Quality assessment of *Clinacanthus nutans* leaf extracts by GC–MS-based metabolomics

Kogilavanee Devasvaran[1], Nabilah Hanim Baharom[1], Hui Wen Chong[1], Rozi Nuraika Ramli[2], Hock Ing Chiu[1], Chee Keong Lee[2] and Vuanghao Lim[1,*]

[1]Integrative Medicine Cluster, Advanced Medical and Dental Institute (AMDI), Universiti Sains Malaysia, Bertam, 13200 Kepala Batas, Pulau Pinang, Malaysia
[2]Bioprocess Technology Division, School of Industrial Technology, Universiti Sains Malaysia, 11800 Pulau Pinang, Malaysia

Clinacanthus nutans **is used in tropical Asia and Southeast Asia as a medicinal plant. Traditionally *C. nutans* had been used to treat diabetes, insect bites, inflammatory diseases, skin problems and cancer. To ensure accurate and consistent preparations of phytomedicines worldwide, quality assurance is crucial. Gas chromatography coupled with mass spectrometry (GC–MS) is a time-saving, specific and reliable technique extensively used for screening, diagnostics and functional genomics. After the extraction and derivatization of *C. nutans*, a vast number of metabolites from various chemical groups can be characterized in one analytical run. We used the GC–MS-based metabolomics approach to assess the quality of *C. nutans* leaf extracts from eight dissimilar locations in Malaysia. To differentiate the groups in each sample, four biomarkers were used, namely betulin, lupeol, stigmasterol and squalene. To differentiate and acquire a concise distribution of samples from different locations, multivariate analysis was used, viz. PCA and OPLS-DA. The biplot results suggested that betulin, stigmasterol and squalene were present in samples from Manjung, Alor Setar and Kangar, while lupeol was present in samples from Kepala Batas and Sungai Ara. Besides, squalene was also identified in a sample from the Herbal Garden, AMDI, USM. The proposed markers provide a useful identification tool for the quality assessment of raw materials from *C. nutans*.**

Keywords: *Clinacanthus nutans*, gas chromatography mass spectrometry, leaf extracts, quality assessment, metabolomics.

PLANTS with medicinal properties are well known as they possess ineffable healing properties[1] and therapeutic values[2]. The understanding and proper utilization of medicinal plants enable the efficacy, safety and quality of the plants to be maintained[3], which are vital to ameliorate or sustain human health[4]. Hence, quality control in plants is crucial to provide a constant chemical profile, up-to-the-mark biological activities and assurance in the quality of herbal medicine. This method is crucial for prescribing safe and effective herbal medicines with a constant parameter as well as standard qualitative and quantitative values[5].

Metabolomics, known as the comprehensive categorization of small metabolites present in the biological system, is an ideal method of obtaining medicinal plants with standardized quality[6]. In this method, metabolite composition analysis is used for the characterization of plant phenotype[7]. Additionally, the metabolic phenotypes are able to accurately show the physiology of an organism as well as environmental changes[8]. This method has been widely used for the characterization of different ecotypes and species of medicinal plants. At present, the metabolomic approach is utilized for metabolite content analysis of individual groups with various natural variances, thereby potentially enhancing crop quality[9]. In metabolomics, an ideal, fast and accurate quality assessment tool is gas chromatography–mass spectrometry (GC–MS), which is user-friendly, cost-efficient, and has high sensitivity and good separation[10]. This technique also has the ability to characterize hundreds of metabolites from various chemical groups in one analytical run, which makes it an efficient tool for screening, diagnostics and functional genomics[11]. GC–MS generates reproducible molecular fragmentation patterns suitable for metabolite identification, making it the most effective and sensitive tool for analysis[12].

Clinacanthus nutans, also known as Belalai Gajah or Sabah snake grass, is a popular medicinal plant in Malaysia[13,14]. This plant is consumed as a tea or salad for its remarkable healing properties, namely in treating hypercholesterolemia, diabetes mellitus, skin rashes, and insect or snake bites[15,16]. *C. nutans* extract reportedly possesses anti-inflammatory, antiviral and analgesic effects[17]. The extraordinary activities reported are mainly due to the phytochemical contents of the plant, including cerebrosides, flavonoids, glycerides, glycoglycerolipids, steroids and triterpenoids[18]. Besides, the high phenolic content in *C. nutans* extract contributes to the management of oxidative stress-related diseases[16].

Major concerns are related to the effect of plant phytomedicine quality; varied geographical locations may contribute to the difference in phytochemical

*For correspondence. (e-mail: vlim@usm.my)

compounds[19], thereby resulting in the variation in metabolite composition[20] and affecting the standardization of herbal drugs[21]. Thus, quality assessment using metabolomics in an unbiased and comprehensive way is crucial to analyse all the metabolites present in *C. nutans*, including those of low molecular number commonly found in biological samples[22]. GC–MS is the most reliable technique with high-throughput for the quantification of metabolites to standardize *C. nutans*, since it combines electron ionization (EI) and has repeatable retention time and high separation power, including versatile, sensitive and selective mass detection[23].

In the present study, soft independent modelling of class analogy (SIMCA) was employed to analyse the normalized GC–MS data to obtain the correlations, clusters and patterns in them[24]. The Pareto scale parameter of SIMCA-P software version 13 (Umetrics AB, Umeå, Sweden) was used for multivariate analysis of processed data[25]. The initial step of data analysis included the principal component analysis (PCA) model[26], and separation of sample in the population was represented by the score plot[27].

Additionally, in statistical analysis, a significant sample discrimination within populations in the respective groups is crucial, thus preferably employing orthogonal projections to latent structures discriminant analysis (OPLS-DA) as an ideal method of analysis[28]. The efficacy of the model was measured via R^2 and Q^2 values, which signified the goodness-of-fit and prediction accuracy respectively. Moreover, PCA and OPLS-DA models were used for the analysis of betulin, lupeol, stigmasterol and squalene, which were the compounds with significant discrimination of samples.

Materials and methods

Consumables and solvents

Solvents like 95% ethanol of analytical grade was purchased from Qrec, New Zealand, while bis(trimethylsilyl)trifluoracetamine + trimethylchlorosilane [BSTFA + TMCS (99 : 1)] and pyridine were purchased from Sigma-Aldrich, Steinheim, Germany.

Plant identification

The authentication of *C. nutans* was done at the Herbarium Unit, School of Biological Sciences, Universiti Sains Malaysia (USM), Penang, Malaysia (voucher no. SK 1980/11).

Sample preparation

C. nutans leaves were collected from eight locations around states in the north of Peninsular Malaysia (Table 1). The freshly harvested leaves were completely dried in an oven at 40°C, before being pulverized into a powder for the extraction step.

C. nutans extraction

C. nutans powder (25 g) was added to a conical flask containing 95% ethanol (250 ml) and sonicated at room temperature for 30 min. The mixture was transferred to a 50 ml centrifuge tube and centrifuged at 6000 rpm for 10 min. The supernatant was collected in a Schott bottle, while the pellet was re-extracted following the above-mentioned procedure for another two cycles. Upon completion of the three cycles, the supernatant was filtered using filter paper in a vacuum filtration system. The filtrates were evaporated in a rotary evaporator at a temperature of 60°C and pressure of 175 mbar. The dried extracts were then stored in a vacuum desiccator prior to use. The percentage yield of the extract was calculated as follows

$$\text{Percentage yield of extract} = \frac{\text{Weight of extract}}{\text{Weight of leaves}} \times 100.$$

Derivatization of C. nutans extracts

C. nutans extract (5 mg) was added to pyridine (50 µl) and BSTFA 75 µl, and the mixture was heated at 80°C for 20 min in a sonicator. The derivatized samples were filtered to remove any particulate residues using a 0.45 µm nylon membrane filter. The samples were then transferred to a micro-volume vial insert for GC–MS analysis[29].

GC–MS analysis of C. nutans

The GC–MS analysis was performed using the GC-System (Agilent Technologies 7890 A, New York, USA). Briefly, an aliquot (2 µl) of the derivatized sample was injected into the GC column at a split ratio of 1 : 20 (ref. 29). Helium was used as the carrier gas at a flow rate of

Table 1. Collection of *Clinacanthus nutans* leaf samples from eight different locations

Sample no.	Group	State	Location
1	G1	Perlis	Kangar
2	G2	Perlis	Beseri
3	G3	Penang	Kepala Batas
4	G4	Kedah	AlorSetar
5	G5	Penang	Sungai Ara
6	G6	Penang	Herbal Garden, IPPT
7	G7	Penang	TasekGelugor
8	G8	Perak	Manjung

1 ml/min. The injector was operated at 200°C in a 100–300°C oven, with a programmed temperature at a rate of 1.2°C/min (ref. 29). The compound identification was done following the National Institute of Standards and Technology/Gaithersburg MD USA (NIST MS) search 2.0 (ref. 30). The compounds were individually identified in a quantitative manner by estimating the percentage of their peak areas.

Data normalization

Normalization of the dataset was done using Microsoft Excel Professional Plus 2010. The information used included molecular mass, retention time and peak area detected by GC–MS.

Statistical analysis

After normalization, SIMCA version 13.0.3 (Umetrics AB, Umea, Sweden) was used for multivariate data analysis (MVDA) employing the PCA and OPLS-DA models[31,32]. PCA, the initial stage for multivariate analysis, was employed to acquire an overview as well as a summary of data to evaluate the key differences between samples. This unsupervised model classifies the data, identifies patterns and trends and finds outliers. OPLS-DA is an extension of partial least squares-discriminant analysis (PLS-DA), that is employed for model interpretation, classification studies along with biomarker identification. This supervised model is led by known information of classes that discovers responsible variables for class discrimination[33]. Generally, this model divides the variations into two groups, namely the correlated and uncorrelated variations to response. Therefore, irrelevant variations are filtered out.

Results and discussion

Plant extraction

Table 2 shows the percentage yield of *C. nutans* extracts from eight different locations. The sample from Kangar (G1) had the highest percentage yield (21.4%), while the lowest was recorded in G4 from Alor Star (14.2). The ultrasonication technique was used to extract the *C. nutans* leaves. Based on these findings, the percentage yield for *C. nutans* extracts using ultrasonication method was high compared to the maceration technique, which gave a value of less than 3% (ref. 34). Ultrasonication is a new extraction method that is efficient, straightforward, rapid and cost-effective. It requires minimum amount of organic solvents compared to conventional methods.

Table 2. Percentage yield of extract

Sample (location)	Percentage
G1 (Kangar Perlis)	21.4
G2 (Beseri Perlis)	13.9
G3 (Kepala Batas)	20.8
G4 (Alor Setar)	14.2
G5 (Sungai Ara)	16.0
G6 (Herbal Garden, IPPT)	20.0
G7 (Tasek Gelugor)	15.0
G8 (Manjung)	14.3

GC–MS profiling

The sample derivatization process prior to GC–MS analysis provides valuable information on the structure and composition of the sample. Derivatization is a process that uses an active agent to change the chemical characteristics of an analyte. In GC, the derivatization process can increase volatility and thermal stability, especially in the analysis of fatty acids and sugars. The reactivity of highly polar compounds can be decreased, and sensitivity towards the compound of interest can be increased[35]. In this study, BSTFA + TMCS were used as the silylation agent. It is an ideal reagent for trimethylsilylation of alcohols, amines, alkaloids, biogenic amines, carboxylic acids, phenols and steroids. In GC analysis, silylation is the most common derivatization practice. Active hydrogen is replaced by an alkyl silyl group, normally trimethylsilyl (TMS). Silyl derivatives are more thermally stable, more volatile and less polar compared to their parent compounds. Pyridine was used to facilitate the derivatization reaction. It acts as a catalyst to increase the reactivity of the reagent. Figure 1 shows the GC–MS chromatogram of sample no. 8. Squalene (molecular mass 410.73) was detected at retention time t_R = 107 min, whereas stigmasterol (molecular mass 412.69) was detected at t_R = 135 min. Betulin with molecular mass 442.72 was detected at t_R = 138 min. These findings have been corroborated by a previous study, in which squalene and stigmasterol were present in the methanol extract of *C. nutans*[36].

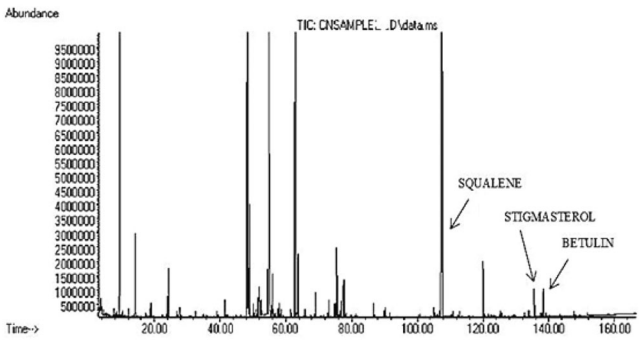

Figure 1. Total ion chromatogram of *Clinacanthus nutans* from sample no. 8.

PCA

PCA is a model that is unsupervised and exists in an orthogonal transformation, thereby making it possible to correlate variables to obtain a smaller number of uncorrelated variables known as principal components (PCs)[37]. Using PCA, the dimensionality of the dataset is reduced, and identification of new important underlying variables is more clear. The main concept of the PCA model is to identify the data pattern and detect correlation among the variables[38]. Normalization of the variables is necessary for PCA, because the original database may have different scales. The unnormalized variables using PCA might lead to the large loading of a dataset with high variance. In multivariate analysis, normalization is important to reach a linear and more robust relationship. Normalization is done to minimize the amount of duplicate data in a dataset, organize the data and provide easy access to them without compromising data integrity. Hence, normalization can be summed up as a systematic way of ensuring that the dataset is suitable for all purposes and free of any undesirable characteristics that can influence data integrity[39].

In the SIMCA software, pre-processing of data involves the scaling and autofitting phase. For scaling, the pareto was chosen as the default to set variation, because data entry was obtained from GC–MS. The Autofitting phase involves removal of components with eigenvalues less than 2. The eigenvalue provides information on how much variance exists in the data. The eigenvalue will have the largest variation in the first PC and gradually reduce in the subsequent PCs. An eigenvalue greater than 1 suggests that PCs account for additional variance than accounted for by one of the original variables in the standardized data. This is typically utilized as a cut-off point where PCs are retained.

PCA images display an outline of the data, in which each sample is denoted by a point in the multidimensional space, projected in a plane. In PCA three main aspects must be considered, namely the model, observation and variable diagnostics. In the model diagnostics, R^2 represents the goodness-of-fit and Q^2 represents the accuracy of the prediction[40]. In the observation diagnostics, strong and moderate outliers were identified using DModX column plot and Hotelling's T2 range column plot respectively.

The PCA model generated six PCs subsequent to the pre-processing step. Observations representing *C. nutans* extract were 24, and variables with reference to the spectra were 267 in number ($N = 24$, $K = 267$). For analysing the goodness-of-fit and accuracy of prediction, the R^2 and Q^2 values were considered. The values were between 0 and 1, where higher values were desired. R^2X value near 1 and Q^2X value of more than 0.5 indicated that the model was good to fit and predict the data. The preferable difference between the R^2X and Q^2X was 0.2–0.3 as the R^2X value must be larger than Q^2X. Based on the results, the R^2X value was 0.936 and the Q^2X value was 0.671, thereby displaying goodness-of-fit and accurate predictability of the model.

In score plot analysis, the relationship between eight different groups of *C. nutans* from different locations was correlated. The groups, trends and patterns of observations were summarized as scores obtained by the reduction of dimensionality and separation of noise. Table 3 shows the percentage variation between samples according to the applied PCs. The best principal component was PC1 (t[1]) versus PC2 (t[2]), with the percentage variation of 40.6 and which provided better cluster separation. Hence, both of these components were employed for the remaining models studied.

Figure 2 shows the score scatter plot using different components. The score plot chosen by components PC1 versus PC2 exhibited four individual groups. Based on the results, samples of groups 1, 4 and 8 were clustered together, whereas samples of groups 3 and 5 formed another cluster. Samples of groups 2, 6 and 7 were grouped together based on the score plot attained. Hence, using this scatter plot, the trend and distribution of group samples were discriminated within the dataset.

Figure 3 is the loading scatter plot which shows the relationship among the variables of a model. Betulin, lupeol, stigmasterol and squalene acted as biomarkers in *C. nutans* in the form of secondary metabolites.

The Hotelling's T2 range and DModX were employed to examine the quality of the PCA model. The former displayed the distance from the model plane origin for each selected observation, whereas the latter corresponded to the residual standard deviation of every observation present in the X or Y block. The Hotelling's T2 range column plot showed that every single group sample value was below than 99% confidence interval, thereby signifying that all the scores were in the normal site of the score plot. The DModX column plot detected moderate outliers within the sample. Based on the results, G2 and G7 were found to be the moderate outliers, because the significance level was higher than 0.05. This model was acceptable, because the Hotelling's T2 range column plot did not detect any strong outlier in all samples (Figure 4). Therefore, based on PCA analysis, it was concluded that the PCA model is acceptable, as the observations and variables were well explained.

Table 3. Principal component analysis (PCA) model – percentage of variation based on different principal components

Principle component	Percentage variation
t [1] versus t [2]	40.6
t [1] versus t [3]	38.9
t [2] versus t [3]	33.5

A biplot was used to show the relationship between loading and score, and exhibited using correlation scaling. Observations located closer to the variables were high in these variables and low in variables located opposite to them. Figure 5 shows the distribution of the metabolites within the samples. From the biplot, it can be

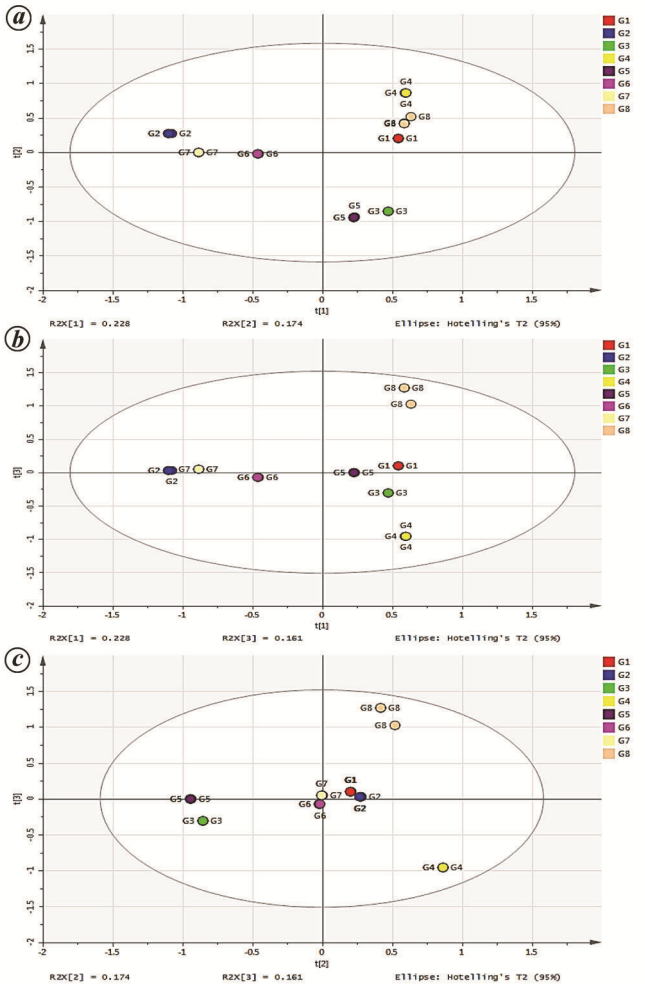

Figure 2. Principal component analysis (PCA) score scatter plot comparison of different principal components (PCs). ***a***, PC1 versus PC2; ***b***, PC1 versus PC3; ***c***, PC2 versus PC3.

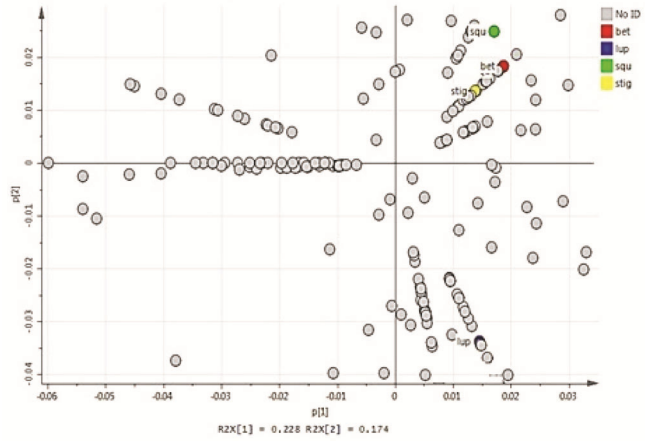

Figure 3. PCA loading scatter plot (bet, betulin; squ, squalene; stig, stigmasterol; lup, lupeol).

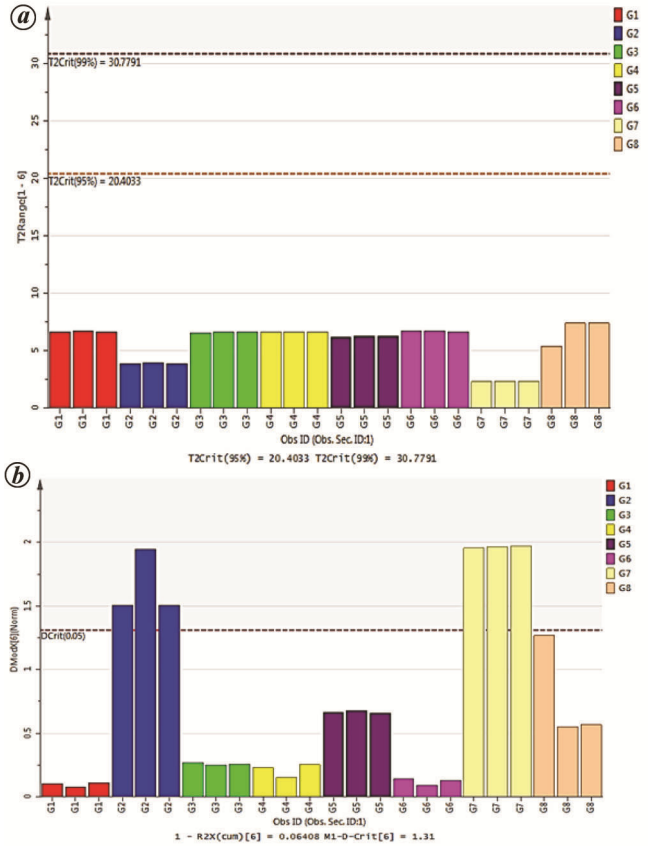

Figure 4. ***a***, The Hotelling's T2 range column plot. ***b***, DmodX column plot of the PCA model.

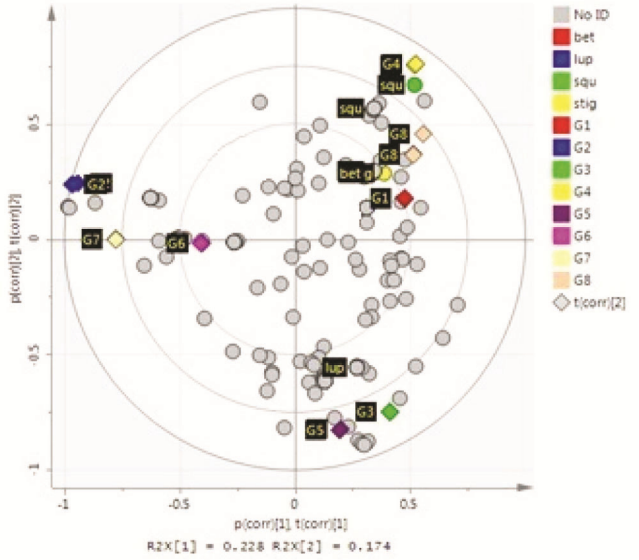

Figure 5. Score versus loading in a biplot.

seen that squalene, stigmasterol and betulin are present in G1, G4 and G8, whereas lupeol is present in G3 and G5. Squalene is also detected in G6, but none of the metabolites is observed in G2 and G7. This activity of these secondary metabolites might be hidden, affected by primary metabolites or disregarded because of their small amounts in the crude extract[41].

OPLS-DA

OPLS is a prediction and regression technique which detects information in the *X* data, that is associated to *Y* data, the known information. The OPLS-DA model is used when working with discrete variables, and it is useful in biomarker identification and classification studies. This model is similar to PCA, as it depends on the projection of the *X* data. The PCA model focuses on the trends and outlier identification, whereas the OPLS-DA model uses multivariate data to discriminate between two or more groups. In OPLS-DA, a regression model is constructed between the multivariate data and a response variable that only contains class information. The OPLS-DA is a variant of PLS that employs orthogonal signal correction to maximize the explicated covariance between *X* and *Y* data on the first latent variable. Thus, it improves data interpretation and reduces the possibility of misleading interpretations.

In this study OPLS-DA was used to identify biomarkers in *C. nutans* samples. OPLS-DA analysed 24 observations and generated a model with six predictive components. Based on the results, the R^2X value was 0.936, R^2Y was 0.856 and Q^2 was 0.715. The R^2X value was near 1 and Q^2 value was more than 0.5, thereby displaying goodness-of-fit and accurate predictability of the model.

The Hotelling's T2 range column plot in Figure 6 *a* shows that G8 could possibly be a strong outlier, because the value is higher compared to the other samples. However, it is an unacceptable range as the value does not exceed 99% of the confidence interval. Based on the results obtained, the value of T2 Crit for 99% was 20.3812, while it was 13.184 for 90%. In Figure 6 *b*, the DModX column plot reveals that G6 shows a moderate outlier within the samples. This is acceptable for the OPLS-DA model, because the Hotelling's T2 column plot did not detect strong outliers.

Figure 7 shows the scatter plot for the OPLS-DA model. Using this model, interpretation could be made more clearly as the sample groups were classified more accurately. Hence, from the results, G5 and G3 were in the same cluster while G1, G4 and G8 formed another cluster. Additionally, samples of G2, G6 and G7 were clustered together. These clusters were formed due to the distribution of metabolite compounds in the samples, suggesting that samples with the same metabolite compounds formed a cluster.

Variable influence on projection (VIP) plot summarizes the significance of the variables to clarify the *X* data, correlate them to the *Y* data and introduce variables with

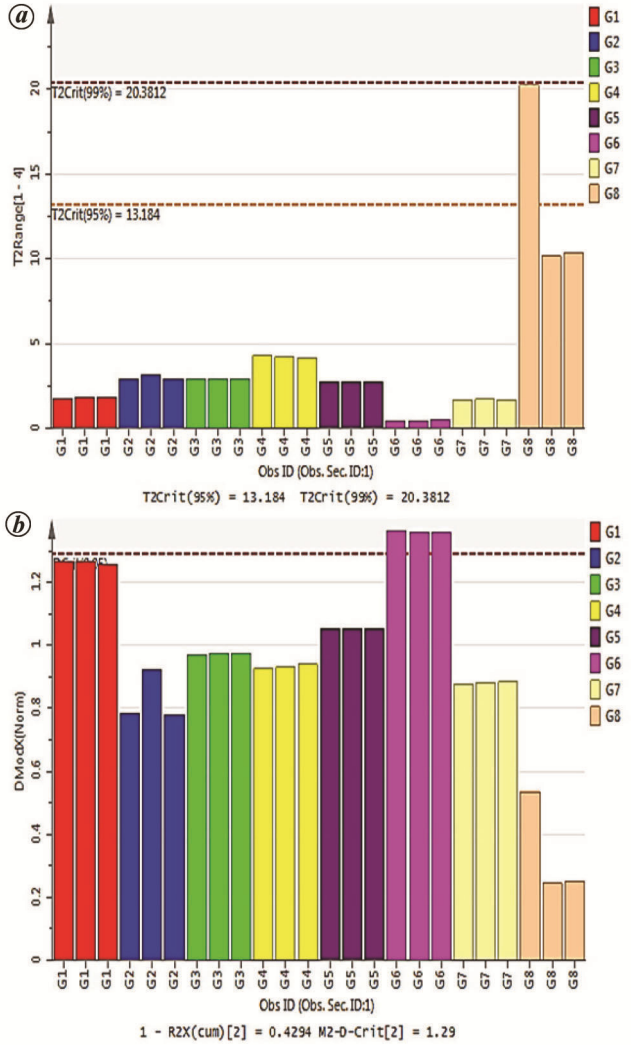

Figure 6. ***a***, The Hotelling's T2 range column plot. ***b***, DModX column plot for the orthogonal projections to latent structures discriminant analysis (OPLS-DA) model.

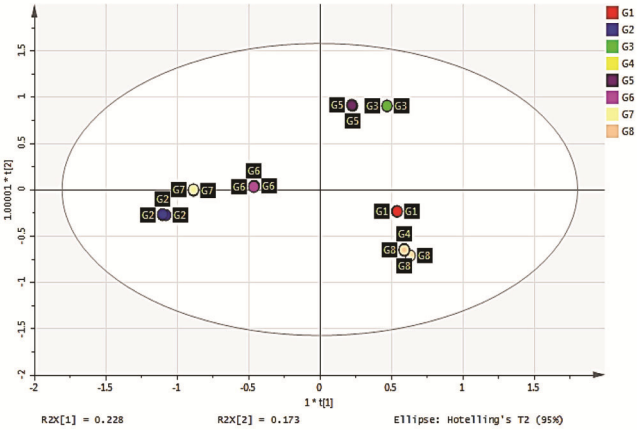

Figure 7. Scatter plot for the OPLS-DA model.

Table 4. Variable influence on projection (VIP) values for the nine most significant variables

Variance ID (primary)	M2.VIP	M2.VIP cv SE
540.261/18.219	2.94438	2.86469
540.261/21.559	2.84088	1.91624
540.261/21.526	2.29774	1.77143
418.247/36.81	1.70096	1.20693
918.432/35.494	1.60876	1.43341
308.166/6.377	1.50908	1.22810
356.311/25.473	1.17124	1.17390
612.301/23.398	1.07009	0.876565
328.28/21.981	1.02196	0.907825

the highest influence in the model. This plot was sorted from the highest to lowest values and displayed with 95% confidence interval. VIP values above 1.0 were chosen as the metabolite cut-off to deduce the crucial peak signals of absorbance present in the loading scatter plots[42]. In Table 4, nine variance IDs are listed that represent the first nine most significant variables in the dataset with VIP values greater than 1.0.

For ANOVA of the cross-validated residuals (CV-ANOVA), the *P*-value indicated statistical significance of the studied model. From the results, the *P*-value obtained was 4.081511e-006, which is less than 0.05. Hence, the OPLS-DA model proves significant for the analysis of this dataset[43].

Conclusion

Metabolomics using GC–MS was employed to standardize the *C. nutans* leaf extracts as the CV-ANOVA stated that the *P*-value was less than 0.05. This shows that the OPLS-DA model could differentiate the extracts significantly well. Multivariate data analysis with a combination of PCA and OPLS-DA models proved to be an effective tool for metabolomics studies, especially for the discrimination and quality assessment of *C. nutans*. The *C. nutans s*amples from Kangar, Alor Setar and Manjung contained squalene, stigmasterol and betulin, whereas those from Sungai Ara and Kepala Batas contained lupeol. However, samples from Beseri Perlis and Tasek Gelugor did not contain the secondary metabolite markers used in this study. Therefore, betulin, lupeol, stigmasterol and squalene could be used as markers for the GC–MS analysis of *C. nutans* leaf extracts from different locations to ensure that quality control is maintained.

1. Ong, M. G., Mat Yusuf, S. N. A. and Lim, V., Pharmacognostic and antioxidant properties of dracaena sanderiana leaves. *Antioxidants*, 2016, **5**, 28.
2. Kamarudin, M. N. A., Sarker, M. M. R., Kadir, A. H. and Ming, L. C., Ethnopharmacological uses, phytochemistry, biological activities, and therapeutic applications of *Clinacanthus nutans* (Burm. f.) Lindau: a comprehensive review. *J. Ethnopharmacol.*, 2017, **206**, 245–266.
3. Kumari, R., A review on the standardization of herbal medicines. *Int. J. Pharm. Sci. Res.*, 2016, **7**, 97–106.
4. Lim, V., Schneider, E., Wu, H. and Pang, I. H., Cataract preventive role of isolated phytoconstituents: findings from a decade of research. *Nutrients*, 2018, **10**(11), 1580.
5. Folashade, O., Omoregie, H. and Ochogu, P., Standardization of herbal medicines – a review. *Int. J. Biodivers. Conserv.*, 2012, **4**, 101–112.
6. Commisso, M., Strazzer, P., Toffali, K., Stocchero, M. and Guzzo, F., Untargeted metabolomics: an emerging approach to determine the composition of herbal products. *Comput. Struct. Biotechnol. J.*, 2013, **4**, 1–7.
7. Dai, H., Xiao, C., Liu, H. and Tang, H., Combined NMR and LC–MS analysis reveals the metabonomic changes in *Salvia miltiorrhiza* Bunge induced by water depletion. *J. Proteome Res.*, 2010, **9**, 1460–1475.
8. Fiehn, O., Metabolomics – the link between genotypes and phenotypes. *Plant Mol. Biol.*, 2002, **48**, 155–171.
9. Schauer, N. and Fernie, A. R., Plant metabolomics: towards biological function and mechanism. *Trends Plant Sci.*, 2006, **11**, 508–516.
10. Kanani, H., Chrysanthopoulos, P. K. and Klapa, M. I., Standardizing GC–MS metabolomics. *J. Chromatogr. B*, 2008, **871**, 191–201.
11. Rohloff, J., Analysis of phenolic and cyclic compounds in plants using derivatization techniques in combination with GC–MS-based metabolite profiling. *Molecules*, 2015, **20**, 3431–3462.
12. Qiu, Y. and Reed, D., Gas chromatography in metabolomics study. In *Advances in Gas Chromatography* (ed. Guo, X.), InTech, London, UK, 2014, pp. 83–101.
13. Yakop, F. *et al*., Silver nanoparticles *Clinacanthus nutans* leaves extract induced apoptosis towards oral squamous cell carcinoma cell lines. *Artif. Cells, Nanomed. Biotechnol.*, 2018, **46**, 131–139.
14. Phua, Q. Y., Subramaniam, S., Lim, V. and Chew, B. L., The establishment of cell suspension culture of Sabah snake grass (*Clinacanthus nutans* (Burm. F.) Lindau). *In Vitro Cell. Dev. Biol. – Plant*, 2018, **54**, 413–422.
15. Kunsorn, P., Ruangrungsi, N., Lipipun, V., Khanboon, A. and Rungsihirunrat, K., The identities and anti-herpes simplex virus activity of *Clinacanthus nutans* and *Clinacanthus siamensis*. *Asian Pac. J. Trop. Biomed.*, 2013, **3**, 284–290.
16. Sarega, N. *et al*., Phenolic rich extract from *Clinacanthus nutans* attenuates hyperlipidemia-associated oxidative stress in rats. *Oxid. Med. Cell. Longev.*, 2016, **2016**, 4137908.
17. Mai, C. W. *et al*., Mechanisms underlying the anti-inflammatory effects of *Clinacanthus nutans* Lindau extracts: inhibition of cytokine production and toll-like receptor-4 activation. *Front. Pharmacol.*, 2016, **7**, 1–11.
18. Sakdarat, S., Shuyprom, A., Pientong, C., Ekalaksananan, T. and Thongchai, S., Bioactive constituents from the leaves of *Clinacanthus nutans* Lindau. *Bioorg. Med. Chem.*, 2009, **17**, 1857–1860.
19. Wang, Y. *et al*., Metabolomic strategy for the classification and quality control of phytomedicine: a case study of chamomile flower (*Matricaria recutita* L.). *Planta Med.*, 2004, **70**, 250–255.
20. Hanan, N. A. *et al*., Cytotoxicity of plant-mediated synthesis of metallic nanoparticles: a systematic review. *Int. J. Mol. Sci.*, 2018, **19**, 1725.
21. Chawla, R. *et al*., Evidence based herbal drug standardization approach in coping with challenges of holistic management of diabetes: a dreadful lifestyle disorder of 21st century. *J. Diabetes Metab. Disord.*, 2013, **12**, 35.
22. Fiehn, O., Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp. Funct. Genomics*, 2001, **2**, 155–168.

23. Koek, M. M., Muilwijk, B., van der Werf, M. J. and Hankemeier, T., Microbial metabolomics with gas chromatography–mass spectrometry. *Anal. Chem.*, 2006, **78**, 1272–1281.
24. Kamboj, A., Analytical evaluation of herbal drugs. In *Drug Discovery Research in Pharmacognosy* (ed. Vallisuta, O.), InTech, London, UK, 2012, pp. 23–55.
25. Madala, N. E., Piater, L. A., Steenkamp, P. A. and Dubery, I. A., Multivariate statistical models of metabolomic data reveals different metabolite distribution patterns in isonitrosoacetophenone-elicited *Nicotiana tabacum* and *Sorghum bicolor* cells. *Springerplus*, 2014, **3**, 254.
26. Zul, M., Bin, F., Rahman, A., Chong, H. W. and Lim, V., UV–visible chemometrics approach for the determination of selected adulterants in claimed premixed coffee. *Malays. J. Med. Health Sci.*, 2018, **14**, 147–152.
27. Kowalczuk, A. P., Łozak, A., Kiljan, M., Mętrak, K. and Zjawiony, J. K., Application of chemometrics for identification of psychoactive plants. *Acta Pol. Pharm. – Drug Res.*, 2015, **72**, 517–525.
28. Javadi, N. *et al*., GC–MS-based metabolite profiling of *Cosmos caudatus* leaves possessing alpha-glucosidase inhibitory activity. *J. Food Sci.*, 2014, **79**, 1130–1136.
29. Graikou, K. *et al*., Chemical analysis of Greek pollen – antioxidant, antimicrobial and proteasome activation properties. *Chem. Cent. J.*, 2011, **5**, 33.
30. Revathi, D. and Rajeswari, M. Chemical profiling of *Guettarda speciosa* Linn. by GC–MS. *Int. J. Emerg. Technol. Adv. Eng.*, 2015, **5**, 5–9.
31. Zhao, Y. *et al*., A metabolomics study delineating geographical location-associated primary metabolic changes in the leaves of growing tobacco plants by GC–MS and CE–MS. *Sci. Rep.*, 2015, **5**, 16346.
32. Worley, B. and Powers, R., Multivariate analysis in metabolomics. *Curr. Metabolomics*, 2013, **1**, 92–107.
33. Chong, H. W., Rezaei, K., Chew, B. L. and Lim, V., Chemometric profiling of *Clinacanthus nutans* leaves possessing antioxidant activities using ultraviolet–visible spectrophotometry. *Chiang Mai J. Sci.*, 2018, **45**, 1–12.
34. Sulaiman, I. S. C., Basri, M., Chan, K. W., Ashari, S. E., Masoumi, H. R. F. and Ismail, M., *In vitro* antioxidant, cytotoxic and phytochemical studies of *Clinacanthus nutans* Lindau leaf extract. *Afr. J. Pharm. Pharmacol.*, 2015, **9**, 861–874.
35. Michel, F., Derivatization of polar compounds for GC. sigma-aldrich.com, 2010, pp. 1–38.
36. Cheong, B. E., Ho, S. Y., Wong, D. V. F., Yong, W. T. L., Teoh, P. L. and How, S. E., Chemical profiling of Sabah snake grass, *Clinacanthus nutans*. In Proceedings of the 11th Seminar on Science & Technology, Sabah, Malaysia, 2013, pp. 1–7.
37. Vargas, J., Quiroga, J. A. and Belenguer, T., Analysis of the principal component algorithm in phase-shifting interferometry. *Opt. Lett.*, 2011, **36**, 2215–7.
38. Yeung, K. Y. and Ruzzo, W. L., Principal component analysis for clustering gene expression data. *Bioinformatics*, 2001, **17**, 763–774.
39. Reisetter, A. C. *et al*., Mixture model normalization for non-targeted gas chromatography/mass spectrometry metabolomics data. *BMC Bioinformatics*, 2017, **18**, 84.
40. Khoo, L. W. *et al*., Phytochemical diversity of *Clinacanthus nutans* extracts and their bioactivity correlations elucidated by NMR based metabolomics. *Phytochem. Lett.*, 2015, **14**, 123–133.
41. Sampaio, B. L., Edrada-Ebel, R. and Da Costa, F. B., Effect of the environment on the secondary metabolic profile of *Tithonia diversifolia*: a model for environmental metabolomics of plants. *Sci. Rep.*, 2016, **6**, 29265.
42. Yusof, N. A. T. *et al*., Infrared-metabolomics approach in detecting changes in *Andrographis paniculata* metabolites due to different harvesting ages and times. *J. Sci. Food Agric.*, 2015, **95**, 2533–2543.
43. Dahiru, T., *P*-value, a true test of statistical significance? A cautionary note. *Ann. Ibadan Postgrad. Med.*, 2008, **6**, 21.

ACKNOWLEDGEMENTS. We thank Universiti Sains Malaysia for funding support from the Research University Grant (RUI, 1001/CIPPT/811318). We also acknowledge financial support from USM Graduate Assistant Scholarship (K.D.) and USM Fellowship (H.I.C.).

Received 10 July 2019; revised accepted 26 February 2020

doi: 10.18520/cs/v119/i4/641-648