$	0.210	-	-	-	-
29.782	*trans*-4-propenylsyringol	92	$C_{11}H_{14}O_3$	0.560	-	1.060	0.530	0.990
31.320	syringylacetone	92	$C_{12}H_{16}O_4$	-	-	-	-	0.410
	Total			3.080	0.000	5.890	4.600	4.130
			Extractive and others					
6.380	*o*-xylene	92	C_8H_{10}	-	-	0.570	-	-
10.169	2,2-diethyl-3-methyl-1,3-oxazolidine	88	$C_8H_{17}NO$	0.680	-	-	0.450	-
16.419	2-(hydroxymethyl)-2-nitropropane-1,3-diol	83	$C_4H_9NO_5$	-	-	-	1.760	-
19.262	1*H*-indole	92	C_8H_7N	-	1.250	1.470	-	-
21.897	3-methyl-1*H*-indole	96	C_9H_9N	-	-	0.400	0.410	-
26.275	3,4-diacetyloxy-6,8-dioxabicyclo[3.2.1]octan-2-yl acetate	82	$C_{12}H_{16}O_8$	-	-	-	1.310	-
27.026	pentadecan-1-ol	85	$C_{15}H_{32}O$	-	-	-	0.210	-
32.85	3,7,11,15-tetramethylhexadec-2-ene	94	$C_{20}H_{40}$	-	-	0.500	-	-
34.833	(*Z*)-18-octadec-9-enolide	93	$C_{18}H_{32}O_2$	-	0.460	-	-	-
34.963	(8*Z*)-1-oxacycloheptadec-8-en-2-one	91	$C_{16}H_{28}O_2$	-	1.820	-	-	-
35.188	(*Z*)-18-octadec-9-enolide	90	$C_{18}H_{32}O_2$	-	2.730	-	-	-
38.401	phytol (alkenol)	96	$C_{20}H_{40}O$	-	-	0.400	0.630	
38.995	(9*Z*,12*Z*,15*Z*)-octadeca-9,12,15-trienoic acid	85	$C_{18}H_{30}O_2$	-	-	-	0.530	-
39.068	(7*Z*,10*Z*,13*Z*)-hexadeca-7,10,13-trienal	90	$C_{16}H_{26}O$	-	-	0.930	-	-
39.88	hexadecanamide	88	$C_{16}H_{33}NO$	-	-	0.660	-	-
44.976	6,7-dimethoxy-1-phenyl-3,4-dihydroisoquinoline	61	$C_{17}H_{17}NO_2$	-	0.630	-	0.950	-
45.688	2,6,10,15,19,23-pentamethyl-2,6,18,22-tetracosatetraen-10,15-diol	86	$C_{30}H_{54}O_2$	-	-	-	-	1.330

t_R (min) [a]	Pyrolysis product	SI (%)[b]	Molecular formula	Relative abundance (%)				
				Roots	Flowers	Leaves	Bark	Twigs
48.224	(*E*)-3,3'-dimethoxy-4,4'-dihydroxystilbene	91	$C_{16}H_{16}O_4$	1.510	-	-	-	0.970
48.284	methyl 2-phenylquinoline-7-carboxylate	67	$C_{17}H_{13}NO_2$	-	-	-	0.830	-
49.421	clionasterol	93	$C_{29}H_{50}O$	1.870	-	-	-	-
50.815	squalene	95	$C_{30}H_{50}$	-	-	0.500	-	-
50.404	heptacosyl heptafluorobutyrate	95	$C_{31}H_{55}F_7O_2$	-	-	-	-	0.530
Total				4.060	6.890	5.430	7.080	2.830

[a] SI (%) = Similarity index based on NIST 2017 library (%)

[b] t_R (min) = retention time in minutes

hemicellulose contents in the hardwood are 39–53 and 19–36%, respectively, while lignin is 17–24%. Meanwhile, the percentages of cellulose and hemicellulose in the herbaceous plants are 24–50 and 12–38%, respectively, whereas lignin is 6–29% [24].

Based on Table 1, carbohydrates generate several classes of compounds during pyrolysis, such as anhydrous sugars, carbonyls, lactones, furans, pyrans, carboxylic acids, and esters [17]. The examples of anhydrous sugars are 1,4:3,6-dianhydro-α-D-glucopyranose and 2,3-anhydro-D-mannosan, whereas those of carbonyls are 2-oxopropanal and butane-2,3-dione.

During the Py-GCMS analysis, lignin is fragmented into its monomers: H (*p*-hydroxyphenyl unit), G (guaiacyl unit), and S (syringyl unit). The concentration of lignin and its monomeric composition change between plant species, tissues, cell types, and different cell wall layers during development [25]. Based on the relative abundance (%) of lignin monomers, among the 5 botanical parts of *E. crista-galli*, the twigs have the highest total lignin content (45.59%), followed by roots (23.56%), bark (18.26%), and leaves (18.03%), while flowers have the lowest total lignin content (5.56%). Lignin accumulates in the cell walls of specialized cell types to enable plants to stand upright and conduct water and minerals [26]. Twigs provide mechanical support and transport water, carbohydrates, and nutrients [27]. This explains why the twigs have the highest total lignin content among the other botanical parts of *E. crista-galli*.

Py-GC/MS provides a complete overview of global metabolite fingerprints to characterize botanical parts of *E. crista-galli*. Through pattern recognition analysis, the multivariate data obtained from Py-GC/MS analysis can be useful to provide information on how each botanical part of *E. crista-galli* is different from one another based on the metabolite fingerprint. Therefore, we coupled the Py-GC/MS results with multivariate analysis in the next step.

Multivariate analysis is concerned with datasets having several response variables for each observational or experimental. The commonly used multivariate data analysis for pattern recognition are PCA and HCA. These are examples of unsupervised learning techniques in which the objective is to identify previously unknown structures in the data set, as well as to identify clusters in a given dataset without using class membership information in the calculations [28].

Multivariate Analysis

Multivariate analysis with all pyrolysis products

PCA is a statistical method that can be used to visualize information in a data set by describing how each sample differs from another, which variables contribute significantly to this difference, as well as to identify sample patterns. In our research, in order to easily identify which metabolite contribute to the similarity or differences between 5 botanical parts of *E. crista-galli* based, we use the relative abundance data of metabolites as variables for PCA analysis, as done by several previous studies [29].

PCA minimizes the data dimension by creating the so-called principal components (PCs), which are linear combinations of the variables in the data set to

summarize the data [30]. Fig. 2(a). shows the scree plot, which is a line plot of the principal components along with the percentage of explained variance from the principal component analysis of the data set. Cross-validation was subjected to the data set in order to determine the number of PCs that should be retained in order to account for most of the data variability. The result from cross-validation suggests that at least the first three PCs should be retained to fulfill a variance of 80% (Fig. 2(b)).

Fig. 3 shows the score plot of the botanical parts of *E. crista-galli* on PC.1, PC.2, and PC.3. PC.1 accounts for 36.5% of the total variance, while PC.2 27.2% and the PC.3 20.3%. Together, the first three PCs explain 84% of the total variance. Each PC can be described by the origin variables (Rts). Variables described the best in each PC can be identified by the correlation coefficient and the coordinates of the botanical parts on a PC. Correlation coefficients are calculated for all the variables, followed by testing the significance of each correlation coefficient and sorting the variables from the most to the less correlated. The most significant variables then describe each PC; such a method is beneficial for interpreting the dimensions with many variables [20]. Table 2 shows a list of significantly correlated variables to PC.1, 2, and 3 from the PCA.

According to Table 2, eugenol, 4-ethylguaiacol, *trans*-isoeugenol, and (*E*)-4-(3-hydroxyprop-1-en-1-yl)-2-methoxyphenol are pyrolysates that have a positive correlation to PC.1. Therefore, samples with a high score

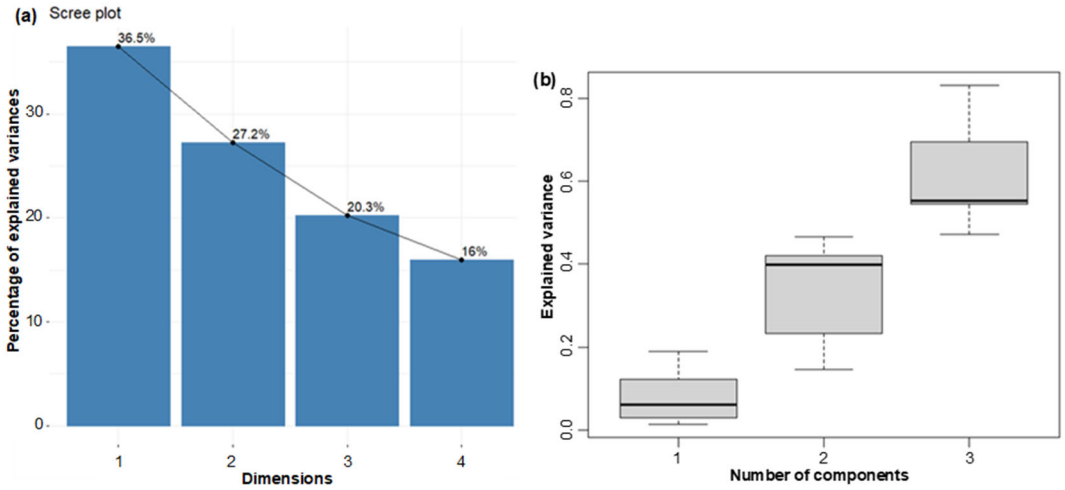

Fig 2. (a) Scree plot and (b) box plot of cumulative variances resulted from leave-one-out cross-validation

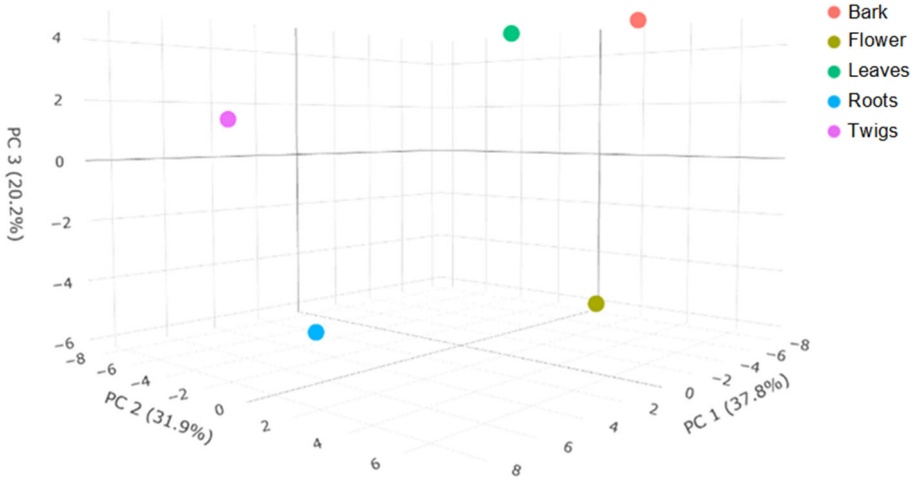

Fig 3. PCA score plot of botanical parts of *E. crista-galli* on PC.1, PC.2, and PC.3

Table 2. List of significantly correlated variables to PC.1, PC.2, and PC.3

t_r (min)	Corr.	*p*-value	Pyrolysis product	Origin
			PC.1	
20.821000	0.979400	0.003541	eugenol	Lignin-G
18.545000	0.978600	0.003747	4-ethylguaiacol	Lignin-G
23.508000	0.970400	0.006087	trans-isoeugenol	Lignin-G
30.973000	0.934900	0.019742	(*E*)-4-(3-hydroxyprop-1-en-1-yl)-2-methoxyphenol	Lignin G
9.731000	-0.992500	0.000776	phenol	Lignin-H
			PC.2	
6.268000	0.974000	0.005015	1-(acetyloxy)-2-propanone	Linear ketone derivatives
			PC.3	
26.422000	0.924300	0.024734	4-vinylsyringol	Lignin-S
13.370000	-0.966000	0.007472	3-methylbutyl 2-methylpropanoate	Linear ketone derivatives

on PC.1 will have a high relative abundance of these pyrolysis products. On the other hand, phenol has a negative correlation to PC.1. Thus, any sample with a high score on PC.1 will have a low relative abundance of that pyrolysis product. For PC.2, only 1-(acetyloxy)-2-propanone has a significant positive correlation to the second latent variable. Meanwhile, for PC.3, 4-vinylsyringol and 3-methylbutyl 2-methylpropanoate are positively and negatively correlated to the third latent variable, respectively.

The PCA score plot (Fig. 3) shows that roots have the highest score on PC.1, followed by twigs, bark, leaves, and flowers. Revering to Table 2, most of the significantly correlated variables on PC.1 come from the pyrolysis products of lignin G. Roots have the highest score on PC.1, while flowers have the lowest score. Thus, roots are characterized by high lignin G content, whereas flowers are low lignin G content, which is also confirmed by Table 1. Similarly, since the bark owns a high score on PC.2, it has the highest relative abundance of 1-(acetyloxy)-2-propanone, whereas twigs have the lowest relative abundance of this pyrolysis product.

Visualization provided by PCA score plots may facilitate clustering in pyrolysis product data. Nonetheless, PCA does not explicitly define clusters. More formal approaches can be used by clustering methods. Cluster analysis divides observations into groups that are related to one another. In terms of specific characteristics, each group or cluster is homogeneous and should be distinct from others. The closeness of two objects is expressed by similarity or dissimilarity, which can be computed by mathematical methods, and eventually displayed in a dendrogram based on the features of individual objects [30]. HCPC is a clustering approach that allows to combine principal component method, hierarchical clustering, and partitioning clustering method to identify clusters within a data set. The combination of the principal component method along with the clustering method is useful in a situation where the data set contains multiple continuous variables. The PCA can be used to reduce the dimension of the data, and then clustering can be performed on the PCA result [15].

From the PCA and LOOCV analysis, at least the first three PCs should be retained to cover 80% of the variance (Fig. 2). Therefore, we performed the HCPC analysis from the first three (84% total variance) and four principal components (100% total variance). Fig. 4 shows the dendrogram of botanical parts of *E. crista-galli* resulting from HCPC analysis.

HCPC analysis from the first three and four principal components shows that the botanical parts of *E. crista-galli* are divided into three different clusters. Fig. 4 show that in HCPC analysis with the first three PCs, cluster 1 consists of flower, cluster 2 consists of leaves and bark, and cluster 3 consists of twig and root, whereas in HCPC analysis with four PCs, Cluster 1 consists of flowers and leaves, cluster 2 consists of twigs and cluster 3 consists of roots and bark. Since the first four PCs cover 100% variability, HCPC analysis from the first four PCs is used to cluster botanical parts of *E. crista-galli*.

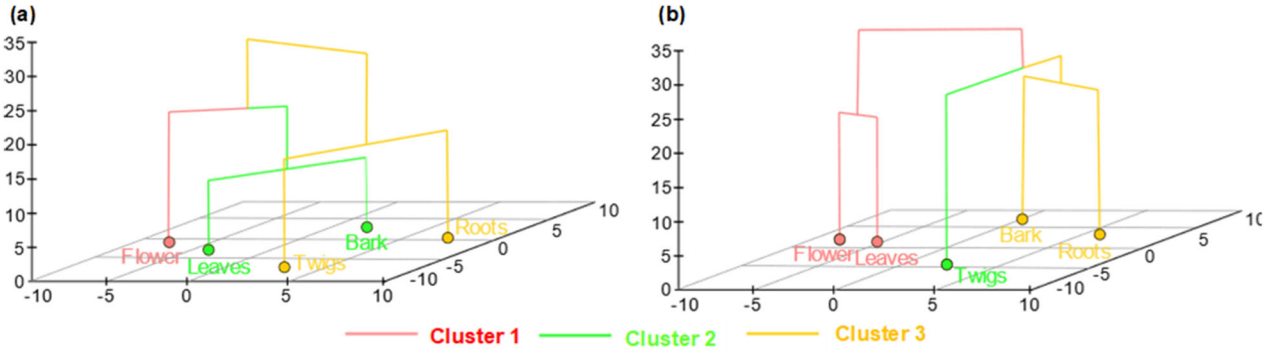

Fig 4. Hierarchical clustering on the factor map of botanical parts of *Erythrina crista-galli* with (a) three PCs and (b) four PCs. Clusters 1, 2, and 3 are denoted by pink, green, and yellow, respectively

Table 3. Variables that describe the most each cluster

t_R (min)	Mean in category	Overall mean	*p*-value	Pyrolysis product	Origin
			Cluster 1		
19.2620	1.3600	0.5440	0.046681	1*H*-indole	Unknown
			Cluster 2		
50.4040	0.5300	0.1060	0.045500	heptacosyl heptafluorobutyrate	Extractive/Unknown?
45.6880	1.3300	0.2660	0.045500	2,6,10,15,19,23-pentamethyl-2,6,18,22-tetracosatetraen-10,15-diol	Unknown
31.3200	0.4100	0.0820	0.045500	syringylacetone	Lignin-S
31.1100	0.9200	0.1840	0.045500	tetradecanoic acid	Linear ketone derivatives
11.9010	11.8900	2.7306	0.045795	2-methylphenol	Lignin-H
2.9180	4.8400	1.4280	0.046366	2,5-dihydrofuran	Furan derivatives
12.6420	12.5700	3.7718	0.046617	*p*-cresol	Lignin-H
2.4720	6.4700	2.4040	0.047257	2,3-butanedione	Linear ketone derivatives
			Cluster 3		
15.1300	0.5900	0.2360	0.045707	7-methyl-1,4-dioxaspiro[2.4]heptan-5-one	Lactone derivatives
17.0240	0.6250	0.2500	0.046065	(1*S*,3*R*,6*R*,7*R*,9*R*)-2,5,8-trioxatricyclo[4.2.1.03,7]nonan-9-ol	Anhydro sugars
4.6030	2.2950	1.4200	0.046812	methyl 2-oxopropanoate	Linear ketone derivatives
8.9510	0.4750	0.1900	0.048895	5-methylfuran-2-carbaldehyde	Cyclopentenone derivatives

Table 3 shows a list of variables that describe the most exact cluster. Variables that are significantly associated with specific clusters have higher mean category values than the overall mean. Thus, it could be said that cluster one (*i.e.,* flowers and leaves) is characterized by the higher content of 1*H*-indole pyrolysate. 1*H*-indole is assumed to be a minor pyrolysis product that originated from protein [31] or extractive as an alkaloid after fragmentation of the pyrolysis process and was detected by Py-GC/MS [32]. 1*H*-Indole is produced by the pyrolysis of the amino acid tryptophan. It undergoes thermal degradation at a temperature above 800 °C. Three main pyrolysates of indole are phenylacetonitrile, 2- methylbenzonitrile, and 3-methylbenzonitrile which formed due to the opening of the pyrrole ring [25]. Since, in our research, the pyrolysis was performed at the temperature of 500 °C, the indole might not undergo a pyrolytic reaction. That's why 1*H*-indole (retention time, t_R = 19.267 min) and 3-methyl-1-*H*-indole (t_R = 21.897 min) pyrolysate are still detected. Those pyrolysates might also indicate the presence of indole alkaloids such as 1*H*-indole-3-propanamide,

abrine, and hypaphorine (Fig. 5) that has been identified in *Erythrina* genus [33]. Since indole pyrolysate is associated with the presence of indole alkaloids, the flowers and leaves contain a higher amount of indole alkaloids compared to the other clusters.

Other pyrolysates that could indicate the presence of alkaloids are 6,7-dimethoxy-1-phenyl-3,4-dihydroisoquinoline (t_R = 44.976 min) and methyl 2-phenylquinoline-7-carboxylate (t_R = 48.284 min). Isoquinoline is one compound that is very stable at elevated temperatures. It undergoes pyrolysis at a temperature above 900 °C to produce benzene, toluene, naphthalene, phenanthrene, and anthracene, as well as the isomer of the other quinoline, indole, and several nitriles, including benzonitrile, and several isomers of cyanostyrene and cyanonaphthalene [25].

Cluster 2 (*i.e.,* twigs) is characterized by a higher relative abundance of heptacosyl heptafluorobutyrate, 2,6,10,15,19,23-pentamethyl-2,6,18,22-tetracosatetraen-10,15-diol, syringylacetone, tetradecanoic acid, 2-methylphenol, 2,5-dihydrofuran, *p*-cresol, and 2,3-butanedione pyrolysate. Heptacosyl heptafluorobutyrate and 2,6,10,15,19,23-pentamethyl-2,6,18,22-tetracosatetraen-10,15-diol were detected at the end of pyrogram as minor pyrolysis products from amino acids of lignocellulose biomass samples [34]. Cluster 3 (*i.e,* roots and barks) is characterized by a higher relative abundance of 7-methyl-1,4-dioxaspiro[2.4]heptan-5-one, (1*S*,3*R*,6*R*,7*R*,9*R*)-2,5,8-trioxatricyclo[4.2.1.03,7]nonan-9-ol, methyl 2-oxopropanoate, and 5-methylfuran-2-carbaldehyde pyrolysate.

Tables 2 and 3 show that the distribution of samples in the score plot of PC.1, PC.2, and PC.3, as well as the clustering, are mainly influenced by the polysaccharide and lignin content in those samples. This is mainly true since polysaccharides and lignin are relatively abundant compared to extractives in higher plants, whether in softwood, hardwood, or even in herbaceous plants [24].

Multivariate analysis with only extractive pyrolisate

The second principal component analysis was performed on the relative abundance (%) of extractive pyrolysates. Fig. 6 shows the score plot of the samples for the second PCA.

For the second principal component analysis, PC.1 and PC.2 account for 37.8 and 31.9% of the total variance, respectively, while PC.3 contributes to 20.2% of the total variance. Together the first three PCs account for 89.9% of the total variance. The analysis shows that phytol (t_R = 38.401 min) is the variable that significantly correlated to PC.1 (corr. = 0.938, p val. = 0.0179). Since bark has the highest score on PC.1, therefore it has the highest relative abundance of phytol.

Indole (t_R = 19.262 min) is the pyrolysate that is significantly correlated to PC.2 (corr. = −0.949, *p*-value = 0.0136) and the correlation of indole with PC.2 is negative. Thus, samples with the smallest score in PC.2 (*i.e.*, leaves and flowers) have the highest relative abundance of this pyrolysate. (*Z*)-18-Octadec-9-enolide, (8*Z*)-1-oxacycloheptadec-8-en-2-one, and (*Z*)-18-Octadec-9-enolide are pyrolysates that significantly correlated to PC.3 with correlation value of −0.925. Since the correlation value is negative, indicating samples that have a positive value on PC.3 will have a small relative abundance of those pyrolysates.

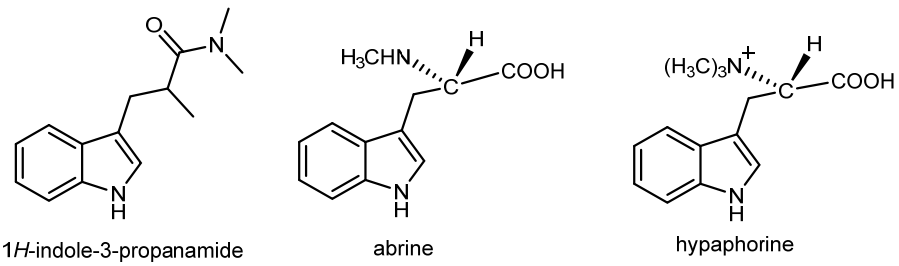

Fig 5. Indole alkaloids identified in *Erythrina*

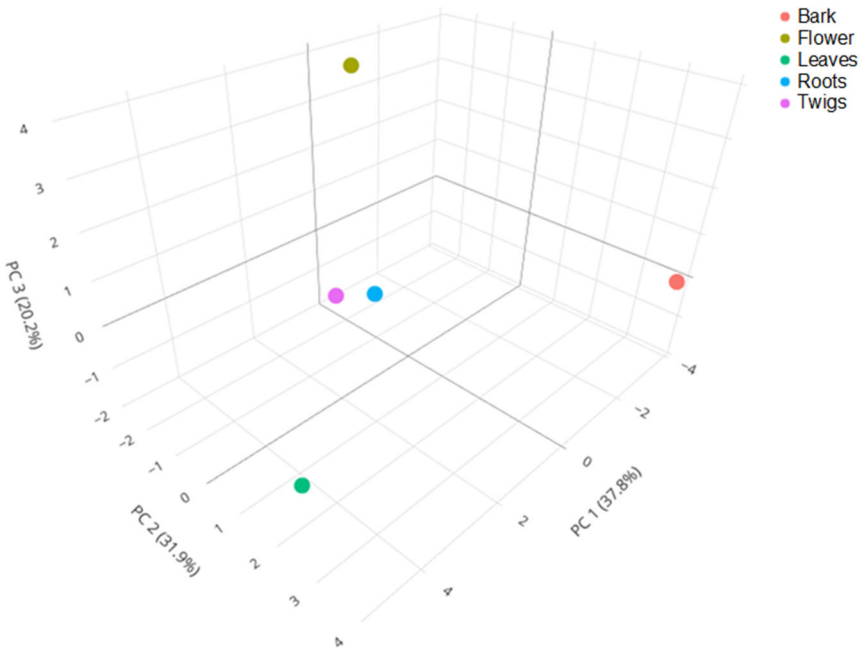

Fig 6. Score plot of botanical parts of *E. crista-galli* for the second principal component analysis

■ CONCLUSION

Py-GC/MS analysis can be used in conjunction with multivariate data analysis to characterize the botanical parts of *E. crista-galli*. The Py-GC/MS shows that most pyrolysis products or pyrolysate are originated from polysaccharides and lignin. PCA shows that the roots of *E. crista-galli* is characterized by the highest relative abundance of lignin G, while the flowers have the least relative abundance of lignin G. Hierarchical cluster analysis shows that the botanical parts of *E. crista-galli* are clustered in three different clusters based on their similarity. Cluster 1 consists of flowers and leaves and is characterized by the higher content of indole pyrolysate. Cluster 2 consist of twigs and characterized by higher relative abundance of heptacosyl heptafluorobutyrate, 2,6,10,15,19,23-pentamethyl-2,6,18,22-tetracosatetraen-10,15-diol, syringylacetone, tetradecanoic acid, 2-methylphenol, 2,5-dihydrofuran, *p*-cresol, and 2,3-butanedione pyrolysate, and cluster 3 consist of roots and barks is characterized with higher relative abundance of 7-methyl-1,4-dioxaspiro[2.4]heptan-5-one, (1*S*,3*R*,6*R*,7*R*,9*R*)-2,5,8-trioxatricyclo[4.2.1.03,7]nonan-9-ol, methyl 2-oxopropanoate, and 5-methylfuran-2-carbaldehyde pyrolysate.

■ ACKNOWLEDGMENTS

The authors are grateful to the Universitas Padjadjaran for providing funds through the Outstanding Padjadjaran Postgraduate Scholarships (BUPP) scheme and Academic Leader Grant (ALG) by Tati Herlina (No. 1959/UN6.3.1/PT.00/2021).

■ AUTHOR CONTRIBUTIONS

Conceptualization, Tati Herlina and Ari Hardianto; Data curation, Abd. Wahid Rizaldi Akili and Maya Ismayati; Formal analysis, Ari Hardianto, Abd. Wahid Rizaldi Akili, Jalifah binti Latip, Maya Ismayati, Tati Herlina; Funding acquisition, Tati Herlina; Investigation, Maya Ismayati, Abd. Wahid Rizaldi Akili; Methodology, Ari Hardianto, Tati Herlina, and Maya Ismayati; Software, Abd. Wahid Rizaldi Akili and Ari Hardianto; Validation, Maya Ismayati, Ari Hardianto, and Tati Herlina; Visualization, Abd. Wahid Rizaldi Akili and Ari Hardianto; Writing – original draft, Abd. Wahid Rizaldi Akili and Ari Hardianto; Writing – review & editing, Tati Herlina, Jalifah binti Latip and Ari Hardianto.

■ REFERENCES

[1] Tan, Q.W., Ni, J.C., Fang, P.H. and Chen, Q.J., 2017, A New Erythrinan alkaloid glycoside from the seeds

of *Erythrina crista-galli*, *Molecules*, 22 (9), 1558.

[2] Fahmy, N.M., Al-Sayed, E., El-Shazly, M., and Singab, A.N., 2018, Comprehensive review on flavonoids biological activities of *Erythrina* plant species, *Ind. Crops Prod.*, 123, 500–538.

[3] Kaushal, A., Sharma, M., Navneet, N., and Sharma, M., 2020, Ethnomedicinal, phytochemical, therapeutic and pharmacological review of the genus *Erythrina*, *Int. J. Bot. Stud.*, 5 (6), 642–648.

[4] Kim, D.S., Na, H., Kwack, Y., and Chun, C., 2014, Secondary metabolite profiling in various parts of tomato plants, *Korean J. Hortic. Sci. Technol.*, 32 (2), 252–260.

[5] Yang, H., Huang, Z., Huang, Y., Dong, W., Pan, Z., and Wang, L., 2015, Characterization of Chinese crude propolis by pyrolysis-gas chromatography/mass spectrometry, *J. Anal. Appl. Pyrolysis*, 113, 158–164.

[6] Youssef, F.S., Mamatkhanova, M.A., Mamadalieva, N.Z., Zengin, G., Aripova, S.F., Alshammari, E., and Ashour, M.L., 2020, Chemical profiling and discrimination of essential oils from six *Ferula* species using GC analyses coupled with chemometrics and evaluation of their antioxidant and enzyme inhibitory potential, *Antibiotics*, 9 (8), 518.

[7] Tanuwidjaja, I., Svečnjak, L., Gugić, D., Levanić, M., Jurić, S., Vinceković, M., and Fuka, M.M., 2021, Chemical profiling and antimicrobial properties of honey bee (*Apis mellifera* L.) venom, *Molecules*, 26 (10), 3049.

[8] Ikram, M.M.M., Ridwani, S., Putri, S.P., and Fukusaki, E., 2020, GC-MS based metabolite profiling to monitor ripening-specific metabolites in pineapple (*Ananas comosus*), *Metabolites*, 10 (4), 134.

[9] Duan, S.G., Hong, K., Tang, M., Tang, J., Liu, L.X., Gao, G.F., Shen, Z.J., Zhang, X.M., and Yi, Y., 2021, Untargeted metabolite profiling of petal blight in field-grown *Rhododendron agastum* using GC-TOF-MS and UHPLC-QTOF-MS/MS, *Phytochemistry*, 184, 112655.

[10] Qian, Y., Wang, Y., Sa, R., Yan, H., Pan, X., Yang, Y., and Sun, Y., 2013, Metabolic fingerprinting of *Angelica sinensis* during growth using UPLC-TOFMS and chemometrics data analysis, *Chem. Cent. J.*, 7 (1), 42.

[11] Picó, Y., and Barceló, D., 2020, Pyrolysis gas chromatography-mass spectrometry: Focus on organic matter and microplastics, *TrAC, Trends Anal. Chem.*, 130, 115963.

[12] Traoré, M., Kaal, J., and Martínez Cortizas, A., 2018, Chemometric tools for identification of wood from different oak species and their potential for provenancing of Iberian shipwrecks (16^{th}-18^{th} centuries AD), *J. Archaeol. Sci.*, 100, 62–73.

[13] Szczepanik, M., Szyszlak-Bargłowicz, J., Zając, G., Koniuszy, A., Hawrot-Paw, M., and Wolak, A., 2021, The use of multivariate data analysis (HCA and PCA) to characterize ashes from biomass combustion, *Energies*, 14 (21), 6887.

[14] Greenacre, M., Groenen, P.J.F., Hastie, T., D'Enza, A.I., Markos, A., and Tuzhilina, E., 2022, Principal component analysis, *Nat. Rev. Methods Primers*, 2 (1), 100.

[15] Wehrens, R., 2020, *Chemometrics with R Multivariate: Data Analysis in the Natural and Life Sciences*, Springer-Verlag, Berlin, Heidelberg.

[16] Madala, N.E., Piater, L.A., Steenkamp, P.A., and Dubery, I.A., 2014, Multivariate statistical models of metabolomic data reveals different metabolite distribution patterns in isonitrosoacetophenone-elicited *Nicotiana tabacum* and *Sorghum bicolor* cells, *Springerplus*, 3 (1), 254.

[17] Moldoveanu, S.C. 2020, *Analytical Pyrolysis of Natural Organic Polymers*, Elsevier Science, New York, US.

[18] Husson, F., Le, S., and Pagès, J., 2017, *Exploratory Multivariate Analysis by Example Using R*, 2^{nd} Ed., Chapman and Hall/CRC, New York, US.

[19] Husson, F., Josse, J., Le, S., and Mazet, F., 2016, Package 'factominer', *An R package*, 96, 698.

[20] Kassambara, A., 2016, *Practical Guide to Principal Component Methods in R*, STHDA.

[21] Wickham, H., 2016, *ggplot2: Elegant Graphics for Data Analysis*, Springer, Cham, Switzerland.

[22] Garcia, H., and Filzmoser, P., 2017, *Multivariate Statistical Analysis using the R package*

chemometrics, Vienna University of Technology, Austria.

[23] Kang, X., Kirui, A., Dickwella Widanage, M.C., Mentink-Vieger, F., Cosgrove, D.J., and Wang, T., 2019, Lignin-polysaccharide interactions in plant secondary cell walls revealed by solid-state NMR, *Nat. Commun.*, 10 (1), 347.

[24] Tarasov, D., Leitch, M., and Fatehi, P., 2018, Lignin-carbohydrate complexes: Properties, applications, analyses, and methods of extraction: A review, *Biotechnol. Biofuels*, 11 (1), 269.

[25] Decou, R., Labrousse, P., Béré, E., Fleurat-Lessard, P., and Krausz, P., 2020, Structural features in tension wood and distribution of wall polymers in the G-layer of *in vitro* grown poplars, *Protoplasma*, 257 (1), 13–29.

[26] Blaschek, L., Champagne, A., Dimotakis, C., Nuoendagula, C., Decou, R., Hishiyama, S., Kratzer, S., Kajita, S., and Pesquet, E., 2020, Cellular and genetic regulation of coniferaldehyde incorporation in lignin of herbaceous and woody plants by quantitative Wiesner staining, *Front. Plant Sci.*, 11, 00109.

[27] Yan, Z., Li, P., Chen, Y., Han, W., and Fang, J., 2016, Nutrient allocation strategies of woody plants: An approach from the scaling of nitrogen and phosphorus between twig stems and leaves, *Sci. Rep.*, 6 (1), 20099.

[28] Granato, D., Santos, J.S., Escher, G.B., Ferreira, B.L., and Maggio, R.M., 2018, Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective, *Trends Food Sci. Technol.*, 72, 83–90.

[29] Park, Y.J., Baek, S.A., Choi, Y., Kim, J.K., and Park, S.U., 2019, Metabolic profiling of nine *Mentha* species and prediction of their antioxidant properties using chemometrics, *Molecules*, 24 (2), 258.

[30] Ebeling, B., Vargas, C., and Hubo, S., 2013, Combined cluster analysis and principal component analysis to reduce data complexity for exhaust air purification, *Open Food Sci. J.*, 7 (1), 8–22.

[31] Kebelmann, K., Hornung, A., Karsten, U., and Griffiths, G., 2013, Intermediate pyrolysis and product identification by TGA and Py-GC/MS of green microalgae and their extracted protein and lipid components, *Biomass Bioenergy*, 49, 38–48.

[32] El Hayany, B., El Fels, L., Dignac, M.F., Quena, K., Rumpel, C., and Hafidi, M., 2021, Pyrolysis-GCMS as a tool for maturity evaluation of compost from sewage sludge and green waste, *Waste Biomass Valorization*, 12 (5), 2639–2652.

[33] Rambo, D.F., Biegelmeyer, R., Toson, N.S.B., Dresch, R.R., Moreno, P.R.H., and Henriques, A.T., 2019, The genus *Erithrina* L.: A review on its alkaloids, preclinical, and clinical studies, *Phytother. Res.*, 33 (5), 1258–1276.

[34] Chen, H., Xie, Y., Chen, W., Xia, M., Li, K., Chen, Z., Chen, Y., and Yang, H., 2019, Investigation on co-pyrolysis of lignocellulosic biomass and amino acids using TG-FTIR and Py-GC/MS, *Energy Convers. Manage.*, 196, 320–329.