Strathprints Institutional Repository

SCIENTIFIC REPORTS

OPEN

Effect of the environment on the secondary metabolic profile of *Tithonia diversifolia*: a model for environmental metabolomics of plants

Bruno Leite Sampaio[1,2,*], RuAngelie Edrada-Ebel[2,*] & Fernando Batista Da Costa[1,*]

Tithonia diversifolia is an invasive weed commonly found in tropical ecosystems. In this work, we investigate the influence of different abiotic environmental factors on the plant's metabolite profile by multivariate statistical analyses of spectral data deduced by UHPLC-DAD-ESI-HRMS and NMR methods. Different plant part samples of *T. diversifolia* which included leaves, stems, roots, and inflorescences were collected from two Brazilian states throughout a 24-month period, along with the corresponding monthly environmental data. A metabolomic approach employing concatenated LC-MS and NMR data was utilised for the first time to study the relationships between environment and plant metabolism. A seasonal pattern was observed for the occurrence of metabolites that included sugars, sesquiterpenes lactones and phenolics in the leaf and stem parts, which can be correlated to the amount of rainfall and changes in temperature. The distribution of the metabolites in the inflorescence and root parts were mainly affected by variation of some soil nutrients such as Ca, Mg, P, K and Cu. We highlight the environment-metabolism relationship for *T. diversifolia* and the combined analytical approach to obtain reliable data that contributed to a holistic understanding of the influence of abiotic environmental factors on the production of metabolites in various plant parts.

Specimens of same plant species growing under different environmental conditions show significant differences in the production and accumulation of the primary and secondary metabolites[1–8]. Influenced by environmental factors, respective group of secondary metabolites act as a chemical interface between the plant and its environment. The chemical interaction between plants and their environment is mediated mainly by the biosynthesis of secondary metabolites, which exert their biological roles, as a plastic adaptive response to their environment. Such chemical interaction often includes variations in the production of plant metabolites[1–3,9–12]. Therefore, the study of these variations is very useful in the chemical characterization of plants of the same species which are collected from different regions and this is when the different geographical origin of a plant material is taken into account[13–15]. Some processes can be the main sources of variation in the levels of metabolites for individual plant species. These processes involve long term acclimation or local adaptation, seasonal differences related to phenology or environmental changes in the biotic and abiotic factors, geographical differences involving different populations (genetic differences within a plant species), or different environmental conditions of the growth location of the species individuals, especially when they have genetic homogeneity (i.e. cultivars and/or clones)[16–17].

Metabolomics is an interesting approach that has been used in the plant sciences, especially in ecological studies, in investigating the effects of environmental factors on plant metabolism. Metabolomic profile data have been utilised to compare different species from the same family, or individuals from populations from the same species

---

[1]AsterBioChem Team, Laboratory of Pharmacognosy, School of Pharmaceutical Sciences of Ribeirão Preto, University of São Paulo. Av. do Café s/n, Monte Alegre, Ribeirão Preto, SP, 14040-903, Brazil. [2]Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde. 161 Cathedral Street, Glasgow, G4 0RE, Scotland, United Kingdom. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to R.E.-E. (email: ruangelie.edrada-ebel@strath.ac.uk)

1

growing under different environmental conditions, or changes in metabolite production of individuals within the same population at different seasons[18–21].

Tithonia diversifolia (Helms.) A. Gray, commonly known as tree marigold or Mexican sunflower, is a perennial herb from the family Asteraceae, tribe Heliantheae, native to both Mexico and Central America. *T. diversifolia* is an interesting example or representative of a plant species that can adapt to different types of environment. It is widely distributed around the world, mainly in tropical and sub-tropical areas of America, Africa and Asia[22–24]. *T. diversifolia* has been described as an invasive weed that can establish itself in different types of ecosystems[25–27]. This plant species has been associated with ecological imbalance. Due to its high adaptability to different environmental conditions (climate and soil), high dispersion and rapid growth rate, *T. diversifolia* has an advantage to occupy degraded areas and dominate over native species[26,28].

The chemical constituents of *T. diversifolia* have been well described in the literature. Most of these reported compounds belong to the class of sesquiterpene lactones, flavonoids[23] and *trans*-cinnamic acid derivatives (mainly caffeoylquinic acid)[29]. The major occurrence of hydrocarbon mono- and sesquiterpenes, especially α-pinene and β-caryophyllene are essential oils commonly found in leaves and inflorescences of *T. diversifolia*[24]. Several studies in recent years have shown that extracts and secondary metabolites isolated from leaves and inflorescences of *T. diversifolia* exhibited interesting biological activities such as anti-inflammatory[30,31], antimalarial[32], cytotoxic[33], gastro-protective[22], antimicrobial[34,35], chemopreventive[36] and antihyperglycemic[37,38].

Due to the wide-distribution and great adaptive response capability of *T. diversifolia* to different environmental conditions, and its proficiency to yield potential therapeutic natural products, a specific and innovative metabolomic approach was used to obtain the respective chemical profiles of the samples and to correlate them with the environmental data (climate and soil) from distinct regions at different seasons. The metabolomic approach is useful to understand how variations in plant metabolism can be a response to changes in the surrounding environmental conditions and to be able to propose a statistically sound experimental model that can be applied to environmental metabolomics of plants in the fields of ecology, agriculture, medicinal plants research and other related fields of research.

Considering the ecological adaptability features of *T. diversifolia* as an invasive weed, this species was proven as an adequate object of study to establish an experimental model for environmental metabolomics of plants through a new concatenated approach combining UHPLC-DAD-(ESI)-HRMS and NMR spectral data. Mass spectrometry has the advantage of detecting metabolites at micro- to nanogram concentrations while NMR provides more information on the identity of the metabolites. The new concatenated approach was an efficient method in directly matching the mass to ratio data with the structure of the respective compounds which can only be specified from the NMR datasets that expedites the dereplication step. This approach was used for the first time to carry out a comparative study of different plant part samples of specimens obtained from two environmentally diverse regions of Brazil and collected at different seasons throughout a 24-month period. For this study, metabolomic profiling was focused on a set of compounds that correspond to the main classes of secondary metabolites encountered in *T. diversifolia*, particularly sesquiterpene lactones and phenolics, which are related to the most relevant biological properties presented by this species; in addition, primary metabolites also compose the set of analysed compounds. This study is strategic in provision to bioprospecting the worthwhileness of invasive weeds in yielding bioactive natural products at different environmental conditions. A regular future harvest of such invasive weeds for bioprospecting purposes will maintain ecological balance of the natural home species.

Results

Environmental data.
The environmental data (climate and soil parameters) were obtained during the 24 months of the study. The obtained climate (see Fig. S1) and soil (see Tables S1 and S2) data are shown under Supplementary Information.

UHPLC-DAD-(ESI)-HRMS and NMR data.
Analyses and data processing of 170 extracts by UHPLC-DAD-(ESI)-HRMS recorded 1,277 and 1,084 peaks in the positive and the negative mode, respectively. The negative mode results were used as a basis for proposing the groups for variation analysis of *T. diversifolia* metabolites. On the other hand, the processed ^{1}H and *J*-resolved NMR spectral data were also subjected to PCA and OPLS-DA. Results of the multivariate analysis of the UHPLC-DAD-(ESI)-HRMS and *J*-resolved NMR spectral data were complementary.

The high-resolution MS and NMR spectral data were concatenated[39] and subjected to the same multivariate methods of analyses and strategies. It was observed that all acquired results from the different ways of data analyses (using separate or concatenated data) were very similar. However, multivariate analyses of the concatenated data gave better separation of the proposed groups which could further demonstrate and explain the relationship between the occurrence of the metabolites in *T. diversifolia* and the abiotic environmental factors. The dendrograms obtained by HCA analysis of the respective crude extracts of leaves, stems, roots and inflorescences of *T. diversifolia* as well as the clustering proposal are shown under Supplementary Information (see Figures S2 to S5). PCA and OPLS-DA score scatter plots of the concatenated data of various plant parts according to different environmental conditions are shown in Figs 1 and 2, respectively.

S-line analyses of the *J*-resolved spectral data sets of respective proposed groups (see Figures S6 to S9) provided characteristic chemical shift signals for each group of samples of *T. diversifolia* which can be identified with standard reference compounds previously isolated from this plant. These characteristic chemical shifts from the crude extracts were used in multivariate analyses to provide quick qualitative information to relate the occurrence of respective metabolites in coherence with environmental changes. The PCA loading plots of the concatenated and *J*-resolved data for the respective plant parts are shown in Figs 3–6. Dereplication identified a class of 55 major metabolites used as discriminants in the multivariate analysis (Table S3).
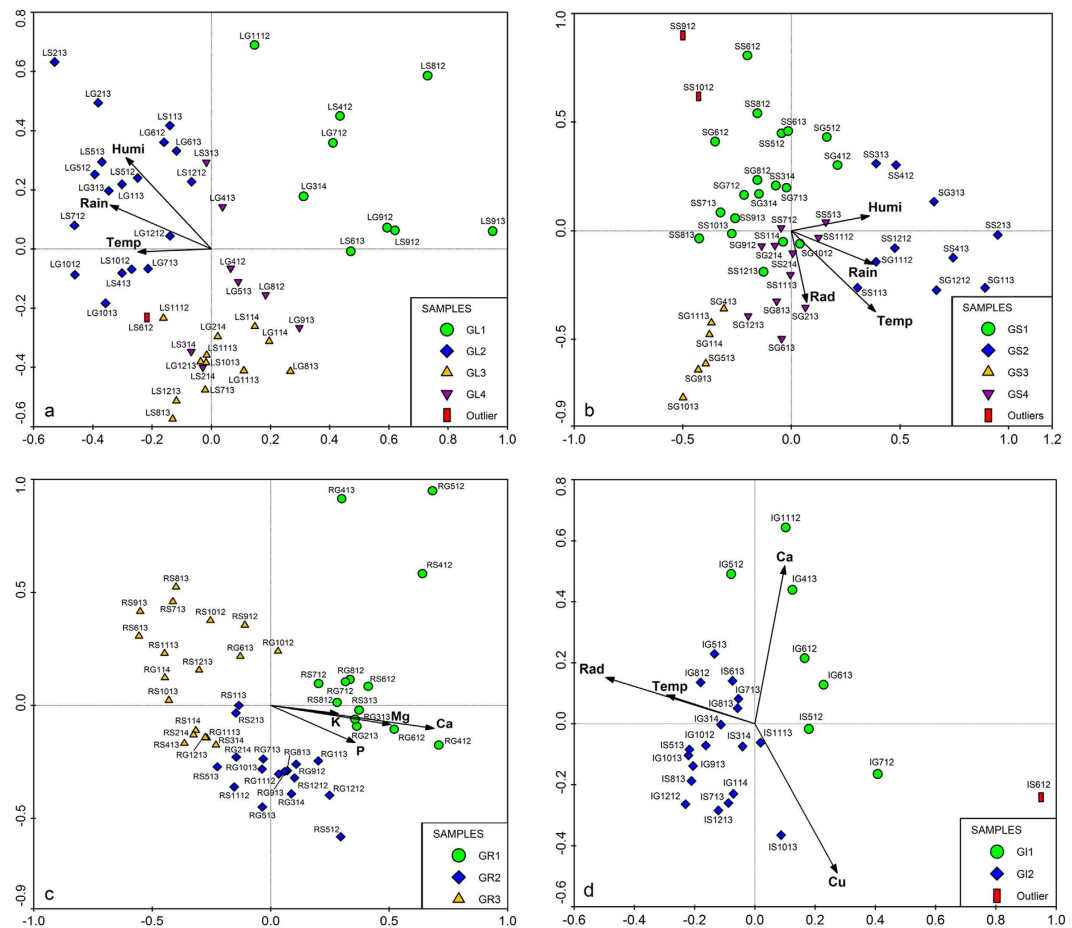
Figure 1. PCA score scatter plots of the concatenated data exhibiting the correlation of the environmental factors with the occurrence of the metabolites from various plant part extracts of *T. diversifolia:* leaves (**a**), stems (**b**), roots (**c**) and inflorescences (**d**). Legend: G = Group; L = Leaves; S = Stems; R = Roots; I = Inflorescences; Number (eg. 112 for Jan 2012 or 1114 for Nov 2014) represents the month (1 for January until 12 for December) and year (12 for 2012 until 14 for 2014); Humi = Humidity; Rain = Rainfall; Temp = Temperature; Rad = Solar radiation; Ca = Soil calcium; Mg = Soil magnesium; P = Soil phosphorus; K = Soil potassium; Cu = Soil copper.

Discussion

Based on the results, it is possible to predict and hypothesize the chemical features of *T. diversifolia* in ration-ale with changes in environmental factors by taking into account the variations in metabolic profile perceived through the occurrence of major metabolites in various plant parts. Regarding the metabolic composition, con-sidering the role of the genetic differences on the production of metabolites in *T. diversifolia*, the obtained results, especially for leaf and stem samples, showed that the genetic differences between the two groups of matrices do not seem to be as significant as the changes in environmental conditions. When analysing the clusters projected for leaves (Figs 1a and S2) and stems (Figs 1b and S3), it can be observed that samples of two different groups of matrices were clustered together, so that the main common feature between samples within the cluster is the time of the year when the samples were collected. It could be also highlighted that, even for the roots and inflo-rescences samples, it was observed that the main pattern of separation between the clusters was not based on the geographical origin of samples, but on the seasonal variation of the environmental factors, so that, biochemically, samples were not grouped primarily according to the group of matrices of which they were obtained. Therefore, the results of this study certainly added new and relevant information on the seasonal variation of the major classes of metabolites found in *T. diversifolia*.

The environmental factors monitored in this study include temperature, rainfall, humidity and solar radiation as well as the amount of soil macro- and micronutrients which proved to correlate to variations in the sugar and nucleoside content as well as secondary metabolite composition of *T. diversifolia*. It should also be kept in mind that other environmental factors that may affect soil conditions in the two remotely distant regions have not demonstrated significant influence on the metabolic profile of the plant.

An interesting aspect that can be observed is that the environmental factors do not seem to exert a constant effect on the entire metabolism of the plant, so that each respective plant part responds differently to changes in environmental conditions. This difference in the metabolic response according to the plant parts has been described in a comparative study of the metabolism of *Holcus lanatus* L. and *Alopecurus pratensis* L. (Poaceae).
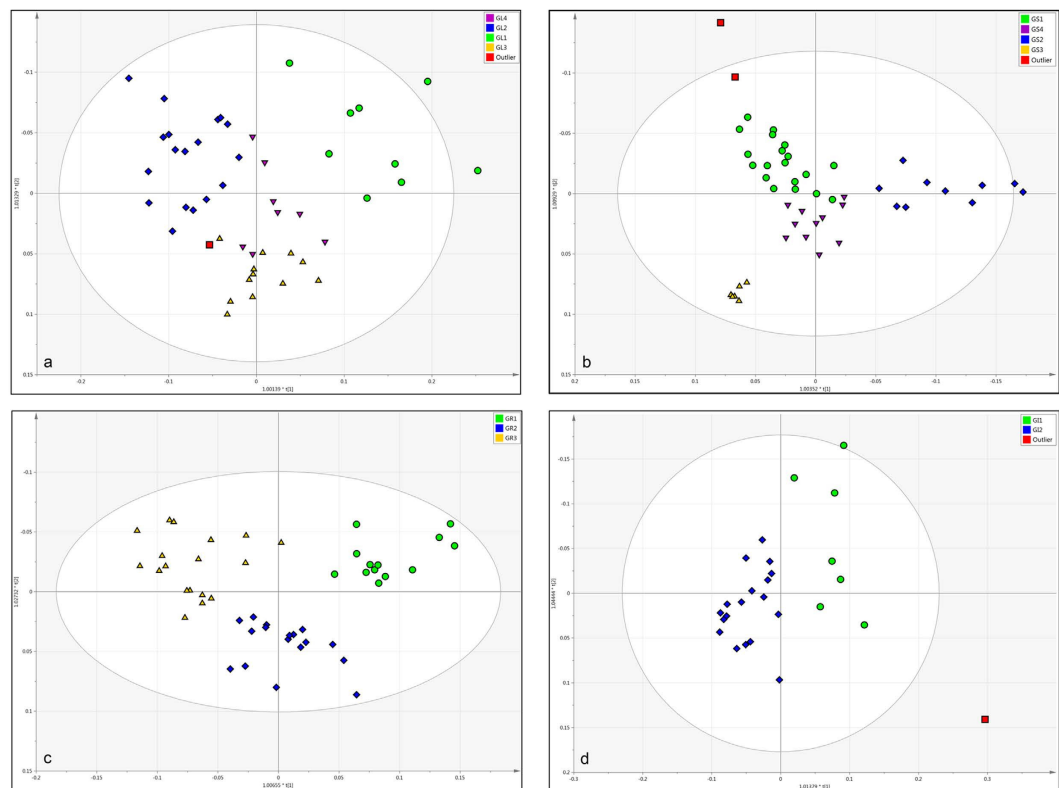
Figure 2. OPLS-DA score scatter plots of the concatenated data obtained from various plant extracts of *T. diversifolia*: leaves (**a**), stems (**b**), roots (**c**) and inflorescences (**d**). G = Group; L = Leaves; S = Stems; R = Roots; I = Inflorescences.

Differences in the accumulation of primary and secondary metabolites between the shoots and roots of both species in different seasons and conditions of drought were observed through LC-MS and NMR spectral data obtained for the samples[40].

Based on the results of the PCA and HCA performed on the concatenated data obtained by UHPLC-DAD-(ESI)-HRMS and *J*-resolved NMR analyses, it was possible to propose groups in agreement to the sample's metabolic profiles in relation with the accrued environmental data for the respective plant part as follows: leaf samples yielded four groups with one outlier (Fig. 1a), stem samples were grouped into four with two outliers (Fig. 1b), root samples indicated three groups (Fig. 1c) while inflorescences resulted to two groups with one outlier (Fig. 1d). The grouping proposal of the various *T. diversifolia* extracts (Fig. 1) showed that, variation in metabolic profile in the leaves and stems seems to be related mainly on rainfall and humidity levels, with temperature and solar radiation also exerting some influence on the metabolic profile. Inflorescences and roots were grouped according to the availability of certain nutrients in the soil, especially soil macronutrients (Ca, Mg, P and K) and the micronutrient Cu, with solar radiation and temperature also significantly affecting the metabolic profile in the inflorescences.

Several climate factors such as water availability, temperature and solar radiation are well described as being able to influence the production of metabolites[2,21,41–43]. Thus, plants under conditions of stress induced by climate factors (i.e. drought, high temperatures, freezing, wide thermal amplitude, high levels of solar radiation) may show changes in the production of different metabolite classes.

Specimens of *T. diversifolia* used for this study were cultivated in a phytogeographic domain indicated as the Brazilian Cerrado, a floristically diverse savannah covering approximately an area larger than two million km² of Central Brazil, representing about 23% of the territory of the country (the same size as Western Europe)[44,45]. The predominant climate in the Brazilian Cerrado is semi-humid tropical, which is characterized by strong seasonality of rainfall, with two distinct seasons: the dry season that typically extends from May to September and a rainy season from October to April, with an annual average temperature ranging from 21.3–27.2 °C[46].

The availability of water is a known factor related to the variation in the production of metabolites in plants[2,21]. Specimens used in this study were grown in places that have a patchy distribution of rainfall throughout the year. Analyses of the PCA plots determined the relationship between the seasonality of the rain in the area and the metabolic profile of the leaves and stems (Fig. 1a,b). It can be observed that samples collected during the rainiest periods tend to cluster and were rich in primary metabolites such as sugars and nucleosides. Leaf and stem samples collected during the drier periods were rich in secondary metabolites such as sesquiterpene lactones particularly in the cold months while there was an increase production of *trans*-cinammic acid ester derivatives during the hot months. Collected samples of leaves and stems during the transition period between the dry and rainy seasons yielded intermediate levels of both primary and secondary metabolites.
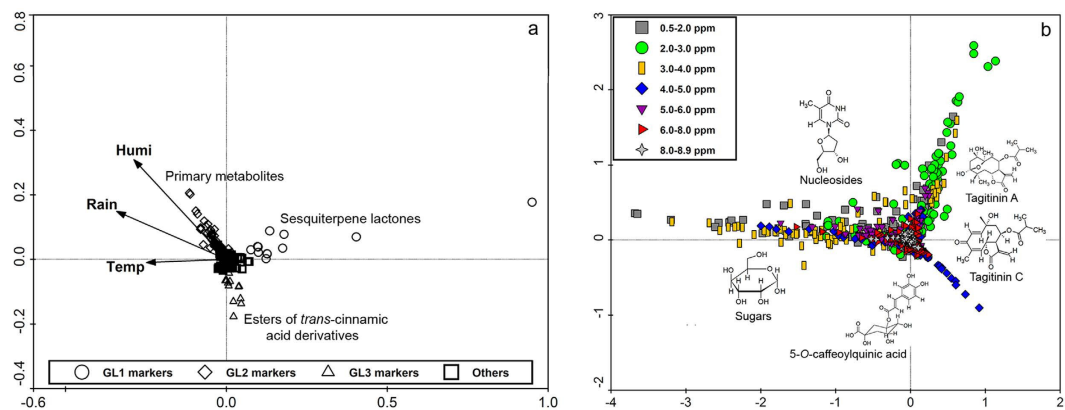
Figure 3. PCA loading plots of the concatenated data (**a**) and *J*-resolved chemical shift data (**b**) of *T. diversifolia* leaf extracts to correlate discriminant classes of metabolites with the various environmental factors. G = Group; L = Leaves; Rain = Rainfall; Humi = Humidity; Temp = Temperature.

In accordance to the PCA (Fig. 1c,d) and OPLS-DA (Fig. 2c,d) plots, root and inflorescence samples of *T. diversifolia* were grouped in a pattern related mainly to the availability of certain soil nutrients. For inflorescences, temperature and solar radiation also seem to influence metabolite production. As clearly shown in Figs 1c and 2c, root samples were distributed into three large groups. The first group, designated as GR1, composed mostly of samples from the state of Goiás which were mainly collected in the first year of study and samples from state of São Paulo were collected in the first half of the first year of study. During this collection period, there were higher levels of soil macronutrients in both regions and GR1 was characterized by lower accumulation of secondary metabolites. The second group, GR2 consisted of a larger number of samples from Goiás, mainly collected in the second year of the study, and samples from São Paulo collected in the second half of the first year. The observed levels of soil macronutrients decreased during this period and in parallel, GR2 samples indicated low but significant accumulation of phenolic compounds. For the third group, GR3, composed mainly of samples from São Paulo collected during the second year of study, it was observed the lowest levels of soil macronutrients and, consequently, the high accumulation of phenolic compounds, especially esters of *trans*-cinnamic acid derivatives, in the group samples.

For the inflorescences, it was possible to observe the existence of two main groups (Figs 1d and 2d). Group GI1 was primarily composed of samples from Goiás, where the soil showed higher levels of Ca, and the temperature and levels of solar radiation were low during the flowering season. The accumulation of sugars was observed along with high Ca concentration in the soil, while, the occurrence of sesquiterpene lactones was correlated with the increase of Cu levels. Group GI2 consisted mostly of samples collected in São Paulo described to have higher temperatures and higher levels of solar radiation. The collected inflorescences indicated the accumulation of esters of *trans*-cinnamic acid derivatives. Similar to the collection from Goiás, the presence of certain soil nutrients can also be associated with the occurrence of some metabolites. Moderate levels of primary metabolites and sesquiterpene lactones were detected in inflorescence samples of plants from soils rich in Ca or Cu, respectively.

The four aerial part outliers (one for the leaves, two for the stems and one for the inflorescence) were characterized by high accumulation of two specific sesquiterpene lactones: namely tagitinin A but mainly tagitinin C. All these outlying samples were collected from São Paulo during the dry season of the first year of study. The increased accumulation of tagitinins A and C was observed to be related to a specific combination of environmental conditions, in this case, low water availability, low temperatures and low levels of solar radiation.

The PCA loading plots obtained from the NMR (*J*-resolved) data showed signals with chemical shift values representing major classes of metabolites used as discriminants for the respective proposed groups. For example, as shown in Figs 3 and 4, resonances between 5 to 6 ppm were related to olefinic double bonds present in an aliphatic carbon skeleton while those between 2 to 3 ppm corresponded to hydrogens attached to a saturated carbon in the environment of a carbonyl group typical to unsaturated fatty acids and sesquiterpene lactones, as found in tagitinin C in samples of leaves and stems grouped under GL1 and GS1, which were characterized by the accumulation of these substances.

The characterization of groups rich in esters of *trans*-cinnamic acid congeners was not so clear by using only the NMR *J*-resolved data results because of overlapping resonances belonging to different classes of metabolites, such as signals between 6 to 7 ppm for phenolic hydrogens which overlap with hydrogens of conjugated double bonds present in some type of sesquiterpene lactones. The use of the UHPLC-DAD-(ESI)-HRMS data were very useful to discriminate both groups. On the other hand, the NMR *J*-resolved data were very useful in characterizing groups of samples rich in sugars indicated by proton resonances from 3 to 5 ppm typical for hydroxylated methine protons common in carbohydrate molecules. The OPLS-DA S-line plots obtained from the NMR *J*-resolved data, shown in the Figures S6 to S9 (supplementary information), illustrate how some ^{1}H-NMR signals from different classes of metabolites can be useful to separate two different groups of samples within the same species, as a good discriminant between diverse chemical profiles (Figs 3–6).
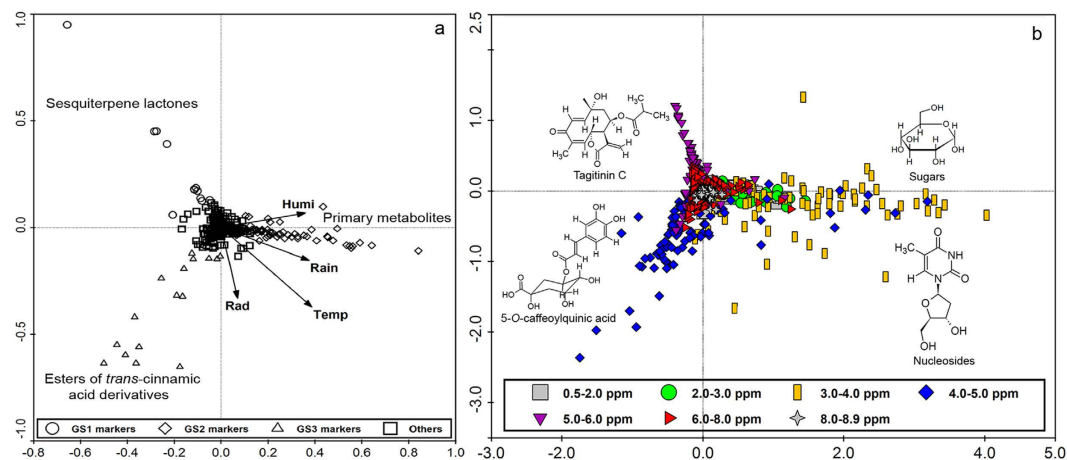
Figure 4. PCA loading plots of the concatenated data (**a**) and *J*-resolved chemical shift data (**b**) of *T. diversifolia* stem extracts to correlate discriminant classes of metabolites with the various environmental factors. G = Group; S = Stems; Rain = Rainfall; Humi = Humidity; Temp = Temperature; Rad = Solar radiation.
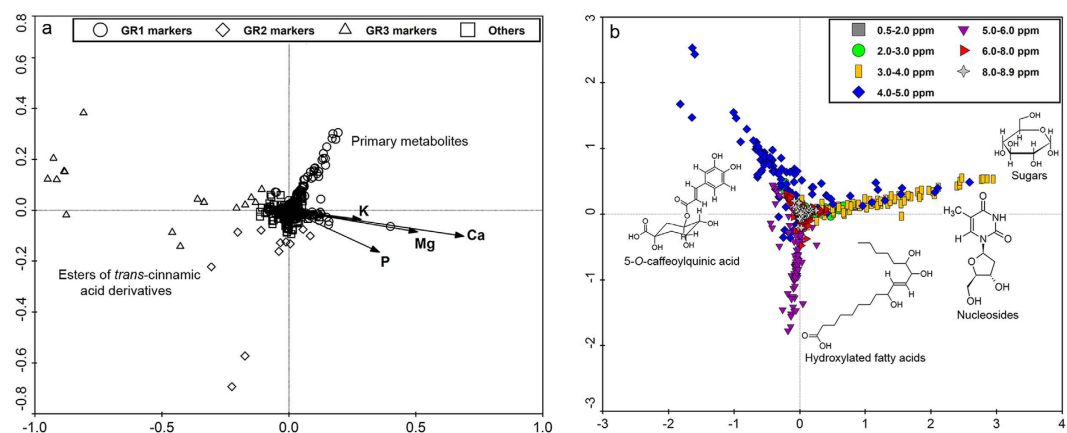

Figure 5. PCA loading plots of the concatenated data (**a**) and *J*-resolved chemical shift data (**b**) of *T. diversifolia* root extracts to correlate discriminant classes of metabolites with the variations in soil nutrients. G = Group; R = Roots; Ca = Soil calcium; Mg = Soil magnesium; P = Soil phosphorus; K = Soil potassium.

The mechanism involved in the influence of climate factors on *T. diversifolia* metabolism appears to be related to seasonal induction of stress factors in the plant, such as: (i) drought stress, which can cause a reduction in photosynthetic rate with a consequent increase in production of reactive oxygen species (ROS), therefore resulting in an increase in the production of phenolic compounds (natural antioxidants) during the dry season as a defense mechanism[2,21,47]; (ii) thermal stress caused by large temperature variation range during the year, especially during drier periods, which can affect metabolic regulation, permeability to water and CO_2 and the rate of intracellular reactions[2,48] as well as improve the antioxidant properties with an increase availability of carbohydrates[48]; and (iii) stress by solar radiation, particularly UV-B radiation (280–320 nm), which mainly affects the production of phenolic compounds (tannins, anthocyanins, flavonoids and other *trans*-cinnamic acid derivatives) which, besides contributing to absorption and/or dissipation of solar energy, provide protection against the deleterious effects of UV radiation, such as tissue damage induced directly by UV-B or formation of free radicals and other oxidative species[49–51].

The influence that the soil nutrients Ca, Mg, P, K and Cu exert on the roots and inflorescences of *T. diversifolia* can be related to several physiological roles that such nutrients play in higher plants, since these nutrients are essential for normal growth and functioning of metabolism[52]. The accumulation of esters of *trans*-cinnamic acid derivatives in roots from soils with lower levels of macronutrients may be a response to the reduced nutrient availability, since polyphenols bound to sesquioxides have the ability to prevent adsorption of phosphate. Moreover, phenolic acids also have the ability to desorb the bound phosphate, thus increasing the availability of inorganic P in the soil, aside from the fact that phenolic compounds can also retain inorganic exchangeable cations (Ca, Mg and K) providing sorption sites on acid soils[53].

Considering all the results presented herein, we can highlight the relationships between environment and the metabolic profile of *T. diversifolia*, in which the variation in the production of certain classes of metabolites in the plant appeared to be a direct response to changes in conditions of its surrounding environment. It can also
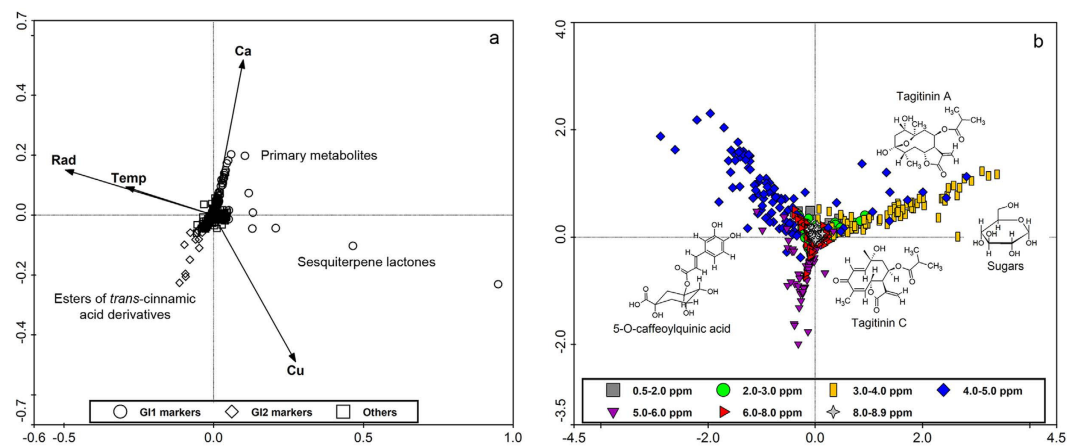
Figure 6. PCA loading plots of the concatenated data (**a**) and *J*-resolved chemical shift data (**b**) of *T. diversifolia* inflorescence extracts to correlate discriminant classes of metabolites with the various environmental factors. G = Group; I = Inflorescences; Ca = Soil calcium; Cu = Soil copper; Temp = Temperature; Rad = Solar radiation.

be highlighted that the set of analytical techniques selected for this work combined with proper multivariate analysis allowed us to obtain statistically-sound, reliable data that contribute to a holistic understanding of how some primary but mainly secondary metabolism are affected in the species according to the changes of the abiotic environmental factors of the location where it can be encountered and how respective plant parts is differently affected by the environment.

There are different works reporting the use of a metabolomic approach to study the influence of the environmental factors on the metabolite production of some economically important plants, such as rice cultivars[54], genetic modified maize[55–57], citrus rootstocks[58] and green tea[59]. However, for the first time, metabolic information was obtained by the fusion of two data sets originated from high resolution LC-MS and NMR, to perform the statistical analyses. The fused concatenated data set was an efficient method in directly matching the *m/z* signals with the structure of the respective compounds within a cluster which expedites the dereplication step.

Thus, the approach reported herein may be useful in the study of other invasive and highly adaptable species as well as in the study of the effects of environmental changes in the development of plants, which can also be extended to explore quality control or biological properties of further economically important crops, food species or medicinal plants, the latter responsible for the accumulation of active substances that exert pharmacological effect in humans.

Methods

Collection of plant material. For this study, samples were collected in condition ex situ from two groups of matrices, each composed of 48 individuals, cultivated in two different states of Brazil, one located in the Garden of Medicinal Plants of the School of Pharmaceutical Sciences of Ribeirão Preto, University of São Paulo, Ribeirão Preto, state of São Paulo (latitude 21° 10′ 07.4″ S, longitude 47° 50′ 49.1″ W), and the other at Fazenda Santo Antonio (Saint Anthony Farm), city of Pires do Rio, state of Goiás (latitude 17° 12′ 38.8″ S, longitude 48° 18′ 24.9″ W). Cultivation was carried out in two Brazilian states with the aim of analysing the influence of two diverse geographical locations on the metabolic homeostasis of the species.

The 48 individuals of each group of matrices were cultivated from seedlings obtained by stem cuttings of two parent plants: PP-1 from the city of Ribeirão Preto, state of São Paulo; and PP-2 from the city of Pires do Rio, state of Goiás. PP-1 was the source of the seedlings used for the cultivation of the group of matrices for Ribeirão Preto while PP-2 was the seedling source for the group of matrices for Pires do Rio. A more detailed description of the cultivation of the groups of matrices of *T. diversifolia* is shown in the Supplementary Information (Method S1: Cultivation of *Tithonia diversifolia* by vegetative propagation).

The collections at both sites started in April 2012, three months after the *T. diversifolia* seedlings were transferred to the cultivation areas, with individuals in a young stage (with floral buds), and were consequently carried out monthly from May 2012 with individuals in an adult stage (bearing inflorescences), for a two-year period until March 2014. The collected material was separated according to plant parts (leaves, stems, roots, and inflorescences), packed in labelled paper bags, immediately frozen on dry ice (−78 °C) after collection, dried in freeze dryer and stored in a refrigerator at −20 °C until subjected to extraction, following an adapted standard operating procedure[19].

Voucher specimens for each sampled population were deposited and identified at Herbarium SPF, Institute of Biosciences, University of São Paulo, under the responsibility of the curation of Dr. Renato de Mello Silva. Voucher numbers for each specimens were assigned as Sampaio #01 (samples from Pires do Rio-GO) and Sampaio #02 (samples from Ribeirão Preto-SP). All the samples and the voucher specimens were prepared according to the standard operating procedure for the access and shipment of component of genetic heritage, No. 010319/2013–1, as issued by the National Council for Scientific and Technological Development (*Conselho Nacional de Desenvolvimento Científico e Tecnológico* – CNPq), under authorization of Genetic Heritage Management Council (*Conselho de Gestão do Patrimônio Genético* – CGEN).

Collection of soil samples and climate data. Commencing from April 2012, soil collection was performed every six months, following the standard procedure for soil analysis recommended by the Agronomic Institute of Campinas (*Instituto Agronômico de Campinas*, IAC). Soil samples were placed in labelled plastic bags (four replicates for each geographical location) and a 200 g aliquot was separated and sent for soil analysis of macro and micronutrients, including sulphur and aluminium composition[60].

For all the plant collection areas covered in this study, climate data information between March 2012 and March 2014 was obtained from the available online climate database of the National Institute of Meteorology (*Instituto Nacional de Meteorologia* – INMET). Climate data included temperature (°C), humidity (%), solar radiation (kJ/m^2) and rainfall (mm). For data analysis, monthly averages of temperature, humidity and radiation, and accumulated rainfall for each month were used.

Preparation of crude extracts of *T. diversifolia*. To prepare the crude extracts used in this study, freeze-dried samples of different plant parts (leaves, stems, roots and inflorescences) were ground in a mill grinder (11 Basic IKA Model), and was sieved through a size 42 mesh (opening = 0.355 mm), packed in plastic microtubes and then stored at −20 °C.

From each collected sample, 100 mg aliquots were weighed into plastic microtubes (a total of 170 samples - 48 samples each of leaves, stems and roots, and 26 samples of inflorescences), 2 mL of 70% ethanol (v/v) were added and the extraction was performed in an ultrasonic bath for 20 min. After the extraction procedure, the material was centrifuged for 3 min at 13,000 rpm, the supernatant was filtered through a syringe filter (0.22 μm pore diameter), the filtrate was further cleaned-up by partitioning with *n*-hexane (95% HPLC grade) to remove excess of fats (fatty acids, waxes, etc.) and pigments, followed by centrifugation to separate the *n*-hexane partition which was then discarded. The fat-free plant extracts were dried in a centrifugal vacuum prior to metabolite variation analyses.

Preparation of extracts for analysis by UHPLC-DAD-(ESI)-HRMS. For the preparation of the samples for UHPLC-DAD-(ESI)-HRMS analysis, the obtained extracts from the previous section were dissolved in water (Milli-Q)/acetonitrile (LC-MS grade) (7:3, v/v) at a concentration of 1 mg/mL then filtered through a syringe filter (0.22 μm). A solvent blank of water/acetonitrile (7:3), was also prepared in order to subtract any solvent interfering signals in the data processing step. A reference sample containing a mixture of extracts from various plant parts was used to align the chromatograms during data processing.

Preparation of extracts for analysis by ¹H NMR and J-resolved. A total of 170 *T. diversifolia* extracts was analysed by nuclear magnetic resonance spectroscopy (¹H NMR and *J*-resolved) at the Strathclyde Institute of Pharmacy and Biomedical Sciences, Glasgow, UK. Access and sample shipment were in accordance with the Brazilian laws (Term of Authorization for Access and Shipment No. 010319/2013-1 issued by CNPq and approved by the letter COAPG No. 182/2014).

For NMR analysis, 5 mg of each of the sample were dissolved in 650 μL of methanol-d$_4$/D$_2$O (600:50) and transferred to 5 mm diameter 7″ NMR tubes.

Analysis of the extracts by UHPLC-DAD-(ESI)-HRMS. The 170 extracts were analysed by UHPLC-DAD-(ESI)-HRMS Thermo Scientific® Accela, equipped with Accela 1250 quaternary pumps, coupled to a Accela photo-diode array detector and a Thermo Scientific® Exactive Plus mass spectrometer with an Orbitrap® analyzer.

Chromatographic analysis of the extracts was performed on an ACE® analytical column (3.0 mm internal diameter × 150 mm length, 3 μm particle size) placed in an oven maintained at 35 °C using H$_2$O with 0.1% formic acid as solvent A and acetonitrile with 0.1% formic acid as solvent B with the following gradient elution program; at 0 min, it started with 5% B which was increased to 20% B in 10 min , then to 25% B in 15 min to 45% B in 25 min, and to 100% B in 30 min, continued at 100% B to 38 min for washing and was equilibrated back to 5% B to 44 min at a solvent flow rate of 400 μL/min.

Electrospray ionization method was used for mass spectrometry under the following conditions: spray voltage (positive mode = 3.6 kV; negative mode = 3.2 kV); temperature of the capillary in positive mode at 300 °C and 320 °C in the negative mode. The analyzer used was an Orbitrap®.

Analysis of the extracts by NMR. NMR was carried out for all the 170 *T. diversifolia* extracts on a 400 MHz Jeol-LA400 FT-NMR spectrometer system equipped with a 40TH5AT/FG probe (JEOL, Tokyo, Japan). The acquisition of the one-dimensional proton spectra (¹H NMR) was performed by the pre-saturation pulse sequence using 16 scans per analysis. Subsequently, two-dimensional *J*-resolved NMR spectra were also acquired with 32 scans and 64 increments per scan. Data points were collected into a plot using spectral widths of 3.56 kHz for F2 (chemical shift axis) and 50 Hz for F1 (spin–spin coupling constant axis). The pre-saturation method was used to suppress the solvent signal during acquisition and the two-dimensional *J*-resolved spectra provided separation of overlapping chemical shift values (δ) against corresponding coupling constants (*J*) at different axes.

Dereplication of the extracts by UHPLC-DAD-(ESI)-HRMS. The data obtained for the different *T. diversifolia* extracts, after analysis by UHPLC-DAD-(ESI)-HRMS, were used in the step of dereplication, based on the comparison of the values of accurate mass and the profiles of absorbance into UV/Vis for the peaks detected with in-house compound libraries, AsterDB Asteraceae Database, http://www.asterbiochem.org/asterdb) and TitDB (Tithonia Database), and the Dictionary of Natural Products (http://dnp.chemnetbase.com/).

Data processing and multivariate analysis. The data obtained after analysis of the extracts by UHPLC-DAD-(ESI)-HRMS were pre-treated and pre-processed for multivariate analysis. MS data was acquired

on switch mode. The obtained data were separated between positive and negative mode of ionization and converted to mzXML format with ProteoWizard 3.0.6002 package MSConvert software (Proteowizard Software Foundation). The sorted data was processed by MZmine 2.10 (MZmine 2 project) for peak detection, peak filtering, chromatogram construction, chromatogram deconvolution, isotopic peak grouping, chromatogram alignment, gap filling, and the search for adducts and peak identification using an in-house compounds database. The following MZmine parameters were used for the data processing: noise level at 10^6; Lorentzian function for the peak shape; minimum peak height at 5×10^6; and m/z tolerance at 0.002 m/z or 5.0 ppm.

The processed data were then exported as tables categorized according to peak areas, exact mass, and retention times for each sample extract, which were then subjected for multivariate statistical analysis.

The acquired 2D J-resolved spectrum were first tilted at 45° to remove the effect of constant couplings over the shaft with the chemical shift values then symmetrized. The spectra on the one-dimensional projection were extracted then processed following the same steps described for ^{1}H spectra.

The data obtained by ^{1}H NMR and J-resolved analyses were processed with MestReNova version 8 (Mestrelab Research S.L.©) prior to multivariate analysis. The ^{1}H NMR spectra of respective extracts were stacked and processed in active spectrum mode. Pre-processing included baseline correction using the Whittaker-Smoother method, apodization was at a Gaussian function of 1 GB [Hz], normalization was on the highest signal equivalent to 100, smoothing was done with the Savitzky-Golay method, binning to the full spectrum used a value of 0.01 ppm for each bin, then subjected to a second normalization step to normalize the intensity of each bin.

The processed NMR data were exported as electronic tables containing the intensities of the signals for each chemical shift at 0.01 ppm intervals. Chemical shift values between 0.5 to 9.25 ppm and 0.5 to 8.9 ppm were used from the ^{1}H and J-resolved spectral data, respectively. The solvent peaks for methanol-d_4 and D_2O were manually deleted in MS-Excel®.

The environmental data sets for the soil nutrients and climate from the two collection sites of *T. diversifolia* were processed according to the data type - data expressed as percentages (relative humidity and percentage of saturation for bases) and data not expressed as percentages (organic matter, levels of P, K, Ca, Mg, S, B, Cu, Fe, Mn and Zn, soil total acidity, cation exchange capacity and the sum of the bases). The environmental data expressed in percentages were transformed by the equation (1), while the other experimental data not expressed in percentages were calculated using the equation (2). For the soil pH, unprocessed data were used.

$$\hat{x}_n = \arcsin\left(\sqrt{\frac{x_n}{100}} \right)$$

(1). Transformation of variables by the arcsin method.

$$\hat{x}_n = \log_{10}(x_n + 1)$$

(2). Logarithmic transformation for the variables.

The obtained processed and normalized spectral data from UHPLC-DAD-(ESI)-HRMS and ^{1}H-NMR were combined using a data fusion method known as concatenation[39]. The data were divided into two blocks consisting of the LC-MS data block and ^{1}H-NMR data block). Each block was scaled by the equation below (3). After the scaling procedure, both blocks were combined into one single data matrix which was subjected to multivariate analysis.

$$\hat{x}_n = \frac{x_n}{\sum \sigma_{block}}$$

(3) Equation of variable scaling for data fusion by concatenation method.

For the study of metabolite variation in *T. diversifolia,* after appropriate treatment of the separate and concatenated data obtained by UHPLC-DAD-(ESI)-HRMS and NMR against the climate and soil data, the results were analysed with CANOCO 4.5 (Biometris – Plant Research International©), OriginPro 9.6 (OriginLab Corporation©) and SIMCA-P (Umetrics©). The data were divided into two sets of variables: chemical (data obtained from the analysis of the extracts) and environmental variables (climate and soil data). These variables were used to perform multivariate analysis by unsupervised method through Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) by the Ward's method[61] as well as a supervised method by Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA). These methods are ordering methods, with the aim of reducing the data set dimension to be able to explain the total variation of the system and clustering of the samples according to their chemical profile.

References

1. Pavarini, D. P., Pavarini, S. P., Niehues, M. & Lopes, N. P. Exogenous influences on plant secondary metabolite levels. *Anim. Feed Sci. Technol.* **176,** 5–16 (2012).
2. Ramakrishna, A. & Ravishankar, G. A. Influence of abiotic stress signals on secondary metabolites in plants. *Plant Signal. Behav.* **6,** 1720–1731 (2011).
3. Gutbrodt, B., Dorn, S., Unsicker, S. B. & Mody, K. Species-specific responses of herbivores to within-plant and environmentally mediated between-plant variability in plant chemistry. *Chemoecology* **22,** 101–111 (2012).
4. Edreva, A. *et al.* Stress-protective role of secondary metabolites: diversity of functions and mechanisms. *Gen. Appl. Plant Physiol.* **34,** 67–78 (2008).
5. Oh, M.-M., Trick, H. N. & Rajashekar, C. B. Secondary metabolism and antioxidants are involved in environmental adaptation and stress tolerance in lettuce. *J. Plant Physiol.* **166,** 180–191 (2009).

6. Theis, N. & Lerdau, M. The Evolution of Function in Plant Secondary Metabolites. *Int. J. Plant Sci.* **164,** S93–S102 (2003).
7. Bennett, R. N. & Wallsgrove, R. M. Secondary Metabolites in Plant Defence Mechanisms. *New Phytol.* **127,** 617–633 (1994).
8. Wink, M. Plant breeding: importance of plant secondary metabolites for protection against pathogens and herbivores. *Theor. Appl. Genet.* **75,** 225–233 (1988).
9. Szakiel, A., Pączkowski, C. & Henry, M. Influence of environmental abiotic factors on the content of saponins in plants. *Phytochem. Rev.* **10,** 471–491 (2010).
10. Miranda, M. A. F. M. *et al.* Phytotoxins from *Tithonia diversifolia*. *J. Nat. Prod.* **78,** 1083−1092 (2015).
11. Treutter, D. Significance of flavonoids in plant resistance and enhancement of their biosynthesis. *Plant Biol.* (*Stuttg*). **7,** 581–591 (2005).
12. Winkel-Shirley, B. Biosynthesis of flavonoids and effects of stress. *Curr. Opin. Plant Biol.* **5,** 218–223 (2002).
13. Stashenko, E. E. *et al. Lippia origanoides* chemotype differentiation based on essential oil GC-MS and principal component analysis. *J. Sep. Sci.* **33,** 93–103 (2010).
14. Lukas, B., Schmiderer, C., Franz, C. & Novak, J. Composition of essential oil compounds from different Syrian populations of *Origanum syriacum* L. (Lamiaceae). *J. Agric. Food Chem.* **57,** 1362–1365 (2009).
15. Vilela, E. C. *et al.* Spatial chemometric analyses of essential oil variability in *Eugenia dysenterica*. *J. Braz. Chem. Soc.* **24,** 873–879 (2013).
16. Telascrea, M. *et al.* Essential oil from leaves of *Cryptocarya mandioccana* Meisner (Lauraceae): Composition and intraspecific chemical variability. *Biochem. Syst. Ecol.* **35,** 222–232 (2007).
17. Rahimmalek, M. *et al.* Essential oil variation among and within six *Achillea* species transferred from different ecological regions in Iran to the field conditions. *Ind. Crops Prod.* **29,** 348–355 (2009).
18. Jones, O. A. H. *et al.* Metabolomics and its use in ecology. *Austral Ecol.* **38,** 713–720 (2013).
19. Kim, H. K., Choi, Y. H. & Verpoorte, R. NMR-based metabolomic analysis of plants. *Nat. Protoc.* **5,** 536–549 (2010).
20. Bundy, J. G., Davey, M. P. & Viant, M. R. Environmental metabolomics: a critical review and future perspectives. *Metabolomics* **5,** 3–21 (2009).
21. Arbona, V., Manzi, M., Ollas, C. De & Gómez-Cadenas, A. Metabolomics as a tool to investigate abiotic stress tolerance in plants. *Int. J. Mol. Sci.* **14,** 4885–4911 (2013).
22. Sánchez-Mendoza, M. E. *et al.* Bioassay-guided isolation of an anti-ulcer compound, tagitinin C, from *Tithonia diversifolia*: role of nitric oxide, prostaglandins and sulfhydryls. *Molecules* **16,** 665–674 (2011).
23. Chagas-Paula, D. A., Oliveira, R. B., Rocha, B. A. & Da Costa, F. B. Ethnobotany, chemistry, and biological activities of the genus *Tithonia* (Asteraceae). *Chem. Biodivers.* **9,** 210–235 (2012).
24. Moronkola, D. O., Ogunwande, I. A., Walker, T. M., Setzer, W. N. & Oyewole, I. O. Identification of the main volatile compounds in the leaf and flower of *Tithonia diversifolia* (Hemsl) Gray. *J. Nat. Med.* **61,** 63–66 (2006).
25. Sun, W., Chen, G. & Wang, S. Characteristics of *Tithonia diversifolia*: an alien invasive plant in Yunnan, south-west China. In *3rd Glob. Bot. Gard. Congr.* **2,** 1–7 (Botanic Gardens Conservation International, 2007).
26. Muoghalu, J. I. Growth, reproduction and resource allocation of *Tithonia diversifolia* and *Tithonia rotundifolia*. *Weed Res.* **48,** 157–162 (2008).
27. Ayeni, A. O., Lordbanjou, D. T. & Majek, B. A. *Tithonia diversifolia* (Mexican sunflower) in south-western Nigeria: occurrence and growth habit. *Weed Res.* **37,** 443–449 (1997).
28. Muoghalu, J. I. & Chuba, D. K. Seed germination and reproductive strategies of *Tithonia diversifolia* (Hemsl.) Gray and *Tithonia rotundifolia* (P.M) Blake. *Appl. Ecol. Environ. Res.* **3,** 39–46 (2005).
29. Zhao, G.-J. *et al.* Chemical constituents from *Tithonia diversifolia* and their chemotaxonomic significance. *Biochem. Syst. Ecol.* **44,** 250–254 (2012).
30. Owoyele, V. B., Wuraola, C. O., Soladoye, A. O. & Olaleye, S. B. Studies on the anti-inflammatory and analgesic properties of *Tithonia diversifolia* leaf extract. *J. Ethnopharmacol.* **90,** 317–321 (2004).
31. Chagas-Paula, D. A. *et al.* Chlorogenic acids from *Tithonia diversifolia* demonstrate better anti-inflammatory effect than indomethacin and its sesquiterpene lactones. *J. Ethnopharmacol.* **136,** 355–362 (2011).
32. Oyewole, I. O., Adeoye, G. O., Anyasor, G. & Obansa, J. A. Anti-malarial and repellent activities of *Tithonia diversifolia* (Hemsl.) leaf extracts. *J. Med. Plants Res.* **2,** 171–175 (2008).
33. Kuroda, M. *et al.* Sesquiterpenoids and flavonoids from the aerial parts of *Tithonia diversifolia* and their cytotoxic activity. *Chem. Pharm. Bull.* (*Tokyo*). **55,** 1240–1244 (2007).
34. Liasu, M. & Ayandele, A. Antimicrobial activity of aqueous and ethanolic extracts from *Tithonia diversifolia* and *Bryum coronatum* collected from Ogbomoso, Oyo state, Nigeria. *Adv. Nat. Appl. Sci.* **2,** 31–34 (2008).
35. Obafemi, C., Sulaimon, T., Akinpelu, D. & Olugbade, T. Antimicrobial activity of extracts and a germacranolidetype sesquiterpene lactone from *Tithonia diversifolia* leaf extract. *African J. Biotechnol.* **5,** 1254–1258 (2006).
36. Gu, J.-Q. *et al.* Sesquiterpenoids from *Tithonia diversifolia* with potential cancer chemopreventive activity. *J. Nat. Prod.* **65,** 532–536 (2002).
37. Zhao, G., Li, X., Chen, W., Xi, Z. & Sun, L. Three new sesquiterpenes from *Tithonia diversifolia* and their anti-hyperglycemic activity. *Fitoterapia* **83,** 1590–1597 (2012).
38. Li, X., Huang, G., Zhao, G. & Chen, W. Two New Monoterpenes from *Tithonia diversifolia* and Their Anti-Hyperglycemic Activity. *Rec. Nat. Prod.* **4,** 351–354 (2013).
39. Forshed, J., Idborg, H. & Jacobsson, S. P. Evaluation of different techniques for data fusion of LC/MS and 1H-NMR. *Chemom. Intell. Lab. Syst.* **85,** 102–109 (2007).
40. Gargallo-Garriga, A. *et al.* Opposite metabolic responses of shoots and roots to drought. *Sci. Rep.* **4,** 6829 (2014).
41. Shulaev, V., Cortes, D., Miller, G. & Mittler, R. Metabolomics for plant stress response. *Physiol. Plant.* **132,** 199–208 (2008).
42. Graham, D. & Patterson, B. D. Responses of Plants to Low, Nonfreezing Temperatures: Proteins, Metabolism, and Acclimation. *Annu. Rev. Plant Physiol.* **33,** 347–372 (1982).
43. Jakobsen, H. & Olsen, C. Influence of climatic factors on emission of flower volatiles *in situ*. *Planta* **192,** 365–371 (1994).
44. Ratter, J. The Brazilian Cerrado Vegetation and Threats to its Biodiversity. *Ann. Bot.* **80,** 223–230 (1997).
45. Simon, M. F. *et al.* Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by *in situ* evolution of adaptations to fire. *Proc. Natl. Acad. Sci.* **106,** 20359–20364 (2009).
46. EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária), O Cerrado. Available at http://www.cpac.embrapa.br/unidade/ocerrado/ (Accessed: 13th July 2014) (2012).
47. Reddy, A. R., Chaitanya, K. V. & Vivekanandan, M. Drought-induced responses of photosynthesis and antioxidant metabolism in higher plants. *J. Plant Physiol.* **161,** 1189–1202 (2004).
48. Morison, J. I. L. & Lawlor, D. W. Interactions between increasing CO_2 concentration and temperature on plant growth. *Plant, Cell Environ.* **22,** 659–682 (1999).
49. Grace, S. C. & Logan, B. a. Energy dissipation and radical scavenging by the plant phenylpropanoid pathway. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **355,** 1499–1510 (2000).
50. Jansen, M. A., Gaba, V. & Greenberg, B. M. Higher plants and UV-B radiation: balancing damage, repair and acclimation. *Trends Plant Sci.* **3,** 131–135 (1998).

51. Rozema, J., van de Staaij, J., Björn, L. O. & Caldwell, M. UV-B as an environmental factor in plant life: stress and regulation. *Trends Ecol. Evol.* **12,** 22–28 (1997).
52. Marschner, H. Functions of Mineral Nutrients in *Marschner's Mineral Nutrition of Higher Plants* 229–404 (Elsevier Science Ltd, 1995).
53. Hättenschwiler, S. & Vitousek, P. M. The role of polyphenols in terrestrial ecosystem nutrient cycling. *Trends Ecol. Evol.* **15,** 238–242 (2000).
54. Hu, C. *et al.* Metabolic variation between *japonica* and *indica* rice cultivars as revealed by non-targeted metabolomics. *Sci. Rep.* **4,** 5067 (2014).
55. Baniasadi, H., Vlahakis, C., Hazebroek, J., Zhong, C. & Asiago, V. Effect of environment and genotype on commercial maize hybrids using LC/MS-based metabolomics. *J. Agric. Food Chem.* **62,** 1412–1422 (2014).
56. Asiago, V. M., Hazebroek, J., Harp, T. & Zhong, C. Effects of genetics and environment on the metabolome of commercial maize hybrids: A multisite study. *J. Agric. Food Chem.* **60,** 11498–11508 (2012).
57. Frank, T., Röhlig, R. M., Davies, H. V., Barros, E. & Engel, K.-H. Metabolite Profiling of Maize Kernels—Genetic Modification versus Environmental Influence. *J. Agric. Food Chem.* **60,** 3005–3012 (2012).
58. Argamasilla, R., Gómez-Cadenas, A. & Arbona, V. Metabolic and Regulatory Responses in Citrus Rootstocks in Response to Adverse Environmental Conditions. *J. Plant Growth Regul.* **33,** 169–180 (2014).
59. Lee, L. S. *et al.* Metabolomic analysis of the effect of shade treatment on the nutritional and sensory qualities of green tea. *J. Agric. Food Chem.* **61,** 332–338 (2013).
60. Silva, F. C. da. Manual de análises químicas de solos, plantas e fertilizantes. 107–190 (EMBRAPA, 2009).
61. Ward, J. H. Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* **58,** 236–244 (1963).

Acknowledgements

Author Contributions

B.L.S. carried out the experiments, analysed the data and wrote the manuscript; F.B.D.C. and R.E.E. conceived and designed the research work and corrected the manuscript. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Sampaio, B. L. *et al.* Effect of the environment on the secondary metabolic profile of *Tithonia diversifolia*: a model for environmental metabolomics of plants. *Sci. Rep.* **6**, 29265; doi: 10.1038/srep29265 (2016).

1 **Effect of the environment on the secondary metabolic profile of *Tithonia***

2 ***diversifolia*: a model for environmental metabolomics of plants.**

3 Bruno Leite Sampaio[1,2+]; RuAngelie Edrada-Ebel[2*+]; Fernando Batista Da Costa[1+].

4

5 <u>**Supplementary Information**</u>

6

7 **Method S1: Cultivation of *Tithonia diversifolia* by vegetative propagation**

8

9 The two groups of matrices of *Tithonia diversifolia* used in this study were cultivated

10 (*ex situ* condition) from seedlings obtained by vegetative propagation of cuttings from stems

11 collected from two distinct parent plants (*in situ* condition) separated by a distance of

12 approximately 500 km. Parent plant PP-1 was from the city of Ribeirão Preto, state of São

13 Paulo while PP-2 was from the city of Pires do Rio, state of Goiás. Both counties are located

14 within the Brazilian territory. The stem cuttings from PP-1 and PP2 were used to obtain the

15 seedlings cultivated at Ribeirão Preto and Pires do Rio, respectively.

16 To obtain the seedlings of *T. diversifolia*, stem cuttings containing at least two axillary

17 buds were transferred to plastic pots and partially covered with soil collected from the original

18 cultivation sites (Garden of Medicinal Plants of the School of Pharmaceutical Sciences of

19 Ribeirão Preto, University of São Paulo, Ribeirão Preto, and Fazenda Santo Antonio, county

20 of Pires do Rio). Forty-eight pots of stem cuttings were prepared for each cultivation site. The

21 method of vegetative propagation was chosen to obtain the seedlings of *T. diversifolia* due to

22 the great capacity of this species to grow clonally[1], which ensures an intra-group genetic

23 homogeneity to the specimens of both groups of matrices.

24 The pots containing the seedlings were kept outdoors, so that they could receive

25 sunlight. Water was added once a day to ensure that the soil within the pots was always wet.

26 The seedlings obtained from the stem cuttings were kept in pots for two months until

27 complete rooting; and then were transferred to the soil of respective cultivation areas. Each

1   cultivation areas were divided into four quadrants, 12 seedlings of *T. diversifolia* were planted

2   in each quadrant, totaling to 48 specimens per area.

3   The seedlings were transferred to the cultivation areas and were kept for three

4   months under local environmental conditions to acclimatise. The first sample collection was

5   done after the emergence of the floral buds (April 2012). However, inflorescence samples

6   were obtained only from the second collection (May 2012), when they were full-grown.

7

8   **Method S2: UHPLC-DAD-(ESI)-HRMS and NMR (J-resolved) data fusion by**

9   **concatenation method**

10

11   The data obtained by UHPLC-DAD-(ESI)-HRMS and *J*-resolved NMR were combined

12   in one single data matrix using the concatenation method described by Forshed et al. (2007)[2]

13   prior to multivariate statistical analysis.

14   The two sets of non-scaled data were divided in two groups according to the type of

15   data (UHPLC-DAD-(ESI)-HRMS data – Block 1; and *J*-resolved NMR data – Block 2). Both

16   data blocks were scaled by the formula described in Equation (1), such that X_n represents

17   the chromatographic peak area value for UHPLC-DAD-(ESI)-HRMS data or signal intensity

18   for *J*-resolved data for each observation of a given variable, and $\Sigma\sigma_{block}$ is the sum of all the

19   standard deviations of each variable in each block. The scaled data ($\hat{x}_n$) were obtained by

20   dividing the value of each peak area (Block 1) or signal intensity (Block 2) by the sum of

21   standard deviations of the respective block.

$$\hat{x}_n = \frac{x_n}{\sum \sigma_{block}}$$

22

23   **(1).** Equation of variable scaling for data fusion by concatenation method.
24

1   After the scaling procedure, the data of the two blocks were arranged as a single data

2   matrix (for each respective plant part), merging both blocks in the same data sheet on an MS

3   Excel© software. The matrix obtained by the fusion of both type of data (LC-MS and NMR)

4   was named as concatenated data matrix and it was used to perform multivariate analysis.

5

6   **Table S1:** Macro and micronutrients levels in soils of the cultivation sites in the states of

7   Goiás (GO) and São Paulo (SP).

Sampling	P[*]	K[**]	Ca[**]	Mg[**]	S[*]	B[*]	Cu[*]	Fe[*]	Mn[*]	Zn[*]
GO-Apr/12	15,9	3,1	40,0	11,0	6,0	0,27	1,8	50,0	5,2	4,8
GO-Oct/12	12,0	2,9	32,0	11,0	6,0	0,24	1,3	23,0	4,9	2,2
GO-Apr/13	13,0	1,8	34,0	7,0	5,0	0,20	1,2	22,0	1,9	0,6
GO-Oct/13	24,0	3,2	27,0	9,0	5,0	0,22	2,0	38,0	5,7	2,3
SP-Apr/12	13,0	3,3	25,0	8,0	8,0	0,22	6,2	20,0	14,6	0,5
SP-Oct/12	12,0	2,8	27,0	9,0	7,0	0,24	4,1	10,0	10,7	0,4
SP-Apr/13	9,0	2,5	20,0	8,0	5,0	0,23	3,8	11,0	10,0	0,3
SP-Oct/13	7,0	2,6	22,0	7,0	6,0	0,20	5,2	14,0	13,6	0,6

8   [*] P, Fe, Mn, Cu, Zn, B e S expressed as mg/dm^3; [**] K, Ca, Mg expressed as mmolc/dm^3.

9

10

11   **Table S2:** Other indicators of the mineral composition of the soils of the cultivation sites in

12   the states of Goiás (GO) and São Paulo (SP).

| Sampling	pH (CaCl$_2$)	OM[*]	H+Al	CEC[**]	BS[***]
GO-Apr/12	5,8	29,0	19,0	73,0	74,0
GO-Oct/12	5,6	23,0	22,0	67,0	67,9
GO-Apr/13	6,0	22,0	20,0	63,0	68,1
GO-Oct/13	5,5	25,0	21,0	60,0	64,9
SP-Apr/12	5,2	30,0	31,0	67,0	54,0
SP-Oct/12	5,2	28,0	32,0	70,0	54,4
SP-Apr/13	5,1	27,0	34,0	65,0	47,5
SP-Oct/13	5,3	25,0	29,0	60,0	52,6

13   [*] OM expressed as g/dm^3; [**] CEC and H+Al expressed as mmolc/dm^3; BS expressed as percentage.

14   Legend: OM = Organic matter; H+Al = Value of potential acidity of soil; CEC = Cation exchange

15   capacity; BS = Base saturation of soil.
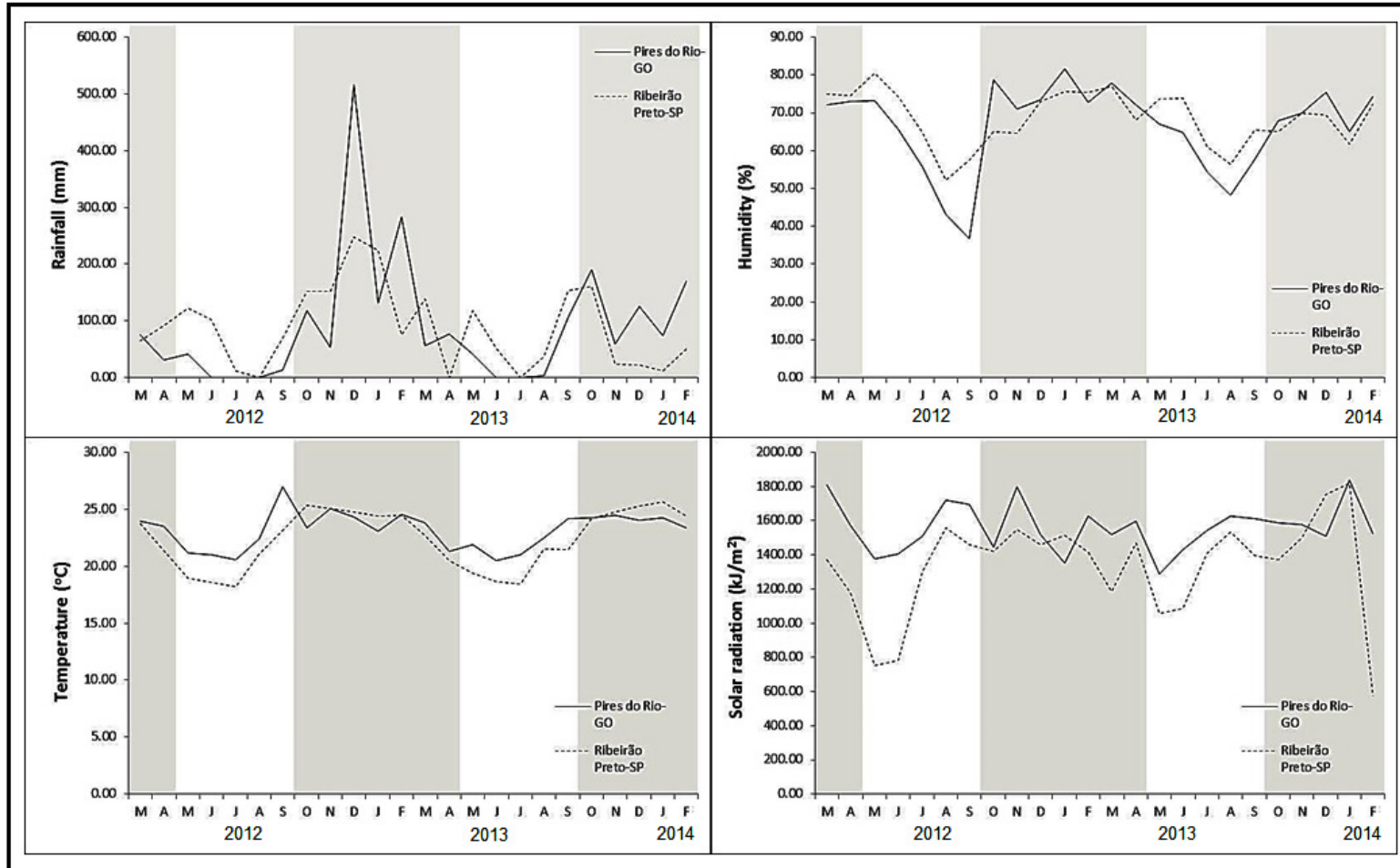
Figure S1: Climate data obtained from the both collection sites during the 24 months of study. Temperature, humidity and solar radiation were expressed as monthly average values and rainfall as accumulated rain per month. Continuous line = Pires do Rio-GO; Dashed line = Ribeirão Preto-SP. Gray areas in the graph indicate the rainy season and the white areas indicate the dry season.
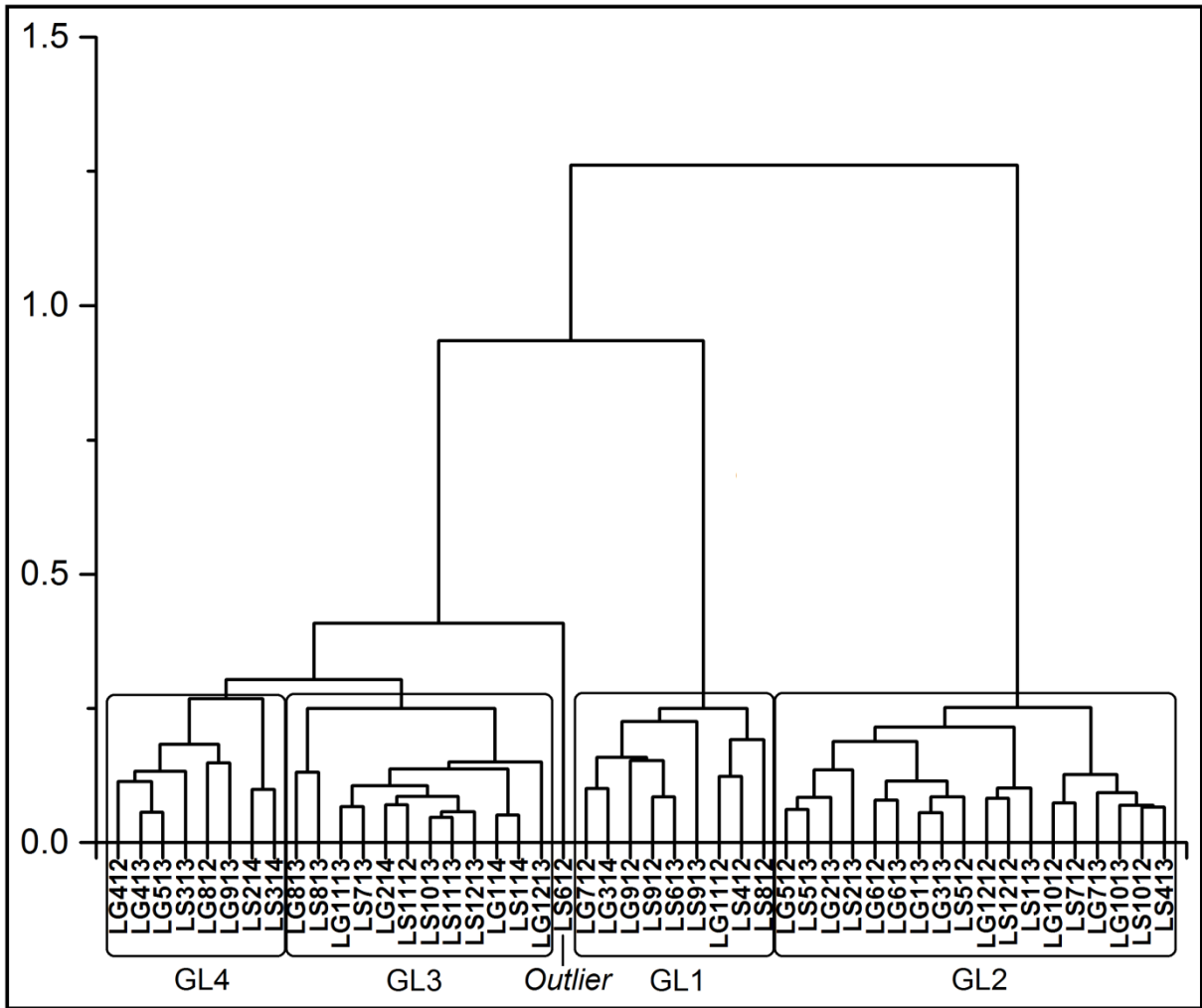
Figure S2: HCA dendrogram for concatenated data obtained from *T. diversifolia* leaf extracts. Box indicates grouping proposal. Legend: GL = Group of leaves extracts; LS = Leaves collected from São Paulo state; LG = Leaves collected from Goiás state; Number (eg. 112 for Jan 2012 or 1114 for Nov 2014)) represents the month (1 for January until 12 for December) and year (12 for 2012 until 14 for 2014).
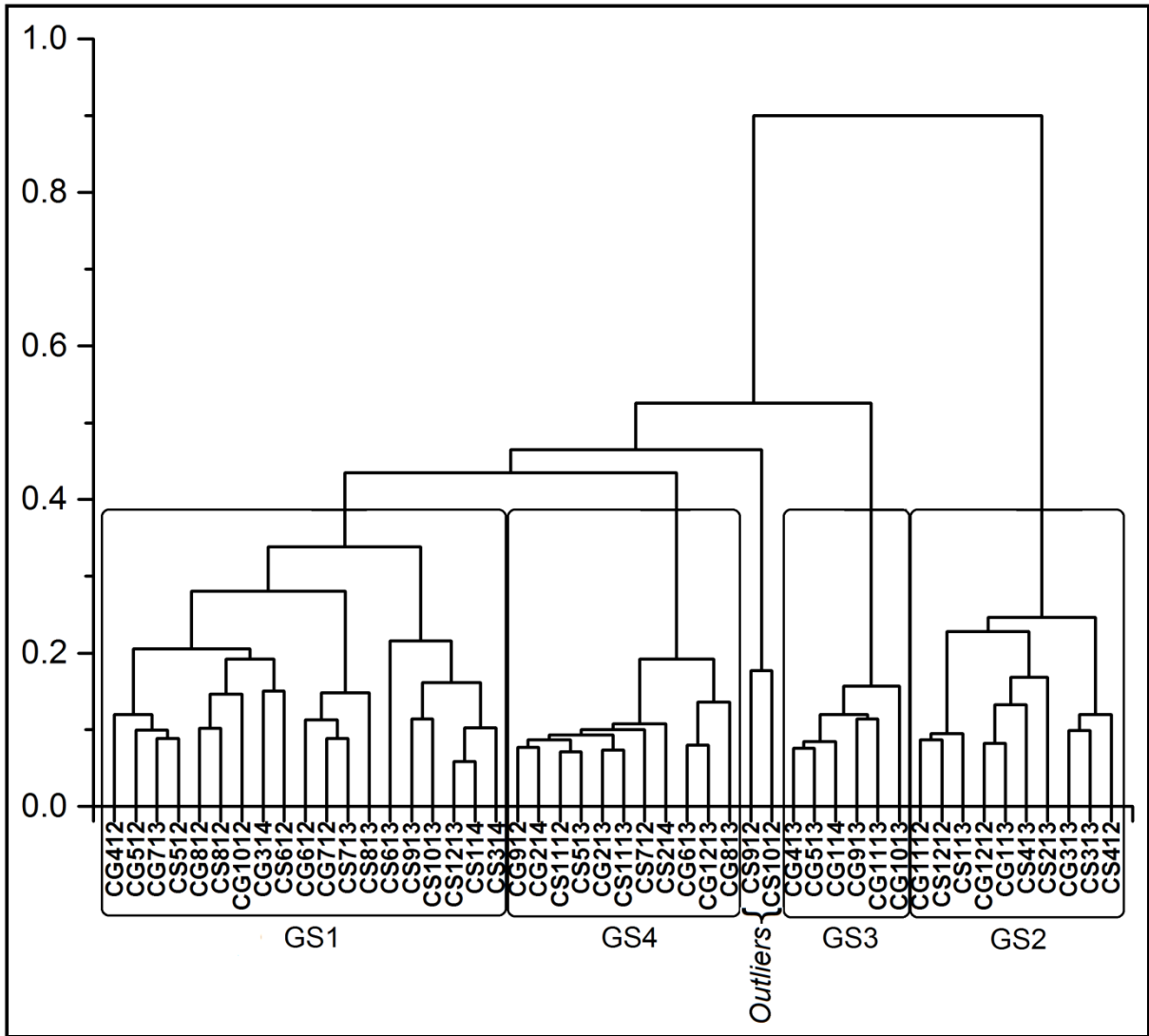
Figure S3: HCA dendrogram for the concatenated data obtained from *T. diversifolia* stem extracts. Box indicates grouping proposal. Legend: GS = Group of stems extracts; SS = Stems collected from São Paulo state; SG = Stems collected from Goiás state; Number (eg. 112 for Jan 2012 or 1114 for Nov 2014)) represents the month (1 for January until 12 for December) and year (12 for 2012 until 14 for 2014).
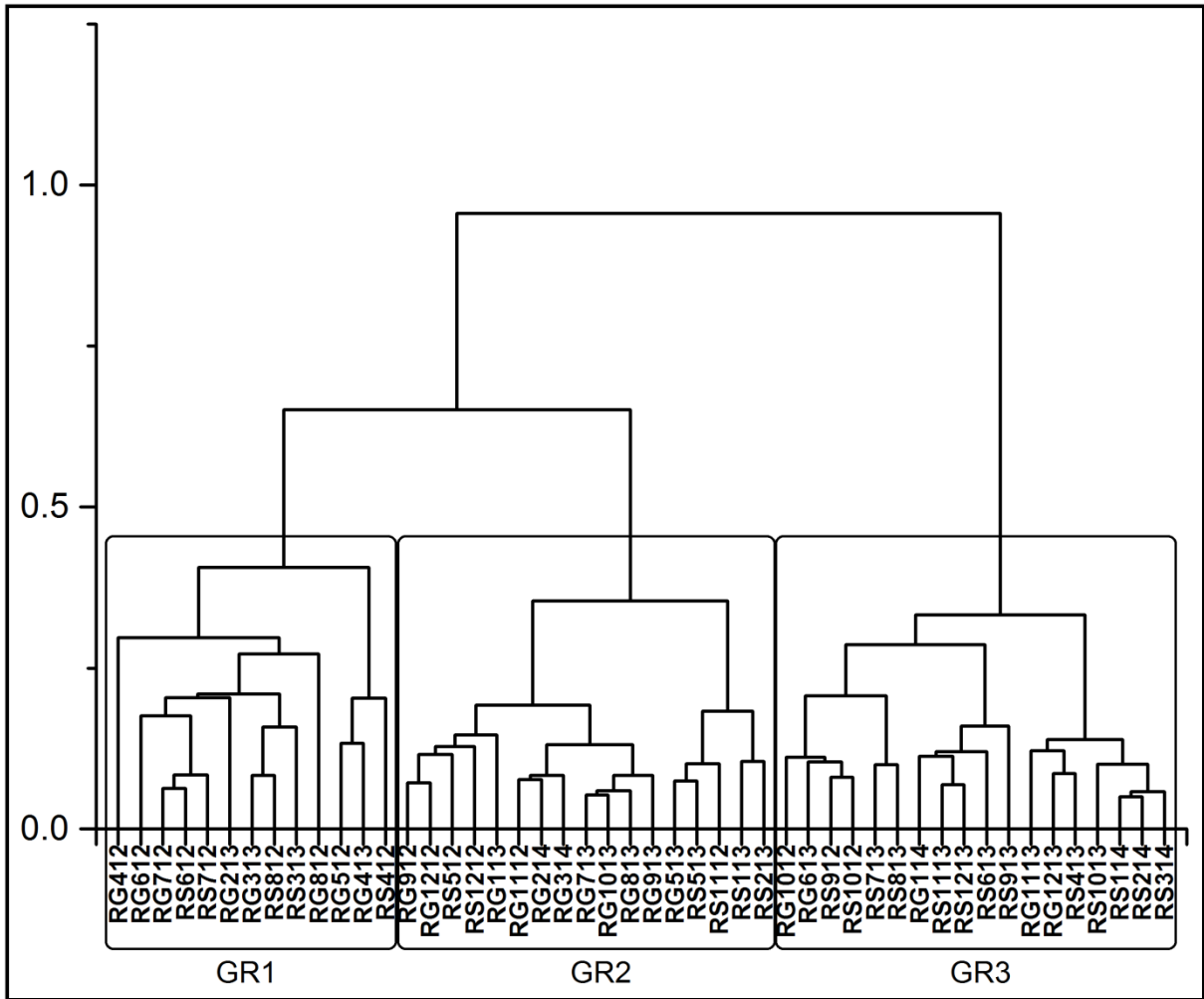
Figure S4: HCA dendrogram for the concatenated data obtained from *T. diversifolia* root extracts. Box indicates grouping proposal. Legend: GR = Group of roots extracts; RS = Roots collected from São Paulo state; RG = Roots collected from Goiás state; Number (eg. 112 for Jan 2012 or 1114 for Nov 2014)) represents the month (1 for January until 12 for December) and year (12 for 2012 until 14 for 2014).

Figure S5: HCA dendrogram for the concatenated data obtained from *T. diversifolia* root extracts Box indicates grouping proposal. Legend: GI = Group of inflorescences extracts; IS = Inflorescences collected from São Paulo state; IG = Inflorescences collected from Goiás state; Number (eg. 112 for Jan 2012 or 1114 for Nov 2014)) represents the month (1 for January until 12 for December) and year (12 for 2012 until 14 for 2014).

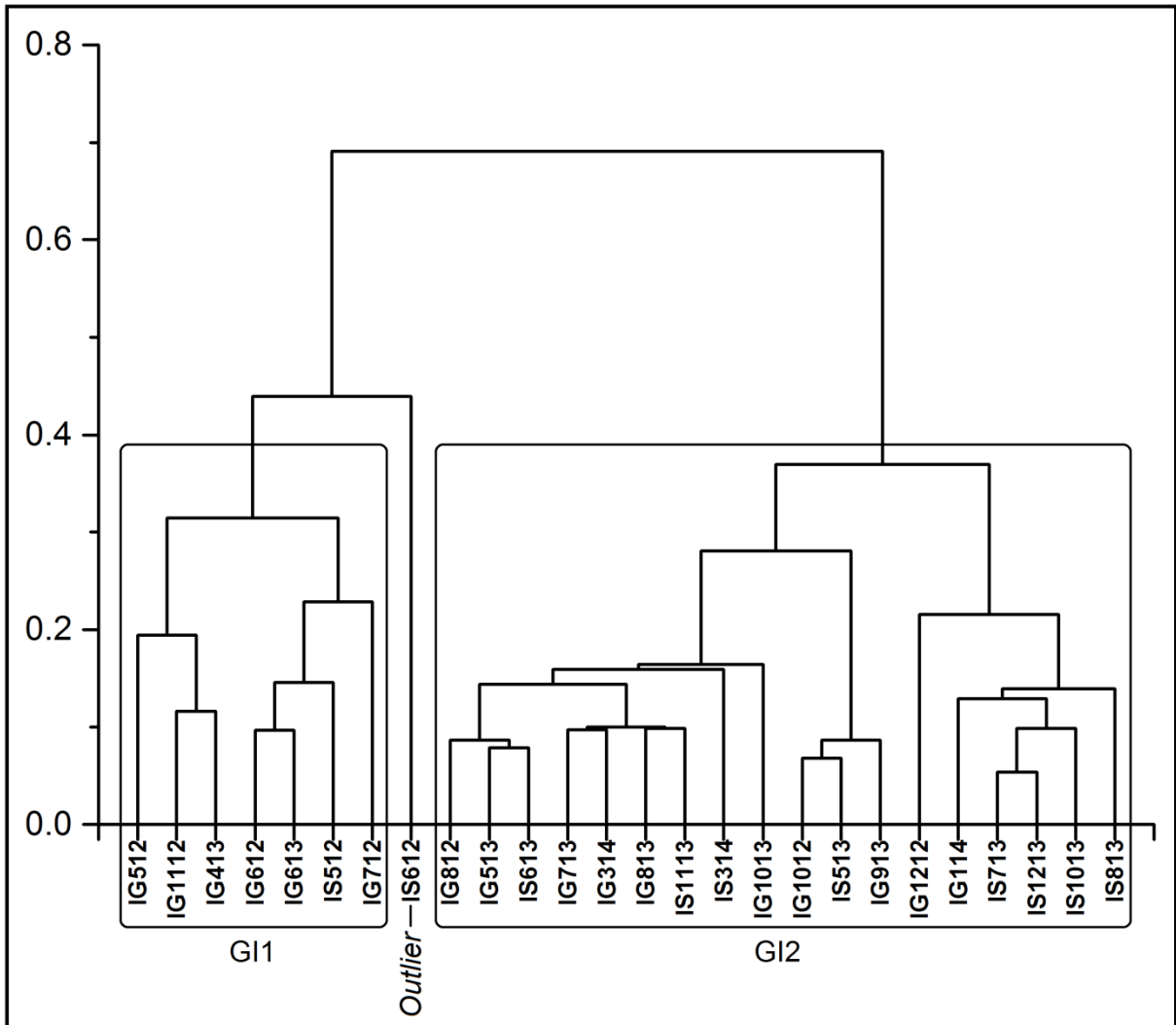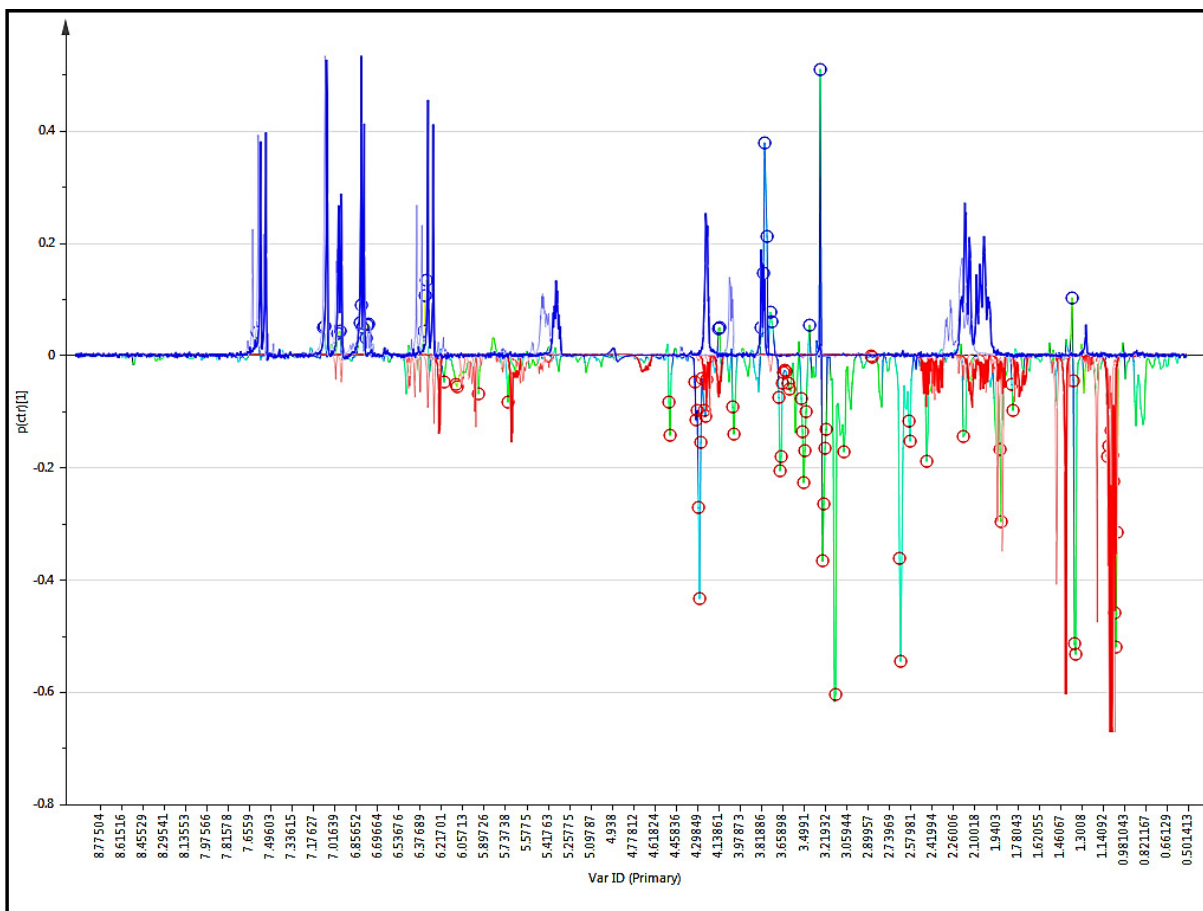**Figure S6:** S-line plot for *J*-resolved data of leaf extracts overlaying the ^{1}H-NMR spectra of the major compounds for groups GL1 (red) and GL3 (blue). The circles indicate the discriminant chemical shifts for the two proposed groups. The spectra in red were obtained from sesquiterpene lactones tagitinin A, tagitinin C and tagitinin C epoxide, and the spectra in blue belong to caffeoylquinic derivatives chlorogenic acid and 3,5-O-dicaffeoylquinic acid.
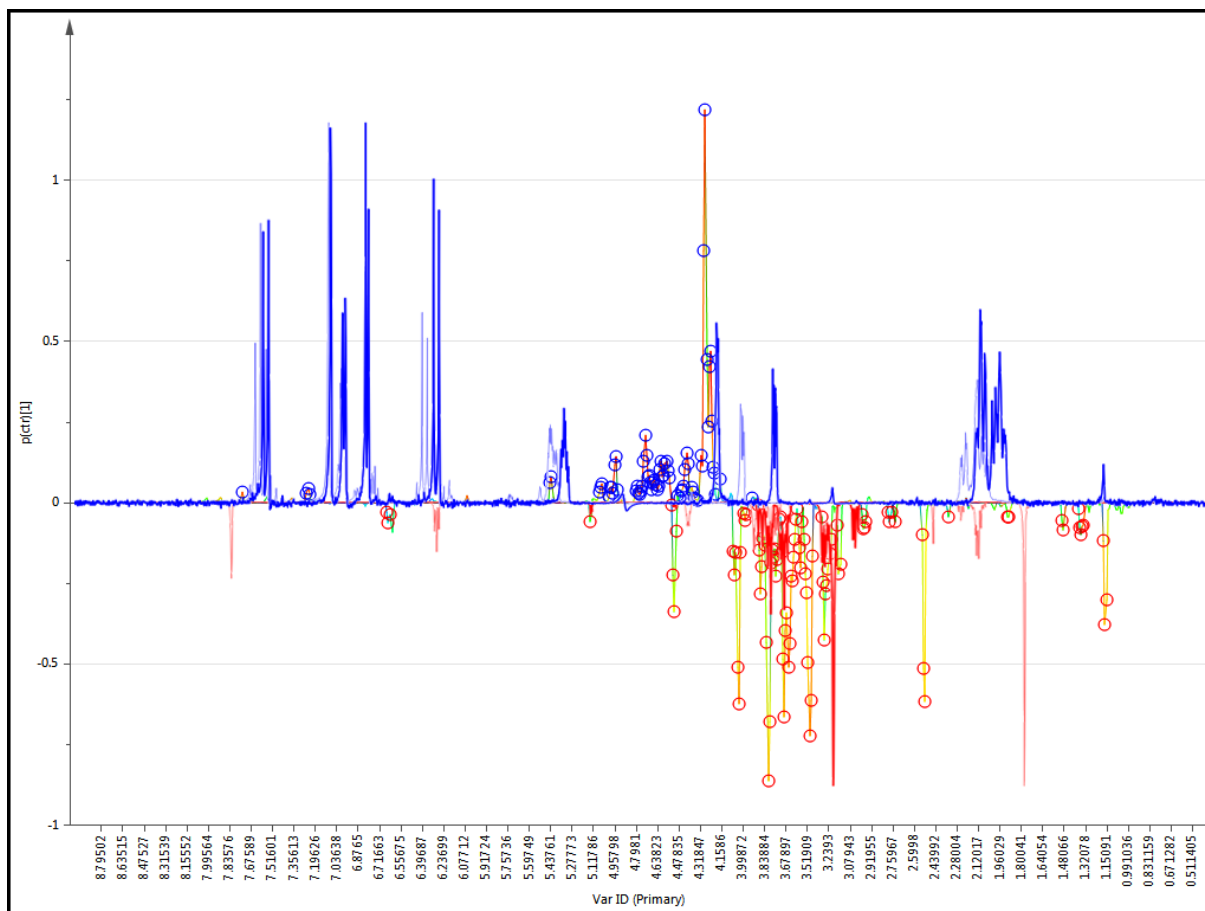
Figure S7: S-line plot for the *J*-resolved data of stem extracts overlaying the ^{1}H-NMR spectra of the major compounds for groups GS2 (red) and GS3 (blue). The circles indicate the discriminant chemical shifts for the two proposed groups. The spectrum in red represents the primary metabolite glucose and the spectra in blue belong to caffeoylquinic derivatives chlorogenic acid and 3,5-O-dicaffeoylquinic acid.
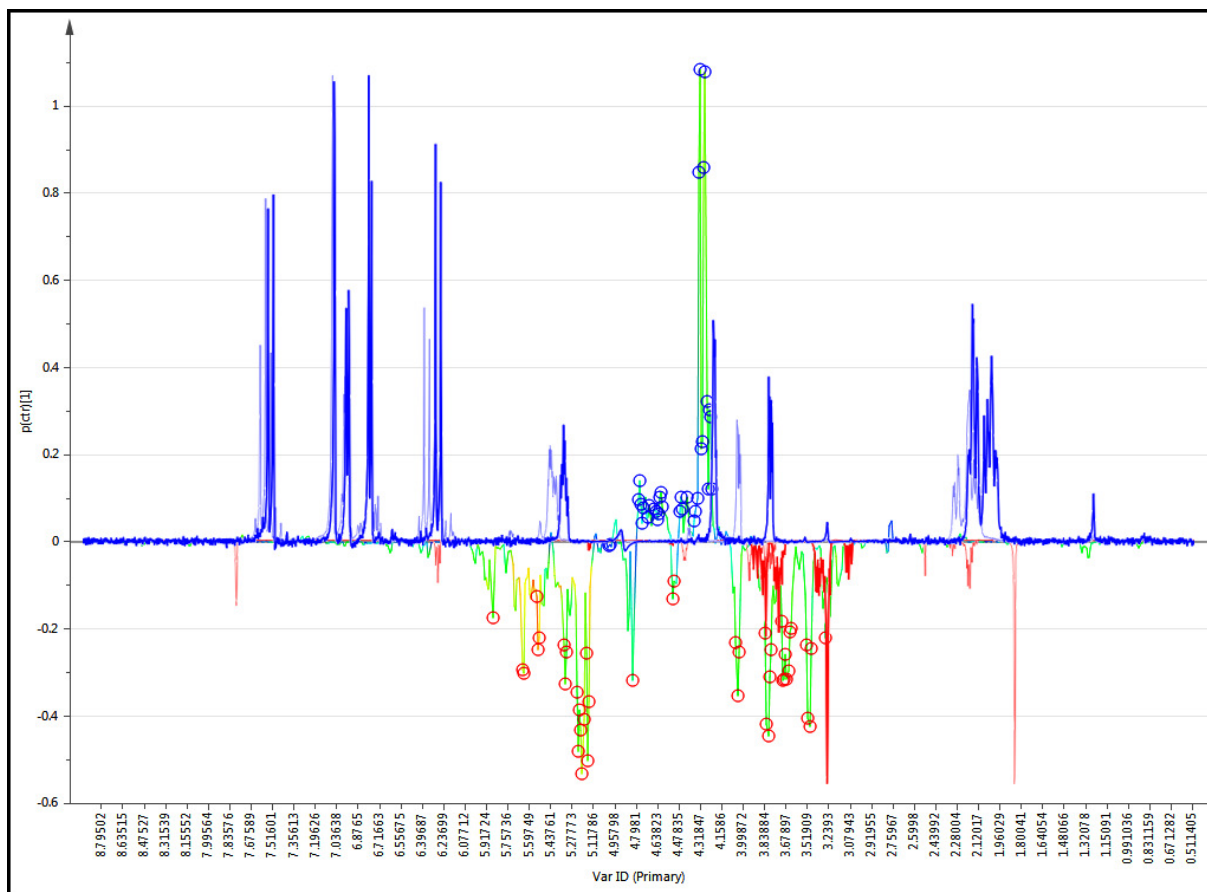
Figure S8: S-line plot for *J*-resolved data of root extracts overlaying the ^{1}H-NMR spectra of the major compounds for groups GR1 (red) and GR3 (blue). The circles indicate the discriminant chemical shifts for the two proposed groups. The spectra in red represent the primary metabolites glucose and thymidine, and the spectra in blue are caffeoylquinic derivatives, chlorogenic acid and 3,5-O-dicaffeoylquinic acid.

Figure S9: S-line plot for the J-resolved data of the extracts of inflorescences overlapped by the ^{1}H-NMR spectra of the main compounds for the groups GI1 (blue) and GI2 (red). The circles indicate the discriminant chemical shifts for the two proposed groups. The spectra in blue were obtained from the primary metabolite glucose and the sesquiterpene lactones tagitinin A, tagitinin C and tagitinin C epoxide, and the spectra in red from the caffeoylquinic derivatives chlorogenic acid and 3,5-O-dicaffeoylquinic acid.
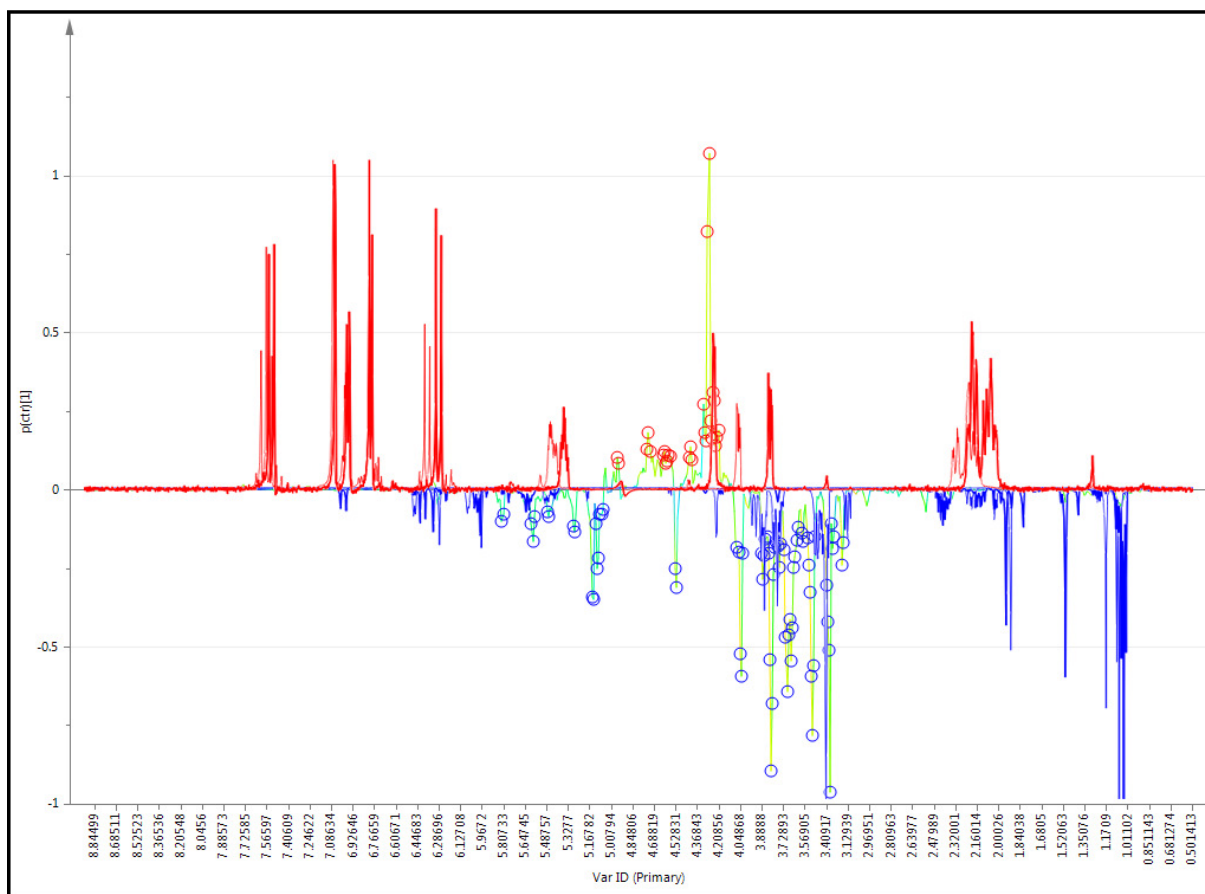
Table S3: Result of the dereplication step performed with the extracts of *T. diversifolia* by UHPLC-DAD-(ESI)-HRMS.

ID	Molecular Formula	Monoisotopic Mass	Suggested metabolite class or compound	RT[1] (min)	Additional Information[2]	Plant Part
1	$C_6H_{12}O_7$	196.0575	Hamamelonic acid (*D*-form)	1.80	-	L; S; R; I
2	$C_5H_{13}O_7P$	216.0395	2-Methyl-1,2,3,4-butanetetrol; (2S,3R)-form, 4-Phosphate	1.80	Intermediate in the MEP pathway	L; S; R; I
3	$C_6H_{12}O_6$	180.0634	Sugar	1.84	-	L; S; R; I
4	$C_7H_{12}O_6$	192.0634	Quinic acid	1.84	-	L; S; R; I
5	$C_7H_{12}O_7$	208.0578	4-*O*-Methyl-*D*-glucuronic acid	1.99	-	L; S; R; I
6	$C_6H_8O_7$	192.0265	Citric acid	2.07	-	L; S; R; I
7	$C_7H_6O_4$	154.0266	Protocatechuic acid	6.05	Compared with a pure standard compound	L; S; I
8	$C_{16}H_{18}O_9$	354.0951	*Trans*-cinnamic acid derivative	6.30	UV = 324, 301sh, 247 nm	L; S; R; I
9	$C_7H_{12}O_6$	192.0367	Quinic acid	8.30	Resulting from the fragmentation of chlorogenic acid	L; S; R; I
10	$C_{16}H_{18}O_9$	354.0951	Chlorogenic acid (5-*O*-caffeoylquinic acid)	8.32	Compared with a pure standard compound	L; S; R; I
11	$C_{16}H_{18}O_9$	354.0951	*Trans*-cinnamic acid derivative	8.77	UV = 326, 300sh, 221 nm	L, S; R; I
12	$C_{24}H_{22}O_{14}$	534.1018	*Trans*-cinnamic acid derivative	9.77	UV = 323, 300sh, 248 nm	L, S; R; I
13	$C_9H_8O_4$	180.0415	Caffeic acid	9.80	Compared with a pure standard compound	L, S; R; I
14	$C_{14}H_{16}O_8$	312.0845	4-*O*-caffeoyl-2-*C*-methyl-*D*-threonic acid	11.90	Compared with a pure standard compound	L

15	$C_{27}H_{30}O_{16}$	610.1537	Rutin	12.90	Compared with a pure standard compound	I
16	$C_{21}H_{20}O_{12}$	464.0960	Isoquercitrin	13.75	UV = 364, 342sh, 260 nm	L; I
17	$C_{22}H_{22}O_{13}$	494.1060	Glycosylated flavonoid	14.00	UV = 354, 347, 293, 258 nm	L
18	$C_{25}H_{24}O_{12}$	516.1268	4,5-*O*-dicaffeoylquinic acid	14.80	Compared with a pure standard compound	L; S; R; I
19	$C_{25}H_{24}O_{12}$	516.1268	3,5-*O*-dicaffeoylquinic acid	15.40	Compared with a pure standard compound	L; S; R; I
20	$C_{33}H_{28}O_{17}$	696.1336	*Trans*-cinnamic acid derivative	15.50	UV = 328, 303sh, 247 nm	S; R
21	$C_{21}H_{20}O_{11}$	448.1010	Quercitrin	15.65	Compared with a pure standard compound	L; R; I
22	$C_{25}H_{24}O_{12}$	516.1273	Dicaffeoylquinic derivative	15.90	UV = 327, 300, 253 nm	L; S; R; I
23	$C_{24}H_{22}O_{15}$	550.0967	Glycosylated flavonoid	16.00	UV = 368, 348, 255 nm	L; I
24	$C_{25}H_{24}O_{12}$	516.1268	3,4-*O*-dicaffeoylquinic acid	16.60	Compared with a pure standard compound	L; S; R; I
25	$C_{33}H_{28}O_{17}$	696.6340	*Trans*-cinnamic acid derivative	16.90	UV = 328, 304sh, 245, 216	R
26	$C_{33}H_{28}O_{17}$	696.1323	*Trans*-cinnamic acid derivative	17.30	UV = 328, 303sh, 248	S; R
27	$C_{43}H_{38}O_{19}$	858.1656	*Trans*-cinnamic acid derivative	20.30	UV = 328, 304sh, 244	S; R
28	$C_{15}H_{10}O_{6}$	286.0480	Luteolin	20.90	UV = 346, 266, 254	L; S; I
29	$C_{15}H_{10}O_{7}$	302.0428	Quercetin	21.02	Compared with a pure standard compound	I
30	$C_{34}H_{37}N_{3}O_{7}$	599.2637	Spermidine; *N'*-(3,4-Dihydroxycinnamoyl), *N,N''*-	21.10	UV = 306, 294	I

			bis(4-hydroxycinnamoyl)			
31	$C_{16}H_{12}O_7$	316.0583	Nepetin	21.20	UV = 346, 271, 255	L; S; I
32	$C_{16}H_{12}O_7$	316.0583	Flavonoid	21.30	UV = 346, 269, 256 nm	L; S; I
33	$C_{19}H_{26}O_7$	366.1679	Tagitinin B or 4,8,10-Trihydroxy-3-oxo-11(13)-guaien-12,6-olide; (1α,4β,5α,6α,8β,10β)-form, 8-O-(2-Methylpropanoyl)	21.30	-	L; S; I
34	$C_{19}H_{28}O_7$	368.1835	Tagitinin A ou 2α-hydroxytirotundin	21.80	-	L; S; I
35	$C_{34}H_{37}N_3O_6$	583.2682	Spermidine; *N,N',N''*-Tris(4-hydroxy-*E*-cinnamoyl)	22.40	UV = 306, 294	I
36	$C_{19}H_{28}O_7$	368.1835	Tagitinin A or 2α-hydroxytirotundin	22.70	-	L; S; I
37	$C_{19}H_{26}O_7$	366.1679	Tagitinina B or4α,10α-dihydroxy-8β-isobutyroyloxy-3-oxoguai-11(13)-en-6α,12-olide	22.70	-	L; S; I
38	$C_{46}H_{50}N_4O_9$	802.3588	Monocaffeoyl-tri-p-coumaroyl spermine	23.05	UV = 305, 297	I
39	$C_{15}H_{10}O_5$	270.0525	Apigenin	23.45	UV = 338, 266 nm	L
40	$C_{16}H_{12}O_6$	300.0634	Hispidulin	23.80	UV = 335, 273 nm	L; S; R
41	$C_{46}H_{50}N_4O_8$	786.3629	Spermine; *N1,N5,N10,N14*-Tetrakis(4-hydroxy-E-cinnamoyl)	24.00	UV = 305, 297	I
42	$C_{18}H_{32}O_5$	328.2252	Phytoprostane *F1* type I	24.13	UV = 203 nm	L; S; I
43	$C_{19}H_{26}O_7$	366.1679	Tagitinin B or 4α,10α-dihydroxy-8β-isobutyroyloxy-3-oxoguai-11(13)-en-6α,12-olide	24.77	-	L; S; I

#	Formula	Mass	Name	RT	Additional Information	Tissue
44	$C_{19}H_{24}O_6$	348.1573	Tagitinin C	24.90	Compared with a pure standard compound	L; S; I
45	$C_{19}H_{26}O_6$	350.1729	Tagitinin E ou diversifolin	24.90	-	L; S; I
46	$C_{20}H_{28}O_7$	380.1835	2-O-methyltagitinin B ou 3,10-Epoxy-1,3,8-trihydroxy-4,11(13)-germacradien-12,6-olide; (1β,3α,4Z,6α,8β)-form, 3-Me ether, 8-O-(2-methylpropanoyl)	25.30	-	L; S; I
47	$C_{12}H_{20}O_4$	228.1359	Hydroxylated fatty acid	25.30	-	L
48	$C_{18}H_{34}O_5$	330.2409	Hydroxylated fatty acid	25.50	-	L; S; R; I
49	$C_{19}H_{26}O_7$	366.1680	STL	26.50	-	L; I
50	$C_{20}H_{26}O_6$	362.1729	STL	27.30	-	L; S; I
51	$C_{20}H_{26}O_6$	362.1731	STL	27.60	UV = 223 nm	L; I
52	$C_{18}H_{32}O_4$	312.2303	Hydroxylated fatty acid	30.40	UV = 224 nm	S; R; I
53	$C_{18}H_{30}O_3$	294.2198	Fatty acid	30.80	-	L; R; I
54	$C_{27}H_{46}O_9$	514.3142	Triterpene or fatty acid	31.30	UV = 227 nm	L; R
55	$C_{20}H_{40}O_4$	344.2926	Fatty acid	34.60	-	L; R. I

[1]RT: Retention time of the metabolites in minutes. [2]Additional Information: Additional data about the metabolites useful for the dereplication step, such as the maximum absorbance in the UV, metabolites identified by comparison with a pure standard compound and other information.

1 **References**

2

3 1. Sun, W., Chen, G. & Wang, S. Characteristics of *Tithonia diversifolia*: an alien invasive

4   plant in Yunnan, south-west China. in *3rd Glob. Bot. Gard. Congr.* **2,** 1–7 (Botanic

5   Gardens Conservation International, 2007).

6 2. Forshed, J., Idborg, H. & Jacobsson, S. P. Evaluation of different techniques for data

7   fusion of LC/MS and 1H-NMR. *Chemom. Intell. Lab. Syst.* **85,** 102–109 (2007).