Journal of Advanced Research 34 (2021) 123–136

Contents lists available at ScienceDirect

Journal of Advanced Research

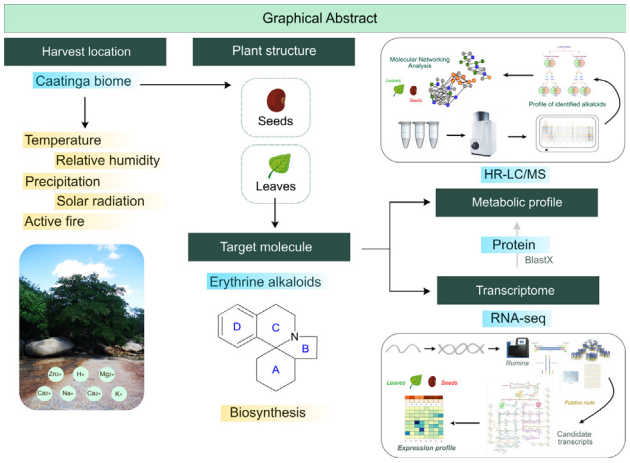
journal homepage: www.elsevier.com/locate/jare

Erythrina velutina Willd. alkaloids: Piecing biosynthesis together from transcriptome analysis and metabolite profiling of seeds and leaves

Daisy Sotero Chacon [a], Taffarel Melo Torres [b], Ivanice Bezerra da Silva [a], Thiago Ferreira de Araújo [a], Alan de Araújo Roque [c], Francisco Ayrton Senna Domingos Pinheiro [a], Denise Selegato [d], Alan Pilon [e], Fernanda Priscila Santos Reginaldo [a], Cibele Tesser da Costa [f], Johnatan Vilasboa [f], Rafael Teixeira Freire [g], Eduardo Luiz Voigt [h], José Angelo Silveira Zuanazzi [i], Renata Libonati [j], Julia Abrantes Rodrigues [j], Filippe Lemos Maia Santos [j], Kátia Castanho Scortecci [h], Norberto Peporine Lopes [e], Leandro De Santis Ferreira [a], Leandro Vieira dos Santos [k], Alberto José Cavalheiro [l], Arthur Germano Fett-Neto [f], Raquel Brandt Giordani [a,*]

[a] *Department of Pharmacy, Federal University of Rio Grande do Norte (UFRN), Natal, RN, Brazil*
[b] *Bioinformatics, Biostatistics and Computer Biology Nucleus, Rural Federal University of the Semiarid, Mossoró, RN, Brazil*
[c] *Institute for Sustainable Development and Environment, Dunas Park Herbarium, Natal, RN, Brazil*
[d] *Magnetic Resonance Center (CERM), University of Florence, Sesto Fiorentino, FI, Italy*
[e] *Faculty of Pharmaceutical Sciences of Ribeirão Preto, University of São Paulo (FCFRP-USP), Ribeirão Preto, SP, Brazil*
[f] *Plant Physiology Laboratory, Center for Biotechnology and Department of Botany, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil*
[g] *Signal and Information Processing for Sensing Systems, Institute for Bioengineering of Catalonia (IBEC), Barcelona Institute of Science and Technology, Barcelona, Spain*
[h] *Department of Cell Biology and Genetics, Center for Biosciences, Federal University of Rio Grande do Norte, Natal, RN, Brazil*
[i] *Laboratory of Pharmacognosy, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil*
[j] *Department of Meteorology, Institute of Geoscience, Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil*
[k] *Brazilian Biorenewables National Laboratory (LNBR), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas, SP, Brazil*
[l] *Chemistry Institute, São Paulo State University (UNESP), Araraquara, SP, Brazil*

GRAPHICAL ABSTRACT



Peer review under responsibility of Cairo University.

* Corresponding author at: Gal. Gustavo Cordeiro de Farias Street, s/n, 59012-570 Natal, RN, Brazil.
E-mail address: raquebg@hotmail.com (R.B. Giordani).

https://doi.org/10.1016/j.jare.2021.01.017

ARTICLE INFO

Article history:
Received 9 November 2020
Revised 1 January 2021
Accepted 30 January 2021
Available online 10 February 2021

Keywords:
Erythrina velutina
Transcriptome
Targeted metabolite profile
Benzylisoquinoline Alkaloids
Caatinga

ABSTRACT

Introduction: Natural products of pharmaceutical interest often do not reach the drug market due to the associated low yields and difficult extraction. Knowledge of biosynthetic pathways is a key element in the development of biotechnological strategies for plant specialized metabolite production. *Erythrina* species are mainly used as central nervous system depressants in folk medicine and are important sources of bioactive tetracyclic benzylisoquinoline alkaloids (BIAs), which can act on several pathology-related biological targets.

Objectives: In this sense, in an unprecedented approach used with a non-model Fabaceae species grown in its unique arid natural habitat, a combined transcriptome and metabolome analyses (seeds and leaves) is presented.

Methods: The Next Generation Sequencing-based transcriptome (*de novo* RNA sequencing) was carried out in a NextSeq 500 platform. Regarding metabolite profiling, the High-resolution Liquid Chromatography was coupled to DAD and a micrOTOF-QII mass spectrometer by using electrospray ionization (ESI) and Time of Flight (TOF) analyzer. The tandem MS/MS data were processed and analyzed through Molecular Networking approach.

Results: This detailed macro and micromolecular approach applied to seeds and leaves of *E. velutina* revealed 42 alkaloids, several of them unique. Based on the combined evidence, 24 gene candidates were put together in a putative pathway leading to the singular alkaloid diversity of this species.

Conclusion: Overall, these results could contribute by indicating potential biotechnological targets for modulation of erythrina alkaloids biosynthesis as well as improve molecular databases with omic data from a non-model medicinal plant, and reveal an interesting chemical diversity of *Erythrina* BIA harvested in Caatinga.

Introduction

Plants are sessile organisms capable of producing over 200,000 specialized metabolites (also known as natural products of secondary metabolites) which represent chemical strategies for survival under environmental stresses and for interaction with other organisms [1]. Chemical diversity afforded by the specialized metabolism is a major source of bioactive molecules from plants, e.g., which have been used in folk medicine since ancient times, and it resonates with target molecules of biotechnological interest [2]. Specialized plant metabolites themselves or structures derived therefrom have been essential for medicinal chemistry and drug discovery [3].

The use of an integrative approach to understand a species in the context of its environment can accelerate and improve the identification of new chemical entities. During the course of evolution, plants have gained the ability to produce chemical compounds that are active against several threats, including biotic and abiotic stresses. This process has been driven by selective pressures inherent to the wide array of plant habitats, thereby contributing to the diversity of plant specialized metabolites. This is particularly true for regions characterized by extreme environments, such as the Caatinga, a Brazilian unique and understudied biome with semiarid edaphoclimatic features. Plants of this biome developed many metabolic adaptations to survive and resist in the face of several stress factors. The specialized metabolism fine-tunes plant fitness to habitat, requiring precise control of metabolic flux through specific pathways derived from central metabolism, and contributing to overall ecosystem functional dynamics. Responses to abiotic stresses encompass complex metabolite regulatory networks which can only be dissected using integrated omics analyses [4]. This scenario leads to the hypothesis that plants of the Caatinga constitute an important reservoir of novel bioactive molecules with an array of potential applications.

Biosynthetic pathways and metabolic regulation at different levels are complex issues that have been the subject of investigation mainly in model plants. Although fundamental, these studies fall short of providing a more representative picture of the variability and molecular wealth of all plants [5]. Targeted genome or transcriptome sequencing of medicinal plant species could pave the way for synthetic biology approaches aiming at producing specific bioactive compounds in microbial platforms, as has been done with artemisinin, a potent anti-malarial drug of *Artemisia annua* [6]. The expansion of knowledge will enhance our ability to modify plants and make rational use of plant biodiversity [5,7].

A major research effort involves molecular and biochemical integrated omic evaluations beyond plant model systems, as well as gauging the effective potential of these medicinal species to afford new bioactive natural products. As part of this effort, seeds and leaves of *Erythrina velutina* Willd., a species widely used in folk medicine were harvested for evaluation. The species was also chosen considering its family, Fabaceae, an alkaloid rich taxon. Alkaloids are organic nitrogenous bases, an important group of diversely distributed, chemically, biologically, and commercially significant natural products, whose nitrogen atoms originate from an amino acid. Their protective functions *in planta* range from defense against pathogens and herbivores to abiotic stress mitigation, e.g. drastic changes in light irradiance, water availability and temperature [8]. *E. velutina* is a native species of the Caatinga [9] having sedative, anticonvulsant, neuroprotective, and anxiolytic activities, mainly acting on central nervous system (CNS) targets [10,11].

This work reports on the combined omic approach of Next Generation Sequencing (NGS)-based transcriptome dataset production and targeted metabolic profiling of a non-model Fabaceae species grown in its natural Caatinga habitat. The results revealed the important chemodiversity of the species and provided insights on the biosynthetic pathway of its unique isoquinoline alkaloids.

Methodology

Plant material

Natural populations of *Erythrina velutina* Willd, from four locations were selected: two in the city of Acari (location 1: 6°27′44.5″S 36°38′30.6″W and location 2: 6°28′19.5″S 36°38′32.5″W at 270 m

above sea level) and two in the city of Jardim do Seridó (location 1: 6°33′58.2″S 36°43′00.6″W and location 2: 6°33′48.6″S 36°43′11.0″W at 218 m above sea level). Leaves and seeds were collected separately (Fig. S1) and samples composed of 5 to 10 trees for each population were used in a multi-omic approach (transcriptome and targeted metabolite profiling) organized in pools 1 to 8 (Table 1), in which pools 1 to 4 corresponded to seed samples (pool 2 and pool 3 - two harvest locations in the city of Acari; pool 1 and pool 4 - two harvest locations in the city of Jardim do Seridó) and 5 to 8 corresponded to leaf samples (pool 6 and pool 7 - two locations in the city of Acari; pool 5 and pool 8 - two locations in the city of Jardim do Seridó) in the state of Rio Grande do Norte, Brazil, in March 2018, during the dry season at daytime period. The following criteria were adopted for selecting samples: dark green fully expanded apical leaves from top branches with no signs of herbivory, located at the ends of the branches at the top of the tree most exposed to solar radiation, and red seeds attached to ripe pods still hanging from the trees. Samples were quickly washed with distilled water, immediately frozen in liquid nitrogen at the collection site and then stored at −80 °C until extraction. A voucher specimen was deposited at the Herbarium of the Federal University of Rio Grande do Norte, Brazil, under the reference number UFRN16079. The authorization for harvesting the plant material was granted by SISBIO (327493) and the permission to access the Brazilian genetic heritage was provided by SISGEN (A8E4663). The experimental methodology is summarized in Fig. 5.

Metabolite profiling

Metabolite extraction

Frozen seeds, and leaves of *E. velutina* were ground with mortar and pestle in liquid nitrogen, followed by lyophilization. Samples containing 10 mg of pulverized tissue were mixed with 100 μL of ethanol:water [70:30 (v/v)] by vortex for 10 s, incubated for 30 min in an ultrasound bath (40 kHz, 200 W), and then filtered through 0.22 μm PTFE membrane. The extraction procedure was repeated three times for each sample and 3.0 μL of each, at 10 mg/mL, was injected in the LC-MS/MS system and analyzed in the NMR spectrometer.

NMR analysis

Nuclear Magnetic Resonance (NMR) 2D ^{1}H- ^{1}H TOCSY analyses (three independent experiments) were performed with 10 mg of the extract in DMSO d_6 (1 mL). The spectra were acquired at 300 K in an Advance 600 spectrometer (Bruker Biospin, Coventry, UK) using a 5 mm TCI probe and the dipsi2esgpph pulse sequence was employed to obtain spectra 2D TOCSY. Each spectrum consisted of 192 scans of 2048 memory points in F2 and 256 in F1 with a spectral width of 16 ppm. All the spectra FIDs were automatically Fourier transformed after the application of an exponential window function with a line broadening of 0.3 Hz. Phasing and baseline correction were carried out with the instrument software. The chemical shifts for all spectra were referenced for TMS (tetramethylsilane) at δ 0.00 as internal reference. The metabolite identifications were performed using an integrated NMR platform. The *in-house* database of ^{1}H NMR data was compiled the alkaloids previously identified and reported for this plant genera. The software detects the substances by comparing the peaks of the standard compounds with the peaks presented in the raw extract. The thresholds of similarity to find a similar peak were 0.01 ppm, 0.05 ppm and 0.1 ppm. The 2D ^{1}H-TOCSY NMR was processed in the MestreNova 10.0 according to the parameters described above and then exported to MATLAB R2015. The noise was removed by a threshold line consisting of twice the intensity of the noise. The peaks were extracted using the FastPeakFind function and the pattern recognition function was applied to the peaks obtained in the F1 dimension providing higher resolution when compared with the F2 dimension. Finally, the results of a substance found were given by the percentage of values, which consists in the number of peaks found in the raw extract and the total number of peaks presented in the standard compound.

HR-LC-DAD-ESI-MS/AutoMS data acquisition

Separation and detection of medium-polar compounds were performed in a Shimadzu Class-LC 10, CBM 20A controller and a micrOTOF-QII (Bruker Daltonics) mass spectrometer using electrospray ionization (ESI) and Time of Flight (TOF) analyzer. An XB C18 Kinetex column (100 × 2.1 mm, 2.6 μm, 100 Å), kept at 40 °C, flow rate of 0.400 mL/min, and UV-DAD detector operating between 190 and 400 nm were used. The chromatographic method was based on previous studies [12,13]. The mobile phase was composed of H_2O (A) and acetonitrile (B), both with 0.1% (v/v) formic acid using the following gradient: 5–35% B (23 min); 35.0–100% B (17 min), holding at 100% B for 8 min followed by reducing from 100 – 5% B (2 min) and maintained in the initial condition (5% B) for 10 min, totaling 60 min of analysis. Mixed test solutions (mixture of standards - amino acids, flavonoids, etc.) and quality controls (QC - mixture of samples – 40 μL each) were injected to ensure reproducibility of the method, instrument, and chromatographic conditions.

The HRMS experiments were carried out in a micrOTOF-QII mass spectrometer (Bruker Daltonics) in positive ionization mode. The acquisition parameters were capillary and cone voltage, 4500 V and 500 V, respectively, capillary temperature at 200 °C, N_2 as a drying gas at 4.5 bar and a flow of 9.0 L/min. The spectra were acquired between 100 and 1500 Da. For high resolution mass analysis instruments and experiments were calibrated using an internal standard Na-TFA solution (4 mg/mL) at the end of each run.

The MS/MS experiments were carried out in Auto-MS/MS mode (positive mode) selecting up to 5 precursor ions within a cycle of 3.0 s. An energy gradient of 20 to 50 eV was used for fragmentation of precursor ions. Initially, the blank samples were analyzed to identify interferences and contaminants from the instrument and the chromatographic column. These were then excluded from the Auto-MS/MS selection and fragmentation cycle by a function that allows their exclusions.

Table 1
Identification and location of collection sites.

Pool set	Pools		Harvest site	Geographical coordinates
	Seeds	Leaves		
A	p1	p5	Jardim do Seridó I	6°33′58.2″S 36°43′00.6″W
B	p2	p6	Acari I	6°27′44.5″S 36°38′30.6″W
C	p3	p7	Acari II	6°28′19.5″S 36°38′32.5″W
D	p4	p8	Jardim do Seridó II	6°33′48.6″S 36°43′11.0″W

MS/MS molecular networking analysis

To perform the Molecular Networking analysis, all MS/MS data were initially converted to mzXML format using MSConvert software and processed using MzMine package (java script) for mass detection and calibration, peak deconvolution, mass grouping (isotopic pattern), and spectral alignment [14]. Once all the data were processed using MzMine software (http://mzmine.github.io/), the spectra were exported in a text format (.mgf) and uploaded to the Global Natural Products Social Molecular Networking (GNPS) platform in order to initiate the calculation of similarity among spectra (cosine similarity score). In this procedure, the spectra are converted into vectors which are paired and compared by a dot product calculation. The cosine similarity score assigned to each pair of vectors ranges from 0 to 1, in which 0 represents completely different and 1 represents identical spectra. The data were filtered by removing all MS2 peaks within ± 17 u of the precursor *m*/*z* value. MS2 spectra were filtered by choosing only the top six peaks in the ± 50 u window throughout the spectrum. The data were clustered with MSCluster with a parent mass tolerance of 1.0 Da and a MS2 fragment ion tolerance of 0.3 Da to create consensus spectra (for consensus spectra, see https://ccms-ucsd.github.io/GNPSDocumentation/networking). The network was created by filtering the edges to have a cosine score above 0.65 and more than 6 matched peaks. The spectra in the network were searched against the GNPS spectral libraries (GNPS: https://gnps.ucsd.edu/ProteoSAFe/libraries.jsp), using the same setup as that for the input data. Network visualization was performed in Cytoscape 2.8.3 and 3.4.3 (https://cytoscape.org). To avoid misinterpretation of HPLC contaminants, blank (mobile phase) injections were uploaded as a distinct sample group on GNPS workflow and excluded from the networks.

In addition, the nodes were colored based on the plant organ (leaf and/or seed) of *Erythrina* and the thickness of the edge attributed to the cosine similarity scores (thickness from 0.95 to 1). Thicker lines were used to indicate greater similarity. Subnets were generated in Cytoscape from isolated parts of the molecular family, in order to improve the analysis of metabolic dynamics, provide more elucidation of molecules, better visibility of edge thickness, and node connectivity. The subnets were based on classification among the *Erythrina* alkaloid classes.

Transcriptome analysis

Total RNA extraction and cDNA library preparation

Total RNA was isolated from seeds and leaves using the Trizol™ Plus RNA Purification kit (Life Technologies) according to the manufacturer's protocol. To obtain high quality RNA, the samples were treated with the RNAse-Free DNAse set (Qiagen) and subsequently purified with the RNeasy Mini kit (Qiagen). RNA quality was assessed using the Bioanalyzer 2100 RNA Nano Chip (Agilent) and quantification was performed using the NanoDrop ND-1000 (NanoDrop) spectrophotometer and the Qubit (Thermo Fischer Scientific) fluorometer. All samples that had an RNA Integrity Number (RIN) value above 6.0 were used in the next steps. Each pool was composed of three RNA samples and a total of 0.5 μg of RNA was equally pooled for the preparation of cDNA libraries. The sample sets were indexed with the TruSeq Stranded mRNA Library Prep Kit according to the protocol (Illumina). The libraries were quantitated using the Kapa Library Quantification kit for the Illumina Genome Analyzer (Kapa Biosystems) platform on the Applied Biosystems ViiA7 machine (Life Technologies) according to the manufacturer's instructions.

RNA sequencing analysis and annotation

Total RNA was sequenced according to the standard protocols of the NextSeq 500 (Illumina) platform, generating 150 paired-end readings. Composite samples obtained from at least three specimens yielded four seed RNA pools (1–4) and four leaves RNA pools (5–8), in accordance with the above mentioned harvest scheme. After RNA sequencing, FastQ files were generated and transferred from the GENIAL facility at CEFAP-USP to the Bioinformatics, Biostatistics and Computational Biology Core-B3C at Federal University of Semiarid (UFERSA-RN-Brazil). The massive analysis was performed in the Computational Cluster of High-Performance Computing Center-NPAD at UFRN. Initially, rRNA removal was performed using the SortMeRNA software (v2.1) (https://bioinfo.lifl.fr/RNA/sortmerna/) for all readings that combine the Forward (R1) and Reverse (R2) strands. This step was necessary to improve the quality of the readings. Sequence quality was evaluated by the FastQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The reading libraries were trimmed by Trimmomatic (http://www.usadellab.org/cms/index.php?page = trimmomatic) to remove poor quality adapters and readings, in which the sliding window cut was 2:20.

The range of final readings was between 21 and 151 base pairs and phred up to 20. The FastQ files described above were used in a Dassembly in the Trinity software (https://github.com/trinityrnaseq/trinityrnaseq/wiki) because the *E. velutina* full genome sequence is not available yet. A k-mer size range 21 was used to obtain the best results for mapping genes from plant orthologs in the BUSCO software (https://busco.ezlab.org/) aiming at validating the assemblies generated in Trinity. This tool provides quantitative measures for the assessment of genome assembly, gene set, and transcriptome completeness, based on evolutionarily-informed expectations of gene content from near-universal single-copy orthologs selected from OrthoDB v9 (https://www.orthodb.org/v9/index.html) and crossing with the Unigene database. Additionally, the software Bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) was used to align sequencing reads with final assembled sequences aiming to validate the transcriptome.

The assembled transcripts were submitted to the InterProScan database strings (v5.30–69.0) (http://www.ebi.ac.uk/interpro/), which combines signatures from various databases (PROSITE, Pfam, PRINTS, ProDom, SMART, TIGRFAMs, PIRSF, SUPERFAMILY, GENE3D and PANTHER) for a primary classification. In this step, an identification parameter of at least 50% coverage and 70% identity of the functional domain was used before data input onto the platform of the aforementioned databases. After domain identification, transcripts containing sequences corresponding to candidate protein domains were selected to participate in the study of alkaloid biosynthesis according to data from public databases and from the literature. All identified sequences of *E. velutina* are "–like sequences", but the suffix was omitted for the sake of simplicity.

The sequences were subjected to a second run of BLASTx annotation against the 469,154 strings of UniProtKB/SWISS-PROT bank (https://www.uniprot.org/). A minimum of 30% identity, 50% coverage and 10^{-10} E-value were used as parameters, in addition to the manual analysis of the alignment, in which sequence size and coverage were checked for the presence of substrate/active site. All transcripts were quantified as the number of transcripts per million (TPM) by the Salmon software [15].

Climatic parameters

Trimestral values of temperature, accumulated precipitation, and accumulated surface net solar radiation in the area where plants were harvested were obtained by using the European Centre for Medium-Range Weather Forecasts' (ECMWF) ERA-Interim reanalysis [16]. For the same year, fire activity was analyzed based

on remote sensing information from the Suomi-National Polar-orbiting Partnership Visible Infrared Imaging Radiometer Suite (VIIRS) 375-m daily active fire product (VNP14IMG) from the National Atmospheric Space Agency (NASA).

Results and discussion

Targeted metabolite profiling and alkaloid diversity

Erythrina alkaloids can be classified into dienoid, alkenoid, and lactone subclasses, based on their core moiety [10,11], a conserved and characteristic chemical nucleus. The molecular diversity among them is attributed, except for the lactone ring, to the substitution patterns, such as –OH, $–CH_3$, $–CH_3O$ and glucosyl residues, usually found at C3, C8 and C11 of the alkaloidal main core skeleton. Herein, the targeted metabolic analysis was used in a metabolomic perspective to investigate the chemical diversity of the alkaloids in the crude extract, in a comparative way among leaves and seeds.

An initial analysis of the samples by 2D TOCSY-NMR ($^1H–^1H$) enabled us to compare the experimental data with an *in-house* database by including all the experimental ^1H NMR spectral data reported to the alkaloids previously identified in *Erythrina* spp. The analysis criteria were related to spectral similarity values based on the algorithm of pattern recognition between the *in-house* database and the experimental data. These data pointed out the occurrence of alkenoid and dienoid scaffolds even though the unequivocal identification of the alkaloids was not possible due the overlapping of chemical shifts present in different but similar molecules. Therefore, a more sensitive and orthogonal metabolomic approach was used in a complementary perspective. In order to obtain a more comprehensive analysis and to visualize structurally related molecules, an MS/MS Molecular Networking (MolNet) approach was employed as a dereplication strategy for metabolite annotation The MolNet protocol was based on cosine similarity score and fragmentation pattern with comparative targeted analysis, using the web-based platform Global Natural Products Social Molecular Networking (GNPS) [17], as well as an *in-house* database that compiles the previously reported mass spectral data on these alkaloids. The level of identification confidence in this study was 2 or 3 in a scale that varies from 1 to 4 [18,19] where it was possible to putatively annotated compounds or characterized compound classes. The 42 alkaloids detected **(**Fig. 1**)** were segregated in two clusters (both clusters have alkaloids of the dienoid and alkenoid subclass, but not lactonic) **(**Fig. 2**B)**; these, 25 of them (60%) were annotated **(Table S1)** and 17 (40%) were unannotated **(Table S11 and S12)**.

As the network produced by GNPS platform groups compounds with similar MS/MS spectra, the evaluation of MS/MS spectra could display information about Erythrina alkaloid scaffolds (dienoid, alkenoid, and lactonic), once each subclass has different fragmentation pathways [20]. Retro Diels-Alder (RDA) is noted usually in alkenoid subclass while substituents elimination in Ring A, as for example CH_3OH (32 u), are commonly noted in dienoid subclass due to the resonance produced that aids to stabilize the charge [20]. Surely, other Erythrina alkaloid fragmentations in the mass spectra were observed as water, indole and sugar elimination although they are not always related to a specific subclass [20]. The analysis of these characteristic fragments of each subclass and the others fragments present in mass spectra aid to understand the compounds clustering including the unannotated compounds (Fig. 2C, **Table S2,** Table S11 **and** Table S12). The cluster 1 could be related mainly to dienoid alkaloids while other alkaloids including alkenoid subclass are included in cluster 2. The 17 unannotated alkaloids display characteristic mass spectrum of the Erythrina alkaloids, therefore were grouped in the clusters after MolNet analysis, however they did not match fully the previously reported alkaloid to the *Erythrina* genus. It underlines the chemical diversity of *E. velutina* when collected in its natural semiarid environment with its unique edaphoclimatic features. Such unexplored chemodiversity constitutes a reservoir of potentially useful metabolites still to be isolated, unequivocally assigned, and tested against pharmacological targets.

We hypothesized that the edaphoclimatic features in the Caatinga biome could select and elicit biosynthesis of novel and diverse metabolites in local species. Furthermore, native species from Caatinga have been scarcely studied so far at the metabolites level. These data provide relevant information, affording a representative picture of plant chemistry/interaction with the environment, which is particularly relevant under the multifactorial biotic and abiotic elicitors imposed by semiarid conditions. Details about the edaphoclimatic conditions prevailing at the harvesting site are provided in Supplementary Material (Fig. S3).

Given that trees used in the present experiments are under natural growth conditions, some alkaloids were heterogeneously distributed among the samples (**Fig. S4** and Table S3). These differences were expected, in contrast to most experiments performed under controlled conditions, in which less variation is apparent. Modulation of plant secondary metabolites, including alkaloids, by macro and micronutrients, light intensity, temperature, and soil salinity has been shown in some plant species [21]. Although it is not possible to establish a clear correlation between the production of different alkaloids and soil features in the present investigation, these factors must be considered as part of the multifactorial environmental equation. The highest chemical diversity was observed on the pools 2, 3 and 7, in which 11, 13 and 17 different structures of alkaloids were identified, respectively (Table S3). Interestingly, pools 3 and 7 (pool set C) were harvested at the same site in Acari. Soil analysis showed that this location has some peculiar characteristics, such as high soil pH, high calcium and potassium contents, low amounts of iron and the presence of chromium, a heavy metal not detected elsewhere (Table S7). Mineral nutrition, pH, and heavy metals have been shown to affect alkaloid accumulation. Levels of indole alkaloids in *Psychotria brachyceras*, for example, were significantly increased by exposure to NaCl or to micromolar concentrations of aluminum and silver, which has been partly attributed to moderate intensity oxidative stress [22].

Even though glycosylation constitutes an effective tool to increase bioavailability and bioactivity of benzylisoquinoline alkaloids [23], it is rare to detect the glycosyl form of Erythrina alkaloids and few of them have been reported so far [11]. Zhang and colleagues [24] identified five glycosylated Erythrina alkaloids by 1D and 2D NMR; in all of them glycosylation was at C11, whereas C15 and C16 had methoxy or methylenedioxy groups. This low availability of data on glycosyl form of alkaloids in the genus could be related to differences in post- harvest procedures, as well to the analytical strategy used in most studies, e.g. varying the pH for high and low levels, that are not suitable for preserve the original chemical structure for some substances [25]. It is likely that the rigorous experimental procedures involved in the present analysis with a targeted metabolome perspective contributed to preserve the original sample; in addition, the environmental conditions of the Caatinga could be involved in promoting the production of this kind of metabolite. Some general features of the annotated alkaloids can be highlighted regarding structure and organ distribution. Seed-exclusive alkaloids showed hydroxyl substituents at C3, C11, and, to a lesser extent, at C2, whereas C3 could also present an *O*-methylation (numbering alkaloid main scaffold in Fig. 2A). Interestingly, seven glycosylated alkaloids were detected exclusively in seeds. To the best of our knowledge, their proposed structures

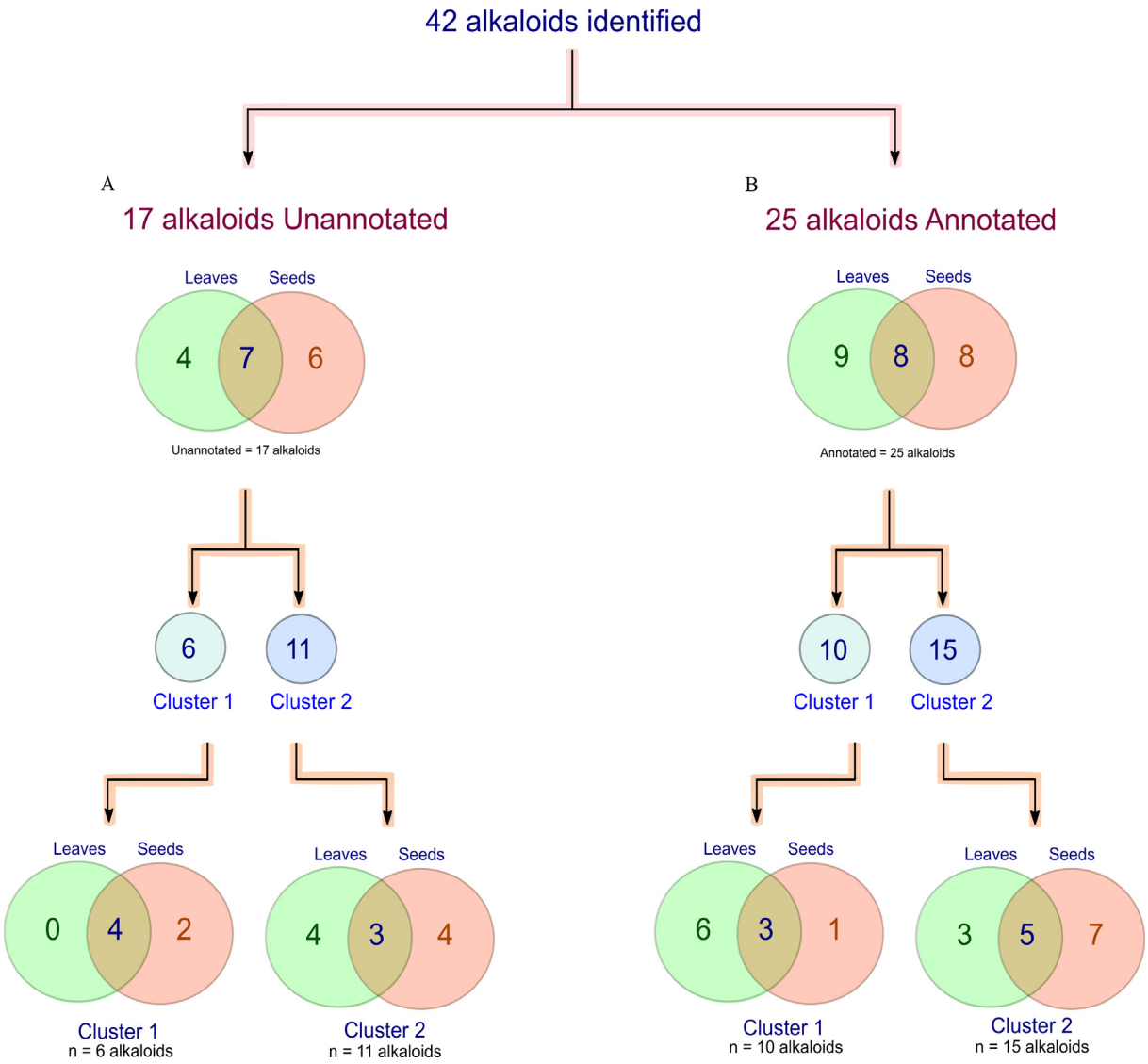


Fig. 1. Overview of *Erythrina velutina* alkaloids by plant structure (leaves and seeds) and presence in cluster 1 and 2.

are unprecedented (Table S1 - alkaloid number 11, 12, 15, 19, 20, 21, 22). Glycosyl residues frequently were attached at C15 and C16; otherwise these positions could be involved in a methylenedioxy bridge (R-O-CH_2-O-R) (Fig. 4B). Although the biological functions of alkaloids in seeds are not completely defined [26], they may provide defense against herbivores or pathogens and perhaps take part in nitrogen metabolism during germination [8]. Erythraline-11β-*O*-glucopyranoside, identified in seeds of *E. crista-galli*, showed inhibitory effect against tobacco mosaic virus [27]. The effects of glycosylation of specialized metabolites are manifold, including modulation of transport, stability, storage, and stress hormones [28]. The presence of glycosylated alkaloids in the examined *E. velutina* organs may be relevant as non-enzymatic antioxidants, as reported for other classes of glycosylated alkaloids [29]. The seasonally anticipated or directly acting extreme conditions of the Caatinga dry season (heat, excess irradiance and low water availability**,** Fig. S3) may also have promoted alkaloid accumulation aiming at redox regulation in plants. An additional reason for the relatively higher abundance of glycosylated alkaloids in seeds of *E. velutina* is the fact that lipids represent the second most abundant reserve in this species. Lipid reserve mobilization may generate hydrogen peroxide during germination and early seedling establishment, particularly at higher temperatures [30,31].

Alkaloids solely found and annotated in leaves share the feature of *O*-methylation at C3, and the oxidized form (C═O) is predominant at C11 and, to a lesser extent, at C2. In the alkaloids shared by both organs main characteristics included predominance of OH at C2, an *O*-methylation pattern at C3, whereas C15 and C16 were mainly involved in a methylenedioxy ring. The role of BIA *in planta* includes protection against herbivores, fungi and bacteria, which contributes to increased plant reproductive success [8].

Divergent metabolite accumulation profiles could be observed in leaves and seeds and establishing intrinsic patterns is nor trivial. The relative signal intensity in *m*/*z* 492.113 (alkaloid 22) with peak height of 3.69E+03 was used to normalize the abundance of all alkaloids (Table S3 and Fig. S4). The *m*/*z* 298.0966 alkaloid, tentatively annotated as coccolinine or isococcolinine (alkaloid 8, Table S3) at peak height of 1.75E+05, was the most prominent alkaloid present in both leaves and seeds. In addition to coccudinone (alkaloid 2 – Table S1), this molecule was recorded for the first time in the *Erythrina* genus. Previously, these structures were identified in *Cocculus laurifolius* (Ranunculales order, Menispermaceae). Despite the taxonomic distance, they share several structural fea-

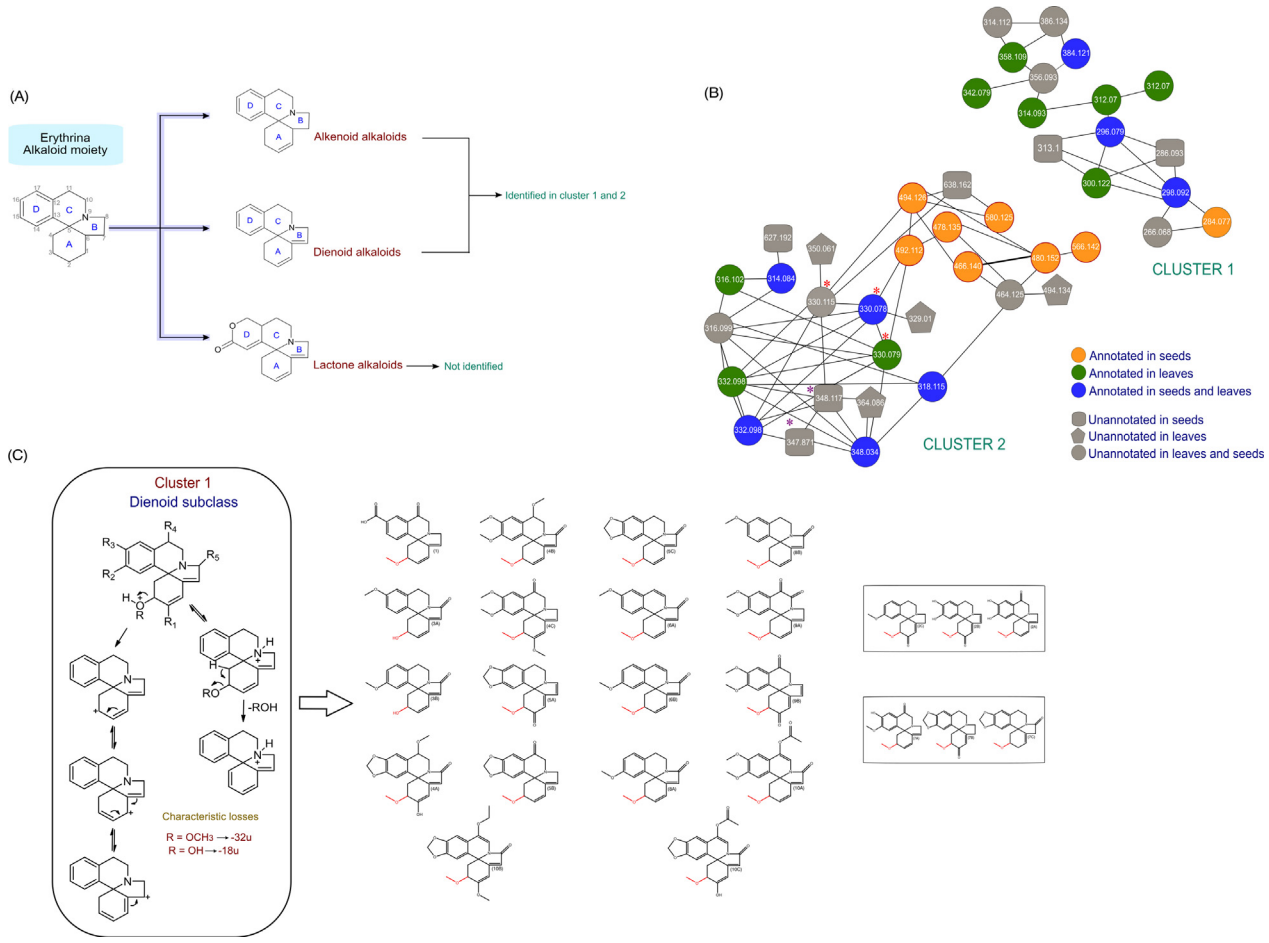


Fig. 2. Molecular networking and annotated alkaloids. (A) Representation of the tetracyclic spiroamine of the alkaloids of the genus *Erythrina* and their respective classes (B) Molecular networking with MS/MS data from hydroethanolic crude extracts analysis of *Erythrina velutina*. Green nodes: alkaloids exclusively in leaves; orange nodes: alkaloids exclusively in seeds; blue nodes: alkaloids in both leaves and seeds, and gray nodes (leaves: pentagon, seeds: square): alkaloids identified but not annotated (they did not match fully the mass spectrum of alkaloids previously reported in *Erythrina* spp.). The high resolution mass value for each alkaloid is shown in the node. Thicker and closer lines indicate greater similarity between the chemical entities. (*) highlight the mass values that were repeated in the cluster (may represent isomers). Red-bordered nodes indicate glycosylated alkaloids. Clusters 1 and 2 are not related to the division of categories of the dienoid and alkenoid classes. (**C**) Structures and characteristic losses due to substituent elimination observed in dienoid subclass fragmentation that were also present in MS/MS spectra of the alkaloids annotated in cluster 1.

tures with Erythrina alkaloids (Fabaceae). These "abnormal Erythrina alkaloids" exhibit a certain stereospecificity for the precursor (*S*)-norprotosinomenine [32,33].

Alkaloids represent one of the most diverse classes of plant metabolites both in terms of structure and bioactivity. About 20% of plant species accumulate alkaloids [34], which can also be found in several classes of other living organisms, forming a group of chemically diverse natural products, abundant both in number and form [35]. These molecules also constitute an interesting source of chemical structures with medicinal properties [36]. Some of the best-known alkaloid-based medicines are the benzylisoquinoline alkaloids (BIAs) morphine and codeine. Indeed, opium poppy (*Papaver somniferum*) has emerged as a powerful "non-model model" system for studying biosynthesis and production of BIAs relevant in the pharmaceutical field [37,38]. BIA diversity is remarkable, spanning over 2,000 known structures, also present in related taxa, such as *Coptis japonica, Thalictrum flavum* and *Eschscholzia californica* [39]. However, beyond the classical Ranunculales distribution of BIAs, lie unique molecules, such as those found in Fabaceae (Erythrina alkaloids). The comparative analysis based on gene similarity data from BIA-producing plants represents a useful tool to suggest candidate genes in *E. velutina* and associate them with steps in the proposed alkaloid biosynthetic pathway.

Candidate genes for a biosynthetic pathway of E. velutina alkaloids

Data from RNA-seq analysis has been used to identify several genes encoding biosynthetic pathways from across plant specialized metabolism, including alkaloids [40,41]. However, this scenario is more complex in non-model plants that do not yet have a genome sequence available. Methods based on alignment with reference species and *de novo* assembly can generate specific and testable hypotheses at the molecular level about metabolic pathways in non-model plants [42]. Therefore, the transcriptome of seeds and leaves of *E. velutina* was analyzed trying to assess main candidate genes involved in the alkaloid biosynthesis.

The primary analysis generated a total of 6,251,021 (Table S5) transcripts among the eight pools. The numbers of transcripts identified in each pool are listed in Table S4. Databases and previously reported information on other BIA-producing species were also searched in order to select suitable domains for candidate genes involved in *E. velutina* alkaloid biosynthetic pathways. The annotated and manually selected transcripts were then subjected to a local analysis via BLASTx. Several transcripts showed similarity with species-specific sequences of other BIA-producing plants (Table S8). Based on the transcripts identified as candidate genes, on the annotated alkaloid features, and on previous reports, a puta-

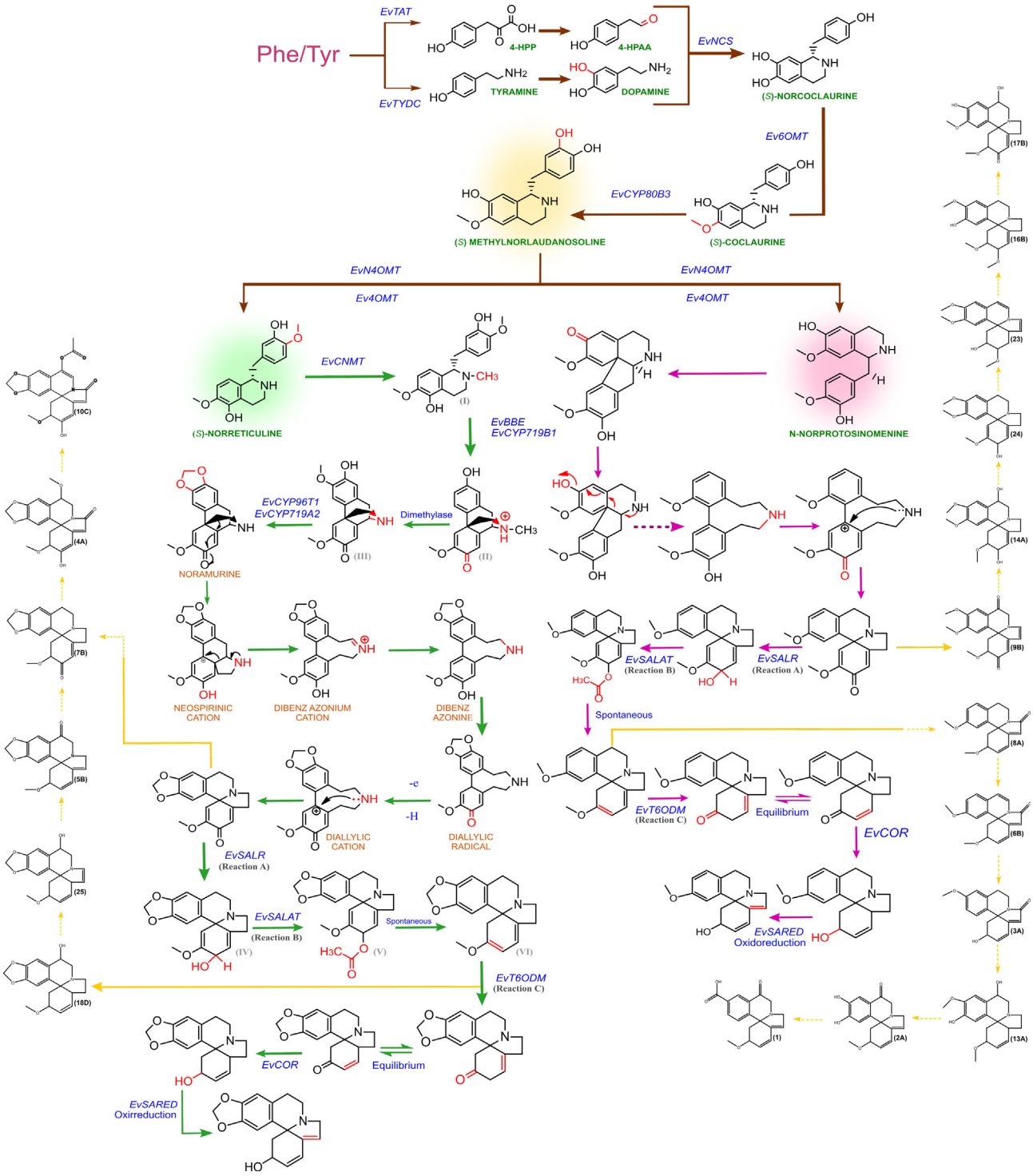


Fig. 3. Putative biosynthetic pathway of *Erythrina* alkaloids. *Erythrina* alkaloids are isoquinoline-type, forming a component of the phenylalanine precursor. The pathway begins with branches initiated with 4-HPP and tyrosine, which by enzymatic action form 4-HPAA and dopamine, respectively, producing intermediate (s)-norcoclaurine and coclaurine. The CYP80B3 part (S)-methylnorlaudanosoline will be formed and is one of the crucial points in the chemical cascade of *Erythrina* alkaloids, as part of it may generate a bifurcation to (S)-norreticuline (pathways in green) and norprotosinomenine (pathways in pink). Pathways in yellow color represent the alkaloids annotated in this study (in accordance with Table S1); Dotted arrows indicate reactions not yet fully established; Blue transcripts were identified in our transcriptome data set according to the Blastx result. Structures of the seven glycosylated alkaloids [11,12,15,19,20,21,22] are shown in Fig. 4. Abbreviations: 4-HPP, 4-Hydroxyphenylpyruvate; 4-HPAA, 4-hydroxyphenylacetaldehyde; *TYDC*, TYROSINE/DOPA DECARBOXYLASE; *TAT*, TYROSINE AMINOTRANSFERASE; *NCS*, NORCOCLAURINE SYNTHASE; *6OMT*, (S)-NORCOCLAURINE 6-O-METHYLTRANSFERASE; *7OMT*, (R,S)-RETICULINE 7-O-METHYLTRANSFERASE; *CYP80B3*, (S)-*N*-METHYLCOCLAURINE 3′-HYDROXYLASE; *4OMT*, 3′-HYDROXY-*N*-METHYL-(S)-COCLAURINE 4′-O-METHYLTRANSFERASE; *N4OMT*, NORBELLADINE 4′-O-METHYLTRANSFERASE; *CNMT*, (S)-COCLAURINE *N*-METHYLTRANSFERASE; *BBE*, BERBERINE BRIDGE ENZYME; *CYP719B1*, SALUTARIDINE SYNTHASE; *CYP712A2*, (S)-STYLOPINE SYNTHASE; *SALR*, SALUTARIDINE REDUCTASE; *SALAT*, SALUTARIDINOL 7-O-ACETYLTRANSFERASE; *CYP96T1*, NOROXOMARITIDINE SYNTHASE; *T6ODM*, THEBAINE 6-O-DEMETHYLASE; *COR*, NADPH-DEPENDENT CODEINONE REDUCTASE; *SARED*, SANGUINARINE REDUCTASE.

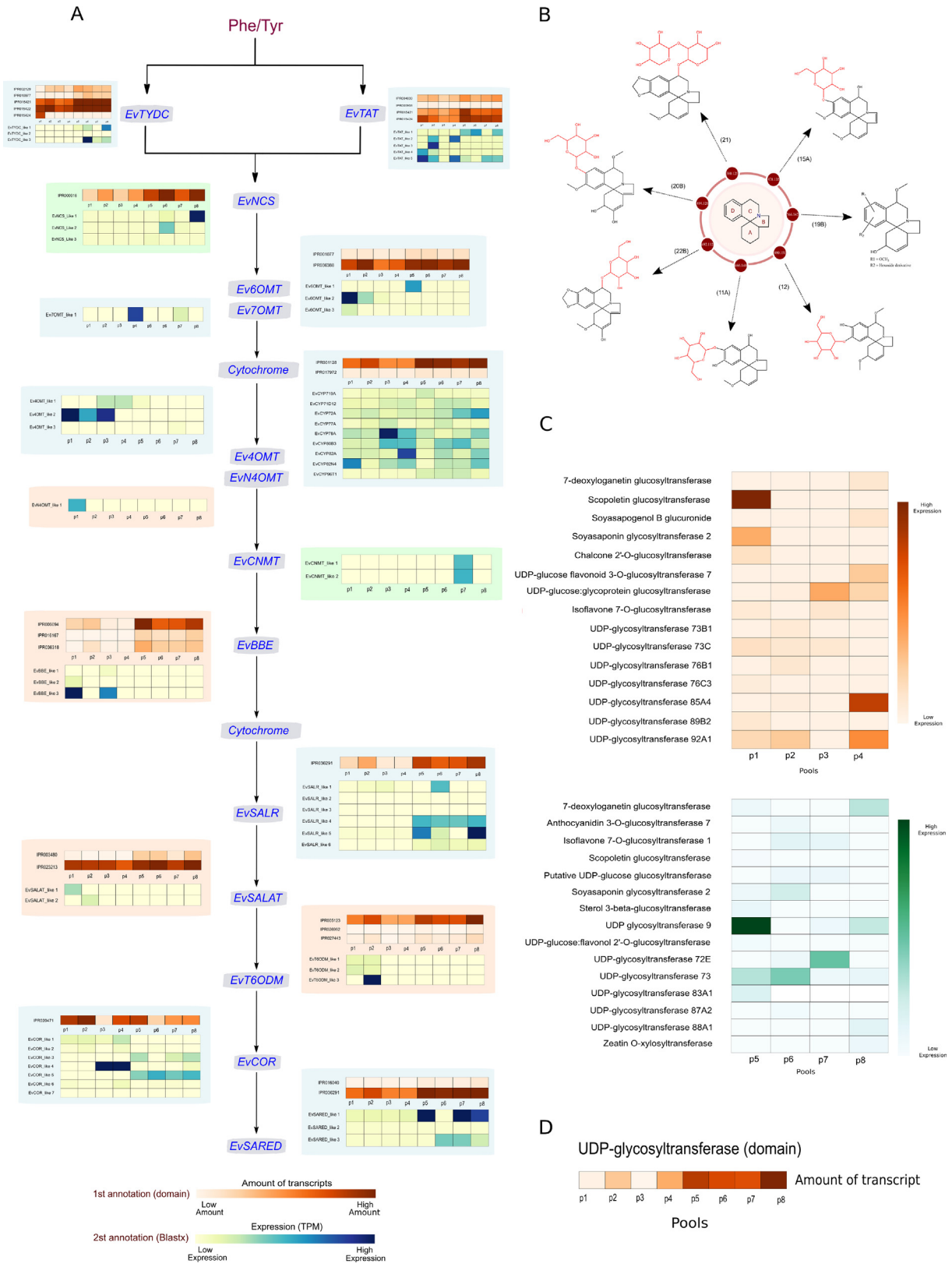


Fig. 4. Gene expression heatmap. The transcriptional expression profile of the candidate genes involved in the Erythrina alkaloid biosynthesis pathway is illustrated by a color gradient. All transcript expression values were scaled by TPM between low and high expression. (A) Orange different color gradients indicate low or high number of transcripts in each identified pool after annotation by domain; Palette with different color gradients in blue indicate the low or high expressed transcripts. Transcripts considered as possible isoforms were identified and named as "like". Heatmap in green and orange box indicate transcripts identified exclusively in leaves and seeds, while in blue box, found in both plant structures. (B) Structure of the glycosylated alkaloids identified in the metabolic profile. (C) Heatmap of transcripts selected by domain with catalytic glycosylation function and subsequently identified using Blastx in seeds and leaves (UDP-glycosyltransferase). (D) Heatmap of the number of transcripts identified with the UDP-glycosyltransferase domain. P1-p4: transcripts expressed in seeds and p5-p8: transcripts expressed in leaves.

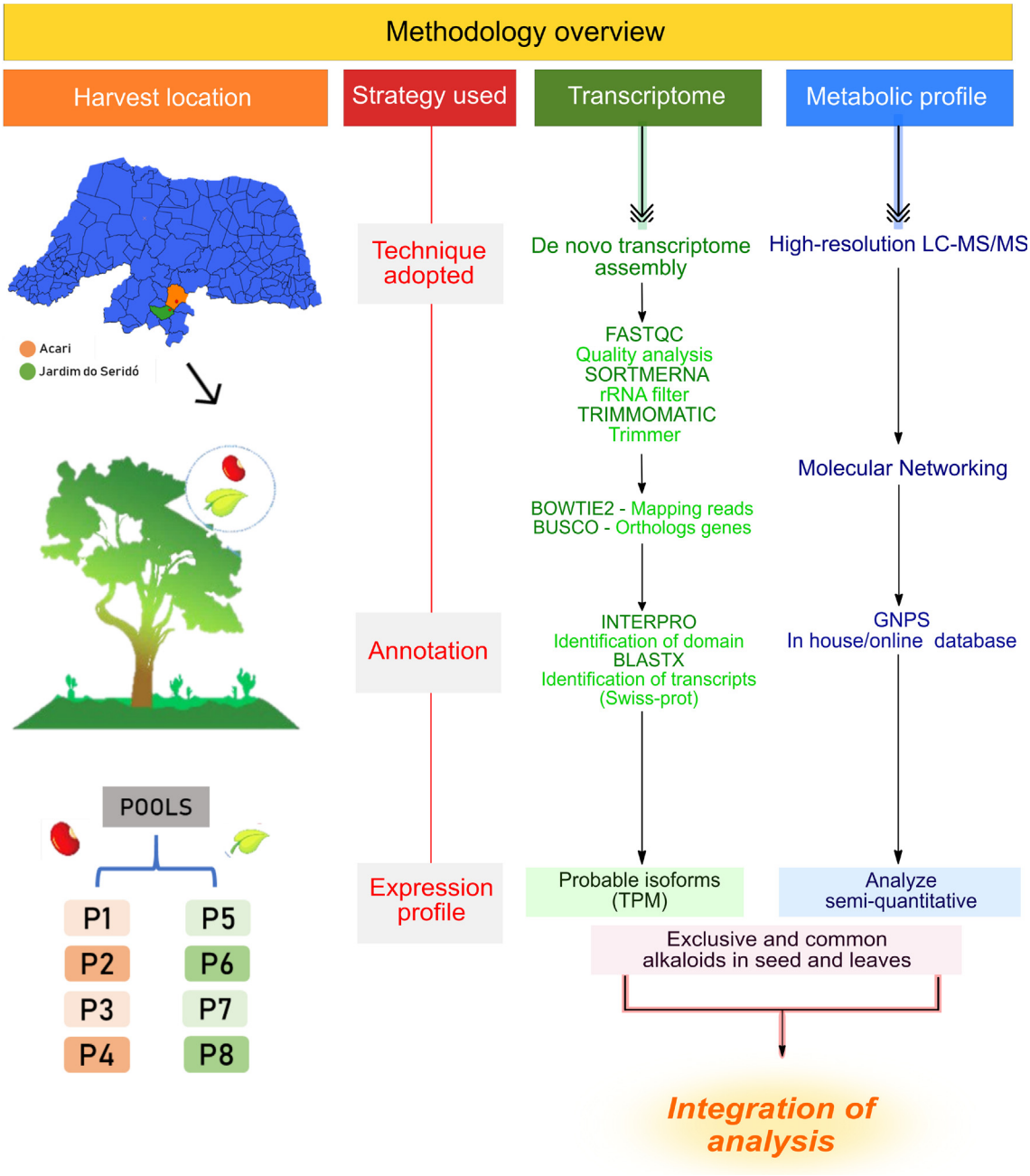


Fig. 5. Experimental strategy and integration of omic analyzes. Prospecting bioactive compounds from biodiversity is a challenging task, and the tools that can most adequately assist in this process become increasingly important. Data integration contributes to a better understanding of the biological system being examined, leading to more detailed understanding of target compound production by the organs of the plant growing in its natural habitat.

tive biosynthetic route for *Erythrina velutina* alkaloids was proposed (Fig. 3).

The biosynthetic pathway for benzylisoquinoline alkaloids derives from aromatic amino acids, starting with the activity of enzymes encoded by *EvTAT* and *EvTYDC*. Next, the condensation of 4-hydroxyphenylacetaldehyde (4-HPAA) and dopamine, two L-Tyr derivatives, is catalyzed by norcoclaurine synthase (*EvNCS*), which produces (*S*)-norcoclaurine [43]. The transcript corresponding to *EvNCS* was identified only in leaves (Fig. 4). Subsequent methylation of (*S*)-norcoclaurine produces (*S*)-coclaurine [44,45], followed by the synthesis of the intermediate (*S*)-methylnorlaudanosoline by (*S*)-*N*-methylcoclaurine 3′-hydroxylase (encoded by *EvCYP80B3*, which was expressed in both leaves and seeds). In opium poppy, the overexpression of *PsCYP80B3* increased total alkaloid yield, as well as morphine biosynthesis [46].

The transcriptome analysis also yielded the gene candidates *Ev6OMT* and *Ev7OMT* as involved in the biosynthesis of (*S*)-coclaurine. *Ev7OMT* was compatible with that of *P. somniferum* [47,48]; however, *Ev6OMT* was more common (Fig. 4). OMTs contribute to the formation of several alkaloid chemical skeletons, such as the isoquinoline monoterpene alkaloid emetine [49] and the last step of the biosynthesis of vindoline, a *Catharanthus roseus* indole alkaloid [50]. Expression of OMT genes seems to be tightly regulated, and the involvement of miRNAs in their regulation has been shown in *P. somniferum* [51].

(*S*)-methylnorlaudanosoline could be one of the pivotal branching points in the biosynthesis of Erythrina alkaloids. This com-

pound is a common intermediate in the production of both (*S*)-norprotosinomenine – part of the "abnormal" Erythrina alkaloids (cocculine alkaloids pathway) - and (*S*)-norreticuline (Fig. 3).

(*S*)-norreticuline, produced by 4′-*O*-methyltransferase (encoded by *Ev4OMT*), present in leaves and seeds, is involved in the pathway of the so-called "common Erythrina alkaloids". These are metabolites usually found in the genus *Erythrina* which contain a tetracyclic spiroamine in their chemical skeleton [52]. *4OMT* has been previously described in BIA-producing species, such as *Coptis japonica* [53], *Papaver somniferum* [54], *Eschscholzia californica* [55] and *Narcissus* spp. [56]. A transcript present in one seed pool, encoding norbelladine 4'-O-methyltransferase (*N4OMT*), was previously reported in Amaryllidaceae as being involved in galantamine biosynthesis, a BIA alkaloid clinically used to treat Alzheimer's disease [57].

In general, methyltransferases previously described in BIA biosynthesis accept a wide variety of multi-structure alkaloid substrates. An example is coclaurine *N*-methyltransferase (encoded by *EvCNMT*), previously characterized as part of the biosynthesis of papaverine, an important isoquinoline opium poppy alkaloid [47,58]. The enzyme acts on the central pathway of BIA and converts a non-methylated substance from secondary amine to a tertiary one [59,48].

Subsequently, after formation of the tertiary amine, C-C bonds can be catalyzed by the proteins encoded by berberine bridge enzyme (*EvBBE*)/*EvCYP719B1*, through the stereospecific conversion of the methyl portion of intermediate (I) to (II) and methylene dioxi bridges formation catalyzed by *EvCYP96T1*/*EvCYP719A,* that converts the intermediate (III) to noramurine. BBE of cultured opium poppy cells was characterized and its gene successfully expressed in *Saccharomyces cerevisiae*, yielding catalytically active enzyme [60–62]. Transcripts encoding cytochrome P_{450} enzymes are important in contributing to the structural diversity of molecules, including BIAs [63]. Several enzyme members of the subfamily [CYP719A9, CYP719A11, CYP719A2, CYP719A3, CYP719A1 and CYP80B3] are involved in isoquinoline moiety biosynthesis [64,46], whereas CYP96T1 was capable of catalyzing the formation of C-C coupling in Amaryllidaceae alkaloid biosynthesis [65]. Most transcripts encoding *CYP719A* and *CYP80B* enzymes were highly expressed in both leaves and seeds. *CYP719A2* and *CYP719A3* cloned from *Eschscholzia californica* cells showed catalytic activity in methylenedioxy bridge formation and methyl jasmonate treatment resulted in the coordinated induction of these genes [66].

The core skeleton of Erythrina alkaloids, a tetracyclic spiroamine, would be produced by subsequent steps of rearrangement to form the cation neospirinic, dibenz azonine to yield a diallylic cation from noramurine [52,67]. We propose that the result of this pathway would be salutaridine (*EvSALR*) for form the intermediate (IV) (reaction A), an enzyme dehydrogenase member and short chain reductase family that in *P. somniferum* use the NADPH as coenzyme, to afford salutaridinol, mainly involved in the morphine and codeine biosynthesis [68].

The next step could be catalyzed by salutaridinol 7-*O*-acetyltransferase (*EvSALAT*) in our proposal through reaction B, which specifically acetylates salutaridinol epimer-7 into salutaridinol-7-*O*-acetate. This step was previously described in the formation of BIA in opium poppy. The acetyl group could be spontaneously eliminated (catalyzing the conversion of intermediate (V) to (VI), in a pH-dependent way as reported in Opium poppy, leading to the formation of an oxide bridge between C4 and C5 to produce thebaine, the first pentacyclic alkaloid prior to morphine formation [37]. Evidence from a gene silencing study suggests that SALR and SALAT interact and may act as an enzyme complex in opium poppy [69]. *EvSALAT* was found only in seeds. After spontaneous deacetylation, thebaine 6-*O*-dimethylase (*EvT6ODM*), a member of the Fe(II)/2-oxoglutarate-dependent dioxygenase family what form neopinone [70], would removes the methyl group and changes the unsaturated bonds in the neighboring carbon (reaction C). Regarding Erythrina alkaloid biosynthesis, these kinds of reactions are also necessary to produce the oxygenated and unsaturated A ring, present in several alkaloids isolated from this genus. Therefore, we propose *SALAT*, and *T6ODM*, both of which seed-specific, as candidate genes in the pathway (Fig. 4).

Finally, the unsaturation of a C-C bond is necessary in the heterocycle of several Erythrina alkaloid moieties. Therefore, we propose the occurrence of an equilibrium between two *Erythrina* intermediates, like the previously observed between neopinone/-codeinone intermediates in *P. somniferum* [71]. The subsequent steps involve oxidoreduction reactions; transcripts compatible with codeinone reductase (*EvCOR*) [72] and sanguinarine reductase isolated from cell cultures of *Eschscholzia californica* (*EvSARED*) [73] were found in the leaves and seeds transcriptome dataset. These candidate genes encode enzymes that would enable oxidation and reduction reactions to achieve the final structure of the Erythrina alkaloids.

Identifying cytochrome P450 enzymes in plants is a challenge, and the information about them is still rather limited; among all crystallized CYPs only five of them are from plants [74]. Therefore, besides the CYPs discussed and pointed as gene candidates in the Erythrina alkaloids biosynthesis since they matched our methodologic parameters, other CYPs reported in plant metabolism and identified in the *E. velutina* transcriptome data set were highlighted (Fig. 4). Among them, *CYP82N4* and *CYP71D12* are also involved in the alkaloid biosynthetic pathways [75,76].

As pointed out above, predominant annotation of seed-exclusive glycosylated alkaloids in *Erythrina* was observed, however, almost twice as many transcripts with a functional domain of UDP-glycosyltransferase were detected in leaves (Fig. 4 and Table S9). Although UDP-glycosyltransferases are a large family of proteins involved in diverse metabolic processes, these results might indicate the possibility of translocation of alkaloids between organs and/or variations of aglycone modifying enzymes in different plant cell types.

Some alkaloids detected in the metabolic profiling dataset belong to a class named cocculine alkaloids or "abnormal Erythrina alkaloids" (Fig. 3). The chemical skeleton of such alkaloids contains a tetracyclic spiroamine moiety, but they were originally characterized from *Cocculus laurifolius*. Data regarding the biosynthesis of the cocculine alkaloids are scarce, but norprotosinomenine has been indicated as an intermediate with stronger isotope incorporation than norreticuline [32,33]. The occurrence of both uncommon and common Erythrina alkaloids in *E. velutina* and the presence of candidate genes possibly involved in both pathways (Fig. 3) suggest that the most appropriate term to refer to these alkaloids is "tetracyclic spiroamine BIA". This would acknowledge the fact that these alkaloids could be synthesized from both precursors: (*S*)-norprotosinomenine as well as (*S*)-norcoclaurine.

The species *Cocculus laurifolius* (order Ranunculales) and *Erythrina velutina* (order Fabales) are included in the great clade of eudicots. The Ranunculales order is a valuable source of benzylisoquinoline alkaloids with important alkaloid-producing families: Menispermaceae, that includes species of the genus *Cocculus.* Berberidaceae, such as the genus *Berberis*, which comprises *B. vulgaris* (the main natural source of berberine) [77], and Papaveraceae (*Papaver somniferum*), which accumulate BIAs, such as morphine, codeine, and noscapine [39]. Fabaceae is taxonomically more distant but is included in the same monophyletic group [78].

Phylogenetic, biochemical and molecular approaches were considered to conduct an empirical investigation on the evolution of BIA synthesis in plants indicating that the ability of angiosperms to produce BIAs appears to have originated shortly before the appearance of eudicots [79]. These studies support a monophyletic

evolution of proteins involved in the biosynthesis of BIAs in angiosperms and suggest that the ecophysiological functions of specific BIAs was generated by selective environmental pressures and/or the emergence of specialized physiological mechanisms that contributed to the maintenance or reactivation of the pathway. The large clade of eudicotyledons seems to be an important evolutionary branch for the synthesis of several bioactive tyrosine derivatives and formation of alkaloids.

In recent years, efforts have been made on biosynthesis research and pharmaceutical use of isoquinoline alkaloids; however, the results are mostly focused on opium poppy and some Amaryllidaceae species. The present study provides a first omic analysis of the importance of Fabaceae and *E. velutina* in this scenario, bringing to the spotlight a non-model plant growing in its naturally harsh environment. Overall, there was good agreement between transcriptome and metabolome data, affording a reasonable picture of BIA biosynthesis and part of its regulation in this species.

Conclusion

This detailed macro and micro molecular approach applied to seeds and leaves of *E. velutina* harvested in Caatinga from four different locations revealed 42 alkaloids by HR-LC-MS/MS and MolNet in a targeted metabolomic approach. Furthermore, based on the combined evidence, 24 gene candidates were put together in a putative biosynthetic pathway. These results provide a framework to advance on further findings to understand the real role of each step in the route as well as the flux controller enzymes leading to the singular alkaloid diversity of this species. In addition, these results contribute to show potential biotechnological targets and to stimulate future studies on Brazilian biodiversity detailed at the level of biosynthesis of bioactive natural products, notably alkaloids. Finally, in an important and necessary way, the datasets presented herein add new data to public databases on metabolome and transcriptome from non-model medicinal plants.

Accession numbers

The transcriptome sequence readings were deposited in the Sequence Read Archive (SRA) under the BioProject accession number: PRJNA668524.

Financing

This work was supported by the Serrapilheira Institute (grant number Serra-1709-19691), by the Ministry of Science, Technology, Innovation and Communications – MCTIC, CNPq/National Council of Science and Technology - INCT BioNat, [grant number 465637/2014-0], São Paulo Research Foundation (FAPESP grant INCTBioNat 2014/50926-0) and by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil [(CAPES) - Finance Code 001)].

Author contributions

D.S.C., I.B.S., F.P.S.R., A.A.R., D.S. and A.P. carried out the experimental procedures. D.S.C., F.P.S.R., D.S., A.P., F.A.S.D.P., R.F., A.S.P, J. A.S.Z., A.J.C., L.S.F. and N.P.L. analyzed and discussed chemical data. D.S.C., I.B.S., C.T.C., J.V., T.M.T., T.F.A., E.L.V., K.C.S., L.V.S. analyzed and discussed molecular data. R.L., J.A.R. and F.L.M.S. analyzed and discussed the edaphoclimatic data. R.B.G. and A.G.F.N., planned the study, coordinated the project, and analyzed all experimental data. All authors discussed the major conclusions and contributed to write the manuscript.

Compliance with Ethics Requirements

This article does not contain any studies with human or animal subjects.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank the research supported by LNBR – Brazilian Bioenewables National Laboratory (CNPEM/MCTIC) during the use of the High Throughput Sequencing (NGS) open access facility. Also, we thank the Core Facility for Scientific Research–University of São Paulo (CEFAP-USP/GENIAL (Genome Investigation and Analysis Laboratory)) for the NextSeq 500 (Illumina) analysis. Ministry of Science, Technology, Innovation and Communications – MCTIC, FAPESP (FAPESP; grant numbers 2009/54083-9R, 2013/07600–3, 2014/05935–0 [DMS], 2017/06446–2 [DMS], and 2016/13292–8 [ACP]) and INCT BioNat. D.S.C and F.P.S.R. are recipients of fellowship from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). JASZ acknowledges CNPQ for the research fellowship.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jare.2021.01.017.

References

[1] Yonekura-Sakakibara K, Saito K. Functional genomics for plant natural product biosynthesis. Nat Prod Rep 2009:1466–87. doi: https://doi.org/10.1039/b817077k.
[2] Matsuura HN, Malik S, de Costa F, Yousefzadi M, Mirjalili MH, Arroo R, et al. Specialized Plant Metabolism Characteristics and Impact on Target Molecule Biotechnological Production. Mol Biotechnol 2018:169–83. doi: https://doi.org/10.1007/s12033-017-0056-1.
[3] Rodrigues T, Reker D, Schneider P, Schneider G. Counting on natural products for drug design. Nat Chem 2016:531–41. doi: https://doi.org/10.1038/nchem.2479.
[4] Debnath M, Pandey M, Bisen PS. An Omics Approach to Understand the Plant Abiotic Stress, OMICS: A Journal of. Integr Biol 2011:739–62. doi: https://doi.org/10.1089/omi.2010.0146No.11.
[5] Chang C, Bowman JL, Meyerowitz EM. Field Guide to Plant Model Systems. Cell 2016:325–39. doi: https://doi.org/10.1016/j.cell.2016.08.031.
[6] Pham JV, Yilma MA, Feliz A, Majid MT, Maffetone N, Walker JR, et al. A Review of the Microbial Production of Bioactive Natural Products and Biologics. Front Microbiol 2019. doi: https://doi.org/10.3389/fmicb.2019.01404.
[7] Zhu H, Li C, Gao C. Applications of CRISPR–Cas in agriculture and plant biotechnology. Nat Rev Mol Cell Biol 2020:1–17. doi: https://doi.org/10.1038/s41580-020-00288-9.
[8] Matsuura HN, Fett-Neto AG. Plant alkaloids: main features, toxicity, and mechanisms of action. Plant toxins 2015:1–15. doi: https://doi.org/10.1007/978-94-007-6728-7_2-1.
[9] Bessa MC, Lacerda CF, Amorim AV, Bezerra AME, Lima AD. Mechanisms of salt tolerance in seedlings of six woody native species of the Brazilian semi-arid 2017:157–65.
[10] Rambo DF, Biegelmeyer R, Tonson NSB, Dresch RR, Moreno PRH, Henriques AT. The genus Erythrina L.: A review on its alkaloids, preclinical, and clinical studies. Phytother Res 2019:258–1276. doi: https://doi.org/10.1002/ptr.6321.
[11] Fahmy NM, Al-sayed E, El-shazly M, Singab AN. Alkaloids of genus *Erythrina*: An updated review. Nat Prod Res 2019:1–22. doi: https://doi.org/10.1080/14786419.2018.1564300.
[12] De Vos RC, Moco S, Lommen A, Keurentjes JJ, Bino RJ, Hall RD. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. Nat Protoc 2007:778–91. doi: https://doi.org/10.1038/nprot.2007.95.
[13] Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR. Gas chromatography mass spectrometry–based metabolite profiling in plants. Nat Protoc 2006:387–96. doi: https://doi.org/10.1038/nprot.2006.59.

[14] Pluskal T, Castillo S, Villar-Briones A, Orešič M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinf 2010:395–405. doi: https://doi.org/10.1186/1471-2105-11-395.
[15] Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 2017:417–9. doi: https://doi.org/10.1038/nmeth.4197.
[16] Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, et al. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorolog Soc 2011:553–97. doi: https://doi.org/10.1002/qj.828.
[17] Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol 2016:828–37. doi: https://doi.org/10.1038/nbt.3597.
[18] Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics 2007:211–21. doi: https://doi.org/10.1007/s11306-007-0082-2.
[19] Dunn WB, Erban A, Weber RMJ, Creek DJ, Brown M, Breitling R, et al. Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. Metabolomics 2013:44–66. doi: https://doi.org/10.1007/s11306-012-0434-4.
[20] Guaratini T, Feitosa LGP, Silva DB, Lopes NP, Lopes JLCL, Vessecchi R. Gas-phase dissociation study of erythrinian alkaloids by electrospray ionization mass spectrometry and computational methods. J Mass Spectrom 2017:571–9. doi: https://doi.org/10.1002/jms.3962.
[21] Yang L, Wen K, Ruan X, Zhao Y, Wei F, Wango Q. Response of Plant Secondary Metabolites to Environmental Factors. Molecules 2018:762–87. doi: https://doi.org/10.3390/molecules23040762.
[22] Nascimento NC, Menguer PK, Henriques AT, Fett-Neto AG. Accumulation of brachycerine, an antioxidant glucosidic indole alkaloid, is induced by abscisic acid, heavy metal, and osmotic stress in leaves of *Psychotria brachyceras*. Plant Physiol Biochem 2013:33–40. doi: https://doi.org/10.1016/j.plaphy.2013.08.007.
[23] Zhang Y, Xie K, Liu A, Chen R, Chen D, Yang L, et al. Enzymatic biosynthesis of benzylisoquinoline alkaloid glycosides via promiscuous glycosyltransferases from *Carthamus tinctorius*. Chin Chem Lett 2018:1–6. doi: https://doi.org/10.1016/j.cclet.2018.05.010.
[24] Zhang B-J, Xiao W-N, Chen J, Bao M-F, Schinnerl J, Wang Q, et al. Erythrina alkaloids from leaves of *Erythrina arborescens*. Tetrahedron 2019:. doi: https://doi.org/10.1016/j.tet.2019.130515130515.
[25] Basiliere S, Kerrigan S. Temperature and pH-Dependent Stability of Mitragyna Alkaloids. J Anal Toxicol 2020:314–24. doi: https://doi.org/10.1093/jat/bkz103.
[26] Ali AH, Abdelrahman M, El-Sayed MA. Alkaloid Role in Plant Defense Response to Growth and Stress. In: Jogaiah S, Abdelrahman M, editors. Bioactive Molecules in Plant Defense. Cham: Springer; 2019.
[27] Tan Q-W, Ni JC, Fang P-H, Chen Q-J. A New Erythrinan Alkaloid Glycoside from the Seeds of *Erythrina crista-galli*. Molecules 2017:1558. doi: https://doi.org/10.3390/molecules22091558.
[28] Wang S, Alseekh S, Fernie AR, Luo J. The Structure and Function of Major Plant Metabolite Modifications. Molecular Plant 2019:899–919. doi: https://doi.org/10.1016/j.molp.2019.06.001.
[29] Matsuura HN, Rau MR, Fett-Neto AG. Oxidative stress and production of bioactive monoterpene indole alkaloids: biotechnological implications. Biotechnol Lett 2013:191–200. doi: https://doi.org/10.1007/s10529-013-1348-6.
[30] Felix FC, Medeiros JAD, Ferrari CS, Pacheco MV, Torres SB. Molecular aspects during seed germination of *Erythrina velutina* Willd. under different temperatures (Part 1): reserve mobilization, Journal of Seed. Science 2020:. doi: https://doi.org/10.1590/2317-1545v42239839e202042029.
[31] Han C, Yin X, He D, Yang P. Analysis of Proteome Profile in Germinating Soybean Seed, and Its Comparison with Rice Showing the Styles of Reserves Mobilization in Different Crops. PLoS ONE 2013:. doi: https://doi.org/10.1371/journal.pone.0056947e56947.
[32] Bhakuni D, Jain S. Late stages in the biosynthesis of abnormal Erythrina alkaloids. Pergamon Press 1980:2153–6. doi: https://doi.org/10.1016/0040-4020(80)80108-6.
[33] Bhakuni DS, Singh AN. Biosynthesis of the abnormal Erythrina alkaloids, cocculidine and cocculine. J Chem Soc 1978:618–22. doi: https://doi.org/10.1039/p19780000618.
[34] De Luca V, St B. Pierre, The cell and developmental biology of alkaloid biosynthesis. Trends Plant Sci 2000:168–73. doi: https://doi.org/10.1016/s1360-1385(00)01575-2.
[35] Goyal S. Ecological Role of Alkaloids. Natural Products 2013:149–71. doi: https://doi.org/10.1007/978-3-642-22144-6-98.
[36] Hussain MM, Tuhin MTH, Akter F, Rashid MA. Constituents of *Erythrina* - a Potential Source of Secondary Metabolities: A Review, Bangladesh. Pharmaceutical Journal 2016:237–53. doi: https://doi.org/10.3329/bpj.v19i2.29287.
[37] Facchini PJ, De Luca V. Opium poppy and Madagascar periwinkle: model non-model systems to investigate alkaloid biosynthesis in plants. Plant J 2008:763–84. doi: https://doi.org/10.1111/j.1365-313x.2008.03438.x.
[38] Labanca F, Ovesnà J, Milella L. *Papaver somniferum* L. taxonomy, uses and new insight in poppy alkaloid pathways. Phytochem Rev 2018:853–71. doi: https://doi.org/10.1007/s11101-018-9563-3.
[39] Hagel JM, Morris JS, Lee EJ, Penix ID, Bross CD, Chang L, et al. Transcriptome analysis of 20 taxonomically related benzylisoquinoline alkaloid-producing plants. BMC Plant Biol 2015:1–16. doi: https://doi.org/10.1186/s12870-015-0596-0.
[40] Rai A, Saito K, Yamazaki M. Integrated omics analysis of specialized metabolism in medicinal plants. Plant J 2017:764–87. doi: https://doi.org/10.1111/tpj.13485.
[41] Wurtzel ET, Kutchan TM. Plant metabolism, the diverse chemistry set of the future. Science 2016:1232–6. doi: https://doi.org/10.1126/science.aad2062.
[42] Ward JA, Ponnala L, Weber CA. Strategies for transcriptome analysis in non-model plants. Am J Bot 2012:267–76. doi: https://doi.org/10.3732/ajb.1100334.
[43] Lee E, Facchini PJ. Norcoclaurine Synthase is a member of the pathogenesis-Related 10/Bet v1 Protein Family. Plant Cell 2010:3489–503. doi: https://doi.org/10.1105/tpc.110.077958.
[44] Ziegler J, Voigtländer S, Schmidt J, Kramell R, Miersch O, Ammer C, et al. Comparative transcript and alkaloid profiling in *Papaver* species identify a short chain dehydrogenase/reductase involved in morphine biosynthesis. Plant J 2006:177–92. doi: https://doi.org/10.1111/j.1365-313x.2006.02860.x.
[45] Zulak KG, Khan MF, Alcantara J, Schriemer DC, Facchini PJ. Plant Defense Responses in Opium Poppy Cell Cultures Revealed by Liquid Chromatography-Tandem Mass Spectrometry Proteomics. Mol Cell Proteomics 2008:86–98. doi: https://doi.org/10.1074/mcp.m800211-mcp200.
[46] Frick S, Kramell R, Kutchan TM. Metabolic engineering with a morphine biosynthetic P450 in opium poppy surpasses breeding. Metab Eng 2007:169–76. doi: https://doi.org/10.1016/j.ymben.2006.10.004.
[47] Pathak S, Lakhwani D, Gupta P, Mishra BK, Shukla S, Asif MH, et al. Comparative transcriptome analysis using high papaverine mutant of *Papaver somniferum* reveals pathway and uncharacterized steps of papaverine biosynthesis. PLoS ONE 2013:1–12. doi: https://doi.org/10.1371/journal.pone.0065622.
[48] Ounaroon A, Decker G, Schmidt J, Lottspeich F, Kutchan TM. (R, S)-Reticuline 7-O-methyltransferase and (R, S)-norcoclaurine 6-O-methyltransferase of *Papaver somnifer*um -cDNA cloning and characterization of methyl transfer enzymes of alkaloid biosynthesis in opium poppy. Plant J 2003:808–19. doi: https://doi.org/10.1046/j.1365-313x.2003.01928.x.
[49] Nomura T, Kutchan TM. Three new O-Methyltransferases are sufficient for all *O*-Methylation reactions of ipecac alkaloid biosynthesis in root culture of *Psychotria ipecacuanha*. J Biol Chem 2010:7722–38. doi: https://doi.org/10.1074/jbc.m109.086157.
[50] Levac D, Murata J, Kim WS, De Luca V. Application of carborundum abrasion for investigating the leaf epidermis: molecular cloning of *Catharanthus roseus* 16-hydroxytabersonine-16-O-methyltransferase. Plant J 2007;225:236. doi: https://doi.org/10.1111/j.1365-313x.2007.03337.x.
[51] Boke H, Ozhuner E, Turktas M, Parmaksiz I, Ozcan S, Unver T. Regulation of the alkaloid biosynthesis by miRNA in opium poppy. Plant Biotechnol J. 2015:409–20. doi: https://doi.org/10.1111/pbi.12346.
[52] Maier UH, Rodl W, Deus-Neumann B. M, H, Zenk, Biosynthesis of *Erythrin*a alkaloids in Erythrina crista-galli. Phytochemistry 1999:373–82. doi: https://doi.org/10.1016/s0031-9422(99)00230-7.
[53] Morishige T, Tsujita T, Yamada Y, Sato F. Molecular characterization of the S-adenosyl-l-methionine: 3′-hydroxy-*N*-methylcoclaurine 4′-O-Methyltransferase involved in isoquinoline alkaloid biosynthesis in *Coptis japonica*. J Biol Chem 2000:23398–405. doi: https://doi.org/10.1074/jbc.m002439200.
[54] Gurkok T, Ozhuner E, Parmaksiz I, Özcan S, Turktas M, Ipek A, et al. Functional characterization of 4′OMT and 7OMT genes in BIA Biosynthesis. Front Plant Sci 2016:1–11. doi: https://doi.org/10.3389/fpls.2016.00098.
[55] Purwanto R, Hori K, Yamada Y, Sato F. Unraveling Additional *O*-methylation steps in benzylisoquinoline alkaloid biosynthesis in California Poppy (*Eschscholzia californica*). Plant Cell Physiol 2017:1528–40. doi: https://doi.org/10.1093/pcp/pcx093.
[56] Kilgore MB, Augustin MM, Starks CM, O'Neil-Johnson M, May GD, Crow JA, et al. Cloning and characterization of a norbelladine 4′-O-methyltransferase involved in the biosynthesis of the Alzheimer's drug galanthamine in *Narcissus sp. aff. Pseudonarcissus*, Plos One 2014:1–11. doi: https://doi.org/10.1371/journal.pone.0103223.
[57] Li W, Qiao C, Pang J, Zhang G, Luo Y. The versatile O-methyltransferase LrOMT catalyzes multiple *O*-methylation reactions in Amaryllidaceae alkaloids biosynthesis. International Journal of Biological Macromolecule 2019:680–92. doi: https://doi.org/10.1016/j.ijbiomac.2019.09.011.
[58] Facchini PJ, Morris JS. Molecular origins of functional diversity in benzylisoquinoline alkaloid methyltransferases. Front Plant Sci 2019:1–36. doi: https://doi.org/10.3389/fpls.2019.01058.
[59] Morris JS, Yu L, Facchini PJ. A single residue determines substrate preference in benzylisoquinoline alkaloid *N*-methyltransferases. Phytochemistry 2020:. doi: https://doi.org/10.1016/j.phytochem.2019.112193112193.
[60] Alcantara J, Bird DA, Franceschi VR, Facchini PJ. Sanguinarine biosynthesis is associated with the endoplasmic reticulum in cultured Opium Poppy Cells after elicitor treatment. Plant Physiol 2005:173–83. doi: https://doi.org/10.1104/pp.105.059287.
[61] Dittrich H, Kutchan TM. Molecular cloning, expression, and induction of berberine bridge enzyme, an enzyme essential to the formation of

benzophenanthridine alkaloids in the response of plants to pathogenic attack, Proc Natl Acad Sci. 1991;9969–73. http://doi.org/10.1073/pnas.88.22.9969.
[62] Winkler A, Puhl M, Weber H, Kutchan TM, Gruber K, Macheroux P. Berberine bridge enzyme catalyzes the six-electron oxidation of (S)-reticuline to dehydroscoulerine. Phytochemistry 2009:1092–7. doi: https://doi.org/10.1016/j.phytochem.2009.06.005.
[63] Zeng J, Liu Y, Liu W, Liu X, Liu F, Huang P, et al. Integration of transcriptome, proteome and metabolism data reveals the alkaloids biosynthesis in *Macleaya cordata and Macleaya microcarpa*. PLoS ONE 2013:1–18. doi: https://doi.org/10.1371/journal.pone.0053409.
[64] Ikezawa N, Iwasa K, Sato F. CYP719A subfamily of cytochrome P450 oxygenases and isoquinoline alkaloid biosynthesis in *Eschscholzia californica*. Plant Cell Rep 2008:123–33. doi: https://doi.org/10.1007/s00299-008-0624-8.
[65] Kilgore MB, Augustin MM, May GD, Crow JA, Kutchan TM. *CYP96T1 of Narcissus sp. aff. pseudonarcissus* Catalyzes Formation of the Para-Para' C-C Phenol Couple in the Amaryllidaceae Alkaloids, Frontiers. Plant Sci 2016:7. doi: https://doi.org/10.3389/fpls.2016.00225.
[66] Ikezawa N, Iwasa K, Sato F. Molecular cloning and characterization of methylenedioxy bridge-forming enzymes involved in stylopine biosynthesis in *Eschscholzia californica*. Febs J 2007:1019–35. doi: https://doi.org/10.1111/j.1742-4658.2007.05652.x.
[67] Barton DHR, James R, Kirby GW, Turner DW, Widdowson DA. Phenol Oxidation and Biosynthesis. Part XVlll. The Structure and Biosynthesis of Erythrina Alkaloids. J Chem Soc 1968:1529–37. doi: https://doi.org/10.1039/j39680001529.
[68] Higashi Y, Kutchan TM, Smith TJ. Atomic Structure of Salutaridine Reductase from the Opium Poppy (*Papaver somniferum*). J Biol Chem 2010:6532–41. doi: https://doi.org/10.1074/jbc.m110.168633.
[69] Kempe K, Higashi Y, Frick S, Sabarna K, Kutchan TM. RNAi suppression of the morphine biosynthetic gene *SALAT* and evidence of association of pathway enzymes. Phytochemistry 2009:579–89. doi: https://doi.org/10.1016/j.phytochem.2009.03.002.
[70] Hagel JM, Facchini PJ. Dioxygenases catalyze the O-demethylation steps of morphine biosynthesis in opium poppy. Nat Chem Biol 2010:273–5. doi: https://doi.org/10.1038/nchembio.317.
[71] Larkin PJ, Miller JAC, Allen RS, Chitty JA, Gerlach WL, Frick S, et al. Increasing morphinan alkaloid production by over-expressing codeinone reductase in transgenic *Papaver somniferum*. Plant Biotechnol J 2007:26–37. doi: https://doi.org/10.1111/j.1467-7652.2006.00212.x.
[72] Unterlinner B, Lenz R, Kutchan TM. Molecular cloning and functional expression of codeinone reductase: the penultimate enzyme in morphine biosynthesis in the opium poppy *Papaver somniferum*. Plant J 1999:465–75. doi: https://doi.org/10.1046/j.1365-313x.1999.00470.x.
[73] Vogel M, Lawson M, Sippl W, Conrad U, Roos W. Structure and mechanism of sanguinarine reductase, an enzyme of alkaloid detoxification. J Biol Chem 2010:18397–406. doi: https://doi.org/10.1074/jbc.m109.088989.
[74] Liu X, Zhu X, Wang H, Liu T, Cheng J, Jiang H. Discovery and modification of cytochrome P450 for plant natural products biosynthesis. Synth Syst Biotechnol 2020:187–99. doi: https://doi.org/10.1016/j.synbio.2020.06.008.
[75] Beaudoin GAW, Facchini PJ. Isolation and characterization of a cDNA encoding (S)-cis-N-methylstylopine 14-hydroxylase from opium poppy, a key enzyme in sanguinarine biosynthesis. Biochem Biophys Res Commun 2013:597–603. doi: https://doi.org/10.1016/j.bbrc.2012.12.129.
[76] Besseau S, Kellner F, Lanoue A, Thamm AMK, Salim V, Schneider B, et al. A Pair of Tabersonine 16-Hydroxylases Initiates the Synthesis of Vindoline in an Organ-Dependent Manner in *Catharanthus roseus*. Plant Physiol 2013:1792–803. doi: https://doi.org/10.1104/pp.113.222828.
[77] Neag MA, Mocan A, Echeverría J, Pop RM, Bocsan CI, Crişan G, et al. Berberine: Botanical Occurrence, traditional uses, extraction methods, and relevance in cardiovascular, Metabolic, Hepatic, and Renal Disorders. Front Pharmacol 2018:1–30. doi: https://doi.org/10.3389/fphar.2018.00557.
[78] Iv APG. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc 2016:1–20. doi: https://doi.org/10.1111/boj.12385.
[79] Liscombe DK, MacLeod BP, Loukanina N, Nandi OI, Facchini PJ. Evidence for the monophyletic evolution of benzylisoquinoline alkaloid biosynthesis in angiosperms. Phytochemistry 2005:1374–93. doi: https://doi.org/10.1016/j.phytochem.2005.04.029.