The application of pattern recognition techniques in metabolite fingerprinting of six different *Phyllanthus* spp.

Saravanan Dharmaraj ^{a,*}, Lay-Harn Gam ^b, Shaida Fariza Sulaiman ^c, Sharif Mahsufi Mansor ^a and Zhari Ismail ^b

^a Centre of Drug Research, Universiti Sains Malaysia, Pulau Pinang, Malaysia

^b School of Pharmaceutical Sciences, Universiti Sains Malaysia, Penang, Malaysia

^c School of Biological Sciences, Universiti Sains Malaysia, Penang, Malaysia

Abstract. FTIR spectroscopy was used together with multivariate analysis to distinguish six different species of *Phyllanthus*. Among these species *P. niruri*, *P. debilis* and *P. urinaria* are morphologically similar whereas *P. acidus*, *P. emblica* and *P. myrsinifolius* are different. The FTIR spectrometer was used to obtain the mid-infrared spectra of the dried powdered leaves in the region of 400–4000 cm⁻¹. The region of 400–2000 cm⁻¹ was analyzed with four different pattern recognition methods. Initially, principal component analysis (PCA) was used to reduce the spectra to six principal components and these variables were used for linear discriminant analysis (LDA). The second technique used LDA on most discriminating wavenumber variables as searched by genetic algorithm using canonical variate approach for either 30 or 60 generations. SIMCA, which consisted of constructing an enclosure for each species using separate principal component models, was the third technique. Finally, multi-layer neural network with batch mode of backpropagation learning was used to classify the samples. The best results were obtained with GA of 60 gens. When LDA was run with the six wavenumbers chosen (1151, 1578, 1134, 609, 876 and 1227), 100% of the calibration spectra and 96.3% of the validation spectra were correctly assigned.

Keywords: FTIR, genetic algorithm, neural networks, SIMCA, *Phyllanthus*

1. Introduction

Phyllanthus niruri Linn (synonym: *P. amarus*) is widely found in tropical regions of the world. Although various activity in animals such as lipid lowering [14,27], contraceptive [20], antiplasmodial [24] and antitumor effect [21] have been reported, its often used by human for beneficial effect on kidney stones in the region of south east Asia. In Malaysia, there are two other similar species of *P. debilis* and *P. urinaria* for which the same health benefit is attributed although usage of *P. niruri* is more popular [28]. It has been mentioned that current methods to distinguish these three species depend greatly on morphological and phytochemical methods, which are not adequate to separate them [25]. However, another group [6] was able to distinguish *P. emblica* from *P. niruri* and *P. urinaria* but their approach required DNA isolation and use of randomly amplified polymorphic DNA-polymerase chain reaction (RAPD-PCR). The RAPD-PCR was later reported to be capable in distinguishing *P. niruri* from *P. urinaria* and *P. debilis* [26].

*Corresponding author: Saravanan Dharmaraj, Centre of Drug Research, Universiti Sains Malaysia, 11800 Pulau Pinang, Malaysia. Tel.: +604 6533888 ext. 3259; Fax: +604 6568669; E-mail: saravanandharmaraj@yahoo.com.

The quality control methods in the herbal industry usually involves visual inspection at the macroscopical as well as microscopical level as an initial first step and later analytical inspection with thin layer chromatography (TLC) or high performance liquid chromatography (HPLC). The first step is subjective whereas the chemical analysis involves analyzing the herbs for presence of chemical markers. The detection of chemical markers need to take into consideration the possibility of spiking or adulteration and the choice of markers from a wide range of chemicals such as terpenoids, flavonoids, etc. Furthermore, it has been reported in food industry that sometimes identity could not be ascertained even after analyzing various markers [3,7]. Other than having supposedly taxonomic significance, the marker compounds should be responsible for activity but often the contribution of a particular compound is not known and activity is due to synergistic activity of various components [12].

Recent development is the use of metabolite fingerprinting of samples for determining origin as well as identification or taxonomic purposes. The approach of metabolite fingerprinting involves obtaining information to unravel metabolic alterations without the need to obtain quantitative data for all the metabolites. This approach is often performed via rapid analytical methods such as nuclear magnetic resonance [1,8,15], mass spectrometry [9,23] or Fourier transform infrared (FTIR) spectroscopy [29]. The approach of FTIR is often selected as this rapid technique measures the vibrational of bonds within functional groups for carbohydrates, amino acids, lipids, fatty acids as well as the secondary metabolites in plants simultaneously.

Metabolite fingerprinting with FTIR spectroscopy in combination with multivariate analysis of principal component analysis (PCA) or genetic algorithm (GA) in combination with linear discriminant analysis (LDA) was capable of differentiating *P. niruri* according to locations [5] and this approach was tried to distinguish the closely resembling *P. niruri*, *P. debilis* and *P. urinaria* as well as three other morphologically different species of *P. acidus*, *P. emblica* and *P. myrtifolius*. Holmes and coworkers [13] mentioned that simple unsupervised method such as PCA might work well with data classes with limited number of well defined classes whereas more complex and hard to distinguish spectra would require more sophisticated statistical approach. Therefore, the approach here was to combine metabolite fingerprinting using FTIR with multivariate analysis of PCA-LDA, GA-LDA, SIMCA and neural networks to distinguish the six different species.

2. Experimental and methods

2.1. Chemicals

Potassium bromide (KBr) for infrared spectroscopy (Sigma-Aldrich).

2.2. Samples of *Phyllanthus* species

Three different batches for each of the six species were collected from Pulau Pinang. Fresh samples of leaves were dried to dryness at 40°C and ground. Techniques of pooling and quartering as described in quality control of medicinal plant materials by World Health Organization [2] were used to obtain the three average samples from each batch. This process was repeated to obtain eighteen spectra for each species and from these, odd numbered spectra were used for calibration set and even numbered set in the validation set.

2.3. KBr method

Dried powdered herbal sample weighing 2 mg was mixed with 98.0 mg of potassium bromide (KBr) powder with a pestle and mortar. This mixture was used for the preparation of herbal KBr tablet at 10 tons of pressure for 2 min. Eighteen KBr discs were used for each species.

2.4. Spectral acquisition

All spectra in the region of 400–4000 cm^{-1} were collected using a Nexus model FTIR spectrometer (Thermo Nicolet Corp., WI, USA), which was equipped with a deuterated triglycine-sulphate (DTGS) detector. The instrument was controlled by OMNICTM code and the infrared measurements were performed at spectral resolution of 4 cm^{-1} with 32 interferograms co-added before Fourier transformation. Happ–Genzel apodization was applied and spectra were encoded every 1.928 cm^{-1} .

2.5. Data processing software

Chemometric analysis was carried out to visualize groupings within the various samples. Two preliminary steps were carried out that is all the spectra were smoothed and normalized with Spectrum Version 3.02 (PerkinElmer, Inc.). The data obtained were then further processed by using tools in Microsoft[®] Excel 2000 (Microsoft Corp., WA, USA), SPSS for Windows Version 11.5 (SPSS Inc., Chicago, IL, USA), The Unscrambler[®] Version 8.05 (Camo Process AS, Oslo, Norway) and Matlab Version 6.5.1 (The MathWorks Inc., Natick, MA, USA).

PCA and LDA of the normalized spectra in the region of 400–2000 cm^{-1} were analyzed by SPSS. Wavelength selection for the above IR region by GA was carried out using Matlab and the selected wavelengths used for LDA by SPSS. SIMCA was carried out using the Unscrambler. The neural network computations were performed with the neural network toolbox in Matlab.

2.6. Multivariate analysis

2.6.1. Principal component analysis–linear discriminant analysis (PCA–LDA)

The FTIR spectra used was the region from 400 to 2000 cm^{-1} consisting of 831 variables. This region was chosen to reduce the computation required and furthermore, this range was capable of differentiating spectra according to region. Both the calibration and validation set consisted of 54 spectra each and was combined into a single matrix. An initial PCA was carried out on the 108×831 data matrix. Later, discriminant analysis was carried out using the first six principal component scores as variables. The classification of samples in the calibration set and more importantly the validation set where groups are not initially assigned was monitored.

2.6.2. Classification by genetic algorithm–linear discriminant analysis (GA–LDA)

The GA was used on the calibration data set to find the most discriminating variables. The approach was similar to the one earlier which separated samples according to region [5]. The Xdata consisted of 54×831 spectral data, whereas the Ydata of dummy variables of species assignment consisted of a matrix of 54×6 . The GA parameters were an initial number of chromosomes of 80, number of generations of terminations of algorithms after 30 or 60 generations and number of canonical variate loadings calculated was four. The spectral variables chosen were the ones with highest magnitude for the first loadings and these variables were used for linear discrimination analysis to obtain classification according to species for both the calibration data and validation data set.

2.6.3. Classification by soft-independent modeling of class analogy (SIMCA)

SIMCA was also carried out by building a separate principal component model for each species by using cross-validation method on the calibration data set. The second step in SIMCA classification consisted in classifying both the calibration and validation spectra using the six principal component models built.

2.6.4. Classification by neural networks

There are various architecture for neural networks and this study used an $831 \times 6 \times 6$ plus bias, that is 831 neurons in the input layer corresponding to the number of wavenumbers selected. Six neurons in the hidden layer was empirically chosen whereas, six neurons in the output layer represented the six species or targets concerned. The whole spectral data of the six species was divided into calibration and validation sets. Both sets consisted of 831×54 matrix. As this network used six output neurons, a target vector T (size of 6×54) was presented in which, row of the matrix represented the species.

The training of the neural network using calibration data set was done with batch mode of backpropagation algorithm. Information on training of the network is explained in detail elsewhere [10,11]. The training set was employed to adjust the weights using the Levenberg–Marquardt algorithm for backpropagation of error. The learning used was gradient descent with momentum. Two different systems of this were tried where in both the hyperbolic tangent sigmoid (tansig) transfer function was used for hidden neurons but differed where either log-sigmoid (logsig) or hyperbolic tangent sigmoid (tansig) was the transfer function for the output neurons. The other training parameters used were common for both types. The parameters of epoch, show, goal, time, min_grad, max_fail, mu, mu_dec, mu_inc, mu_max, and mem_red are explained in detail elsewhere [4]. Except for epoch and show, which used values of 20 and 5, respectively, all other parameters used were of default settings. After the network has been trained, it was stimulated using the validation set and correct classification was noted.

3. Results and discussion

3.1. Principal component analysis–linear discriminant analysis (PCA–LDA)

The concept used here was to reduce the dimensionality of the data by performing PCA and then using the principal components as variables for LDA. The PCA was carried out on the combined calibration and validation set. The number of principal components chosen for the subsequent LDA was based on the number of components that gave a total explained variance of at least ninety percent. Six principal components were required to fulfill these criteria and their variance explained was 34.3, 25.2, 16.7, 8.6, 3.7 and 3.5%. These six principal components accounted for 92.1% of the total variability.

LDA showed good recognition ability according to species by using the six principal components as variables. This was shown by 98.1% of the calibration set being correctly classified. Only one out of nine spectra of *P. niruri* was wrongly classified as belonging to *P. urinaria*. The high percentage of correct classification was also confirmed in the validation set which had 96.3% correct classification. Only two out of nine spectra of *P. niruri* in the validation set were wrongly classified as *P. debilis*. The other five species had 100% correct classification in both calibration and validation set.

3.2. Genetic algorithm–linear discriminant analysis (GA–LDA)

The GA for differentiation of species used the canonical variate concept, which maximized the ratio of variance between groups to within groups to find the best chromosomes that could separate spectra

Fig. 1. GA classification of different *Phyllanthus* spp. The canonical variate (CV) scores for combined calibration and validation data of each species at 30 and 60 generation. PAC is *P. acidus*, PDE is *P. debilis*, PEM is *P. emblica*, PMY is *P. myrtifolius*, PNI is *P. niruri* and PUR is *P. urinaria*.

according to species. The GA was run 10 times for the 30 gens and 6 times for 60 gens using the calibration set data. The most common results for both of the runs were identified and using their loadings, the canonical variate scores were calculated for both calibration and validation set. This was done by multiplying each variable's loading with the autoscaled reading for the respective variable and the total sum of this for the 831 variables gave the canonical variates scores for the dimension concerned. The canonical variate scores were calculated for the first and second dimensions for both the 30 gens as well as 60 gens runs and were plotted. The plot in Fig. 1 shows groupings according to species and the plot for 60 gens shows slightly better separation of species. The calibration and validation data points fall in the same region of the plot for each of the respective species. Therefore, the plot combines the data for calibration and validation set and shows the locations on the plot according to the species.

The discriminating ability was confirmed by running LDA using either four or six most discriminating wavenumber variables from the 60 gens run. It is the magnitude of loadings that show importance of a particular variable for discrimination of species and the wavenumber variables were sorted from lowest value to highest. The wavenumber variables with the three lowest and three highest loadings were chosen. The three lowest to three highest were 1151, 1578, 1134, ..., 609, 876, 1227.

In the first LDA using four wavenumbers of 1151, 1578, 876 and 1227 as variables, 94.4% of the calibration data and 90.7% of the validation set were correctly assigned. Two spectra of *P. niruri* in both calibration and validation set were wrongly assigned as *P. debilis*. Furthermore, two spectra in calibration and three spectra in validation belonging to *P. debilis* were wrongly classified as *P. niruri*.

When LDA was run with the six wavenumbers (1151, 1578, 1134, 609, 876 and 1227), 100% of the calibration spectra and 96.3% of the validation spectra was correctly assigned. In the validation set, one spectrum out of nine of *P. debilis* was assigned as *P. niruri* whereas one spectrum of *P. niruri* was misidentified as *P. debilis*. Further optimization to improve the classification would involve selection of different fitness function as well as change in approach for crossover or mutation [16]. There are other approaches for GA to be implemented and some of these [17–19,22] might improve the classification of very closely related spectra.

3.3. Soft-independent modeling of class analogy (SIMCA)

SIMCA classification was also utilized for discrimination of the six different *Phyllanthus* species. Six models of principal components were developed for each of the category and the number of principal components used was as suggested. The number of principal components for *P. acidus*, *P. debilis*, *P. emblica*, *P. myrtifolius*, *P. niruri* and *P. urinaria* were 4, 7, 7, 5, 6 and 5, respectively. SIMCA classification gave 100% sensitivity with no spectra belonging to their class was rejected. However, sensitivity (where spectra were not classified into other groups) was low with values of 74.1% for calibration set and 72.2% for validation set.

In the calibration set, seven out of nine spectra of *P. debilis* were also assigned as *P. niruri*, whereas three out of nine spectra of *P. niruri* were also assigned as *P. urinaria*. In the case of *P. urinaria*, four out of nine was co-assigned as *P. niruri*, whereas two others also assigned to *P. debilis*.

In the validation set, the whole nine spectra of *P. debilis* was also classified into *P. niruri*, three out of nine spectra of *P. niruri* was also designated as *P. urinaria*, whereas another was co-assigned to *P. debilis*. Out of the nine *P. urinaria* spectra, three were co-assigned to *P. niruri* whereas two were also assigned to *P. debilis*.

3.4. Neural networks

The neural networks use the patterns in the training or calibrations set, and learn to enable the network to make predictions. Two types of neural networks, which differed only in the function for the output neurons, where in the first log-sigmoid (logsig) was used, whereas in the second hyperbolic tangent sigmoid (tansig) transfer functions was used. The overall prediction ability of the first neural network using Levenberg–Marquardt algorithm and gradient descent with momentum is shown in Table 1. The final assignment success rates for the calibration and validation set was 98.2% and 83.3%, respectively. Overall, the network performed well as seen from the calibration set. The low correct classification in *P. urinaria* in the validation set could be overcome by using a larger sample set. The average score for the six output neurons using the validation set, which represents the classification achieved by the neural network, is shown in Fig. 2. The graph displays the prediction of the network for each selected species as a bar chart. It can be noticed that the similarity between *P. niruri* and *P. urinaria* can be seen from the outputs for the fifth and sixth neurons.

The second neural network used had hyperbolic tangent sigmoid (tansig) transfer functions in both the hidden and output neurons. The network was also trained using Levenberg–Marquardt algorithm

Table 1

Classification with neural network using feed-forward backpropagation (FFBP) with transfer function of hyperbolic tangent sigmoid (tansig) for hidden neurons and log-sigmoid (logsig) for output neurons

Group	Calibration set			Validation set		
	Number classed correct	Number classed wrong	Classified correctly (%)	Number classed correct	Number classed wrong	Classified correctly (%)
<i>P. acidus</i>	9	0	100.0	8	1	88.9
<i>P. debilis</i>	9	0	100.0	7	2	77.8
<i>P. emblica</i>	9	0	100.0	9	0	100.0
<i>P. myrtifolius</i>	9	0	100.0	9	0	100.0
<i>P. niruri</i>	9	0	100.0	8	1	88.7
<i>P. urinaria</i>	8	1	88.9	4	5	44.4

Fig. 2. The average scores of the six output neurons for each species. The neural network used transfer function of hyperbolic tangent sigmoid (tansig) for hidden neurons and log sigmoid (logsig) for output neurons. In ideal classification output neuron associated with species would give value of 1; other neurons would give value of 0.

Table 2

Classification with neural network using feed-forward backpropagation (FFBP) with transfer function of hyperbolic tangent sigmoid (tansig) for both hidden and output neurons

Group	Calibration set			Validation set		
	Number classed correct	Number classed wrong	Classified correctly (%)	Number classed correct	Number classed wrong	Classified correctly (%)
<i>P. acidus</i>	9	0	100.0	9	0	100.0
<i>P. debilis</i>	6	3	66.7	7	2	77.8
<i>P. emblica</i>	9	0	100.0	9	0	100.0
<i>P. myrtifolius</i>	9	0	100.0	9	0	100.0
<i>P. niruri</i>	6	3	66.7	6	3	66.7
<i>P. urinaria</i>	9	0	100.0	7	2	77.8

and gradient descent with momentum. The prediction ability of this network was almost similar to the earlier one but was better in handling the *P. urinaria* samples. The prediction ability for the calibration and validation set is shown in Table 2. The success rates for prediction of the calibration and validation set was 88.9% and 87.1%. The average score for the six output neurons with the validation data set is shown in Fig. 3. Although the sixth neuron was able to distinguish *P. niruri* from *P. urinaria*, the fifth neurons showed that both these samples have quite similar spectra as compared to the other four species.

It has been stated [30] that in an ideal classification, only one output neuron associated with the particular species from which the data was used, should have an output of one. At the same time, all

Fig. 3. The average scores for the six output neurons for neural network with transfer functions of hyperbolic tangent sigmoid (tansig) for both the hidden and output neurons. In ideal classification output neuron associated with species would give value of 1; other neurons would give values of 0 or less than 0.

other output neurons should have an output of zero. This could be achieved perhaps, if a larger data set is used and the number of epoch for the algorithm is increased.

4. Conclusion

Four different chemometric approaches consisting of PCA-LDA, GA-LDA, SIMCA and neural networks were evaluated with the best classification according to species achieved by GA-LDA. PCA-LDA gave the next best results with neural networks showing classification ability close to PCA-LDA. However, the advantage of neural networks is that the underlying variables that are important for discrimination is not known and this would discourage adulteration if the approach were used in an industrial setting. *P. niruri*, *P. debilis* and *P. urinaria* not only possessed morphological similarities but also showed similar FTIR spectra which were only distinguished easily by the more sophisticated method of GA-LDA. Future work will attempt to improve classification of larger set of samples using different GA approaches to fine tune the important GA parameters of fitness function, crossover and mutation.

Acknowledgements

An Intensifying Research Priority Areas (IRPA) grant from Ministry of Science Technology and Innovation (MOSTI), Malaysia supported the study. Saravanan Dharmaraj wishes to thank Universiti Sains Malaysia for a post-doctoral fellowship.

References

- [1] G.B. Alcantara, N.K. Honda, M.M.C. Ferreira and A.G. Ferreira, Chemometric analysis applied in ¹H HR-MAS NMR and FT-IR data for chemotaxonomic distinction of intact lichen samples, *Anal. Chim. Acta* **595** (2007), 3–8.
- [2] Anonymous, *Quality Control Methods for Medicinal Plant Materials*, World Health Organization, Geneva, 1998.
- [3] Y. Chen, G. Fan, Q. Zhang, H. Wu and Y. Wu, Fingerprint analysis of the fruits of *Cnidium monnieri* by high-performance liquid chromatography–diode array detection–electrospray ionization tandem mass spectrometry, *J. Pharmaceut. Biomed. Anal.* **43** (2007), 926–936.
- [4] H. Demuth, M. Beale and M. Hagan, *Neural Network Toolbox*, The MathWorks, Natick, 2005.
- [5] S. Dharmaraj, A.S. Jamaludin, H.M. Razak, R. Valliappan, N.A. Ahmad, G.L. Harn and Z. Ismail, The classification of *Phyllanthus niruri* Linn. according to location by infrared spectroscopy, *Vib. Spectrosc.* **41** (2006), 68–72.
- [6] W. Dnyaneshwar, C. Preeti, J. Kalpana and P. Bhushan, Development and application of RAPD-SCAR marker for identification of *Phyllanthus emblica* Linn, *Biol. Pharm. Bull.* **29** (2006), 2313–2316.
- [7] G. Downey, Food and food ingredient authentication by mid-infrared spectroscopy and chemometrics, *Trends Anal. Chem.* **17** (1998), 418–424.
- [8] M. Frederich, Y.H. Choi, L. Angenot, G. Harnischfeger, A.W.M. Lefeber and R. Verpoorte, Metabolomic analysis of *Strychnos nux-vomica*, *Strychnos icaia* and *Strychnos ignatii* extracts by ¹H nuclear magnetic resonance spectrometry and multivariate analysis techniques, *Phytochem.* **65** (2004), 1993–2001.
- [9] R. Goodacre, E.V. York, J.K. Heald and I.M. Scott, Chemometric discrimination of unfractionated plant extracts analyzed by electrospray mass spectrometry, *Phytochem.* **62** (2003), 859–863.
- [10] M.T. Hagan, H.B. Demuth and M. Beale, *Neural Network Design*, Thompson Learning, Singapore, 2004.
- [11] S. Haykin, *Neural Network. A Comprehensive Foundation*, Prentice-Hall of India, New Delhi, 2005.
- [12] M.M.W.B. Hendriks, L. Cruz-Juarez, D. De Bont and R.D. Hall, Preprocessing and exploratory analysis of chromatographic profiles of plant extracts, *Anal. Chim. Acta* **545** (2005), 53–64.
- [13] E. Holmes and H. Antti, Chemometric contributions to the evaluation of metabonomics: mathematical solutions to the characterizing and interpreting complex biological NMR spectra, *Analyst* **127** (2002), 1549–1557.
- [14] A.K. Khanna, F. Rizvi and R. Chander, Lipid lowering activity of *Phyllanthus niruri* in hyperlipidemic rats, *J. Ethnopharmacol.* **82** (2002), 19–22.
- [15] H.K. Kim, Y.H. Choi, C. Erkelens, A.W.M. Lefeber and R. Verpoorte, Metabolic fingerprinting of *Ephedra* species using ¹H NMR spectroscopy and principal component analysis, *Chem. Pharm. Bull.* **53** (2005), 105–109.
- [16] B.K. Lavine, C.E. Davidson and A.J. Moores, Genetic algorithms for spectral pattern recognition, *Vib. Spectrosc.* **28** (2002), 83–95.
- [17] B.K. Lavine, A.J. Moores, H.T. Mayfield and A. Faraque, Fuel spill identification by gas chromatography-genetic algorithms/pattern recognition techniques, *Anal. Lett.* **31** (1998), 2805–2822.
- [18] B.K. Lavine, A.J. Moores, H. Mayfield and A. Faraque, Genetic algorithms applied to pattern recognition analysis of high-speed gas chromatograms of aviation turbine fuels using an integrated jet-A/JIP-8 database, *Microchem. J.* **61** (1999), 69–78.
- [19] A.E. Nikulin, B. Dolenko, T. Bezabeh and R.L. Somorjai, Near-optimal region selection for feature space reduction: novel preprocessing methods for classifying MR spectra, *NMR Biomed.* **11** (1998), 209–216.
- [20] A.W. Obianime and F.I. Uche, The comparative effects of methanol extract of *Phyllanthus amarus* leaves and vitamin E on the sperm parameters of male guinea pigs, *J. Appl. Sci. Environ. Manag.* **13** (2009), 37–41.
- [21] N.V. Rajeshkumar, K.L. Joy, G. Kuttan, R.S. Ramsewak, M.G. Nair and R. Kuttan, Antitumor and anticarcinogenic activity of *Phyllanthus amarus* extract, *J. Ethnopharmacol.* **81** (2002), 17–22.
- [22] C. Reynes, S. de Souza, R. Sabatier, G. Figueres and B. Vidal, Selection of discriminant wavelength intervals in NIR spectrometry with genetic algorithms, *J. Chemom.* **20** (2006), 136–145.
- [23] A.R. Robinson, R. Gheneim, R.A. Kozaak, D.D. Ellis and S.D. Mansfield, The potential of metabolite profiling as a selection tool for genotype discrimination in *Populus*, *J. Exp. Bot.* **56** (2005), 2807–2819.
- [24] P.N. Soh, J.T. Banzouzi, H. Mangombo, M. Lusakibanza, F.O. Bulubulu, L. Tona, A.N. Diamuini, S.N. Luyindula and F. Benoit-Vical, Antiplasmodial activity of various parts of *Phyllanthus niruri* according to its geographical distribution. *Afr. J. Pharm. Pharmacol.* **3** (2009), 598–601.
- [25] S.F. Sulaiman and A.S. Othman, DNA fingerprinting of three morphological confusing species from the genus *Phyllanthus*: a plant genus with medicinal properties, in: *Proceedings of the International Conference on Traditional/Complementary Medicine*, Ministry of Health Malaysia, Kuala Lumpur, November 13–15, 2000.
- [26] P. Theerakulpisut, N. Kanawapee, D. Maensiri, S. Bunnag and P. Chantaranonthai, Development of species specific SCAR markers for identification of three medicinal species of *Phyllanthus*, *J. Systemat. Evol.* **46** (2008), 614–621.
- [27] R.P. Umbare, G.S. Mate, D.V. Jawalkar, S.M. Patil and S.S. Dongare, Quality evaluation of *Phyllanthus amarus* (Schumacher) leaves extract for its hypolipidemic activity, *Biol. Med.* **1** (2009), 28–33.

- [28] F.L. Van Holthoon, *Phyllanthus* L., in: *Plant Resources of South-East Asia, No. 12(1), Medicinal and Poisonous Plants 1*, L.S. de Padua, N. Bunyaphatsara and R.H.M.J Lemmens, eds, Backhuys Publishers, Leiden, 1999. pp. 381–392.
- [29] Y.A. Woo, H.J. Kim, K.R. Ze and H. Chung, Near-infrared (NIR) spectroscopy for the non-destructive and fast determination of geographical origin of *Angelicae gigantis* Radix, *J. Pharmaceut. Biomed. Anal.* **36** (2005), 955–959.
- [30] J. Zupan, M. Novic, X. Li and J. Gasteiger, Classification of multicomponent analytical data of olive oils using different neural networks, *Anal. Chim. Acta* **292** (1994), 219–234.