



# Phenolic variation among *Chamaecrista nictitans* subspecies and varieties revealed through UPLC-ESI(-)-MS/MS chemical fingerprinting

Luis Quirós-Guerrero<sup>1,2</sup> · Federico Albertazzi<sup>3,4</sup> · Emanuel Araya-Valverde<sup>5,7</sup> · Rosaura M. Romero<sup>1,2</sup> · Heidy Villalobos<sup>2,3</sup> · Luis Poveda<sup>6</sup> · Max Chavarría<sup>1,2,5</sup> · Giselle Tamayo-Castillo<sup>1,2</sup>

Received: 28 August 2018 / Accepted: 11 January 2019 / Published online: 19 January 2019  
© Springer Science+Business Media, LLC, part of Springer Nature 2019

## Abstract

**Introduction** Comparative analysis of metabolic features of plants has a high potential for determination of quality control of active ingredients, ecological or chemotaxonomic purposes. Specifically, the development of efficient and rapid analytical tools that allow the differentiation among species, subspecies and varieties of plants is a relevant issue. Here we describe a multivariate model based on LC–MS/MS fingerprinting capable of discriminating between subspecies and varieties of the medicinal plant *Chamaecrista nictitans*, a rare distributed species in Costa Rica.

**Methods** Determination of the chemical fingerprint was carried out on a LC–MS (ESI-QTOF) in negative ionization mode, main detected and putatively identified compounds included proanthocyanidin oligomers, several flavonoid C- and O-glycosides, and flavonoid acetates. Principal component analysis (PCA), partial least square-discriminant analysis (PLS-DA) and cluster analysis of chemical profiles were performed.

**Results** Our method showed a clear discrimination between the subspecies and varieties of *Chamaecrista nictitans*, separating the samples into four fair differentiated groups: M1 = *C. nictitans* ssp. *patellaria*; M2 = *C. nictitans* ssp. *disadena*; M3 = *C. nictitans* ssp. *nictitans* var. *jaliscensis* and M4 = *C. nictitans* ssp. *disadena* var. *pilosa*. LC–MS/MS fingerprint data was validated using both morphological characters and DNA barcoding with ITS2 region. The comparison of the morphological characters against the chemical profiles and DNA barcoding shows a 63% coincidence, evidencing the morphological similarity in *C. nictitans*. On the other hand, genetic data and chemical profiles grouped all samples in a similar pattern, validating the functionality of our metabolomic approach.

**Conclusion** The metabolomic method described in this study allows a reliably differentiation between subspecies and varieties of *C. nictitans* using a straightforward protocol that lacks extensive purification steps.

**Keywords** *Chamaecrista nictitans* · Chemotaxonomy · Metabolite fingerprinting · LC–MS/MS · Chemical fingerprinting

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s11306-019-1475-8>) contains supplementary material, which is available to authorized users.

✉ Giselle Tamayo-Castillo  
giselle.tamayo@ucr.ac.cr

<sup>1</sup> Centro de Investigaciones en Productos Naturales (CIPRONA), Universidad de Costa Rica, San Jose 11501-2060, Costa Rica

<sup>2</sup> Escuela de Química, Universidad de Costa Rica, Sede Central, San Pedro de Montes de Oca, San Jose 11501-2060, Costa Rica

<sup>3</sup> Centro de Investigación en Biología Celular y Molecular (CIBCM), Universidad de Costa Rica, San Jose 11501-2060, Costa Rica

<sup>4</sup> Escuela de Biología, Universidad de Costa Rica, San Jose 11501-2060, Costa Rica

<sup>5</sup> Centro Nacional de Innovaciones Biotecnológicas (CENiBiot), CeNAT-CONARE, San Jose 1174-1200, Costa Rica

<sup>6</sup> Herbario Juvenal Valerio Rodríguez, Universidad Nacional, Heredia, Costa Rica

<sup>7</sup> Escuela de Biología, Instituto Tecnológico de Costa Rica, Cartago 159-7050, Costa Rica

## 1 Introduction

The metabolite fingerprinting aims to classify a large number of samples using multivariate statistics analysis, which reveal particular chemical patterns within a complex mixture, with or without identification of their components, and usually with a preliminary separation of the matrix (e.g. use of chromatographic techniques) (Frag et al. 2017; Luthria et al. 2008; Ryan and Robards 2006; Salvador et al. 2016). The patterns are associated to the presence and concentration of certain type of natural products biosynthesized according to the expression of genes in a reduced population associated to a common ancestor (Glaser et al. 2013; Sarker and Nahar 2012; Semmar et al. 2007) and to the plant environment (Hao et al. 2015; Williams et al. 1989; Wink 2010). The metabolite fingerprinting could assist in the determination of food quality (Hummer et al. 2014; Kashif et al. 2009; Klockmann et al. 2016; Lee et al. 2009; Liu et al. 2017; Magagna et al. 2017; Mayorga-Gross et al. 2016; Pongsuwan et al. 2007; Ronningen et al. 2017; Zieliński et al. 2014), geographical origin (Consonni et al. 2012; Kim et al. 2013; Mehari et al. 2016; Panet et al. 2016; Watson et al. 2006), ecological and evolution studies or assessment of plant specimens classification (Hegnauer 1986; Luthria et al. 2008; Mattoli et al. 2006; Musah et al. 2015; Wink 2003).

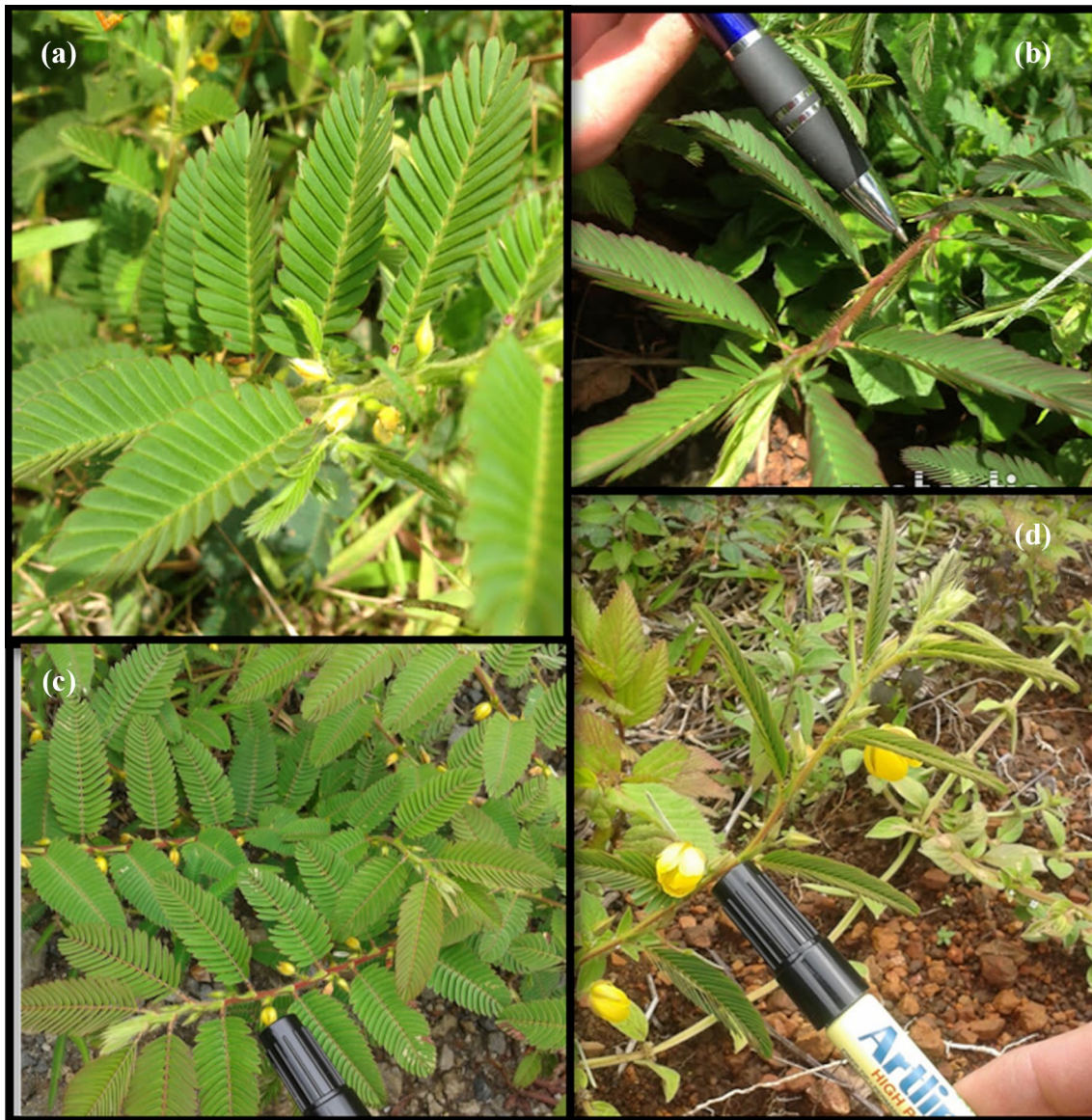
After morphological characters were established as a general rule for plant taxonomy, the independent work of some botanists caused a certain level of discrepancies principally due to the lack of communication and the huge diversity of plant families (Godfray and Knapp 2004). This problem was lessened with arisen of molecular taxonomy techniques based on the analysis of specific DNA regions (Govindaraghavan et al. 2012; Walker 2014). Even so, there are some gaps associated to the usual protocols followed specially with the freshness of collected plant tissue or the size of the sample or when morphological characters are not enough, which could be surpassed with the chemical fingerprinting using a sensible technique, such as Liquid Chromatography coupled to Mass Spectrometry (LC–MS). This technique is ideal for identification of secondary metabolites in complex biological matrices, such as, fruits, vegetables, seeds and plants extracts without any purification procedure. In addition, tandem mass spectrometry experiments provide useful structural information of compounds when their identification is needed (Del Bubba et al. 2012; Stobiecki 2000; Stobiecki et al. 2015).

Several metabolomic methods have been reported for chemotaxonomic purposes. For example, *Fabaceae* super family exhibits a huge diversity of nitrogen-rich natural products such as pyrrolizidine, quinolizidine, indol-type and imidazol-type alkaloids, protein, non-protein

amino acids and polyamides, which have been proposed as chemical markers for chemotaxonomic classification (Wink 2003, 2013). Specific use of phenolic compounds as markers for discrimination between species is reported in several families including *Fabaceae* (Visnevschi-Necrasov et al. 2015), *Lamiaceae* (Kharazian 2014), *Araceae* (Clark et al. 2014), *Apiaceae* (Güzel et al. 2011) and *Moraceae* (Venkataraman 1972). Visnevschi-Necrasov et al. (2015) observed genus level discrimination when using quantitative HPLC determination of isoflavones present in nine species of four different genera of the *Fabaceae* family. The ubiquity of polyphenols and their value as markers could be advantageous in the discrimination of phylogeny of the plants (Wink 2003, 2013) instead of the use of specific compounds such as alkaloids, lactones or terpenes, which have a patchy and irregular distribution into the same family.

*Chamaecrista nictitans* (Fabaceae/Caesalpiaceae) is reported to be an annual, herbaceous, leguminous plant native to the United States, Mexico, parts of South and Central America according to Tropicos database (“Tropicos” 2017). It is found in the wild, growing alongside rural roadsides. It has been recently associated with antiviral activity (Herrero Uribe et al. 2004). Phytochemical screening of *C. nictitans* (Fabaceae/Caesalpiaceae) in Costa Rica was reported by Herrero Uribe et al. (2004), whom associated the mechanism of action against *Herpes simplex II* virus of purified fractions of *C. nictitans* with the presence of certain polyphenolic compounds. Chemical constituents were described as a mixture of complex tannins according to their <sup>1</sup>H-NMR fingerprint and qualitative TLC analysis. Nevertheless, several subspecies and varieties of *C. nictitans* are found in Costa Rica (see Fig. 1) and the lack of agreement among taxonomists conspired to make a proper classification of the species (Hammel et al. 2010). These constraints made difficult our efforts to isolate active components. Further studies using LC–MS/MS demonstrated that all *C. nictitans* exhibit a high content of proanthocyanidin oligomers, whereas *C. nictitans* subspecies *patellaria* presents particular monohydroxyphenol substructures such as (epi)gubourtinidol, (epi)afzelechin and (epi)fisetinidol, with B-type and A-type linkages (Mateos-Martín et al. 2014).

The aim of this study was to validate the use of a metabolite fingerprinting method based on LC–MS/MS, capable of discriminating between subspecies and varieties of *C. nictitans*, using a straightforward protocol. Results demonstrate that our approach was able to discriminate between the subspecies and varieties measured, separating the samples into four differentiated groups: M1 = *C. nictitans* ssp. *patellaria*; M2 = *C. nictitans* ssp. *disadena*; M3 = *C. nictitans* ssp. *nictitans* var. *jalisensis* and M4 = *C. nictitans* ssp. *disadena* var. *pilosa*. Thus, this study reports a useful tool to discriminate



**Fig. 1** Representative pictures of collected specimens: **a** *C. nictitans* var. *patellaria*; **b** *C. nictitans* ssp. *disadena* var. *pilosa*; **c** *C. nictitans* ssp. *nictitans* var. *jaliscensis*; and **d** *C. nictitans* ssp. *disadena*

between subspecies and varieties of *C. nictitans* in a simple and fast way.

## 2 Materials and methods

### 2.1 Plant material collection

Forty samples were collected in different sites of the Central Valley and the Pacific coast of Costa Rica from September 2013 until February 2014, from October 2014 until November 2014 and July 2015, (see Table S1 in Supplementary Information) including subspecies *patellaria* and *disadena*;

ssp. *nictitans* var. *jaliscensis* and ssp. *disadena* var. *pilosa* all identified by Prof. Luis Poveda (National University of Costa Rica) in the field. Green fresh aerial parts were sampled for all specimens (including stems, leaves and flowers) except for collections from Puntarenas, which were collected without flowers in an early stage of development. Each sample was split into two: one for chemical analysis and another for genetic taxonomy. For most samples, a voucher specimen was collected and kept at the Herbarium of the National University of Costa Rica. Centro de Investigaciones en Productos Naturales (CIPRONA, UCR) provided four specimens cultivated in a greenhouse (*C. nictitans* ssp. *patellaria* n=3 and *C. nictitans* ssp. *disadena* n=1), and two in vitro

samples in early stage (*C. nictians* ssp. *patellaria* n=1 and *C. nictians* ssp. *disadena* n=1). Herbarium of the University of Costa Rica (USJ, UCR) provided 5 samples of genus *Chamaecrista* with three different species than *C. nictitans* (*C. rotundifolia* n=1, *C. diphyllia* n=1, *C. flexuosa* n=4) and seven samples of *C. nictitans* species not further classified including two samples collected in United States (see Table S2). A grand total of 58 samples were included in the experiment.

## 2.2 In vitro cultures and greenhouse conditions

For the establishment of in vitro cultures, seeds were removed from fruits of *C. nictitans* ssp. *patellaria*, *C. nictitans* ssp. *disadena*, *C. nictitans* ssp. *nictitans* var. *jalisensis* and *C. nictitans* ssp. *disadena* var. *pilosa*. The seeds were rigorously washed with water and soap. Afterwards they were dipped into ethanol solution (70% v/v) followed with sodium hypochlorite solution (3% v/v, 15 min) followed by triplicate washes with sterile water. Seeds from *C. nictitans* ssp. *patellaria* and *C. nictitans* ssp. *disadena* germinated after 2 months of culture on a MS medium (Murashige and Skoog 1962) devoid of growth regulators, with activated charcoal (1.5 g/L) and solidified agar (7 g/L). The pH of the culture media was adjusted to 5.7. Culture media was autoclaved (121 °C, 20 min). The cultures were kept in a controlled environment (25 °C with 16-h photoperiod). For the establishment of greenhouse cultures, seeds were planted in plastic pots (30 cm of diameter), containing disinfected sandy loam soil under greenhouse conditions. Plants from *C. nictitans* ssp. *patellaria* and *C. nictitans* ssp. *disadena* emerged after 4 months.

## 2.3 Plant material extraction

Fresh chopped material (1 g) was extracted once in a 25 mL Erlenmeyer with methanol (10 mL, HPLC grade, J.T. Baker, Center Valley, USA) with sonication (10 min). The alcoholic extract was recovered by filtration on a glass funnel with cotton and completely dried under vacuum in a R200 rotary evaporator (Büchi, Flawil, Switzerland) at 40 °C. After evaporation, 100 mg of the solid residue were dissolved in methanol (10 mL, HPLC grade, J.T. Baker, Center Valley, USA) and eluted through a 100 mg STRATA-C18 cartridge (Phenomenex, Torrance, USA) fitted on a solid phase extraction system (SPE-SUPELCO, Darmstadt, Germany); the resulting filtrate yields a brownish to yellowish color, depending on the sample. The filtrate was recovered and evaporated to dryness under vacuum in a centrifugal evaporator (SpeedVac SC200, Waltham, USA) equipped with a RVT400 refrigerated vapor trap (Thermo Savant, Waltham, USA).

Samples from the herbaria (ca 250 mg) were extracted twice with 10 mL of methanol (HPLC grade, J.T. Baker,

Center Valley, USA) and once with 5 mL of methanol (HPLC grade, J.T. Baker, Center Valley, USA) with sonication (15 min each); extracts were combined and concentrated to 10 mL. The remaining extract was filtrated through a 50 mg STRATA-C18 cartridge (Phenomenex, Torrance, USA) as described above. Filtrates were dry under same conditions as above. Two subsamples of each filtrate were weighed out (2 mg each) and dissolved to a 2 mg/mL concentration with a mixture of acetonitrile:water (8:2; Optima, Fisher Scientific, Waltham, USA). The injected sample consisted of a mixture of 800 µL of the filtrate solution and 70 µL of an acyclovir (USP Reference standard, purity 99.3%), solution (1 mg/mL, in acetonitrile:water 8:2) as internal standard.

## 2.4 Chromatographic and Mass Spectrometry conditions

Chromatographic profiles were performed on a 50×2.1 mm Waters® ACQUITY™ 1.7 µm BEH C<sub>18</sub> column (Waters, Milford, USA) in an ACQUITY Ultra Performance LC™ system equipped with an auto sampler and PDA detector. The column was kept at 40 °C and the auto sampler at 10 °C. Injection volume was 2.0 µL. The flow rate was set to 550 µL/min and a gradient elution was carried out with a binary system consisting of [A] 0.1% aqueous formic acid (Optima, Fisher Scientific, USA) and [B] 0.1% formic acid (Optima, Fisher Scientific, Waltham, USA) in acetonitrile (Optima, Fisher Scientific, Waltham, USA). An increasing linear gradient (v/v) of [B] was used as follow [t(min), %B]: 0.00, 8; 4.30, 20; 10.00, 35; 12.00, 70; 16.00, 92; followed by re-equilibration steps (20.00, 8; 22.00, 8). PDA detector was set from 190 to 600 nm with a resolution of 1.2 nm. Mass spectrometry was performed on a Waters® SYNAPT ESI-QToF system (Waters, Milford, USA). Desolvation gas (N<sub>2</sub>) flow was set to 450 L/h and a desolvation temperature of 150 °C, cone gas (N<sub>2</sub>) flow was set to 10 L/h and Source temperature was set to 200 °C. The capillary voltage and sampling cone voltage were set to 3.0 kV and 35 V respectively. Extraction cone voltage was set to 4.0 V for negative operation mode. Detection was performed using two different MS functions, one of low collision energy (4.0 V) to record the exact mass and a second one of high energy using a collision energy ramp (10–40 V) for obtaining preliminary fragmentation patterns. Both with a scan time and an inter scan time delay of 0.2 and 0.02 s respectively. MS/MS experiments were obtained using collision induced dissociation (CID) functions with collision energy from 25 to 55 eV depending of the molecule.

All analysis were acquired using Lock Spray™, Leucine-enkephalin was used as lock mass (V<sup>-</sup>: 554.2615). Data was collected in centroid mode, with a lock spray frequency of 10 s, and data was averaged over 10 scans. The Synapt was

calibrated in negative mode with sodium formate (reference mass 860.8467 uma) for an  $m/z$  range from 100 to 1200 in negative ionization mode. MassLynx software (version 4.1, Waters) was used for acquisition and data processing.

## 2.5 MZmine data processing and multivariate data analysis

Data of LC–MS runs were treated with MZmine v2.33 software (Pluskal et al. 2010) for data mining, considering all peaks with intensity above 20, using the Grid Mass (Treviño et al. 2015) algorithm with an  $m/z$  tolerance of 8 ppm and a min–max width time of 0.03–0.5 min, ignoring detection after 11 min. Afterwards, deisotoping and filtering procedures were performed to remove all peaks without isotopic pattern. Alignment was performed using Join Aligner algorithm with a retention time tolerance of 0.1 min and  $m/z$  tolerance of 8 ppm. Gap filling was performed using the Same RT and  $m/z$  Range Gap Filler algorithm with a RT tolerance of 0.3 min and an  $m/z$  tolerance of 8 ppm.

The processed data (\*.CSV format) was fed to SIMCA-P 13.0.2.0 software (Umetrics, Umea, Sweden) for multivariate data analysis using the Principal Component Analysis (PCA) algorithm for studying tendency of different samples to group. Afterwards, a Partial Least Square Discriminant Analysis (PLS-DA) algorithm was applied. The significance of the model was determined with a cross-validated analysis of variance (CV-ANOVA)  $p$ -value < 0.001 and its quality was evaluated by the  $R^2X$  (0.837) and  $R^2Y$  (0.909) values. Goodness of fit and percentage of explained variability, and the  $Q^2X$  (0.842) for predictive capacity of the model were determined. A permutation test was performed to evaluate data over fitting ( $n = 200$ ). The degree of discrimination of each ion between the different groups was assessed by performing and assigning a Variable Importance Projection index (VIP index) to each variable. The studied ions considered important have a VIP value higher than 1.00 and ANOVA  $p$ -value lower than 0.05.

## 2.6 Metabolite identification

Structures were tentatively proposed based on fragmentation patterns generated due to certain types of fragmentation reactions (Retro Diels–Alder [RDA] and heterolytic ring fission [HRF]), which provide information on the hydroxylation pattern on A, B and C-ring (Figure S1a) of the flavonoid unit, bonds between two units (Quinone Methide [QM]) and the unit itself (Jaiswal et al. 2012). Fragmentation takes place from the top unit (more prone to fragment) to the base unit in the compound, giving information about the order of the subunits in the polyphenolic structure (Calderón et al. 2009; Friedrich et al. 2000). The presence of *C*- or *O*-glycosides gives characteristic fragments through the cleavage

of two C–C bonds in different positions of the sugar moiety. The classical nomenclature was employed to describe the fragmentations pathways of the compounds and position of *C*-glycosides were assigned by comparing relative intensities of the fragments according to literature (Figure S1.b) (Benayad et al. 2014; Farias and Mendez 2014; March et al. 2004, 2006; Stobiecki 2000; Stobiecki et al. 2015).

## 2.7 DNA barcoding analysis

The plants ITS2 region was sequenced over a subset of 20 *C. nictitans* samples including subspecies and varieties (Table S1). Selection of *C. nictitans* samples was based on identified metabolites data matrix as described chemo types in Fig. 1 (M1 = *C. nictitans* ssp. *patellaria*; M2 = *C. nictitans* ssp. *disadena*; M3 = *C. nictitans* ssp. *nictitans* var. *jaliscensis*; M4 = *C. nictitans* ssp. *disadena* var. *pilosa*). The former identification by Prof. Luis Poveda (Table S3) was kept in the phylogenetic analysis to compare with molecular and chemical analysis. DNA was isolated from approximately 50 mg of disrupted leaf tissue with a micro mortar. 750  $\mu$ L of lysis buffer (20 mM sodium EDTA pH 8.0, 100 mM Tris–HCl pH 8.0, 1.4 M NaCl, 2.0% CTAB, 2% PVP and 0.2% of beta-mercaptoethanol) was added. After an incubation at 60 °C for 20 min, 750  $\mu$ L of chloroform:octanol (24:1) was added and tubes mixed gently by inverting 20 times. The mixture was centrifuged at 13,000 rpm for 5 min and 400  $\mu$ L of the aqueous phase was transferred into a new 1.5 mL reaction tube. One volume of cold isopropanol was added to precipitate the DNA at room temperature for 5 min. The solution was centrifuged at 13,000 rpm for 5 min, the supernatant was poured off and the pellet kept at the bottom of the tube. The pellet was washed with 70% ethanol (500  $\mu$ L) and centrifuged at 13,000 rpm for 2 min. The ethanol was removed, and the pellet dried at 37 °C. Finally, the pellet was dissolved in 50  $\mu$ L of TE buffer with treatment of 1  $\mu$ L of RNase (Fermentas). After an incubation at 37 °C for 1 h, the DNA was measured spectrophotometrically and diluted at 50 ng/ $\mu$ L for PCR amplification. Isolated DNA was amplified with primers for ITS2 region. We used S2F forward primer (5'ATG CGATACTTGGTGTGAAT'3) designed by Chen et al. (2010) and S3R reverse primer (5'GACGCTTCTCCAGAC TACAAT'3) designed by Chiou et al. (2007). PCR reactions were conducted in a final volume of 20  $\mu$ L, containing 1 $\times$  of HotStart Ready Mix (Fermentas) and 0.250 mM of each primer. Finally, 2  $\mu$ L of DNA (ranging 50–100 ng/ $\mu$ L) were added to the final reaction. PCR were performed with an initial denaturalization step of 95 °C for 5 min, followed by 35 cycles of 95 °C for 30 s, 55 °C for 30 s. and 72 °C for 1 min, and a final elongation step of 72 °C for 10 min. The Veriti™ 96-Well Thermal Cycler (Applied Biosystems, USA) were used for PCR reactions and for further steps for sequencing.

For sequencing, 10  $\mu\text{L}$  of PCR products were purified using 2  $\mu\text{L}$  of Exonuclease I (Thermo Scientific Fermentas, USA) and 1  $\mu\text{L}$  of Phosphatase Alkaline (Thermo Scientific Fermentas, USA). The mixture was incubated with an initial temperature of 37  $^{\circ}\text{C}$  for 15 min, followed by 85  $^{\circ}\text{C}$  for 15 min. Sequencing reactions were conducted with Big Dye Terminator 3,1<sup>®</sup> Kit (Applied Biosystems, USA) with final primer concentrations of 0.320  $\mu\text{M}$  of either forward or reverse primer. The thermal profile was as follows: 1 cycle of initial temperature of 96  $^{\circ}\text{C}$  for 1 min; 15 cycles of 96  $^{\circ}\text{C}$  for 10 s, 50  $^{\circ}\text{C}$  for 0.05 s and 60  $^{\circ}\text{C}$  for 75 s; 5 cycles of 96  $^{\circ}\text{C}$  for 10 s, 50  $^{\circ}\text{C}$  for 0.05 s and 60  $^{\circ}\text{C}$  for 90 s; 5 cycles of 96  $^{\circ}\text{C}$  for 10 s, 50  $^{\circ}\text{C}$  for 0.05 s and 60  $^{\circ}\text{C}$  for 120 s. Sequencing products were purified with Big Dye XTerminator<sup>®</sup> kit (Applied Biosystems, USA) according to manufacturer instructions. Purified sequencing reactions were analysed in a 3130xl Genetic Analyzer (Applied Biosystems, USA) with 50 cm capillary and POP7 polymer (Applied Biosystems, USA).

Forward and reverse strands were manually inspected in FinchTV (Geospiza) and a consensus sequence was obtained in BioEdit (Hall 1999). Multiple sequence alignments of ITS2 region were conducted using MEGA version 6 (Tamura et al. 2013) with MUSCLE option (Edgar 2004) using sequences from this study. Besides, sequences of *C. nictitans* collected in Costa Rica previously deposited by Prof. Federico Albertazzi were retrieved from GeneBank database (Accession codes: KU720152.1, KU720153.1, KU720154.1, KU720155.1, KU720156.1, KU720157.1, KU720158.1, KU720159.1, KU720160.1, KU720161.1, KU720162.1, KU720163.1, KU720164.1, KU720165.1, KU720166.1, KU720167.1). These sequences included 18S, ITS1, 5.8S and ITS2 regions of the ribosomal RNA gene. Retrieved sequences were trimmed to align the ITS2 region to sequences in the current study. After running the best nucleotide substitution models in MEGA 6 (Tamura et al. 2013), “Tamura three parameters” model was selected to build the phylogenetic tree with the Neighbour Joining (NJ) grouping method (with 1000 bootstrap replicates). The same model allowed the estimation of distances between groups, the interspecific and intraspecific diversity.

### 3 Results and discussion

#### 3.1 Statistical analysis: construction and evaluation of the model

A principal component algorithm was applied to a matrix based on over 2000 ions detected in negative ionization mode (both modes were tested and negative mode was preferred given its higher total intensity, see Figure S2), obtaining the scatter and loadings projections (Figure S3 a

and b) showing four defined groups with a fair separation and resolution through an untargeted clustering analysis in SIMCA-P software. The space built by the first two components explains 52% of the total variation. Outliers from group M3 and M4 were evaluated and no chemical differences were found according to chromatograms and mass spectra; hence, they were conserved for further statistical analysis; their location may be a direct consequence of variation in concentration of the metabolites in their profiles.

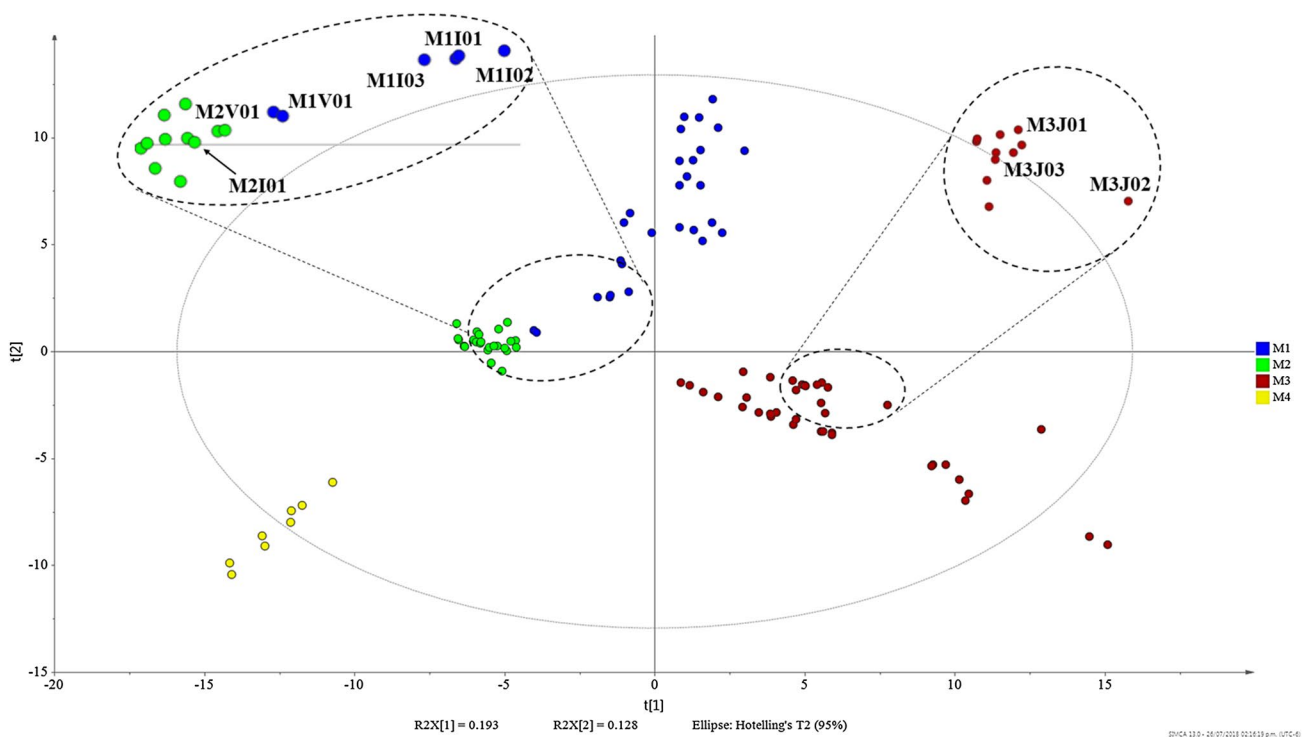
PLS-DA algorithm on the same variables generated the projection shown in Fig. 2. Validity of the model was determined acceptable by assessing cumulative ratio Q2 (cum) (0.842) and the analysis of variance of cross-validation predictive residuals (CV-ANOVA  $p$ -value < 0.001). A later Ward’s minimum-variance clustering based on the PLS-DA analysis was performed (figure S4) showing a good separation between subspecies and varieties, therefore grouping of the samples is suggested to be tightly related to their chemical composition.

PLS-DA separates group M1, M2 and M3 from M4 in the first component, this separation is also observed in the HCA in figure S4, where groups M2 and M4 are derived from the same branch and are more related to M1 than M3.

Observed segregation is in accordance with chemical profile variation observed in LC–MS profiles (See figure S5). Each subspecies and variety produces specific metabolites but also share some of them (See Table S3). For example, *C. nictitans* ssp. *patellaria* (M1) and *C. nictitans* ssp. *nictitans* var. *jaliscensis* (M3), both produce proanthocyanidin oligomers up to tetramers and some specific *C*-flavonoid glycosides, such as Cassiaoccidentalinal A (35), Cassiaoccidentalinal B (34) and their supposed isomers. However both differ in the presence of a trimer with  $[\text{M}-\text{H}]^{-}$  at  $m/z$  817.21 (retention time 6.84 min, 10) described as (epi)afz-(epi)fis-(epi)fis, which is a major component in the former species, and the presence of the *C*-flavonoid glycoside Luteolin-6-*C*-hexosyl-(1 $\rightarrow$ 2)-rhamnoside (28) in *C. nictitans* ssp. *nictitans* var. *jaliscensis*.

*Chamaecrista nictitans* ssp. *disadena* (M2) shares mainly proanthocyanidin trimeric compounds with *C. nictitans* ssp. *nictitans* var. *jaliscensis* and produces *C*-flavonoid glycosides in high concentration. On the other hand, *C. nictitans* ssp. *disadena* var. *pilosa* (M4) possess a highly specific metabolism, centered on the production of *C*- and *O*-flavonoid glycosides, and some acetylated compounds such as flavonoid acetates.

Thus, it is possible to differentiate up to the subspecies and variety level, plants belonging to *C. nictitans* species thanks to their distinctive chemical profiles, given that the statistical analysis yields good separation between subgroups. The proposed methodology proves to be simpler and more straightforward than morphological and molecular taxonomy for these specific samples. The comparison of the



**Fig. 2** Partial least square—discriminant analysis score plot (C1 vs. C2) showing 46 samples from *C. nictitans* including varieties and subspecies, based on identified metabolites data matrix. (M1 = *C. nic-*

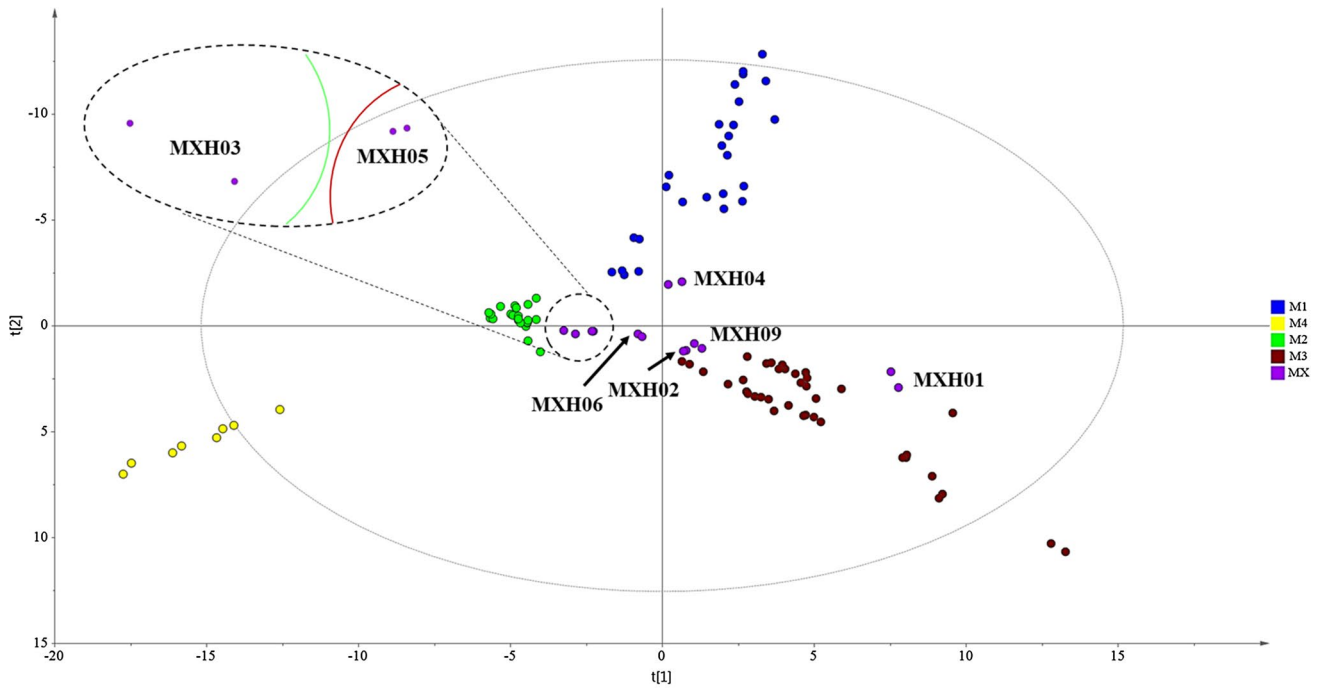
*titans* var. *patellaria*; M2 = *C. nictitans* ssp. *disadena*; M3 = *C. nictitans* ssp. *nictitans* var. *jaliscensis*; M4 = *C. nictitans* ssp. *disadena* var. *pilosa*)

morphological characters against the chemical profiles and DNA barcoding shows a 63% coincidence, evidencing the morphological similarity in *C. nictitans*.

The validity and usefulness of this analysis rests on the ability of identifying species and varieties not depending on age, fertility, or even chemical profile variation due to environmental factors such as geographical distribution, soil and direct sun light exposure or even multi-year effect for collection in different years as ours samples. For this, a myriad of samples was obtained and evaluated. The results are in good agreement with the corresponding cluster (Fig. 2). First, greenhouse cultivated plants of *C. nictitans* ssp. *disadena* ( $n = 1$ , M2I01) and *C. nictitans* ssp. *patellaria* ( $n = 3$ , MI01–MI03) were tested, as well as immature specimens of *C. nictitans* ssp. *nictitans* var. *jaliscensis* [ $n = 3$ , M3J01–M3J03, taxonomy associated based on geographical distribution (Hammel et al. 2010)], which were collected in Jacó, Puntarenas at sea level. Lastly, two early stage in vitro samples (M1V01, M2V02) were analyzed to evaluate the dependency on the ontogeny of the plant and results may suggest that the chemical composition is relatively constant since the early stages of the plant, as the major characteristic compound (14) for *C. nictitans* ssp. *patellaria* was detected in in vitro plants in high concentration. These results suggest that LC–MS/MS analysis of crude extracts is capable of distinguishing amongst subspecies and varieties even when

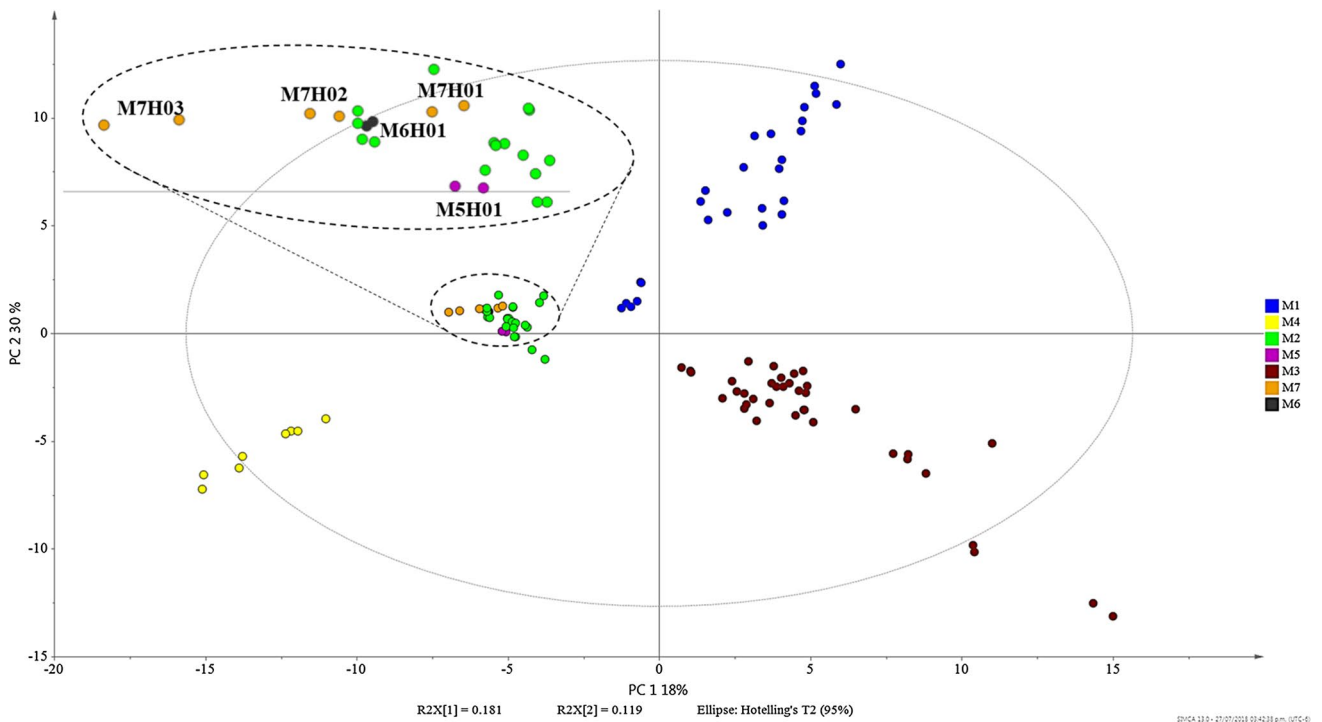
they lack of no further morphological characters, imperative in classical taxonomy.

Additional testing and evaluation were performed by analyzing 12 samples provided for the Herbarium of the University of Costa Rica (USJ, UCR) including seven samples classified no further than to the species level (*C. nictitans*) and five other species of the same genus. Scatter plot of the PLS-DA in Fig. 3 shows that samples of different species (M5, M6 and M7) were grouped within M2 (*C. nictitans* ssp. *disadena*) space, this is in accordance with the few detected compounds in all these samples compared to M1 and M4, as explain above for the case of the in vitro samples, and also that these different species mainly produce flavonoids glycosides similar to M2 and M4. The model was able to keep discriminating M1, M3 and M4 subspecies/varieties without losing resolution. Clearly, the model could eventually be refined upon chemical characterization of new variables from other subspecies, varieties or even different species tight related. An important aspect to highlight is the fact that preserved herbarium material with a minimum weight (ca. 200 mg or less, see figure S5) could be analyzed and classified without trouble (the oldest sample collected in 1949). Remaining samples (seven specimens), classified to the species level (*C. nictitans*) were include into the model and afforded the PLS-DA score plot shown in Fig. 4, where they were spread within the M1, M2 and M3 clusters, five



**Fig. 3** Partial least square—discriminant analysis score plot (C1 vs. C2) showing 51 samples including *C. nictitans* varieties and subspecies (n=46), three different *Chamaecrista* species (n=5, M5-M7), (M1 = *C. nictitans* var. *patellaria*; M2 = *C. nictitans* ssp. *disadana*;

M3 = *C. nictitans* ssp. *nictitans* var. *jaliscensis*; M4 = *C. nictitans* ssp. *disadana* var. *pilosa*, M5 = *C. rotundifolia*, M6 = *C. diphyllia*, M7 = *C. flexuosa* L.)



**Fig. 4** Partial least square—discriminant analysis score plot (C1 vs. C2) showing 53 samples including *C. nictitans* varieties and subspecies (n=46) and seven samples classified to *Chamaecrista nictitans* specie level from Herbarium. (M1 = *C. nictitans* var. *patellaria*;

M2 = *C. nictitans* ssp. *disadana*; M3 = *C. nictitans* ssp. *nictitans* var. *jaliscensis*; M4 = *C. nictitans* ssp. *disadana* var. *pilosa*., MX = Samples from herbarium)



of them were grouped into the *C. nictitans* ssp. *nictitans* var. *jaliscensis* subgroup (M3). It is highly noteworthy that two of the samples, which were collected in 1965 and in 1957 in the United States (MXH02 and MXH06 respectively, see Table S2) conserved their chemical fingerprint that allowed its placement in the model. This may suggest that the chemical profile and not necessarily concentration of the compounds is genetically stabilized and may not be strongly dependent of external factors.

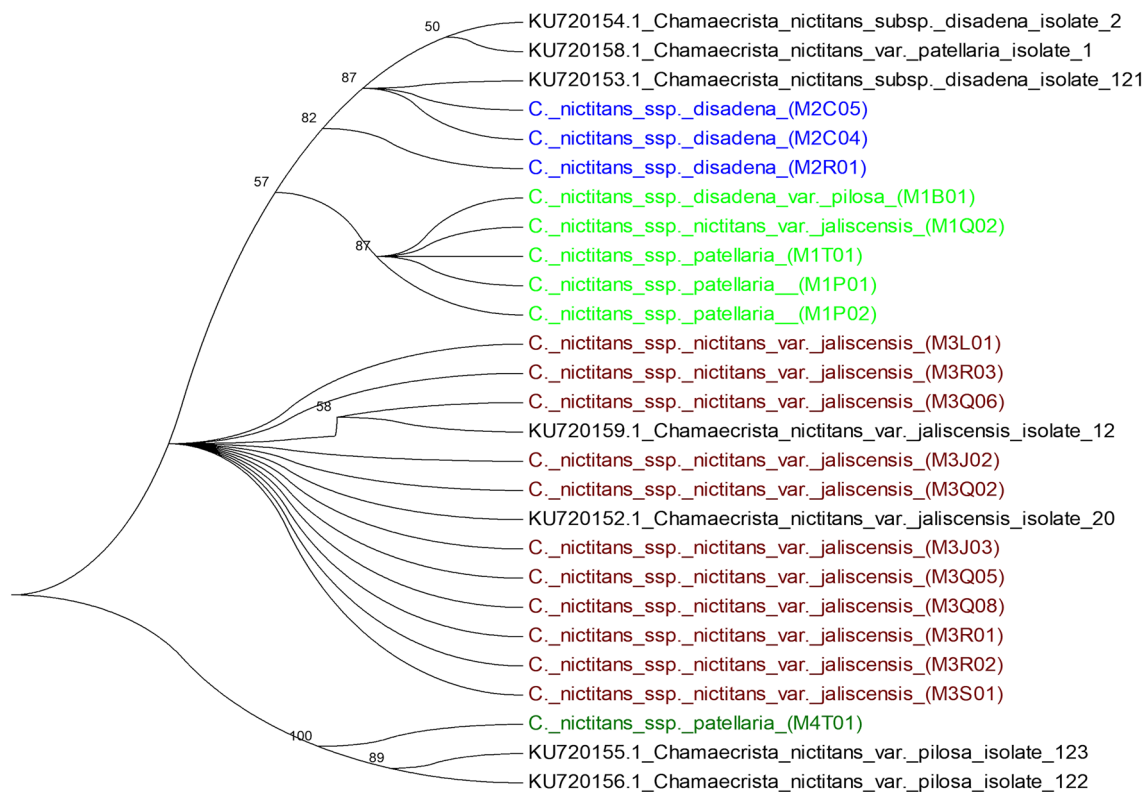
Just one sample (MXH03) provided by the herbarium was classified to variety level as *C. nictitans* ssp. *disadena* var. *pilosa*. However, our statistical analysis grouped it closed to the M2 cluster (*C. nictitans* ssp. *disadena*) (Fig. 4), which may indicate a taxonomical misplacement based on the morphological characteristics. Last sample, MXH04, was grouped into the M1 cluster (Fig. 5), indicating that this is a *C. nictitans* ssp. *patellaria*. Further examination of the chromatograms and mass spectra of these samples corroborated the observed grouping.

The highest discriminant ions with a VIP > 1 and a p-value < 0.05 are shown in the table S4. These ions correspond mainly to the proanthocyanidin oligomers present in all subspecies and varieties although others are also present, such as several flavonoid glycosides e.g. compound 21, present only in *C. nictitans* ssp. *disadena* var. *pilosa*.

### 3.2 Identification of chemical markers

Analysis of LC–MS and LC–MS/MS data for samples collected afforded the tentative identification of 44 polyphenols (Table 1). Putatively identified compounds correspond to seven B-type tetramers (1–7), eleven B-type trimers (8–14, 17–20), two A-type trimers (15–16), two B-type dimers (33, 38), several flavonoid glycosides and acetates (21–32, 33–37 and 39–43) and one flavanol (44). All molecular formulas were confirmed through accurate mass measurement and detected with an error less than 2.0 ppm, thus corroborating their elemental composition.

Compounds 14, 15, 18, 34, and 41 to 44 were previously reported by Mateos-Martin et al. (2014), while compounds 21 (Le Gall et al. 2003; Sakushima and Nishibe 1988), 27 (Alston et al. 1965; Herrmann 1988; Saleh et al. 1982; Sharaf et al. 1997), 33 (Del Bubba et al. 2012; Hatano et al. 2002; Lin et al. 2014; Onagas et al. 2007), and 38 (Ferreira and Li 2000; Sobeh et al. 2018) and 40 (Peng et al. 2015) were previously reported in the literature but not for these species. Accordingly, to the proposed structures based on their fragmentation pattern on MS/MS experiments, the remaining proanthocyanidin oligomers and flavonoid glycosides and acetates (1–13, 16, 17, 19, 20, 22–26, 28–32, 35–37, 39), to the best of our



**Fig. 5** Neighbor-joining tree constructed based on the sequence of the ITS2 region of subspecies and varieties of *Chamaecrista nictitans* collected in Costa Rica

**Table 1** Characterization of polyphenols found in all samples of *Chamaecrista nictitans* using UPLC-PDA-ESI-MS/MS-QToF in negative ionization mode

ID	Compound	RT (min)	MF	Experimental accurate $m/z$ [M-H] <sup>-</sup>	Error (ppm)	Product Ions MS/MS in negative mode (relative intensity, %)
1	(epi)fis-(epi)fis-(epi)cat-(epi)fis	6.95	C <sub>60</sub> H <sub>50</sub> O <sub>21</sub>	1105.2761	-0.5	953(30), 935(11), 833(54), 681(67), 543(70), 433(34), 391(75), 271(100), 161(55)
2	(epi)fis-(epi)cat-(epi)afz-(epi)gui	5.79	C <sub>60</sub> H <sub>50</sub> O <sub>20</sub>	1089.2799	-1.7	953(20), 833(40), 707(60), 681(30), 561(85), 435(40), 409(100), 272(95), 161(70)
3	(epi)gui-(epi)afz-(epi)fis-(epi)cat	6.01	C <sub>60</sub> H <sub>50</sub> O <sub>20</sub>	1089.2797	-1.8	937(15), 801(30), 691(10), 649(15), 529(25), 393(76), 377(31), 257(21), 137(15)
4	(epi)fis-(epi)fis-(epi)fis-(epi)fis	7.07	C <sub>60</sub> H <sub>50</sub> O <sub>20</sub>	1089.2838	1.9	953(1), 817(10), 665(15), 545(100), 435(7), 393(36), 273(73), 161(20)
5	(epi)gui-(epi)fis-(epi)fis-(epi)fis	7.40	C <sub>60</sub> H <sub>50</sub> O <sub>19</sub>	1073.2863	-0.5	921(21), 801(17), 691(23), 649(15), 529(100), 409(25), 375(41), 255(41), 137(9)
6	(epi)afz-(epi)afz-(epi)fis-(epi)gui	7.90	C <sub>60</sub> H <sub>50</sub> O <sub>19</sub>	1073.2868	0.0	937(3), 817(5), 707(2), 665(8), 545(100), 409(16), 393(12), 272(18), 137(3)
7	(epi)gui-(epi)cat-(epi)gui-(epi)gui	8.32	C <sub>60</sub> H <sub>50</sub> O <sub>18</sub>	1057.2911	-0.8	921(31), 801(30), 649(21), 545(100), 393(40), 289(30), 255(47), 137(14)
8	(epi)cat-(epi)fis-(epi)fis	4.36	C <sub>45</sub> H <sub>38</sub> O <sub>16</sub>	833.2071	1.2	681(25), 561(100), 409(33), 391(60), 289(87), 271(30), 137(13)
9	(epi)fis-(epi)cat-(epi)fis	5.79	C <sub>45</sub> H <sub>38</sub> O <sub>16</sub>	833.2083	-0.1	723(1), 681(2), 561(16), 409(46), 393(16), 289(75), 272(90), 137(10)
10	(epi)afz-(epi)cat-(epi)afz	6.92	C <sub>45</sub> H <sub>38</sub> O <sub>16</sub>	833.2088	0.7	707(3), 697(6), 561(30), 409(47), 393(13), 271(91), 137(15)
11	(epi)gui-(epi)cat-(epi)cat	7.01	C <sub>45</sub> H <sub>38</sub> O <sub>16</sub>	833.2078	-0.5	681(1), 545(18), 419(35), 401(8), 393(4), 255(82), 137(7)
12	(epi)fis-(epi)cat-(epi)gui	8.10	C <sub>45</sub> H <sub>38</sub> O <sub>15</sub>	817.2128	-0.5	681(2), 561(14), 545(10), 409(36), 271(55), 161(57)
13	(epi)gui-(epi)cat-(epi)fis	7.91	C <sub>45</sub> H <sub>38</sub> O <sub>15</sub>	817.2127	-0.6	665(2), 545(15), 393(71), 256(84), 137(4)
14	(epi)afz-(epi)fis-(epi)fis	6.84	C <sub>45</sub> H <sub>38</sub> O <sub>15</sub>	817.2125	0.2	707(9), 681(26), 665(51), 545(100), 435(28), 409(68), 393(29), 271(47), 137(24)
15	(epi)fis-(epi)fis-A-(epi)afz	6.98	C <sub>45</sub> H <sub>36</sub> O <sub>15</sub>	815.1985	1.1	690(44), 663(10), 543(77), 432(54), 378(43), 272(100), 161(96)
16	(epi)fis-A-(epi)fis-(epi)afz	6.13	C <sub>45</sub> H <sub>36</sub> O <sub>15</sub>	815.1976	0.0	689(23), 679(43), 663(15), 543(56), 433(40), 272(66)
17	(epi)gui-(epi)fis-(epi)fis	7.71	C <sub>45</sub> H <sub>38</sub> O <sub>14</sub>	801.2193	01.2	649(60), 529(100), 419(69), 377(28), 271(56), 255(37), 137(48)
18	(epi)fis-(epi)fis-(epi)gui	8.32	C <sub>45</sub> H <sub>38</sub> O <sub>14</sub>	801.2185	0.2	691(7), 665(20), 545(100), 393(50), 271(30), 161(28)
19	(epi)fis-(epi)afz-(epi)gui	7.47	C <sub>45</sub> H <sub>38</sub> O <sub>14</sub>	801.2176	-0.9	691(4), 665(60), 545(100), 411(18), 409(70), 271(56), 255(37), 161(53)
20	(epi)gui-(epi)cat-(epi)gui	9.06	C <sub>45</sub> H <sub>38</sub> O <sub>14</sub>	801.2188	0.6	691(2), 665(1), 545(12), 393(26), 256(100), 137(14)
21	Kaempferol-O-hexoside-O-rhamnosyl-hexoside	4.20	C <sub>33</sub> H <sub>40</sub> O <sub>20</sub>	755.2121	0.7	635(3), 593(100), 431(1), 285(35), 255(13), 227(3)
22	Quercetin-O-rhamnoside-O-rhamnosylpentoside	4.00	C <sub>32</sub> H <sub>38</sub> O <sub>19</sub>	725.1936	0.0	593(2), 463(14), 389(1), 301(9), 284(27), 255(7), 243(0.5), 151(2)
23	Rhamnetin-O-pentoside-hexoside-O-acetate	5.23	C <sub>29</sub> H <sub>32</sub> O <sub>17</sub>	651.1557	-0.6	609(5), 519(2), 314(100), 299(38), 285(3), 255(1), 151(3)
24	Kaempferol-O-rhamnosylhexoside-O-acetate	7.28	C <sub>29</sub> H <sub>32</sub> O <sub>16</sub>	635.1625	0.4	593(1), 431(2), 285(100), 255(5), 227(4)

**Table 1** (continued)

ID	Compound	RT (min)	MF	Experimental accurate $m/z$ [M-H] <sup>-</sup>	Error (ppm)	Product Ions MS/MS in negative mode (relative intensity, %)
25	Rhamnetin- <i>O</i> -hexosylrhamnoside	4.47	C <sub>28</sub> H <sub>32</sub> O <sub>16</sub>	623.1617	0.8	609(30), 447(5), 343(6), 315(35), 300(2), 299(22), 285(5), 151(1)
26	Kaempferol- <i>O</i> -(acetyl)hexoside-pentoside	5.00	C <sub>28</sub> H <sub>30</sub> O <sub>16</sub>	621.1459	0.5	561(8), 285(61), 255(22), 227(6)
27	Quercetin- <i>O</i> -rhamnosyl-(1→6)-hexoside	3.75	C <sub>27</sub> H <sub>30</sub> O <sub>16</sub>	609.1254	0.8	535(4), 505(6), 463(13), 403(5), 343(8), 329(10), 301(66), 285(100), 255(17), 179(9), 151(10)
28	Luteolin-6- <i>C</i> -hexosyl-(1→2)-rhamnoside	4.89	C <sub>27</sub> H <sub>30</sub> O <sub>15</sub>	593.1506	0.0	575(24), 557(1), 473(28), 411(100), 337(46), 327(10), 298(9), 285(17), 151(6), 133(4)
29	Luteolin-6- <i>C</i> -hexosyl-(1→2)-rhamnoside isomer	4.53	C <sub>27</sub> H <sub>30</sub> O <sub>15</sub>	593.1494	-1.2	575(56), 557(2), 473(28), 411(100), 337(38), 327(7), 298(13), 285(15), 151(4), 133(1)
30	Kaempferol-6- <i>C</i> -hexosyl-(1→2)-rhamnoside	2.60	C <sub>27</sub> H <sub>30</sub> O <sub>15</sub>	593.1511	0.8	473(38), 447(1), 429(10), 357(28), 327(60), 298(100), 285(2), 255(5)
31	Kaempferol-6- <i>C</i> -hexosyl-(1→2)-rhamnoside isomer	2.55	C <sub>27</sub> H <sub>30</sub> O <sub>15</sub>	593.1515	0.1	473(100), 447(4), 429(38), 357(31), 327(35), 298(25), 285(5), 255(1)
32	Apigenin-6- <i>C</i> -hexosyl-(1→2) rhamnoside	2.95	C <sub>27</sub> H <sub>30</sub> O <sub>14</sub>	577.1556	-0.2	431(2), 413(70), 311(18), 293(100), 283(3), 174(1)
33	(epi)cat-(epi)cat	3.40	C <sub>30</sub> H <sub>26</sub> O <sub>12</sub>	577.1342	-0.7	451(12), 425(75), 289(100), 161(72), 137(10)
34	Cassiaoccidentalinalin B	4.91	C <sub>27</sub> H <sub>28</sub> O <sub>14</sub>	575.1402	0.2	473(34), 471(2), 429(1), 411(48), 383(14), 337(81), 327(48), 309(45), 298(100), 285(70), 257(3)
35	Cassiaoccidentalinalin A	5.40	C <sub>27</sub> H <sub>28</sub> O <sub>13</sub>	559.1452	0.0	457(8), 395(37), 377(5), 367(15), 321(69), 310(25), 293(100), 282(17), 269(35), 151(3)
36	Cassiaoccidentalinalin A isomer	5.65	C <sub>27</sub> H <sub>28</sub> O <sub>13</sub>	559.1454	0.4	457(5), 395(18), 377(5), 367(3), 321(64), 310(29), 293(100), 282(28), 269(47), 151(4)
37	Cassiaoccidentalinalin A isomer	5.87	C <sub>27</sub> H <sub>28</sub> O <sub>13</sub>	559.1445	-0.7	457(4), 395(18), 377(5), 367(2), 321(58), 310(27), 293(100), 282(24), 269(38), 151(2)
38	(epi)gui-(epi)cat	5.15	C <sub>30</sub> H <sub>26</sub> O <sub>10</sub>	545.1445	-0.3	393(7), 289(8), 255(11), 137(36)
39	Rhamnetin- <i>O</i> -(acetyl)hexoside	5.58	C <sub>24</sub> H <sub>24</sub> O <sub>13</sub>	519.1140	0.2	459(0.6), 357(3), 314(100), 299(18), 285(50), 271(53), 257(16), 243(41), 151(2)
40	Quercetin- <i>O</i> -hexoside- <i>O</i> -acetate	4.82	C <sub>23</sub> H <sub>22</sub> O <sub>13</sub>	505.0995	0.3	463(4), 343(0.8), 300(100), 271(18), 255(8), 151(4)
41	Quercetin- <i>O</i> -hexoside	4.06	C <sub>21</sub> H <sub>20</sub> O <sub>12</sub>	463.0558	0.6	343(10), 301(100), 285(12), 255(8), 243(5), 151(6)
42	Luteolin- <i>C</i> -hexoside	4.76	C <sub>21</sub> H <sub>20</sub> O <sub>11</sub>	447.0939	0.2	429(15), 411(2), 357(90), 327(100), 298(13), 285(10)
43	Apigenin- <i>C</i> -hexoside	3.51	C <sub>21</sub> H <sub>20</sub> O <sub>10</sub>	437.1058	0.2	341(40), 311(100), 283(18), 269(5)
44	Fisetin	4.81	C <sub>15</sub> H <sub>10</sub> O <sub>6</sub>	285.0397	-0.7	239(3), 228(8), 162(6), 149(10), 135(100), 121(90)

(epi)afz, (epi)cat, (epi)gui and (epi)fis are abbreviations for (epi)afzelequin, (epi)catechin, (epi)guibourtinidol and (epi)fisetinidol respectively

knowledge, are reported for the first time. Moreover, compounds **29**, **31** and **36–37** seem to be regioisomers of **28**, **30** and **35** respectively, in consistency with the observed fragmentation patterns in MS/MS experiments.

All structures were proposed according to the above explanation (see Sect. 2.6). For example, compound **21** has a [M-H]<sup>-</sup> at  $m/z$  755 and was characterized as kaempferol-*O*-hexoside-*O*-rhamnosyl-hexoside in accordance with the

fragments detected. The mass spectra showed ions at  $m/z$  635 [(M-H)-120] ( $^{-0.2}X_2$ , characteristic of an hexose sugar moiety),  $m/z$  593 [(M-H)-162] ( $Y_2^-$ , loss of a hexose moiety),  $m/z$  431 [ $Y_2$ -162] ( $Y_1^-$ , loss of a hexose moiety),  $m/z$  285 ( $Y_0^-$ , aglycone, loss of a rhamnose moiety),  $m/z$  255 ( $Y_0^-$ -CH<sub>2</sub>O), 227 (loss of water on ion 255) and 151 ( $^{-1.3}A$ ).  $Y_0^-$  is produced from ion at  $m/z$  431 by loss of 146 Da, suggesting the presence of a rhamnose sugar. The relative intensities of ions  $m/z$  593 and  $m/z$  285 indicate the presence of the *O*-glycosidic linkage, also the presence of the ion  $m/z$  593 is an indicative that there are two different sugar moieties attached to the aglycone (figure S6).

The putatively identified compounds across different *C. nictitans* subspecies and varieties are shown in table S3 and detailed description of other compounds can be seen in supporting information.

### 3.3 DNA barcoding analysis

Twenty *C. nictitans* species (including subspecies and varieties) were selected for DNA barcoding analysis (Table S1) to validate the data obtained by LC-MS/MS. Samples were selected based on the chemo types determined from the identified metabolites data matrix. We obtained 100% efficiency of PCR amplification and sequencing with the use of a primer combination for ITS2 designed for DNA barcoding identification of medicinal plants (Chen et al. 2010; Chiou et al. 2007). We certainly had a relatively small sample set, nonetheless, this is not a surprising result considering that with the same primer combination efficiency of PCR amplification was 89,6% in a set of 992 samples (Chen et al. 2010) and 91% in a different set of 192 samples (Han et al. 2013). Ashfaq et al. (2013) reported 100% of sequence recovery when the ITS2F primer was used in cotton species.

The estimation of diversity parameters of our sequences joined with those retrieved from the GeneBank displayed larger inter-specific (0.113) than intra-specific (0.011) diversity. Gao et al. (2010) also described larger inter-specific variation in the Fabaceae using ITS2 region. The sequence diversity of ITS2 (figure S7) allowed the separation of *C. nictitans* subspecies and varieties in the present study. An in silico study showed that ITS2 inter-specific divergence of congeneric species was greater than intra-specific variation which allow a high rate of correct identification of closely related species into several dicotyledons families, including the Fabaceae family (Yao et al. 2010).

### 3.4 Comparison of morphological identification with DNA barcoding

When the morphological classification of samples was considered as reference, the comparison with the DNA barcoding (ITS2 sequence) revealed that 17 out of 20 samples

grouped accordingly (Neighbor-joining tree in Fig. 5). So that, three discrepancies were observed in the phylogenetic tree. Samples M1B01 (*C. nictitans* subsp. *disadena* var. *pilosa*) and M1Q02 (*C. nictitans* subsp. *nictitans* var. *jaliscensis*) were clustered into the group of *C. nictitans* subsp. *patellaria*. Sample M4T01 (identified as *C. nictitans* subsp. *disadena* based on morphology) grouped with two of the available Genebank accessions (KU720155 and KU720156) described as *C. nictitans* var. *pilosa*. For the retrieved sequences from Genebank, only accession KU720158 (*C. nictitans* ssp. *patellaria*) did not grouped accordingly and it was located with accessions and samples of *C. nictitans* ssp. *disadena* (Fig. S6). Thus, DNA barcoding analysis suggest that samples M1B1, M1Q02 and M4T01 are morphologically misidentified and belong to the group suggested in the phylogenetic tree. The classification of a morphotype with shared characteristics of *C. cultrifolia* and *C. diphylla* allowed the re-establishment of the samples as *C. cultrifolia*, based on molecular and morphological comparisons (Barrbosa et al. 2016).

### 3.5 Correlation of chemical profile and genetic data

When chemical classification of samples was used as reference, the subset of 20 samples grouped accordingly. As shown in figure S4 (Hierarchical Cluster Analysis) the chemical profile revealed a clear separation of groups M1, M2, M3 and M4. Similarly, the sequence of the ITS2 region revealed a separation of the same groups (Fig. S6). However, the pattern of separation in the phylogenetic tree had some differences with the chemical profile. Groups M2 and M4 in figure S4 derived from the same branch but with the ITS2 region M4 (*C. nictitans* ssp. *disadena* var. *pilosa*) were widely separated which is explained by large differences between sequences (Figure S6).

Differences between sequences of ITS2 (figure S6) were reflected in larger mean distances between groups M1, M2 and M3 ( $d=0.471$ ,  $0.493$  and  $0.454$  respectively). In this case, for DNA barcoding analysis we had only one sample of M4. To estimate the mean distance for this group, two references of Genebank (KU720155 and KU720156) were included in the estimation of distances. Besides, the groups M1 (*C. nictitans* ssp. *patellaria*), M2 (*C. nictitans* ssp. *disadena*) and M3 (*C. nictitans* ssp. *nictitans* var. *jaliscensis*) derived from the same branch when ITS2 region was used. This branch was divided in two subgroups: M1 y M2 clustered with a clear separation of both subspecies (bootstrap values > 80). The other subgroup revealed the similarity of the samples of *C. nictitans* ssp. *nictitans* var. *jaliscensis*. This pattern of groups separation was based on mean distances. The least distance ( $d=0.0140$ ) was between M1 and M3. The M2 group showed similar distances from M1 to M3 ( $d=0.024$  and  $0.021$  respectively).

A similar approach of chemotaxonomy and genetic analysis has been described for plants of genus such as *Lespedeza* (Kim et al. 2012), *Rhodiola* (Liu et al. 2013) and *Zingiber* (Jiang et al. 2006). Similar to our results, Jiang et al. (2006) found that the chemical characters of the investigated species of *Zingiber* were able to generate essentially the same phylogenetic relationships as the DNA sequences but using trnL and rps16 sequences. Kim et al. (2012) used a combination of ITS and cpDNA (trnL–trnF) for classifying genotypes of *Lespedeza sp.* They found that in both the genetic and chemotaxonomic classification methods, the distance between species *L. cyrtobotrya* and *L. bicolor* was the closest between species and *L. cuneata* was the farthest away from the other three species.

Our results are also very similar to the study in *Rhodiola*. Liu et al. (2013) used a combination of morphological characteristics, genetic analysis with sequences obtained from cDNA and phytochemical analysis. Samples of *Rhodiola* were accordingly classified by genetic taxonomy and with four types of bioactive compound as reference markers. Liu et al. (2013) described that HCA results showed considerably comparable results for both the geno type- and chemo type-based classification methods, as also described in our results.

## 4 Conclusions

Comparative analysis of metabolic features of plants has a high potential for multiple purposes including chemotaxonomy. Here we describe a multivariate model based on LC–MS/MS fingerprinting capable of discriminating between subspecies and varieties of the medicinal plant *C. nictitans*. Results demonstrate that our metabolomic approach was able to discriminate between the subspecies and varieties of this plant, separating the samples into four differentiated groups: M1 = *C. nictitans* var. *patellaria*; M2 = *C. nictitans* ssp. *disadena*; M3 = *C. nictitans* ssp. *nic-titans* var. *jaliscensis* and M4 = *C. nictitans* ssp. *disadena* var. *pilosa*. Chemical LC–MS/MS fingerprint results were confirmed using both morphological characters and DNA barcoding with ITS2 region. In conclusion, the metabolomic approach described in this study allows an efficient and reliably differentiation between subspecies and varieties of *C. nictitans* using a straightforward protocol that lacks extensive purification steps.

**Acknowledgements** The authors wish to thank Vicerrectoría de Investigación of the University of Costa Rica, for their financial support (project number 809-B3-082) and the Herbarium of the University of Costa Rica (USJ) for supplying samples from the collection.

**Author contributions** LQG, EAV, MC, RMR and GTC wrote the paper; LQG designed and did the LC–MS/MS experiments and MS

data analysis; LQG, FA, LP and GTC collected the wild plant material; LP did the botanical identification; RMR did the in vitro culturing of wild plant material and provided greenhouse plantlets for the experiments, FA, HV, EAV and MC performed DNA extraction and ITS2 amplification and sequence analysis, LQG did the data mining and statistical analysis of MS data. GTC, FA and MC did the experimental design of the research.

## Compliance with ethical standards

**Conflict of interest** There is no conflict of interest.

**Ethical statements** This article does not contain any studies with human and/or animal participants performed by any of the authors.

## References

- Alston, R. E., Rösler, H., Naifeh, K., & Mabry, T. J. (1965). Hybrid compounds in natural interspecific hybrids. *Proceedings of the National Academy of Sciences of the United States of America*, 54(5), 1458–1465.
- Ashfaq, M., Asif, M., Anjum, Z. I., & Zafar, Y. (2013). Evaluating the capacity of plant DNA barcodes to discriminate species of cotton (*Gossypium*: Malvaceae). *Molecular Ecology Resources*, 13(4), 573–582. <https://doi.org/10.1111/1755-0998.12089>.
- Barbosa, A. R., Machado, M. C., Lewis, G. P., Forest, F., & De Queiroz, L. P. (2016). Re-establishment of *Chamaecrista cultrifolia* (Leguminosae, Caesalpinioideae) based on morphological and molecular analyses. *Phytotaxa*, 265(3), 183–203. <https://doi.org/10.11646/phytotaxa.265.3.1>.
- Benayad, Z., Gómez-Cordovés, C., & Es-Safi, N. (2014). Characterization of flavonoid glycosides from fenugreek (*Trigonella foenum-graecum*) crude seeds by HPLC–DAD–ESI/MS analysis. *International Journal of Molecular Sciences*, 15(11), 20668–20685. <https://doi.org/10.3390/ijms151120668>.
- Calderón, A. I., Wright, B. J., Hurst, W. J., & van Breemen, R. B. (2009). Screening antioxidants using LC-MS: case study with cocoa. *Journal of Agricultural and Food Chemistry*, 57(13), 5693–5699. <https://doi.org/10.1021/jf9014203>.
- Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., et al. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE*, 5(1), 1–8. <https://doi.org/10.1371/journal.pone.0008613>.
- Chiou, S. J., Yen, J. H., Fang, C. L., Chen, H. L., & Lin, T. Y. (2007). Authentication of medicinal herbs using PCR-amplified ITS2 with specific primers. *Planta Medica*, 73(13), 1421–1426. <https://doi.org/10.1055/s-2007-990227>.
- Clark, B. R., Bliss, B. J., Suzuki, J. Y., & Borris, R. P. (2014). Chemotaxonomy of Hawaiian Anthurium cultivars based on multivariate analysis of phenolic metabolites. *Journal of Agricultural and Food Chemistry*, 62, 11323–11334. <https://doi.org/10.1021/jf502187c>.
- Consonni, R., Cagliani, L. R., & Cogliati, C. (2012). NMR based geographical characterization of roasted coffee. *Talanta*, 88, 420–426. <https://doi.org/10.1016/j.talanta.2011.11.010>.
- Del Bubba, M., Checchini, L., Chiuminatto, U., Doumet, S., Fibbi, D., & Giordani, E. (2012). Liquid chromatographic/electrospray ionization tandem mass spectrometric study of polyphenolic composition of four cultivars of *Fragaria vesca* L. berries and their comparative evaluation. *Journal of Mass Spectrometry*, 47(9), 1207–1220. <https://doi.org/10.1002/jms.3030>.

- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792–1797. <https://doi.org/10.1093/nar/gkh340>.
- Farag, M. A., Maamoun, A. A., Ehrlich, A., Fahmy, S., & Wesjohann, L. A. (2017). Assessment of sensory metabolites distribution in 3 cactus *Opuntia ficus-indica* fruit cultivars using UV fingerprinting and GC/MS profiling techniques. *LWT-Food Science and Technology*, 80, 145–154. <https://doi.org/10.1016/j.lwt.2017.02.014>.
- Farias, LdaS., & Mendez, A. (2014). LC/ESI-MS Method applied to characterization of flavonoids glycosides in *B. forficata* subsp. pruinosa. *Quimica Nova*, 37(3), 483–486. <https://doi.org/10.5935/0100-4042.20140069>.
- Ferreira, D., & Li, X. (2000). Oligomeric proanthocyanidins: Naturally occurring O-heterocycles. *Natural Product Reports*, 17, 193–212. <https://doi.org/10.1039/a705728h>.
- Friedrich, W., Eberhardt, a, & Galensa, R. (2000). Investigation of proanthocyanidins by HPLC with electrospray ionization mass spectrometry. *European Food Research and Technology*, 211(1), 56–64. <https://doi.org/10.1007/s002170050589>.
- Gao, T., Yao, H., Song, J., Liu, C., Zhu, Y., Ma, X., et al. (2010). Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2. *Journal of Ethnopharmacology*, 130(1), 116–121. <https://doi.org/10.1016/j.jep.2010.04.026>.
- Glauser, G., Boccard, J., Wolfender, J.-L., & Rudaz, S. (2013). Metabolomics: Application in plant sciences. In M. Lämmerhofer & W. Weckwerth (Eds.), *Metabolomics in practice: Successful strategies to generate and analyze metabolic data* (First., pp. 313–343). Hoboken: Wiley. <https://doi.org/10.1002/9783527655861.ch13>.
- Godfray, H. C. J., & Knapp, S. (2004). Taxonomy for the 21th century. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359(1444), 559–569. <https://doi.org/10.1098/rstb.2003.1457>.
- Govindaraghavan, S., Hennell, J. R., & Sucher, N. J. (2012). From classical taxonomy to genome and metabolome: Towards comprehensive quality standards for medicinal herb raw materials and extracts. *Fitoterapia*, 83(6), 979–988. <https://doi.org/10.1016/j.fitote.2012.05.001>.
- Güzel, Y., Aktoklu, E., Roumy, V., Alkhatib, R., Hennebelle, T., Bailleul, F., & Şahpaz, S. (2011). Chemotaxonomy and flavonoid profiling of *Torilis* species by HPLC/ESI/MS2. *Biochemical Systematics and Ecology*, 39(4–6), 781–786. <https://doi.org/10.1016/j.bse.2011.07.012>.
- Hall, T. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, 95–98.
- Hammel, B. E., Grayum, M. H., Herrera, C., Zamora, N., & Troyo, S. (2010). *Manual de plantas de Costa Rica, Volumen VI: Dicotiledoneas (Clusiaceae-Gunneraceae)*. Missouri: Monographs in Systematic Botany from Missouri Botanical Garden Press.
- Han, J., Zhu, Y., Chen, X., Liao, B., Yao, H., Song, J., et al. (2013). The short ITS2 sequence serves as an efficient taxonomic sequence tag in comparison with the full-length ITS. *BioMed Research International*, 2013, 3–10. <https://doi.org/10.1155/2013/741476>.
- Hao, D. C., Gu, X.-J., & Xiao, P. G. (2015). Chemotaxonomy: A phylogeny-bases approach. *Medicinal Plants*. <https://doi.org/10.1016/B978-0-08-100085-4.00001-3>.
- Hatano, T., Miyatake, H., Natsume, M., Osakabe, N., Takizawa, T., Ito, H., & Yoshida, T. (2002). Proanthocyanidin glycosides and related polyphenols from cacao liquor and their antioxidant effects. *Phytochemistry*, 59(7), 749–758. [https://doi.org/10.1016/S0031-9422\(02\)00051-1](https://doi.org/10.1016/S0031-9422(02)00051-1).
- Hegnauer, R. (1986). Phytochemistry and plant taxonomy—An essay on the chemotaxonomy of higher plants. *Phytochemistry*, 25(7), 1519–1535. [https://doi.org/10.1016/S0031-9422\(00\)81204-2](https://doi.org/10.1016/S0031-9422(00)81204-2).
- Herrero Uribe, L., Chaves Olarte, E., & Tamayo Castillo, G. (2004). In vitro antiviral activity of *Chamaecrista nictitans* (Fabaceae) against herpes simplex virus: Biological characterization of mechanisms of action. *Revista de biología tropical*, 52(3), 807–816.
- Herrmann, K. (1988). On the occurrence of flavonol and flavone glycosides in vegetables. *Zeitschrift für Lebensmittel-Untersuchung und -Forschung*, 186(1), 1–5. <https://doi.org/10.1007/BF01027170>.
- Hummer, K., Durst, R., Zee, F., Atnip, A., & Giusti, M. M. (2014). Phytochemicals in fruits of Hawaiian wild cranberry relatives. *Journal of the Science of Food and Agriculture*, 94(8), 1530–1536. <https://doi.org/10.1002/jsfa.6453>.
- Jaiswal, R., Kuhnert, N., Jayasinghe, L., & Kuhnert, N. (2012). Identification and characterization of proanthocyanidins of 16 members of the *Rhododendron* genus (Ericaceae) by tandem LC-MS. *Journal of Mass Spectrometry: JMS*, 47(12), 502–515. <https://doi.org/10.1002/jms.541>.
- Jiang, H., Xie, Z., Koo, H. J., McLaughlin, S. P., Timmermann, B. N., & Gang, D. R. (2006). Metabolic profiling and phylogenetic analysis of medicinal Zingiber species: Tools for authentication of ginger (*Zingiber officinale* Rosc.). *Phytochemistry*, 67(15), 1673–1685. <https://doi.org/10.1016/j.phytochem.2005.08.001>.
- Kashif, A., Federica, M., Eva, Z., Martina, R., Young, H. C., & Robert, V. (2009). NMR metabolic fingerprinting based identification of grapevine metabolites associated with downy mildew resistance. *Journal of Agricultural and Food Chemistry*, 57(20), 9599–9606. <https://doi.org/10.1021/jf902069f>.
- Kharazian, N. (2014). Chemotaxonomy and flavonoid diversity of *Salvia* L. (Lamiaceae) in Iran. *Acta Botanica Brasílica*, 28(2), 281–292. <https://doi.org/10.1590/S0102-33062014000200015>.
- Kim, J., Jung, Y., Song, B., Bong, Y.-S., Ryu, D. H., Lee, K.-S., & Hwang, G.-S. (2013). Discrimination of cabbage (*Brassica rapa* ssp. *pekinensis*) cultivars grown in different geographical areas using <sup>1</sup>H NMR-based metabolomics. *Food Chemistry*, 137(1–4), 68–75. <https://doi.org/10.1016/j.foodchem.2012.10.012>.
- Kim, Y. M., Lee, J., Park, S. H., Lee, C., Lee, J. W., Lee, D. H., et al. (2012). LC-MS-based chemotaxonomic classification of wild-type *Lespedeza* sp. and its correlation with genotype. *Plant Cell Reports*, 31(11), 2085–2097. <https://doi.org/10.1007/s00299-012-1319-8>.
- Klockmann, S., Reiner, E., Bachmann, R., Hackl, T., & Fischer, M. (2016). Food fingerprinting: Metabolomic approaches for geographical origin discrimination of hazelnuts (*Corylus avellana*) by UPLC-QTOF-MS. *Journal of Agricultural and Food Chemistry*, 64(48), 9253–9262. <https://doi.org/10.1021/acs.jafc.6b04433>.
- Le Gall, G., Dupont, M. S., Mellon, F. A., Davis, A. L., Collins, G. J., Verhoeven, M. E., & Colquhoun, I. J. (2003). Characterization and content of flavonoid glycosides in genetically modified tomato (*Lycopersicon esculentum*) fruits. *Journal of Agricultural and Food Chemistry*, 51(9), 2438–2446. <https://doi.org/10.1021/jf025995e>.
- Lee, E. J., Rustem, S., Weljie, A. M., Vogel, H. J., Facchini, P. J., Park, S. U., et al. (2009). Quality assessment of ginseng by <sup>1</sup>H NMR metabolite fingerprinting and profiling analysis. *Journal of Agricultural and Food Chemistry*, 57(16), 7513–7522. <https://doi.org/10.1021/jf901675y>.
- Lin, L., Sun, J., Chen, P., Monagas, M. J., & Harnly, J. M. (2014). UHPLC-PDA-ESI/HRMSn profiling method to identify and quantify oligomeric proanthocyanidins in plant products. *Journal of Agricultural and Food Chemistry*, 62, 9387–9400.
- Liu, Z., Liu, Y., Liu, C., Song, Z., Li, Q., Zha, Q., et al. (2013). The chemotaxonomic classification of *Rhodiola* plants and its correlation with morphological characteristics and genetic taxonomy. *Chemistry Central Journal*, 7(1), 118. <https://doi.org/10.1186/1752-153X-7-118>.
- Liu, Z., Wang, D., Li, D., & Zhang, S. (2017). Quality evaluation of *Juniperus rigida* Sieb. et Zucc. based on phenolic profiles, bioactivity, and HPLC fingerprint combined with chemometrics.

- Frontiers in Pharmacology*, 8(APR), 1–13. <https://doi.org/10.3389/fphar.2017.00198>.
- Luthria, D. L., Lin, L. Z., Robbins, R. J., Finley, J. W., Banuelos, G. S., & Harnly, J. M. (2008). Discriminating between cultivars and treatments of broccoli using mass spectral fingerprinting and analysis of variance-principal component analysis. *Journal of Agricultural and Food Chemistry*, 56(21), 9819–9827. <https://doi.org/10.1021/jf801606x>.
- Magagna, F., Guglielmetti, A., Liberto, E., Reichenbach, S. E., Allegrucci, E., Gobino, G., et al. (2017). Comprehensive chemical fingerprinting of high-quality cocoa at early stages of processing: Effectiveness of combined untargeted and targeted approaches for classification and discrimination. *Journal of Agricultural and Food Chemistry*. <https://doi.org/10.1021/acs.jafc.7b02167>.
- March, R. E., Lewars, E. G., Stadey, C. J., Miao, X.-S., Zhao, X., & Metcalfe, C. D. (2006). A comparison of flavonoid glycosides by electrospray tandem mass spectrometry. *International Journal of Mass Spectrometry*, 248(1–2), 61–85. <https://doi.org/10.1016/j.ijms.2005.09.011>.
- March, R. E., Miao, X.-S., & Metcalfe, C. D. (2004). A fragmentation study of a flavone triglycoside, kaempferol-3-O-robinoside-7-O-rhamnoside. *Rapid Communications in Mass Spectrometry: RCM*, 18(9), 931–934. <https://doi.org/10.1002/rcm.1428>.
- Mateos-Martín, M. L., Fuguet, E., Jiménez-Ardón, A., Herrero-Urbe, L., Tamayo-Castillo, G., & Torres, J. L. (2014). Identification of polyphenols from antiviral *Chamaecrista nictitans* extract using high-resolution LC-ESI-MS/MS. *Analytical and Bioanalytical Chemistry*. <https://doi.org/10.1007/s00216-014-7982-6>.
- Mattoli, L., Cangi, F., Maidecchi, A., Ghiara, C., Raggazi, E., Tubaro, M., et al. (2006). Metabolic fingerprinting of plant extracts. *Journal of Mass Spectrometry: JMS*, 41, 1534–1545. <https://doi.org/10.1002/jms>.
- Mayorga-Gross, A. L., Quirós-Guerrero, L. M., Fourny, G., & Vaillant, F. (2016). An untargeted metabolomic assessment of cocoa beans during fermentation. *Food Research International*, 89, 901–909. <https://doi.org/10.1016/j.foodres.2016.04.017>.
- Mehari, B., Redi-Abshiro, M., Chandravanshi, B. S., Combrinck, S., Atlabachew, M., & McCrindle, R. (2016). Profiling of phenolic compounds using UPLC-MS for determining the geographical origin of green coffee beans from Ethiopia. *Journal of Food Composition and Analysis*, 45, 16–25. <https://doi.org/10.1016/j.jfca.2015.09.006>.
- Murashige, T., & Skoog, F. (1962). A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiologia Plantarum*, 15(3), 473–497. <https://doi.org/10.1111/j.1399-3054.1962.tb08052.x>.
- Musah, R. A., Espinoza, E. O., Cody, R. B., Lesiak, A. D., Christensen, E. D., Moore, H. E., et al. (2015). A high throughput ambient mass spectrometric approach to species identification and classification from chemical fingerprint signatures. *Nature Publishing Group*, 5(May), 1–16. <https://doi.org/10.1038/srep11520>.
- Onagas, M. A. M., Arrido, I. G. G., Guilar, R. O. S. A. L. E., Artolome, B. E. B., & Ordovés, C. A. G. Ó. (2007). Almond (*Prunus dulcis* (Mill.) D. A. Webb) skins as a potential source of bioactive polyphenols. *Journal of Agricultural and Food Chemistry*, 55, 8498–8507.
- Pan, Y., Zhang, J., Li, H., Wang, Y. Z., & Li, W. Y. (2016). Characteristic fingerprinting based on macamides for discrimination of maca (*Lepidium meyenii*) by LC/MS/MS and multivariate statistical analysis. *Journal of the Science of Food and Agriculture*. <https://doi.org/10.1002/jsfa.7660>.
- Peng, F., Cheng, C., Xie, Y., & Yang, Y. (2015). Optimization of microwave-assisted extraction of phenolic compounds from “Anli” pear (*Pyrus ussuriensis* Maxim). *Food Science and Technology Research*, 21(3), 463–471. <https://doi.org/10.3136/fstr.21.463>.
- Pluskal, T., Castillo, S., Villar-Briones, A., & Oresic, M. (2010). MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*, 11, 395. <https://doi.org/10.1186/1471-2105-11-395>.
- Pongsuwan, W., Fukusaki, E., Bamba, T., Yonetani, T., Yamahara, T., & Kobayashi, A. (2007). Prediction of Japanese green tea ranking by gas chromatography/mass spectrometry-based hydrophilic metabolite fingerprinting. *Journal of Agricultural and Food Chemistry*, 55(2), 231–236. <https://doi.org/10.1021/jf062330u>.
- Ronningen, I., Miller, M., Xia, Y., & Peterson, D. G. (2017). Identification and validation of sensory-active compounds from data-driven research: A flavoromics approach. *Journal of Agricultural and Food Chemistry*. <https://doi.org/10.1021/acs.jafc.7b00093>.
- Ryan, D., & Robards, K. (2006). Metabolomics: The greatest omics of them all? *Analytical Chemistry*, 78(23), 7954–7958. <https://doi.org/10.1021/ac0614341>.
- Sakushima, A., & Nishibe, S. (1988). Mass spectrometry in the structural determination flavonol triglycosides from VLNCA major. *Phytochemistry*, 27(3), 915–919. [https://doi.org/10.1016/0031-9422\(88\)84119-0](https://doi.org/10.1016/0031-9422(88)84119-0).
- Saleh, N. A. M., Ahmed, A. A., & Abdalla, M. F. (1982). Flavonoid glycosides of *Tribulus pentandrus* and *T. terrestris*. *Phytochemistry*, 21(8), 1995–2000. [https://doi.org/10.1016/0031-9422\(82\)83030-6](https://doi.org/10.1016/0031-9422(82)83030-6).
- Salvador, A. C., Rudnitskaya, A., Silvestre, A. J. D., & Rocha, S. M. (2016). Metabolomic-based strategy for fingerprinting of *Sambucus nigra* L. berry volatile terpenoids and norisoprenoids: Influence of ripening and cultivar. *Journal of Agricultural and Food Chemistry*, 64(26), 5428–5438. <https://doi.org/10.1021/acs.jafc.6b00984>.
- Sarker, S. D., & Nahar, L. (2012). *Chapter 12 Hyphenated techniques and their applications in natural products analysis* (Vol. 864). New York: Springer. <https://doi.org/10.1007/978-1-61779-624-1>.
- Semmar, N., Jay, M., & Nouira, S. (2007). A new approach to graphical and numerical analysis of links between plant chemotaxonomy and secondary metabolism from HPLC data smoothed by a simplex mixture design. *Chemoecology*, 17(3), 139–155. <https://doi.org/10.1007/s00049-007-0374-z>.
- Sharaf, M., El-Ansari, M. A., & Saleh, N. A. M. (1997). Flavonoids of four Cleome and three Capparis species. *Biochemical Systematics and Ecology*, 25(2), 161–166. [https://doi.org/10.1016/S0305-1978\(96\)00099-3](https://doi.org/10.1016/S0305-1978(96)00099-3).
- Sobeh, M., Mahmoud, M. F., Abdelfattah, M. A. O., Cheng, H., El-Shazly, A. M., & Wink, M. (2018). A proanthocyanidin-rich extract from *Cassia abbreviata* exhibits antioxidant and hepatoprotective activities in vivo. *Journal of Ethnopharmacology*, 213(May 2017), 38–47. <https://doi.org/10.1016/j.jep.2017.11.007>.
- Stobiecki, M. (2000). Application of mass spectrometry for identification and structural studies of flavonoid glycosides. *Phytochemistry*, 54(3), 237–256. [https://doi.org/10.1016/S0031-9422\(00\)00091-1](https://doi.org/10.1016/S0031-9422(00)00091-1).
- Stobiecki, M., Kachlicki, P., & Wojakowska, A. (2015). Application of LC/MS systems to structural characterization of flavonoid glycoconjugates. *Phytochemistry Letters*, 11, 358–367. <https://doi.org/10.1016/j.phytol.2014.10.018>.
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30(12), 2725–2729. <https://doi.org/10.1093/molbev/mst197>.
- Treviño, V., Yañez-Garza, I., Rodríguez-López, C. E., Urrea-López, R., Garza-Rodríguez, M.-L., Tamez-Peña, J. G., et al. (2015). GridMass: A fast two-dimensional feature detection method for LC/MS. *Journal of Mass Spectrometry*. <https://doi.org/10.1002/jms.3512>.

- Tropicos. (2017). <http://www.tropicos.org/home.aspx>. Accessed 20 July 2007.
- Venkataraman, K. (1972). Wood Phenolics in the chemotaxonomy of the Moraceae. *Phytochemistry*, *11*(1964), 1571–1586. [https://doi.org/10.1016/0031-9422\(72\)85002-7](https://doi.org/10.1016/0031-9422(72)85002-7).
- Visnevschi-Necrasov, T., Barreira, J. C. M., Cunha, S. C., Pereira, G., Nunes, E., & Oliveira, M. B. P. P. (2015). Phylogenetic insights on the isoflavone profile variations in *Fabaceae* spp.: Assessment through PCA and LDA. *Food Research International*, *76*, 51–57. <https://doi.org/10.1016/j.foodres.2014.11.032>.
- Walker, J. M. (2014). *Methods in molecular biology 1115: Molecular Plant Taxonomy*. (P. Besse, Ed.). Clifton, NJ: Humana Press. <https://doi.org/10.1007/978-1-62703-767-9>.
- Watson, D. G., Peyfoon, E., Zheng, L., Lu, D., Seidel, V., Johnston, B., et al. (2006). Application of principal components analysis to <sup>1</sup>H-NMR data obtained from propolis samples of different geographical origin. *Phytochemical Analysis*, *17*(5), 323–331. <https://doi.org/10.1002/pca.921>.
- Williams, D., Stone, M., Hauck, P., & Rahman, S. (1989). Why are secondary metabolites (natural products) biosynthesized? *Journal of Natural Products*, *52*(6), 1189–1208. <https://doi.org/10.1021/np50066a001>.
- Wink, M. (2003). Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry*, *64*(1), 3–19. [https://doi.org/10.1016/S0031-9422\(03\)00300-5](https://doi.org/10.1016/S0031-9422(03)00300-5).
- Wink, M. (2010). *Biochemistry of plant secondary metabolism*. Chichester: Wiley. <https://doi.org/10.1002/9781444320503>.
- Wink, M. (2013). Evolution of secondary metabolites in legumes (Fabaceae). *South African Journal of Botany*, *89*, 164–175. <https://doi.org/10.1016/j.sajb.2013.06.006>.
- Yao, H., Song, J., Liu, C., Luo, K., Han, J., Li, Y., et al. (2010). Use of ITS2 region as the universal DNA barcode for plants and animals. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0013102>.
- Zieliński, Ł., Deja, S., Jasicka-Misiak, I., & Kafarski, P. (2014). Chemometrics as a tool of origin determination of polish monofloral and multifloral honeys. *Journal of Agricultural and Food Chemistry*, *62*(13), 2973–2981. <https://doi.org/10.1021/jf4056715>.

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