Bukti Submit

[Molecules] Manuscript ID: molecules-1470133 - Article Processing Charge Confirmation

Inbox

Molecules Editorial Office <molecules@mdpi.com> Thu, Nov 4, 2021, 2:28 PM

to Anjar, me, Molecules

Dear Prof. Rohman,

Thank you very much for submitting your manuscript to Molecules:

Journal name: Molecules
Manuscript ID: molecules-1470133
Type of manuscript: Article
Title: Metabolite Fingerprinting using 1H-NMR Spectroscopy and Chemometrics for Classification of Three Curcuma Species from Different Origins
Authors: Laela Hayu Nurani, Abdul Rohman *, Anjar Windarsih, Any Guntarti, Florentinus Dika Octa Riswanto, Endang Lukitaningsih, Nurrulhidayah Ahmad Fadzillah, Mohamad Rafi
Received: 3 November 2021
E-mails: laelafarmasi@yahoo.com, abdulkimfar@gmail.com, anjarwindarsih2@gmail.com, any_guntarti@yahoo.co.id, dikaocta@uad.ac.id, lukitaningsih_end@ugm.ac.id, nurrulhidayah@iium.edu.my, mra@apps.ipb.ac.id
Submitted to section: Analytical Chemistry, https://www.mdpi.com/journal/molecules/sections/Analytical_Chemistry

We confirm that, if accepted for publication, the following Article Processing Charges (APC), 2000 CHF, will apply to your article:

Journal APC: 2000 CHF
Total APC: 2000 CHF

Please note that you may be entitled to a discount if you have previously received a discount code. Also note that reviewer vouchers must be applied before acceptance for publication. Vouchers can no longer be applied once an APC invoice has been issued. Reviewer vouchers, IOAP discounts, and vouchers offered by the Editorial Office cannot be applied to one invoice at the same time. You need to select one type of voucher to use. If you need to add any discount or replace the current discount with another type of discount, please contact the Molecules Editorial Office as soon as possible.

Please confirm that you support open access publishing, which allows unlimited access to your published paper and that you will pay the Article Processing Charge if your manuscript is accepted.

Thank you in advance for your cooperation. I look forward to hearing from you.

Kind regards,
Ms. Jelena Grujic
Assistant Editor, MDPI Belgrade

E-Mail: jelena.grujic@mdpi.com
Bulevar Mihajla Pupina 115, 11000 Belgrade, Serbia
+381 11 414 75 49

PAPER YANG DISUBMIT

Article

Metabolite Fingerprinting Using ^{1}H-NMR Spectroscopy and Chemometrics for Classification of Three Curcuma Species from Different Origins

Laela Hayu Nurani [1], Abdul Rohman [2,3,*], Anjar Windarsih [4], Any Guntarti [1], Florentinus Dika Octa Riswanto [5], Endang Lukitaningsih [2], Nurrulhidayah Ahmad Fadzillah [6] and Mohamad Rafi [7]

[1] Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta 55281, Indonesia; laelafarmasi@yahoo.com (L.H.N.); any_guntarti@yahoo.co.id (A.G.)
[2] Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia; lukitaningsih_end@ugm.ac.id
[3] Center of Excellence Institute for Halal Industry and Systems (IHIS), Universitas Gadjah Mada, Yogyakarta 55281, Indonesia
[4] Research Division for Natural Product Technology (BPTBA), National Research and Innovation Agency (BRIN), Yogyakarta 55861, Indonesia; anjarwindarsih2@gmail.com
[5] Division of Pharmaceutical Analysis and Medicinal Chemistry, Faculty of Pharmacy, Universitas Sanata Dharma, Yogyakarta 55282, Indonesia; dikaocta@uad.ac.id
[6] International Institute for Halal Research and Training (INHART), International Islamic University
Malaysia (IIUM), Kuala Lumpur 50728, Malaysia; nurrulhidayah@iium.edu.my
[7] Department of Chemistry, Faculty of Mathematics and Natural Sciences, IPB University, Bogor 16680,
Indonesia; mra@apps.ipb.ac.id

Citation: Nurani, L.H.; Rohman, A.; Windarsih, A.; Guntarti, A.; Riswanto, F.D.O.; Lukitaningsih, E.; Fadzillah, N.A.; Rafi, M. Metabolite Fingerprinting Using ^{1}H-NMR Spectroscopy and Chemometrics for Classification of Three Curcuma Species from Different Origins. *Molecules* **2021**, *26*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s): Davide Bertelli

Received: 03 November 2021
Accepted: 02 December 2021
Published: date

* Correspondence: abdulkimfar@gmail.com

Abstract: *Curcuma longa, Curcuma xanthorrhiza,* and *Curcuma manga* have been widely used for herbal or traditional medicine purposes. It was reported that turmeric plants provided several biological activities such as antioxidant, anti-inflammatory, hepatoprotector, cardioprotector, and anticancer activities. Authentication of the Curcuma species is important to ensure its authenticity and to avoid adulteration practices. Plants from different origins will have different metabolite compositions because metabolites are affected by soil nutrition, climate, temperature, and humidity. ^{1}H-NMR spectroscopy, principal component analysis (PCA), and orthogonal projections to latent structures-discriminant analysis (OPLS-DA) were used for authentication of *C. longa, C. xanthorrhiza,* and *C. manga* from seven different origins in Indonesia. From the ^{1}H-NMR analysis it was obtained that 14 metabolites were responsible for generating classification model such as curcumin, demethoxycurcumin, alanine, methionine, threonine, lysine, alpha-glucose, beta-glucose, sucrose, alpha-fructose, beta-fructose, fumaric acid, tyrosine, and formate. Both PCA and OPLS-DA model demonstrated goodness of fit (R^2 value more than 0.8) and good predictivity (Q^2 value more than 0.45). All OPLS-DA models were validated by assessing the permutation test results with high value of original R^2 and Q^2. It can be concluded that metabolite fingerprinting using ^{1}H-NMR spectroscopy and chemometrics provide a powerful tool for authentication of herbal and medicinal plants.

Keywords: authentication; curcuma; ^{1}H-NMR spectroscopy; chemometrics; metabolite
fingerprinting

1. Introduction

For hundreds of years, herbal medicines and their preparations have been widely used in folk medicines over the world. The preparation of herbal medicine preparations is typically presented either as single herbs or several herbs in a composite formulae, and it is reported that about 92% of herbal medicine formulas are a combination of less than 13 herbs [1] [2]. Annually, the market growth of herbal products has increased in which the raw material for most herbal products come from South Asian and Southeast Asian countries, including Indonesia [3]. Products of natural origin, such as supplements, herbal products, or herbal preparations, are increasingly widespread and used to maintain health or for the treatment of minor diseases. However, natural is not necessarily synonymous with safe. The adulteration of herbal preparations, together with contamination, sophistication, and degradation is a problem of global interest. In recent years, public awareness on herbal authentication and species admixtures in the raw herbal has increased significantly, because the adverse consequences of adulterated herbal components on consumer safety has been recognized [4]. Adulteration in herbal medicine involves replacing botanical materials, diluting high quality herbal medicines with lower grade ones, and mislabeling herbal medicine. Therefore, it is essential to have monitoring and pharmacovigilance systems [5].

For species authentication of herbal medicine, the World Health Organization (WHO), the United States Food and Drug Administration (USFDA), and the European Medicines Agency (EMEA) have regulated

that the identification of herbal medicines should be made to ensure their quality and to discriminate them from related species or adulterated samples [6]. Among herbal medicines components, Curcuma species including *C. longa* (turmeric), *C. xanthorrhiza* (Java Turmeric), and *C. manga* have been widely applied as medicinal plants for herbal or traditional medicine purposes [7]. These Curcumas have been reported to have some biological effects which are beneficial to human health including antioxidant, anticancer anti-inflammatory, hepatoprotector, cardioprotector, antibacterial activities, and wound healing [8] [9] [10].

Some analytical methods have been developed for analysis of Curcuma species. Most of the methods use chromatography-based methods such as high performance-thin layer chromatography (TLC) [11], high performance liquid chromatography (HPLC), ultra-high performance liquid chromatography (UHPLC) [12], gas chromatography-mass spectrometry (GC-MS) for analysis volatile compounds in Curcuma species [13], and liquid chromatography-mass spectrometry (LC-MS/MS) [14]. Chromatographic-based methods typically involved complex sample preparation technique and resulted huge number of responses which make difficulty in data analysis. Therefore, spectroscopic-based methods in combination with multivariate data analysis (MDA) or chemometrics were potential to be employed since this combination method was provided the way to analyze such an environmental big data [15]. Ultraviolet, visible, and vibrational spectroscopy (infrared and Raman) [16], [17], and NMR spectroscopy [18] are widely reported for authentication of Curcuma species. NMR spectroscopy offers some advantages for authentication of medicinal plants such as fast time analysis, simple in sample preparation, high reproducibility, and high robust. Moreover, NMR spectroscopy can be used for simultaneous analysis either primary or secondary metabolites comprehensively in certain samples [19,20]. Combined with chemometrics of multivariate analysis such as principal component analysis (PCA), partial least square-discriminant analysis (PLS-DA), and orthogonal projections to latent structures-discriminant analysis (OPLS-DA) which can manage the huge data generated from NMR measurement, it becomes a powerful analytical tool for metabolite fingerprinting of medicinal plants [21,22]. Combination of ^{1}H-NMR spectroscopy and chemometrics of PLS-DA and OPLS-DA has been used for authentication of Saffron adulteration [23]. ^{1}H-NMR spectroscopy and chemometrics have also been used for authentication of *C. longa* adulterated with *C. manga* and *C. heyneana* [7,24]. Authentication of *C. xanthorrhiza* from *C. aeruginosa* has been successfully investigated using ^{1}H-NMR and multivariate analysis [18]. However, study on authentication of Curcuma species from different origins using ^{1}H-NMR spectroscopy is still limited. Therefore, the objective of this study was to use ^{1}H-NMR spectroscopy in combination with chemometrics for authentication of *C. longa, C. xanthorrhiza,* and *C. manga* from different origins.

2. Results and Discussion

2.1.1. H-NMR Spectra Analysis

^{1}H-NMR spectra can be used for authentication of medicinal plants because it offers fingerprinting which mean that each sample has specific ^{1}H-NMR spectra pattern. Generally, metabolites of plants extracted using deuterated methanol and deuterium oxide measured using ^{1}H-NMR

spectroscopy are divided into three main regions, namely amino acid and organic acids (0.20–3.00 ppm), carbohydrate or sugar (3.01–5.00 ppm), and aromatic compounds (6.00–8.00 ppm) [25]. Different origins have different conditions such as soil condition, soil nutrition (macro and micronutrients), humidity, light, salinity, and temperature as well as internal developmental genetic circuits including regulated gene, and enzyme which can obviously affect the metabolite formation either primary or secondary metabolites [26]. The ^{1}H-NMR spectra of *C. longa* (CL), *C. xanthorrhiza* (CX), and *C. manga* (CM) are shown in Figure 1. It can be observed that *C. longa*, *C. xanthorrhiza*, and *C. manga* have different spectra pattern indicating different metabolite contents. Specifically observed, *C. longa* and *C. xanthorrhiza* have higher signal intensities in the region of amino acid and organic acid (0.20–3.00 ppm) as well as in the aromatic region (6.00–8.00 ppm) than *C. manga*. On the other hand, the signal intensities in the region of glucose (3.01–5.00 ppm) are higher in *C. manga* compared to *C. longa* and *C. xanthorrhiza*. Fourteen metabolites in Curcuma species obtained from ^{1}H-NMR measurement are shown in Table 1. Curcumin, demethoxycurcumin, alanine, methionine, threonine, lysine, alpha-glucose, beta-glucose, sucrose, alpha-fructose, beta-fructose, fumaric acid, tyrosine, and formate were stated as metabolites which play important roles in generating an OPLS-DA model. Investigation on each species obtained from different regions resulted in different signal patterns, especially in intensities, indicating the variations in metabolite contents in each species from different origins. It indicated that different origins affect the metabolite contents in each Curcuma rhizome. For example, *C. longa* from Blitar (CL7) has the lowest signal intensities in the aromatic region and *C. xanthorrhiza* from Gunungkidul (CX2) has the lowest signal intensities in the whole regions. However, the spectra patterns of each species are quite similar, therefore, for deeper classification of *C. longa*, *C. xanthorrhiza*, and *C. manga* from different regions powerful statistical tool such as chemometrics is required to obtain clear classification.

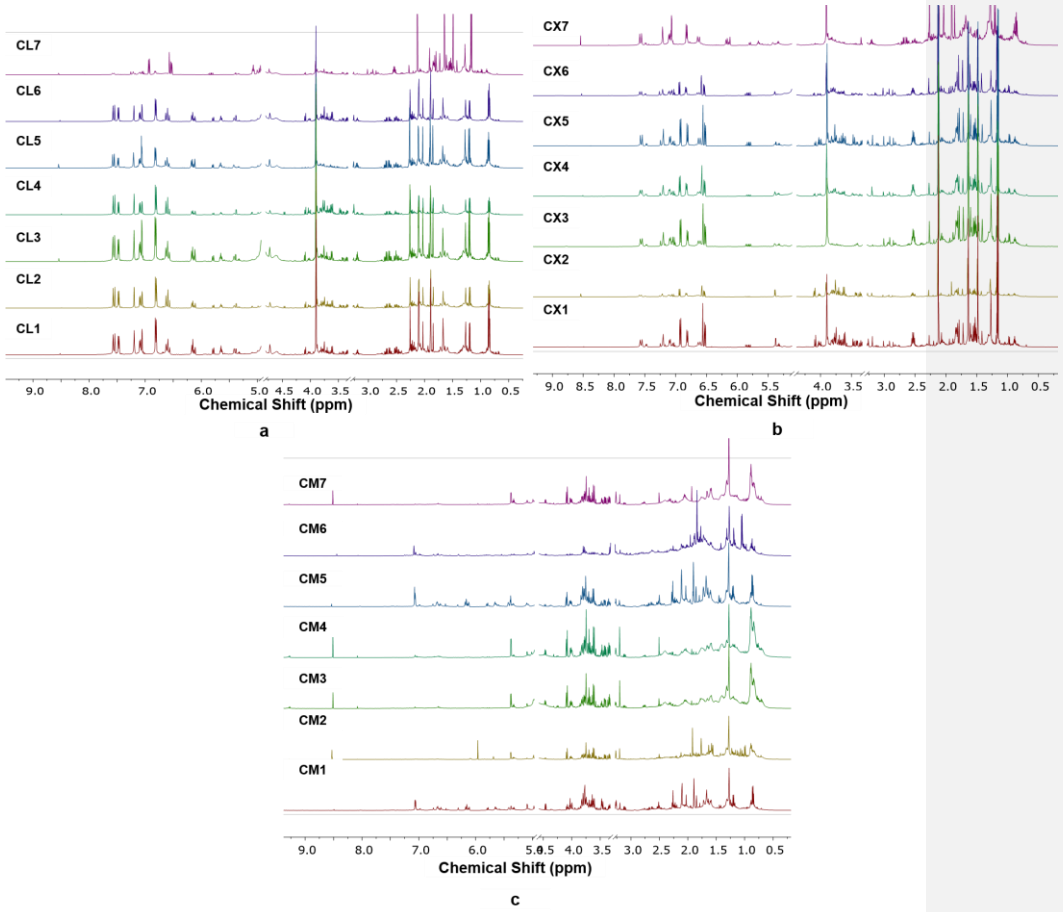


Figure 1. ^{1}H-NMR spectra of *C. longa* (**a**), *C. xanthorrhiza* (**b**), and *C. manga* (**c**) from different origins.

Table 1. Several metabolites of Curcuma species observed using ^{1}H-NMR spectra obtained from this study. The assignment of the metabolites refers to the previous published literature by Jung et.al and Awin et al. [27, 28].

No.	Chemical Shift (ppm)	Multiplicity	Metabolite
1.	7.57	Singlet	Curcumin
	7.28	Singlet	
	7.21	Doublet	
	6.77	Doublet	
	3.68	Singlet	
2.	6.92	Doublet	Demethoxycurcumin
	5.89	Singlet	

	3.94	Singlet	
3.	1.49	Doublet	Alanine
	3.72	Quartet	
4.	2.11	Singlet	Methionine
5.	1.33	Doublet	Threonine
	3.53	Doublet	
6.	3.81	Triplet	Lysine
	1.5	Multiplet	
7.	5.19	Doublet	Alpha-Glucose
	3.46	Doublet of Doublet	
	3.67	Triplet	
	3.35	Triplet	
8.	4.59	Doublet	Beta-Glucose
	3.19	Doublet of Doublet	
	3.44	Triplet	
	3.71	Doublet of Doublet	
9.	5.42	Doublet	Sucrose
	3.74	Triplet	
	3.43	Triplet	
	3.80	Multiplet	
	3.84	Multiplet	
10.	4.07	Doublet	Alpha-Fructose
	3.82	Doublet of Doublet	
	3.53	Doublet	
	3.55	Doublet	
	3.63	Quartet	
11.	3.95	Multiplet	Beta-Fructose
	3.52	Doublet	
	4.02	Doublet of Doublet	
12.	6.57	Singlet	Fumaric acid
13.	6.81	Doublet	Tyrosine
	7.14	Doublet	

14.	8.42	Singlet	Formate

Curcuminoids have been reported as the active compound in Curcuma species. The content of curcuminoids is varied among Curcuma species and it is reported that curcuminoid content in *C. longa* and *C. xanthorrhiza* is higher among other Curcuma species. Curcuminoids consist of curcumin, demethoxycurcumin, and bisdemethoxycurcumin which curcumin possess the highest concentration. However, not all Curcuma species contain these three types of curcuminoids, for instance *C. xanthorrhiza* does not contain bisdemethoxycurcumin. Curcuminoids are aromatic molecules therefore most of the signals appeared in the chemical shift of aromatic regions. Curcumin signal could be observed in the chemical shift of 7.57 ppm (singlet), 7.28 ppm (singlet), 7.22 ppm (doublet), 6.77 ppm (doublet), and 3.90 ppm (singlet) whereas demethoxycurcumin could be found in the chemical shift of 6.92 ppm (doublet), 5.89 ppm (singlet), and 3.94 ppm (singlet) [28]. From the ^{1}H-NMR spectra, higher signal intensities in the aromatic region of *C. longa* and *C. xanthorrhiza* supports that curcuminoids content in *C. longa* and *C. xanthorrhiza* is higher than in *C. manga*.

2.2. Chemometrics Analysis

Rhizomes of *C. longa*, *C. xanthorrhiza*, and *C. manga* are often used in a powder form as well as in an extract form for their herbal and traditional medicine applications. Both powder and extract are susceptible to adulteration because of their similar appearance especially in the adulterated form it is challenging to state whether the unknown sample is authentic or adulterated [29]. Chemometrics of PCA could not differentiate *C. longa*, *C. xanthorrhiza*, and *C. manga* clearly (data not shown). It might be caused by the large variations of the variables; therefore, the principal components (PC) were not able to represent the original variables. Observation using supervised pattern recognition, namely PLS-DA using 7 PC, could classify *C. longa, C. xanthorriza*, and *C. manga* resulting in three different classifications. However, several misclassifications occurred between *C. longa* and *C. xanthorrhiza* (Figure 2a). In the PLS-DA score plot, several *C. longa* samples appear in the region of *C. xanthorrhiza* and several *C. xanthorrhiza* samples appear in the region of *C. longa*. It can be explained that some of the metabolite compositions of *C. longa* and *C. xanthorrhiza* were similar especially in curcuminoid contents in which curcumin and demethoxycurcumin were the major active compounds in *C. longa* and *C. xanthorrhiza*. In addition, it is often reported that adulteration or substitution of *C. longa* with *C. xanthorrhiza* is often difficult to detect because the appearance of *C. longa* and *C. xanthorrhiza* in powder and extract form are quite similar [17]. Therefore, another supervised pattern recognition chemometrics, namely OPLS-DA, was performed to obtain better classification of *C. longa, C. xanthorrhiza,* and *C. manga* extracts. The OPLS-DA model demonstrated good capability to differentiate three different species of *C. longa*, *C. heyneana*, and *C. manga* from different origins as shown in the OPLS-DA score plot (Figure 2b). The OPLS-DA model successfully classified *C. longa*, *C. xanthorrhiza*, and *C. manga* samples. The samples were successfully classified using first PC and first X-orthogonal components which accounted for 80.8% of the variance with R^2X (cum) of 0.808, R^2Y (cum) of 0.776, and Q^2 (cum) of 0.767. A high value of R^2X (cum) and R^2Y (cum) (close to 1) indicated goodness of fit of the OPLS-DA model,

whereas the value of Q^2 greater than 0.45 indicated goodness of predictivity of the models [30]. The S-line correlation plot (Figure 2c) variables which have roles in the differentiation of *C. longa*, *C. xanthorrhiza*, and *C. manga*. It was found that alanine, curcumin, demethoxycurcumin, fumaric acid, sucrose, and tyrosine had p (corr) values of more than 0.5, indicating their important roles in separating samples. Moreover, using variable importance in projection (VIP) value, chemical shifts of 6.77, 3.89, 7.57, 6.81, 6.57, 7.21, 1.49, 6.49, 6.13, 0.85, 6.09, 5.29, 5.25, and 6.92 ppm were found to have important roles for the classification between three Curcuma species in OPLS-DA models. Variables with VIP values greater than 1 are considered to have important roles for differentiation. Some of the variables correspond to the metabolites of curcumin, tyrosine, fumaric acid, alanine, and demethoxycurcumin. The receiver operating characteristic curve (Figure 2d) for differentiating and classifying Curcuma longa, Curcume xanthorrhiza, and Curcuma manga from different origins was also depicted. The ROC analysis represents the probability of the model by plotting the value of true positivity rate (TPR) against the value of false positivity rate (FPR) [31].

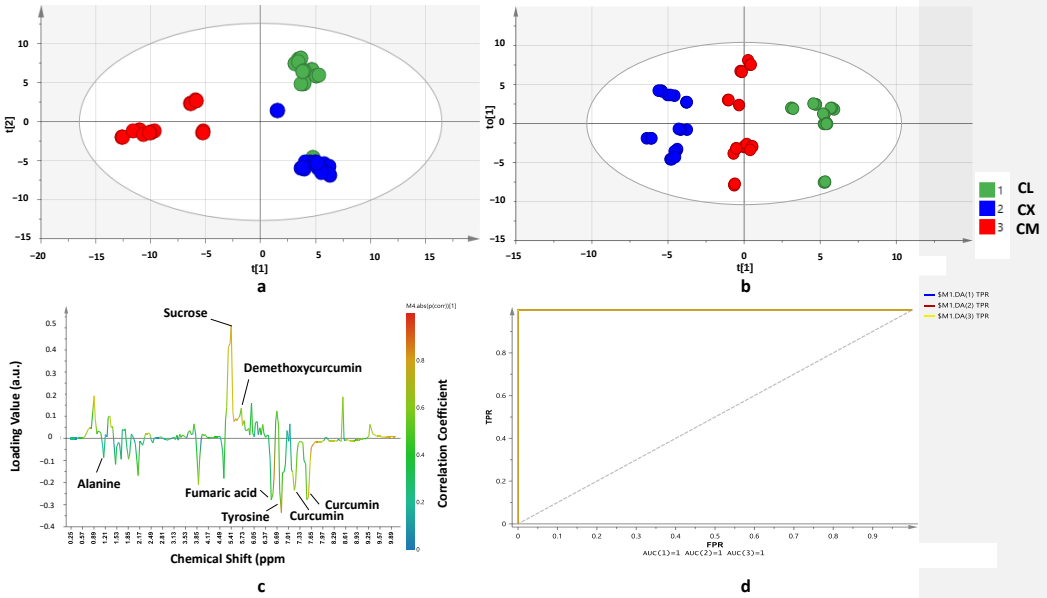


Figure 2. PLS-DA score plot (**a**), OPLS-DA score plot (**b**), OPLS-DA S-line correlation plot (**c**), and ROC curve (**d**) for differentiation and classification *of C. longa, C. xanthorrhiza,* and *C. manga* from different origins. Statistical parameters of the models: (a) Number of samples = 42; R^2X = 0.603; R^2Y = 0.783; Q^2 = 0.755 for t[1] and t[2] components; (b,c) Number of samples = 42; R^2X = 0.808; R^2Y = 0.776; Q^2 = 0.767 for t[1] and to[1] components.

PCA using number of PC 8 could differentiate *C. longa* from seven different origins as shown in the PCA score plot (Figure 3a). The PCA model provided high confidence for its fitting and predictivity capacity, shown by its R^2 value (0.770) and Q^2 value (0.650) for PC1 and PC2, respectively, accounting for 77.0% of the variance. The score plots which

appear close to each other indicates high similarity between samples, especially their metabolite compositions. The PCA score plot result shows that CL2 and CL4 possessed high similarity and CL1 has high similarity with CL6. Meanwhile, CL3 appeared closely to CL5. CL7 appeared far from all CL samples from other regions meaning that the metabolites composition of CL7 differs from other *C. longa* used in this research. Classification of *C. longa* samples from seven different origins using OPLS-DA demonstrated different pattern with PCA result (Figure 3b). OPLS-DA was created using first PC and first orthogonal-X component resulting R^2X (cum) of 0.748, R^2Y (cum) of 0.754 and Q^2 (cum) of 0.639. The first PC and first X-orthogonal component explained 74.8% of the total variance. There were three main groups obtained from OPLS-DA classification. The first group was CL2 and CL7 which appeared close to each other. The second group consisted of CL1 and CL5, and the last group of CL3, CL4, and CL6. The important variables for differentiation and classification of *C. longa* between groups observed using S-line correlation plot (Figure 3c) were alanine, β-fructose, curcumin, demethoxycurcumin, fumaric acid, and tyrosine. Meanwhile, investigation using VIP value found that variables of 6.92, 6.54, 6.57, 1.50, 1.66, 1.54, 1.18, 7.46, 6.62, 0.86, 7.10, 7.50, 7.57, and 7.14 ppm had important roles for *C. longa* differentiation and classification from seven different origins classified using OPLS-DA model. Some of the variables correspond to the molecule signals of curcumin, demethoxycurcumin, fumaric acid, lysine, and tyrosine. It demonstrated that different origins affect the composition of some metabolites. It is in accordance with research by Jung et al. [27] on the metabolite compositions of *C. longa* from several regions in China. The condition of geographical origin and environmental conditions such as temperature, humidity, and rainfall rate affect the metabolite composition of plants.

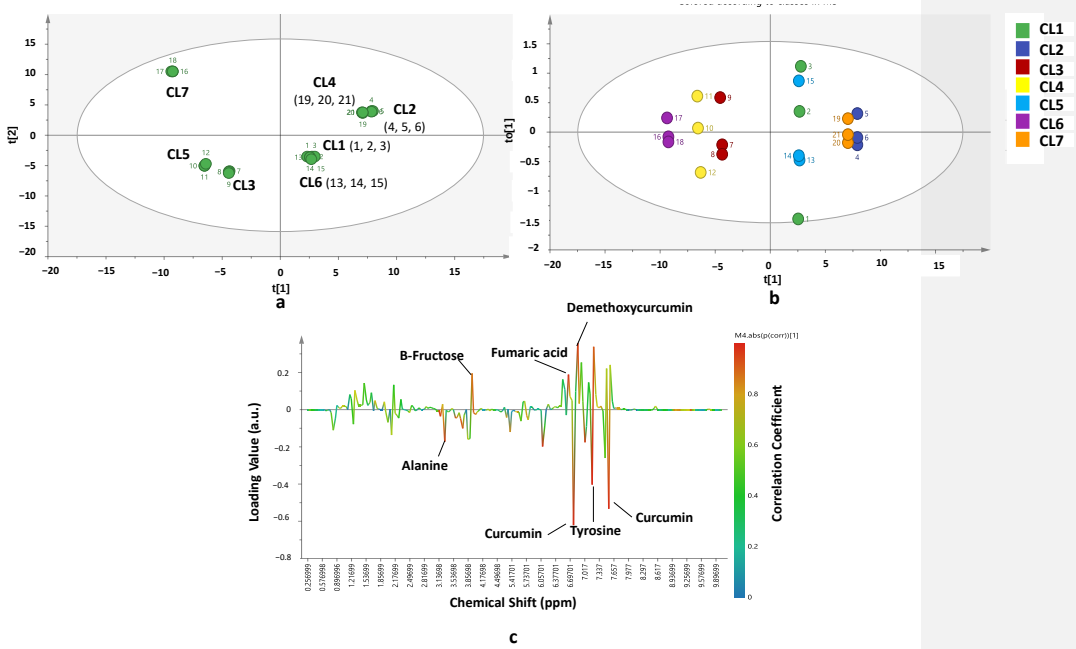


Figure 3. PCA score plot (**a**) OPLS-DA score plot (**b**) and OPLS-DA S-line correlation plot (**c**) for differentiation and classification *of C. longa* from different origins. Statistical parameter of the models: (a,) Number of samples = 21; Number of PC = 8; $R^2X = 0.770$; $Q^2 = 0.650$; (b,c) Number of samples = 21; $R^2X = 0.748$; $R^2Y = 0.754$; $Q^2 = 0.639$ for t[1] and to[1] components.

Chemometrics of PCA and OPLS-DA was also successfully used for differentiation and classification of *C. xanthorrhiza* from seven different origins. The result of differentiation of samples using PCA was slightly different with the result from OPLS-DA. PCA performed using first and second principal components demonstrated goodness of fit (R^2cum = 0.743) and predictivity (Q^2cum = 0.678) with total variance of 74.3%.Meanwhile, OPLS-DA was created using first PC and first orthogonal-X components. A high value of R^2X (cum) (0.743) and R^2Y (cum) (0.833) indicated good model fitness while a high value of Q^2 (cum) (0.626) indicated predictivity of the OPLS-DA model. The first PC and first X-orthogonal-X component demonstrated 74.3% of the total variance. There were four main groups which appeared in the PCA score plot (Figure 4a), whereas in OPLS-DA (Figure 4b) there were three groups. *C. xanthorrhiza* of CX3 and CX5 appeared close to each other in both PCA and OPLS-DA score plot results. Meanwhile, *C. xanthorrhiza* of CX1 was found in a separate group with others observed both in PCA and OPLS-DA. Samples of CX2, CX4, CX6, and CX7 appeared in the same group observed using OPLS-DA; however, from the PCA result, a sample of CX7 appeared in a different group. From these results, it is suggested that different locations have significant effects on metabolites' compositions in *C. xanthorrhiza*. The S-line correlation plot (Figure

4c) shows that methionine, β-glucose, sucrose, fumaric acid, curcumin, demethoxycurcumin, and tyrosine were the important variables for CX differentiation. Moreover, using a VIP value, it can be found that sthe variables important for classifying samples were 4.61, 6.57, 6.92, 4.59, 2.13, 6.92, 7.01, 7.21, 6.53, 1.69, 1.25, and 1.77 ppm. Some of the variables corresponded to the metabolites of curcumin, fumaric acid, demethoxycurcumin, and beta-glucose. It is presumed that these metabolites have higher scores and significantly affect the differentiation and classification of *C. xanthorrhiza* from different origins.

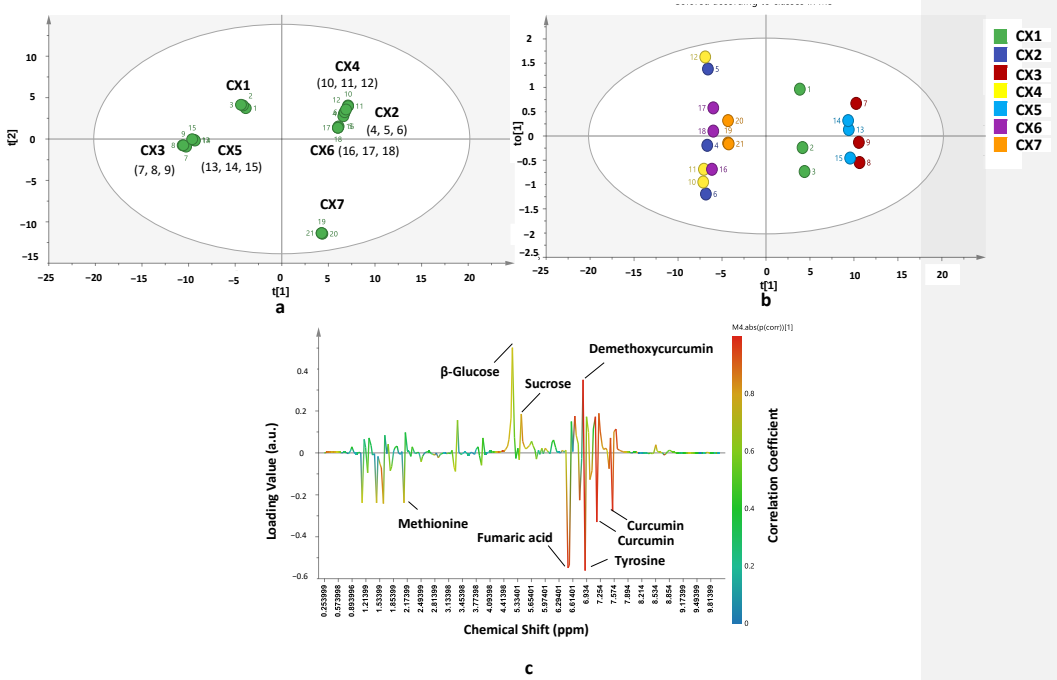


Figure 4. PCA score plot (**a**), OPLS-DA score plot (**b**), and OPLS-DA S-line correlation plot (**c**) for differentiation and classification *of C. xanthorrhiza* from different origins. Statistical parameter of the models: (a) Number of samples = 21; Number of PC = 6; $R^2X = 0.743$; Q2 = 0.678; (b,c) Number of samples = 21; $R^2X = 0.743$; $R^2Y = 0.833$; $Q^2 = 0.626$ for t[1] and to[1] components.

Chemometrics of PCA and OPLS-DA was also successfully applied for differentiation and classification of *C. manga* from seven different origins. Different classification results were observed between PCA and OPLS-DA. The PCA model was created using first and second principal components resulting R^2X (cum) of 0.661 and Q^2 (cum) of 0.501 indicating goodness of fit and good predictivity of the PCA model, respectively. The first and second PCs showed 66.1% of the total variance. From the PCA score plot (Figure 5a), *C. manga* were classified in five classes as follows: CM1 (first class), CM5 (second class), CM6 (third class), CM2 (fourth class), and the rest of the samples was in the last class (CM3, CM4, and CM7). Samples of CM3, CM4, and CM7 have similar chemical or

metabolite compositions because they appeared in the same location in PCA score plot. It is presumed that the conditions in the region of Malang (CM3), Tulung Agung (CM4), and Blitar (CM7) are similar resulting in the similar metabolites of *C. manga* rhizomes. On the other hand, OPLS-DA was performed using first principal components and first orthogonal-X component which presented 66.1% of the total variance. The obtained R^2X (cum) (0.661) and R^2Y (cum) (0.667) indicated goodness of fit whereas the value of Q^2 (cum) (0.707) demonstrated goodness of model predictivity. Three main groups were found in the OPLS-DA score plot (Figure 5b), namely CM1 and CM6 as the first group, CM5 in the second group, and CM2, CM3, CM4, and CM7 in the last group. Observation using an S-line correlation plot (Figure 5c) demonstrated that β-glucose, sucrose, curcumin, tyrosine, and format had important roles for CM differentiation. The VIP value showed that some variables were found to have significant contributions in the differentiation of *C. manga* samples from different origins, namely: 7.05, 5.29, 5.97, 5.33, 5.37, 8.53, 1.25, 1.81, 5.42, 5.65, 7.09, 6.17, 6.69, 0.85, 0.89, 3.33, and 5.41 ppm. Some of the variables are associated with curcumin, demethoxycurcumin, sucrose, and fumaric acid.

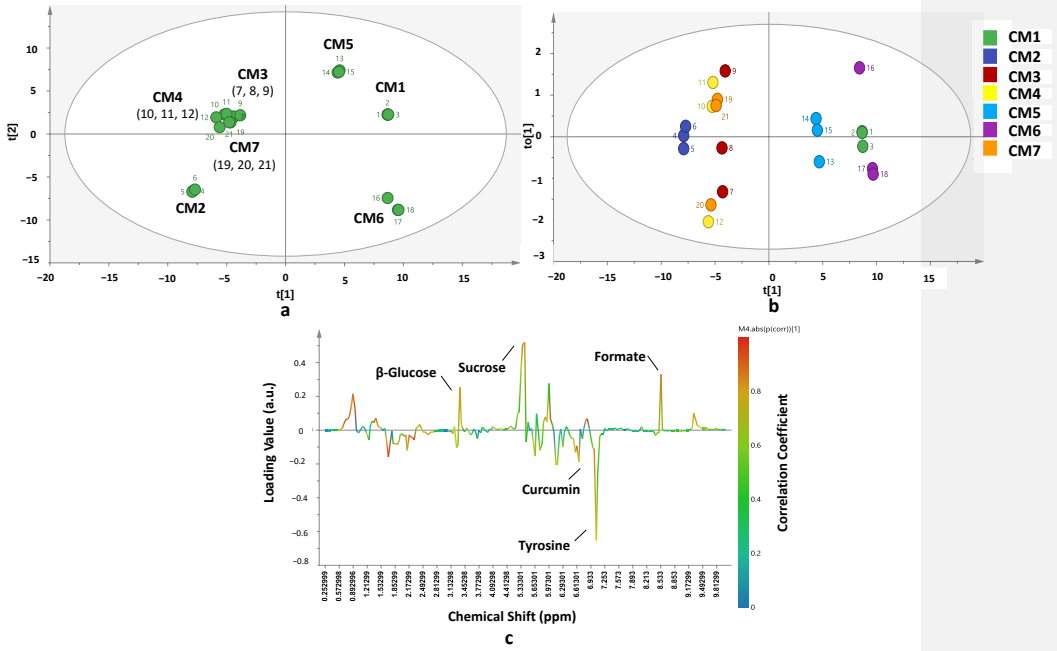


Figure 5. PCA score plot (**a**) OPLS-DA score plot (**b**) and OPLS-DA S-line correlation plot (**c**) for differentiation and classification *of C. manga* from different origins. Statistical parameter of the models: (a) Number of samples = 21; Number of PC = 7; R^2X = 0.661; Q^2 = 0.501; (b,c) Number of samples = 21; R^2X = 0.661; R^2Y = 0.667; Q^2 = 0.707 for t[1] and to[1] components.

2.3. Validation of OPLS-DA Model. Using Permutation Test.

Supervised pattern recognition of chemometrics such as PLS-DA and OPLS-DA requires a test to confirm the model's validity because of

its potential for overfitting. Validation is a confirmation step to ensure that the models have goodness of fit. A permutation test is one of validation testing which used a permutated model. Models of R^2 and Q^2 are permutated and compared to the original models of R^2 and Q^2. A good model is obtained when all the permutated models of R^2 and Q^2 values are lower than the R^2 and Q^2 original values. Moreover, the validation was also determined using intersection value of Q^2. The intersection value should be zero or lower than zero to be categorized as valid models. The result of the permutation test from 999 permutations of OPLS-DA models were demonstrated in Figure 6. The permutated models of R^2 and Q^2 are on the left side while the original R^2 and Q^2 models are on the right side. The models were permutated for 100 permutations. Results showed that all permutation tests confirmed the validity of the OPLS-DA model demonstrated by the value of R^2 and Q^2 in all permutated models being below the value of original R^2 and Q^2 models. On the other hand, the intersection values of Q^2 for all four OPLS-DA models were also zero and lower than zero, as follows: (0.0, −0.473) for a classification model between three Curcuma species; (0.0, −1.02) for classification model of *C. longa* from different origins; (0.0, −0.896), for classification model of *C. xanthorrhiza* from different origins; and (0.0, −0.904) for classification model of *C. manga* from different origins. It is suggested that OPLS-DA could be used as a powerful statistical tool for classification of different Curcuma species from different origins with high validity.

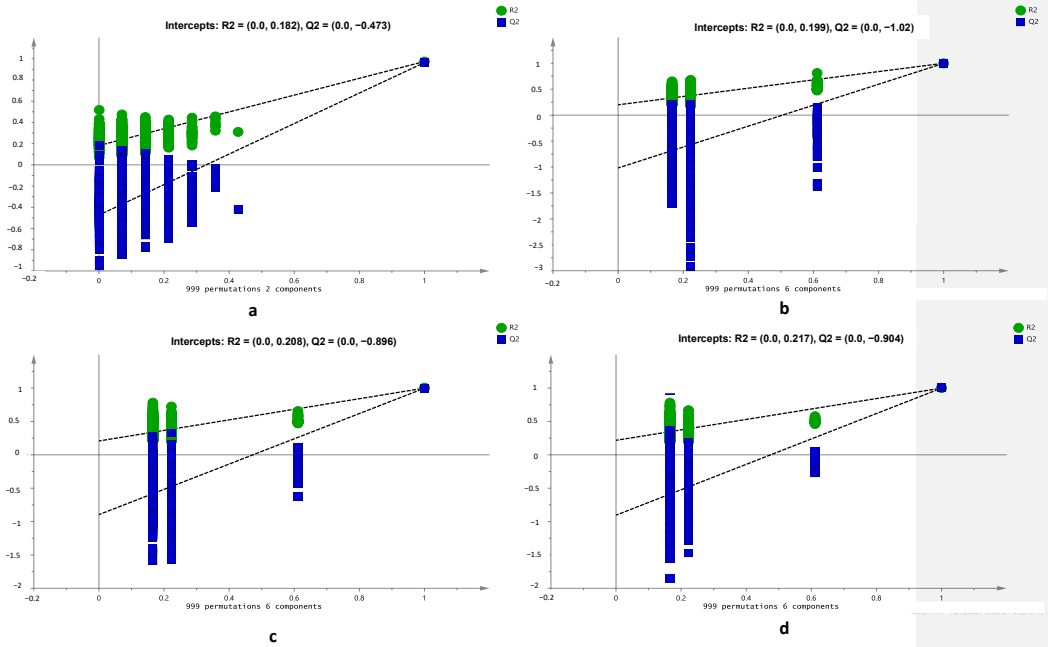


Figure 6. Permutation test using 999 permutations of OPLS-DA result from three Curcuma species (**a**), *C. longa* from different origins (**b**), *C. xanthorrhiza* from different origins (**c**), and *C. manga* from different origins (**d**).

3. Materials and Methods

3.1. Sample Collection and Preparation

Rhizome of *C. longa* (CL), *C. xanthorrhiza* (CX), and *C. manga* (CM) were collected from seven different regions in Indonesia, namely: Boyolali (1), Gunungkidul (2), Ngawi (3), Malang (3), Tulung Agung (5), Karang Anyar (6), and Blitar (7). Determination of plant species used in this study has been carried out at the Pharmaceutical Biology Department, Faculty of Pharmacy, Universitas Gadjah Mada, Indonesia. Rhizomes were cleaned using running water then chopped into small pieces. Subsequently, the rhizomes were dried using an oven at 50 °C for 48 h. The dried rhizomes were then ground into powder.

3.2. Preparation of Curcuma Rhizome Methanolic Extract

The powdered rhizome of *C. longa*, *C. xanthorrhiza*, and *C. manga* were extracted using methanol pro analysis using sample to solvent ratio of 1:10. Extraction was performed using a maceration technique for 3 days. The supernatant was collected and evaporated using a vacuum rotary evaporator to obtain a concentrated methanolic extract.

3.3. ^{1}H-NMR Analysis

Sample preparation was carried out according to Kim et al. [22] with modifications. An amount of 5 mg extract was weighed and placed into a 2 mL microtube. Subsequently the extract was added to 0.5 mL of deuterated methanol (CD_3OD) and 0.5 mL of deuterium oxide (D_2O) containing TMSP (trimethylsilyl propionic acid) 0.01%. The mixture was vortexed for 30 s and ultrasonicated for 20 min at room temperature. The sample was then centrifuged at 12,000 rpm for 10 min at room temperature. An amount of 800 μL of the supernatant was taken and transferred into an NMR tube. The sample was measured using a JEOL ECZ-R 500 MHz NMR spectrometer (JEOL, Tokyo, Japan). The NMR spectra acquisition was performed with the field strength of 11.74736 T, relaxation delay of 5s, and X_offset of 5.0 ppm. Each spectrum was acquired for a 3.53 min acquisition time which consisted of 128 scans and a width of 12 ppm. Each sample was measured in three replicates.

3.4. Data Analysis

The ^{1}H-NMR spectra were analyzed using MestreNova 12.0 Software (Mestrelab Research, S.L., Santiago de Compostela, Spain). Spectra were manually phase-corrected. Automatic baseline correction was performed using polynomial fit using degree of 3. The binning of the spectra was then performed for every 0.04 ppm from the chemical shift of 0.2–10 ppm excluding the region of residual water and methanol. Meanwhile, the chemometrics of multivariate analysis was performed using SIMCA 14.0 (Umetrics, Umeå, Sweden) software.

3.5. Chemometrics Analysis

Chemometrics of pattern recognition were used to analyze the data obtained from NMR measurements, namely principal component analysis (PCA), partial least square-discriminant analysis (PLS-DA), and orthogonal projections to latent structures-discriminant analysis (OPLS-DA). The data were processed using MestreNova 12.0 software (Mestrelab Research, S.L., Santiago de Compostela, Spain) for binning to extract the ^{1}H-NMR data to obtain a dataset for chemometrics analysis.

The data were normalized using total area. The variables used were the intensity values from the chemical shift of 0–10 ppm excluding the area of methanol and water residual. Prior to PCA, PLS-DA, and OPLS-DA analysis, Pareto scaling was performed to the dataset. The result was observed using a score plot, S-line correlation plot, variable importance in projection (VIP) value, and permutation test. Variables with a p (corr) value of more than 0.5 observed in an S-line correlation plot were important variables in OPLS-DA. Meanwhile, variables with a VIP value greater than 1 were considered to have important role in samples' differentiation. In addition, in evaluation using a permutation test, the value of the original R2 and Q2 must have the highest value among permutated models.

4. Conclusions

Authentication of Curcuma species is important to ensure the quality, safety, and authenticity of the products. ^{1}H-NMR spectroscopy method could be employed at the stage of sample fingerprinting for authentication purpose both for herbal and medicinal plants. Combined with chemometrics of PCA, PLS-DA, and OPLS-DA, ^{1}H-NMR spectroscopy method is a powerful analytical tool for authentication of Curcuma species from different origins.

Exploratory data and classification models were successfully built. Several useful plots of output from the chemometrics models were also presented to visually assess the classification analysis. Predictive models for each species including *C. longa, C. xanthorrhiza*, and *C. manga* were evaluated according to the high values of the R^2 and Q^2. Other statistical and visual observations made considering the ROC curve and permutation test proved the probability and performance quality of the model. Hence, there is promise and potential to develop a combinational method with data fusion of ^{1}H-NMR spectroscopy and chemometrics technique for the authentication of medicinal plants and herbal products.

Author Contributions: Conceptualization, A.R. and A.W.; methodology, A.R. and A.W.; writing—original draft preparation, A.W., L.H.N.; writing—review and editing, L.H.N., F.D.O.R., A.G., E.L., N.A.F., and M.R.; funding acquisition, A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Ministry of Education and Culture through World Class Research, grant number 4507/UN1/DITLIT/DIT-LIT/PT/2021.

Data Availability Statement: The data presented in this study are available in the article.

Acknowledgments: The authors thank to Faculty of Pharmacy Universitas Gadjah Mada and Integrated Research and Testing Laboratory, Universitas Gadjah Mada for facilitating instrument for this research.

Conflicts of Interest: The authors declare no conflict of interest.

Sample Availability: Samples of the extracts of *Curcuma longa, Curcuma xanthorrhiza* and *Curcuma manga* are available from the authors.

References

1. Gad, H.A.; El-Ahmady, S.H.; Abou-Shoer, M.I.; Al-Azizi, M.M. Application of chemometrics in authentication of herbal medicines: A review. *Phytochem. Anal.* **2013**, *24*, 1–24, doi:10.1002/pca.2378.

2. Srirama, R.; Santhosh Kumar, J.U.; Seethapathy, G.S.; Newmaster, S.G.; Ragupathy, S.; Ganeshaiah, K.N.; Uma Shaanker, R.; Ravikanth, G. Species adulteration in the herbal trade: Causes, consequences and mitigation. *Drug Saf.* **2017**, *40*, 651–661, doi:10.1007/s40264-017-0527-0.
3. Subositi, D.; Wahyono, S. Study of the genus curcuma in Indonesia used as traditional herbal medicines. *Biodiversitas* **2019**, *20*, 1356–1361, doi:10.13057/biodiv/d200527.
4. Rohman, A.; Rawar, E.A.; Sudevi, S.; Nurrulhidayah, A.F.; Windarsih, A. The use of chemometrics in combination with molecular spectroscopic and chromatographic methods for authentication of curcuma species: A review. *Food Res.* **2020**, *4*, 1850–1858, doi:10.26656/fr.2017.4(6).345.
5. Rohman, A.; Nugroho, A.; Lukitaningsih, E.; Sudjadi Application of vibrational spectroscopy in combination with chemometrics techniques for authentication of herbal medicine. *Appl. Spectrosc. Rev.* **2014**, *49*, 603–613, doi:10.1080/05704928.2014.882347.
6. Wang, P.; Yu, Z. Species authentication and geographical origin discrimination of herbal medicines by near infrared spectroscopy: A review. *J. Pharm. Anal.* **2015**, *5*, 277–284, doi:10.1016/j.jpha.2015.04.001.
7. Windarsih, A.; Rohman, A.; Swasono, R.T. Application of 1H-NMR based metabolite fingerprinting and chemometrics for authentication of Curcuma longa adulterated with C. heyneana. *J. Appl. Res. Med. Aromat. Plants* **2019**, *13*, 100203, doi:10.1016/j.jarmap.2019.100203.
8. Jugreet, B.S.; Mahomoodally, M.F.; Sinan, K.I.; Zengin, G.; Abdallah, H.H. Chemical variability, pharmacological potential, multivariate and molecular docking analyses of essential oils obtained from four medicinal plants. *Ind. Crops Prod.* **2020**, *150*, 112394, doi:10.1016/j.indcrop.2020.112394.
9. Lukitaningsih, E.; Rohman, A.; Rafi, M.; Nurrulhidayah, A.F.; Windarsih, A. In vivo antioxidant activities of Curcuma longa and Curcuma xanthorrhiza: A review. *Food Res.* **2020**, *4*, doi:10.26656/fr.2017.4(1).172.
10. Ayati, Z.; Ramezani, M.; Amiri, M.S.; Moghadam, A.T.; Rahimi, H.; Abdollahzade, A.; Sahebkar, A.; Emami, S.A. Ethnobotany, phytochemistry and traditional uses of Curcuma spp. and pharmacological profile of two important species (C. longa and C. zedoaria): A review. *Curr. Pharm. Des.* **2019**, *25*, 871–935, doi:10.2174/1381612825666190402163940.
11. Kharade, S.S.; Samal, K.C.; Rout, G.R. High performance thin layer chromatography fingerprint profile of rhizome extracts of five important Curcuma species. *Proc. Natl. Acad. Sci. India Sect. B—Biol. Sci.* **2017**, *87*, 1335–1341, doi:10.1007/s40011-016-0709-z.
12. Cheng, J.; Weijun, K.; Yun, L.; Jiabo, W.; Haitao, W.; Qingmiao, L.; Xiaohe, X. Development and validation of UPLC method for quality control of Curcuma longa Linn.: Fast simultaneous quantitation of three curcuminoids. *J. Pharm. Biomed. Anal.* **2010**, *53*, 43–49, doi:10.1016/j.jpba.2010.03.021.
13. Xu, L. lu; Shang, Z. peng; Lu, Y. ying; Li, P.; Sun, L.; Guo, Q. lei; Bo, T.; Le, Z. yong; Bai, Z. li; Zhang, X. li; et al. Analysis of curcuminoids and volatile components in 160 batches of turmeric samples in China by high-performance liquid chromatography and gas chromatography mass spectrometry. *J. Pharm. Biomed. Anal.* **2020**, *188*, 113465, doi:10.1016/j.jpba.2020.113465.
14. Jia, S.; Du, Z.; Song, C.; Jin, S.; Zhang, Y.; Feng, Y.; Xiong, C.; Jiang, H. Identification and characterization of curcuminoids in turmeric using ultra-high performance liquid chromatography-quadrupole time of flight tandem mass spectrometry. *J. Chromatogr. A* **2017**, *1521*, 110–122, doi:10.1016/j.chroma.2017.09.032.
15. Gismondi, A., D'Agostino, A., Marco, G. Di, Martínez-Labarga, C., Leonini, V., Rickards, O. & Canini, A. (2020). Back to the roots: Dental calculus analysis of the first documented case of coeliac disease. *Archaeological and Anthropological Sciences*, **12**, 1–10, https://doi.org/10.1007/s12520-019-00962-w.
16. Wu, X.M.; Zuo, Z.T.; Zhang, Q.Z.; Wang, Y.Z. FT-MIR and UV–vis data fusion strategy for origins discrimination of wild Paris Polyphylla Smith var. yunnanensis. *Vib. Spectrosc.* **2018**, *96*, 125–136, doi:10.1016/j.vibspec.2018.04.001.
17. Eti Rohaetia, Mohamad Rafi, Utami Dyah Syafitric, R.H. Fourier transform infrared spectroscopy combined with chemometrics for discrimination of Curcuma longa, Curcuma xanthorrhiza and Zingiber cassumunar. *J. Chem. Inf. Model.* **2015**, 1244–1249.
18. Rohman, A.; Wijayanti, T.; Windarsih, A.; Riyanto, S. The authentication of Java turmeric (Curcuma xanthorrhiza) using thin layer chromatography and 1H-NMR based-metabolite fingerprinting coupled with multivariate analysis. *Molecules* **2020**, *25*, 3928, doi:10.3390/molecules25173928.
19. Deborde, C.; Fontaine, J.-X.; Jacob, D.; Botana, A.; Nicaise, V.; Richard-Forget, F.; Lecomte, S.; Decourtil, C.; Hamade, K.; Mesnard, F.; et al. Optimizing 1D 1 H-NMR profiling of plant samples for high throughput analysis: Extract preparation, standardization, automation and spectra processing. *Metabolomics 2019 153* **2019**, *15*, 1–12, doi:10.1007/S11306-019-1488-3.
20. Gholkar, M.S.; Li, J.V.; Daswani, P.G.; Tetali, P.; Birdi, T.J. 1H nuclear magnetic resonance-based metabolite profiling of guava leaf extract: An attempt to develop a prototype for standardization of plant extracts. *BMC Complement. Med. Ther.* **2021**, *21*, 1–20, doi:10.1186/s12906-021-03221-5.
21. Hatumura, P.H.; de Oliveira, G.S.; Marcheafave, G.G.; Rakocevic, M.; Bruns, R.E.; Scarminio, I.S.; Terrile, A.E. Chemometric Analysis of 1H NMR Fingerprints of Coffea arabica Green Bean Extracts Cultivated under Different Planting Densities. *Food Anal. Methods* **2018**, *11*, 1906–1914, doi:10.1007/s12161-017-1104-y.

22. Belmonte-Sánchez, E.; Romero-González, R.; Garrido Frenich, A. Applicability of high-resolution NMR in combination with chemometrics for the compositional analysis and quality control of spices and plant-derived condiments. *J. Sci. Food Agric.* **2021**, *101*, 3541–3550, doi:10.1002/jsfa.11051.
23. Petrakis, E.A.; Cagliani, L.R.; Polissiou, M.G.; Consonni, R. Evaluation of saffron (Crocus sativus L.) adulteration with plant adulterants by1H NMR metabolite fingerprinting. *Food Chem.* **2015**, *173*, 890–896, doi:10.1016/j.foodchem.2014.10.107.
24. Windarsih, A.; Rohman, A.; Swasono, R.T. Application of H-NMR metabolite fingerprinting and chemometrics for the authentication of Curcuma longa adulterated with Curcuma manga. *J. Appl. Pharm. Sci.* **2018**, *8*, 75–81, doi:10.7324/JAPS.2018.8610.
25. Kim, H.K.; Choi, Y.H.; Verpoorte, R. NMR-based metabolomic analysis of plants. *Nat. Protoc.* **2010**, *5*, 536–549, doi:10.1038/nprot.2009.237.
26. Li, Y.; Kong, D.; Fu, Y.; Sussman, M.R.; Wu, H. The effect of developmental and environmental factors on secondary metabolites in medicinal plants. *Plant. Physiol. Biochem.* **2020**, *148*, 80–89, doi:10.1016/j.plaphy.2020.01.006.
27. Jung, Y.; Lee, J.; Kim, H.K.; Moon, B.C.; Ji, Y.; Ryu, D.H.; Hwang, G.S. Metabolite profiling of Curcuma species grown in different regions using 1H NMR spectroscopy and multivariate analysis. *Analyst* **2012**, *137*, 5597–5606, doi:10.1039/C2AN35397K.
28. Awin, T.; Mediani, A.; Shaari, K.; Faudzi, S.M.M.; Sukari, M.A.H.; Lajis, N.H.; Abas, F. Phytochemical profiles and biological activities of Curcuma species subjected to different drying methods and solvent systems: NMR-based metabolomics approach. *Ind. Crops Prod.* **2016**, *94*, 342–352.
29. Parvathy, V.A.; Swetha, V.P.; Sheeja, T.E.; Sasikumar, B. Detection of plant-based adulterants in turmeric powder using DNA barcoding. *Pharm. Biol.* **2015**, *53*, 1774–1779, doi:10.3109/13880209.2015.1005756.
30. Worley, B.; Powers, R. Multivariate analysis in metabolomics. *Curr. Metabolomics* **2013**, *1*, 92–107, doi:10.2174/2213235x11301010092.
31. Narkhede, S. Understanding AUC—ROC curve [ONLINE]. *Available* via *https//towardsdatascience.com/understanding-auc-roccurve-68b2303cc9c5 (accessed 27 Sept. 2020)* **2018**.

Paper untuk direvisi: major revision

[Molecules] Manuscript ID: molecules-1470133 - Major Revisions (Due date: 26 November)

Inbox

Molecules Editorial Office <molecules@mdpi.com> Tue, Nov 16, 2021, 5:03 PM

to Anjar, Laela, me, Any, Florentinus, Endang, Nurrulhidayah, Mohamad, Molecules

Dear Prof. Rohman,

Thank you again for your manuscript submission:

Manuscript ID: molecules-1470133
Type of manuscript: Article
Title: Metabolite Fingerprinting using 1H-NMR Spectroscopy and Chemometrics for Classification of Three Curcuma Species from Different Origins
Authors: Laela Hayu Nurani, Abdul Rohman *, Anjar Windarsih, Any Guntarti, Florentinus Dika Octa Riswanto, Endang Lukitaningsih, Nurrulhidayah Ahmad Fadzillah, Mohamad Rafi
Received: 3 November 2021
E-mails: laelafarmasi@yahoo.com, abdulkimfar@gmail.com, anjarwindarsih2@gmail.com, any_guntarti@yahoo.co.id, dikaocta@uad.ac.id, lukitaningsih_end@ugm.ac.id, nurrulhidayah@iium.edu.my, mra@apps.ipb.ac.id
Submitted to section: Analytical Chemistry, https://www.mdpi.com/journal/molecules/sections/Analytical_Chemistry

Your manuscript has now been reviewed by experts in the field. Please find your manuscript with the referee reports at this link:

https://susy.mdpi.com/user/manuscripts/resubmit/042ab89649528c19d9b247ee0526ffac

Please revise the manuscript according to the referees' comments and upload the revised file within 10 days (Due date: 26 November).

Please use the version of your manuscript found at the above link for your revisions.

(I) Any revisions to the manuscript should be marked up using the “Track Changes” function if you are using MS Word/LaTeX, such that any changes can be easily viewed by the editors and reviewers.
(II) Please provide a cover letter to explain, point by point, the details of the revisions to the manuscript and your responses to the referees’ comments.
(III) If you found it impossible to address certain comments in the review

reports, please include an explanation in your rebuttal.
(IV) The revised version will be sent to the editors and reviewers.

If one of the referees has suggested that your manuscript should undergo extensive English revisions, please address this issue during revision. We propose that you use one of the editing services listed at https://www.mdpi.com/authors/english or have your manuscript checked by a native English-speaking colleague.

Do not hesitate to contact us if you have any questions regarding the revision of your manuscript. We look forward to hearing from you soon.

Kind regards,
Ms. Jelena Grujic
Assistant Editor, MDPI Belgrade

E-Mail: jelena.grujic@mdpi.com
Bulevar Mihajla Pupina 115, 11000 Belgrade, Serbia
+381 11 414 75 49

Komentar Reviewer

Response to Reviewer 1 Comments

Comment 1: The revised version of the manuscript "Metabolite Fingerprinting using 1H-NMR Spectroscopy and Chemometrics for Classification of Three Curcuma Species 3 from Different Origins" is not accompanied by a cover letter addressing the issues remaining and not supporting the changes made.

Response 1: Actually, we have included a cover letter addressing the issue when we resubmit our manuscript.

I am reporting various issues with this version:

Comment 2: line 178 - "The samples were perfectly classified" In chemometrics such exclamations should be avoided

Response 2: We have revised to "The samples were successfully classified"

Comment 3: line 179 -180 "using 2 PC and 5 X-orthogonal components with R2X(cum) of 0.808, R2Y(cum) 179 of 0.978 and Q2 (cum) of 0.971" Which components did the authors use? I have informed them from version 1 that they must emphasize the components they used and report the variance these two specific components explain. This is valid for all models

Response 3: We have revised and emphasized the components we used in the manuscript in all models.

Comments 4: line182-184 "The OPLS-DA model showed better classification than in PLS-DA model because OPLS-DA not only used latent variables but also orthogonal components to deeply search important variables for classification" I have reported in previous versions that this remark is not necessary since the OPLS-DA is expected to deliver such results

Response 4: We have removed this remark from the manuscript.

Comments 5: line 187 "variable importance in projection (VIP) value" The authors do not desceribe the workflow of their data treatment in the materials and methods section they must discuss how they implemented VIP values. Needless to say I have asked this from the previous version.

Response 5: We have added the explanation in implementing VIP values in the methods section, subsection of chemometrics analysis.

Comments 6: The authors have corrected the scores plot but I did not receive any comments on them.

Response 6: We have made corrections to the score plot and we have corrected the elaboration in the text.

Comments 7: There are several issues with the annotation of the loading plots
e.g, Figure 2 loading score the Demethoxycurcumin is in both directions of the loading that is probably due to an alignment issue.

Response 7: We have made the correction in Figure 2 by changing to S-line correlation plot

Comments 8: Figure 3 loading plot lysine is in both directions of the loading that is probably due to an alignment issue.

Response 8: We have made the correction in Figure 3 by changing to S-line correlation plot

Comments 9: Figure 4 Demethoxycurcuminis in both directions of the loading that is probably due to an alignment issue.

Response 9: We have made the correction in Figure 4 by changing to S-line correlation plot

Comments 10: A better way to elaborate on the results is the S-line plot through pairwise comparison. If the authors do not wish that they must incorporate in the loading plots both components in order to elaborate on the parameters affecting the sample's localization in each quadrant.

Response 10: We have made revision to provide S-line correlation plot

Comments 11: In line 346 "result was observed using score plot, loading score, and permutation test" What about the VIP? what about the S-line plot the authors stated in a previous version they employed?

Response 11: We have revised to "result was observed using score plot, S-line correlation plot, variable importance in projection (VIP) value and permutation test".

We have made revisions in our manuscript. We change the loading plot to S-line correlation plot.

Comments 12: In the conclusion "Exploratory data and classification analysis were successfully built. " I think the word models should be added and after reworking the models this section should be thought over

Response 12: We have revised the conclusion as follow: Exploratory data and classification model were successfully built. Several useful plots as output from the chemometrics models were also presented to visually assess the classification analysis.

Response to Reviewer 2 Comments

The paper has been improved as requested.

I have just two minor comments. After the addressing of these modifications, the paper can be published.

Comments 1

> (4th comment of the previous revision) among the cited articles the authors selected a paper that is interesting but not required. I indicated https://doi.org/10.1007/s12520-019-00962-w that should be considered

Response 1: We apologize and we have added the suggested reference as follow: Gismondi, A., D'Agostino, A., Di Marco, G. et al. Back to the roots: dental calculus analysis of the first documented case of coeliac disease. Archaeol Anthropol Sci 12, 6 (2020). https://doi.org/10.1007/s12520-019-00962-w in the text (Reference 15).

Comments 2

> The term Curcuma (after the first time) should be reported in abbreviated form C. always in the text (probably the authors misunderstood my previous request).

Response 2: We have changed the term Curcuma to abbreviated form C. after the first time in the manuscript.

Response to Reviewer 3 Comments

Comments

Products of natural origin are generally considered totally harmless, but the active ingredients they contain carry out a biological action on the body and therefore can cause unwanted effects or interact with synthetic drugs, not to mention the presence of possible contaminants. Therefore, it is extremely important to have analysis tools for identifying plant matrices. In this work, 1H-NMR spectroscopy and principal component analysis (PCA) was used, as well as orthogonal projections to latent structures-discriminant analysis (OPLS-DA) for the authentication of three turmeric species.

The manuscript is well laid out.

In particular

The Introduction section introduces the reader to the current thematic problem.

The Materials and Methods section is written in detail.

The Results and Discussion section is written correctly and the references are sufficiently up to date.

The manuscript requires no further modification.

Response:

Thank you very much for your positive comments.

Paper yang direvisi

Article

Metabolite Fingerprinting using ^{1}H-NMR Spectroscopy and Chemometrics for Classification of Three Curcuma Species from Different Origins

Laela Hayu Nurani [1], Abdul Rohman [2,3,*], Anjar Windarsih [4], Any Guntarti [1], Florentinus Dika Octa Riswanto [5], Endang Lukitaningsih [2], Nurrulhidayah Ahmad Fadzillah [6], and Mohamad Rafi [7]

Citation: Nurani, L.H.; Rohman, A.; Windarsih, A.; Guntarti, A. Metabolite Fingerprinting using ^{1}H-NMR Spectroscopy and Chemometrics for Classification of Three Curcuma Species from Different Origins. *Molecules* **2021**, *26*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s): name

Received: 17 August 2021
Accepted: date
Published: date

[1] Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta, Indonesia; laelafarmasi@yahoo.com (L.H.N.); any_guntarti@yahoo.co.id (A.G.)
[2] Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Universitas Gadjah Mada, Yogyakarta, 55281
[3] Center of Excellence Institute for Halal Industry and Systems (IHIS), Universitas Gadjah Mada, Yogyakarta, 55281
[4] Research Division for Natural Product Technology (BPTBA), National Research and Innovation Agency (BRIN), Yogyakarta, Indonesia 55861; anjarwindarsih2@gmail.com (A.W.)
[5] Division of Pharmaceutical Analysis and Medicinal Chemistry, Faculty of Pharmacy, Universitas Sanata Dharma, Yogyakarta, Indonesia, 55282
[6] International Institute for Halal Research and Training (INHART), International Islamic University
Malaysia (IIUM), Gombak, 50728, Kuala Lumpur, Malaysia
[7] Department of Chemistry, Faculty of Mathematics and Natural Sciences, IPB University, Bogor, 16680,
Indonesia
* Correspondence: abdulkimfar@gmail.com

Abstract: *Curcuma longa, Curcuma xanthorrhiza,* and *Curcuma manga* have been widely used as medicinal plants for herbal or traditional medicine purposes. It was reported that turmeric plants provided several biological activities such as antioxidant, anti-inflammatory, hepatoprotector, cardioprotector, and anticancer activities. Authentication of Curcuma species is important to ensure its authenticity as well as to avoid adulteration practices. Different origins will have different metabolite compositions because metabolites are affected by soil nutrition, climate, temperature, and humidity. ^{1}H-NMR spectroscopy and principal component analysis (PCA) as well as orthogonal projections to latent structures-discriminant analysis (OPLS-DA) could be used for authentication of *C. longa, C. xanthorrhiza ,* and *C. manga* from seven different origins in Indonesia. From the ^{1}H-NMR analysis it was obtained that 14 metabolites were responsible for generating classification model such as curcumin, demethoxycurcumin, alanine, methionine, threonine, lysine, alpha-glucose, beta-glucose, sucrose, alpha-fructose, beta-fructose, fumaric acid, tyrosine, and formate. Both PCA and OPLS-DA model demonstrated good of fitness (R^2 value more than 0.8) and good predictivity (Q^2 value more than 0.45). All OPLS-DA models were validated by assessing the permutation test results with high value of original R^2 and Q^2. It can be concluded that metabolite fingerprinting using ^{1}H-NMR spectroscopy and chemometrics provide a powerful tool for authentication of herbal and medicinal plants.

Keywords: authentication; curcuma; ^{1}H-NMR spectroscopy; chemometrics; metabolite fingerprinting

1. Introduction

Since hundreds of years, herbal medicines and their preparations have been widely used in folk medicines over the world. The preparation of herbal medicine preparations is typically presented either as single herbs or as a combination of several herbs in composite formulae, and it is reported that about 92% of herbal medicine formulas are a combination of less than thirteen herbal medicines [1] [2]. Annually, the market growth of herbal products has increased in which the raw material for most herbal products come from South Asian and Southeast Asian countries including Indonesia [3]. Products of natural origin, such as supplements,

herbal products or herbal preparations, are increasingly widespread and used to maintain the state of health or for the treatment of small diseases. However, natural is not necessarily synonymous with safe. The adulteration of herbal preparations, together with contamination, sophistication, degradation is a problem of global interest. In recent years, the public awareness on herbal authentication and species admixtures in the raw herbal has increased significantly, because the adverse consequences of adulterated herbal components on the consumer safety has been well recognized [4]. Sophistication in herbal medicine involves replacing and adulterating botanical materials, diluting high quality herbal medicines with lower grade ones, and mislabeling herbal medicine. Therefore, it is essential to have monitoring and pharmacovigilance systems [5].

For species authentication of herbal medicine, the World Health Organization (WHO), the United States Food and Drug Administration (USFDA) and the European Medicines Agency (EMEA) have regulated that the identification of herbal medicines is one of the first identification which should be made to ensure their quality and to make discrimination from related species or adulterated samples [6]. Among herbal medicines components, Curcuma species including *C. longa* (turmeric), *C. xanthorrhiza* (Java Turmeric), and *C. manga* have been widely applied as medicinal plants for herbal or traditional medicine purposes [7]. These Curcumas have been reported to have some biological effects which are beneficial to human health including antioxidant, anticancer anti-inflammatory, hepatoprotector, cardioprotector, antibacterial activities, and wound healing [8] [9] [10].

Some analytical methods have been developed for analysis of Curcuma species. Most of the methods are using chromatography-based methods such as high performance-thin layer chromatography (TLC) [11] , high performance liquid chromatography (HPLC), ultra-high performance liquid chromatography (UHPLC) [12], gas chromatography-mass spectrometry (GC-MS) for analysis volatile compounds in Curcuma species [13] and liquid chromatography-mass spectrometry (LC-MS/MS) [14]. Chromatographic-based methods typically involved complex sample preparation technique and resulted huge number of responses which make difficulty in data analysis. Therefore, spectroscopic-based methods in combination with multivariate data analysis (MDA) or chemometrics were potential to be employed since this combination method was provided the way to analyze such an environmental big data [15]. Ultraviolet, visible and vibrational spectroscopy (infrared and Raman) [16], [17], and NMR spectroscopy [18] are widely reported for authentication of Curcuma species. NMR spectroscopy offers some advantages for authentication of medicinal plants such as fast time analysis, simple in sample preparation, high reproducibility, and high robust. Moreover, NMR spectroscopy can be used for simultaneous analysis either primary or secondary metabolites comprehensively in certain samples [19,20]. Combined with chemometrics of multivariate analysis such as principal component analysis (PCA), partial least square-discriminant analysis (PLS-DA), and orthogonal projections to latent structures-discriminant analysis (OPLS-DA) which can manage the huge data generated from NMR measurement, it becomes a powerful analytical tool for metabolite fingerprinting of medicinal plants [21,22]. Combination of ^{1}H-NMR spectroscopy and chemometrics of PLS-DA and OPLS-DA has been used

for authentication of Saffron adulteration [23]. ^{1}H-NMR spectroscopy and chemometrics have also been used for authentication of *C. longa* adulterated with *C. manga* and *C. heyneana* [7,24]. Authentication of *C. xanthorrhiza* from *C. aeruginosa* has been successfully investigated using ^{1}H-NMR and multivariate analysis [18]. However, study on authentication of Curcuma species from different origins using ^{1}H-NMR spectroscopy is still limited. Therefore, the objective of this study was to use ^{1}H-NMR spectroscopy in combination with chemometrics for authentication of *C. longa, C. xanthorrhiza* , and *C. manga* from different origins.

2. Results and Discussion

2.1.1. H-NMR Spectra Analysis

^{1}H-NMR spectra can be used for authentication of medicinal plants because it offers fingerprinting which mean that each sample has specific ^{1}H-NMR spectra pattern. Generally, metabolites of plants extracted using deuterated methanol and deuterium oxide measured using ^{1}H-NMR spectroscopy are divided into three main regions, namely amino acid and organic acids (0.20-3.00 ppm), carbohydrate or sugar (3.01-5.00 ppm), and aromatic compounds (6.00-8.00 ppm) [25]. Different origins have different conditions such as soil condition, soil nutrition (macro and micronutrients), humidity, light, salinity, and temperature as well as internal developmental genetic circuits including regulated gene, and enzyme which can obviously affect the metabolite formation either primary or secondary metabolites [26]. The ^{1}H-NMR spectra of *C. longa* (CL), *C. xanthorrhiza (*CX), and *C. manga* (CM) are shown in Figure 1. It can be observed that *C. longa*, *C. xanthorrhiza*, and *C. manga* have different spectra pattern indicating different metabolite contents. Specifically observed, *C. longa* and *C. xanthorrhiza* have higher signal intensities in the region of amino acid and organic acid (0.20-3.00 ppm) as well as in the aromatic region (6.00-8.00 ppm) than *C. manga*. On the other hand, the signal intensities in the region of glucose (3.01-5.00 ppm) are higher in *C. manga* compared to *C. longa* and *C. xanthorrhiza*. Fourteen metabolites in Curcuma species obtained from ^{1}H-NMR measurement are shown in Table 1. Curcumin, demethoxycurcumin, alanine, methionine, threonine, lysine, alpha-glucose, beta-glucose, sucrose, alpha-fructose, beta-fructose, fumaric acid, tyrosine, and formate were stated as metabolites which playing important role in generating OPLS-DA model. Investigation on each species obtained from different regions resulted different signal patterns especially in intensities indicating the variations in metabolite contents in each species from different origins. It indicated that different origins affect the metabolite contents in each Curcuma rhizome. For example, *C. longa* from Blitar (CL7) has the lowest signal intensities in the aromatic region and *C. xanthorrhiza* from Gunungkidul (CX2) has the lowest signal intensities in the whole regions. However, the spectra patterns of each species are quite similar, therefore, for deeper classification of *C. longa*, *C. xanthorrhiza* , and *C. manga* from different regions powerful statistical tool such as chemometrics is required to obtain clear classification.

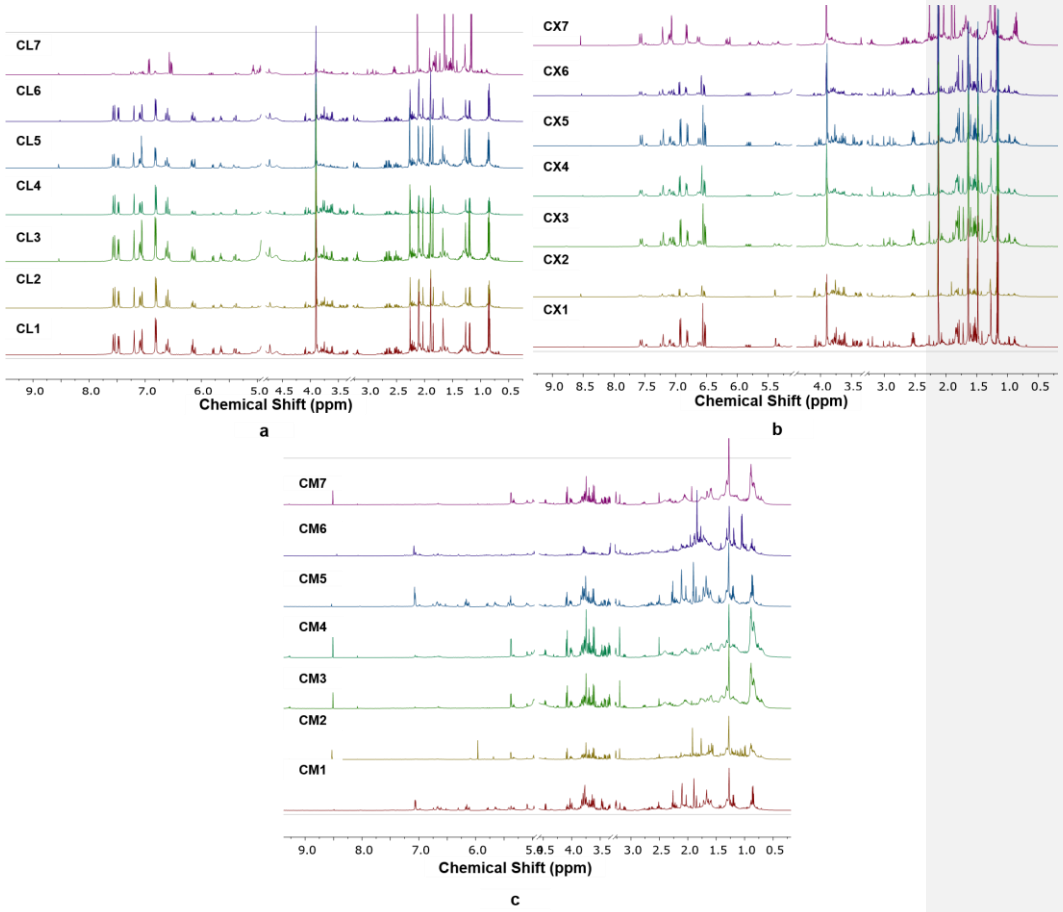


Figure 1. ^{1}H-NMR spectra of *C. longa* (a), *C. xanthorrhiza* (b), and *C. manga* (c) from different origins.**Table 1.** Several metabolites of Curcuma species observed using ^{1}H-NMR spectra [23,24].

No.	Chemical Shift (ppm)	Multiplicity	Metabolite
1.	7.57	Singlet	Curcumin
	7.28	Singlet	
	7.21	Doublet	
	6.77	Doublet	
	3.68	Singlet	
2.	6.92	Doublet	Demethoxycurcumin
	5.89	Singlet	
	3.94	Singlet	

3.	1.49	Doublet	Alanine
	3.72	Quartet	
4.	2.11	Singlet	Methionine
5.	1.33	Doublet	Threonine
	3.53	Doublet	
6.	3.81	Triplet	Lysine
	1.5	Multiplet	
7.	5.19	Doublet	Alpha-Glucose
	3.46	Doublet of Doublet	
	3.67	Triplet	
	3.35	Triplet	
8.	4.59	Doublet	Beta-Glucose
	3.19	Doublet of Doublet	
	3.44	Triplet	
	3.71	Doublet of Doublet	
9.	5.42	Doublet	Sucrose
	3.74	Triplet	
	3.43	Triplet	
	3.80	Multiplet	
	3.84	Multiplet	
10.	4.07	Doublet	Alpha-Fructose
	3.82	Doublet of Doublet	
	3.53	Doublet	
	3.55	Doublet	
	3.63	Quartet	
11.	3.95	Multiplet	Beta-Fructose
	3.52	Doublet	
	4.02	Doublet of Doublet	
12.	6.57	Singlet	Fumaric acid
13.	6.81	Doublet	Tyrosine
	7.14	Doublet	
14.	8.42	Singlet	Formate

Curcuminoids have been reported as the active compound in Curcuma species. The content of curcuminoids is varied among Curcuma species and it is reported that curcuminoid content in *C. longa* and *C. xanthorrhiza* is higher among other Curcuma species. Curcuminoids consist of curcumin, demethoxycurcumin, and bisdemethoxycurcumin which curcumin possess the highest concentration. However, not all

Curcuma species contain these three types of curcuminoids, for instance *C. xanthorrhiza* does not contain bisdemethoxycurcumin. Curcuminoids are aromatic molecules therefore most of the signals appeared in the chemical shift of aromatic regions. Curcumin signal could be observed in the chemical shift of 7.57 ppm (singlet), 7.28 ppm (singlet), 7.22 ppm (doublet), 6.77 ppm (doublet) and 3.90 ppm (singlet) whereas demethoxycurcumin could be found in the chemical shift of 6.92 ppm (doublet), 5.89 ppm (singlet), and 3.94 ppm (singlet) [28]. From the ^{1}H-NMR spectra, higher signal intensities in the aromatic region of *C. longa* and *C. xanthorrhiza* supports that curcuminoids content in *C. longa* and *C. xanthorrhiza* is higher than in *C. manga*.

2.2. Chemometrics Analysis

Rhizomes of *C. longa*, *C. xanthorrhiza*, and *C. manga* are often used in a powder form as well as in an extract form for their herbal and traditional medicine applications. Both in powder and extract, are susceptible for adulteration because of their similar appearance especially in the adulterated form it is challenging to state the unknown sample whether it is authentic or adulterated sample [29]. Chemometrics of PCA could not differentiate *C. longa*, *C. xanthorrhiza*, and *C. manga* clearly (data not shown). It might be caused by the large variations of the variables therefore the principal components (PC) were not able to represent the original variables. Observing using supervised pattern recognition, namely PLS-DA using 7 PC, could classify *C. longa*, *C. xanthorriza*, and *C. manga* resulting three different classifications. However, several misclassifications occurred between *C. longa* and *C. xanthorrhiza* (Fig. 2a). In the PLS-DA score plot, several *C. longa* samples appeared in the region of *C. xanthorrhiza* and several *C. xanthorrhiza* samples appeared in the region of *C. longa*. It can be explained that some of the metabolite compositions of *C. longa* and *C. xanthorrhiza* are similar especially in curcuminoid contents which curcumin and demethoxycurcumin are the major active compounds *in C. longa* and *C. xanthorrhiza*. In addition, it is also often reported about adulteration or substitution of *C. longa* with *C. xanthorrhiza* and the adulteration is often difficult to be detected because the appearance of *C. longa* and *C. xanthorrhiza* in powder and extract form are quite similar [17]. Therefore, another supervised pattern recognition chemometrics, namely OPLS-DA is performed to obtain better classification of *C. longa*, *C. xanthorrhiza*, and *C. manga* extracts. The OPLS-DA model demonstrated good capability to differentiate three different species of *C. longa*, *C. heyneana*, and *C. manga* from different origins as shown in OPLS-DA score plot (Fig. 2b). OPLS-DA model successfully classified *C. longa*, *C. xanthorrhiza*, and *C. manga* samples. The samples were successfully classified without misclassification using first PC and first X-orthogonal components accounted for 80.8% of the variance with $R^2X(cum)$ of 0.808, $R^2Y(cum)$ of 0.776 and $Q^2(cum)$ of 0.767. High value of $R^2X(cum)$ and $R^2Y(cum)$ (close to 1) indicated good of fitness of the OPLS-DA model whereas the value of Q^2 greater than 0.45 indicated good of predictivity of the models [30]. The S-line correlation plot (Fig. 2c) variables which have roles in the differentiation of *C. longa*, *C. xanthorrhiza*, and *C. manga*. It was found that alanine, curcumin, demethoxycurcumin, fumaric acid, sucrose and tyrosine had p(corr) value more than 0.5 indicating their important roles in separating samples. Moreover, observed using variable importance in projection (VIP) value, the chemical shift of 6.77, 3.89, 7.57, 6.81, 6.57, 7.21, 1.49, 6.49,

6.13, 0.85, 6.09, 5.29, 5.25, and 6.92 ppm were found to have important roles for the classification between three Curcuma species in OPLS-DA models. Variables with VIP values greater than 1 are considered to have important roles for differentiation. Some of the variables correspond to the metabolites of curcumin, tyrosine, fumaric acid, alanine, and demethoxycurcumin. The receiver operating characteristic curve (Fig. 2d) for differentiating and classifying Curcuma longa, Curcume xanthorrhiza, and Curcuma manga from different origins was also depicted. The ROC analysis represents the propability of the model by plotting the value of true positivity rate (TPR) against the value of false positivity rate (FPR) [31].

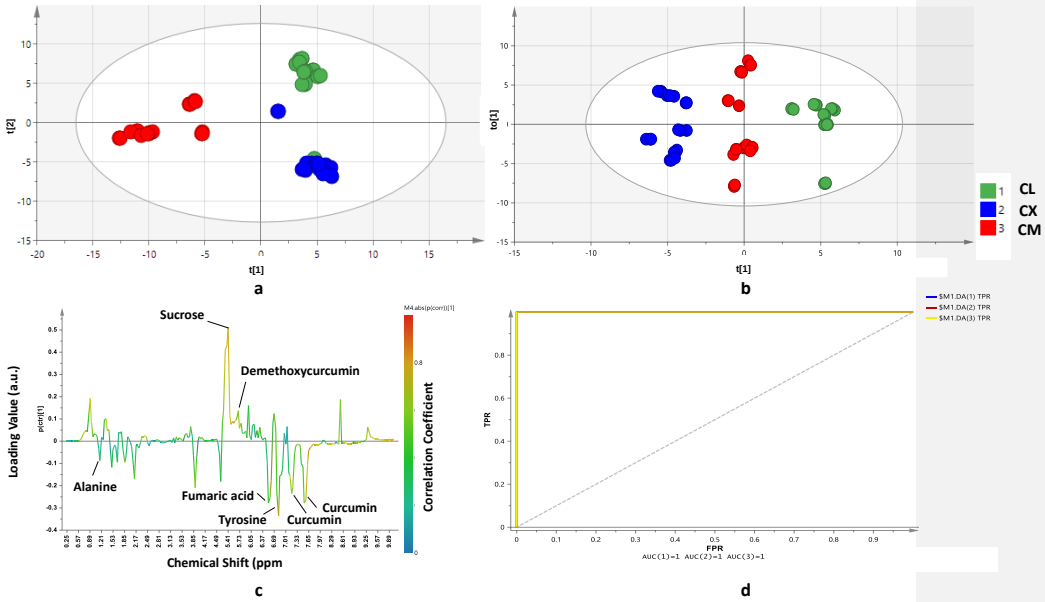


Figure 2. PLS-DA score plot (a), OPLS-DA score plot (b), OPLS-DA S-line correlation plot (c), and ROC curve (d) for differentiation and classification *of C. longa, C. xanthorrhiza,* and *C. manga* from different origins. Statistical parameters of the models: (a) Number of samples = 42; $R^2X = 0.603$; $R^2Y = 0.783$; $Q^2 = 0.755$ for t[1] and t[2] components; (b,c) Number of samples = 42; $R^2X = 0.808$; $R^2Y = 0.776$; $Q^2 = 0.767$ for t[1] and to[1] components.

PCA using number of PC 8 could differentiate *C. longa* from seven different origins as shown in PCA score plot (Fig 3a). The PCA model provided high confident for its fitting and predictivity capacity showed by its R^2 value (0.770) and Q^2 value (0.650) for PC1 and PC2, respectively accounted for 77.0% of the variance. The score plot which appears close each other indicated high similarity between samples especially their metabolite compositions. From the PCA score plot result, it shows that CL2 and CL4 possessed high similarity and CL1 has high similarity with CL6. Meanwhile, CL3 appeared closely to CL5. CL7 appeared far from all CL samples from other regions meaning that the metabolites composition of CL7 differs from other *C. longa* used in this research. Classification of

C. longa samples from seven different origins using OPLS-DA demonstrated different pattern with PCA result (Fig. 3b). OPLS-DA was created using first PC and first orthogonal-X component resulting $R^2X(cum)$ of 0.748, $R^2Y(cum)$ of 0.754 and $Q^2(cum)$ of 0.639. The first PC and first X-orthogonal component explained 74.8% of the total variance. There were three main groups obtained from OPLS-DA classification. The first group was CL2 and CL7 which appeared close to each other. The second group consist of CL1 and CL5 whereas the last group was CL3, CL4, and CL6. The important variables for differentiation and classification of *C. longa* between groups observed using S-line correlation plot were alanine, β-fructose, curcumin, demethoxycurcumin, fumaric acid, and tyrosine. Meanwhile, investigation using VIP value found that variables of 6.92, 6.54, 6.57, 1.50, 1.66, 1.54, 1.18, 7.46, 6.62, 0.86, 7.10, 7.50, 7.57, and 7.14 ppm had important roles for *C. longa* differentiation and classification from seven different origins classified using OPLS-DA model. Some of the variables correspond to the molecule signals of curcumin, demethoxycurcumin, fumaric acid, lysine, and tyrosine. It demonstrated that different origins affect the composition of some metabolites. It is in accordance with the study of Jung et al. [27] that studying the metabolite compositions of *C. longa* from several regions in China. The condition of geographical origin, environmental condition such as temperature, humidity, and rainfall rate definitely affect the metabolite composition of plants.

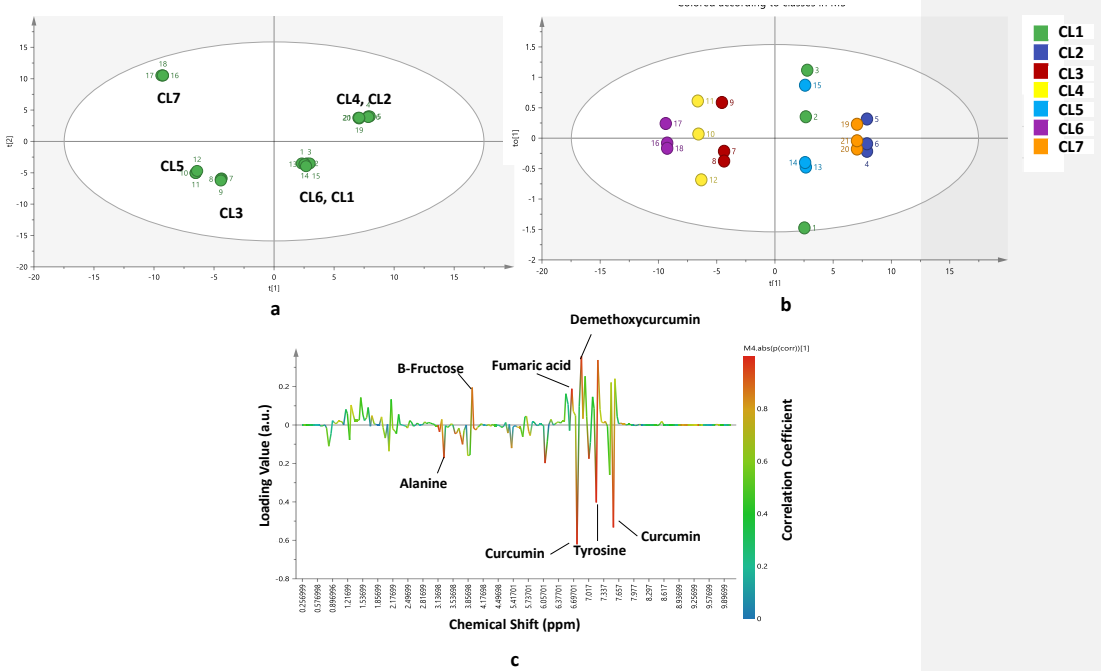


Figure 3. PCA score plot (a) OPLS-DA score plot (b) and OPLS-DA S-line correlation plot (c) for differentiation and classification *of C. longa* from different origins. Statistical parameter of the models:

(a,) Number of samples = 21; Number of PC = 8; $R^2X = 0.770$; $Q^2 = 0.650$; (b,c) Number of samples = 21; $R^2X = 0.748$; $R^2Y = 0.754$; $Q^2 = 0.639$ for t[1] and to[1] components.

Chemometrics of PCA and OPLS-DA was also successfully used for differentiation and classification of *C. xanthorrhiza* from seven different origins. The result of differentiation of samples using PCA was slightly different with the result from OPLS-DA. PCA performed using first and second principal components demonstrated good of fitness ($R^2cum = 0.743$) and predictivity ($Q^2cum = 0.678$) with total variance of 74.3%. Meanwhile, OPLS-DA was created using first PC and first orthogonal-X components. High value of $R^2X(cum)$ (0.743) and $R^2Y(cum)$ (0.833) indicated good of model fitness while high value of $Q^2(cum)$ (0.626) indicated good of predictivity of the OPLS-DA model. The first PC and first X-orthogonal-X component demonstrated 74.3% of the total variance. There were four main groups appeared in the PCA score plot whereas in OPLS-DA there were just divided into three groups. *C. xanthorrhiza* of CX3 and CX5 appeared close each other both in PCA and OPLS-DA score plot result. Meanwhile, *C. xanthorrhiza* of CX1 was found in a separate group with others observed both in PCA and OPLS-DA. Samples of CX2, CX4, CX6 and CX7 appeared in the same group observed using OPLS-DA, however, from the PCA result, sample of CX7 appeared in a different group. From these results, it suggested that different locations have significant effects to the metabolite's compositions of *C. xanthorrhiza*. S-line correlation plot showed that methionine, β-glucose, sucrose, fumaric acid, curcumin, demethoxycurcumin, and tyrosine were the important variables for CX differentiation. Moreover, observed using VIP value, the variables important for classifying samples were 4.61, 6.57, 6.92, 4.59, 2.13, 6.92, 7.01, 7.21, 6.53, 1.69, 1.25, and 1.77 ppm. Some of the variables corresponded to the metabolites of curcumin, fumaric acid, demethoxycurcumin, and beta-glucose. It is presumed that these metabolites have higher score and significantly affect the differentiation and classification of *C. xanthorrhiza* from different origins.

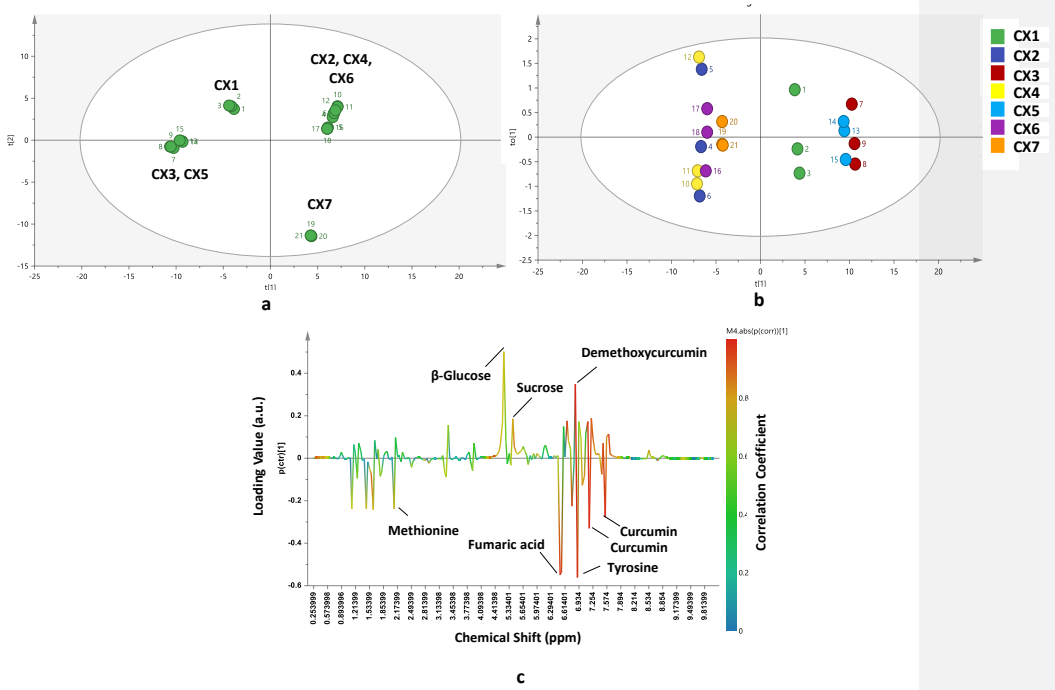


Figure 4. PCA score plot (a) OPLS-DA score plot (b) and OPLS-DA S-line correlation plot (c) for differentiation and classification *of C. xanthorrhiza* from different origins. Statistical parameter of the models: (a) Number of samples = 21; Number of PC = 6; $R^2X = 0.743$; Q2 = 0.678; (b,c) Number of samples = 21; $R^2X = 0.743$; $R^2Y = 0.833$; $Q^2 = 0.626$ for t[1] and to[1] components.

Chemometrics of PCA and OPLS-DA was also successfully applied for differentiation and classification of *C. manga* from seven different origins. Different classification results were observed between PCA and OPLS-DA. The PCA model was created using first and second principal components resulting R^2X(cum) of 0.661 and Q^2(cum) of 0.501 indicating good of fitness and good of predictivity of PCA model, respectively. The first and second PCs showed 66.1% of the total variance. From the PCA score plot, *C. manga* were classified in five classes as follows: CM1 (first class), CM5 (second class), CM6 (third class), CM2 (fourth class), and the rest of the samples was in the last class (CM3, CM4, and CM7). Samples of CM3, CM4, and CM7 have similar chemical or metabolite compositions because they appeared in the same location in PCA score plot. It is presumed that the conditions in the region of Malang (CM3), Tulung Agung (CM4), and Blitar (CM7) are similar resulting the similar metabolites of *C. manga* rhizomes. On the other hand, OPLS-DA was performed using first principal components and first orthogonal-X component which presented 66.1% of the total variance. The obtained R^2X(cum) (0.661) and R^2Y(cum) (0.667) indicated good of fitness whereas the value of Q^2(cum) (0.707) demonstrated good of model predictivity. Three main groups were found in the OPLS-DA score plot, namely CM1 and CM6 as the first group, CM5 in the second group, and in the last

group consist of CM2, CM3, CM4 and CM7. Observed using S-line correlation plot demonstrated that β-glucose, sucrose, curcumin, tyrosine, and format had important roles for CM differentiation. The VIP value showed that some variables were found to have significant contributions in the differentiation of *C. manga* samples from different origins, namely: 7.05, 5.29, 5.97, 5.33, 5.37, 8.53, 1.25, 1.81, 5.42, 5.65, 7.09, 6.17, 6.69, 0.85, 0.89, 3.33, and 5.41 ppm. Some of the variables are associated with curcumin, demethoxycurcumin, sucrose and fumaric acid.

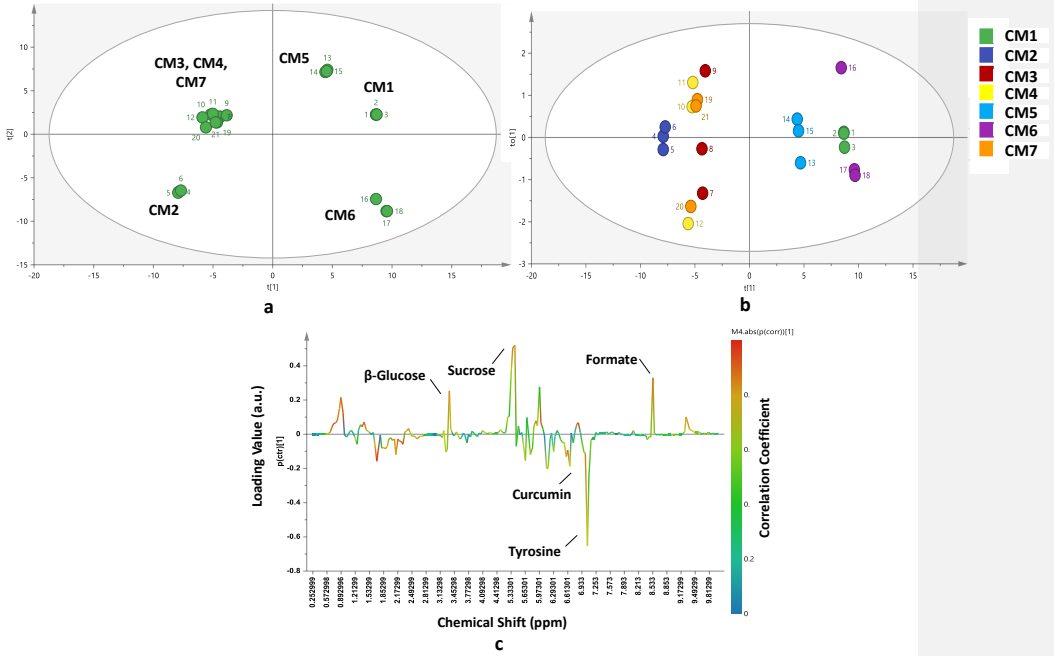


Figure 5. PCA score plot (a) OPLS-DA score plot (b) and OPLS-DA S-line correlation plot (c) for differentiation and classification *of C. manga* from different origins. Statistical parameter of the models: (a) Number of samples = 21; Number of PC = 7; $R^2X = 0.661$; $Q^2 = 0.501$; (b,c) Number of samples = 21; $R^2X = 0.661$; $R^2Y = 0.667$; $Q^2 = 0.707$ for t[1] and to[1] components.

2.3. *Validation of OPLS-DA Model Using Permutation Test*

Supervised pattern recognition of chemometrics such as PLS-DA and OPLS-DA requires validation test to confirm the model's validity because it is potential to have overfitting model. Validation is a confirmation step to ensure that the models have good of fitness. Permutation test is one of validation testing which used permutated model. Models of R^2 and Q^2 are permutated and compared to the original models of R^2 and Q^2. Good model is obtained when all the permutated models of R^2 and Q^2 values are lower than the R^2 and Q^2 original values. Moreover, the validation was also determined using intersection value of Q^2. The intersection value should be zero or lower than zero to be categorized as valid models. The result of permutation test from 999

permutations of OPLS-DA models were demonstrated in Figure 6. The permutated models of R^2 and Q^2 were located in the left side while the original R^2 and Q^2 models were shown in the right side. The models were permutated for 100 permutations. Results showed that all permutation test confirmed the validity of the OPLS-DA model demonstrated by the value of R^2 and Q^2 in all permutated models were below the value of original R^2 and Q^2 models. On the other hand, the intersection values of Q^2 for all four OPLS-DA models were also zero and lower than zero, as follows; (0.0, -0.473) for classification model between three Curcuma species; (0.0, -1.02) for classification model of *C. longa* from different origins; (0.0, -0.896), for classification model of *C. xanthorrhiza* from different origins; and (0.0, -0.904) for classification model of *C. manga* from different origins. It is suggested that OPLS-DA could be used as a powerful statistical tool for classification of different Curcuma species from different origins with high validity.

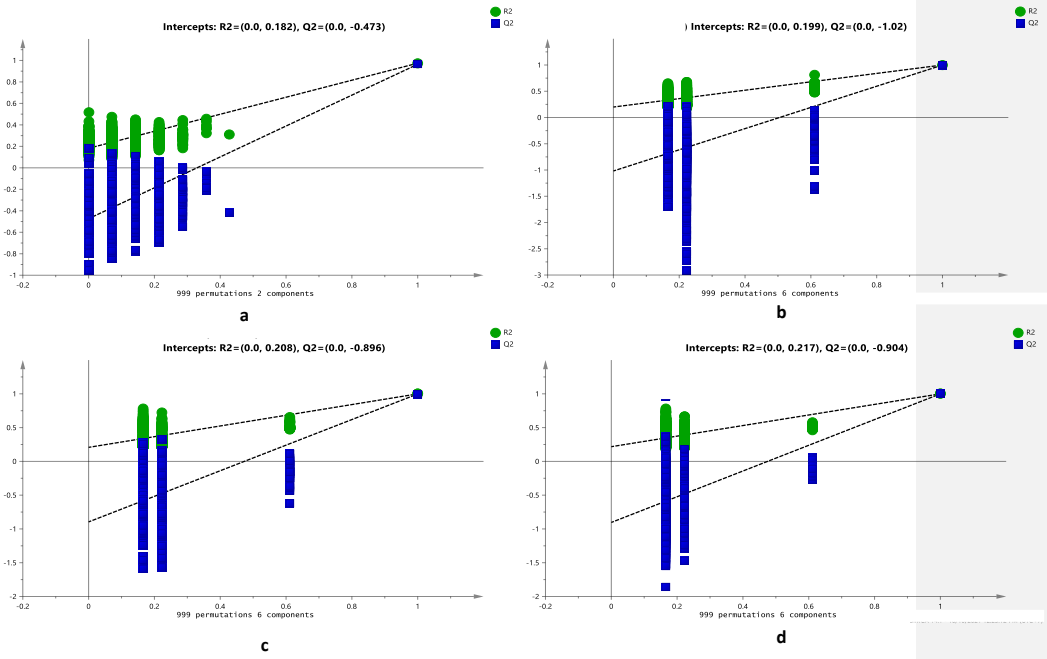


Figure 6. Permutation test using 999 permutations of OPLS-DA result from three Curcuma species (a), *C. longa* from different origins (b), *C. xanthorrhiza* from different origins (c), and *C. manga* from different origins (d).

3. Materials and Methods

3.1. Sample Collection and Preparation

Rhizome of *C. longa* (CL), *C. xanthorrhiza* (CX), and *C. manga* (CM) were collected from seven different regions in Indonesia, namely: Boyolali (1), Gunungkidul (2), Ngawi (3), Malang (3), Tulung Agung (5), Karang Anyar (6), and Blitar (7). Determination of plant species used in this study has been carried out in Pharmaceutical Biology Department,

Faculty of Pharmacy, Universitas Gadjah Mada, Indonesia. Rhizomes were cleaned using running water then chopped into small pieces. Subsequently, the rhizomes were dried using oven at 50°C for 48 hours. The dried rhizomes were then ground into powder.

3.2. Preparation of Curcuma rhizome methanolic extract

The powdered rhizome of *C. longa*, *C. xanthorrhiza*, and *C. manga* were extracted using methanol pro analysis using sample to solvent ratio of 1:10. Extraction was performed using maceration technique for 3 days. The supernatant was collected and evaporated using vacuum rotary evaporator to obtain concentrated methanolic extract.

3.3. ^{1}H-NMR Analysis

Sample preparation was carried out according to Kim et al. [22] with modifications. Amount of 5 mg extract was weighed and placed into a 2 mL microtube. Subsequently the extract was added with 0.5 mL of deuterated methanol (CD_3OD) and 0.5 mL of deuterium oxide (D_2O) containing TMSP (trimethylsilyl propionic acid) 0.01%. Mixture was then vortexed for 30 s and ultrasonicated for 20 min at room temperature. Sample was then centrifuged performed at 12,000 rpm for 10 min at room temperature. Amount of 800 µL of the supernatant was taken and transferred into an NMR tube. Sample was measured using JEOL ECZ-R 500 NMR spectrometer (500 MHz). The NMR spectra acquisition was performed with the field strength of 11.74736 T, relaxation delay of 5s and X_offset of 5.0 ppm. Each spectrum was acquired for 3.53 min acquisition time which consist of 128 scans and a width of 12 ppm. Each sample was measured in three replicates.

3.4. Data Analysis

The ^{1}H-NMR spectra were analyzed using MestreNova 12.0 Software. Spectra were manually phase-corrected. Automatic baseline correction was performed using polynomial fit using degree of 3. The binning of the spectra was then performed for every 0.04 ppm from the chemical shift of 0.2 ppm-10 ppm excluding the region of residual water and methanol. Meanwhile, the chemometrics of multivariate analysis was performed using SIMCA 14.0 (Umetrics, Sweden) software.

3.5. Chemometrics Analysis

Chemometrics of pattern recognition was used to analyze the data obtained from NMR measurements, namely principal component analysis (PCA), partial least square-discriminant analysis (PLS-DA) and orthogonal projections to latent structures-discriminant analysis (OPLS-DA). The data were processed using MestreNova 12.0 software for binning to extract the ^{1}H-NMR data to obtain dataset for chemometrics analysis. The data were normalized using total area. The variables used were the intensity values from the chemical shift of 0-10 ppm excluding the area of methanol and water residual. Prior to PCA, PLS-DA, and OPLS-DA analysis, Pareto scaling was performed to the dataset. The result was observed using score plot, S-line correlation plot, variable importance in projection (VIP) value and permutation test. Variables with p(corr) value more than 0.5 observed in S-line correlation plot were important variables in OPLS-DA. Meanwhile, variables with VIP value greater than 1 is considered to have important role in samples

differentiation. In addition, evaluation using permutation test, the value of original R2 and Q2 must have the highest value among permutated models.

4. Conclusions

Authentication of Curcuma species is very important in order to ensure the quality, safety, and authenticity of the products. ^{1}H-NMR spectroscopy method could be employed in the stage of sample fingerprinting for authentication purpose both for herbal and medicinal plants. Combined with chemometrics of PCA, PLS-DA, and OPLS-DA, ^{1}H-NMR spectroscopy method becomes a powerful analytical tool for authentication of Curcuma species from different origins.

Exploratory data and classification models were successfully built. Several useful plots as output from the chemometrics models were also presented to visually assess the classification analysis. Predictive model for each species including *C. longa, C. xanthorrhiza*, and *C. manga* were evaluated according to the high values of the R^2 and Q^2. Other statistical and visual inspection by considering the ROC curve and permutation test proved the probability and performance quality of the model. Hence, it is promising and potential to develop combinational method with data fusion of ^{1}H-NMR spectroscopy and chemometrics technique for authentication of medicinal plants and herbal products.

Author Contributions: Conceptualization, A.R. and A.W.; methodology, A.R. and A.W.; writing—original draft preparation, A.W.; writing—review and editing, L.H.A., F.D.O.R., A.G., E.L., N.A.F., and M.R.; funding acquisition, A.R. All authors have read and agreed to the published version of the manuscript.

Funding: Please add: This research was funded by Ministry of Education and Culture through World Class Research, grant number 4507/UN1/DITLIT/DIT-LIT/PT/2021.

Acknowledgments: The authors thank to Faculty of Pharmacy Universitas Gadjah Mada and Integrated Research and Testing Laboratory, Universitas Gadjah Mada for facilitating instrument for this research.

Conflicts of Interest: The authors declare no conflict of interest.

Sample Availability: Samples of the extracts of *Curcuma longa, Curcuma xanthorrhiza* and *Curcuma manga* are available from the authors.

References

32. Gad, H.A.; El-Ahmady, S.H.; Abou-Shoer, M.I.; Al-Azizi, M.M. Application of chemometrics in authentication of herbal medicines: A review. *Phytochem. Anal.* **2013**, *24*, 1–24, doi:10.1002/pca.2378.
33. Srirama, R.; Santhosh Kumar, J.U.; Seethapathy, G.S.; Newmaster, S.G.; Ragupathy, S.; Ganeshaiah, K.N.; Uma Shaanker, R.; Ravikanth, G. Species adulteration in the herbal trade: Causes, consequences and mitigation. *Drug Saf.* **2017**, *40*, 651–661, doi:10.1007/s40264-017-0527-0.
34. Subositi, D.; Wahyono, S. Study of the genus curcuma in Indonesia used as traditional herbal medicines. *Biodiversitas* **2019**, *20*, 1356–1361, doi:10.13057/biodiv/d200527.
35. Rohman, A.; Rawar, E.A.; Sudevi, S.; Nurrulhidayah, A.F.; Windarsih, A. The use of chemometrics in combination with molecular spectroscopic and chromatographic methods for authentication of curcuma species: A review. *Food Res.* **2020**, *4*, 1850–1858, doi:10.26656/fr.2017.4(6).345.
36. Rohman, A.; Nugroho, A.; Lukitaningsih, E.; Sudjadi Application of vibrational spectroscopy in combination with chemometrics techniques for authentication of herbal medicine. *Appl. Spectrosc. Rev.* **2014**, *49*, 603–613, doi:10.1080/05704928.2014.882347.
37. Wang, P.; Yu, Z. Species authentication and geographical origin discrimination of herbal medicines by near infrared spectroscopy: A review. *J. Pharm. Anal.* **2015**, *5*, 277–284, doi:10.1016/j.jpha.2015.04.001.

38. Windarsih, A.; Rohman, A.; Swasono, R.T. Application of 1H-NMR based metabolite fingerprinting and chemometrics for authentication of Curcuma longa adulterated with C. heyneana. *J. Appl. Res. Med. Aromat. Plants* **2019**, *13*, 100203, doi:10.1016/j.jarmap.2019.100203.
39. Jugreet, B.S.; Mahomoodally, M.F.; Sinan, K.I.; Zengin, G.; Abdallah, H.H. Chemical variability, pharmacological potential, multivariate and molecular docking analyses of essential oils obtained from four medicinal plants. *Ind. Crops Prod.* **2020**, *150*, 112394, doi:10.1016/j.indcrop.2020.112394.
40. Lukitaningsih, E.; Rohman, A.; Rafi, M.; Nurrulhidayah, A.F.; Windarsih, A. In vivo antioxidant activities of Curcuma longa and Curcuma xanthorrhiza: A review. *Food Res.* **2020**, *4*, doi:10.26656/fr.2017.4(1).172.
41. Ayati, Z.; Ramezani, M.; Amiri, M.S.; Moghadam, A.T.; Rahimi, H.; Abdollahzade, A.; Sahebkar, A.; Emami, S.A. Ethnobotany, phytochemistry and traditional uses of Curcuma spp. and pharmacological profile of two important species (C. longa and C. zedoaria): A review. *Curr. Pharm. Des.* **2019**, *25*, 871–935, doi:10.2174/1381612825666190402163940.
42. Kharade, S.S.; Samal, K.C.; Rout, G.R. High performance thin layer chromatography fingerprint profile of rhizome extracts of five important Curcuma species. *Proc. Natl. Acad. Sci. India Sect. B - Biol. Sci.* **2017**, *87*, 1335–1341, doi:10.1007/s40011-016-0709-z.
43. Cheng, J.; Weijun, K.; Yun, L.; Jiabo, W.; Haitao, W.; Qingmiao, L.; Xiaohe, X. Development and validation of UPLC method for quality control of Curcuma longa Linn.: Fast simultaneous quantitation of three curcuminoids. *J. Pharm. Biomed. Anal.* **2010**, *53*, 43–49, doi:10.1016/j.jpba.2010.03.021.
44. Xu, L. lu; Shang, Z. peng; Lu, Y. ying; Li, P.; Sun, L.; Guo, Q. lei; Bo, T.; Le, Z. yong; Bai, Z. li; Zhang, X. li; et al. Analysis of curcuminoids and volatile components in 160 batches of turmeric samples in China by high-performance liquid chromatography and gas chromatography mass spectrometry. *J. Pharm. Biomed. Anal.* **2020**, *188*, 113465, doi:10.1016/j.jpba.2020.113465.
45. Jia, S.; Du, Z.; Song, C.; Jin, S.; Zhang, Y.; Feng, Y.; Xiong, C.; Jiang, H. Identification and characterization of curcuminoids in turmeric using ultra-high performance liquid chromatography-quadrupole time of flight tandem mass spectrometry. *J. Chromatogr. A* **2017**, *1521*, 110–122, doi:10.1016/j.chroma.2017.09.032.
46. Gismondi, A., D'Agostino, A., Marco, G. Di, Martínez-Labarga, C., Leonini, V., Rickards, O. & Canini, A. (2020). Back to the roots: dental calculus analysis of the first documented case of coeliac disease. *Archaeological and Anthropological Sciences*, **12**, 1–10, https://doi.org/10.1007/s12520-019-00962-w.
47. Trant, P.L.K.; Kristiansen, S.M.; Sindbæk, S.M. Visible near-infrared spectroscopy as an aid for archaeological interpretation. *Archaeol. Anthropol. Sci.* **2020**, *12*, 1–19, doi:10.1007/s12520-020-01239-3.
48. Wu, X.M.; Zuo, Z.T.; Zhang, Q.Z.; Wang, Y.Z. FT-MIR and UV–vis data fusion strategy for origins discrimination of wild Paris Polyphylla Smith var. yunnanensis. *Vib. Spectrosc.* **2018**, *96*, 125–136, doi:10.1016/j.vibspec.2018.04.001.
49. Eti Rohaetia, Mohamad Rafi, Utami Dyah Syafitric, R.H. Fourier transform infrared spectroscopy combined with chemometrics for discrimination of Curcuma longa, Curcuma xanthorrhiza and Zingiber cassumunar. *J. Chem. Inf. Model.* **2015**, 1244–1249.
50. Rohman, A.; Wijayanti, T.; Windarsih, A.; Riyanto, S. The authentication of Java turmeric (Curcuma xanthorrhiza) using thin layer chromatography and 1H-NMR based-metabolite fingerprinting coupled with multivariate analysis. *Molecules* **2020**, *25*, 3928, doi:10.3390/molecules25173928.
51. Deborde, C.; Fontaine, J.-X.; Jacob, D.; Botana, A.; Nicaise, V.; Richard-Forget, F.; Lecomte, S.; Decourtil, C.; Hamade, K.; Mesnard, F.; et al. Optimizing 1D 1 H-NMR profiling of plant samples for high throughput analysis: extract preparation, standardization, automation and spectra processing. *Metabolomics 2019 153* **2019**, *15*, 1–12, doi:10.1007/S11306-019-1488-3.
52. Gholkar, M.S.; Li, J.V.; Daswani, P.G.; Tetali, P.; Birdi, T.J. 1H nuclear magnetic resonance-based metabolite profiling of guava leaf extract: an attempt to develop a prototype for standardization of plant extracts. *BMC Complement. Med. Ther.* **2021**, *21*, 1–20, doi:10.1186/s12906-021-03221-5.
53. Hatumura, P.H.; de Oliveira, G.S.; Marcheafave, G.G.; Rakocevic, M.; Bruns, R.E.; Scarminio, I.S.; Terrile, A.E. Chemometric Analysis of 1H NMR Fingerprints of Coffea arabica Green Bean Extracts Cultivated under Different Planting Densities. *Food Anal. Methods* **2018**, *11*, 1906–1914, doi:10.1007/s12161-017-1104-y.
54. Belmonte-Sánchez, E.; Romero-González, R.; Garrido Frenich, A. Applicability of high-resolution NMR in combination with chemometrics for the compositional analysis and quality control of spices and plant-derived condiments. *J. Sci. Food Agric.* **2021**, *101*, 3541–3550, doi:10.1002/jsfa.11051.
55. Petrakis, E.A.; Cagliani, L.R.; Polissiou, M.G.; Consonni, R. Evaluation of saffron (Crocus sativus L.) adulteration with plant adulterants by1H NMR metabolite fingerprinting. *Food Chem.* **2015**, *173*, 890–896, doi:10.1016/j.foodchem.2014.10.107.
56. Windarsih, A.; Rohman, A.; Swasono, R.T. Application of H-NMR metabolite fingerprinting and chemometrics for the authentication of Curcuma longa adulterated with Curcuma manga. *J. Appl. Pharm. Sci.* **2018**, *8*, 75–81, doi:10.7324/JAPS.2018.8610.
57. Kim, H.K.; Choi, Y.H.; Verpoorte, R. NMR-based metabolomic analysis of plants. *Nat. Protoc.* **2010**, *5*, 536–549, doi:10.1038/nprot.2009.237.

58. Li, Y.; Kong, D.; Fu, Y.; Sussman, M.R.; Wu, H. The effect of developmental and environmental factors on secondary metabolites in medicinal plants. *Plant Physiol. Biochem.* **2020**, *148*, 80–89, doi:10.1016/j.plaphy.2020.01.006.
59. Jung, Y.; Lee, J.; Kim, H.K.; Moon, B.C.; Ji, Y.; Ryu, D.H.; Hwang, G.S. Metabolite profiling of Curcuma species grown in different regions using 1H NMR spectroscopy and multivariate analysis. *Analyst* **2012**, *137*, 5597–5606, doi:10.1039/C2AN35397K.
60. Awin, T.; Mediani, A.; Shaari, K.; Faudzi, S.M.M.; Sukari, M.A.H.; Lajis, N.H.; Abas, F. Phytochemical profiles and biological activities of Curcuma species subjected to different drying methods and solvent systems: NMR-based metabolomics approach. *Ind. Crops Prod.* **2016**, *94*, 342–352.
61. Parvathy, V.A.; Swetha, V.P.; Sheeja, T.E.; Sasikumar, B. Detection of plant-based adulterants in turmeric powder using DNA barcoding. *Pharm. Biol.* **2015**, *53*, 1774–1779, doi:10.3109/13880209.2015.1005756.
62. Worley, B.; Powers, R. Multivariate analysis in metabolomics. *Curr. Metabolomics* **2013**, *1*, 92–107, doi:10.2174/2213235x11301010092.
63. Narkhede, S. Understanding AUC - ROC curve [ONLINE]. *Available via https//towardsdatascience.com/understanding-auc-roccurve-68b2303cc9c5 (accessed 27 Sept. 2020)* **2018**.

Acceptance Letter

[Molecules] Manuscript ID: molecules-1470133 - Accepted for Publication

Inbox

Molecules Editorial Office <molecules@mdpi.com> Thu, Dec 2, 2021, 1:34 PM

to Anjar, Laela, me, Any, Florentinus, Endang, Nurrulhidayah, Mohamad, Molecules, Jelena

Dear Prof. Rohman,

Congratulations on the acceptance of your manuscript, and thank you for your interest in submitting your work to Molecules:

Manuscript ID: molecules-1470133
Type of manuscript: Article
Title: Metabolite Fingerprinting using 1H-NMR Spectroscopy and Chemometrics for Classification of Three Curcuma Species from Different Origins
Authors: Laela Hayu Nurani, Abdul Rohman *, Anjar Windarsih, Any Guntarti, Florentinus Dika Octa Riswanto, Endang Lukitaningsih, Nurrulhidayah Ahmad Fadzillah, Mohamad Rafi
Received: 3 November 2021
E-mails: laelafarmasi@yahoo.com, abdulkimfar@gmail.com, anjarwindarsih2@gmail.com, any_guntarti@yahoo.co.id, dikaocta@uad.ac.id, lukitaningsih_end@ugm.ac.id, nurrulhidayah@iium.edu.my, mra@apps.ipb.ac.id
Submitted to section: Analytical Chemistry, https://www.mdpi.com/journal/molecules/sections/Analytical_Chemistry
https://susy.mdpi.com/user/manuscripts/review_info/042ab89649528c19d9b247ee0526ffac

We will now edit and finalize your paper, which will then be returned to you for your approval. Within the next couple of days, an invoice concerning the article processing charge (APC) for publication in this open access journal will be sent by email from the Editorial Office in Basel, Switzerland.

If, however, extensive English edits are required to your manuscript, we will need to return the paper requesting improvements throughout.

We encourage you to set up your profile at SciProfiles.com, MDPI's researcher network platform. Articles you publish with MDPI will be linked to your SciProfiles page, where colleagues and peers will be able to see all of your publications, citations, as well as other academic contributions.

We also invite you to contribute to Encyclopedia (https://encyclopedia.pub), a scholarly platform providing accurate information about the latest research

results. You can adapt parts of your paper to provide valuable reference information, via Encyclopedia, for others both within the field and beyond.

Kind regards,
Farid Chemat
Editor-in-Chief

Proof Article

Article

Metabolite Fingerprinting Using ^{1}H-NMR Spectroscopy and Chemometrics for Classification of Three Curcuma Species from Different Origins

Laela Hayu Nurani [1], Abdul Rohman [2,3,*], Anjar Windarsih [4], Any Guntarti [1], Florentinus Dika Octa Riswanto [5], Endang Lukitaningsih [2], Nurrulhidayah Ahmad Fadzillah [6] and Mohamad Rafi [7]

1 Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta 55281, Indonesia; laelafarmasi@yahoo.com (L.H.N.); any_guntarti@yahoo.co.id (A.G.)
2 Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia; lukitaningsih_end@ugm.ac.id
3 Center of Excellence Institute for Halal Industry and Systems (IHIS), Universitas Gadjah Mada, Yogyakarta 55281, Indonesia; abdulkimfar@gmail.com
4 Research Division for Natural Product Technology (BPTBA), National Research and Innovation Agency (BRIN), Yogyakarta 55861, Indonesia; anjarwindarsih2@gmail.com
5 Division of Pharmaceutical Analysis and Medicinal Chemistry, Faculty of Pharmacy, Universitas Sanata Dharma, Yogyakarta 55282, Indonesia; dikaocta@uad.ac.id
6 International Institute for Halal Research and Training (INHART), International Islamic University Malaysia (IIUM), Gombak 50728, Kuala Lumpur, Malaysia; nurrulhidayah@iium.edu.my
7 Department of Chemistry, Faculty of Mathematics and Natural Sciences, IPB University, Bogor 16680, Indonesia; mra@apps.ipb.ac.id

Commented [M51]: Please carefully check the accuracy of names and affiliations.

Commented [a2R1]: Names and affiliations are correct

Commented [M53]: Please provide intitutional emails. Same applies for Ms. Anjar Windarsih in affiliation 4 and corresponding author Prof. Dr. Abdul Rohman.

Commented [a4R3]: It is fine to use this email.

Commented [M55]: Please confirm if we can delete this and leave only Kuala Lumpur with 50728 as a postcode.

Commented [a6R5]: Yes, it can be deleted

Citation: Nurani, L.H.; Rohman, A.; Windarsih, A.; Guntarti, A.; Riswanto, F.D.O.; Lukitaningsih, E.; Fadzillah, N.A.; Rafi, M. Metabolite Fingerprinting Using ^{1}H-NMR Spectroscopy and Chemometrics for Classification of Three Curcuma Species from Different Origins. *Molecules* **2021**, *26*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s): Davide Bertelli

Received: 03 November 2021
Accepted: 02 December 2021
Published: date

Commented [M57]: Please check all author names carefully.

Commented [a8R7]: All author names are correct

* Correspondence: abdulkimfar@gmail.com

Abstract: *Curcuma longa, Curcuma xanthorrhiza,* and *Curcuma manga* have been widely used for herbal or traditional medicine purposes. It was reported that turmeric plants provided several biological activities such as antioxidant, anti-inflammatory, hepatoprotector, cardioprotector, and anticancer activities. Authentication of the Curcuma species is important to ensure its authenticity and to avoid adulteration practices. Plants from different origins will have different metabolite compositions because metabolites are affected by soil nutrition, climate, temperature, and humidity. ^{1}H-NMR spectroscopy, principal component analysis (PCA), and orthogonal projections to latent structures-discriminant analysis (OPLS-DA) were used for authentication of *C. longa, C. xanthorrhiza,* and *C. manga* from seven different origins in Indonesia. From the ^{1}H-NMR analysis it was obtained that 14 metabolites were responsible for generating classification model such as curcumin, demethoxycurcumin, alanine, methionine, threonine, lysine, alpha-glucose, beta-glucose, sucrose, alpha-fructose, beta-fructose, fumaric acid, tyrosine, and formate. Both PCA and OPLS-DA model demonstrated goodness of fit (R^2 value more than 0.8) and good predictivity (Q^2 value more than 0.45). All OPLS-DA models were validated by assessing the permutation test results with high value of original R^2 and Q^2. It can be concluded that metabolite fingerprinting using ^{1}H-NMR spectroscopy and chemometrics provide a powerful tool for authentication of herbal and medicinal plants.

Keywords: authentication; curcuma; ^{1}H-NMR spectroscopy; chemometrics; metabolite
fingerprinting

1. Introduction

For hundreds of years, herbal medicines and their preparations have been widely used in folk medicines over the world. The preparation of herbal medicine preparations is typically presented either as single herbs or several herbs in a composite formulae, and it is reported that about 92% of herbal medicine formulas are a combination of less than 13 herbs [1] [2]. Annually, the market growth of herbal products has increased in which the raw material for most herbal products come from South Asian and Southeast Asian countries, including Indonesia [3]. Products of natural origin, such as supplements, herbal products, or herbal preparations, are increasingly widespread and used to maintain health or for the treatment of minor diseases. However, natural is not necessarily synonymous with safe. The adulteration of herbal preparations, together with contamination, sophistication, and degradation is a problem of global interest. In recent years, public awareness on herbal authentication and species admixtures in the raw herbal has increased significantly, because the adverse consequences of adulterated herbal components on consumer safety has been recognized [4]. Adulteration in herbal medicine involves replacing botanical materials, diluting high quality herbal medicines with lower grade ones, and mislabeling herbal medicine. Therefore, it is essential to have monitoring and pharmacovigilance systems [5].

For species authentication of herbal medicine, the World Health Organization (WHO), the United States Food and Drug Administration (USFDA), and the European Medicines Agency (EMEA) have regulated

Commented [MOU9]: Please ensure that the meaning has been maintained

Commented [a10R9]: The meaning has been maintained.

that the identification of herbal medicines should be made to ensure their quality and to discriminate them from related species or adulterated samples [6]. Among herbal medicines components, Curcuma species including *C. longa* (turmeric), *C. xanthorrhiza* (Java Turmeric), and *C. manga* have been widely applied as medicinal plants for herbal or traditional medicine purposes [7]. These Curcumas have been reported to have some biological effects which are beneficial to human health including antioxidant, anticancer anti-inflammatory, hepatoprotector, cardioprotector, antibacterial activities, and wound healing [8] [9] [10].

Some analytical methods have been developed for analysis of Curcuma species. Most of the methods use chromatography-based methods such as high performance-thin layer chromatography (TLC) [11], high performance liquid chromatography (HPLC), ultra-high performance liquid chromatography (UHPLC) [12], gas chromatography-mass spectrometry (GC-MS) for analysis volatile compounds in Curcuma species [13], and liquid chromatography-mass spectrometry (LC-MS/MS) [14]. Chromatographic-based methods typically involved complex sample preparation technique and resulted huge number of responses which make difficulty in data analysis. Therefore, spectroscopic-based methods in combination with multivariate data analysis (MDA) or chemometrics were potential to be employed since this combination method was provided the way to analyze such an environmental big data [15]. Ultraviolet, visible, and vibrational spectroscopy (infrared and Raman) [16], [17], and NMR spectroscopy [18] are widely reported for authentication of Curcuma species. NMR spectroscopy offers some advantages for authentication of medicinal plants such as fast time analysis, simple in sample preparation, high reproducibility, and high robust. Moreover, NMR spectroscopy can be used for simultaneous analysis either primary or secondary metabolites comprehensively in certain samples [19,20]. Combined with chemometrics of multivariate analysis such as principal component analysis (PCA), partial least square-discriminant analysis (PLS-DA), and orthogonal projections to latent structures-discriminant analysis (OPLS-DA) which can manage the huge data generated from NMR measurement, it becomes a powerful analytical tool for metabolite fingerprinting of medicinal plants [21,22]. Combination of ^{1}H-NMR spectroscopy and chemometrics of PLS-DA and OPLS-DA has been used for authentication of Saffron adulteration [23]. ^{1}H-NMR spectroscopy and chemometrics have also been used for authentication of *C. longa* adulterated with *C. manga* and *C. heyneana* [7,24]. Authentication of *C. xanthorrhiza* from *C. aeruginosa* has been successfully investigated using ^{1}H-NMR and multivariate analysis [18]. However, study on authentication of Curcuma species from different origins using ^{1}H-NMR spectroscopy is still limited. Therefore, the objective of this study was to use ^{1}H-NMR spectroscopy in combination with chemometrics for authentication of *C. longa, C. xanthorrhiza,* and *C. manga* from different origins.

2. Results and Discussion

2.1.1. H-NMR Spectra Analysis

^{1}H-NMR spectra can be used for authentication of medicinal plants because it offers fingerprinting which mean that each sample has specific ^{1}H-NMR spectra pattern. Generally, metabolites of plants extracted using deuterated methanol and deuterium oxide measured using ^{1}H-NMR

spectroscopy are divided into three main regions, namely amino acid and organic acids (0.20–3.00 ppm), carbohydrate or sugar (3.01–5.00 ppm), and aromatic compounds (6.00–8.00 ppm) [25]. Different origins have different conditions such as soil condition, soil nutrition (macro and micronutrients), humidity, light, salinity, and temperature as well as internal developmental genetic circuits including regulated gene, and enzyme which can obviously affect the metabolite formation either primary or secondary metabolites [26]. The ^{1}H-NMR spectra of *C. longa* (CL), *C. xanthorrhiza* (CX), and *C. manga* (CM) are shown in Figure 1. It can be observed that *C. longa*, *C. xanthorrhiza*, and *C. manga* have different spectra pattern indicating different metabolite contents. Specifically observed, *C. longa* and *C. xanthorrhiza* have higher signal intensities in the region of amino acid and organic acid (0.20–3.00 ppm) as well as in the aromatic region (6.00–8.00 ppm) than *C. manga*. On the other hand, the signal intensities in the region of glucose (3.01–5.00 ppm) are higher in *C. manga* compared to *C. longa* and *C. xanthorrhiza*. Fourteen metabolites in Curcuma species obtained from ^{1}H-NMR measurement are shown in Table 1. Curcumin, demethoxycurcumin, alanine, methionine, threonine, lysine, alpha-glucose, beta-glucose, sucrose, alpha-fructose, beta-fructose, fumaric acid, tyrosine, and formate were stated as metabolites which play important roles in generating an OPLS-DA model. Investigation on each species obtained from different regions resulted in different signal patterns, especially in intensities, indicating the variations in metabolite contents in each species from different origins. It indicated that different origins affect the metabolite contents in each Curcuma rhizome. For example, *C. longa* from Blitar (CL7) has the lowest signal intensities in the aromatic region and *C. xanthorrhiza* from Gunungkidul (CX2) has the lowest signal intensities in the whole regions. However, the spectra patterns of each species are quite similar, therefore, for deeper classification of *C. longa*, *C. xanthorrhiza*, and *C. manga* from different regions powerful statistical tool such as chemometrics is required to obtain clear classification.

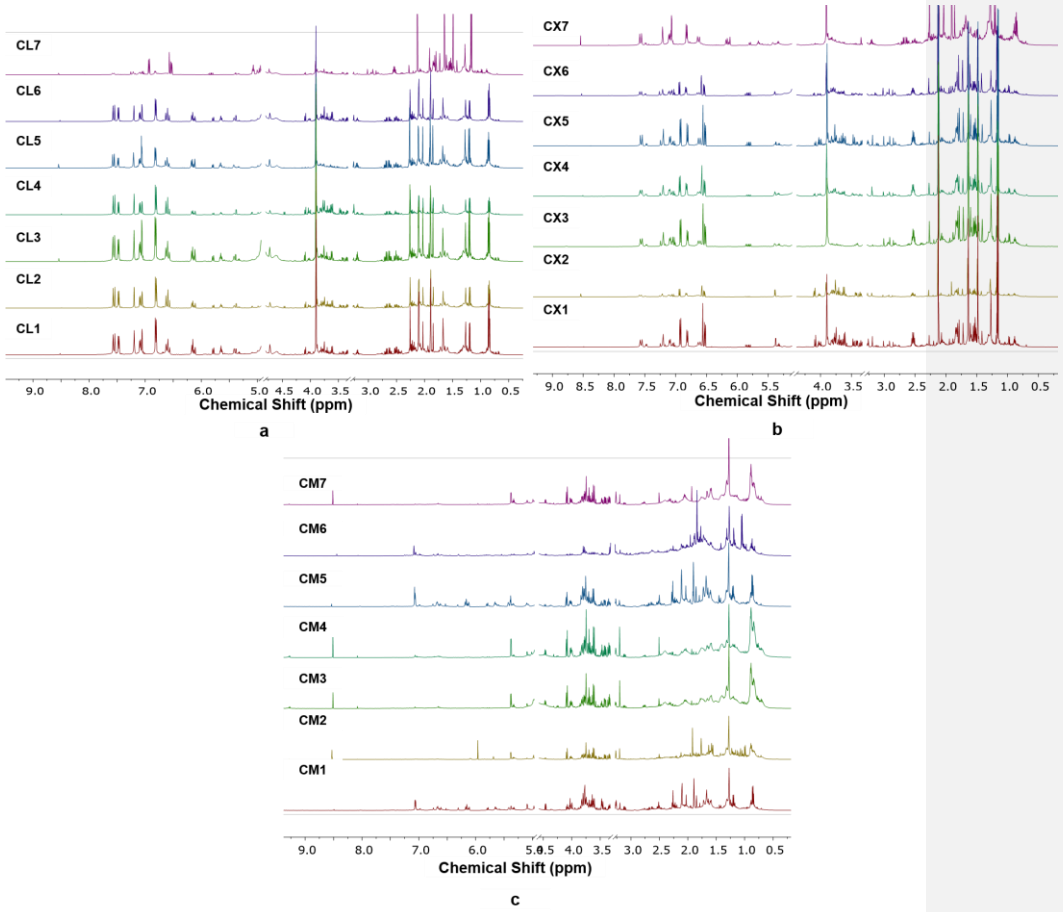


Figure 1. ^{1}H-NMR spectra of *C. longa* (**a**), *C. xanthorrhiza* (**b**), and *C. manga* (**c**) from different origins.

Table 1. Several metabolites of Curcuma species observed using ^{1}H-NMR spectra obtained from this study. The assignment of the metabolites refers to the previous published literature by Jung et.al and Awin et al. [27, 28].

No.	Chemical Shift (ppm)	Multiplicity	Metabolite
1.	7.57	Singlet	Curcumin
	7.28	Singlet	
	7.21	Doublet	
	6.77	Doublet	
	3.68	Singlet	
2.	6.92	Doublet	Demethoxycurcumin
	5.89	Singlet	

Commented [M511]: Please provide copyright permision for the data used in the Table.

Commented [a12R11]: The data we used in the table are original from our own NMR measurement, but the interpretation refers to the previous literature by Jung et al. [27] and Awin et al. [28]

	3.94	Singlet	
3.	1.49	Doublet	Alanine
	3.72	Quartet	
4.	2.11	Singlet	Methionine
5.	1.33	Doublet	Threonine
	3.53	Doublet	
6.	3.81	Triplet	Lysine
	1.5	Multiplet	
7.	5.19	Doublet	Alpha-Glucose
	3.46	Doublet of Doublet	
	3.67	Triplet	
	3.35	Triplet	
8.	4.59	Doublet	Beta-Glucose
	3.19	Doublet of Doublet	
	3.44	Triplet	
	3.71	Doublet of Doublet	
9.	5.42	Doublet	Sucrose
	3.74	Triplet	
	3.43	Triplet	
	3.80	Multiplet	
	3.84	Multiplet	
10.	4.07	Doublet	Alpha-Fructose
	3.82	Doublet of Doublet	
	3.53	Doublet	
	3.55	Doublet	
	3.63	Quartet	
11.	3.95	Multiplet	Beta-Fructose
	3.52	Doublet	
	4.02	Doublet of Doublet	
12.	6.57	Singlet	Fumaric acid
13.	6.81	Doublet	Tyrosine
	7.14	Doublet	

14.	8.42	Singlet	Formate

Curcuminoids have been reported as the active compound in Curcuma species. The content of curcuminoids is varied among Curcuma species and it is reported that curcuminoid content in *C. longa* and *C. xanthorrhiza* is higher among other Curcuma species. Curcuminoids consist of curcumin, demethoxycurcumin, and bisdemethoxycurcumin which curcumin possess the highest concentration. However, not all Curcuma species contain these three types of curcuminoids, for instance *C. xanthorrhiza* does not contain bisdemethoxycurcumin. Curcuminoids are aromatic molecules therefore most of the signals appeared in the chemical shift of aromatic regions. Curcumin signal could be observed in the chemical shift of 7.57 ppm (singlet), 7.28 ppm (singlet), 7.22 ppm (doublet), 6.77 ppm (doublet), and 3.90 ppm (singlet) whereas demethoxycurcumin could be found in the chemical shift of 6.92 ppm (doublet), 5.89 ppm (singlet), and 3.94 ppm (singlet) [28]. From the ^{1}H-NMR spectra, higher signal intensities in the aromatic region of *C. longa* and *C. xanthorrhiza* supports that curcuminoids content in *C. longa* and *C. xanthorrhiza* is higher than in *C. manga*.

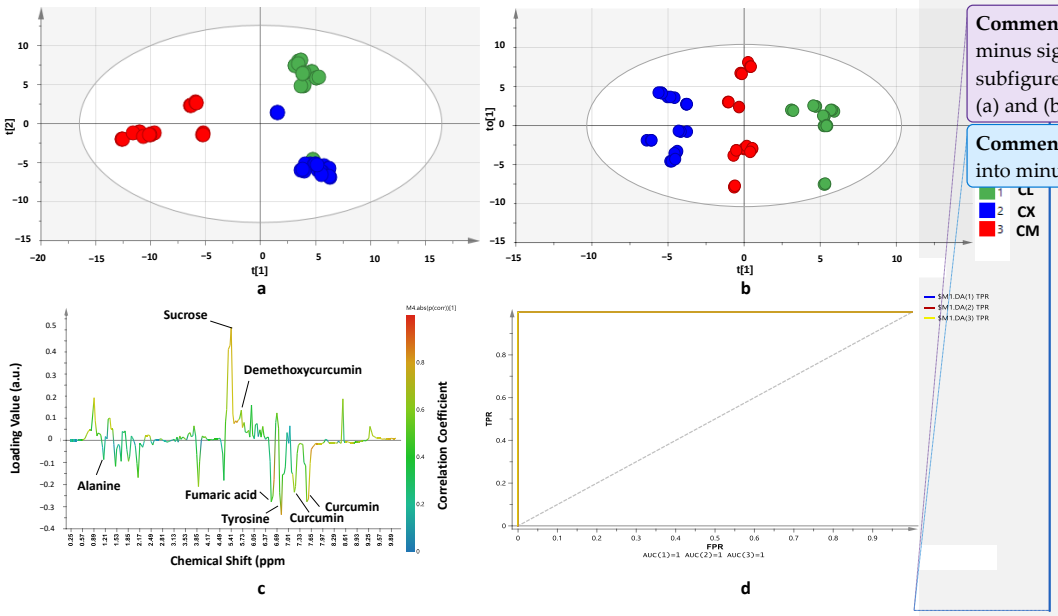
Commented [M513]: Wrong reference number: [28], previous number is [26]. Please revise all references and make sure they follow numerical order.

Commented [a14R13]: We have corrected the previous number (The previous reference in the table [23] [24], must be [27] [28] and we have revised the order of all references.

2.2. Chemometrics Analysis

Rhizomes of *C. longa*, *C. xanthorrhiza*, and *C. manga* are often used in a powder form as well as in an extract form for their herbal and traditional medicine applications. Both powder and extract are susceptible to adulteration because of their similar appearance especially in the adulterated form it is challenging to state whether the unknown sample is authentic or adulterated [29]. Chemometrics of PCA could not differentiate *C. longa*, *C. xanthorrhiza*, and *C. manga* clearly (data not shown). It might be caused by the large variations of the variables; therefore, the principal components (PC) were not able to represent the original variables. Observation using supervised pattern recognition, namely PLS-DA using 7 PC, could classify *C. longa, C. xanthorriza*, and *C. manga* resulting in three different classifications. However, several misclassifications occurred between *C. longa* and *C. xanthorrhiza* (Figure 2a). In the PLS-DA score plot, several *C. longa* samples appear in the region of *C. xanthorrhiza* and several *C. xanthorrhiza* samples appear in the region of *C. longa*. It can be explained that some of the metabolite compositions of *C. longa* and *C. xanthorrhiza* were similar especially in curcuminoid contents in which curcumin and demethoxycurcumin were the major active compounds in *C. longa* and *C. xanthorrhiza*. In addition, it is often reported that adulteration or substitution of *C. longa* with *C. xanthorrhiza* is often difficult to detect because the appearance of *C. longa* and *C. xanthorrhiza* in powder and extract form are quite similar [17]. Therefore, another supervised pattern recognition chemometrics, namely OPLS-DA, was performed to obtain better classification of *C. longa, C. xanthorrhiza*, and *C. manga* extracts. The OPLS-DA model demonstrated good capability to differentiate three different species of *C. longa*, *C. heyneana*, and *C. manga* from different origins as shown in the OPLS-DA score plot (Figure 2b). The OPLS-DA model successfully classified *C. longa*, *C. xanthorrhiza*, and *C. manga* samples. The samples were successfully classified using first PC and first X-orthogonal components which accounted for 80.8% of the variance with R^2X (cum) of 0.808, R^2Y (cum) of 0.776, and Q^2 (cum) of 0.767. A high value of R^2X (cum) and R^2Y (cum) (close to 1) indicated goodness of fit of the OPLS-DA model,

whereas the value of Q^2 greater than 0.45 indicated goodness of predictivity of the models [30]. The S-line correlation plot (Figure 2c) variables which have roles in the differentiation of *C. longa*, *C. xanthorrhiza*, and *C. manga*. It was found that alanine, curcumin, demethoxycurcumin, fumaric acid, sucrose, and tyrosine had p (corr) values of more than 0.5, indicating their important roles in separating samples. Moreover, using variable importance in projection (VIP) value, chemical shifts of 6.77, 3.89, 7.57, 6.81, 6.57, 7.21, 1.49, 6.49, 6.13, 0.85, 6.09, 5.29, 5.25, and 6.92 ppm were found to have important roles for the classification between three Curcuma species in OPLS-DA models. Variables with VIP values greater than 1 are considered to have important roles for differentiation. Some of the variables correspond to the metabolites of curcumin, tyrosine, fumaric acid, alanine, and demethoxycurcumin. The receiver operating characteristic curve (Figure 2d) for differentiating and classifying Curcuma longa, Curcume xanthorrhiza, and Curcuma manga from different origins was also depicted. The ROC analysis represents the probability of the model by plotting the value of true positivity rate (TPR) against the value of false positivity rate (FPR) [31].



Commented [M515]: Please change hyphen into minus sign in front of the numbers on Y-axis in subfigures (a), (b) and (c) and X-axis in subfigures (a) and (b).

Commented [a16R15]: We have changed hyphen into minus sign in subfigures (a), (b), and (c).

Figure 2. PLS-DA score plot (**a**), OPLS-DA score plot (**b**), OPLS-DA S-line correlation plot (**c**), and ROC curve (**d**) for differentiation and classification *of C. longa, C. xanthorrhiza,* and *C. manga* from different origins. Statistical parameters of the models: (a) Number of samples = 42; $R^2X = 0.603$; $R^2Y = 0.783$; $Q^2 = 0.755$ for t[1] and t[2] components; (b,c) Number of samples = 42; $R^2X = 0.808$; $R^2Y = 0.776$; $Q^2 = 0.767$ for t[1] and to[1] components.

PCA using number of PC 8 could differentiate *C. longa* from seven different origins as shown in the PCA score plot (Figure 3a). The PCA model provided high confidence for its fitting and predictivity capacity, shown by its R^2 value (0.770) and Q^2 value (0.650) for PC1 and PC2, respectively, accounting for 77.0% of the variance. The score plots which

appear close to each other indicates high similarity between samples, especially their metabolite compositions. The PCA score plot result shows that CL2 and CL4 possessed high similarity and CL1 has high similarity with CL6. Meanwhile, CL3 appeared closely to CL5. CL7 appeared far from all CL samples from other regions meaning that the metabolites composition of CL7 differs from other *C. longa* used in this research. Classification of *C. longa* samples from seven different origins using OPLS-DA demonstrated different pattern with PCA result (Figure 3b). OPLS-DA was created using first PC and first orthogonal-X component resulting R^2X (cum) of 0.748, R^2Y (cum) of 0.754 and Q^2 (cum) of 0.639. The first PC and first X-orthogonal component explained 74.8% of the total variance. There were three main groups obtained from OPLS-DA classification. The first group was CL2 and CL7 which appeared close to each other. The second group consisted of CL1 and CL5, and the last group of CL3, CL4, and CL6. The important variables for differentiation and classification of *C. longa* between groups observed using S-line correlation plot (Figure 3c) were alanine, β-fructose, curcumin, demethoxycurcumin, fumaric acid, and tyrosine. Meanwhile, investigation using VIP value found that variables of 6.92, 6.54, 6.57, 1.50, 1.66, 1.54, 1.18, 7.46, 6.62, 0.86, 7.10, 7.50, 7.57, and 7.14 ppm had important roles for *C. longa* differentiation and classification from seven different origins classified using OPLS-DA model. Some of the variables correspond to the molecule signals of curcumin, demethoxycurcumin, fumaric acid, lysine, and tyrosine. It demonstrated that different origins affect the composition of some metabolites. It is in accordance with research by Jung et al. [27] on the metabolite compositions of *C. longa* from several regions in China. The condition of geographical origin and environmental conditions such as temperature, humidity, and rainfall rate affect the metabolite composition of plants.

Commented [M517]: Wrong reference number: [28], previous number is [31]. Please revise all references and make sure they follow numerical order.

Commented [a18R17]: We have revised the order of all references.

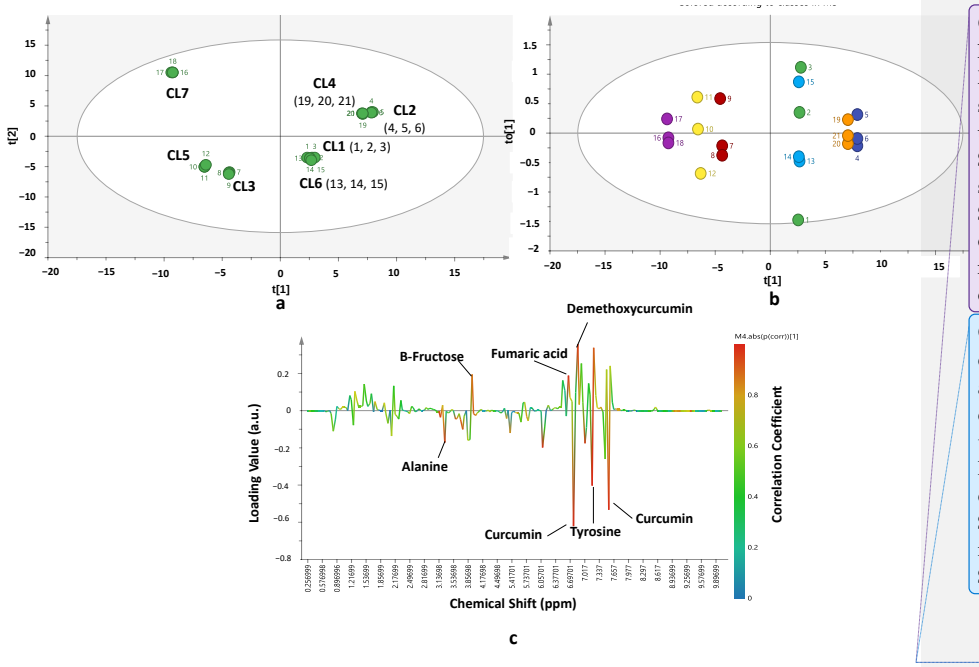


Figure 3. PCA score plot (**a**) OPLS-DA score plot (**b**) and OPLS-DA S-line correlation plot (**c**) for differentiation and classification *of C. longa* from different origins. Statistical parameter of the models: (a,) Number of samples = 21; Number of PC = 8; $R^2X = 0.770$; $Q^2 = 0.650$; (b,c) Number of samples = 21; $R^2X = 0.748$; $R^2Y = 0.754$; $Q^2 = 0.639$ for t[1] and to[1] components.

Chemometrics of PCA and OPLS-DA was also successfully used for differentiation and classification of *C. xanthorrhiza* from seven different origins. The result of differentiation of samples using PCA was slightly different with the result from OPLS-DA. PCA performed using first and second principal components demonstrated goodness of fit (R^2cum = 0.743) and predictivity (Q^2cum = 0.678) with total variance of 74.3%.Meanwhile, OPLS-DA was created using first PC and first orthogonal-X components. A high value of R^2X (cum) (0.743) and R^2Y (cum) (0.833) indicated good model fitness while a high value of Q^2 (cum) (0.626) indicated predictivity of the OPLS-DA model. The first PC and first X-orthogonal-X component demonstrated 74.3% of the total variance. There were four main groups which appeared in the PCA score plot (Figure 4a), whereas in OPLS-DA (Figure 4b) there were three groups. *C. xanthorrhiza* of CX3 and CX5 appeared close to each other in both PCA and OPLS-DA score plot results. Meanwhile, *C. xanthorrhiza* of CX1 was found in a separate group with others observed both in PCA and OPLS-DA. Samples of CX2, CX4, CX6, and CX7 appeared in the same group observed using OPLS-DA; however, from the PCA result, a sample of CX7 appeared in a different group. From these results, it is suggested that different locations have significant effects on metabolites' compositions in *C. xanthorrhiza*. The S-line correlation plot (Figure

Commented [M519]: Subfigure (a): Please change hyphen into minus sign in front of the number on X and Y-axis. Numbers for CL4, CL2 in the same subfigure are overlapping which makes them unclear. Please revise this subfigre if possible. Subfigure (b): Please change hyphen into minus sign in front of the number on X and Y-axis. Subfigure (c): Explanation for Y-axis and numbers on Y-axis are overlapping. Please revise and if number are negative ones, use minus sign instead of hyphen.

Commented [a20R19]: Subfigure (a): We have changed hyphen into minus sign. Plots which appeared overlapping indicated they have similar characteristics such as similar metabolites content, therefore the plot appeared overlap each other. For the overlapping plots, we have revised the code to make them clear.
Subfigure (b): We have changed hyphen into minus sign.
Subfigure (c): We have revised subfigure (c).

4c) shows that methionine, β-glucose, sucrose, fumaric acid, curcumin, demethoxycurcumin, and tyrosine were the important variables for CX differentiation. Moreover, using a VIP value, it can be found that sthe variables important for classifying samples were 4.61, 6.57, 6.92, 4.59, 2.13, 6.92, 7.01, 7.21, 6.53, 1.69, 1.25, and 1.77 ppm. Some of the variables corresponded to the metabolites of curcumin, fumaric acid, demethoxycurcumin, and beta-glucose. It is presumed that these metabolites have higher scores and significantly affect the differentiation and classification of *C. xanthorrhiza* from different origins.

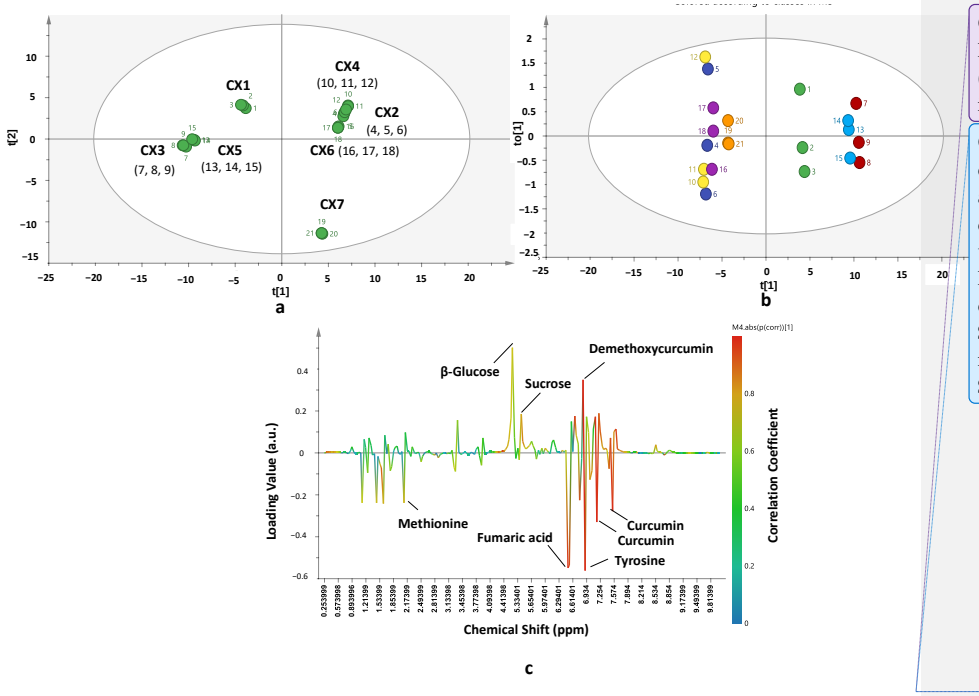


Figure 4. PCA score plot (**a**), OPLS-DA score plot (**b**), and OPLS-DA S-line correlation plot (**c**) for differentiation and classification *of C. xanthorrhiza* from different origins. Statistical parameter of the models: (a) Number of samples = 21; Number of PC = 6; $R^2X = 0.743$; Q2 = 0.678; (b,c) Number of samples = 21; $R^2X = 0.743$; $R^2Y = 0.833$; $Q^2 = 0.626$ for t[1] and to[1] components.

Chemometrics of PCA and OPLS-DA was also successfully applied for differentiation and classification of *C. manga* from seven different origins. Different classification results were observed between PCA and OPLS-DA. The PCA model was created using first and second principal components resulting R^2X (cum) of 0.661 and Q^2 (cum) of 0.501 indicating goodness of fit and good predictivity of the PCA model, respectively. The first and second PCs showed 66.1% of the total variance. From the PCA score plot (Figure 5a), *C. manga* were classified in five classes as follows: CM1 (first class), CM5 (second class), CM6 (third class), CM2 (fourth class), and the rest of the samples was in the last class (CM3, CM4, and CM7). Samples of CM3, CM4, and CM7 have similar chemical or

Commented [M521]: Same comment as for the Figure 3. Please make sure nothing overlaps (subfigure a) and that all hyphens are changed into minus sign.

Commented [a22R21]: Subfigure (a): We have changed hyphen into minus sign. Plots which appeared overlapping indicated they have similar characteristics such as similar metabolites content, therefore the plot appeared overlap each other. For the overlapping plots, we have revised the code to make them clear.
Subfigure (b): We have changed hyphen into minus sign.
Subfigure (c): We have revised subfigure (c).

Commented [M523]: Figure 4 is not cited in the main text. Please revise it and make sure that the Figure 4 placed after its first citation.

Commented [a24R23]: We have cited Figure 4 in the main text.

metabolite compositions because they appeared in the same location in PCA score plot. It is presumed that the conditions in the region of Malang (CM3), Tulung Agung (CM4), and Blitar (CM7) are similar resulting in the similar metabolites of *C. manga* rhizomes. On the other hand, OPLS-DA was performed using first principal components and first orthogonal-X component which presented 66.1% of the total variance. The obtained R^2X (cum) (0.661) and R^2Y (cum) (0.667) indicated goodness of fit whereas the value of Q^2 (cum) (0.707) demonstrated goodness of model predictivity. Three main groups were found in the OPLS-DA score plot (Figure 5b), namely CM1 and CM6 as the first group, CM5 in the second group, and CM2, CM3, CM4, and CM7 in the last group. Observation using an S-line correlation plot (Figure 5c) demonstrated that β-glucose, sucrose, curcumin, tyrosine, and format had important roles for CM differentiation. The VIP value showed that some variables were found to have significant contributions in the differentiation of *C. manga* samples from different origins, namely: 7.05, 5.29, 5.97, 5.33, 5.37, 8.53, 1.25, 1.81, 5.42, 5.65, 7.09, 6.17, 6.69, 0.85, 0.89, 3.33, and 5.41 ppm. Some of the variables are associated with curcumin, demethoxycurcumin, sucrose, and fumaric acid.

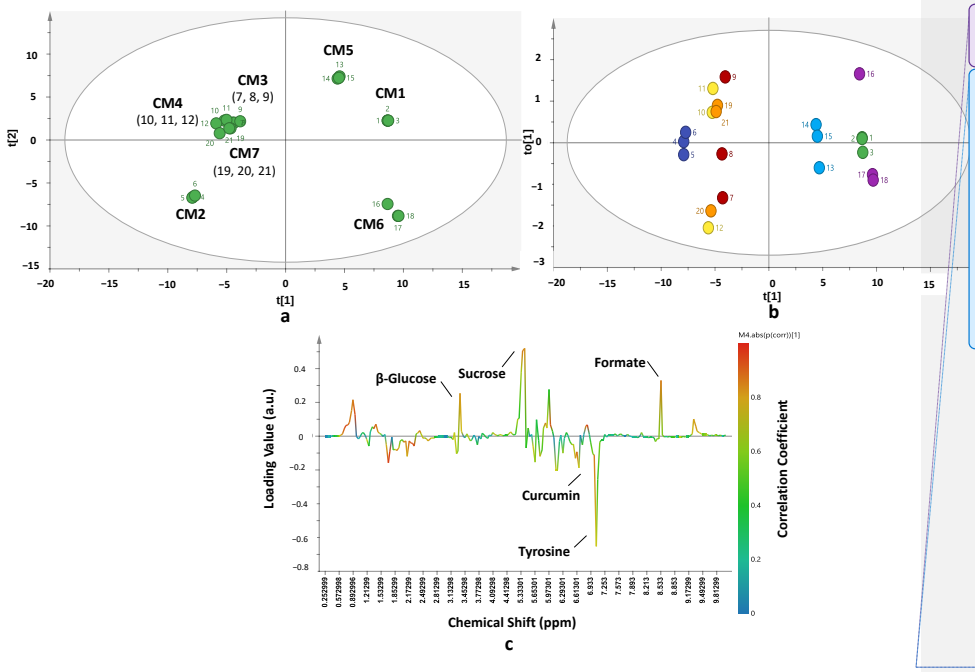


Figure 5. PCA score plot (**a**) OPLS-DA score plot (**b**) and OPLS-DA S-line correlation plot (**c**) for differentiation and classification *of C. manga* from different origins. Statistical parameter of the models: (a) Number of samples = 21; Number of PC = 7; R^2X = 0.661; Q^2 = 0.501; (b,c) Number of samples = 21; R^2X = 0.661; R^2Y = 0.667; Q^2 = 0.707 for t[1] and to[1] components.

2.3. Validation of OPLS-DA Model. Using Permutation Test.

Supervised pattern recognition of chemometrics such as PLS-DA and OPLS-DA requires a test to confirm the model's validity because of

Commented [M525]: Same comment as for the Figures 3 and 4.

Commented [a26R25]: Subfigure (a): We have changed hyphen into minus sign. Plots which appeared overlapping indicated they have similar characteristics such as similar metabolites content, therefore the plot appeared overlap each other. For the overlapping plots, we have revised the code to make them clear.
Subfigure (b): We have changed hyphen into minus sign.
Subfigure (c): We have revised subfigure (c).

Commented [M527]: Figure 5 is not cited in the main text. Please revise it and make sure that the Figure 5 placed after its first citation.

Commented [a28R27]: We have cited Figure 5 in the main text.

its potential for overfitting. Validation is a confirmation step to ensure that the models have goodness of fit. A permutation test is one of validation testing which used a permutated model. Models of R^2 and Q^2 are permutated and compared to the original models of R^2 and Q^2. A good model is obtained when all the permutated models of R^2 and Q^2 values are lower than the R^2 and Q^2 original values. Moreover, the validation was also determined using intersection value of Q^2. The intersection value should be zero or lower than zero to be categorized as valid models. The result of the permutation test from 999 permutations of OPLS-DA models were demonstrated in Figure 6. The permutated models of R^2 and Q^2 are on the left side while the original R^2 and Q^2 models are on the right side. The models were permutated for 100 permutations. Results showed that all permutation tests confirmed the validity of the OPLS-DA model demonstrated by the value of R^2 and Q^2 in all permutated models being below the value of original R^2 and Q^2 models. On the other hand, the intersection values of Q^2 for all four OPLS-DA models were also zero and lower than zero, as follows: (0.0, −0.473) for a classification model between three Curcuma species; (0.0, −1.02) for classification model of *C. longa* from different origins; (0.0, −0.896), for classification model of *C. xanthorrhiza* from different origins; and (0.0, −0.904) for classification model of *C. manga* from different origins. It is suggested that OPLS-DA could be used as a powerful statistical tool for classification of different Curcuma species from different origins with high validity.

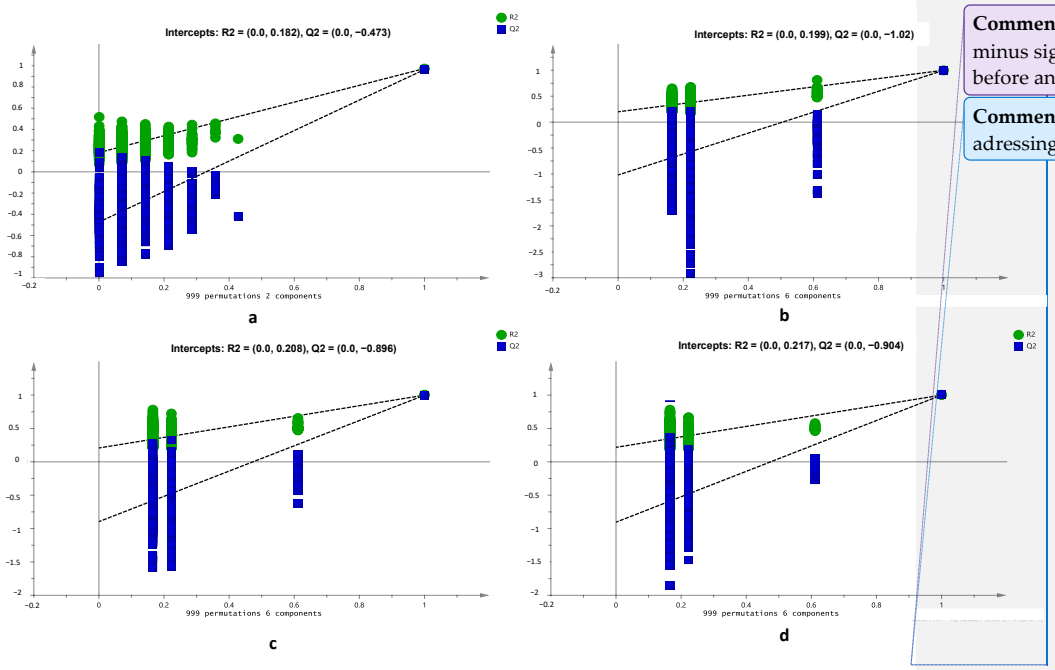


Figure 6. Permutation test using 999 permutations of OPLS-DA result from three Curcuma species (**a**), *C. longa* from different origins (**b**), *C. xanthorrhiza* from different origins (**c**), and *C. manga* from different origins (**d**).

Commented [M529]: Please change hyphen into a minus sign where it is needed and add spaces before and after = (e.g., R2 = (0.0, 0.182)

Commented [a30R29]: We have made revisions adressing the comments.

3. Materials and Methods

3.1. Sample Collection and Preparation

Rhizome of *C. longa* (CL), *C. xanthorrhiza* (CX), and *C. manga* (CM) were collected from seven different regions in Indonesia, namely: Boyolali (1), Gunungkidul (2), Ngawi (3), Malang (3), Tulung Agung (5), Karang Anyar (6), and Blitar (7). Determination of plant species used in this study has been carried out at the Pharmaceutical Biology Department, Faculty of Pharmacy, Universitas Gadjah Mada, Indonesia. Rhizomes were cleaned using running water then chopped into small pieces. Subsequently, the rhizomes were dried using an oven at 50 °C for 48 h. The dried rhizomes were then ground into powder.

3.2. Preparation of Curcuma Rhizome Methanolic Extract

The powdered rhizome of *C. longa*, *C. xanthorrhiza*, and *C. manga* were extracted using methanol pro analysis using sample to solvent ratio of 1:10. Extraction was performed using a maceration technique for 3 days. The supernatant was collected and evaporated using a vacuum rotary evaporator to obtain a concentrated methanolic extract.

3.3. ^{1}H-NMR Analysis

Sample preparation was carried out according to Kim et al. [22] with modifications. An amount of 5 mg extract was weighed and placed into a 2 mL microtube. Subsequently the extract was added to 0.5 mL of deuterated methanol (CD_3OD) and 0.5 mL of deuterium oxide (D_2O) containing TMSP (trimethylsilyl propionic acid) 0.01%. The mixture was vortexed for 30 s and ultrasonicated for 20 min at room temperature. The sample was then centrifuged at 12,000 rpm for 10 min at room temperature. An amount of 800 µL of the supernatant was taken and transferred into an NMR tube. The sample was measured using a JEOL ECZ-R 500 MHz NMR spectrometer (JEOL, Tokyo, Japan). The NMR spectra acquisition was performed with the field strength of 11.74736 T, relaxation delay of 5s, and X_offset of 5.0 ppm. Each spectrum was acquired for a 3.53 min acquisition time which consisted of 128 scans and a width of 12 ppm. Each sample was measured in three replicates.

Commented [M531]: Please state manufacturer, city and country from where equipment has been sourced.

Commented [a32R31]: We have revised it.

3.4. Data Analysis

The ^{1}H-NMR spectra were analyzed using MestreNova 12.0 Software (Mestrelab Research, S.L., Santiago de Compostela, Spain). Spectra were manually phase-corrected. Automatic baseline correction was performed using polynomial fit using degree of 3. The binning of the spectra was then performed for every 0.04 ppm from the chemical shift of 0.2–10 ppm excluding the region of residual water and methanol. Meanwhile, the chemometrics of multivariate analysis was performed using SIMCA 14.0 (Umetrics, Umeå, Sweden) software.

Commented [M533]: Please state manufacturer, city and country from where equipment has been sourced. In case this is an online program, please add link and accessed date.

Commented [a34R33]: We have revised it.

3.5. Chemometrics Analysis

Chemometrics of pattern recognition were used to analyze the data obtained from NMR measurements, namely principal component analysis (PCA), partial least square-discriminant analysis (PLS-DA), and orthogonal projections to latent structures-discriminant analysis (OPLS-DA). The data were processed using MestreNova 12.0 software for binning to extract the ^{1}H-NMR data to obtain a dataset for chemometrics analysis. The data were normalized using total area. The variables used

were the intensity values from the chemical shift of 0–10 ppm excluding the area of methanol and water residual. Prior to PCA, PLS-DA, and OPLS-DA analysis, Pareto scaling was performed to the dataset. The result was observed using a score plot, S-line correlation plot, variable importance in projection (VIP) value, and permutation test. Variables with a p (corr) value of more than 0.5 observed in an S-line correlation plot were important variables in OPLS-DA. Meanwhile, variables with a VIP value greater than 1 were considered to have important role in samples' differentiation. In addition, in evaluation using a permutation test, the value of the original R2 and Q2 must have the highest value among permutated models.

4. Conclusions

Authentication of Curcuma species is important to ensure the quality, safety, and authenticity of the products. ^{1}H-NMR spectroscopy method could be employed at the stage of sample fingerprinting for authentication purpose both for herbal and medicinal plants. Combined with chemometrics of PCA, PLS-DA, and OPLS-DA, ^{1}H-NMR spectroscopy method is a powerful analytical tool for authentication of Curcuma species from different origins.

Exploratory data and classification models were successfully built. Several useful plots of output from the chemometrics models were also presented to visually assess the classification analysis. Predictive models for each species including *C. longa, C. xanthorrhiza*, and *C. manga* were evaluated according to the high values of the R^2 and Q^2. Other statistical and visual observations made considering the ROC curve and permutation test proved the probability and performance quality of the model. Hence, there is promise and potential to develop a combinational method with data fusion of ^{1}H-NMR spectroscopy and chemometrics technique for the authentication of medicinal plants and herbal products.

Author Contributions: Conceptualization, A.R. and A.W.; methodology, A.R. and A.W.; writing—original draft preparation, A.W., L.H.N.; writing—review and editing, L.H.N., F.D.O.R., A.G., E.L., N.A.F., and M.R.; funding acquisition, A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Ministry of Education and Culture through World Class Research, grant number 4507/UN1/DITLIT/DIT-LIT/PT/2021.

Data Availability Statement: The data presented in this study are available in the article.

Acknowledgments: The authors thank to Faculty of Pharmacy Universitas Gadjah Mada and Integrated Research and Testing Laboratory, Universitas Gadjah Mada for facilitating instrument for this research.

Conflicts of Interest: The authors declare no conflict of interest.

Sample Availability: Samples of the extracts of *Curcuma longa, Curcuma xanthorrhiza* and *Curcuma manga* are available from the authors.

References

64. Gad, H.A.; El-Ahmady, S.H.; Abou-Shoer, M.I.; Al-Azizi, M.M. Application of chemometrics in authentication of herbal medicines: A review. *Phytochem. Anal.* **2013**, *24*, 1–24, doi:10.1002/pca.2378.
65. Srirama, R.; Santhosh Kumar, J.U.; Seethapathy, G.S.; Newmaster, S.G.; Ragupathy, S.; Ganeshaiah, K.N.; Uma Shaanker, R.; Ravikanth, G. Species adulteration in the herbal trade: Causes, consequences and mitigation. *Drug Saf.* **2017**, *40*, 651–661, doi:10.1007/s40264-017-0527-0.

Commented [M535]: In this section, please provide details regarding where data supporting reported results can be found, including links to publicly archived datasets analyzed or generated during the study. Please refer to suggested Data Availability Statements in section "MDPI Research Data Policies" at https://www.mdpi.com/ethics. You might choose to exclude this statement if the study did not report any data.

Commented [a36R35]: We have added information regarding Data Availability Statement.

Commented [M537]: Citation for Reference 32 is missing. Please revise all references throughout the whole text and make sure they follow numerical order.

Commented [a38R37]: We have checked and revised the order of all references. Reference 16 is not included in this manuscript and we have removed it, so the total of the references is 31 reference.

66. Subositi, D.; Wahyono, S. Study of the genus curcuma in Indonesia used as traditional herbal medicines. *Biodiversitas* **2019**, *20*, 1356–1361, doi:10.13057/biodiv/d200527.
67. Rohman, A.; Rawar, E.A.; Sudevi, S.; Nurrulhidayah, A.F.; Windarsih, A. The use of chemometrics in combination with molecular spectroscopic and chromatographic methods for authentication of curcuma species: A review. *Food Res.* **2020**, *4*, 1850–1858, doi:10.26656/fr.2017.4(6).345.
68. Rohman, A.; Nugroho, A.; Lukitaningsih, E.; Sudjadi Application of vibrational spectroscopy in combination with chemometrics techniques for authentication of herbal medicine. *Appl. Spectrosc. Rev.* **2014**, *49*, 603–613, doi:10.1080/05704928.2014.882347.
69. Wang, P.; Yu, Z. Species authentication and geographical origin discrimination of herbal medicines by near infrared spectroscopy: A review. *J. Pharm. Anal.* **2015**, *5*, 277–284, doi:10.1016/j.jpha.2015.04.001.
70. Windarsih, A.; Rohman, A.; Swasono, R.T. Application of 1H-NMR based metabolite fingerprinting and chemometrics for authentication of Curcuma longa adulterated with C. heyneana. *J. Appl. Res. Med. Aromat. Plants* **2019**, *13*, 100203, doi:10.1016/j.jarmap.2019.100203.
71. Jugreet, B.S.; Mahomoodally, M.F.; Sinan, K.I.; Zengin, G.; Abdallah, H.H. Chemical variability, pharmacological potential, multivariate and molecular docking analyses of essential oils obtained from four medicinal plants. *Ind. Crops Prod.* **2020**, *150*, 112394, doi:10.1016/j.indcrop.2020.112394.
72. Lukitaningsih, E.; Rohman, A.; Rafi, M.; Nurrulhidayah, A.F.; Windarsih, A. In vivo antioxidant activities of Curcuma longa and Curcuma xanthorrhiza: A review. *Food Res.* **2020**, *4*, doi:10.26656/fr.2017.4(1).172.
73. Ayati, Z.; Ramezani, M.; Amiri, M.S.; Moghadam, A.T.; Rahimi, H.; Abdollahzade, A.; Sahebkar, A.; Emami, S.A. Ethnobotany, phytochemistry and traditional uses of Curcuma spp. and pharmacological profile of two important species (C. longa and C. zedoaria): A review. *Curr. Pharm. Des.* **2019**, *25*, 871–935, doi:10.2174/1381612825666190402163940.
74. Kharade, S.S.; Samal, K.C.; Rout, G.R. High performance thin layer chromatography fingerprint profile of rhizome extracts of five important Curcuma species. *Proc. Natl. Acad. Sci. India Sect. B—Biol. Sci.* **2017**, *87*, 1335–1341, doi:10.1007/s40011-016-0709-z.
75. Cheng, J.; Weijun, K.; Yun, L.; Jiabo, W.; Haitao, W.; Qingmiao, L.; Xiaohe, X. Development and validation of UPLC method for quality control of Curcuma longa Linn.: Fast simultaneous quantitation of three curcuminoids. *J. Pharm. Biomed. Anal.* **2010**, *53*, 43–49, doi:10.1016/j.jpba.2010.03.021.
76. Xu, L. lu; Shang, Z. peng; Lu, Y. ying; Li, P.; Sun, L.; Guo, Q. lei; Bo, T.; Le, Z. yong; Bai, Z. li; Zhang, X. li; et al. Analysis of curcuminoids and volatile components in 160 batches of turmeric samples in China by high-performance liquid chromatography and gas chromatography mass spectrometry. *J. Pharm. Biomed. Anal.* **2020**, *188*, 113465, doi:10.1016/j.jpba.2020.113465.
77. Jia, S.; Du, Z.; Song, C.; Jin, S.; Zhang, Y.; Feng, Y.; Xiong, C.; Jiang, H. Identification and characterization of curcuminoids in turmeric using ultra-high performance liquid chromatography-quadrupole time of flight tandem mass spectrometry. *J. Chromatogr. A* **2017**, *1521*, 110–122, doi:10.1016/j.chroma.2017.09.032.
78. Gismondi, A., D'Agostino, A., Marco, G. Di, Martínez-Labarga, C., Leonini, V., Rickards, O. & Canini, A. (2020). Back to the roots: Dental calculus analysis of the first documented case of coeliac disease. *Archaeological and Anthropological Sciences*, **12**, 1–10, https://doi.org/10.1007/s12520-019-00962-w.
79. Wu, X.M.; Zuo, Z.T.; Zhang, Q.Z.; Wang, Y.Z. FT-MIR and UV–vis data fusion strategy for origins discrimination of wild Paris Polyphylla Smith var. yunnanensis. *Vib. Spectrosc.* **2018**, *96*, 125–136, doi:10.1016/j.vibspec.2018.04.001.
80. Eti Rohaetia, Mohamad Rafi, Utami Dyah Syafitric, R.H. Fourier transform infrared spectroscopy combined with chemometrics for discrimination of Curcuma longa, Curcuma xanthorrhiza and Zingiber cassumunar. *J. Chem. Inf. Model.* **2015**, 1244–1249.
81. Rohman, A.; Wijayanti, T.; Windarsih, A.; Riyanto, S. The authentication of Java turmeric (Curcuma xanthorrhiza) using thin layer chromatography and 1H-NMR based-metabolite fingerprinting coupled with multivariate analysis. *Molecules* **2020**, *25*, 3928, doi:10.3390/molecules25173928.
82. Deborde, C.; Fontaine, J.-X.; Jacob, D.; Botana, A.; Nicaise, V.; Richard-Forget, F.; Lecomte, S.; Decourtil, C.; Hamade, K.; Mesnard, F.; et al. Optimizing 1D 1 H-NMR profiling of plant samples for high throughput analysis: Extract preparation, standardization, automation and spectra processing. *Metabolomics 2019 153* **2019**, *15*, 1–12, doi:10.1007/S11306-019-1488-3.
83. Gholkar, M.S.; Li, J.V.; Daswani, P.G.; Tetali, P.; Birdi, T.J. 1H nuclear magnetic resonance-based metabolite profiling of guava leaf extract: An attempt to develop a prototype for standardization of plant extracts. *BMC Complement. Med. Ther.* **2021**, *21*, 1–20, doi:10.1186/s12906-021-03221-5.
84. Hatumura, P.H.; de Oliveira, G.S.; Marcheafave, G.G.; Rakocevic, M.; Bruns, R.E.; Scarminio, I.S.; Terrile, A.E. Chemometric Analysis of 1H NMR Fingerprints of Coffea arabica Green Bean Extracts Cultivated under Different Planting Densities. *Food Anal. Methods* **2018**, *11*, 1906–1914, doi:10.1007/s12161-017-1104-y.
85. Belmonte-Sánchez, E.; Romero-González, R.; Garrido Frenich, A. Applicability of high-resolution NMR in combination with chemometrics for the compositional analysis and quality control of spices and plant-derived condiments. *J. Sci. Food Agric.* **2021**, *101*, 3541–3550, doi:10.1002/jsfa.11051.

86. Petrakis, E.A.; Cagliani, L.R.; Polissiou, M.G.; Consonni, R. Evaluation of saffron (Crocus sativus L.) adulteration with plant adulterants by1H NMR metabolite fingerprinting. *Food Chem.* **2015**, *173*, 890–896, doi:10.1016/j.foodchem.2014.10.107.
87. Windarsih, A.; Rohman, A.; Swasono, R.T. Application of H-NMR metabolite fingerprinting and chemometrics for the authentication of Curcuma longa adulterated with Curcuma manga. *J. Appl. Pharm. Sci.* **2018**, *8*, 75–81, doi:10.7324/JAPS.2018.8610.
88. Kim, H.K.; Choi, Y.H.; Verpoorte, R. NMR-based metabolomic analysis of plants. *Nat. Protoc.* **2010**, *5*, 536–549, doi:10.1038/nprot.2009.237.
89. Li, Y.; Kong, D.; Fu, Y.; Sussman, M.R.; Wu, H. The effect of developmental and environmental factors on secondary metabolites in medicinal plants. *Plant. Physiol. Biochem.* **2020**, *148*, 80–89, doi:10.1016/j.plaphy.2020.01.006.
90. Jung, Y.; Lee, J.; Kim, H.K.; Moon, B.C.; Ji, Y.; Ryu, D.H.; Hwang, G.S. Metabolite profiling of Curcuma species grown in different regions using 1H NMR spectroscopy and multivariate analysis. *Analyst* **2012**, *137*, 5597–5606, doi:10.1039/C2AN35397K.
91. Awin, T.; Mediani, A.; Shaari, K.; Faudzi, S.M.M.; Sukari, M.A.H.; Lajis, N.H.; Abas, F. Phytochemical profiles and biological activities of Curcuma species subjected to different drying methods and solvent systems: NMR-based metabolomics approach. *Ind. Crops Prod.* **2016**, *94*, 342–352.
92. Parvathy, V.A.; Swetha, V.P.; Sheeja, T.E.; Sasikumar, B. Detection of plant-based adulterants in turmeric powder using DNA barcoding. *Pharm. Biol.* **2015**, *53*, 1774–1779, doi:10.3109/13880209.2015.1005756.
93. Worley, B.; Powers, R. Multivariate analysis in metabolomics. *Curr. Metabolomics* **2013**, *1*, 92–107, doi:10.2174/2213235x11301010092.
94. Narkhede, S. Understanding AUC—ROC curve [ONLINE]. *Available* via *https//towardsdatascience.com/understanding-auc-roccurve-68b2303cc9c5 (accessed 27 Sept. 2020)* **2018**.

Commented [M539]: Citation for the Reference 32 is missing. Please revise.

Commented [a40R39]: We have revised and cited.