UNIVERSITY OF SÃO PAULO

SCHOOL OF PHARMACEUTICAL SCIENCES OF RIBEIRÃO PRETO

Metabolomics and evolution of chemical traits in the subtribe Espeletiinae (Asteraceae)

Metabolômica e evolução de caracteres químicos na subtribo Espeletiinae (Asteraceae)

GUILLERMO FEDERICO PADILLA GONZÁLEZ

Ribeirão Preto

2018

UNIVERSITY OF SÃO PAULO

SCHOOL OF PHARMACEUTICAL SCIENCES OF RIBEIRÃO PRETO

GUILLERMO FEDERICO PADILLA GONZÁLEZ

Metabolomics and evolution of chemical traits in the subtribe Espeletiinae (Asteraceae)

Metabolômica e evolução de caracteres químicos na subtribo Espeletiinae (Asteraceae)

Doctoral thesis presented to the Graduate Program of the School of Pharmaceutical Sciences of Ribeirão Preto/USP to obtain the degree of Doctor in Sciences Concentration Area: Natural and Synthetic Products

Supervisor: Prof. Dr. Fernando Batista da Costa

Corrected version of the doctoral thesis presented to the Post-Graduate Program in Pharmaceutical Sciences on 14/12/2018. The original version is available at the School of Pharmaceutical Sciences of Ribeirão Preto/USP.

Ribeirão Preto

2018

COPYRIGHT PAGE

CERTIFICATE OF APPROVAL

Name: Guillermo Federico Padilla González

Title: Metabolomics and evolution of chemical traits in the subtribe Espeletiinae (Asteraceae).

Doctoral thesis presented to the Graduate Program of the School of Pharmaceutical Sciences of Ribeirão Preto/USP to obtain the degree of Doctor in Sciences.

Concentration Area: Natural and Synthetic Products

Supervisor: Prof. Dr. Fernando Batista da Costa

Approved on:

Examiners

Prof. Dr._____

Institution:_____Signature:_____

Prof. Dr. _____

Institution:_____ Signature:_____

Prof. Dr._____

Institution:_____Signature:_____

Prof. Dr._____

Institution:_____ Signature:_____

Prof. Dr. _____

Institution:_____ Signature:_____

To my mother, whose example, unconditional love and support determined

great part of who I am today.

ACKNOWLEDGMENTS

Nothing makes sense in Biology except in the light of evolution… (Theodosius Dobzhansky)

RESUMO

PADILLA GONZÁLEZ, G. F. **Metabolômica e Evolução de Caracteres Químicos na Subtribo Espeletiinae (Asteraceae).** 2018. 181f. Tese (Doutorado). Faculdade de Ciências Farmacêuticas de Ribeirão Preto – Universidade de São Paulo, Ribeirão Preto, 2018.

A subtribo Espeletiinae (Asteraceae) representa um exemplo clássico de adaptação em ecossistemas tropicais de altitudes elevadas. No entanto, estudos que combinem diferentes campos de pesquisa ainda são necessários para entender este caso proeminente de radiações adaptativas rápidas nos trópicos. Esta tese fornece uma abordagem multidisciplinar combinando informação metabolômica, biogeográfica, taxonômica, evolutiva, química, molecular e ecológica, para um estudo aprofundado da subtribo Espeletiinae e do seu gênero irmão *Smallanthus*. Através de análises metabolômicas baseadas em cromatografia líquida de ultra-alta eficiência acoplada a espectrometria de massas, nós fornecemos, pela primeira vez, evidências metabolômicas de segregação alopátrica em Espeletiinae e evidência metabolômica apoiando a possível segregação do gênero *Espeletia* em dois gêneros diferentes com distintas impressões digitais metabólicas. Em combinação com a filogenia molecular da subtribo e amplificações por PCR, demonstramos que a evolução dos caracteres químicos em Espeletiine seguiu cenários complexos de mudança química com alguns caracteres representando sinapomorfias químicas e outros representando múltiplos ganhos e perdas, implicando em evolução convergente. Por fim, analisando os padrões de expressão dos principais genes envolvidos na biossíntese de ácidos clorogênicos, flavonoides e lactonas sesquiterpênicas, em combinação com análises metabolômicas e informações ambientais, relatamos a regulação ambiental e de desenvolvimento do metabolismo secundário de *Smallanthus sonchifolius*, fornecendo informações relevantes para o entendimento dos fatores regulatórios e possíveis papéis adaptativos dos metabólitos secundários em táxons andinos. Em conclusão, esta tese fornece uma compreensão holística de uma linhagem que representa um exemplo clássico de radiações adaptativas rápidas nos Andes tropicais, abrindo uma nova perspectiva intrigante de pesquisa em outros grupos.

Palavras-chave: Espeletiinae, evolução química, metabolômica, *Smallanthus*.

ABSTRACT

PADILLA GONZÁLEZ, G. F. **Metabolomics and evolution of chemical traits in the subtribe Espeletiinae (Asteraceae).** 2018. 181s. Thesis (Doctorate). Faculty of Pharmaceutical Sciences of Ribeirão Preto – University of São Paulo, Ribeirão Preto, 2018.

The subtribe Espeletiinae (Asteraceae) represents a classic example of adaptation in tropical high-elevation ecosystems. However, studies bringing different research fields are still necessary to understand this prominent case of rapid adaptive radiations in the tropics. This dissertation provides a multidisciplinary approach combining metabolomic, biogeographic, taxonomical, evolutionary, chemical, molecular and ecological information, for an in-depth study of the subtribe Espeletiinae and its sister genus *Smallanthus*. Through metabolomic analyses based on ultrahigh-performance liquid chromatography-mass spectrometry, we provide, for the first time, metabolomic evidence of allopatric segregation in Espeletiinae and metabolomic evidence supporting a putative segregation of the genus *Espeletia* in two different genera with distinctive metabolic fingerprints. In combination with the molecular phylogeny of the subtribe and PCR amplifications, we also demonstrate that the evolution of chemical traits in Espeletiinae followed complex scenarios of chemical change with some traits representing chemical synapomorphies and other traits being gained and lost multiple times implying convergent evolution. Lastly, by analyzing the expression patterns of key genes involved in the biosynthesis of chlorogenic acids, flavonoids and sesquiterpene lactones, in combination with metabolomic analyses and environmental information, we report the developmental and environmental regulation of the secondary metabolism of *Smallanthus sonchifolius*, providing relevant information towards the understanding of the regulatory factors and possible adaptive roles of secondary metabolites in Andean taxa. In conclusion, this dissertation provides a holistic understanding of a lineage representing a classic example of rapid adaptive radiations in the tropical Andes, opening an intriguing new perspective of research in other groups.

Keywords: Espeletiinae, metabolomics, chemical evolution, *Smallanthus*.

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND ACRONYMS

4CL	4-Cinnamoyl CoA ligase
ACT	Actin
AFLP	Amplified fragment length polymorphism
AGC	Automatic gain control
AIF	All ions fragmentation
AsterDB	Asteraceae Database
AT	Annealing temperature
AU	Approximately unbiased p-values
BLAST	Basic local alignment search tool
C3H	p-Coumaroyl-3'-hydroxylase
C4H	Cinnamate 4'-hydroxylase
CADs	Caffeic acid derivatives
CCA	Canonical correspondence analysis
CD$_3$OD	Deuterated methanol
cDNA	complementary DNA
CHS	Chalcone synthase
cpDNA	Chloroplast DNA
dd-MS^2	Data-dependent tandem mass spectrometry
DNP	Dictionary of natural products
EF	Elongation factor
EI	Electron impact ionization source
ESI	Electrospray ionization source
EtOH	Ethanol
ETS	External transcribed spacer
FTMS	Fourier transform mass spectrometry
FWHM	Full width at half maximum
GAO	Germacrene A oxidase
GAS	Germacrene A synthase
GC	Gas chromatography
gDNA	Genomic DNA
GNPS	Global natural product social molecular networking
HCAbp	Hierarchical clustering analysis with bootstrap resampling
HCD	Higher energy collisional dissociation
HCT	hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl transferase
HPLC	High performance liquid chromatography
HQT	Hydroxycinnamoyl CoA:quinate hydroxycinnamoyl transferase
HRMS	High-resolution mass spectrometry
ICR	Ion cyclotron resonance
ID	Identity
IR	Infrared spectroscopy
ITS	Internal transcribed spacer
LC-MS	Liquid chromatography - mass spectrometry

m/z	Mass-to-charge ratio
MALDI	Matrix-assisted laser desorption/ionization
MeCN	Acetonitrile
MeOH	Methanol
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry (in space)
MS^2	Tandem mass spectrometry (in time)
MSI	Metabolomics standards initiative
NCE	Normalized collision energy
NMDS	Non-metric multidimensional scaling
NMR	Nuclear magnetic resonance
nrDNA	Nuclear DNA
OPLS-DA	Orthogonal partial least squares-discriminant analysis
PAL	Phenylalanine ammonia lyase
PCA	Principal component analysis
PCR	Polymerase chain reaction
PDA	Photo-diode array detector
PTFE	Polytetrafluoroethylene
PVP	Polyvinylpyrrolidone
Q^2	Coefficient of prediction
qPCR	Quantitative polymerase chain reaction
R^2	Coefficient of determination
Rt	Retention time
RT-qPCR	Quantitative polymerase chain reaction in real time
SM	Secondary metabolites
STLs	Sesquiterpene lactones
TIC	Total ion current
TOF	Time-of-flight detector
UHPLC	Ultra-high performance liquid chromatography
UV	Ultraviolet
VIP	Variable importance in the projection

TABLE OF CONTENTS

CHAPTER 1: GENERAL INTRODUCTION

METABOLOMICS

As one of the most recent "omics" sciences, metabolomics constitutes a relatively new research field aimed at studying the whole set of metabolites synthesized by a given biological system and their response to genetic or environmental changes (FIEHN, 2002; OLIVER, 1998). The metabolome, similarly to the definition of the proteome (the set of proteins) or the transcriptome (the set of RNAs), can be defined on all levels of complexity, from single cells or cell compartments to tissues and organisms (FIEHN, 2002). Considering that metabolites are the end products of regulatory processes occurring at the cellular level, the types of metabolites and their concentration levels are often the result of complex response mechanisms of biological systems to genetic or environmental disturbance. Thus, the metabolome of a given organism, tissue or cell is often chemically complex in terms of its qualitative and quantitative composition, which represents an important challenge to the current analytical platforms. As a consequence, different metabolomic approaches have been developed to precisely detect, identify and quantify metabolites at different levels, from single compounds from a given chemical class to broad methods aiming to detect as many metabolites as possible in a given organism.

Metabolomic approaches

Given their high structural complexity and the fact that "metabolites have a greater variability of atoms and groups compared to the linear 4-letter codes for genes or the linear 20-letter codes for proteins" (FIEHN, 2002), determining the elemental composition, the linkage order among atoms and the stereochemical configuration of all metabolites in a given biological system remains an elusive task. Furthermore, due to their highly diversified physicochemical properties, varying concentrations and structural complexity, no single analytical platform or extraction protocol is currently able to simultaneously extract and detect all metabolites from a given system. Instead, different metabolomic approaches have been designed in order to answer specific types of questions. In general terms, metabolomic analyses can be categorized in two broad types: targeted and untargeted analyses. Targeted

metabolomics refers to the detection, identification and quantification of a small set of known metabolites in a certain system. On the other hand, untargeted metabolomics is focused on the detection of as many metabolites as possible through comprehensive analyses of complex biological matrices, without previous knowledge of the compounds identities or classes (FIEHN, 2002). More recently, different terms have been adopted by the scientific community and three main metabolomic approaches have been proposed: (1) metabolite targeted analysis, which refers to the "detection and precise quantification of a single or small set of target compounds", (2) metabolic profiling, which focuses on the "identification and approximate quantification of a group of metabolites associated to specific biosynthetic pathways" and (3) metabolic fingerprinting, which is used for "complete metabolome comparison without knowledge of compounds identities", usually by spectral fingerprinting of crude extracts, biological fluids or by direct analysis in real time (KRASTANOV, 2010).

Analytical techniques

By definition, metabolomics, and especially metabolic fingerprinting, must aim at detecting all metabolites present in a given biological system by using comprehensive extraction protocols and analytical techniques. However, this is a very ambitious goal still far from reality when considering the high chemical complexity of biological matrices and the limitations of comprehensive extraction methods and the current analytical platforms in terms of sensitivity and dynamic range of detection (SUMNER; MENDES; DIXON, 2003). Accordingly, no single analytical technique is currently able to detect all metabolites from a complex biological matrix and, similarly, no single extraction protocol is currently able to extract all metabolites from a given system. As a truly comprehensive analysis of the metabolome is currently not feasible, in practice a comprehensive metabolomic analysis should at least include multiple metabolic pathways and cover as many metabolites as possible (SUMNER; MENDES; DIXON, 2003). Therefore, different analytical techniques have been routinely used in metabolomic analyses, including ultraviolet (UV) and infrared (IR) spectroscopy, nuclear magnetic resonance (NMR) and mass spectrometry (MS), among others (SUMNER; MENDES; DIXON, 2003). Since "the dynamic range of many techniques can be severely limited by the sample matrix or the presence of interfering and competing compounds" (SUMNER;

MENDES; DIXON, 2003), in most cases, analytical detectors are coupled to chromatographic separation techniques like gas chromatography (GC) or high-performance liquid chromatography (HPLC) to avoid overlapping or suppression of signals in complex matrices.

Considering that different analytical techniques have different sensitivities and dynamic ranges of detection, the selection of the most suitable technology is generally a compromise between speed, selectivity and sensitivity (SUMNER; MENDES; DIXON, 2003). NMR and MS constitute the two most commonly used methods in metabolomics. However, they still have some limitations. For example, NMR constitutes a widely used technique because of its high dynamic range of detection and speed, but its very low sensitivity (approx. 10^{-6} mol) compared to mass spectrometry (approx. 10^{-15} mol) represents a serious concern as it fails to detect low concentrated metabolites (SUMNER; MENDES; DIXON, 2003). On the other hand, mass spectrometry presents a very good sensitivity able to detect compounds in minor concentrations but its selectivity can be compromised. In order to be detected, metabolites should be first ionized and, for example, in electrospray ionization (ESI), the most commonly used MS-ionization method in metabolomics, differences in the ionization behavior of different chemical classes compromise their detection as compounds in the extremes of the pH scale are more prone to gain or lose protons (or other atoms to form specific adducts) than neutral species favoring their ionization and subsequent detection. Other ionization methods are also possible, but each of them has its own merits and drawbacks are reported by Ernst et al. (2014). In spite of this inherent limitation, MS represents the technique of choice for the analysis of cellular proteins and metabolites (proteome and metabolome respectively) due to the recent development of high-resolution analyzers (such as time-of-flight - TOF, Orbitraps and ion cyclotron resonance - ICR), the different ionization methods available (such as electron impact - EI, ESI, matrix-assisted laser desorption/ionization- MALDI, etc.) and the possibility of performing tandem MS analyses as well as coupling different detectors in a single equipment allowing the detection of a wide range of metabolites on a scale previously unimaginable.

Applications and importance

Metabolomics was first considered as a functional genomics tool aiming to link genotypes with phenotypes by helping in the identification of gene function (FIEHN,

2002). However, in addition to this specific application, which requires its integration with other "omics" sciences (e.g. genomics and transcriptomics), metabolomics is applied in many biological systems, including human, plants and microorganisms as a transdisciplinary science comprising of 'organic chemistry', 'analytical chemistry', 'chemometrics' and 'informatics', among others (KRASTANOV, 2010). Through the use of statistical multivariate pattern recognition methods, metabolomics enables the classification and discrimination of diverse biological status like sample origin and quality, the characterization of metabolic and phenotypic alterations between genetically manipulated and wild-type lineages or between healthy and diseased tissues, the investigation of food composition and quality, the discovery of biomarkers involved in certain biological activity, the monitoring of metabolic responses to developmental or environmental alterations, the understanding of metabolic networks established in symbiotic associations and the fingerprinting of species for taxonomic purposes, among others (KRASTANOV, 2010; SUMNER; MENDES; DIXON, 2003; VILLAS-BÔAS; RASMUSSEN; LANE, 2005). Thus, given its transdisciplinary nature, metabolomics has been applied in several research fields. Yet, plant science is one of the fields where metabolomics has found many of its current applications, especially in the chemical characterization and quality control of botanicals, phytomedicines and crop plants, as well as in the assessment of plant responses to genetic and environmental alterations.

THE SUBTRIBE ESPELETIINAE

The subtribe Espeletiinae Cuatrec. (tribe Millerieae, Asteraceae) represents a morphologically diverse group of 144 species distributed in eight genera: *Carramboa* Cuatrec., *Coespeletia* Cuatrec., *Espeletia* Mutis ex Humb. & Bonpl., *Espeletiopsis* Cuatrec., *Libanothamnus* Ernst, *Paramiflos* Cuatrec., *Ruilopezia* Cuatrec., and *Tamania* Cuatrec., popularly known as "frailejones", endemic to the high altitudes of the tropical Andes of South America (CUATRECASAS, 2013; DIAZGRANADOS, 2012; DIAZGRANADOS; BARBER, 2017). Since the first collection and description of an Espeletiinae species in 1792 by José Celestino Mutis during the "Expedición Botánica del Nuevo Reino de Granada", frailejones have intrigued the scientific community for their unusual life forms, impressive morphological diversity (Figure 1.1) and numerous adaptations to the harsh environmental conditions prevailing at high Andean altitudes (CUATRECASAS, 2013; MONASTERIO; SARMIENTO, 1991).

Figure 1.1 Morphological diversity within the subtribe Espeletiinae Cuatrec. Capitula of: A, *Espeletiopsis pannosa* (Standl.) Cuatrec.; B, *E. purpurascens* (Cuatrec.) Cuatrec.; C, *Tamania chardonii* (A.C.Sm.) Cuatrec.; D, *Espeletia batata* Cuatrec. Leaves of: E, *E. grandiflora* Humb. & Bonpl.; F, *E. chocontana* Cuatrec.; G, *E. barclayana* Cuatrec.; H, *Ruilopezia jabonensis* (Cuatrec.) Cuatrec. Young leaves of:

I, *E. murilloi* Cuatrec.; J, *E. lopezii* Cuatrec.; K, *Coespeletia palustris* Diazgr. & Morillo; L, *E. boyacensis* Cuatrec. Trichomes of: M, *Paramiflos glandulosus* (Cuatrec.) Cuatrec.; N, *R. floccosa* (Standl.) Cuatrec.; O, *E. aristeguietana* Cuatrec.; P, *R. marcescens* (S.F.Blake) Cuatrec. Rosettes of: Q, *R. jahnii* (Standl.) Cuatrec.; R, *E. discoidea* Cuatrec.; S, *Espeletiopsis caldasii* (Cuatrec.) Cuatrec.; T, *T. chardonii.* Habits of: U, *Carramboa badilloi* (Cuatrec.) Cuatrec.; V, *Coespeletia timotensis* (Cuatrec.) Cuatrec.; W, *E. jimenez-quesadae* (Cuatrec.) Cuatrec.; X, *Libanothamnus arboreus* (Aristeg.) Cuatrec. (Photographs from M. Diazgranados.)

Morphologically, frailejones exhibit an exceptional diversity of growth-forms including large trees, dichotomous trees, shrubs and the iconic giant rosettes (Figure 1.1), an unusual life form found only in a few other high elevation ecosystems. Their leaves, reproductive organs and indument display a comparable level of variation (Figure 1.1). The enormous morphological diversity of Espeletiinae is also represented by their remarkable ecological specialization and associations with numerous species of insects and microorganisms (CUATRECASAS, 2013; FAGUA; GONZALEZ, 2007; MONASTERIO; SARMIENTO, 1991). Furthermore, frailejones represent key species of critical ecological importance because they contribute to regulating the hydrologic cycle and prevent soil erosion in the high altitude Andean grasslands (GARCÍA, N., CALDERÓN, E., GALEANO, 2005).

Distribution and biogeography

Espeletiinae comprises a monophyletic group entirely restricted to the high Andean forests and páramos of Colombia, Venezuela and Ecuador, usually located between 2,000 to 4,700 m above the sea level (CUATRECASAS, 2013). Páramos, represent high elevation ecosystems located "along the crests of the highest mountain ranges or on isolated mountaintops, like islands in a sea of forest" (LUTEYN, 1999). The broken topography and altitudinal restriction of Andean páramos, makes them a particular ecosystem that biogeographically function as islands acting as barriers promoting allopatric speciation (LUEBERT; WEIGEND, 2014), a mechanism that has shaped the evolution and biogeographic history of several plant groups including the subtribe Espeletiinae (DIAZGRANADOS; BARBER, 2017; MADRIÑÁN; CORTÉS; RICHARDSON, 2013). Considering that even highly variable molecular markers commonly used to reconstruct phylogenies

and assess biogeography do not have the resolving power needed for a detailed geographic discrimination in Espeletiinae, in chapter 2 we present a metabolomics approach based on ultrahigh-performance liquid chromatography-mass spectrometry to provide metabolomic evidence of allopatric segregation in this Andean-endemic subtribe. Our approach allows distinct degrees of discrimination of Espeletiinae taxa at different geographic scales with characteristic metabolic fingerprints related to the species' country of origin in a global scale, to their páramo massifs in a regional scale and to their páramo complexes in a local scale, revealing inter- and intraspecific metabolic variations, and demonstrating that metabolomic data can provide relevant information for understanding allopatric segregation in recently diversified plant groups where the genetic divergence is still at an early stage.

Evolutionary history and classification

Since the official publication of the genus *Espeletia* by Humboldt and Bonpland in Plantae Aequinoctiales (1809[1808]) and the early works of José Cuatrecasas (1976, 1986), the taxonomy of the frailejones has been the focus of great debate and much disagreement persists over the taxonomic subdivisions of the subtribe, as well as its phylogeny (DIAZGRANADOS; BARBER, 2017; POUCHON et al., 2018). Recent phylogenetic studies based on nuclear (ITS and ETS), chloroplast (rpl16) and AFLPs molecular markers provided the more complete phylogeny of Espeletiinae to date (although with a low resolution), suggesting a very strong influence of geography on phylogenetic relationships and concluding that the current generic classification of the subtribe needs to be deeply revised, in view that most of the genera were recovered as paraphyletic (DIAZGRANADOS; BARBER, 2017). Recently, the first fully resolved phylogeny of all morphological groups of Espeletiinae, using whole plastomes and about one million nuclear nucleotides, was published and it confirmed the strong influence of geography on the evolutionary history of the group and the role of adaptive morphological evolution in the exceptionally rapid radiation of the frailejones (POUCHON et al., 2018). By using a metabolomics, phylogenetic and biosynthetic approach, in chapter 3 we explored chemotaxonomic relationships in Espeletiinae, determined ancestral states of chemical traits and explored the presence of a key gene (Germacrene A oxidase, GAO) involved in the biosynthesis of secondary metabolites in selected members of the subtribe. Our results not only contribute to clarify taxonomic relationships in

Espeletiinae but suggest that specialized metabolites might have contributed to the adaptive success of this subtribe to the Andean páramos.

Secondary metabolites and their regulatory factors

Because secondary metabolites possess a broad range of ecological roles, serving as chemical interface between plants and their environment, it is likely that some of them played a crucial role in evolutionary success of Espeletiinae by helping the plants to survive in their highly specialized ecological niche, the Andean páramos. However, the regulatory factors of the secondary metabolism in Andean taxa from a metabolomics perspective remain poorly understood, and the adaptive roles that secondary metabolites played in Andean groups have yet to be described. By analyzing gene expression patterns, metabolic fingerprints and environmental variables in an integrated approach, in chapter 4 we report the developmental and environmental regulation of the secondary metabolism of *Smallanthus sonchifolius* (yacón), an important Andean crop with medicinal use, considered a sister taxa of the subtribe Espeletiinae. Our results demonstrate that a complex interplay between environmental factors and plant development regulate the secondary metabolism of yacón, providing relevant information towards the understanding of the regulatory factors and possible adaptive roles of secondary metabolites in Andean taxa.

Thus, through a multidisciplinary approach combining metabolomic, biogeographic, taxonomical, evolutionary, chemical, molecular and ecological information, this dissertation provides an in-depth study of the subtribe Espeletiinae and its closely related species, *Smallanthus sonchifolius* (Poepp. & Endl.) H. Robinson. The results obtained herein contribute not only towards a holistic understanding of the biogeographic and taxonomic relationships of a lineage representing a classic example of rapid adaptive radiations in the tropical Andes, but to the understanding of chemical evolution as a whole, including the processes and regulatory factors influencing the metabolic fingerprints of several Asteraceae species, opening an intriguing new perspective of research in other plant groups with adaptive radiations.

CHAPTER 2: METABOLOMIC EVIDENCE FOR ALLOPATRIC SEGREGATION IN ESPELETIINAE (ASTERACEAE): A NEW PERSPECTIVE ON AN ANDEAN ADAPTIVE RADIATION IN SKY ISLANDS

INTRODUCTION

The origin of the current plant diversity in the tropical Andes and the causes of present-day species distribution have long intrigued biogeographers. The high species richness and endemicity of the tropical Andes make it a global plant diversity hotspot, considered one of the most diverse areas on Earth in terms of vascular plant species (BARTHLOTT et al., 2005; MYERS et al., 2000). According to Myers et al. (2000), the Andean mountains represent the most species-rich biodiversity hotspot worldwide with ca. 15% of the world's plant species in only 1% of the world's land surface. Andean uplift and past climatic fluctuations are often considered the causal agents for the explosive speciation and adaptive radiation of several Andean plant groups, as these events provide ecological opportunities by the availability of new habitats with absence of competition (DIAZGRANADOS; BARBER, 2017; GENTRY, 1982; GÓMEZ-GUTIÉRREZ et al., 2017; HUGHES; EASTWOOD, 2006; KOLÁŘ; DUŠKOVÁ; SKLENÁŘ, 2016; VÁSQUEZ et al., 2016). With more than 3,400 species of vascular plants, a large proportion of this diversity is found in the páramo ecosystem (LUTEYN, 1999; PÉREZ-ESCOBAR et al., 2017), a biome regarded as the world's fastest evolving and coolest biodiversity hotspot (MADRIÑÁN; CORTÉS; RICHARDSON, 2013). According to Hughes and Eastwood (2006), "a high proportion of the 3,400 plant species recorded in the northern Andean páramos have evolved in the very recent evolutionary past". Páramos represent high elevation Andean grasslands located between ca. 2,800 and 4,700 m above sea level "along the crests of the highest mountain ranges or on isolated mountaintops, like islands in a sea of forest" (LUTEYN, 1999). Due to their broken topography and altitudinal restriction, páramos form a discontinuous chain of mountain ranges in the South American Andes from western Venezuela to Colombia, Ecuador and northern Perú (Figure 2.1) that biogeographically function as islands, often referred to as "sky islands", that have been proposed to act as barriers promoting allopatric speciation (LUEBERT; WEIGEND, 2014).
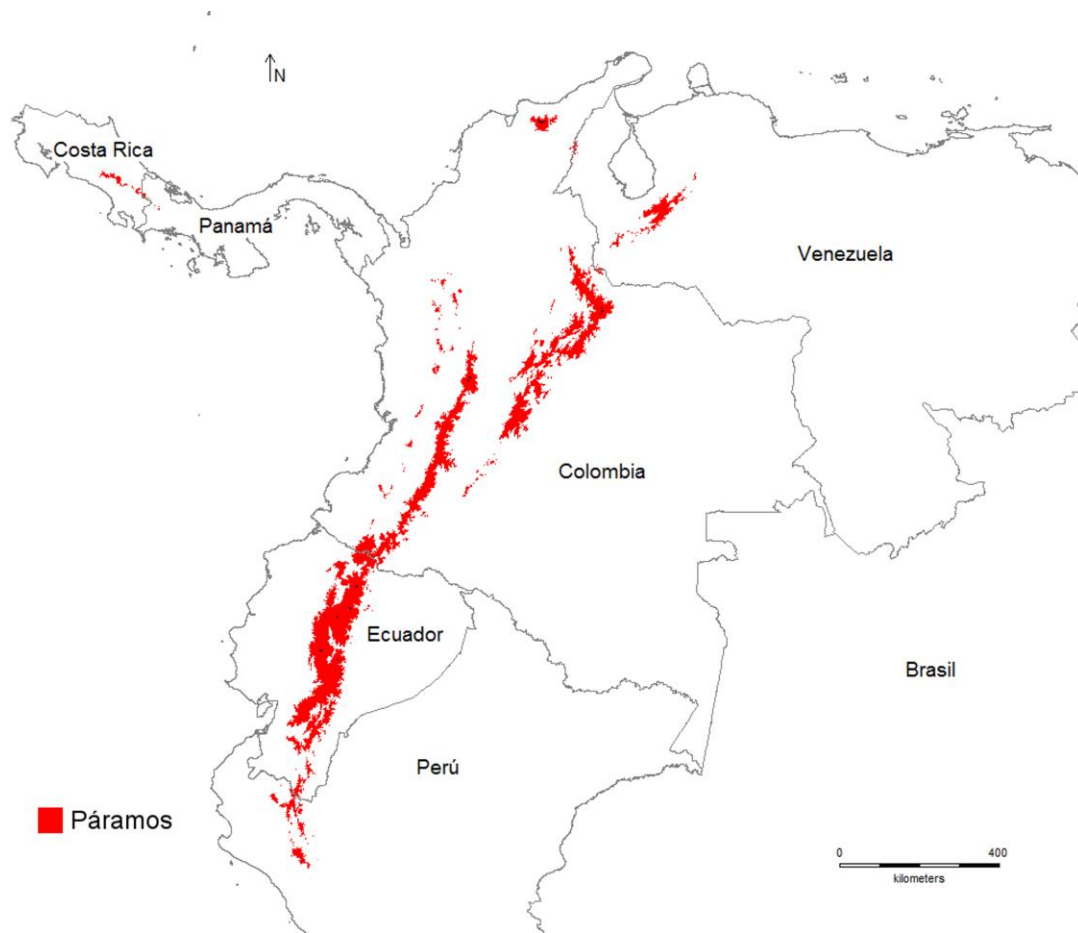
Figure 2.1 Map indicating the present-day area covered by páramos.

With more than 20 known Pleistocene glaciations during the last 3.5 million years, the Andean climate history has been very dynamic (HOOGHIEMSTRA; RAN, 1994; HOOGHIEMSTRA; VAN DER HAMMEN, 2004). Based on palynological records, Hooghiemstra and Ran (1994) determined that Andean vegetation belts moved elevationally by as much as >1000 m in response to the Pleistocene climate oscillations. During glacial maxima, the area of páramos was considerably larger than in interglacial periods, promoting distributional changes of plant communities as repeated events of expansion and contraction of the páramo zone took place (HOOGHIEMSTRA; WIJNINGA; CLEEF, 2006). Andean orogeny and recent climate history are believed to have provided pathways for plant migrations between individual areas (HOOGHIEMSTRA; CLEEF, 1986) and have shaped the evolution and biogeographic history of several plant groups, such as *Loricaria* (KOLÁŘ; DUŠKOVÁ; SKLENÁŘ, 2016), *Oreobolus* (GÓMEZ-GUTIÉRREZ et al., 2017), *Lupinus* (HUGHES; EASTWOOD, 2006; VÁSQUEZ et al., 2016) and Espeletiinae (DIAZGRANADOS; BARBER, 2017).

The subtribe Espeletiinae (Asteraceae, Millerieae) represents the most iconic and representative plant group of the páramo ecosystem with 144+ species, popularly known as frailejones, distributed in eight genera: *Carramboa* Cuatrec., *Coespeletia* Cuatrec., *Espeletia* Mutis ex Humb. & Bonpl., *Espeletiopsis* Cuatrec., *Libanothamnus* Ernst, *Paramiflos* Cuatrec., *Ruilopezia* Cuatrec., and *Tamania* Cuatrec. (CUATRECASAS, 2013; DIAZGRANADOS, 2012; DIAZGRANADOS; BARBER, 2017). As one of the fastest evolving lineages on earth due allopatric speciation, hybrid origin and adaptive morphological evolution (DIAZGRANADOS; BARBER, 2017; MADRIÑÁN; CORTÉS; RICHARDSON, 2013; POUCHON et al., 2018), Espeletiinae is considered an ideal model system for studying rapid adaptive radiations (MONASTERIO; SARMIENTO, 1991). Based on species distribution patterns, frailejones have three main centers of species diversity: The Venezuelan páramo massif of Mérida (with 44 species), and the Colombian massifs of Santander/Norte de Santander (with 41 species) and Boyacá (with 45 species). The Colombian páramos of the Central/Western Cordillera and Cundinamarca represent additional centers with lower diversity (CUATRECASAS, 2013; DIAZGRANADOS; BARBER, 2017). Considering the high endemicity of Espeletiinae taxa and their strong limitations in long-distance seed dispersal and pollination, Cuatrecasas (2013) hypothesized that the evolution of frailejones should reflect their biogeographic history with clades representing their diversification centers. However, based on concatenated analyses including nuclear (ITS and ETS), plastid (rpl16) and AFLP molecular markers, only two large clades were recovered: one of primarily Venezuelan species and a second clade of Colombian species (including one from northern Ecuador) (DIAZGRANADOS; BARBER, 2017). A similar pattern was obtained in a recent phylogenomic analysis based on approximately one million nuclear nucleotides (POUCHON et al., 2018), suggesting that molecular markers are too coarse to reveal subtle biogeographic trends in Espeletiinae.

According to Diazgranados and Barber (2017), evolutionary relationships within numerous Andean recent rapid radiations are still obscure because of the low phylogenetic signal of commonly used molecular markers and the rapid diversification of lineages. In Espeletiinae, the selected markers do not clearly resolve putative genetic-geographic patterns, as the genetic divergence of the subtribe may be too subtle for the studied genes. Therefore, even highly variable molecular markers commonly used to reconstruct phylogenies and assess

biogeography (e.g., ITS and AFLP) do not have the resolving power needed for fine geographic discrimination (DIAZGRANADOS; BARBER, 2017). For instance, in the most complete molecular phylogeny of the frailejones, many species even from different genera share exactly the same sequence in the alignment; trees had low resolution and there was a low percentage of informative characters (13.9%) insufficient to resolve the phylogeny and a detailed geographic structure (DIAZGRANADOS; BARBER, 2017).

Although recent analyses using the latest high-throughput sequencing technologies have facilitated a deeper exploration into the origin and evolution of Espeletiinae and have provided an unprecedented high resolution in the phylogeny of the frailejones (POUCHON et al., 2018), it is still necessary to apply new approaches for understanding biogeographic patterns, especially in recently diversified plant groups where molecular markers are not informative enough to resolve a clear biogeographical segregation (ROSSELLÓ-MORA et al., 2008) or when different types of molecular markers (e.g., plastid and nuclear) are highly discordant, such as in Espeletiinae (POUCHON et al., 2018). In this context, the subtribe Espeletiinae constitutes a good model to evaluate the effect of biogeographic isolation on other genotypic or phenotypic traits. Considering that analyzing the expression of the genotype may lead to a better understanding of the interactions that organisms have with their environment (ROSSELLÓ-MORA et al., 2008), approaches such as metabolomics, which aims to study the set of metabolites synthesized by an organism and their response to genetic or environmental changes, may be considered the link between genotypes and phenotypes (FIEHN, 2002).

Using an ultrahigh-performance liquid chromatography-high-resolution mass spectrometry (UHPLC-HRMS)-based metabolomics approach, we recently showed that a clear biogeographic segregation of 72 individual taxa (120 samples) from the genus *Espeletia* could be discerned at the regional level (PADILLA-GONZÁLEZ; DIAZGRANADOS; DA COSTA, 2017). In the present chapter, we aimed to provide metabolomic evidence for allopatric segregation of the subtribe Espeletiinae at different geographic scales by enlarging the collection to 114 individual taxa (211 samples) including members from five genera: *Espeletia, Espeletiopsis, Libanothamnus, Ruilopezia*, and *Paramiflos.*

METHODS

Plant material

Young leaves from 211 samples of Espeletiinae were collected during major botanical expeditions beginning in 2007 in ca. 70 páramo locations in Colombia and Venezuela (Appendix 1). Samples were collected by Mauricio Diazgranados (Royal Botanic Gardens, Kew, United Kingdom), Petr Sklenář (Charles University, Prague, Czech Republic) and Sandra Castañeda (Jardín Botanico José Celestino Mutis, Bogotá, Colombia) following standard protocols for herbarium and molecular analyses. After collection, all samples were immediately deposited inside resealable zipper storage bags containing silica gel beads with a humidity indicator. Samples contaminated with dirt or fungus were excluded from the extraction process. Taxonomical identification resulted in 169 samples classified to the rank of species (representing 72 different species) and 42 samples classified to the rank of genus (due to their lack of conclusive diagnostic characters), totaling 114 individual taxa belonging to five genera (excluding replicates, Table 2.1).

Table 2.1 Samples of Espeletiinae analyzed in the present study.

Genus	No. species	Unknown species	Individuals/ species[a]	Total samples
Espeletia	65	28	2.3	179
Espeletiopsis	5	0	2.8	14
Libanothamnus	1	8	2.0	10
Ruilopezia	?	6	?	6
Paramiflos	1	0	2.0	2
Total	72	42	2.3	211

[a] Mean values of samples classified to the rank of species.

Information about the exact location, elevation and season of collection of each species is reported in Appendix 1. The same samples used in this study were also used in molecular studies to reconstruct the phylogeny of the group (DIAZGRANADOS; BARBER, 2017). Vouchers of the Colombian species were deposited at the COL, ANDES and JBB herbaria, and the Venezuelan species were deposited at the MER herbarium according to the collection numbers reported in Appendix 1. Collections were made under permits No. 2698 of 09/23/2009 and No. 2

of 02/03/2010 (Ministerio de Ambiente, Colombia) and IE-126 (Venezuela, authorized by Petr Sklenář).

Sample preparation

Two independent experiments were carried out to correlate the metabolome of Espeletiinae with its biogeography. The first experiment focused on the interspecific chemical variability displayed by species collected in five different páramo massifs (Venezuela, Santander/Norte de Santander (hereafter referred to as "Santanderes"), Boyacá, Cundinamarca and Central/Western Cordillera), representing the main centers of biogeographic diversification of the subtribe Espeletiinae (CUATRECASAS, 2013). Thus, different individuals from the same species collected in the same páramo massif (biological replicates) were pooled in a single sample (totaling 114 individual taxa) before grinding and extraction (see below). The second experiment focused on the intraspecific chemical variability displayed by individuals from the same species collected in different páramo complexes within the same massif (lower geographic scale). In this analysis, 25 individuals from three species (*E. argentea*, *E. boyacensis* and *E. grandiflora*) with a wide geographic distribution were extracted and analyzed independently (Appendix 2). In this case, individual samples were collected in seven different páramo localities, five of them located in the massif of Cundinamarca (Páramo of Sumapaz, Chingaza, Cruz Verde, Guerrero and Rabanal) and two in the massif of Boyacá (Iguaque and Guantiva). Three biological replicates were analyzed for each species in each locality.

The extraction process used in this study was based on the protocol of extraction and analysis of plant tissues for metabolomic studies by UHPLC-UV-MS reported by De vos et al. (2007) with some modifications. Plant leaves (20 mg) were ground with liquid nitrogen and transferred to Eppendorf tubes where 2 mL of a 7:3 v/v MeOH-H_2O mixture (LC-MS and Milli-Q grade, respectively) was added. After addition of the extraction solvent, tubes were placed in an ultrasonic bath for 15 min at 25 °C using a frequency of 40 kHz. Samples were subsequently centrifuged at 19,975 × *g* for 10 min, and the supernatant was then partitioned with 0.5 ml of *n*-hexane (HPLC grade) to eliminate fats and pigments and filtered through a 0.2 µm PTFE filter.

To check for the reproducibility of the extraction process, 10 mg of each of the 211 samples were mixed and placed into three different 50 mL falcon tubes where

the extraction process previously described was carried out while keeping the same ratio of plant material and solvent. These three samples represent three different analytical replicates (pool1, pool2 and pool3), which were separately and randomly analyzed in the UHPLC-UV-HRMS, for quality control purposes.

UHPLC-UV-HRMS analysis

The UHPLC-UV-HRMS experiments were performed on an Accela UHPLC (Thermo Scientific, USA) apparatus with an 80 Hz photodiode array detector (PDA) coupled to an ESI-Orbitrap mass spectrometer Exactive Plus (Thermo Scientific).

Chromatographic separation of each plant extract (4 µL) was performed using a core shell C18 column (Kinetex 1.7 µm XB, 150 × 2.1 mm, Phenomenex, USA) connected to a C18 guard cartridge (Security GuardTM Ultra cartridge, Phenomenex). The mobile phase consisted of purified H_2O with 0.1% formic acid (line A) and MeCN with 0.1% formic acid (line B). Separation was performed at a flow rate of 400 µL.min^{-1}. The elution gradient was 0-15 min, 5-50% B; 15-20 min, 50% B; 20-30 min, 50-100% B; 30-35 min (column washing), 100% B; 35-40 min (column equilibration), 100-5% B. The oven temperature was set at 45 ºC, while the auto-injector temperature was kept at 10 ºC. The PDA detector was set to record between 200 and 600 nm.

The column effluent was analyzed by ESI-HRMS (resolution of 70,000 FWHM) and by ESI-HCD MS/MS (resolution of 35,000 FWHM) in both positive and negative ionization modes using the *Full scan* and *All Ions Fragmentation* (AIF) modes. The mass spectra were acquired and processed using the software provided by the manufacturer. Total Ion Current (TIC) chromatograms were recorded between 100 and 1,500 *m/z*. The spray voltage was programmed to +3.6 kV and -3.2 kV for each ionization mode. The capillary and heater temperatures were set at 320 and 300 °C, respectively. Additional parameters for the mass spectrometer include: automatic gain control (AGC) target, 1.0E6; maximum inject time, 100 ms; sheath gas flow rate, 30; auxiliary gas flow rate, 10; sweep gas flow rate, 11; S-lens RF level, 50; and HCD, normalized collision energy (NCE), 35 eV. N_2 was used as the drying, nebulizer and fragmentation gas.

All plant extracts were randomly analyzed, and each of the three analytical replicates (pool1, pool2 and pool3) were injected five times along the

chromatographic sequence to serve as technical replicates (quality control samples) to check for reproducibility in the extraction process and chromatographic analysis.

Data preprocessing

Each set of chromatographic raw data was sliced into two sets according to the ionization mode (positive and negative) and transformed to .mzXML format using the MSConvert package from the software ProteoWizard 3.0.9798 (Proteowizard Software Foundation, USA). Peaks detected in the last five minutes of the chromatographic run (equilibration stage) were excluded from the analyses. The sorted data were processed by MZMine 2.21 (PLUSKAL et al., 2010), to perform peak detection, peak filtering, chromatogram construction, chromatogram deconvolution, isotopic peak grouping, chromatogram alignment, gap filling, duplicate peaks filter, fragment search and the search for adducts and peak identities using an in-house chemical structure database (see below). The following key MZmine parameters were used for data processing: Exact mass for mass detection considering a noise level at 1.0E6; Lorentzian function for the peak shape considering a resolution of 70,000; minimum peak height at 5.0E6; m/z tolerance at 0.002 or 5.0 ppm; retention time tolerance of 0.7 min; XCMS algorithm for chromatogram deconvolution; Ransac method for chromatogram alignment and duplicate peaks filter as filtering algorithm (Appendix 3).

After raw data preprocessing in MZmine, the acquired data were exported in.csv format and edited as Excel spreadsheets (Microsoft Windows, USA), where peaks detected in the blank sample (extraction solvent) were subtracted from the peaks detected in the samples to eliminate possible interfering variables. For each ionization mode (positive and negative), a data matrix containing the species (as rows) versus their peak areas (as columns) was built. A third matrix combining both ionization modes was also considered for the multivariate analyses.

Multivariate analyses and biogeography correlation

The generated matrices were scaled by the Pareto method and submitted to unsupervised and supervised multivariate statistical analyses in R 3.0.3 (R Foundation for Statistical Computing, Austria), Weka 3.6 (University of Waikato, New Zealand), and SIMCA 13.0.3.0 (Umetrics, Sweden). Initially, a Principal Component Analysis (PCA) was performed as an exploratory method to determine trends and

outliers. Subsequently, a supervised analysis by decision trees was carried out to establish correlations between the metabolic fingerprints of the species and their centers of biogeographic diversification. Decision trees were built using the C4.5 algorithm (J48 in Weka 3.6) and the following parameters: Classifier = Trees – J48, BinarySplits = False, ConfidenceFactor = 0.25, Debug = False, MinNumObj = 5, NumFolds = 3, ReducedErrorPruning = False, SaveInstanceData = False, SubtreeRaising = True, Unpruned = False, and UseLaplace = False. To avoid overfitting, a pruning strategy was adopted by configuring MinNumObj to 5, which prevents the software from continuing splitting if the nodes become very small (WITTEN; EIBE; HALL, 2005). In this analysis, 66% of the observations were used to build the model, while the remaining 33% were considered an external validation set. As a complementary validation strategy, repeated training and testing was performed by changing the random seed value from 1 to 10. In all cases, the generated models had a similar performance of correctly classified instances from the external validation group, with an average of 63.3% of correctly classified instances. All multivariate analyses were performed with datasets generated in the positive and negative ionization modes and with a combined dataset including both ionization modes. In all cases, similar percentages of correctly classified instances were obtained in the decision trees. However, considering that nodes splitting pure classes are desired and that in the negative mode dataset a lower number of nodes were achieved, this mode was chosen for further analysis.

Considering that plant samples were collected in different months and years for each geographic locality, we investigated the potential influence of seasonality on the metabolic fingerprints of Espeletiinae. To assign the most likely season to the collected samples, we linked the collection sites and dates with the climatological information of the closest páramos with available climate data. Seasons were coded as "dry season", "rainy season", "rainy season to dry season" and "dry season to rainy season" based on the literature reports (RANGEL-CH, 2000). Subsequently, with the resulting table, including the locality, elevation, season of collection and taxonomic genus, we ran a nonmetric multidimensional scaling (NMDS) analysis in the R package vegan using the Bray-Curtis dissimilarity index and 100,000 permutations. This analysis was performed only for the Colombian samples due to the lack of available climatological information for the Venezuelan páramos.

Additionally, we excluded six taxa from Boyacá (2, 9, 23, 29, 39 and 46, Table S1) whose biological replicates were collected in different seasons.

Identification of discriminant metabolites

Discriminant variables indicated by the multivariate statistical analyses were tentatively identified by comparing accurate mass measurements, UV spectra, fragmentation patterns, and database searches. Compounds were identified with varying levels of confidence according to the four levels reported in the Metabolomics Standards Initiative (MSI) (SUMNER et al., 2007). Level 1 corresponds to the metabolites identified by co-characterization with authentic samples analyzed under identical experimental conditions. Metabolites reported in level 2 were characterized upon accurate mass, UV and fragmentation pattern comparisons with compounds reported in the literature, in the Dictionary of Natural Products (DNP, http://dnp.chemnetbase.com) or in our in-house chemical structure database of the Asteraceae family (AsterDB, www.asterbiochem.org/asterdb). Lastly, the chemical class of metabolites reported in level 3 was tentatively proposed by comparison of physicochemical properties (e.g., UV maximum absorptions, fragmentation patterns and retention time values) with characteristic values of a chemical class, while compounds reported in level 4 correspond to unknown metabolites. Detailed information including chemical name, retention time, *m/z* values, fragmentation patterns and confidence level achieved in the putative identification of metabolites is reported in Appendix 5.

RESULTS

Metabolic fingerprinting and biogeography correlation

UHPLC-UV-HRMS-based metabolic fingerprinting of the 114 taxa considered in the first experiment resulted in the detection of 6,042 and 4,181 mass signals in the positive and negative ionization modes, respectively. Two types of multivariate analyses were performed with all detected mass features in each ionization mode and with a combined dataset including both ionization modes: unsupervised methods by PCA and a supervised method using decision trees. In all cases, similar results were obtained. However, as the negative mode dataset performed better in the decision trees when compared to the positive and combined datasets (Appendix 4), it was chosen for further analysis.

PCA of the negative mode dataset grouped plant extracts by chemical fingerprint similarity. This analysis showed segregation of the samples into two main groups related to the country of origin of each species (Colombia and Venezuela, Figure 2.2). All Venezuelan samples were grouped in the left quadrant of PC1, while Colombian samples were grouped mainly in the right quadrant (Figure 2.2). Analysis of the PCA loadings plot (Figure 2.2) revealed that different mass features characterize each group, indicating a differential accumulation of selected metabolites according to the country of origin of each species. Remarkably, all replicates of the quality control samples were clustered together in the PCA score scatter plot (Figure 2.2), thus suggesting good reproducibility both in the extraction process and in the chromatographic analyses.
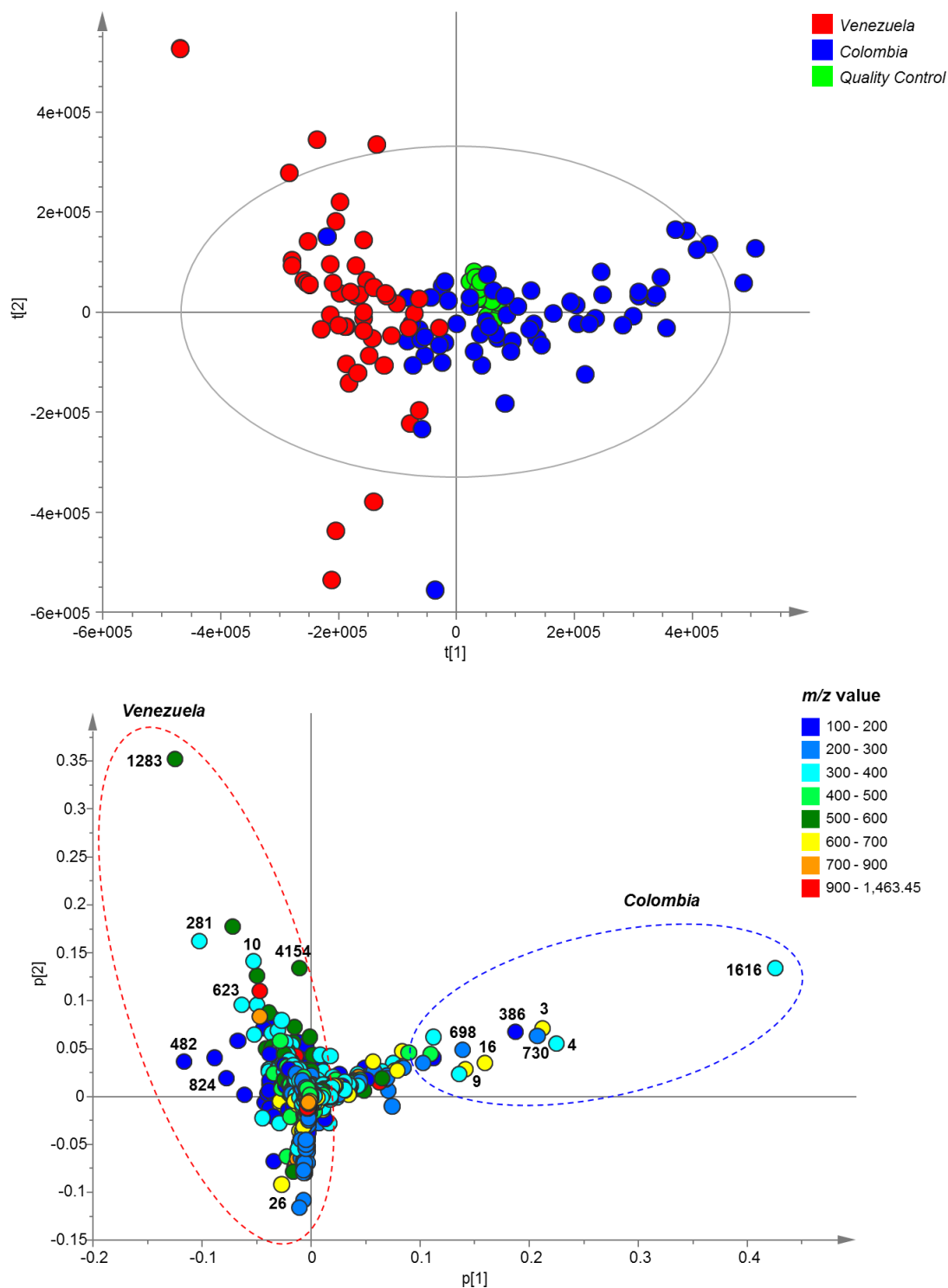
Figure 2.2 PCA based on metabolic fingerprinting in negative mode of 114 Espeletiinae taxa analyzed by UHPLC-UV-HRMS. (A) Scores scatter plot showing the clustering of samples according to their country of origin and (B) loadings plot. Discriminant mass signals in the loadings plot are labeled according to their MZmine ID. For compound identities, refer to Table 2.2.

To further explore the correlation between the metabolic fingerprints of the species and their biogeography, a supervised method by decision trees was performed considering the main páramo massifs where the biogeographic diversification of Espeletiinae took place as the Y variable: Venezuela, Santanderes, Boyacá, Cundinamarca and Central/Western Cordillera (CUATRECASAS, 2013). The generated decision tree (Figure 2.3) showed a high percentage of correctly classified instances (71.8%) of the external validation group, which suggest a high correlation between the metabolic fingerprints of species of Espeletiinae and their páramo massifs of origin. Considering that the decision tree algorithm chooses the variables of the data that most effectively split its set of samples into subsets enriched in one class (WITTEN; EIBE; HALL, 2005), the nodes of the tree are associated to the metabolites (represented by their MZmine ID) that discriminate samples from each geographic locality (Figure 2.3).
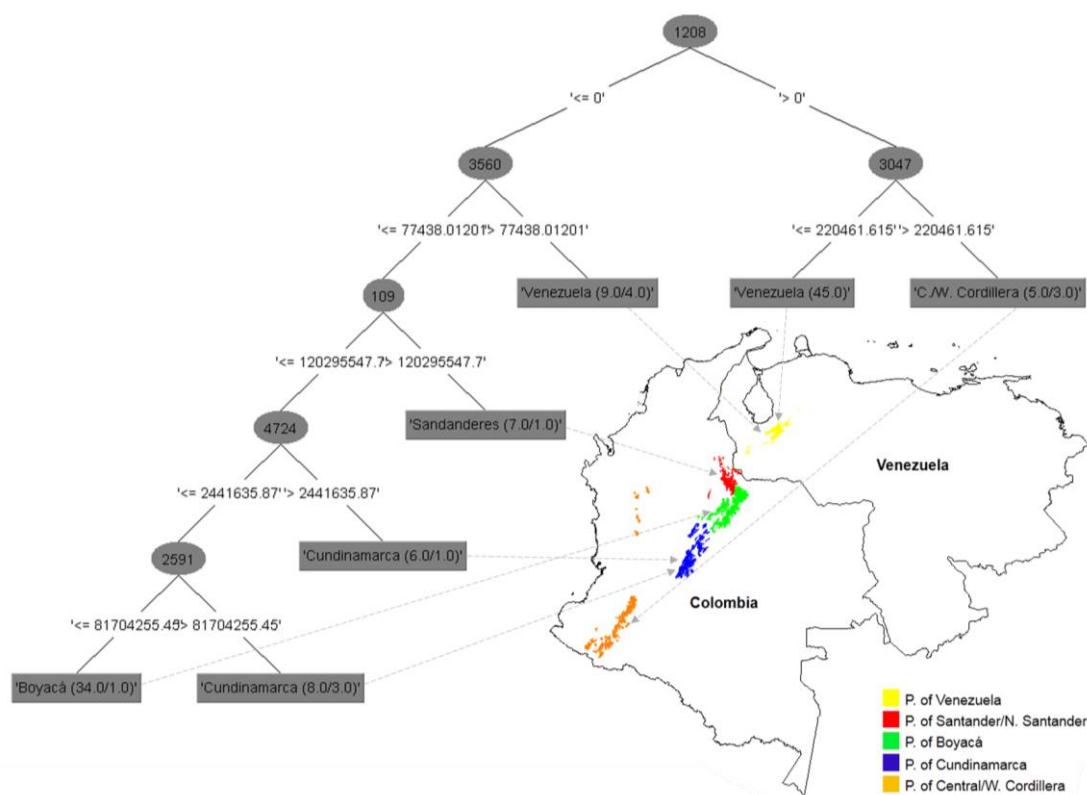
Figure 2.3 Decision-tree based on metabolic fingerprinting in negative mode of 114 Espeletiinae taxa analyzed by UHPLC-UV-HRMS. J48 tree showing the correlation of metabolic fingerprints of the species with the main páramo massifs in Colombia and Venezuela. For compound identities, refer to Table 2.3.

As observed in Figures 2.3 and 2.4, in most cases, the metabolic clustering of species of Espeletiinae reflects the geographic proximity of their respective páramo locations. However, in a particular case, groups of species located in distant geographic localities clustered nearby. Species located in the Colombian Central and Western Cordillera are grouped closely to the Venezuelan taxa (Figure 2.3), despite the large geographic distance between these páramo massifs, while the grouping among the species located in Santanderes, Boyacá and Cundinamarca reflects the geographic proximity of these páramo massifs (Figure 2.3).
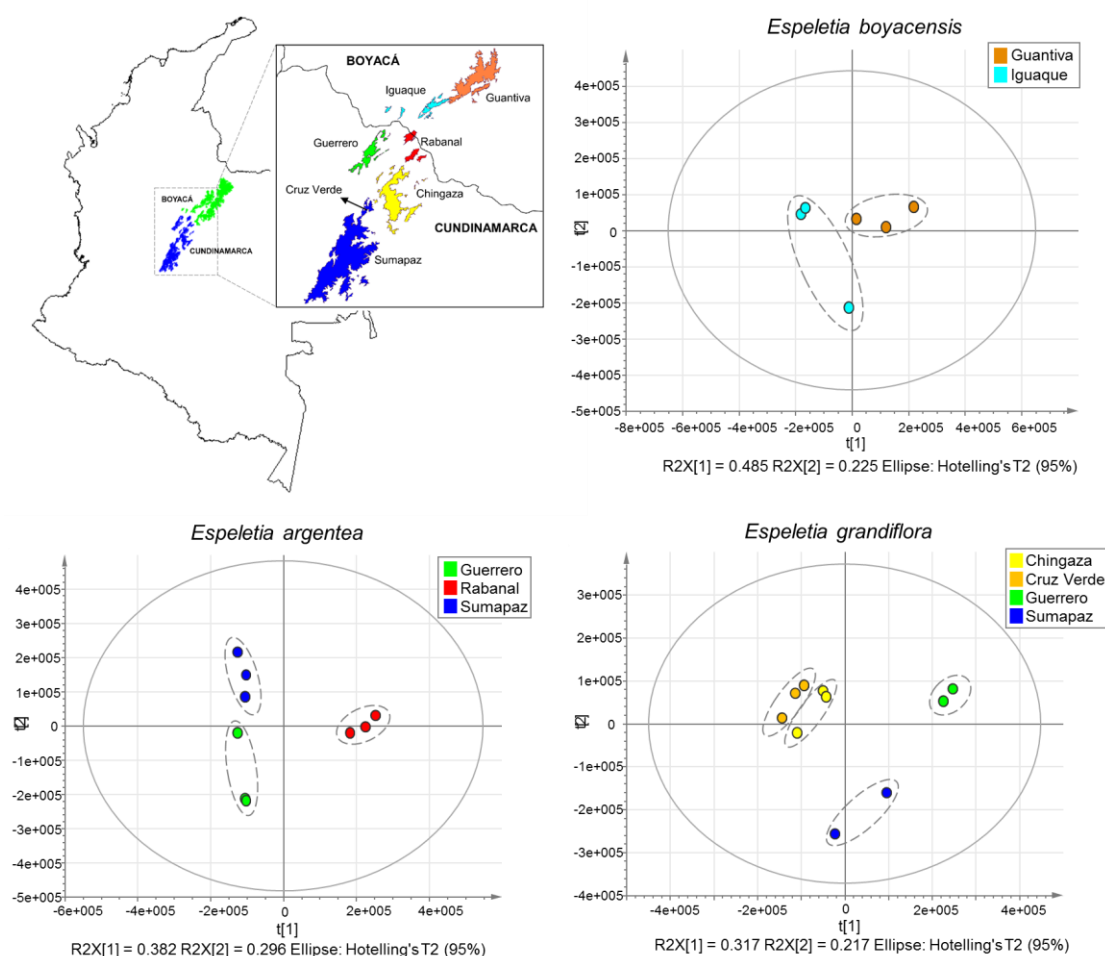
Figure 2.4 PCA based on metabolic fingerprinting in negative mode of different individuals of *Espeletia*. (A) Map of Colombia with different páramo complexes within the Boyacá and Cundinamarca massifs highlighted in different colors. Score plots showing the metabolic clustering of (B) *E. boyacensis*, (C) *E. argentea* and (D) *E. grandiflora* according to their páramo complex of origin.

The UHPLC-UV-HRMS analysis of the second dataset comprising 25 individuals of *Espeletia* belonging to three different species (*E. argentea*, *E.*

boyacensis and *E. grandiflora*) resulted in the detection of 917 mass signals in the negative ionization mode and 1,563 signals in the positive mode. PCA of the negative mode dataset showed that individuals of each species cluster according to their respective páramo complex of origin based on their similar metabolic fingerprints (Figure 2.4). Interestingly, in this analysis even different species collected in the same páramo complex tend to cluster together in a combined PCA scores plot including all individuals from the three species. These results suggest that differences in the metabolic fingerprints of Espeletiinae are not only related to the country and páramo massifs of origin of each species (Figures 2.2 and 2.3) but also, on a lower geographic scale, to the individual páramo complexes within a given massif where the species grow (Figure 2.4). Therefore, our metabolomics approach allows distinct degrees of discrimination of Espeletiinae taxa at different geographic scales with quantitative rather than qualitative differences in metabolite patterns.
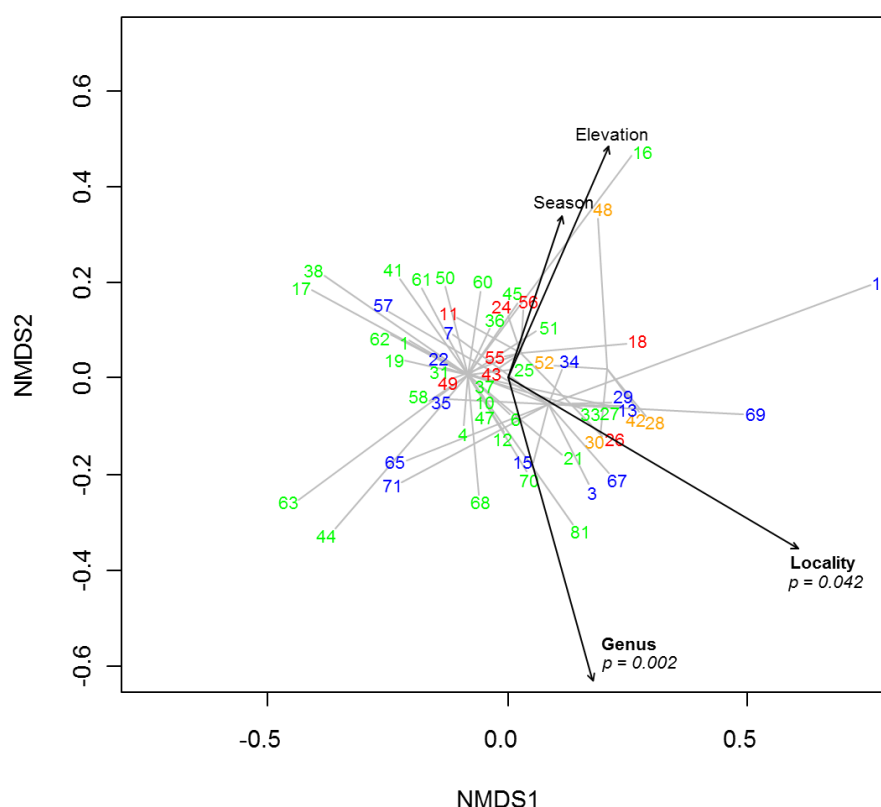
Figure 2.5 NMDS showing the correlation between metabolic fingerprints of Espeletiinae and external variables. Elevation, season, taxonomic genus and locality of collection were considered external variables (*p*-values shown only for significant factors).

To assess the significance of the locality, season, elevation and taxonomic genus (Appendix 1) on the metabolic fingerprints of each species, the same data matrix used in the correlation with biogeography was submitted to NMDS using the Bray-Curtis dissimilarity index (Figure 2.5). This method revealed that among the studied factors, the taxonomic genus and locality of collection were the only statistically significant factors influencing the metabolomes of each species ($p < 0.05$). Therefore, metabolic variations introduced due to the elevation and season of collection (if any) may still be considered minor relative to interspecific variation (represented by taxonomy) or the variation induced by the collection locality on the metabolic fingerprints of species of Espeletiinae.

Accumulation patterns of SM by geography

Analysis of PCA loadings plot (Figure 2.2) and the nodes of the decision tree (Figure 2.3) pointed the discriminant metabolites associated with the geographic origin of each species. As observed in Table 2.2, which shows the discriminant metabolites associated with each country of origin, the clustering of Colombian taxa of Espeletiinae is mainly due to the accumulation of large amounts of the flavonoids 3-methoxy quercetin, quercetin and a putative biflavonoid with a molecular formula of $C_{31}H_{20}O_{14}$. On the other hand, the clustering of the Venezuelan taxa is mainly due to the accumulation of large amounts of the chlorogenic acids 3,5-*O*-dicaffeoylquinic acid, 5-*O*-caffeoylquinic acid and 1,5-*O*-dicaffeoylquinic acid (Table 2.2).

The discrimination between the páramo massif of origin of each species can be related to the differential accumulation of the six metabolites reported in Table 2.3. Interestingly, the decision tree algorithm, which has the advantage of being easily interpreted, usually selects a few variables to effectively segregate samples into subsets. Therefore, in this case, the presence of the metabolite with the molecular formula $C_{27}H_{44}O_9$ (ID 1208, Table 2.3) was sufficient to separate most of the Venezuelan and the Colombian Central/Western Cordillera taxa from the remaining species of Espeletiinae. On the other hand, species from the Colombian Central/Western Cordillera differ from their Venezuelan congeners because the first group accumulates a greater quantity of 5-hydroxyanthraquinone-1,3-dicarboxylic acid relative to the Venezuelan taxa. The clustering of the species from Santanderes can be explained by the accumulation of higher amounts of a glycosylated metabolite with the molecular formula $C_9H_{16}O_9$ when compared to the species from Boyacá and

Cundinamarca (Table 2.3), while these last two groups of species can be segregated by the differential accumulation of methyl 3-*O*-caffeoyl-4-*O*-feruloylquinate, which is higher in species from Cundinamarca.

Table 2.1 Discriminant metabolites from the country of origin of 114 Espeletiinae taxa analyzed by UHPLC-UV-HRMS. Metabolites sorted by their order of importance.

ID	*m/z*	Rt	Identity or molecular formula	MSI level
			Colombia	
1616	315.052	11.0	3-methoxy quercetin	1
4	301.036	10.7	quercetin	1
3	631.109	11.0	dimer of 3-methoxy quercetin	1
730	299.020	15.0	biflavonoid ($C_{31}H_{20}O_{14}$) fragment	3
386	153.018	2.8	protocatechuic acid	1
16	615.079	15.1	biflavonoid ($C_{31}H_{20}O_{14}$)	3
9	695.126	9.8	2,3,5 or 2,4,5-tricaffeoylaltraric acid	1
698	299.020	11.0	3-methoxy quercetin fragment	1
2217	301.072	13.5	isoetin	2
484	209.030	1.0	altraric acid	2
			Venezuela	
1283	515.120	8.3	3,5-dicaffeoylquinic acid	1
281	353.088	4.5	5-caffeoylquinic acid	1
10	329.067	14.5	dimethoxyquercetin	2
4154	515.120	7.9	1,5-dicaffeoylquinic acid	1
623	317.212	21.9	hydroxy-*ent*-kauren-18-oic acid isomer 1	3
482	191.055	1.0	quinic acid	1
824	171.066	7.6	2-formyl tetrahydropyran-3-yl acetate	2
26	609.126	9.9	quercetin-3-*O*-methylpentoside-hexoside	2

Table 2.2 Discriminant metabolites from the páramo massif of origin of 114 Espeletiinae taxa analyzed by UHPLC-UV-HRMS.

ID	*m/z*	Rt	Identity or molecular formula	MSI level
1208	511.292	12.9	$C_{27}H_{44}O_9$	4
3047	311.020	12.6	5-hydroxyanthraquinone-1,3-dicarboxylic acid	2
3560	657.292	10.6	$C_{30}H_{58}O_{15}$	4
109	267.072	1.0	glycosylated metabolite ($C_9H_{16}O_9$)	3
4724	527.235	4.2	$C_{22}H_{40}O_{14}$	4
2591	543.151	8.0	methyl 3-*O*-caffeoyl-4-*O*-feruloylquinate	2

Lastly, analysis of PCA loading plots allowed the identification of the discriminant metabolites from the páramo complexes of origin of 25 individuals belonging to three species of *Espeletia* (Table 2.4). As observed in Table 2.4,

discriminant substances correspond mainly to flavonoids, *trans*-cinnamic derivatives (especially mono- and dicaffeoylquinic acids) and free organic acids, such as gluconic, malic and altraric acids, to a lesser extent.

Table 2.3 Discriminant metabolites from the páramo complex of 25 individuals of *Espeletia* analyzed by UHPLC-UV-HRMS. Metabolites sorted by their order of importance.

ID	*m/z*	Rt	Identity or molecular formula	MSI level
			***E. argentea* P. de Sumapaz**	
1	315.051	11.0	3-methoxy quercetin	1
2	631.110	11.0	dimer of 3-methoxy quercetin	1
5	191.055	0.9	quinic acid	1
9	515.120	8.1	1,5-dicaffeoylquinic acid	1
20	115.003	1.1	$C_4H_4O_4$	4
12	189.076	4.4	$C_8H_{14}O_5$	4
373	485.282	28.3	putative triterpene ($C_{29}H_{42}O_6$)	3
			***E. argentea* P. de Guerrero**	
29	301.072	13.4	isoetin	2
23	329.234	12.6	$C_{22}H_{34}O_2$	4
11	491.342	35.7	putative triterpene ($C_{29}H_{46}O_6$)	3
			***E. argentea* P. de Rabanal**	
3	195.050	0.9	gluconic acid or isomers	3
18	179.055	0.9	$C_6H_{12}O_6$	4
4	133.013	1.0	malic acid or hydroxypropanedioic acid	3
424	165.040	0.9	xylonic acid or isomers	3
8	191.019	1.1	$C_6H_8O_7$	4
71	485.283	34.8	putative triterpene ($C_{29}H_{40}O_6$)	3
			***E. boyacensis* P. de Guantiva**	
1	315.051	11.0	3-methoxy quercetin	1
2	631.110	11.0	dimer of 3-methoxy quercetin	1
579	353.088	4.4	5-caffeoylquinic acid	1
27	695.126	9.7	2,3,5 or 2,4,5-tricaffeoylaltraric acid	1
436	485.283	35.3	putative triterpene ($C_{29}H_{42}O_6$)	3
50	695.126	9.1	tricaffeoylaltraric acid isomer	3
			***E. boyacensis* P. de Iguaque**	
9	515.120	8.1	1,5-dicaffeoylquinic acid	1
8	191.019	1.1	$C_6H_8O_7$	4
4	133.013	1.0	malic acid or hydroxypropanedioic acid	3
3	195.050	0.9	gluconic acid or isomers	3
22	353.088	4.4	5-caffeoylquinic acid	1
45	515.120	8.6	3,4-dicaffeoylquinic acid	1
38	515.120	5.8	1,3-dicaffeoylquinic acid	1
			***E. grandiflora* P. de Chingaza**	

7	209.030	1.0	altraric acid	2
298	485.283	35.5	putative triterpene ($C_{29}H_{42}O_6$)	3
32	329.067	12.9	dimethoxyflavonoid ($C_{17}H_{14}O_7$)	3
37	235.046	1.0	di-acetyl altraric acid ($C_8H_{10}O_8$)	2
818	389.146	4.0	$C_{17}H_{24}O_{10}$	4
42	629.094	11.2	biflavonoid ($C_{32}H_{22}O_{14}$)	3
819	529.303	12.4	$C_{27}H_{46}O_{10}$	4
656	287.233	13.3	$C_{16}H_{32}O_4$	4
783	505.115	17.1	$C_{27}H_{22}O_{10}$	4
			E. grandiflora P. de Cruz Verde	
9	515.120	8.1	1,5-dicaffeoylquinic acid	1
26	153.018	2.6	protocatechuic acid	1
38	515.120	5.8	1,3-dicaffeoylquinic acid	1
27	695.126	9.7	2,3,5 or 2,4,5-tricaffeoylaltraric acid	1
50	695.126	9.1	tricaffeoylaltraric acid isomer	2
35	313.181	15.8	putative diterpene ($C_{20}H_{26}O_3$)	3
22	353.088	4.4	5-caffeoylquinic acid	1
8	191.019	1.1	$C_6H_8O_7$	4
			E. grandiflora P. de Guerrero	
1	315.051	11.0	3-methoxy quercetin	1
2	631.110	11.0	dimer of 3-methoxy quercetin	1
291	485.283	30.2	putative triterpene ($C_{29}H_{42}O_6$)	3
10	615.078	14.9	biflavonoid ($C_{31}H_{20}O_{14}$)	3
16	333.207	16.2	putative diterpene ($C_{20}H_{30}O_4$)	3
999	525.323	26.2	$C_{32}H_{46}O_6$	4
			E. grandiflora P. de Sumapaz	
29	301.072	13.3	isoetin	2
16	333.207	16.2	putative diterpene ($C_{20}H_{30}O_4$)	3
12	189.076	4.4	$C_8H_{14}O_5$	4
20	115.003	1.1	$C_4H_4O_4$	4
929	335.223	14.7	putative diterpene ($C_{20}H_{32}O_4$)	3
856	509.224	7.0	$C_{29}H_{34}O_8$	4
381	293.213	20.4	putative diterpene ($C_{18}H_{30}O_3$)	3

DISCUSSION

This study provides metabolomic evidence for allopatric segregation in the Andean-endemic subtribe Espeletiinae by identifying quantitative metabolic differences at higher taxonomic levels and lower geographic scales compared to our previous report (PADILLA-GONZÁLEZ; DIAZGRANADOS; DA COSTA, 2017). Distinct degrees of discrimination for Espeletiinae taxa at different geographical scales were achieved. Species displayed characteristic metabolic fingerprints related to their country of origin on a global scale, to their páramo massifs on a regional scale and to their páramo complexes on a local scale, revealing inter- and intraspecific metabolic variations. A previous study based on the metabolic fingerprinting of 72 individual taxa (120 samples) from the genus *Espeletia* indicated that a trend for geographic isolation could be discerned at the regional level with quantitative metabolic differences according to the páramo massif of origin of each species (PADILLA-GONZÁLEZ; DIAZGRANADOS; DA COSTA, 2017). In the present study, we enlarged the collection to 114 individual taxa (211 samples) and included members from the closely related genera *Espeletiopsis*, *Libanothamnus*, *Ruilopezia*, and *Paramiflos*. The results obtained herein corroborate the strong influence that biogeography exerted on the metabolic fingerprints of the frailejones at finer geographic scales.

As reported by Rosselló-Mora et al. (2008), although "phenotype studies are being relegated in favor of those based on genomic information, due to the ease of the latter, standard genotypic techniques may not always help in clearly resolving biogeographical trends". Contrary to studies on other Andean groups, such as *Oreobolus* (GÓMEZ-GUTIÉRREZ et al., 2017) and *Loricaria* (KOLÁŘ; DUŠKOVÁ; SKLENÁŘ, 2016), the use of conventional molecular markers (e.g., ITS and AFLPs) for studying biogeographical trends was not suitable in Espeletiinae due to the high sequence similarity and low resolution of these markers, failing to reveal subtle biogeographical trends (DIAZGRANADOS; BARBER, 2017). A study aiming to reconstruct the evolutionary history of the frailejones based on nrDNA (ITS and ETS), cpDNA (rpl16) and AFLP fragments identified two large clades related to the country of origin of the species (Colombia and Venezuela), suggesting a strong influence of geography in shaping the evolutionary history of the frailejones (DIAZGRANADOS; BARBER, 2017). However, the selected markers were not informative enough to

resolve a clear biogeographical segregation of Espeletiinae related to their páramo massifs or individual páramo complexes (DIAZGRANADOS; BARBER, 2017). This suggests that, in Espeletiinae, the genetic divergence is still at an early stage and molecular markers commonly used in other groups cannot render clearly resolvable trends, confirming the recent rapid radiation of the frailejones and the need for other sources of information to reveal those trends. Although a recent phylogenomic analysis based on whole plastomes and approximately one million nuclear nucleotides produced the first fully resolved phylogeny of all morphological groups of Espeletiinae and confirmed the strong influence of geography (POUCHON et al., 2018), the poor sampling, especially among Colombian taxa, hindered a detailed correlation with geographical data. Whole-genome comparisons of a representative set of Espeletiinae samples covering the entire range of geographic distribution for the subtribe might, in the future, indicate which genes could be useful for understanding allopatric segregation based on genetic drift.

Our data not only corroborates the results obtained by molecular markers, as based on metabolic fingerprints Espeletiinae taxa also cluster in two main groups related to their country of origin (Figure 2.2), but also provide evidence of segregation at finer geographic scales, partially supporting the migration routes and diversification centers proposed by Cuatrecasas (2013). The genetic and metabolic differentiation between Colombian and Venezuelan species of Espeletiinae can be related to the biogeographic history of the subtribe and its limitations in seed dispersal and lack of associated long distance pollinators (BERRY; CALVO, 1989; CUATRECASAS, 2013; FAGUA; GONZALEZ, 2007; SMITH, 1981). Based on palynological evidence and his knowledge of the group, Cuatrecasas (1986, 2013) proposed that the origin of the subtribe should be located in the páramo massif of Mérida (Venezuela) and that the origin of the subtribe occurred between 4.5 and 2.5 million years ago during the last stages of the uplift of the Venezuelan Andes (CUATRECASAS, 2013). Phylogenomic analyses confirmed this hypothesis but restricted the origin of the frailejones to 2.3 million years ago (Late Pliocene – early Pleistocene), with major diversifications occurring during the last one million years (POUCHON et al., 2018). For instance, Espeletiinae is considered one of the fastest evolving plant groups on Earth with an average rate of 3.09 species/Myr (POUCHON et al., 2018). Although a previous hypothesis suggested that Espeletiinae crossed the Táchira depression at the border of Colombia and Venezuela during the Pleistocene glacial periods to subsequently

colonize the Colombian páramos of the Eastern Cordillera (CUATRECASAS, 2013), recent phylogenomic evidence suggest otherwise. According to Pouchon et al. (2018), the common ancestor of Espeletiinae appeared in today's border between Colombia and Venezuela, and two rather independent radiations took place in each country, while dispersal events across the Táchira depression never involved lineages belonging to the genus *Espeletia* or *Espeletiopsis*. Therefore, the Colombian radiation followed mainly southward migrations along the Eastern Cordillera, where frailejones established themselves and diversified in three main páramo massifs - Santander/Norte de Santander, Boyacá and Cundinamarca to reach, finally, the páramos of the Colombian Central and Western Cordillera (Figure 2.3) and those of northern Ecuador (CUATRECASAS, 2013; DIAZGRANADOS; BARBER, 2017). The clustering pattern observed in the present study based on metabolomic evidence resembles this geographic diversification of the subtribe (Figure 2.3). It is likely that the phylogenetic and metabolic clustering of the species represent the most recent vicariance events that occurred after the last glacial when populations of Espeletiinae were already established along the Colombian Andes. According to Diazgranados and Barber (2017) "Expansions (with reconnection) and contractions (with isolation) of the páramo ecosystem during the Pleistocene glaciations and interglaciations, respectively, could have played a major role in the radiation and dispersion of these taxa". Therefore, the biogeographic segregation of Espeletiinae based on metabolomic data suggests that geographic isolation not only favored speciation mechanisms (allopatric speciation) but also the accumulation of metabolic differences among individuals from different localities, possibly as a response to unique selective pressures in these high-elevation environments.

The extensive speciation of Espeletiinae in the Colombian Eastern Cordillera can be traced to the diverse and broken topography of the area, especially in the páramo massif of Boyacá, where the Eastern Cordillera reaches its maximum width and complexity (CUATRECASAS, 2013; DIAZGRANADOS; BARBER, 2017). According to Cuatrecasas (2013), the high diversity of *Espeletia* species in the páramos of Boyacá points to this massif as the most important center of speciation in the Colombian Eastern Cordillera. Metabolomic comparisons among populations from *E. boyacensis* located in two of the five páramo complexes within the Boyacá massif (Guantiva and Iguaque, Figure 2.4) revealed important differences in their metabolic fingerprints, with a higher accumulation of 3-methoxy quercetin by

individuals from the Guantiva complex (Table 2.4). This result indicates that the geographic isolation of the different páramo complexes within the Boyacá massif might influence the metabolic fingerprints of its populations in a similar way to the pattern observed for the different páramo massifs (Figure 2.3). Analogous results were obtained when comparing the metabolic fingerprints of populations from *E. grandiflora* and *E. argentea* collected in the four páramo complexes (Guerrero, Rabanal, Chingaza and Cruz Verde-Sumapaz, Figure 2.4) within the Cundinamarca massif. As reported by Cuatrecasas (2013), "certainly, *Espeletia* has existed in the páramos of Cundinamarca long enough to become well-established and diversify into at least 11 distinct species", some of which are locally endemic to individual páramos within the Cundinamarca massif (e.g., *E. sumapacis*).

In most cases, the metabolic clustering of species of Espeletiinae reflects the geographic proximity of their respective páramo locations. Species located in the páramos of Santanderes, Boyacá and Cundinamarca, corresponding to three neighboring geographic localities, are metabolically very similar and cluster together in the decision tree (Figure 2.3). The close clustering distance among individuals from *E. grandiflora* collected in the páramo of Cruz Verde and Chingaza to those from the páramo of Sumapaz also supports this observation. There is a closer geographic distance among Cruz Verde and Chingaza, although Cruz Verde and Sumapaz currently belong to the same complex (Figure 2.4). This result suggests that a closer genetic distance could also exist among populations of *E. grandiflora* from these two páramo localities (Cruz Verde and Chingaza), but additional analyses at the population level are necessary to shed light on this issue. On the other hand, in some cases, species located in distant geographic localities clustered nearby, such as the clustering observed between the Colombian Central/Western Cordillera and the Venezuelan taxa (Figure 2.3). This observation could be the result of chemical convergence resulting from similar selective pressures acting on plants from different localities.

Interestingly, NMDS analysis considering metabolomic data along with environmental factors and taxonomic information (Figure 2.5) pointed to the locality of collection and taxonomic genus as statistically significant factors influencing the metabolic fingerprints of the species, while the season and elevation had no significant influence. This result indicates that although there is a clear geography-based grouping in which even species from different genera collected in the same

geographic area are metabolically very similar, the different genera considered here share not only diagnostic morphological traits but also characteristic metabolic fingerprints. This suggests that chemical traits might be potentially useful to clarify or support taxonomic relationships in the subtribe. A topic explored in detail in the following chapter.

In conclusion, based on comprehensive metabolomic studies combined with statistical methods and machine learning algorithms, our results provide metabolomic evidence for allopatric segregation in the Andean-endemic subtribe Espeletiinae at different geographical scales revealing inter- and intraspecific metabolic variations, opening an intriguing new perspective for research of other plant groups with adaptive radiations. Together with studies based on the distribution patterns of Espeletiinae as well as its molecular phylogeny, our results suggest that biogeography not only shaped the evolution of the frailejones but also determined its metabolic fingerprints where each location holds different species sharing certain metabolic traits and similar DNA sequences. Considering that metabolic differences were related to quantitative composition yields, rather than qualitative production of distinct metabolites, such differences could be attributed to transcriptional or posttranscriptional regulations (e.g., differential levels of gene expression by geographic locality) rather than composition changes in genes at the genomic level, possibly as a response to unique selective pressures in these high-elevation environments. Therefore, our results suggest that putative changes in gene expression levels could play an important role in the adaptation and diversification of Espeletiinae lineages, similar to the observed trend in Andean lineages on the genus *Lupinus* (NEVADO et al., 2016). However, future studies aiming to identify the present or past selective pressures shaping the species metabolome in each locality are still necessary for a deeper understanding of this remarkable adaptive radiation in sky islands.

CHAPTER 3: METABOLOMICS-BASED CHEMOTAXONOMY OF ESPELETIINAE (ASTERACEAE) AND EVOLUTION OF CHEMICAL TRAITS: AN INTEGRATIVE METABOLOMIC, PHYLOGENETIC AND BIOSYNTHETIC APPROACH

INTRODUCTION

Although the tropical Andes represents a global biodiversity hotspot holding some of the most spectacular plant radiations documented to date (HUGHES; EASTWOOD, 2006; MADRIÑÁN; CORTÉS; RICHARDSON, 2013), relationships among some Andean taxa and local drivers of rapid plant diversifications in this area have been the focus of great debate (GENTRY, 1982; HUGHES; EASTWOOD, 2006; LUEBERT; WEIGEND, 2014; NEVADO et al., 2016). Three broad mechanisms have been proposed to explain the high species richness and impressive endemicity of the tropical Andes: ecological opportunity, given by Andean orogeny and the emergence of new habitats with absence of competition; Pleistocene climate fluctuations, promoting distributional changes in plant communities followed by geographic isolation and allopatric speciation; and evolutionary innovation, providing key adaptive advantages and increased fitness in locally adapted Andean taxa (HOOGHIEMSTRA; CLEEF, 1986; HOOGHIEMSTRA; VAN DER HAMMEN, 2004; HUGHES; EASTWOOD, 2006; MONASTERIO; SARMIENTO, 1991). The páramo ecosystem, located within the tropical Andes, has been the stage for the rapid diversification of several plant lineages, where these mechanisms are considered to have played a major role (MADRIÑÁN; CORTÉS; RICHARDSON, 2013).

With an estimated age of 2 to 4 million years (HOOGHIEMSTRA; RAN, 1994; HOOGHIEMSTRA; VAN DER HAMMEN, 2004), páramo constitutes a recently formed high-elevation ecosystem characterized by its extreme climatic conditions, which include high levels of ultraviolet radiation, daily seasonality, low mean annual temperatures and poor soils (CUATRECASAS, 2013; MONASTERIO; SARMIENTO, 1991). Given its altitudinal restriction, recent origin and climate characteristics, the páramo vegetation contains some of the most remarkable adaptations and astonishing and beautiful life forms on earth. For instance, páramos represent the world's fastest evolving biome (MADRIÑÁN; CORTÉS; RICHARDSON, 2013), where local endemism at the specific level can be as high as 90%, possibly as a result of

island-like radiation on a continental scale (DIAZGRANADOS; BARBER, 2017; HUGHES; EASTWOOD, 2006). According to Hughes and Eastwood (2006) "on recently formed islands and island-like formations such as lakes and mountains, rapid diversification has been attributed to ecological opportunities afforded by the availability of new habitats with the absence of competition". In the tropical Andes, these new habitats (i.e., páramos) became available after Andean uplift, while past climatic fluctuations allowed a progressive colonization of otherwise unreachable geographically isolated páramos, especially in groups with strong limitations in long-distance seed dispersal and pollination (i.e., Espeletiinae). Although ecological opportunity is widely accepted as a determining factor in the rapid diversification of Andean lineages, the role of other factors, such as evolutionary innovation, has not been widely recognized (HUGHES; EASTWOOD, 2006) and was only recently suggested to have played a relevant role in the diversification of Andean taxa, such as the genus *Lupinus* (CONTRERAS-ORTIZ et al., 2018; NEVADO et al., 2018). Furthermore, considering that in rapid adaptive radiations the morphological diversity is usually not accompanied by genetic divergence (GIVNISH, 1997), taxonomic relationships based on traditional classification systems do not usually represent monophyletic clades in numerous Andean groups, such as in the subtribe Espeletiinae (DIAZGRANADOS; BARBER, 2017; POUCHON et al., 2018).

Espeletiinae (Asteraceae) represents a monophyletic subtribe considered a classic example of adaptation in tropical high-elevation ecosystems (DIAZGRANADOS; BARBER, 2017; POUCHON et al., 2018). The subtribe is currently subdivided in eight genera: *Carramboa* Cuatrec. (4 spp.), *Coespeletia* Cuatrec. (8 spp.), *Espeletia* Mutis ex Humb. & Bonpl. (72 spp.), *Espeletiopsis* Cuatrec. (23 spp.), *Libanothamnus* Ernst (11 spp.), *Paramiflos* Cuatrec. (1 sp.), *Ruilopezia* Cuatrec. (24 spp.) and *Tamania* Cuatrec. (1 sp.), and exhibits an enormous morphological diversity, ranging from small sessile rosettes a few centimeters high to well-branched trees over 15 m tall (DIAZGRANADOS; BARBER, 2017). According to Diazgranados and Barber (2017), numerous synapomorphies characterize the subtribe, such as "spiral leaf phyllotaxis; obpyramidal to prismatic shape of the glabrous and epappose cypselae; fertile female ray flowers and functionally male disc flowers; pluriseriate involucre and persistent pales of the receptacle; thick and woody stems; xeromorphic structure; specialized life-forms; and a static chromosome number (n = 19)" (CUATRECASAS, 1976, 2013). Among the

distinctive morphological features of Espeletiinae, some key traits, such as the rosette growth-form, foliar pubescence and insulation of the stem with dead leaves are strongly associated with the high diversification rates seen in Espeletiinae (which can reach 3.1 spp/My), as they enabled the adaptation and colonization of the Andean highlands (MONASTERIO; SARMIENTO, 1991; POUCHON et al., 2018). Thus, Espeletiinae radiation is notable not only in terms of net rates of species diversification but also because of its strong adaptive component.

As one of the fastest evolving lineages on earth due allopatric speciation, hybrid origin and adaptive morphological evolution (CUATRECASAS, 2013; DIAZGRANADOS; BARBER, 2017; MADRIÑÁN; CORTÉS; RICHARDSON, 2013; POUCHON et al., 2018), Espeletiinae is considered an ideal model system for studying rapid adaptive radiations (MONASTERIO; SARMIENTO, 1991). For instance, since the first collection and description of an Espeletiinae species in 1792 by José Celestino Mutis during the "Expedición Botánica del Nuevo Reino de Granada", frailejones have intrigued botanists for over two centuries (CUATRECASAS, 2013). However, disagreement persists over the taxonomic subdivisions of the frailejones, and there is no consensus regarding the current generic limits proposed by Cuatrecasas (2013). Based on ITS, ETS, rpl16 and AFLPs molecular markers, Diazgranados and Barber (2017) concluded that the current generic classification of the subtribe needs to be deeply revised, as most of the genera were recovered as paraphyletic and pointed to geography as a driving factor shaping the phylogeny of the frailejones. Furthermore, a recent phylogenomic analysis based on whole plastomes and approximately one million nuclear nucleotides confirmed the paraphyletic nature of most of the genera and highlighted the influence of some key morphological traits and geographical isolation on the rapid diversification of the frailejones (POUCHON et al., 2018), suggesting that both adaptive (i.e., morphological evolution) and non-adaptive processes (i.e., geographic isolation) contributed to the rapid diversification of Espeletiinae.

Although ecological opportunity, in the form of extrinsic circumstances and events, is considered to play a dominant role in driving peak episodes of species diversifications (HUGHES; EASTWOOD, 2006), evolutionary innovations have also played an important role in adaptation and diversification of several taxa in the Andean highlands, including the subtribe Espeletiinae (CONTRERAS-ORTIZ et al., 2018; MONASTERIO; SARMIENTO, 1991; POUCHON et al., 2018). However, most

studies have focused on morphological innovations, and there are no studies reporting the putative role of chemical evolution in the adaptation and diversification of Andean taxa. Considering that secondary metabolites contribute to the adaptive success of plant species by serving as a chemical interface between the plant and its environment (SAMPAIO; EDRADA-EBEL; DA COSTA, 2016), it is plausible that chemical innovations might have precipitated rapid diversifications in several Andean groups, such as Espeletiinae, by providing increased fitness allowing adaptation to the harsh environmental conditions of the páramo ecosystem or as defense traits involved in ecological interactions (HARBORNE, 1989). For instance, Ehrlich and Raven (1964) hypothesized that plant diversification is often the result of innovation in defenses against natural enemies (EHRLICH; RAVEN, 1964). Since secondary metabolites play a vital role as defense or signal compounds, their occurrence apparently reflects both adaptations to particular life strategies and common descent. Hence, "the distribution of secondary metabolites in plants offers information on the underlying evolutionary and ecological processes and strategies embedded in a phylogenetic framework" (WINK, 2003).

Considering that taxonomic relationships among genera and species of Espeletiinae remain largely unresolved and that other phenotypic traits might be equally involved in the adaptation and rapid diversification of Espeletiinae to the Andean páramos, in this chapter, we applied an untargeted metabolomics approach by UHPLC-HRMS to aid in the taxonomic classification of the frailejones and determined ancestral character states of key metabolic traits potentially involved in the adaptation of the frailejones to the harsh environmental conditions prevailing in the páramo ecosystem. Previous studies using an UHPLC-HRMS-based metabolomics approach have demonstrated the potential of metabolomics as an auxiliary chemotaxonomic tool aiding in the classification of several plants and microorganisms (GALLON et al., 2018; KIM et al., 2016; MARTUCCI et al., 2014). Therefore, as an ideal classification system should consider different sources of information (e.g., morphological, molecular and chemical) in an integrated approach, exploring chemotaxonomic relationships and chemical evolution in Espeletiinae by an untargeted metabolomics, phylogenetic and biosynthetic approach might provide relevant information towards a holistic understanding of a lineage representing a classic example of rapid adaptive radiations in the tropical Andes.

METHODS

Plant material

Two hundred eleven samples of Espeletiinae (Table 3.1) were collected by Dr. Mauricio Diazgranados (Royal Botanic Gardens, Kew, United Kingdom), Dr. Petr Sklenář (Charles University in Prague, Czech Republic) and Sandra Castañeda (Jardín Botanico de Bogotá, Colombia) following standard protocols for herbarium and molecular analyses. Samples were collected during major botanical expeditions beginning in 2007 in ca. 70 páramo locations in Colombia and Venezuela (Appendix 1), under the following collection permits: No. 2698 of 09/23/2009 and No. 2 of 02/03/2010 (Ministerio de Ambiente, Colombia), and IE-126 (Venezuela, authorized by Petr Sklenář). Among the collected samples, 169 were classified to the rank of species and 42 to the rank of genus due to the lack of diagnostic characters in these samples, totaling 114 individual taxa belonging to five genera in addition to 97 biological replicates (mean value of 2.3 individuals per species).

Table 3.1 Samples of Espeletiinae analyzed in the present study.

Genus	No. species	Unknown species	Individuals/ species[a]	Total samples
Espeletia	65	28	2.3	179
Espeletiopsis	5	0	2.8	14
Libanothamnus	1	8	2.0	10
Ruilopezia	?	6	?	6
Paramiflos	1	0	2.0	2
Total	72	42	2.3	211

[a] Mean values of samples classified to the rank of species.

After harvesting the plant leaves in their natural habitat, they were immediately deposited inside resealable zipper storage bags containing silica gel beads with a humidity indicator. The same samples used in this study were also used in molecular studies to reconstruct the phylogeny of the group (DIAZGRANADOS; BARBER, 2017). Contaminated samples were excluded from all analyses. Detailed information about the exact location, elevation and season of collection of each taxa is reported in Appendix 1, along with the collection numbers. Vouchers of the Colombian species were deposited at the COL, ANDES and JBB herbaria, while Venezuelan species were deposited at the MER herbarium.

Sample preparation

In this study, individuals from the same species collected in the same páramo massif (biological replicates) were grouped in a single sample (totaling 114 individual taxa) before grinding and extraction to minimize intraspecific metabolic differences, thereby focusing on interspecific metabolic variations. Homogenization of the silica-dried plant leaves was performed in liquid nitrogen with a mortar and pestle. A total of 20 mg of the resulting plant powder was transferred to Eppendorf tubes and extracted with 2mL MeOH: H2O (7:3, v/v) following the protocol for extraction and analysis of plant tissues for metabolomic studies by UHPLC-UV-MS reported by De Vos et al. (2007). Plant extracts were then vortexed for 10 seconds and subsequently placed in an ultrasonic bath for 15 min at 25 °C, using a frequency of 40 kHz. Lastly, samples were centrifuged at 19,975 × g for 10 min, and the supernatant was then partitioned with 0.5 mL of *n*-hexane (HPLC grade) to eliminate fats and pigments and filtered through a 0.2 μm PTFE filter (DE VOS et al., 2007).

To guarantee the reproducibility of the extraction process, 10 mg of each of the 211 samples were mixed in a single sample and placed into three different 50 mL falcon tubes, where the extraction process previously described was carried out keeping the same plant: solvent ratio. These three samples represent three different analytical replicates (pool1, pool2 and pool3), which were analyzed independently in the UHPLC-UV-HRMS, serving for quality control purposes.

UHPLC-UV-HRMS analysis

UHPLC-UV-HRMS analyses were carried out in a Thermo Scientific (USA) equipment, composed of two Accela 1250 quaternary pumps, an Accela photodiode array detector (PDA), and an *Exactive Plus* high-resolution mass spectrometer equipped with an Orbitrap mass analyzer.

Four microliters of each plant extract were injected in a core shell C18 column (Kinetex 1.7 μm XB, 150 × 2.1 mm, Phenomenex) connected to a C18 guard cartridge (Security GuardTM Ultra cartridge, Phenomenex) for chromatographic separation. The mobile phase flow was 400 μL.min^{-1} and was composed of water (A) and MeCN (B), both with 0.1% formic acid, at the following elution gradient: 0-15 min, 5-50% B; 15-20 min, 50% B; 20-30 min, 50-100% B; 30-35 min (column washing), 100% B; 35-40 min (column equilibration), 100-5% B. The oven temperature was

kept at 45 ºC, while the autoinjector temperature was maintained at 10 ºC. The PDA scan wavelength was set to record between 200 and 600 nm.

Mass spectrometry detection of the column effluent (ionized by electrospray) was operated under fast scan-to-scan polarity switching using the *Full scan* (resolution of 70,000 FWHM) and *All Ions Fragmentation* MS/MS (AIF, resolution of 35,000 FWHM) modes. Total Ion Current (TIC) chromatograms were recorded between 100 and 1,500 *m/z* under the following parameters: spray voltage of +3.6 kV (for the positive mode) and -3.2 kV (for the negative mode), capillary temperature of 320 °C, heater temperature of 300 °C, automatic gain control (AGC) of 1.0e6, maximum injection time of 100 ms, sheath gas flow rate at 30 units, auxiliary gas flow rate at 10 units, sweep gas flow rate at 11 units, S-lens RF level set at 50 and HCD normalized collision energy (NCE) of 35.0 eV. N_2 was used as the drying, nebulizer and fragmentation gas.

Randomization of all plant extracts was performed before the chromatographic analyses, and each of the three analytical replicates (pool1, pool2 and pool3) were injected five times along the chromatographic sequence to check for reproducibility in the chromatographic analyses, thus serving as quality control samples.

Data preprocessing

UHPLC-UV-HRMS datasets obtained in .raw format for each plant sample were sliced into two subsets according to the ionization mode (positive and negative) and transformed to nonproprietary .mzXML format using the MSConvert package from the software ProteoWizard 3.0.9798 (Proteowizard Software Foundation, USA). The sorted data, from 0 to 35 min, were preprocessed by MZMine 2.21 (PLUSKAL et al., 2010), where several steps were performed, including peak detection, peak filtering, chromatogram construction, chromatogram deconvolution, isotopic peak grouping, chromatogram alignment, gap filling, duplicate peaks filter, fragment search, adducts search and the search for peak identities using an *in*-house database containing the retention time and *m/z* value of more than one hundred pure compounds injected in the same conditions. Basic parameters used in MZmine for data preprocessing were: noise level at 1.0E6, Lorentzian function for the peak shape considering a resolution of 70,000, minimum peak height at 5.0E6, *m/z* tolerance at 0.002 *m/z* or 5.0 ppm and retention time tolerance of 0.7 min (Appendix 3).

The exported datasets in .csv format for each ionization mode (positive and negative), obtained after MZmine preprocessing, were edited as Excel spreadsheets (Microsoft Windows, USA), where peaks detected in the blank sample (extraction solvent) were subtracted from the peaks detected in all samples. For each ionization mode (positive and negative), a data matrix containing the species (as rows) and the peak area associated with each *m/z* value (as columns) was built. Initial exploratory analyses demonstrated no advantages of analyzing datasets from each ionization mode independently compared to a concatenated dataset including data from both modes. Therefore, this concatenated matrix was used for further multivariate analyses. Additionally, considering that individual datasets possess the limitation that different chemical classes preferably ionize in a given mode (e.g., most STLs ionize in the positive mode, while phenolic acids ionize in the negative mode), the concatenated matrix results were more comprehensive in terms of chemical diversity and thus favored its use for chemotaxonomic purposes.

Multivariate analyses and correlation with taxonomy

Supervised and unsupervised multivariate statistical analyses were performed in R 3.0.3 (R Foundation for Statistical Computing, Austria), Weka 3.6 (University of Waikato, New Zealand), and SIMCA 13.0.3.0 (Umetrics, Sweden) after submitting the concatenated matrix to Log transformation of metabolite signal intensities. Initially a Hierarchical Clustering Analysis with bootstrap resampling (HCAbp) was performed in the R package *pvclust* using Ward's method and Euclidean distance, with the aim of recognizing the natural clustering tendency of the samples based on their metabolic fingerprint similarity and ultimately to establish overall chemotaxonomic relationships. To assess the reliability of the clusters in HCAbp, 10,000 multiscale bootstrap replicates were performed considering only approximately unbiased *p*-values (AU), as they provide more accurate results than traditional bootstrapping by correcting the bias caused by a constant sample size (SUZUKI; SHIMODAIRA, 2006).

Supervised methods by Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) and decision trees were performed with the aim of maximizing chemotaxonomic segregations and determining whether the taxonomic genera currently recognized accumulate distinctive metabolites allowing their chemical differentiation. Considering that the genus *Espeletia* showed a clear clustering tendency by country of origin (Colombia or Venezuela), with two clear groups based

on the species' metabolic fingerprints, we considered them independent groups in the supervised methods (here named as *Espeletia*-COL and *Espeletia*-VEN). Three different criteria were adopted to identify the discriminating variables in the OPLS-DA model: (1) a significant contribution in the loading or coefficient plot, (2) a high magnitude and reliability jack-knifed confidence intervals in the loading or coefficient plots and (3) a variable importance in the projection (VIP) value greater than 1.0. The performance of the OPLS-DA model was assessed by two different approaches. In the first approach, we built an additional OPLS-DA model by randomly splitting the data into two subgroups. Seventy percent of the observations (80 samples) were used for model building (training set) and the remaining 30% (34 samples) for external validation (test set). The same procedure was repeated for a second time with different samples included in the training and test sets. In both cases, 97% of the observations from the test set were correctly classified, and similar R^2 and Q^2 values were obtained relative to the initial model including all samples ($R^2 = 0.87$, $Q^2 = 0.59$). The second approach consisted of performing ten raw permutations as described by Triba et al. (2015) and checking the stability of the Q^2 value in the resulting ten OPLS-DA models (TRIBA et al., 2015). In all cases, similar results were obtained, validating the performance of the initial model.

Decision trees were built using the C4.5 algorithm (J48 in Weka 3.6) and the following parameters: Classifier = Trees – J48, BinarySplits = False, ConfidenceFactor = 0.25, Debug = False, MinNumObj = 7, NumFolds = 3, ReducedErrorPruning = False, SaveInstanceData = False, SubtreeRaising = True, Unpruned = False, and UseLaplace = False. As a strategy to avoid over fitting, trees were pruned by setting the MinNumObj value to 7, which prevents the software from continuing splitting if the nodes become very small (WITTEN; EIBE; HALL, 2005). To validate the model, 66% of the observations were used as training set, while the remaining 33% were considered to be the external validation group. As a complementary validation strategy, repeated training and testing were performed by changing the random seed value from 1 to 10. In all cases, a similar percentage of correctly classified instances (78.95%) from the external validation group was obtained.

Identification of discriminant metabolites

Discriminant variables pointed by OPLS-DA and decision trees were identified considering the guidelines of the Metabolomics Standards Initiative (MSI), which currently recognizes four levels of accuracy in metabolite identifications (SUMNER et al., 2007). Level 1 corresponds to the most accurate level of identification by comparing chromatographic retention times, high-resolution MS values, UV and MS^2 spectra with reference compounds analyzed in identical experimental conditions. Level 2 represents the metabolites characterized by high-resolution MS values, UV and MS^2 spectral comparisons with information from available and commercial databases, such as AsterDB (www.asterbiochem.org/asterdb), the Dictionary of Natural Products (DNP, http://dnp.chemnetbase.com) and SciFinder Scholar. Level 3 represents the less accurate level of identification, made only to the rank of chemical class (based on characteristic UV spectra or MS^2 fragmentation patterns), in which the obtained data were not sufficiently conclusive to assign a punctual identity. Lastly, Level 4 represents the unknown metabolites. Detailed information about all dereplicated metabolites is reported in Appendix 5.

To improve the level of accuracy in the identification of the main chromatographic peaks for which no standard compound was available, the isolation and structural elucidation of seven common metabolites was accomplished by submitting a sample of *Espeletia barclayana* leaves to classical phytochemical procedures. Briefly, aerial parts of *E. barclayana* (25 g) were ground and extracted with solvents of increasing polarity (n-hexane, dichloromethane, ethyl acetate and ethanol 70%) in three consecutive steps with each solvent for 24 h. The ethanol extract (4.3 g) was submitted to Sephadex LH-20 column chromatography (60 cm × 5 cm i.d.) employing mixtures of water and methanol in different proportions (100:0 to 0:100) to produce 24 fractions (Fr1-Fr24). Fractions 3 to 24 were further purified by semipreparative HPLC (Shimadzu) using a C18 column (Shim-Pack, 5 mm, 25 × 250 mm) and a mixture of MeCN (B) and acidified water (0.1%) (A) as eluent (0-3 min: 9% B, 3-30 min: 9-40% B, 30-40 min: 40-100% B, 40-45 min: 100% B, 45-46 min: 100-9% B, 46-50 min: 9% B) at a flow rate of 10 mL/min. Eleven partially pure subfractions were collected and submitted to additional purification steps by HPLC to yield 5-*O*-(*E*)-caffeoylquinic acid (4.8 mg) (1) (CHOI et al., 2006), 1,3-di-*O*-(*E*)-caffeoylquinic acid (15 mg) (2) (DANINO et al., 2009), 1,5-di-*O*-(*E*)-caffeoylquinic acid

(27.8 mg) (3) (CARNAT et al., 2000), 3,4-di-*O*-(*E*)-caffeoylquinic acid (3 mg) (4) (KIM et al., 2011), quercetin (15.4 mg) (5) (WANG et al., 2010), 3-*O*-methylquercetin (127.7 mg) (6) (WANG et al., 2010), and 3-*O*-methylquercetin-7-*O*-*β*-glucopyranoside (3.8 mg) (7) (KRENN et al., 2003). The structures of the isolated metabolites were identified by 1H NMR experiments (Bruker ARX 500, ARX 400 and ARX 300 spectrometers, CD$_3$OD) and by high resolution MS/MS mass spectrometry (micrOTOF-Q II, Bruker, and Exactive™ Plus Orbitrap, Thermo Scientific). NMR and MS spectra are available in Appendix 6.

Reconstruction of ancestral character states

Ancestral character state reconstructions were performed using a comprehensive dataset of 33 chemical traits (grouped by biosynthetic relatedness) belonging to the main chemical classes commonly found in the subtribe Espeletiinae (free organic acids, *trans*-cinnamic acid derivatives, flavonoids, sesquiterpene lactones (STLs) and diterpenes). Compounds identified by spectral comparisons with authentic standards or data from the literature were scored as present (1) or absent (0) in each taxon by manual inspection of the UHPLC-UV-HRMS data. Subsequently, ancestral reconstructions by maximum likelihood were performed in Mesquite v3.5 (MADDISON; MADDISON, 2008) using a concatenated molecular phylogeny of ITS, ETS and rpl16, including 142 taxa of Espeletiinae reported by Diazgranados and Barber (2017). Reconstructions were based on the Markov k-state 1 parameter model, assuming equal rates of changes for both gains and losses.

Although the phylogeny reported by Pouchon et al. (2018) based on genomic data displays a better resolution and relationships among taxa than the phylogeny based on ITS, ETS and rpl16 markers used here, the poor sampling of the genomic-based phylogeny, especially among Colombian taxa of *Espeletia*, hindered its use in the reconstruction of ancestral chemical traits. Therefore, reconstructions were based on the Diazgranados' phylogeny and were made only for the well-supported major clades to avoid misinterpretation of ancestral traits. Due to the lack of chemical information for some of the species included in the molecular phylogeny, especially for the genera *Carramboa*, *Ruilopezia* and *Libanothmanus*, we restricted tree terminals to the taxa with available molecular and chemical data to perform the reconstructions, including representatives from major lineages and focusing on the genera *Espeletia* and *Espeletiopsis*. The same samples used to reconstruct the

phylogeny of the group by Diazgranados and Barber (2017) were also used in this study to investigate their chemical composition.

DNA extraction and amplification of GAO

Due to the random distribution of STLs in Espeletiinae pointed out by the UHPLC-UV-HRMS analyses and the ambiguous reconstruction of ancestral states for this chemical class, we aimed to determine whether the genetic information necessary for the biosynthesis of this important class of secondary metabolites is conserved in the subtribe or whether it is restricted to the STLs-producing clades/taxa. To that end, we investigated the presence of the key gene involved in the biosynthesis of this chemical class (Germacrene A oxidase, GAO) in genomic DNA (gDNA) extracted from leaves of 29 species, including members from five different genera and individuals producing and lacking STLs (Table 3.2). gDNA was extracted following the protocol of Ristaino et al. (2001) with some modifications. Instead of using ethanol for DNA precipitation, we used isopropyl alcohol, as this solvent provided a higher yield of DNA. Additionally, given the high amounts of phenolic compounds present in the leaves and indument of Espeletiinae, we added 40 mg of insoluble polyvinylpyrrolidone (PVP) after the nucleus lysis buffer to adsorb polyphenols.

Considering that there is no available molecular information about the genes involved in the biosynthesis of secondary metabolites in Espeletiinae, the amplification of GAO was performed by designing degenerate primers on conserved regions from GAO sequences of other Asteraceae species to retrieve a partial sequence of 532 bp located on an exon region. Briefly, GAO sequences were downloaded from GenBank (www.ncbi.nlm.nih.gov/genbank), aligned in BioEdit (HALL, 1999) and inspected for conserved regions. Degenerate primers (Table 3.3) were designed on highly conserved regions using the software Lasergene (Primer select) and then tested on genomic DNA. PCRs were performed in a peq-STAR Thermocycler (Peqlab, Germany) using RNase/DNase-free water and RedTaq MasterMix (Genaxxon Bioscience, Germany). The PCR conditions were 4 min of initial denaturation, followed by 38 cycles of 15 s denaturation (96 °C), 15 s annealing (54 °C, determined by temperature gradient), 1 min x kb-1 elongation (72 °C) and final elongation for 4 min (72 °C). Amplification products were analyzed on 1.5% agarose gels at 160 V for 30 min, including 10 µL of HD green as staining agent.

Documentation was performed under UV light at 302 nm. All PCR products were sequenced by Macrogen (Netherlands), and their identities were confirmed by BLAST searches.

Table 3.2 Species of Espeletiinae selected to amplify GAO and percentage of sequence similarity relative to *Helianthus annuus*.

Code	Species	Locality	STLs[a]	Seq. similarity[b]
PS10527	*Libanothamnus sp7*	Venezuela	Yes	87%
MD4198	*L. neriifolius*	Venezuela	No	87%
PS10545	*Ruilopezia sp6*	Venezuela	No	88%
MD4190	*T. chardonii*	Venezuela	No	88%
MD3942	*Espeletia schultzii*	Venezuela	No	88%
MD3932	*E. cuniculorum*	Venezuela	No	88%
MD4071	*E. semiglobulata*	Venezuela	No	87%
MD4201	*E. praesidentis*	Santander	No	87%
MD3872	*E. dugandii*	Santander	No	88%
MD3869	*E. estanislana*	Santander	No	88%
MD4260	*E. curialensis*	Boyacá	No	87%
MD3663	*E. brachyaxiantha*	Boyacá	Yes	88%
MD3582	*E. boyacensis*	Boyacá	No	87%
MD3640	*E. congestiflora*	Boyacá	No	88%
MD3817	*E. episcopalis*	Boyacá	No	88%
MD3834	*E. frontinoensis*	Antioquia	Yes	88%
GFPG	*E. occidentalis*	Antioquia	Yes	88%
GFPG009	*E. murilloi*	Cundinamarca	Yes	88%
MD3735	*E. summapacis*	Cundinamarca	No	88%
MD3617	*E. uribei*	Cundinamarca	No	88%
MD4197	*Espeletiopsis insignis*	Santander	Yes	87%
D. granados	*Es. santanderensis*	Santander	No	87%
D. granados	*Es. sanchezii*	Santander	Yes	87%
D. granados	*Es. funkii*	Santander	No	87%
MD4285	*Es. petiolata*	San - Boy	No	87%
MD4214	*Es. colombiana*	Boyacá	No	87%
GFPG007	*Es. pleiochasia*	Boyacá	Yes	84%
MD4250	*Es. garciae*	Boy - Cun	No	87%
GFPG015	*Es. corymbosa*	Cundinamarca	No	87%

[a] Presence of sesquiterpene lactones determined by LC-MS
[b] relative to *H. annuus* GAO.

Table 3.3 Primer pair used to amply GAO from Espeletiinae.

Code	Primer (5' - 3')	Size (bp)	AT (ºC)
Esp-GAO	F: CARTCGAATGGGCGATTTC	532	54
	R: CTTTGAACCGTGGCTCCAA		

AT: annealing temperature

RESULTS

Metabolic fingerprinting and correlation with taxonomy

Metabolic fingerprints of 114 individual taxa of Espeletiinae (representing 211 samples including replicates) were obtained by UHPLC-UV-HRMS analyses of crude methanol (70%) extracts. A HCAbp (Figure 3.1) of a concatenated matrix including data from the positive and negative ionization modes allowed establishment of chemotaxonomic relationships by recognizing the natural clustering tendency of the samples based on their metabolic fingerprints' similarity. The results from HCAbp showed a clear clustering tendency by geographic origin and taxonomic genera, with two large clusters representing the species' country of origin (Colombia or Venezuela, Figure 3.1). Within the Colombian cluster, Colombian representatives of the genus *Espeletia* were grouped with good support values with samples belonging to the genus *Espeletiopsis*, which corresponds to a closely related genus mostly restricted to Colombia. Among the Venezuelan cluster, a large subcluster grouped most of the Venezuelan samples of *Espeletia* and all the species belonging to the Venezuelan genus *Ruilopezia* with a support value of 68%. Therefore, based on overall chemical similarity, the genus *Espeletiopsis* and the Colombian representatives of *Espeletia* appeared to be closely related, with similar trends being seen among the genus *Ruilopezia* and the Venezuelan species of *Espeletia*. A clear subcluster within the Venezuelan cluster grouped most of the species belonging to the genus *Libanothamnus* (all of them from Venezuelan origin) with a good support value of 99% (Figure 3.1). In a few cases, however, the metabolic clustering did not reflect the geographic or taxonomic proximity of the studied taxa. For example, five species of the genus *Espeletia* growing in Venezuela were grouped within the Colombian representatives of the same genus, and the *Libanothamnus* cluster also grouped a Venezuelan species of *Espeletia*. Interestingly, the segregation of the genus *Espeletia* in two large clusters in the HCAbp (Figure 3.1) based on overall chemical similarity suggests that geography is strongly influencing the metabolic fingerprints of this genus, showing two large groups of species with very different metabolic compositions. As observed in the HCAbp (Figure 3.1), all quality control samples were grouped together in an independent cluster with a support value of 100%, confirming the reproducibility of the extraction and chromatographic method.
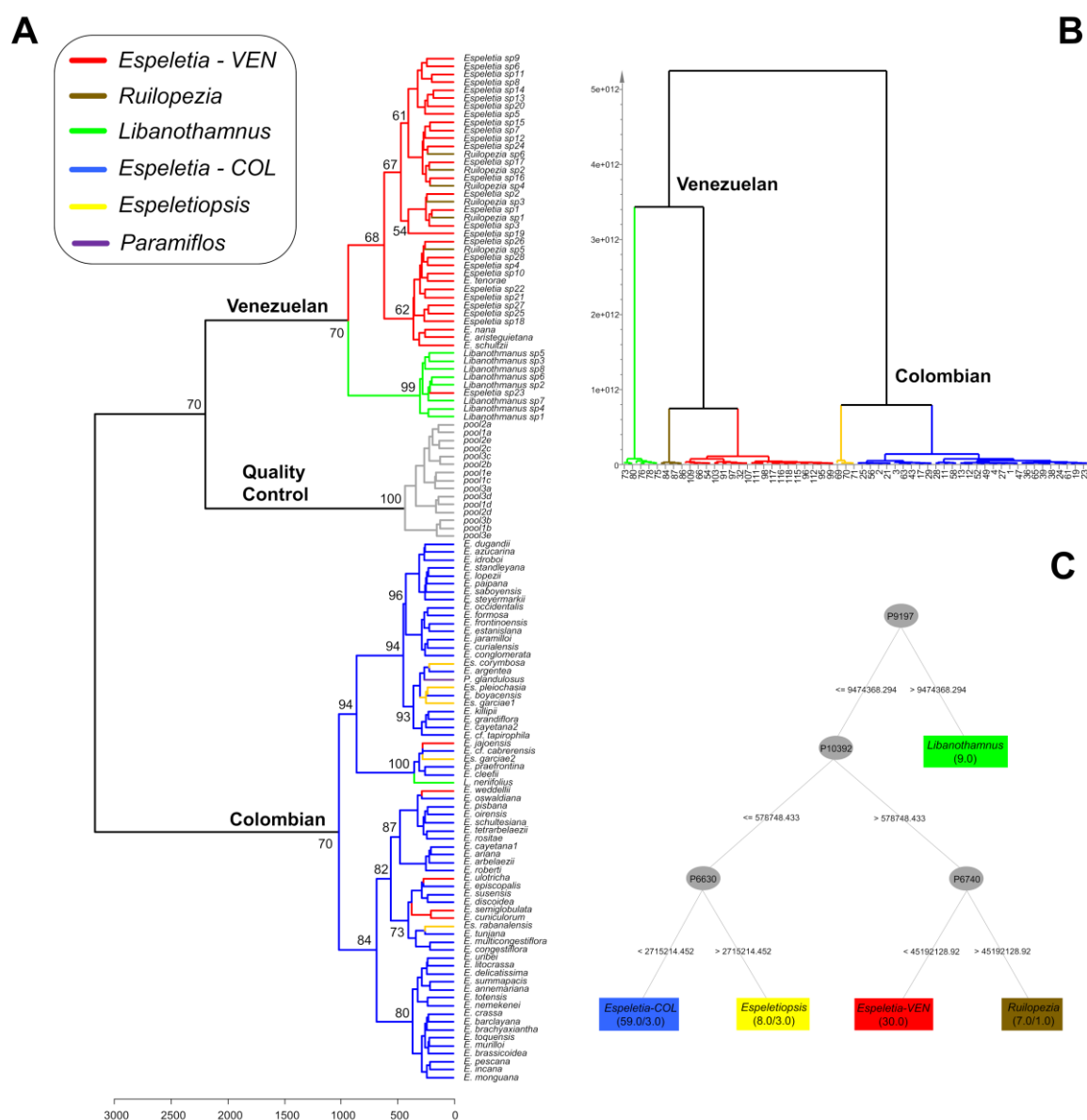
Figure 3.1 Clustering of 114 individual taxa of Espeletiinae based on metabolic fingerprints' similarity including all detected mass features in the positive and negative ionization modes. (A) Hierarchical clustering analysis with bootstrap values above branch. (B) OPLS-DA dendrogram and (C) decision tree showing the metabolic segregation of Espeletiinae taxa by taxonomy. Taxonomic genera highlighted in different colors.

Supervised methods by OPLS-DA and decision trees (Figures 3.1B and 3.1C, respectively) were performed with the aim of maximizing chemotaxonomic segregations and determining whether the taxonomic genera currently recognized produce distinctive metabolites allowing a chemical segregation previously

undetected by the unsupervised methods based on overall chemical similarity (HCAbp, Figure 3.1A). Considering the clear geographic structure seen in the metabolic clustering of the genus *Espeletia* (Figure 3.1), supervised analyses were performed with species of *Espeletia* located in Colombia and Venezuela as independent groups in the Y variable (*Espeletia*–COL and *Espeletia*-VEN), along with the other analyzed genera of Espeletiinae (*Libanothamnus*, *Ruilopezia* and *Espeletiopsis*). The results from OPLS-DA (R^2=0.87, Q^2=0.59) and decision trees (78.95% correctly classified instances) showed clear differences in the metabolic compositions of the different genera of Espeletiinae and a strong geographic structure by country of origin. In both cases, the genus *Libanothamnus* appeared to be the more chemically distinct group compared to the other genera. However, it was still grouped within the Venezuelan cluster, suggesting a closer chemical similarity with species from the genera *Ruilopezia* and *Espeletia*-VEN than with the Colombian taxa of *Espeletiopsis* and *Espeletia*-COL (Figure 3.1B and 3.1C). Therefore, even though based on overall chemical similarity there were two main groups of samples (Venezuelan and Colombian, Figure 3.1A) with a rather low chemical differentiation among the different genera composing each group (except for *Libanothamnus*). There are still important differences in the chemical composition of these genera, allowing their differentiation based on the common presence of certain characteristic metabolites.

Accumulation patterns of SM by taxonomy

Further analysis of the OPLS-DA model and concomitant dereplication allowed the identification of different variables as discriminating metabolites of each genus (Table 3.4). A heatmap based on the OPLS-DA top-ranked mass features (Figure 3.2) showed a differential accumulation of these metabolites by species from each genus. Several flavonoids, caffeic acids esters of quinic and altraric acids, and some diterpenes were identified as the main chemical classes of discriminant metabolites found in Espeletiinae (Table 3.4). Species belonging to the Colombian representatives of the genus *Espeletia* accumulated high amounts of 3-methoxy quercetin, 2,3,5 or 2,4,5-tricaffeoylaltraric acid and a putative biflavonoid with a molecular formula of $C_{31}H_{20}O_{14}$, while the Venezuelan *Espeletia* accumulated mostly 2-formyl tetrahydropyran-3-yl acetate, quinic acid and quercetin-3-*O*-galactoside (Table 3.4). Species from the genus *Espeletiopsis* showed high amounts of an

unknown metabolite with a molecular formula of $C_{26}H_{22}O_{10}$, in addition to 3,4-dihydroxyphenyl caffeate and the flavonoid dihydroxy-methoxyflavanone (Table 3.4). Several diterpenes and some flavonoids, such as 16-hydroxy kauran-19-al, kaurenoic acid, santin and dimethoxyquercetin, were identified as discriminant metabolites for the genus *Libanothamnus*, while the genus *Ruilopezia* accumulated high amounts of 3,5-*O*-dicaffeoylquinic acid and quercetin-3-di-*O*-glycoside (Table 3.4). Lastly, some metabolites, such as quinic acid, 1,5-*O*-dicaffeoylquinic acid and 3',5-dihydroxy-3,4',7-trimethoxyflavone (Var IDs 482, 1283 and 6005, respectively, Figure 3.2), were equally accumulated in high quantities by the different Venezuelan genera.

Figure 3.2 Heatmap showing the differential accumulation of the top ranked OPLS-DA mass features in each genus (for compound identities see Table 3.4).

Table 3.4 Discriminant metabolites of each genus based on the OPLS-DA model.

Var ID	Ionization mode	VIP	Rt	*m/z*	Substance or molecular formula	MSI level[+]
Espeletia - COL						
P6396, N1616	Both	28.94	11.0	317.067	3-methoxy quercetin	1
N9	Neg	16.01	9.8	695.126	2,3,5 or 2,4,5-tricaffeoylaltraric acid	1
N3	Neg	13.67	11.0	631.109	dimer of 3-methoxy quercetin	1
P6004, N16	Both	13.44	15.1	617.093	biflavonoid ($C_{31}H_{20}O_{14}$)	3
N730	Neg	13.69	15.0	299.020	fragment of biflavonoid ($C_{31}H_{20}O_{14}$)	3
N4	Neg	12.26	10.7	301.036	quercetin	1
N386	Neg	12.58	2.8	153.018	protocatechuic acid	1
P6009	Pos	11.07	19.0	337.105	Na^+ adduct of $C_{18}H_{18}O_5$	4
P6509	Pos	9.49	9.6	163.039	fragment of 2,3,5 or 2,4,5-tricaffeoylaltraric acid	1
N2217	Neg	9.65	13.5	301.072	isoetin	2
N698	Neg	9.50	11.2	299.020	fragment of 3-methoxy quercetin	2
N484	Neg	8.51	1.0	209.030	altraric acid	2
Espeletia - VEN						
N824	Neg	10.16	7.6	171.066	2-formyl tetrahydropyran-3-yl acetate	2
N482	Neg	8.67	0.9	191.055	quinic acid	1
N3270, P8861	Both	7.44	7.5	463.089	quercetin-3-*O*-galactoside	1
N1283	Neg	7.21	8.3	515.120	3,5-*O*-dicaffeoylquinic acid	1
P6005	Pos	7.09	16.7	345.097	santin	1
N26, P6030	Both	7.08	9.9	609.126	quercetin 3-*O*-methylpentoside-hexoside	3
N1060	Neg	6.08	6.0	515.120	1,3-*O*-dicaffeoylquinic acid	1
Espeletiopsis						
P6108, N120	Both	19.57	15.21	495.129	$C_{26}H_{22}O_{10}$	4
N3600, P9787	Both	18.52	9.44	287.056	3,4-dihydroxyphenyl caffeate	2
N62, P6076	Both	18.52	7.76	177.019	$C_6H_9O_4$	4
N2272	Neg	10.59	8.08	289.072	$C_{15}H_{14}O_6$	4
P6537	Pos	10.24	27.98	301.216	grandiflorenic acid	2
P6014	Pos	9.64	15.34	287.091	dihydroxy-methoxyflavanone	2
P8732	Pos	8.97	12.42	303.086	hesrepetin	1
N482	Neg	8.67	0.97	191.055	quinic acid	1
Libanothamnus						
P6401	Pos	22.93	27.94	305.247	16-hydroxy kauran-19-al	2
P6005	Pos	18.06	16.74	345.097	santin	1
N10, P6007	Both	15.36	14.51	329.067	dimethoxyquercetin	2
P6406	Pos	12.92	28.04	287.237	fragment of 16-hydroxy kauran-19-al	2
P6232	Pos	12.59	20.37	319.227	hydroxy-*ent*-kauren-18-oic acid isomer 3	3
N1283	Neg	12.23	8.31	515.120	3,5-*O*-dicaffeoylquinic acid	1

P6116	Pos	10.70	23.66	303.232	putative diterpene ($C_{20}H_{30}O_2$)	3
P6683	Pos	10.36	18.92	319.227	hydroxy-*ent*-kauren-18-oic acid isomer 4	3
P6257	Pos	10.07	21.15	319.227	hydroxy-*ent*-kauren-18-oic acid isomer 2	3
P6027	Pos	9.77	15.56	317.102	$C_{17}H_{16}O_6$	4
P6407	Pos	9.56	12.76	359.113	pentahydroxy-tetramethoxyflavone	2
N623	Neg	8.65	21.91	317.212	hydroxy-*ent*-kauren-18-oic acid isomer 1	3
Ruilopezia						
N1283	Neg	13.69	8.31	515.120	3,5-*O*-dicaffeoylquinic acid	1
N28	Neg	10.77	6.24	625.142	quercetin-3-di-*O*-glycoside	3
P6839	Pos	9.42	19.22	233.154	$C_{15}H_{22}O_2$	4
P6005	Pos	8.81	16.74	345.097	santin	1
P6065	Pos	7.89	17.34	231.138	$C_{15}H_{20}O_2$	4
N824	Neg	7.77	7.64	171.066	2-formyl tetrahydropyran-3-yl acetate	2

* Based on the metabolomics standards initiative (Sumner et al., 2007).

Evolution of chemical traits

The reconstruction of ancestral states, displayed in Figures 3.3 to 3.7, indicated that the most recent common ancestor of Espeletiinae had a very complex secondary metabolite chemistry characterized by the presence of several isomers of caffeoylquinic and caffeoylaltraric acids, as well as some free and glycosylated flavonoids (especially quercetin, 3-methoxy quercetin and their glycosil derivatives) and diterpenes of the *ent*-kaurane type (not shown). This analysis showed that whereas most chemical traits were already present or evolved with high likelihood (>50%) in the common ancestor of the subtribe, some compounds seemed to have appeared several times in the common ancestors of more restricted clades or in terminal taxa, showing complex scenarios of chemical change in Espeletiinae with frequent losses and gains of various traits (Figures 3.3 to 3.7).

Among caffeoylquinic acids, our results indicate the occurrence of three putative chemical synapomorphies (1,3-di-*O*-caffeoylquinic acid, 1,5-di-*O*-caffeoylquinic acid and 3,4,5-tri-*O*-caffeoylquinic acid, Figure 3.3) for the whole subtribe Espeletiinae with a few losses in some terminal taxa. None of the caffeoylquinic acids investigated was restricted to smaller clades. Instead, these compounds were found to be very conserved or completely absent in the subtribe, such the isomers 3- and 5-*O*-caffeoylquinic acids (not included in Figure 3.3) present in the external group and evolutionarily conserved in all the investigated species, and the isomer 4-*O*-caffeoylquinic acid, absent in all Espeletiinae but present in the external group (Figure 3.3). Similar patterns were obtained in the reconstructions of

caffeoylaltraric acids (Figure 3.4). Three of the investigated tri-*O*-caffeoylaltraric acid isomers (isomers 2, 3 and 4) appeared to be plesiomorphic traits conserved in the subtribe with putative secondary loss in *Ruilopezia cardonae* and *Libanothamnus neriifolius* (Figure 3.4), while ambiguous reconstructions were obtained for isomer 1 (Figure 3.4). This last isomer was likely present (likelihood >65%) in the common ancestor of only two clades composed of Colombian *Espeletia* (Figure 3.4), although it was also observed in numerous terminal taxa of distantly related species.

Reconstructions of flavones and flavonols (Figure 3.5) showed that whereas both quercetin and 3-methoxy quercetin (not included in Figure 3.5) represent plesiomorphic traits present in the external group and evolutionarily conserved in all Espeletiinae, other flavonols appeared to have been gained and lost multiple times (Figure 3.5). A putative biflavonoid with a molecular formula of $C_{31}H_{20}O_{14}$ represents a characteristic chemical synapomorphie (likelihood of 100%) for the subtribe that appeared once in the common ancestor of Espeletiinae, with secondary loss only in *Ruilopezia cardonae*, while a putative biflavonoid-protocatechuic acid appears to have been gained and lost multiple times (Figure 3.5). This last compound was present with a high likelihood in the common ancestor of most of the clades of Colombian *Espeletia* (which also include a few species currently circumscribed within *Espeletiopsis*), as well as in a subclade of *Espeletia* from Venezuela (Figure 3.5). Similar ancestral states reconstructions were found for 3-*O*-methoxy quercetin-7-*O*-acetylglycoside (Figure 3.6). The flavonoid trihydroxy-pentamethoxy flavone, produced only by the Venezuelan taxa of Espeletiinae, appears to be present with high likelihood (>64%) in the common ancestor of a Venezuelan clade of *Espeletia*, which also includes *R. cardonae* and *Libanothamnus neriifolius* (Figure 3.5). Ancestral reconstructions of different quercetin glycosides pointed to quercetin-3-*O*-diglycoside as a distinctive chemical synapomorphie already present in the common ancestor of the subtribe and evolutionarily conserved in all the investigated species, while its monoglycosil derivative (quercetin-3-*O*-glycoside) represents an ancestral trait that underwent a putative secondary loss only in *Espeletia oswaldiana* (Figure 3.6). Quercetin-7-*O*-glycoside represents an additional chemical synapomorphie in Espeletiinae with secondary loss in a few terminal taxa, while ambiguous reconstructions were obtained for quercetin-3-*O*-methylpentoside-hexoside, especially in Venezuelan taxa and Colombian *Espeletiopsis* (Figure 3.6).

Lastly, ambiguous reconstructions were also found for all of the melampolide-type STLs identified in Espeletiinae: longipilin acetate, fluctuadin and polymatin B (Figure 3.7). Contrary to the other chemical classes, the accumulation of these three STLs was very consistent in the investigated taxa, showing a tendency in which either all three compounds were accumulated or none of them was produced by a certain species, suggesting a shared biosynthetic origin. However, in an evolutionary framework the accumulation of STLs showed to be extremely variable, with a random distribution in Espeletiinae, being present in at least one representative of all major clades (Figure 3.7), which suggests a putative chemical convergence for this chemical class in some terminal taxa or, a more likely scenario, that our ancestral states reconstruction of melampolide-type STLs were not conclusive enough to make inferences about the evolution of this chemical trait in the subtribe.
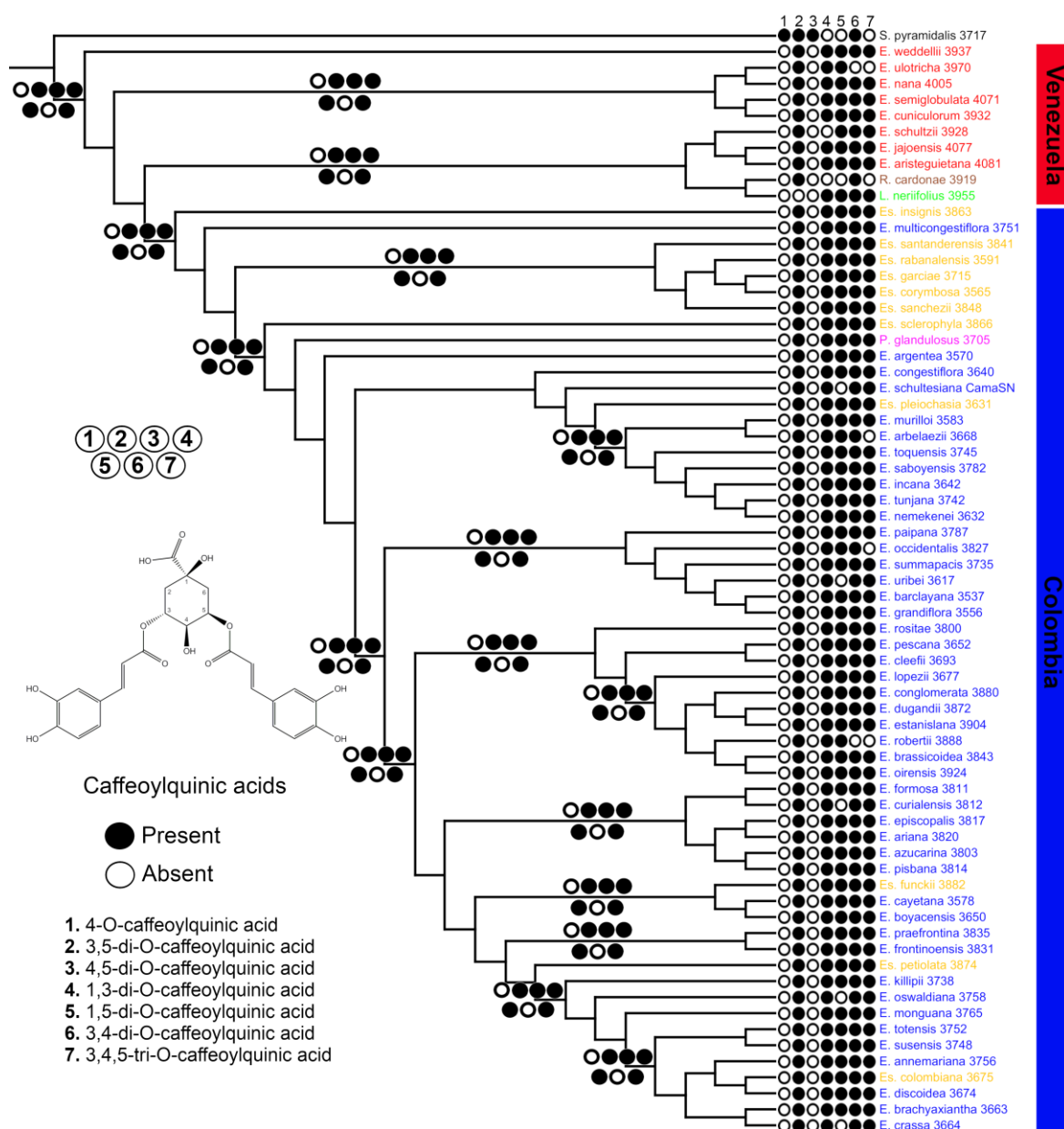
Figure 3.3 Summary of ancestral state reconstructions for caffeoylquinic acids in well-supported major clades of Espeletiinae. Tree obtained by maximum likelihood based on ITS, ETS and rpl16 molecular markers (DIAZGRANADOS; BARBER, 2017). The pie charts along the branches illustrate the likelihood for the ancestral states reconstructed in Mesquite 3.5.
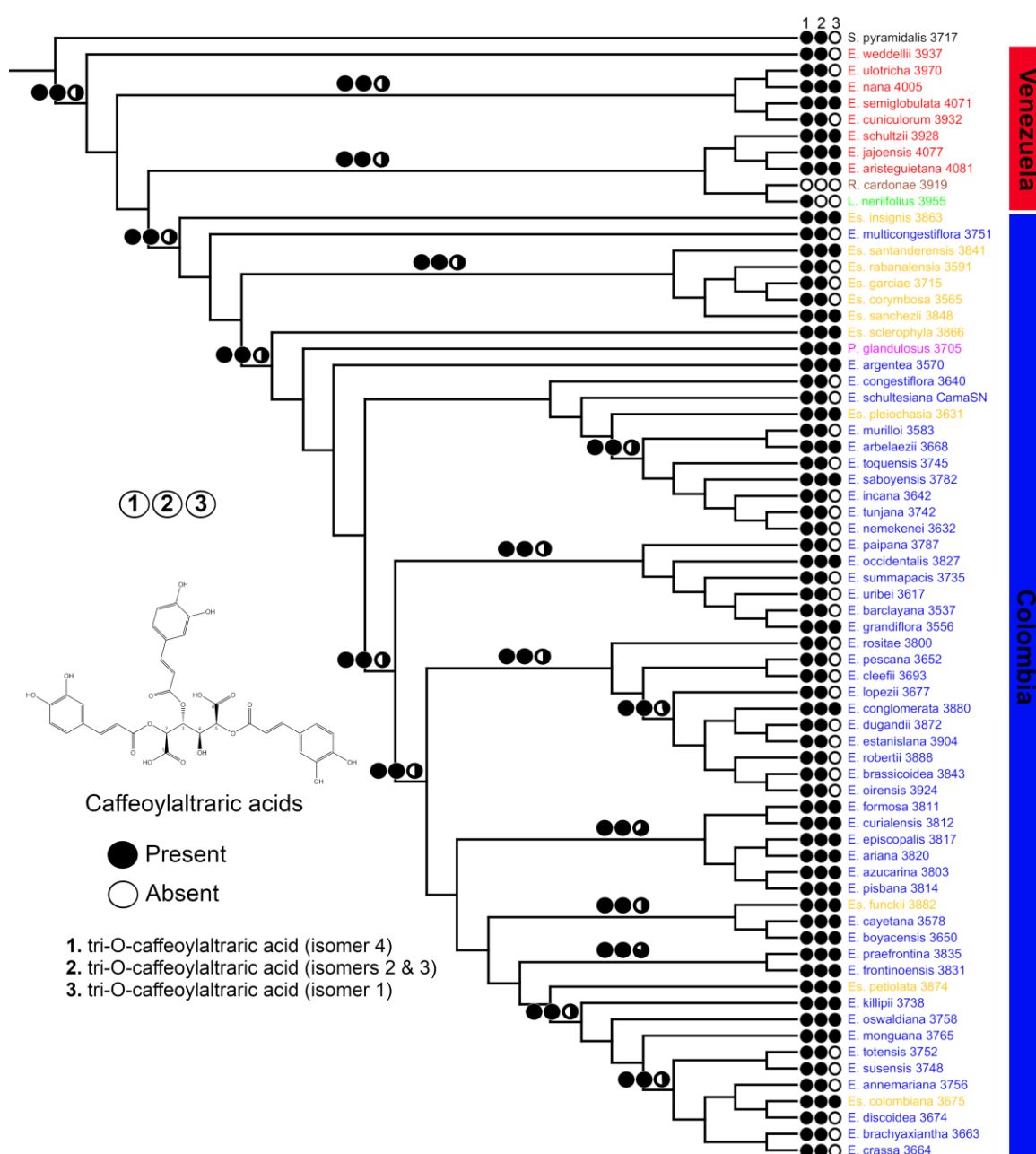
Figure 3.4 Summary of ancestral state reconstructions for caffeoylaltraric acids in well-supported major clades of Espeletiinae. Tree obtained by maximum likelihood based on ITS, ETS and rpl16 molecular markers (DIAZGRANADOS; BARBER, 2017). The pie charts along the branches illustrate the likelihood for the ancestral states reconstructed in Mesquite 3.5.
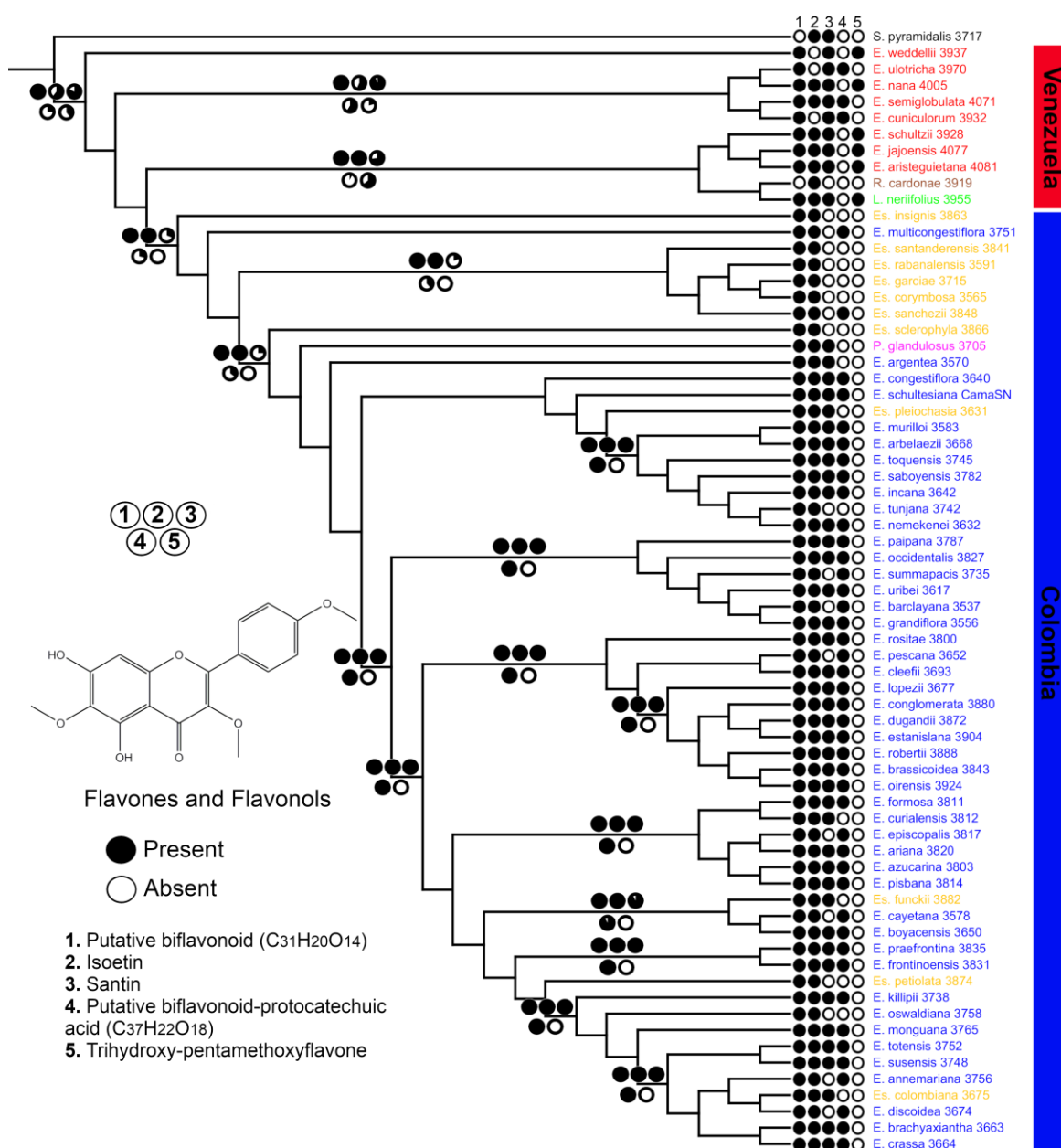
Figure 3.5 Summary of ancestral state reconstructions for flavones and flavonols in well-supported major clades of Espeletiinae. Tree obtained by maximum likelihood based on ITS, ETS and rpl16 molecular markers (DIAZGRANADOS; BARBER, 2017). The pie charts along the branches illustrate the likelihood for the ancestral states reconstructed in Mesquite 3.5.
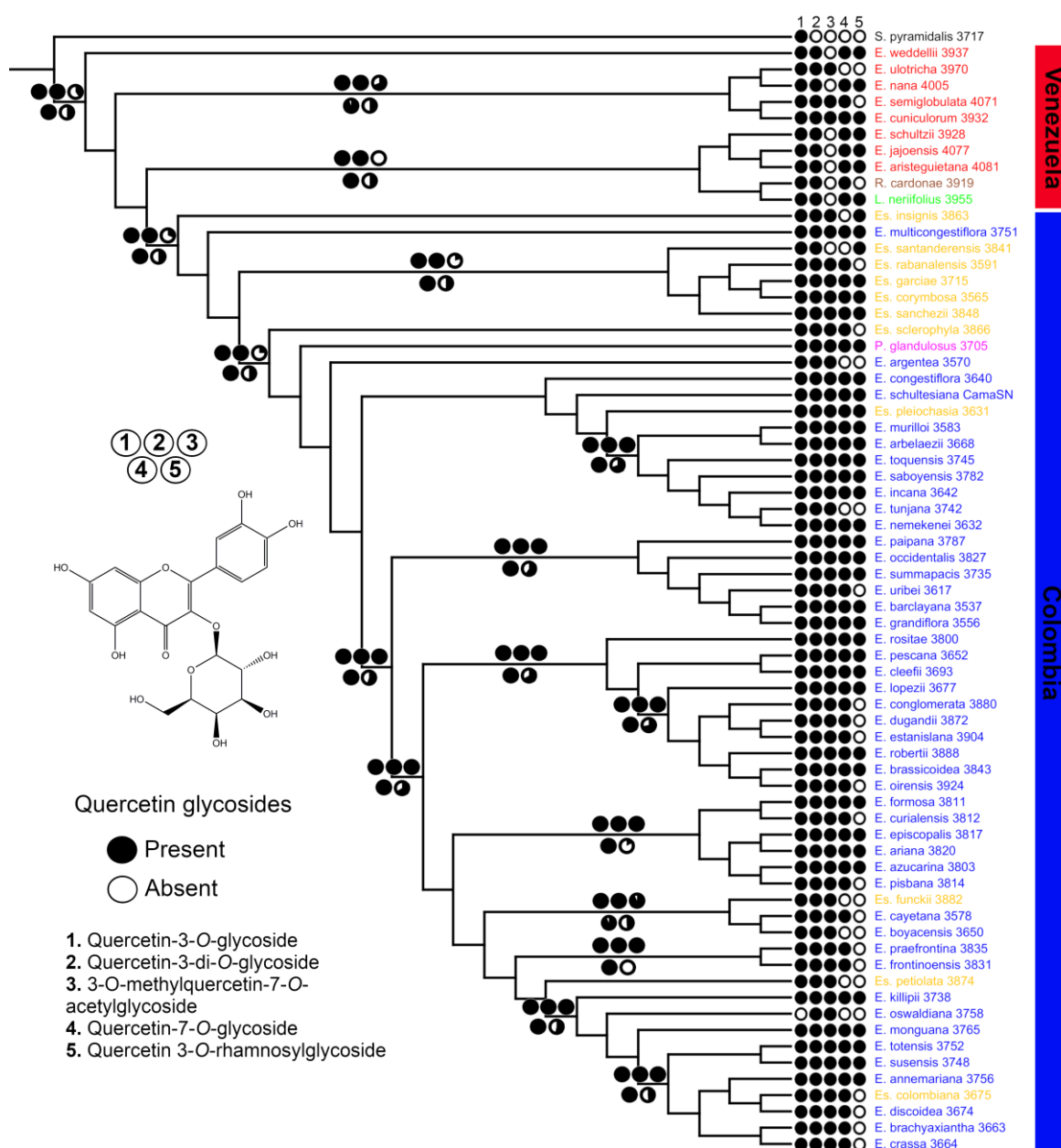
Figure 3.6 Summary of ancestral state reconstructions for quercetin glycosides in well-supported major clades of Espeletiinae. Tree obtained by maximum likelihood based on ITS, ETS and rpl16 molecular markers (DIAZGRANADOS; BARBER, 2017). The pie charts along the branches illustrate the likelihood for the ancestral states reconstructed in Mesquite 3.5.
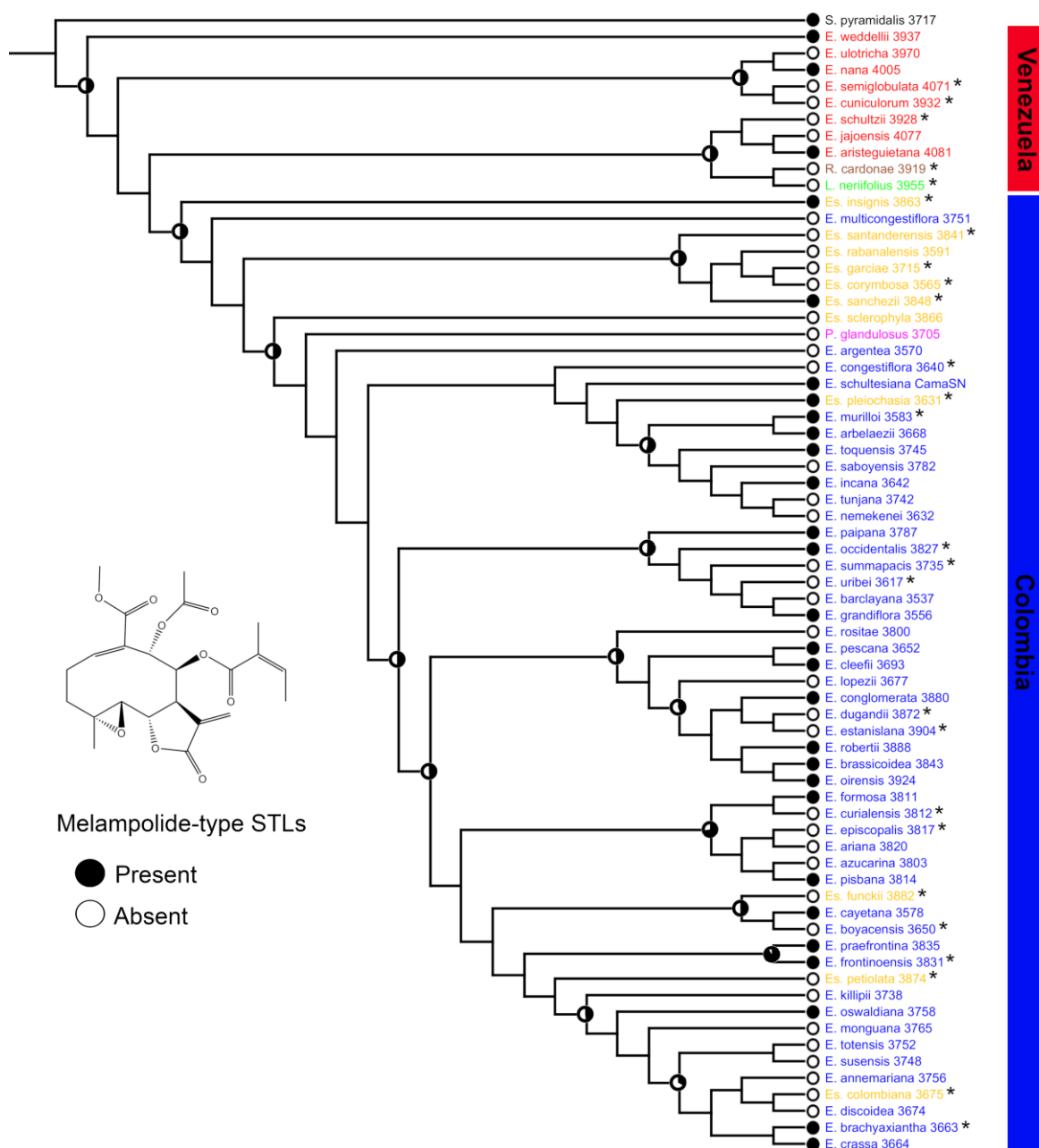
Figure 3.7 Ancestral state reconstructions for melampolide-type sesquiterpene lactones in well-supported major clades of Espeletiinae. Tree obtained by maximum likelihood based on ITS, ETS and rpl16 molecular markers (DIAZGRANADOS; BARBER, 2017). The pie charts along the branches illustrate the likelihood for the ancestral states reconstructed in Mesquite 3.5. Asterisks after species names represent taxa investigated for the presence of GAO.

Sesquiterpene lactone biosynthesis in Espeletiinae

PCR amplifications using primers for GAO (Table 3.3) in 29 species of Espeletiinae gave clear partial sequences of approximately 500 bp (Figure 3.8) representing *ca.* 30% of the GAO coding region (which contains 1467 bp in *Helianthus annuus*).
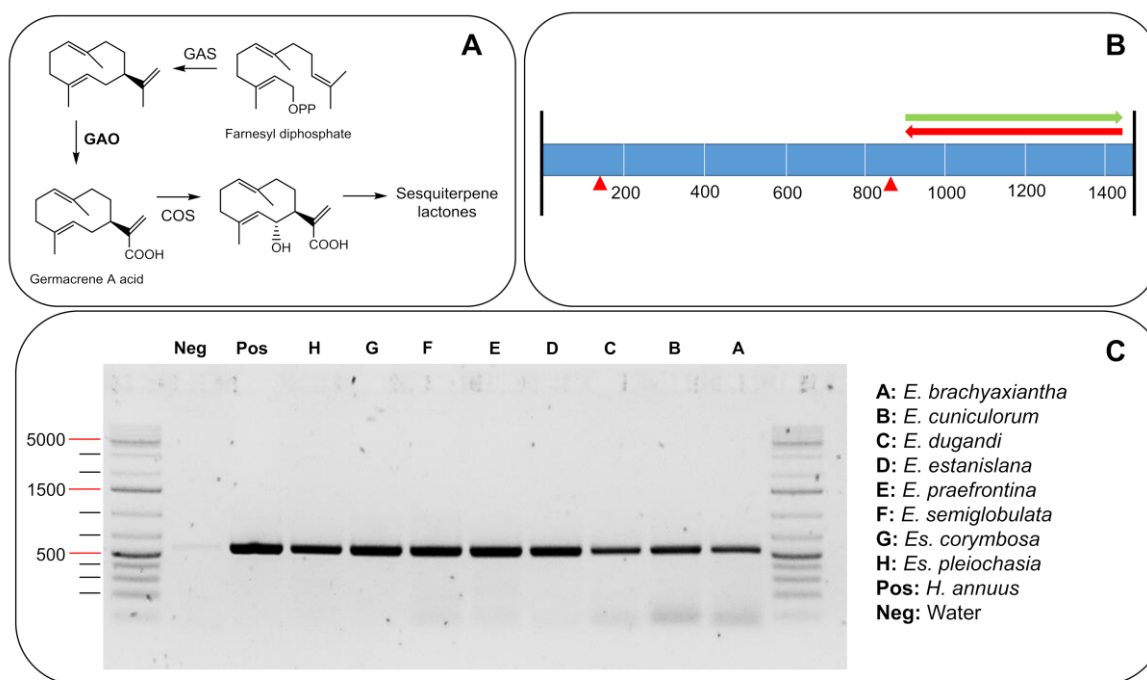
Figure 3.8 (A) Biosynthetic pathway of sesquiterpene lactones in Asteraceae, GAS: germacrene A synthase, GAO: germacrene A oxidase, COS: costunolide synthase. (B) Schematic view of the amplified region of GAO. Red triangles represent intron regions reported by (NGUYEN, 2013), while green and red arrows represent the amplified region linked to the forward and reverse primers, respectively. (C) Amplification products of partial GAO sequences from selected Espeletiinae species separated on capillary gel electrophoresis.

Interestingly, amplification was successful even in species where no STLs were detected by the UHPLC-UV-HRMS analyses, while sequence comparisons by BLAST displayed a similarity of 87 to 88% relative to reported GAO sequence of *H. annuus* (Table 3.2). These results suggest that the key gene involved in the biosynthesis of germacrene A acid, the putative precursor of all STLs in Asteraceae, is also present and is evolutionary conserved in the subtribe Espeletiinae. However, further studies are still necessary to retrieve the whole coding sequence of this gene

in the studied species and perform their functional characterization *in vitro* to unambiguously determine if such genes are functionally active (especially in the STLs-lacking taxa) and capable of catalyzing the first oxidative transformation of germacrene A to yield germacrene A acid, which represents a key step in the biosynthesis of STLs (Figure 3.8).

DISCUSSION

Chemotaxonomic relationships

As a core finding, this study allowed the establishment of chemotaxonomic relationships in Espeletiinae based on whole-metabolome comparisons, providing metabolomic evidence to support a putative segregation of the genus *Espeletia* in two different genera based on their characteristic metabolic fingerprints: one comprising Colombian taxa of *Espeletia* and the other the remaining Venezuelan representatives. Furthermore, our results indicate that such "geography-based" clustering is not only restricted to the genus *Espeletia*. Metabolomic comparisons of 211 taxa of Espeletiinae (representing 114 species from five different genera) showed a clear geographic structure with two large clusters related to the species' country of origin (Colombia or Venezuela, Figure 3.1), where species from each respective country share certain metabolic traits (Table 3.4). Within the "Colombian" cluster, Colombian representatives of the genus *Espeletia* appear to be closely related to the genus *Espeletiopsis* and the monotypic *Paramiflos*, while the "Venezuelan" cluster is composed of the genera *Libanothamnus*, *Ruilopezia* and the Venezuelan representatives of *Espeletia* (Figure 3.1).

Interestingly, our geography-based clustering based on metabolomic data largely agrees with the molecular phylogeny of Espeletiinae based on ITS, ETS, rpl16 and AFLP markers (DIAZGRANADOS; BARBER, 2017), and with the more recent and better-resolved phylogeny based on nuclear genomic evidence (POUCHON et al., 2018). In both phylogenies, two large geographically distinct clades (Colombian and Venezuelan) were recovered, which deeply challenges the current generic classification of the subtribe proposed by Cuatrecasas (1976, 2013) based on morphology (DIAZGRANADOS; BARBER, 2017; POUCHON et al., 2018). For instance, both phylogenies proposed redefining the generic limits of the subtribe by two possible approaches: a splitting approach or a lumping approach. On one hand, the splitting approach would imply: 1) preserving the monophyletic genera *Carramboa* and *Libanothamnus*; 2) including *Espeletia semiglobulata* and *E. cuniculorum* in the genus *Coespeletia*; 3) creating two new genera for the Venezuelan *Espeletia* and *Espeletiopsis*, respectively; (4) splitting Colombian *Espeletiopsis* in three genera; and 5) splitting *Ruilopezia* into at least three different genera, one of them including *Tamania* (DIAZGRANADOS; BARBER, 2017;

POUCHON et al., 2018). On the other hand, the lumping approach proposed by Diazgranados and Barber (2017) would imply "creating two groups (or three if keeping the monophyletic *Carramboa*): one with the mainly Venezuelan species, in which case should be named *Libanothamnus*; and another group with the Colombian taxa, named *Espeletia*", or adopting the view proposed by Pouchon et al. (2018) that "only the genus *Espeletia* should be considered in the subtribe, with the other seven genera considered as synonyms". Our results based on whole-metabolome comparisons tend to support the lumping approach proposed by Diazgranados and Barber (2017), as two geography-based groups of species were consistently recovered in all analyses, but differently from such an approach, the genus *Libanothamnus* was clearly segregated from the other Venezuelan taxa based on its characteristic metabolic fingerprints, suggesting an alternative approach with three groups (or four if keeping the monophyletic *Carramboa* that remained unsampled by us): one with the Colombian species of *Espeletia* and *Espeletiopsis* (including *Paramiflos*), another with the Venezuelan *Espeletia* and *Ruilopezia* and a third group, closely related to the previous group, including the monophyletic and chemically distinct *Libanothamnus*. However, as reported by Diazgranados and Barber (2017), it seems premature to address profound changes in the taxonomy of Espeletiinae without further evidence, especially considering that based on supervised statistical methods (OPLS-DA and decision trees, Figure 3.1B and 3.1C), our results showed important differences in the metabolic compositions of the different taxonomic genera currently recognized in Espeletiinae, while reinforcing the putative segregation of *Espeletia*.

Supervised algorithms, unlike unsupervised methods (e.g., HCAbp), use the class membership of the samples (in this case the taxonomic genera) to find the variables supporting such groupings. Therefore, although based on whole-metabolome comparisons we found two different groups of samples based on their characteristic metabolic fingerprints (Colombian and Venezuelan, Figure 3.1A), there are still important differences in the chemical composition of the species belonging to each taxonomic genus (except for *Espeletia*, which assemblies two clearly different groups), providing chemical evidence to support some of the taxonomic genera currently recognized, as well as chemical evidence supporting the segregation of *Espeletia*. These chemical differences rely on the preferential accumulation of the discriminant metabolites reported in Table 3.4 by species from each genus. The fact

that such metabolic differences are related to quantitative composition yields, rather than qualitative production of distinct metabolites, implies that such differences could be attributed to differential levels in gene expression among the different genera of Espeletiinae, rather than composition changes in genes at the genomic level. Such differences could be related to adaptations to unique selective pressures faced by each taxonomic group in their respective localities or as a consequence of different life strategies, as secondary metabolites are involved in several ecological interactions and in protection mechanisms (HARBORNE, 1989). Considering that a close association between shifts in gene expression and shifts in species diversification has been identified in Andean lineages of the genus *Lupinus* (NEVADO et al., 2016), it is possible that analogous to *Lupinus,* rapid gene expression divergence also drives rapid evolutionary diversification in Espeletiinae. However, further analyses based on correlations between transcriptomic and diversification rates are necessary to shed light on this issue.

We highlight the largely congruent results obtained by different sources of information, such as single molecular markers (DIAZGRANADOS; BARBER, 2017), genomic data (POUCHON et al., 2018), metabolic fingerprints (present study) and morphological characteristics (CUATRECASAS, 2013). The putative segregation of the genus Espeletia in two different geography-based genera represents such a case in which genomic, metabolomic and morphological data appear to converge. For instance, long before the first phylogenetic and metabolomic analyses, Cuatrecasas (2013) recognized the Venezuelan *Espeletia* as a morphologically distinct group of species, and grouped most of them under the section *Weddellia* Cuatrec., while most of the Colombian taxa were grouped under the section *Espeletia*. Morphologically, these two sections differ mainly in the shape of the phyllaries, which are linear, narrowly triangular, lanceolate or narrowly oblong, with a length to width ration of 4-10:1 in *Weddellia*, while the phyllaries in *Espeletia* are ovate, elliptic, triangular or broadly oblong with a length to width ration of 1-3:1 (CUATRECASAS, 2013). Chemically, we found that species belonging to the Colombian *Espeletia* accumulate high amounts of 3-methoxy quercetin, 2,3,5 or 2,4,5-tricaffeoylaltraric acid and a putative biflavonoid ($C_{31}H_{20}O_{14}$), while Venezuelan *Espeletia* accumulate mostly 2-formyl tetrahydropyran-3-yl acetate, quinic acid and quercetin-3-*O*-galactoside (Table 3.4). Therefore, although it is still premature to address profound changes in the taxonomy of Espeletiinae (e.g., the splitting approach previously mentioned), there is

substantial evidence suggesting the possible segregation of the genus *Espeletia* into two different genera by adopting the "lumping approach" previously mentioned.

Although metabolomics constitutes an interesting tool for assessing taxonomic relationships and has been previously used to aid in the classification of several plants and microorganisms (GALLON et al., 2018; KIM et al., 2016; MARTUCCI et al., 2014), care must be taken when comparing its results to phylogenetic hypotheses (MARTUCCI et al., 2018). Metabolomic analyses are usually based on the comparison of hundreds to thousands of metabolites among different taxa by multivariate analyses. Hence, these comparisons are based solely on the phenetic similarity of the secondary metabolite patterns of different taxa and thus are not indicative of species evolution, as they do not distinguish between homology (similar traits inherited from a common ancestor) and homoplasy (similar traits from independent evolutionary origins). Consequently, unless phylogenetic algorithms (e.g., maximum parsimony, maximum likelihood and Bayesian inference) are used to infer relationships or determine trait evolution based on metabolomic data (e.g., PADILLA-GONZÁLEZ et al., 2018), multivariate analyses with metabolomic data offer no information about the evolution of certain traits or taxa.

Chemical evolution

Reconstruction of ancestral states showed complex scenarios of chemical evolution where the common ancestor of Espeletiinae had an intrincate secondary metabolite chemistry with some traits representing direct descent from ancestral taxa (plesiomorphic traits, i.e., quercetin, 3-methoxy quercetin and *ent*-kaurane diterpenes), other traits resulting from evolutionary innovation (chemical synapomorphies, i.e., quercetin-3-di-*O*-glycoside and biflavonoid $C_{31}H_{20}O_{14}$), and several traits being gained and lost multiple times, implying convergent evolution. Considering that the diversification of the frailejones started approximately 2.3 My, after the final uplift of the Andes and the emergence of the páramo conditions (POUCHON et al., 2018), the higher chemical diversity found, especially among phenolics (caffeoylquinic- and caffeoylaltraric acids and flavonoids), compared to the sister genus *Smallanthus*, their conserved nature in the subtribe, as well as their accumulation in higher amounts, suggest that these compounds might play a key role in the adaptation of the frailejones to the harsh environmental conditions characterizing the páramo ecosystem.

Among the identified phenolics, three caffeoylquinic acids: 1,3- and 1,5-di-*O*-caffeoylquinic acids and 3,4,5-tri-caffeoylquinic acid; two glycosil flavonoids: quercetin-3-di-*O*-glycoside and quercetin-7-*O*-glycoside; and a putative biflavonoid with a molecular formula of $C_{31}H_{20}O_{14}$ represent distinctive chemical synapomorphies at the level of subtribe, which provide further evidence supporting the monophyly of the frailejones and suggest a possible adaptive advantage to the producing plants. Previous studies have reported that caffeic acid derivatives and flavonoids provide protection to the producing plants against the harmful effects of the UV radiation by absorbing or dissipating solar energy and neutralizing radiation-induced free radicals and other reactive oxygen species (CLÉ et al., 2008; GRACE; LOGAN, 2000; JANSEN; GABA; GREENBERG, 1998; JULKUNEN-TIITTO et al., 2015; ROZEMA et al., 1997). It is possible that the remarkable structural diversification of caffeoyl derivatives and flavonoids and their accumulation in high amounts in Espeletiinae are the consequence of a specialized protection mechanism developed to withstand the high rates of UV radiation present in the páramo ecosystem, which are considerably higher than the ones observed in lowland ecosystems (LUTEYN, 1999). The fact that flavonoids are preferentially accumulated in the characteristic epidermal hairs of Espeletiinae, which usually form a dense layer of long trichomes covering the entire leaf surface and reproductive organs, known as indument, further supports this hypothesis (PADILLA-GONZÁLEZ, 2014).

In addition to the reported phenolics, the subtribe Espeletiinae is especially rich in diterpenes of the *ent*-kaurane type, where 40 different chemical structures have been reported to date (reviewed in PADILLA-GONZÁLEZ et al., 2018). The occurrence of *ent*-kaurane diterpenes in all of the investigated species as well as in the sister genus *Smallanthus* points to this chemical class as a plesiomorphic trait that is evolutionarily conserved and especially diversified in the frailejones. In Espeletiinae, *ent*-kauranes are usually found as major components of exudates and resins produced by mechanical rupture of the leaf blades or stems (CUATRECASAS, 2013; PIOZZI et al., 1968, 1971, 1972), where they are assumed to play a role in plant-insect interactions. Considering that several *ent*-kaurane diterpenes and their semisynthetic derivatives have been widely reported both, as insect antifeedants and growth inhibitors (BRUNO et al., 2001; COOPER-DRIVER; LEQUESNE, 1987), it is likely that in Espeletiinae, such plesiomorphic traits are evolutionarily conserved as an effective protection mechanism against herbivory, similar to the reported role for

the *ent*-kauranes exuded by the secretory canals of the root cortex from four species of the closely related genus *Smallanthus* (COLL ARAOZ et al., 2010). According to Ehrlich and Raven (1964), plant diversification is often the result of innovation in defenses against natural enemies, and their occurrence apparently reflects both adaptations to particular life strategies and common descent. Hence, in plant groups with high diversification rates, such as Espeletiinae, the evolution of novel chemical defenses could be related to a higher ecological success and increased speciation.

Although Pleistocene climatic fluctuations, geographic isolation, frequent hybridization and adaptive morphological evolution are proposed as the main factors influencing the rapid radiation of Espeletiinae (DIAZGRANADOS; BARBER, 2017; MONASTERIO; SARMIENTO, 1991; POUCHON et al., 2018), the ecological implications and adaptive advantages given by chemical traits may have also played an important role in the diversification of the frailejones, similar to the well-known adaptive advantages given by the rosette growth form, foliar pubescence and insulation of the stem with dead leaves seen in most species of Espeletiinae (MONASTERIO; SARMIENTO, 1991; POUCHON et al., 2018). Such morphological characteristics represent key ensembles of traits that triggered the adaptive radiation of the frailejones (POUCHON et al., 2018) by improving photosynthetic assimilation, insulating the apical bud and the youngest leaves, intercepting and channeling rainfall, providing protection against freezing of the apical meristem and reducing water loss through transpiration (MONASTERIO; SARMIENTO, 1991). In Espeletiinae, the unbranched caulescent rosette growth form appears to have evolved independently three different times in the evolutionary history of the subtribe (one in Colombia and two in Venezuela) in response "to the cold, dry, and irradiated conditions that prevail at high elevations in both countries", suggesting an important role for natural selection in the diversification of Espeletiinae (POUCHON et al., 2018), similar to its role in other Andean taxa, such as the genus *Lupinus* (CONTRERAS-ORTIZ et al., 2018; NEVADO et al., 2016, 2018). For instance, based on genomic and transcriptomic data of multiple species and populations of Andean *Lupinus*, Nevado et al. (2018) concluded that "while geographical isolation may initiate speciation, species differentiation in the face of changing environmental conditions is only achieved once positive selection drives fixation of different adaptive alleles in each species". Thus, considering that secondary metabolites play essential roles in adaptation and defense (HARBORNE, 1989), it appears likely that both the

environment and natural enemies exerted strong selective pressures and shaped the metabolic fingerprints of the different Espeletiinae lineages, contributing to the adaptive success of the subtribe to the Andean highlands and acting as a driving force in adaptation and chemical evolution.

Putative biochemical conservation of GAO

Despite the random distribution of STLs in Espeletiinae and the ambiguous reconstruction of ancestral states for this chemical class, PCR amplification of a partial GAO sequence was successful even in species where no STLs were detected, including members from five different genera (*Espeletia*, *Espeletiopsis*, *Libanothmanus*, *Ruilopezia* and *Tamania*) and representatives from major clades, suggesting that GAO, the key gene involved in STL biosynthesis in Asteraceae, is putatively present and evolutionarily conserved in the subtribe Espeletiinae. The putative presence of GAO in all major clades of Espeletiinae could be explained by inheritance from ancestral taxa producing STLs, although it appears that in most of the species, this gene could be inactive, latent or being simply unexpressed. The *in vitro* functional characterization of GAO from Espeletiinae is still necessary to shed light on this issue.

According to Nguyen et al. (2016), STLs represent specialized metabolites thought to have contributed to the adaptive success of the Asteraceae family, giving an adaptive advantage to the producing plants by acting mainly as allelochemicals and anti-herbivory agents (AMBRÓSIO et al., 2008; MACÍAS et al., 2006; RODRIGUEZ; TOWERS; MITCHELL, 1976). The fact that the genus *Smallanthus*, the sister group of Espeletiinae, produces STLs further supports the hypothesis that the capacity of biosynthesizing STLs was already present in the common ancestor of the frailejones. Therefore, the putative biochemical conservation of GAO in Espeletiinae suggests that this trait might not represent a chemical homoplasy that evolved several times in the evolutionary history of Espeletiinae, as suggested by our ancestral states reconstruction, but represent a plesiomorphic trait inherited from ancestral taxa. Previous studies have identified homologous genes to GAO in the three major subfamilies of Asteraceae (Asteroideae, Cichorioideae, and Carduoideae), as well as in the phylogenetically basal Barnadesioideae, where no STLs have been detected to date (NGUYEN et al., 2010). Furthermore, "the recombinant GAOs from these genes clearly showed germacrene A oxidase

activities, suggesting that GAO activity is widely conserved in Asteraceae including the basal lineage" (NGUYEN et al., 2010). Therefore, the successful amplification of GAO in the recently evolved subtribe Espeletiinae, including species-producing and species-lacking STLs, and in all major subfamilies of Asteraceae, including the basal Barnadesioideae, suggests that the capacity of biosynthesizing germacrene A acid, the putative precursor of all STLs in Asteraceae, is highly conserved throughout the entire family, being present in ancestral and derived taxa (NGUYEN et al., 2010).

Although STLs are likely to be in part attributable to the evolutionary success of Asteraceae (NGUYEN et al., 2010), in Espeletiinae the structural diversity of this chemical class is very low. Only three melampolides have been detected in rather low concentrations, which raises concerns about their role, if any, in the adaptation and diversification of the frailejones. On the other hand, *Smallanthus*, the sister group of the frailejones, presents a higher structural diversity of STLs, where more than 40 different individual compounds belonging to different chemical subclasses (eudesmanolides, leucantholides and especially melampolides) have been reported, suggesting that STLs might not represent key traits triggering the adaptive success of Espeletiinae, contrary to the accepted notion in other Asteraceae lineages. However, "considering that the genes in plant secondary metabolism, such as cytochrome P450s of the CYP71 family, are often rapidly diversified for neofunctionalization", the conservation of GAO for more than 30 million years, since the beginning of the Asteraceae evolution, suggests that GAO offers increased fitness to plants in this family (NGUYEN et al., 2010), at least in some of its widespread lineages that diversified "outside of South America" as proposed by Nguyen et al. (2016). Further analyses are necessary to determine whether the low diversity of STLs seen in Espeletiinae compared to its sister genus *Smallanthus* is due to a putatively lower selective pressure by insect herbivores found in the high Andean forests and páramos compared to the Andean lowlands, where *Smallanthus* thrives better or, in a more likely scenario, is due to the presence of other compounds playing a similar ecological role (e.g., *ent*-kaurane diterpenes).

CHAPTER 4: DEVELOPMENTAL AND ENVIRONMENTAL REGULATION OF THE SECONDARY METABOLISM OF *Smallanthus sonchifolius* (ASTERACEAE): A SISTER TAXON OF ESPELETIINAE

INTRODUCTION

With more than 23,000 species currently described, Asteraceae represents one of the largest angiosperm plant families with a global distribution in which secondary metabolites are attributed to have played a crucial role in its evolutionary success (CALABRIA et al., 2009; CRONQUIST, 1977; JEFFREY, 2007; TURNER, 1977). According to Ehrlich and Raven (1964), plant diversification is often the result of innovation in defenses against natural enemies. Therefore, in plant groups with high diversification rates, such as Asteraceae, the evolution of novel chemical defenses is related to a higher ecological success and increased speciation. The high diversity of Asteraceae, especially in the tropical Andes of South America (PANERO; CROZIER, 2016), is represented by the large array of secondary metabolites that have been identified in some of its members (CALABRIA et al., 2009). Flavonoids, *trans*-cinnamic acid derivatives, polyacetylenes, coumarins and sesquiterpene lactones (STLs), in addition to other terpenoids and acetylenes, constitute characteristic metabolites commonly found in Asteraceae (CALABRIA et al., 2009). These metabolites possess well-known ecological roles, such as pollinator attraction, allelopathy, and protection against parasites, pathogens, herbivores, and abiotic environmental factors (HARBORNE, 1989). Therefore, secondary metabolites constitute the chemical interface between the plant and its environment, an interaction mediated by thousands of metabolites.

Climatic factors such as water availability, temperature and solar radiation are well described as being able to influence the production of several metabolites from different chemical classes (AKULA et al., 2011; JAKOBSEN; OLSEN, 1994; SAMPAIO; EDRADA-EBEL; DA COSTA, 2016). Examples of this are the substantial levels of phenolic compounds accumulated in the upper epidermal layer of plant cells in response to high levels of UV-B radiation (SILVA et al., 2014). The presence of water-soluble polymers of fructose, such as fructans and inulins in some Asteraceae species, e.g., *Smallanthus sonchifolius* (Poepp. & Endl.) H. Robinson, represents another example, as these molecules are associated with the plant ability to

withstand low temperatures and drought conditions (KRASENSKY; JONAK, 2012; LIVINGSTON; HINCHA; HEYER, 2009). The accumulation of these fructose polymers has been hypothesized to help explain the adaptive success of the family in the seasonally dry and cold environments characteristic of some tropical Andean ecosystems (HENDRY, 1996). The capitate glandular trichomes present in aerial parts of many Asteraceae species represent a clear example of protection against herbivores (PEIFFER et al., 2009). High amounts of STLs are biosynthesized and stored in this type of specialized trichomes (AMREHN et al., 2015; ASCHENBRENNER et al., 2015; BOMBO et al., 2015), acting as herbivore-deterrents (PICMAN, 1986) and serving as allelochemicals (MACÍAS et al., 1996). However, in spite of the well-described adaptive roles that secondary metabolites possess in Asteraceae, most of the studies are based on monitoring single compounds or chemical classes under controlled laboratory conditions. Therefore, the regulatory factors present in nature and their influence on the metabolism of Andean taxa from a metabolomics perspective remain poorly understood.

Smallanthus sonchifolius, popularly known as yacón, is a perennial herb native to the South American Andes (GRAU; REA, 1997). Among the closest relatives of yacón, the subtribe Espeletiinae stands out as a recently diversified plant group in the northern Andes (RAUSCHER, 2002). Yacón and Espeletiinae share not only high homology in their ITS sequences, as pointed out by phylogenetic studies, but also share some key morphological and chemical traits (CUATRECASAS, 2013; DIAZGRANADOS; BARBER, 2017). However, considering that yacón constitutes a fast growing species with a high plasticity to adapt in different environments, contrary to Espeletiinae, it can be considered as a good model organism to study the influence of intrinsic (e.g., plant developmental stage) and extrinsic environmental factors on the biosynthesis and accumulation of secondary metabolites in Andean groups. Recent studies suggest *Smallanthus* and Espeletiinae have a very similar secondary metabolite chemistry, characterized by the presence of flavonoids, caffeic acid derivatives, and STLs (LACHMAN; FERNÁNDEZ; ORSÁK, 2003; PADILLA-GONZÁLEZ et al., 2017, 2018). Therefore, as a first step to understand the regulatory factors and adaptive roles secondary metabolites possess under changing environmental conditions in Andean groups, we aimed to study the metabolome of yacón leaves along the development of the plant by correlating metabolic fingerprints and gene expression patterns with environmental variables.

The possibility of monitoring the metabolome of yacón leaves in combination with the expression levels of key genes and their relationships with environmental factors could largely improve the understanding of the adaptation mechanisms Andean plants possess under certain environmental conditions at specific developmental stages. Additionally, considering that yacón constitutes a functional food and medicinal plant commercialized for the treatment of hyperglycemic disorders, among which STLs and CADs play an active role (DELGADO et al., 2013; GENTA et al., 2010), the identification of the regulatory factors modulating the biosynthesis of these metabolites can be used as eliciting treatments to develop advanced cultivars with increased protection and bioactivity (GENTA et al., 2010). Thus, given the ecological significance and medicinal applications of chlorogenic acids (CGAs), flavonoids and STLs, key genes involved in their biosynthetic pathway (hydroxycinnamoyl CoA: quinate hydroxycinnamoyl transferase (HQT), chalcone synthase (CHS) and germacrene A oxidase (GAO), respectively (Figure 4.1)) were chosen to monitor their expression levels during different plant developmental stages.
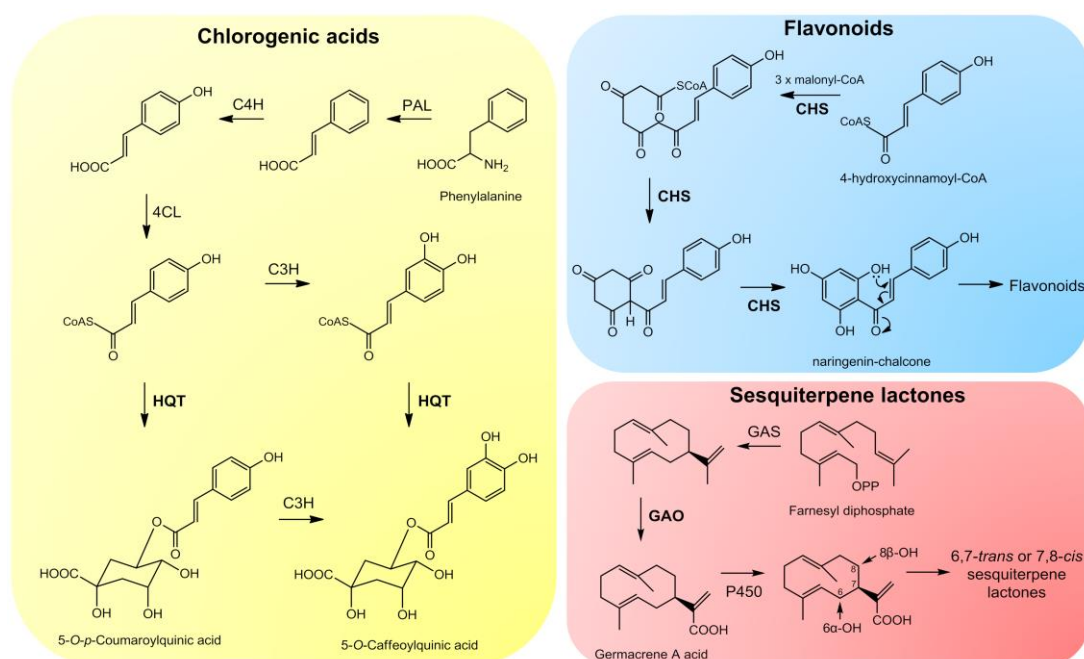
Figure 4.1 Biosynthetic pathways of chlorogenic acids, flavonoids and sesquiterpene lactones in Asteraceae. Chlorogenic acid enzymes: phenylalanine ammonia lyase (PAL), cinnamate 4'-hydroxylase (C4H), 4-cinnamoyl CoA ligase (4CL), hydroxycinnamoyl CoA: quinate hydroxycinnamoyl transferase (HQT), *p*-coumaroyl-3'-hydroxylase (C3H); flavonoid enzymes: chalcone synthase (CHS); sesquiterpene lactone enzymes: germacrene A synthase (GAS), germacrene A oxidase (GAO).

METHODS

Plant material

Smallanthus sonchifolius plants were cultivated on a trial field of the Institute of Botany at the University of Hohenheim, Stuttgart, Germany, from April to October 2017 (during 176 days) until they reached the adult stage. Considering that the investigated clone did not naturally flower in the field, the maturity of the plant was determined in accordance with hallmarks found within the literature: when plants reached more than 1.5 m height, stems began to wither away and bend and when tuberous roots were completely developed (FERNANDEZ, C.E. VIEHMANNOVA et al., 2007). In the conditions of central Europe, the length of the vegetation period of yacón ranges from 149 to 170 days (FERNANDEZ, C.E. VIEHMANNOVA et al., 2007). In this study, all plants came from the same clone held at the Botanic Garden of the University of Hohenheim, and they were propagated by rhizome cuttings in the form of germinated rhizomes to ensure genetic homogeneity in the studied samples.

Collection of yacón leaves started 15 days after sprouting and continued every 15 days for a period of six months. In each collection, the third pair of leaves (from top to bottom), corresponding to the first well-expanded leaf, was sampled and stored at -70°C until further use. Collections were always performed at the same hour of the day (between 17:00 and 18:00 h). The development of yacón can be divided into four different growth phases after sprouting (excluding the flowering phases): leaf development, stem ramification, crop cover and formation of tuberous roots[50]. In each collection, the developmental stage of the plant was recorded. Three biological replicates were collected for each time point.

Environmental data

For each collection date, climate data from the closest climatological station (Wetterstation Hohenheim) to the cultivation field were obtained. Climate data included temperature (°C), air humidity (%), solar radiation (Wh/m2) and rainfall (mm) (Figure 4.2). For data analysis, average (temperature, humidity and rainfall) values for the 15 days prior to each collection were considered. In the case of solar radiation, the accumulated value for the day of collection was considered as previous studies demonstrate that UV-light induces a rapid response in gene expression (BARTLEY et al., 2016).
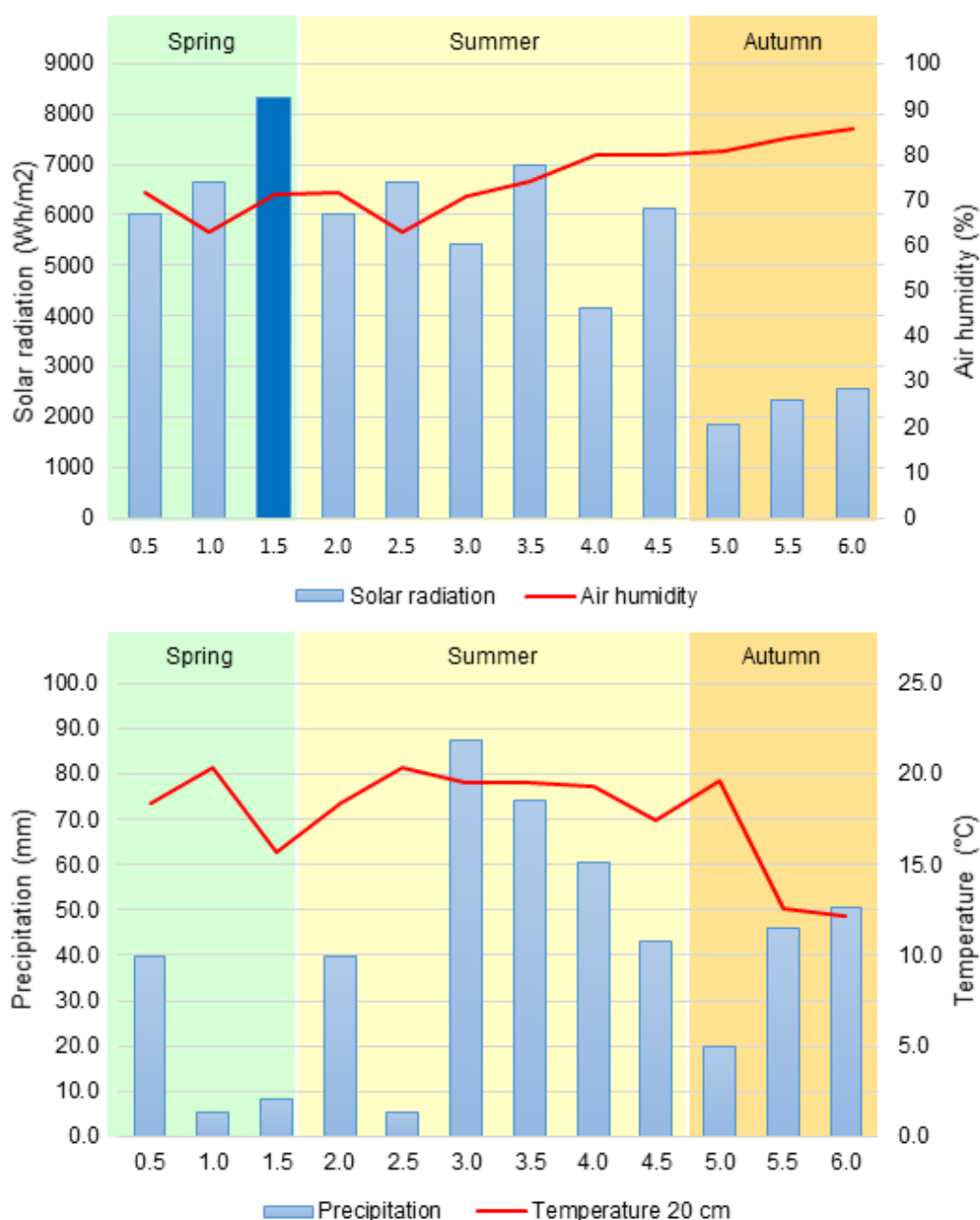
Figure 4.2 Climatograms reporting environmental variables from the Hohenheim climate station for the twelve collections.

Metabolite extraction and UHPLC-UV-HRMS analysis

The collected samples (50 mg/sample – fresh weight) were homogenized in liquid nitrogen and extracted with 70% aqueous ethanol (1 mL) by vortexing for a few seconds followed by ultrasonication at room temperature for 10 min at 40 kHz. After extraction, samples were centrifuged for 5 min at 13,000 rpm and the supernatant was filtered through a 0.2 μm PTFE filter. The extraction process used in this study

was based on the protocol of extraction and analysis of plant tissues for metabolomics studies reported by de Vos et al. (2007).

Metabolic fingerprinting by UHPLC-UV-HRMS was performed on an Agilent 1290 UHPLC system (Agilent, Germany) coupled to an 80 Hz photodiode array detector (PDA) and a *Q Exactive Plus* (Thermo Scientific, USA) high-resolution mass spectrometer. Chromatographic separation of plant extracts (5 µL) was performed on an Acquity CSH C18 column (1.7 µm, 150 mm x 2.1 mm, Waters, Ireland) using water (0.2% formic acid) as solvent A, acetonitrile (0.2% formic acid) as solvent B (flow rate: 400 µL/min) and the following program: 0-15 min, 3-20% B; 15-40 min, 20-95% B; 40-43 min, 95-3% B; 43-45 min, 3% B. UV detection was performed between 190 and 400 nm.

Mass spectrometry detection was performed in both positive and negative ionization modes using the *Fullscan* (resolution of 70,000 FWHM) and data-dependent MS^2 (dd-MS^2, resolution of 17,500 FWHM) methods. Total Ion Current (TIC) chromatograms were obtained over the range of 140-1200 *m/z* using a spray voltage of +4.2 kV and -3.5 kV for the positive and negative ionization modes, respectively. Additional parameters for the mass spectrometer included: capillary temperature, 360°C; heater temperature, 380°C; automatic gain control (AGC) target, 1.0e6 (*Fullscan*) and 5.0e4 (dd-MS^2); maximum inject time, 500 ms (*Fullscan*) and 64 ms (dd-MS^2); sheath gas flow rate, 60; auxiliary gas flow rate, 20; topN for dd-MS^2, 5 and an isolation window of 1.5 *m/z*. Nitrogen was used as the drying, nebulizer and fragmentation gas.

Data preprocessing and multivariate analyses

Chromatographic raw data obtained for each ionization mode (positive and negative) were directly uploaded and processed by MZMine 2.28, where peak detection, peak filtering, chromatogram construction, chromatogram deconvolution, isotopic peak grouping, chromatogram alignment, gap filling, duplicate peaks filter, fragment search and the search for adducts and peak identities were performed. The following MZmine parameters were used for data preprocessing: noise level at 1.0E5; Lorentzian function as peak shape algorithm (resolution of 70,000); minimum peak height at 5.0E5; *m/z* tolerance at 0.002 *m/z* or 5.0 ppm and retention time tolerance of 0.7 min. After MZmine preprocessing, data for each ionization mode were exported as .csv tables with rows representing plant extracts and columns representing ion

peak areas associated with a given mass and retention time value. Before statistical analyses, peaks detected in the blank (extraction solvent) and in the final stages of the chromatographic run were removed from the original matrix and the final matrix was scaled by the Pareto method.

Multivariate statistical analyses were performed in the software R 3.0.3 (R Foundation for Statistical Computing, Austria) and SIMCA 13.0.3.0 (Umetrics, Sweden). Initially, a Hierarchical Clustering Analysis with bootstrap resampling (HCAbp) was performed using 10,000 bootstrap replicates, the Ward's method and a Euclidean distance, with the aim of reducing the dataset dimension and recognizing the natural clustering tendency of the samples based on their metabolic fingerprints. To identify the discriminating variables accounting for the observed groups in the HCAbp, a supervised method by Orthogonal Partial Least Squares-Discriminant Analysis (OPLS-DA) was performed with the same dataset, considering the HCAbp groups as Y variable. Heatmaps were built with all of the identified compounds (see below), using the R package *gplots,* in order to create an overview of the relative accumulation of each metabolite in all samples. Finally, Canonical Correspondence Analysis (CCA) was carried out in the R package *vegan* to correlate the environmental data with the metabolic fingerprints of the samples.

The environmental data sets were processed according to the type of data. Data expressed as percentages (e.g., relative air humidity) were transformed by the arcsine method, while the logarithmic scale was applied to the data not expressed as percentages (e.g., solar radiation, temperature and rainfall) in accordance with the literature (SAMPAIO; EDRADA-EBEL; DA COSTA, 2016).

Dereplication of plant extracts

Dereplication of plant extracts was performed by molecular networking based on MS/MS spectral similarity, following the online workflow at GNPS (WANG et al., 2016). Initially, raw data from each ionization mode (positive and negative) was transformed to nonproprietary .mzXML format using the MSConvert package from the ProteoWizard 3.0.9798 software (Proteowizard Software Foundation, USA) to subsequently upload it to the GNPS platform. Consensus spectra of nearly identical MS/MS spectra were created via MS-Cluster, considering a parent and a fragment ion mass tolerance of 0.02 Da and a minimum of two MS/MS spectra in a consensus to be considered for molecular networking. Networks were created with edges having

a cosine score above 0.65 and a minimum of 4 common fragment ions shared by two separate consensus MS/MS spectra in order to be connected by an edge. Edges between two nodes were kept in the network only if both nodes were within the top 10 most similar nodes to each other. Spectral library annotation was performed considering a minimum of 4 common fragment ions, and Cytoscape 3.6.1 was used to display and edit the final structure of the network.

To confirm the compound annotation made by molecular networking, we performed a systematic comparison of retention time, UV, HRMS and MS/MS spectra of experimental data with authentic standards and data from the literature. Accurate mass comparisons were performed relative to the theoretical monoisotopic masses (<3 ppm accuracy) of the secondary metabolites previously reported in the genus *Smallanthus*. For that, a database containing all secondary metabolite reports in species of the genus *Smallanthus* (Appendix 7) was built as follows: data collection using SciFinder (Chemical Abstract Service, USA) considering the scientific papers published between the years 1970 and 2014, chemical structure drawing using MarvinSketch 14.9.15.0 (ChemAxon Ltd., Hungary) and database creation, handling and visualization using JChem for Excel (ChemAxon Ltd., Hungary). This database contains 414 entries of 298 2D chemical structures and its information is freely available at www.asterbiochem.org/asterdb. To identify compounds previously unreported in the genus *Smallanthus*, accurate mass values were used to perform online database searches in the Dictionary of Natural Products (DNP, http://dnp.chemnetbase.com) and in SciFinder Scholar. To confirm peak assignments, the fragmentation patterns of the identified metabolites were proposed based on MS^2 spectra and reference substances were used to compare accurate mass measurements and MS^2 spectra to those of the detected metabolites. Based on the Metabolomics Standard Initiative (SUMNER et al., 2007), three levels of confidence were adopted to report the accuracy of the identifications: level 1 (identified compounds), based upon co-characterization with authentic samples; level 2 (putatively annotated compounds), "based upon physicochemical properties and/or spectral similarity with public/commercial spectral libraries" and level 3 (putatively characterized compound classes), "based upon characteristic physicochemical properties or spectral similarity to known compounds of a chemical class". In all cases, a minimum of two independent and orthogonal data (e.g., accurate mass and MS^2 fragmentation patterns or retention time and accurate mass, etc.) were

considered. Detailed information about all dereplicated metabolites is reported in Appendix 8.

Primer design and sequencing

To investigate the expression patterns of key genes involved in the secondary metabolism of yacón, a species for which scarce information regarding genomic and molecular traits is available, we designed degenerate primers based on homologous sequences from other Asteraceae species (Table 4.1). Genes involved in the biosynthesis of CGAs, flavonoids and STLs, representing the main classes of SM accumulated in yacón leaves, were investigated. Among the CGA pathways, the enzyme hydroxycinnamoyl CoA: quinate hydroxycinnamoyl transferase (HQT), which catalyzes the biosynthesis of CGA (NIGGEWEG; MICHAEL; MARTIN, 2004) was selected. Among the flavonoid pathways, the gene coding for the enzyme chalcone synthase (CHS), which catalyzes the conversion of 4-coumaroyl-CoA and malonyl-CoA to naringenin chalcone, was chosen, while for the STL pathways, the gene coding for the enzyme germacrene A oxidase (GAO) was investigated (Figure 4.1). GAO catalyzes three consecutive oxidations of germacrene A to yield germacrene A acid, the putative precursor of all STLs in Asteraceae (NGUYEN et al., 2010). Actin (ACT) and elongation factor (EF) were selected as housekeeper genes based on literature reports (LU et al., 2013; TANG et al., 2015) and their high expression stability (ACT = 0.03 and EF = 0.015) (Table 4.2) (BAO et al., 2016).

Table 4.1 Species used to design degenerate primers for GAO, CHS and HQT.

Germacrene A oxidase (GAO)		Chalcone synthase (CHS)	
Species	**Genebank ID**	**Species**	**Genebank ID**
Helianthus annuus	GU256646.1	*Helianthus annuus*	KR921882.1
Lactuca sativa	GU198171.1	*Helianthus annuus*	XM_022178725.1
Tanacetum cinerariifolium	KC441527.1	*Helianthus annuus*	XM_022153962.1
Cichorium intybus	GU256644.1	*Rudbeckia hirta*	EF070339.1
Saussurea costus	GU256645.1	*Echinacea pallida*	KY094648.1
Cynara cardunculus	KF752448.1	*Ageratina adenophora*	FJ913888.2
Barnadesia spinosa	GU256647.1	*Dahlia pinnata*	AB591825.1
Hydroxycinnamoyl-CoA:quinate hydroxycinnamoyltransferase (HQT)		*Dahlia pinnata*	AB591826.1
Cynara cardunculus	DQ915590.1	*Carthamus tinctorius*	LC128420.1
Cynara scolymus	DQ915589.1	*Gynura bicolor*	AB550239.1
Cynara cardunculus	KU711509.1	*Eschenbachia blinii*	KJ155749.1

Cichorium intybus	KT222893.1	*Centaurea jacea*	EF112474.1
Helianthus annuus	XM_022170642.1	*Lactuca sativa*	AB525909.1
Cynara cardunculus	JF338140.1		

For primer design, homologous sequences (complete coding sequences) from several Asteraceae species (Table 4.1) were retrieved from GenBank (www.ncbi.nlm.nih.gov/genbank), aligned in BioEdit (HALL, 1999) and inspected for conserved regions. Degenerate primers (Table 4.2) were designed using the software module Primer Select from Lasergene (V7.0.0, DNASTAR) on highly conserved regions and tested on gDNA extracted from yacón leaves following the protocol of Ristaino et al. (2001) with the following modification: isopropyl alcohol was used instead of ethanol for DNA precipitation and 40 mg of insoluble polyvinylpyrrolidone (PVP) was added after the nucleus lysis buffer to absorb polyphenols.

Table 4.2 Primer combinations used in PCR and qPCR.

Code	Primer (5' - 3')	Size (bp)	AT (ºC)	Efficiency*	R^2
Ss_qGAO	F: CGAAAACGGCAACACCACCATT	162	60	92.3	0.998
	R: GCTCGCACCATTGGGAAGTTTC				
Ss_qCHS	F: GCCGACTACCAGCTCACCAAACTC	193	60	87.2	0,999
	R: CCTCATTAGGGCCACGGAACG				
Ss_qHQT	F: GCTACCTTGGGAACGTGGTCTTTACA	194	60	91.2	0.997
	R: GTAACTGGGTCCACGGATTAGAGCC				
Ha_qACT	F: GCCGTGCTTTCTCTTTATGCCAGCGACC	137	60	95.0	0.998
	R: AGCGAGATCAAGACGAAG				
Ha_qEF	F: ACCAAATCAATGAGCCCAAGAGACCCA	131	60	99.6	0.987
	R: TACCGGGCTTGATCACACCAG				
Ss_GAO	F: CARTCGAATGGGCGATTTC	532	54	-	-
	R: CTTTGAACCGTGGCTCCAA				
Ss_CHS	F: AAGCCATYAAAGAATGGGGAAA	729	54	-	-
	R: CCRCCGACTTCTTCCTCATCTC				
Ss_HQT	F: GATATGGCTCGCGGGTTCTC	617	56	-	-
	R: GCATCATGTACTGGTAGCCTGGTC				

* Determined only for qPCR primers; AT: annealing temperature.

PCRs were performed in a peqSTAR Thermocycler (Peqlab, Germany) using RNase/DNase-free water and RedTaq MasterMix (Genaxxon Bioscience, Germany). The PCR conditions included 4 min of initial denaturation followed by 38 cycles of 15 s denaturation (96°C), 15 s annealing (GAO: 54°C, CHS and HQT: 56ºC), 1 min x kb[-1] elongation (72°C) and final elongation for 4 min (72°C). All PCR products were

sequenced by Macrogen (Netherlands), and their identities were confirmed by BLAST searches. After sequencing the target genes on gDNA, the obtained sequences were used as templates to design specific qPCR primers (Table 4.2) to be used in the gene expression analyses. qPCR primers were tested on yacón cDNA (see below) to confirm the presence of the expected product size and the absence of secondary non-specific bands following the same PCR conditions described above.

RNA extraction and RT-qPCR

For RNA extraction, six different time points were chosen (0.5, 1.5, 2.5, 3.5, 5.0 and 6.0 months), representing important changes in the yacón metabolome, with one collection for each month. Total RNA from fresh yacón leaves (20 mg) was extracted using the EURx GeneMATRIX Universal RNA Purification kit (Roboklon, Germany), following the manufacturer's instructions with the following modification: forty milligrams of insoluble PVP was added after the RL buffer to absorb polyphenols. The quantity and purity of the isolated RNA was verified in a BioPhotometer (Eppendorf, Germany). To eliminate residual genomic DNA, the Perfecta Dnase I (Rnase-free) kit (Quanta Biosciences, USA) was used. Samples were confirmed to be free from gDNA by PCR analysis using actin primers. Before cDNA synthesis, the RNA concentration of all samples was adjusted to 50 ng/mL. For qPCR reactions, cDNA synthesis was performed using the RevertAid cDNA Synthesis Kit (Thermo Scientific). The primer used for the first strand synthesis was the VNdT18-Oligonucleotide at a final concentration of 5 µM.

Real time qPCR analyses were performed in a CFX96 Touch™ qPCR System (BioRad GmbH, Germany) and SYBR Green I was used for visualization (SensiFast™ SYBR No-ROX Kit, Bioline). HPLC-purified qPCR primers in a final concentration of 0.25 µM were used. The qPCR program was 2 min at 95°C for initial denaturation, followed by 45 cycles of 15 s at 95°C and 15 s at 60°C. The gene expression analysis was performed using the Bio-Rad CFX Maestro 1.0 software (Bio-Rad Laboratories GmbH). The relative expression of each gene was normalized relative to the youngest stage and to the housekeeper expression of ACT and EF. A melting curve analysis following each RT-qPCR was performed to assess product specificity (Appendix 9), and the efficiency of all primers was calculated using four serial dilutions of 1:5 (Appendix 10). Gene expression patterns were analyzed in three biological replicates, each with three technical replicates.

RESULTS

Metabolic fingerprinting and correlation with developmental and environmental data

In conducting metabolic fingerprinting by UHPLC-UV-HRMS of yacón collected in different developmental stages, 1,562 and 1,353 mass signals were recorded in the positive and negative ionization modes, respectively. Based on HCAbp, the positive ionization mode dataset was chosen for further analysis. This mode was selected as a basis for proposing different groups of samples with similar metabolic fingerprints (Figure 4.3).
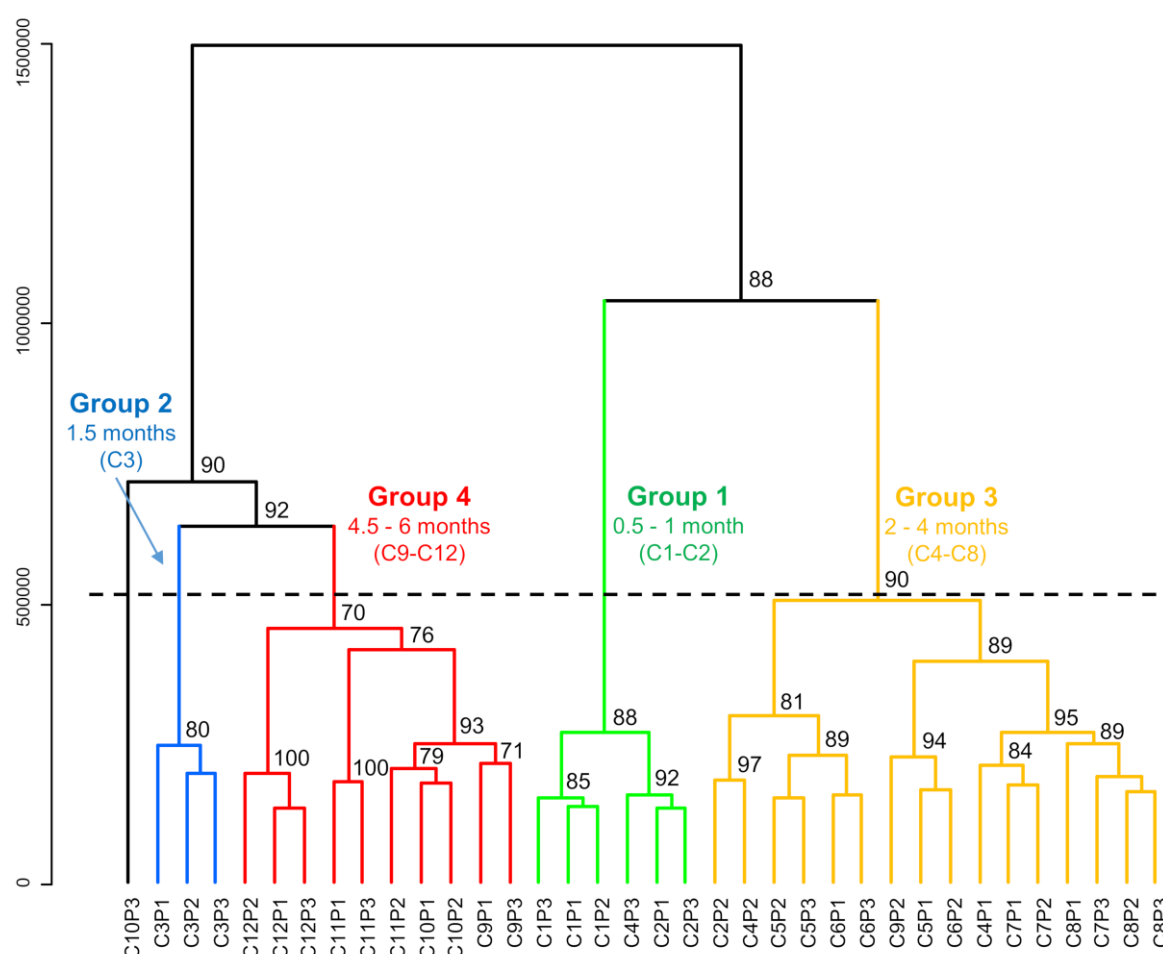
Figure 4.3 HCAbp based on metabolic profiling by UHPLC-UV-HRMS in positive mode of 36 extracts from *Smallanthus sonchifolius* leaves collected along the development of the plant (C1 to C12). Proposed groups of samples with similar metabolic fingerprints are highlighted in different colors.

Results from HCAbp (Figure 4.3) showed a clear clustering tendency by developmental stage in the metabolic fingerprints of yacón leaves, with four main groups of samples differentiating based on their metabolic fingerprints. Group 1 clusters the first two collections (C1 and C2), corresponding to leaves produced by 15 and 30-day-old plants, respectively, while leaves from 1.5-month-old plants formed an independent cluster in the HCApb (Group 2, Figure 4.3). Similarly, leaves from two to four-month-old plants clustered in an independent group (Group 3, Figure 4.3), while leaves from the last stages of plant development (between 4.5 and six-month-old plants) clustered in group 4 (Figure 4.3). Interestingly, leaves from 1.5-month-old plants (Group 2, Figure 4.3) appear to be metabolically more similar to the leaves produced by plants reaching the adult stage (Group 4, Figure 4.3) than to the leaves produced by slightly younger or older plants. Additionally, different subclusters segregate specific collections in groups 3 and 4 (Figure 4.3). For example, collections 7 and 8, corresponding to 3.5 and four-month-old plants (Figure 4.3), tend to form an independent subcluster in group 3, with a similar pattern observed among collections 9 and 10 in group 4 (4.5 and five-month-old plants, respectively, Figure 4.3). In most cases, high bootstrap values (>80%) were obtained, indicating good support for the obtained clusters.

Further analysis of the same dataset by OPLS-DA (R^2=0.93, Q^2=0.74,) and heatmaps (Figure 4.4) pointed to the discriminant metabolites responsible for the observed clustering tendency in the HCAbp. Leaves from 1.5-month-old plants (Group 2, Figure 4.4) accumulate mostly CGAs and other caffeic acid derivatives, of which 1-*O*-caffeoylquinic acid is the major discriminatory compound for this group, along with its isomer 5-*O*-caffeoylquinic acid (Figure 4.4). This class of metabolites is also accumulated in high amounts (although to a lesser extent) by samples belonging to group 4 (Figure 4.4), corresponding to 4.5 to six-month-old plants, explaining the close clustering distance exhibited among samples in groups 2 and 4 (Figure 4.3 and 4.4). However, samples in group 4 accumulate mostly two dicaffeoylquinic acid isomers, namely, 3,5- and 4,5-di-*O*-caffeoylquinic acids (Figure 4.4), as well as several flavonoids, STLs and diterpenes. The STL enhydrin and the diterpene smaditerpenic acid F appear as the main discriminatory substances of group 4 based on the OPLS-DA model. The accumulation of several metabolites from different chemical classes within this group of nearly adult and adult plants (Group 4, Figure 4.4) suggests that plant age affects the metabolism of yacón, as adult plants are

metabolically more diverse than their younger counterparts. The low chemical diversity found among 15 and 30-day-old plants further supports this fact (Group 1, Figure 4.4), as the clustering tendency of samples belonging to this group is rather due to their poor metabolic diversity (Figure 4.4). Leaves collected from two to four-month-old plants (Group 3, Figure 4.4) accumulate intermediate amounts of CGAs, flavonoids and STLs, although some metabolites such as smaditerpenic acid E, trihydroxy trimethoxyflavone, polymatin A and fluctuadin are accumulated in high amounts (Figure 4.4).
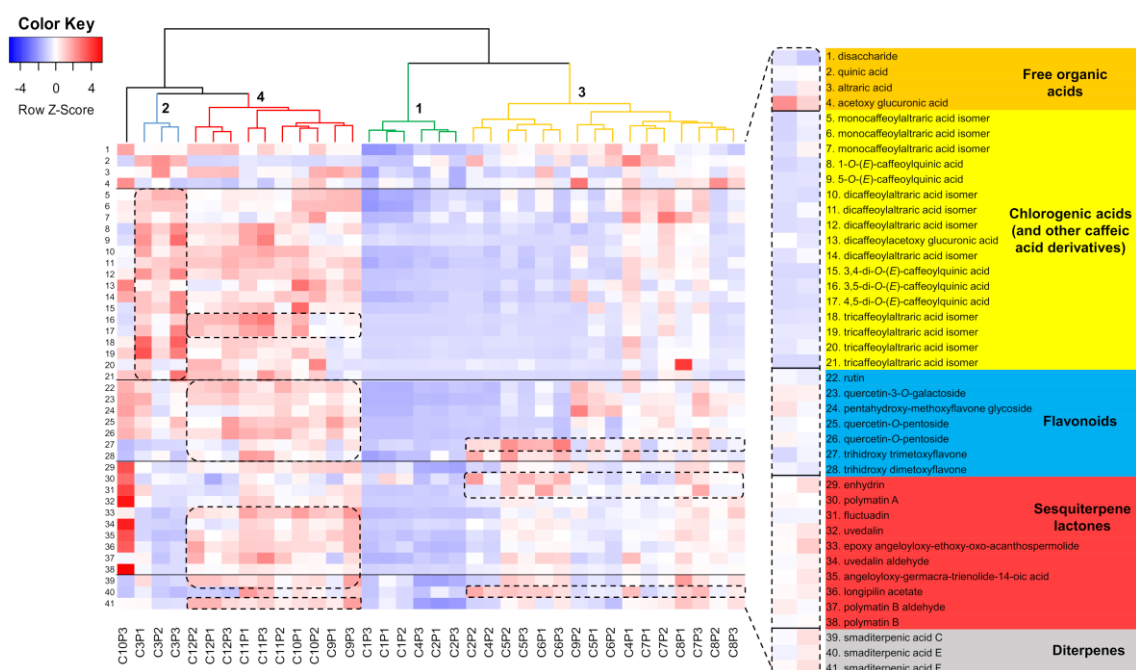
Figure 4.4 Heatmap showing the differential accumulation of 41 metabolites in 36 extracts from *Smallanthus sonchifolius* analyzed by UHPLC-UV-HRMS. Dash boxes highlight the preferential accumulation of metabolites by the four different groups of samples.

Results from CCA (Figure 4.5) allowed for correlating environmental variables and the plant developmental stage at the time of collection with the samples' metabolic fingerprints. This analysis revealed that each of the environmental factors considered had a different impact on the metabolic fingerprints of yacón leaves. Intermediate to high levels of solar radiation and temperature, in addition to early developmental stages and humidity, seem to influence the metabolic fingerprints displayed by yacón leaves from 0.5 and one-month-old plants (Group 1, Figure 4.5). The independent clustering of yacón leaves from 1.5-month-old plants (Group 2,

Figure 4.5) is correlated with the highest levels of solar radiation and low temperatures (Figure 4.2). On the other hand, the metabolic fingerprint displayed by yacón leaves from 5.5 and six months-old plants is associated with a high humidity and developmental stage, in addition to low temperatures and low levels of solar radiation (Figure 4.5). A similar pattern is observed among the other subcluster in group 4, containing 4.5 and five-month-old plants, but this group is also correlated with high temperatures (Figure 4.5) in addition to a high humidity and developmental stage. The metabolic fingerprints displayed by 2.5 and three-month-old plants are correlated with high temperatures and intermediate levels of solar radiation and humidity, while in 3.5 and four-month-old plants, high precipitation and high temperatures seem to be the two most important factors determining their metabolic composition (Figure 4.5).
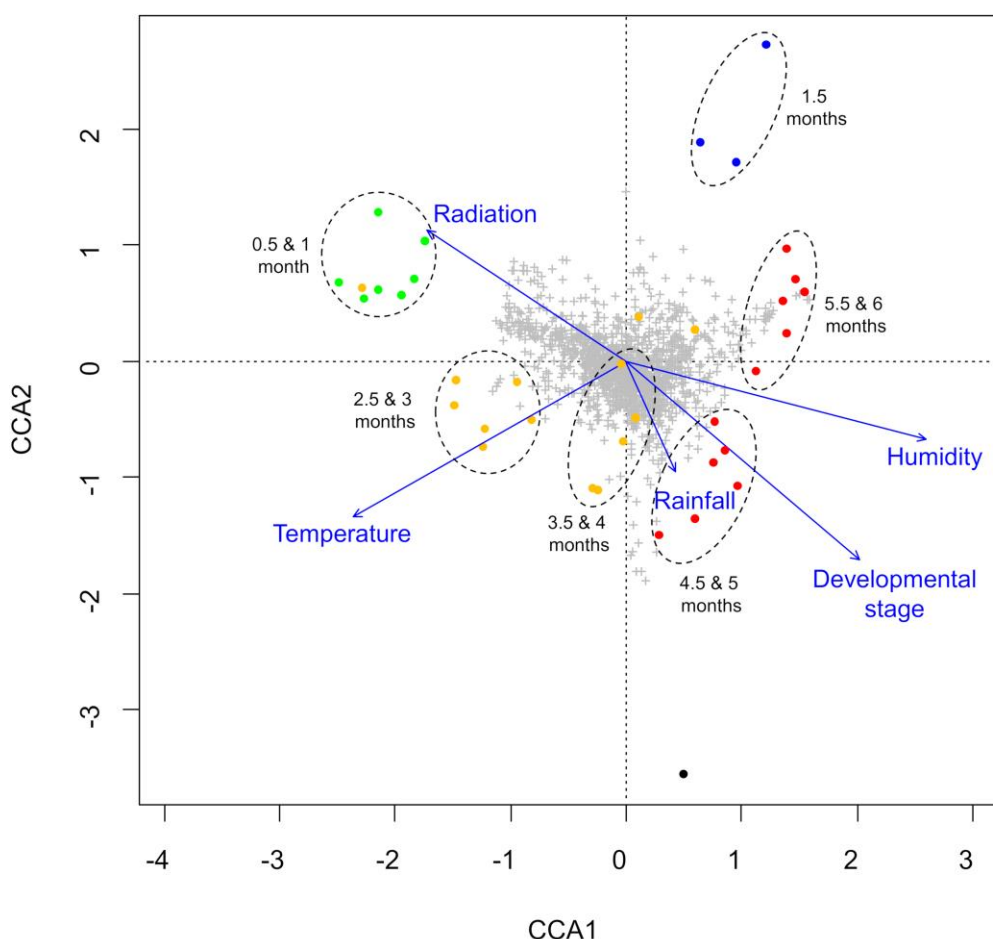

Figure 4.5 CCA based on metabolic profiling by UHPLC-UV-HRMS in positive mode of 36 extracts from *Smallanthus sonchifolius*, showing the correlation of environmental factors and developmental stage with the clustering of samples by

metabolic similarity. Samples colored according to their HCAbp groups (0.5 to one-month-old (green), 1.5-months-old (blue), two to four-months-old (yellow) and 4.5 to six-months-old (red)).

Combined metabolomic and gene expression patterns

We identified the homologous genes of three key enzymes involved in the biosynthesis of CGAs, flavonoids and STLs in yacón. Blast searches revealed high nucleotide identity to the corresponding genes of closely related Asteraceae: 91% for HQT (relative to *Helianthus annuus* L. HQT), 89% for CHS (relative to *Ageratina adenophora* (Spreng.) R.M. King & H. Rob. CHS) and 85% for GAO (relative to *H. annuus* GAO). Although gene expression analyses were carried out for only one gene of each biosynthetic pathway, the selected genes represent key steps in each pathway. Among CGAs, the overall patterns in metabolite accumulation are similar to the HQT expression results (Figure 4.6). There is a statistically significant peak in the accumulation of this class of metabolites in 1.5-month-old plants, which is closely mirrored by the gene expression results (Figure 4.6). Thus, the accumulation of CGAs in leaves of 1.5-month-old yacón plants is closely associated with the levels of HQT expression. However, after reaching a peak in 1.5-month-old plants, the accumulation of CGAs decreases in the subsequent collections to ultimately increase in the final stages of plant development (Figure 4.6), while HQT expression remains low even in the final stages of plant development (Figure 4.6). This suggests that the high amounts of CGAs present in adult yacón plants are not associated with higher levels of HQT expression. Analogous to CGAs, the accumulation of flavonoids is proportional to the levels of CHS expression in leaves of 1.5-month-old yacón plants, showing similar patterns in gene expression and metabolites accumulation (Figure 4.6). The accumulation of flavonoids increases significantly when plants reach 1.5 months old (Figure 4.6); however, contrary to CGAs, no significant changes occur thereafter. Gene expression data for CHS also show a significant upregulation at 1.5 months (Figure 4.6), followed by a significant decrease in the next collection, no significant changes in the following two collections and a significant upregulation at six months (Figure 4.6). On the other hand, the accumulation of STLs is inversely proportional to the levels of GAO expression in yacón leaves (Figure 4.6). The accumulation of STLs starts early in plant development when plants are 15 days old (or probably even before) and remains stable thereafter (Figure 4.6). Gene

expression results show an inverse tendency with a significant GAO expression when plants are 15 days old and a subsequent decrease to the final stages of plant development (Figure 4.6).
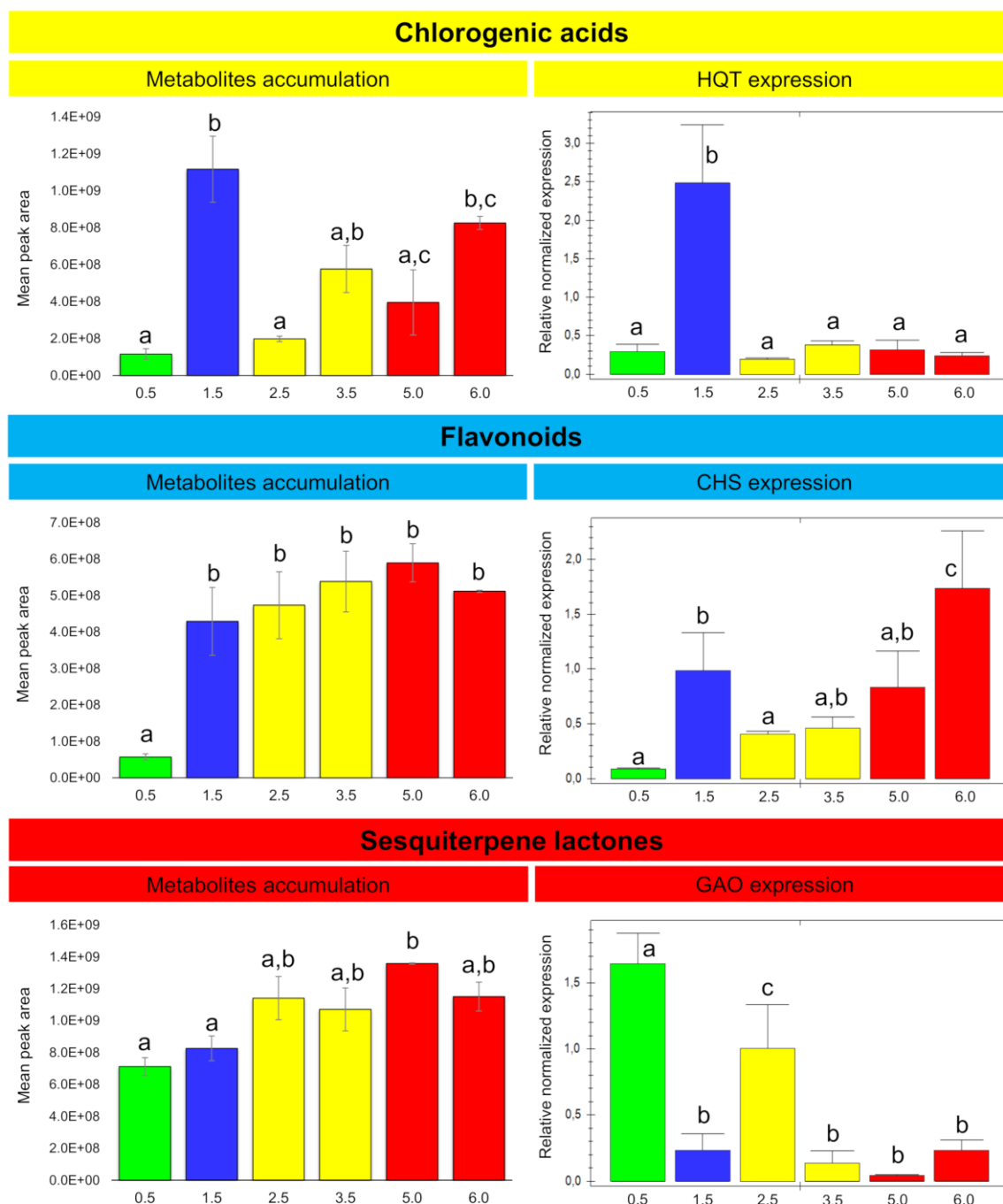
Figure 4.6 Accumulation of chlorogenic acids (8, 9, 15-17), flavonoids (22-28) and STLs (29-38) and gene expression of HQT, CHS and GAO during yacón development. Leaves samples collected from 0.5 to 6 months-old plants. Metabolites identities reported in Fig 3. Normalized gene expression is given relative to the youngest stage and housekeeping genes Act and EF (n=3). Error bar: SEM. HQT:

hydroxycinnamoyl CoA: quinate hydroxycinnamoyl transferase; CHS: chalcone synthase; GAO: germacrene A axidase; ACT: actin; EF: elongation factor 1a. Bars colored according to the HCAbp groups (0.5 to one-month-old (green), 1.5-months-old (blue), two to four-months-old (yellow) and 4.5 to six-months-old (red)).

Molecular networking-based dereplication

Clustering of the MS/MS spectra based on cosine similarity formed molecular networks with 823 and 491 parent ions visualized as nodes in the positive and negative ionization modes, respectively (Figure 4.7). Visual inspection of the molecular networks from both ionization modes showed that compounds from the same chemical class tend to cluster together in the network, where seven different chemical classes were identified: caffeoylaltraric acids and derivatives, free glycosides, glycosylated flavonoids, caffeoylquinic acids, STLs, glycolipids and diterpenes (Figures 4.7 and 4.8).

Analysis of the molecular network from the negative ionization mode (Figure 4.7), showed that caffeoylaltraric acids and derivatives represent a chemical class clustering the highest number of nodes, suggesting a high chemical diversity for these molecules in the studied samples. This chemical class was found to be mainly accumulated in samples belonging to group 4, representing nearly adult and adult plants, and confirming the accumulation patterns seen in the heatmap (Figure 4.4). In addition to the mono-, di- and tri-caffeoylaltraric acids previously reported in yacón, other nodes representing putative analogues were also observed, such as the molecule with the *m/z* value of 547.073 (Figure 4.9A), whose fragmentation pattern is characterized by a neutral loss of 162.032 Da representing the putative loss of a caffeoyl moiety and a base peak at 209.029 *m/z*, representing an ionized molecule of altraric acid. However, the isolation and structure elucidation of this molecule is still necessary to unambiguously determine whether it is a novel metabolite. Interestingly, detailed analysis of the nodes and MS/MS data of the entire network suggest that other cyclic and open chain polyols (in addition to altraric and quinic acid) could be also esterified with caffeic acid molecules at different positions. The putative identification of dicaffeoylacetoxy glucuronic acid (Figure 4.9B) further supports this hypothesis.

Analysis of the networks clustering most of the nodes representing glycosylated flavonoids and CGAs (Figure 4.9C and 4.9D, respectively) allowed the

identification of similar trends to the previous analysis by heatmaps (Figure 4.4). Here, glycosylated flavonoids are more commonly found among plants belonging to groups 3 and 4, corresponding to two to six months-old plants, while chlorogenic acids are found especially in groups 2 (1.5 months-old plants) and 4 (4.5 to six months-old plants). Among the glycosylated flavonoids (Figure 4.9C), the clustering of a previously unidentified molecule with a *m/z* value of 639.157 occurred near the nodes corresponding to rutin and quercetin 3-*O*-galactoside (Figure 4.9C), which allowed the proposal of the identity of this molecule as 7-methoxy gossypetin 3-*O*-rutinoside, which represents a putative new report for the genus *Smallanthus*. This molecule showed a fragmentation pattern characterized by a neutral loss of 308.112 Da, suggesting the putative loss of the disaccharide rutinose, and an intense peak at 331.046 *m/z* representing an ionized molecule of 7-methoxy gossypetin.

Visual inspection of the molecular network resulting from the analysis of the positive ionization data (Figure 4.8), allowed the identification of several STLs previously reported in yacón, in addition to some putative glycolipid analogues and diterpenes based on spectral matches with the GNPS library and detailed manual inspection of the MS/MS data. Among STLs, different adducts of enhydrin and uvedalin, among others (Figure 4.10A), were identified in all the studied samples, confirming previous results suggesting that the biosynthesis of this chemical class starts early in yacón development and its presence remains stable through time. In addition to the well-described yacón's STLs identified in the present study, molecular networks suggest the putative presence of other STLs-analogues (Figure 4.10A), based on their similar fragmentation pattern, which requires further investigations as none of them appears to be reported in yacón to date. Analysis of the diterpenes' cluster (Figure 4.10C) allowed the identification of several *in*-source fragments of two known yacón diterpenes, namely smaditerpenic acid F and E, in addition to a putative labdane diterpene based on its similar fragmentation pattern to the labdanes deported in the GNPS library.
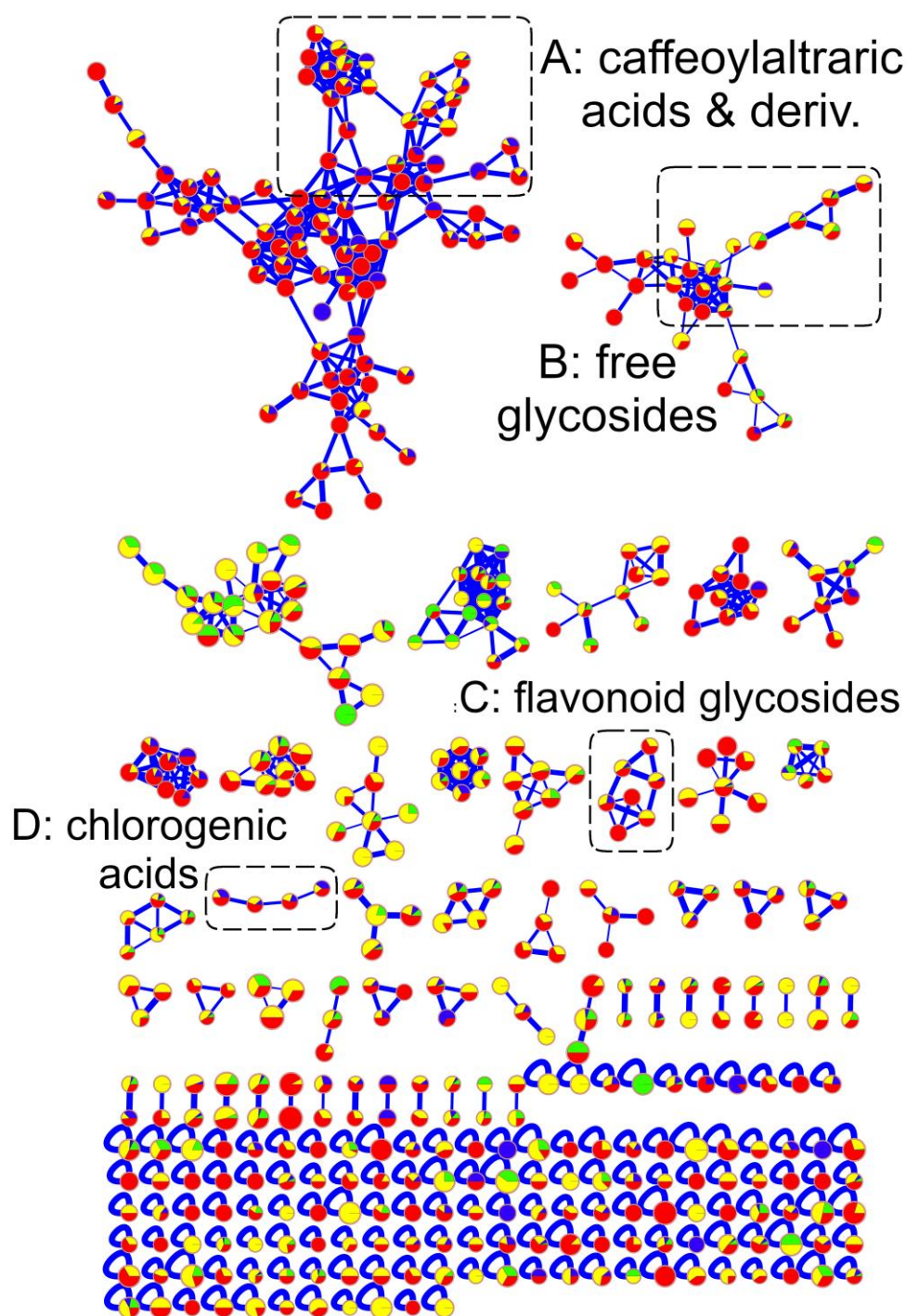
Figure 4.7 Molecular networking based on metabolic fingerprinting by UHPLC-HRMS/MS in negative mode of 36 extracts from *Smallanthus sonchifolius* leaves collected along the development of the plant. Metabolic developmental stage in different colors: 0.5 to one-month-old (green), 1.5-months-old (blue), two to four-months-old (yellow) and 4.5 to six-months-old (red).
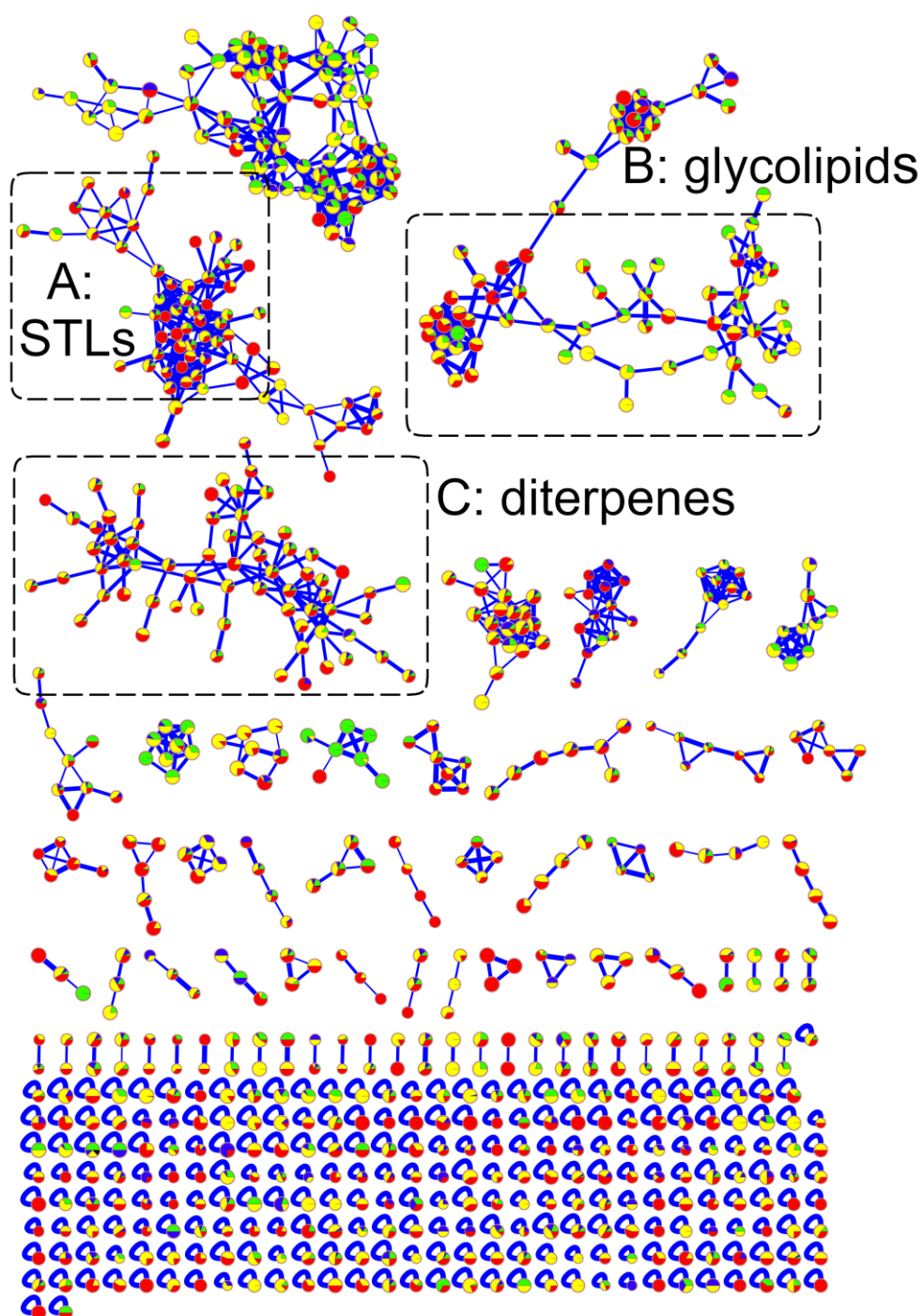
Figure 4.8 Molecular networking based on metabolic fingerprinting by UHPLC-HRMS/MS in positive mode of 36 extracts from Smallanthus sonchifolius leaves collected along the development of the plant. Metabolic developmental stage in different colors: 0.5 to one-month-old (green), 1.5-months-old (blue), two to four-months-old (yellow) and 4.5 to six-months-old (red).
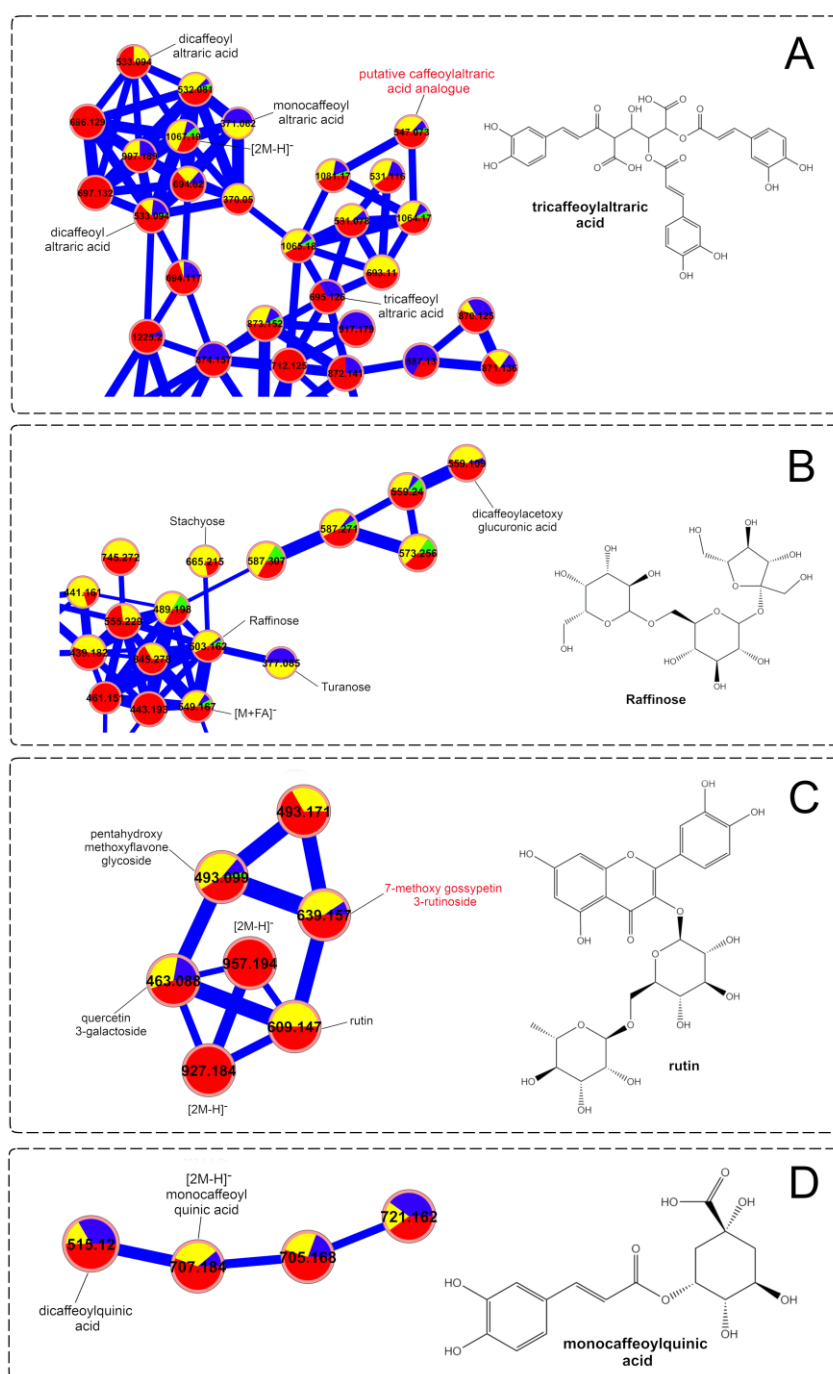
Figure 4.9 Representation of specific network clusters from the negative mode showing (A) caffeoylaltraric acids and derivatives, (B) free glycosides, (C) glycosylated flavonoids and (D) chlorogenic acids. Metabolic developmental stage in different colors: 0.5 to one-month-old (green), 1.5-months-old (blue), two to four-months-old (yellow) and 4.5 to six-months-old (red).
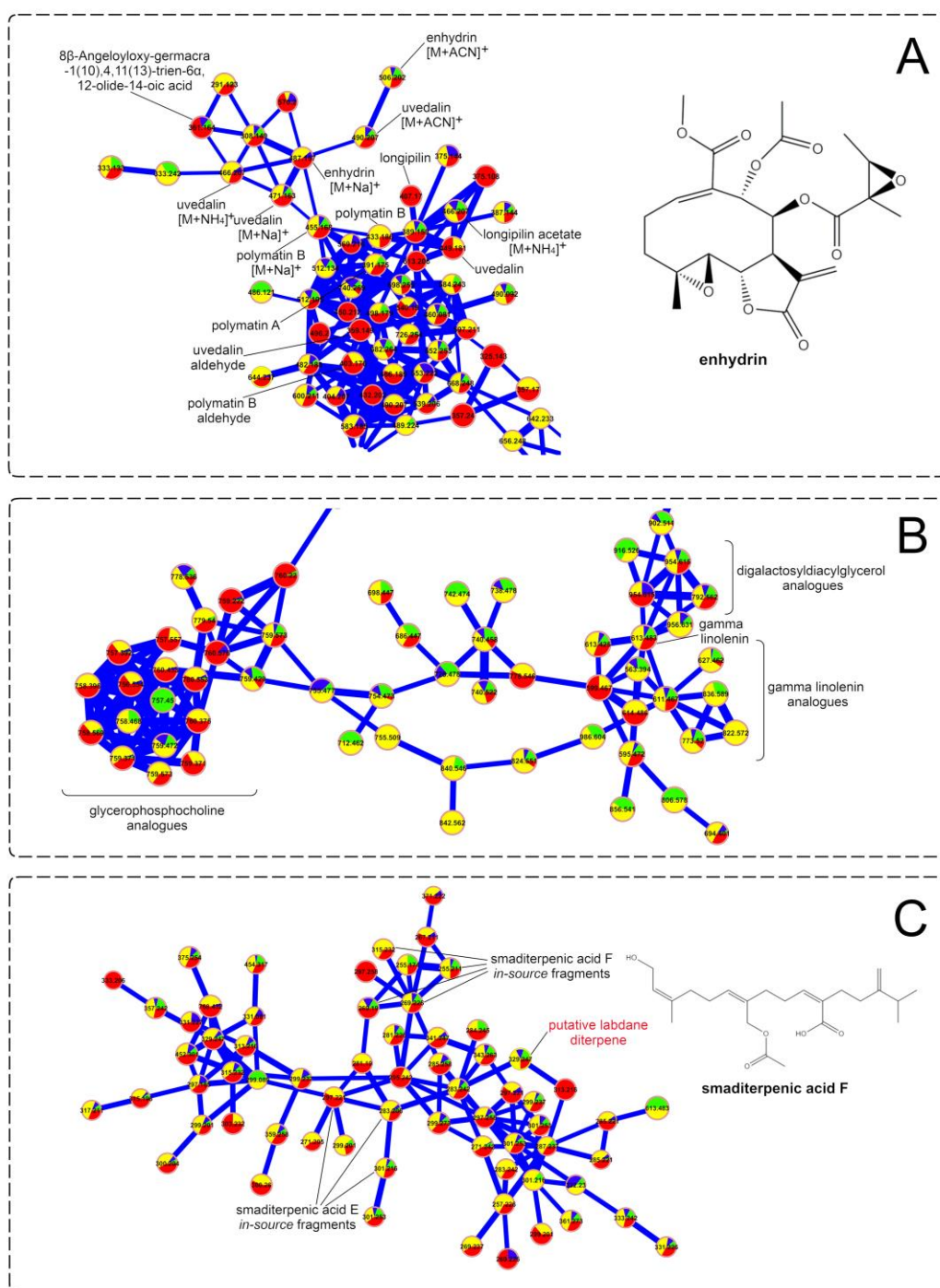
Figure 4.10 Representation of specific network clusters from the positive mode showing (A) sesquiterpene lactones, (B) glycolipids and (C) diterpenes. Metabolic developmental stage in different colors: 0.5 to one-month-old (green), 1.5-months-old (blue), two to four-months-old (yellow) and 4.5 to six-months-old (red).

DISCUSSION

This study is a pioneer in reporting the developmental and environmental regulation of the secondary metabolism in a native Andean plant by analyzing gene expression patterns and metabolic fingerprints in an integrated approach. We found that metabolic differences in leaves of yacón are related to developmental and environmental factors and that different classes of secondary metabolites respond differently to such factors. Our results suggest that high levels of solar radiation are associated with the biosynthesis of flavonoids and CGAs in young yacón leaves (1.5-month-old plants), which is in accordance with the literature (CLÉ et al., 2008; JULKUNEN-TIITTO et al., 2015; MOGLIA et al., 2008; PETRUĽOVÁ; DUČAIOVÁ; REPČÁK, 2014; SAMPAIO; EDRADA-EBEL; DA COSTA, 2016). These two classes of phenolic compounds (flavonoids and CGAs) are well known to act as protectants of internal tissues against UV-induced damage by absorbing or dissipating solar energy and neutralizing radiation-induced free radicals and other reactive oxygen species (GRACE; LOGAN, 2000; JANSEN; GABA; GREENBERG, 1998; ROZEMA et al., 1997).

In yacón, it seems that during the earliest stages of plant development (between 0.5 and one-month-old plants) the biosynthesis and accumulation of flavonoids and CGAs is rather low, but apparently once the young yacón plant is exposed to high levels of solar radiation (in addition to low temperatures) there is a significant upregulation in the expression of CHS and HQT, followed by an increase in the accumulation of flavonoids and CGAs, respectively. This suggests that flavonoids and CGAs might also act as UV protectants in young yacón leaves. Previous studies have shown that an increased accumulation of CGAs and flavonoids is associated with increased UV-protection (CLÉ et al., 2008; JULKUNEN-TIITTO et al., 2015). However, the fact that CHS expression and the accumulation of flavonoids increase over time suggests that the environment might not be the only regulatory factor influencing the biosynthesis and accumulation of this chemical class in yacón leaves and that the plant developmental stage is also an important regulatory factor. Samples collected from adult plants at low levels of solar radiation and low temperatures showed comparable amounts of flavonoids and higher CHS expression when compared to their younger counterparts collected at high levels of solar radiation.

On the other hand, it seems that solar radiation constitutes the main regulatory factor influencing the biosynthesis of CGAs (especially of 5-*O*-caffeoylquinic acid) in yacón leaves. Previous studies have shown that 5-*O*-caffeoylquinic acid protects tomato (*Solanum lycopersicum* L., Solanaceae), artichoke (*Cynara cardunculus* L., Asteraceae) and chamomile (*Matricaria chamomilla* L., Asteraceae) against UV-induced damage (CLÉ et al., 2008; MOGLIA et al., 2008; PETRUĽOVÁ; DUČAIOVÁ; REPČÁK, 2014). In the present study, the highest HQT expression and accumulation of CGAs was observed in 1.5-month-old yacón plants which were collected at the highest level of solar radiation. However, the high amounts of CGAs found in nearly adult and adult individuals are not associated with high levels of HQT expression. The fact that other enzymes, such as hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl transferase (HCT), and alternative pathways are also capable of converting 4-coumaroyl-CoA to CGA represents a possible explanation for this contrasting trend (PAYYAVULA et al., 2015; LALLEMAND et al., 2012). According to Bartley et al. (2016), there are still some uncertainties with respect to the exact nature of the last steps of chlorogenic acids biosynthesis, as a number of possible routes and enzymes exist to convert 4-coumaroyl-CoA to chlorogenic acid (SONNANTE et al., 2010). For example, HQT and HCT constitute similar enzymes with different substrate preferences: HQT prefers quinic acid, while HCT prefers shikimic acid. However, HCT is still capable of producing 3,5-*O*-dicaffeoylquinic acid, indicating a certain amount of flexibility in substrate specificity (LALLEMAND et al., 2012). An alternative route in which caffeoyl-*D*-glucose is converted to CGA through hydroxyl cinnamoyl *D*-glucose:quinate hydroxycinnamoyl transferase (HCGQT) has also been reported (PAYYAVULA et al., 2015). Thus, it is possible that other enzymes or alternative pathways could be acting in the later stages of yacón development, explaining the contrasting trend observed in the accumulation of CGAs and HQT expression in adult yacón plants. However, detailed analyses are still necessary to test these hypotheses, as additional factors such as the transcription factor AN1 (PAYYAVULA et al., 2015), the content of sucrose (PAYYAVULA et al., 2015) or the turnover mechanism between the biosynthesis of CGA and lignin (TAYLOR; ZUCKER, 1966) could also be involved.

Interestingly, a previous study with carrot (*Daucus carota* L.) roots exposed to UVB showed a linear increase in the concentration of chlorogenic acids over a seven-day study period (BARTLEY et al., 2016). At day seven, the average

concentration of chlorogenic acid was six times greater than the concentration found in untreated samples, while the levels of HQT expression in the UVB treatment increased steadily to peak at approximately 9-fold on day 3 and decreased on day 6 to 3.3-fold (BARTLEY et al., 2016). This contrasting trend in the accumulation of chlorogenic acid and HQT expression is similar to the observed trend in our study indicating that a similar, yet unknown, mechanism must be acting on both carrot and yacón.

Contrary to the observed trends for flavonoids and CGAs, the biosynthesis and accumulation of STLs showed an inverse tendency. The highest GAO expression was observed in 15-day-old plants, while the accumulation of STLs showed a rather stable presence (with a slight increase) over time. A possible explanation for this might be attributed to the age of the leaves. The biosynthesis of STLs in Asteraceae is strongly associated with the leaf developmental stage, with higher expression in leaf primordia and young leaves and subsequent accumulation in capitate glandular trichomes (AMREHN et al., 2015; ASCHENBRENNER et al., 2015; BOMBO et al., 2015; GÖPFERT et al., 2005, 2009). Although in the present study we always collected the third leaf (from top to bottom), leaf age was not closely monitored. Thus, it is possible that the highest GAO expression seen in 15-day-old plants is due to the analysis of slightly younger leaves relative to the other collections. In yacón, a slow growth occurs during the early stages of plant development (FERNANDEZ, C.E. VIEHMANNOVA et al., 2007) and in the first collection, leaves could be still in a secretory stage, characterized by high levels of gene expression. In the following collections, the secretory activity could have already concluded and STLs were possibly already stored in capitate glandular trichomes, explaining the inverse tendency observed among GAO expression and STLs accumulation. The relatively constant presence of STLs throughout the development of yacón, and their compartmentalization in capitate glandular trichomes, supports a possible protective role for these molecules in the aerial parts of yacón, as the highly nonspecific toxicity of such compounds necessitates compartmentalization to prevent autotoxicity and to protect the plant against herbivores (PADILLA-GONZALEZ; DOS SANTOS; DA COSTA, 2016).

Given its high adaptability and plastic response to changing environmental conditions, yacón constitutes a good model organism to understand the adaptive roles of secondary metabolites as the chemical interface between the plant and its

environment. Our results suggest that, although plant age seems to strongly influence the metabolism of yacón, it appears that environmental factors such as solar radiation induce a significant upregulation in CHS and HQT expression and in the accumulation of flavonoids and CGAs in young individuals. The accumulation of CGAs can even exceed the extent observed in adult individuals. In yacón, adult plants are metabolically richer and more diverse than their young counterparts are, as demonstrated by the higher number of nodes present in adult plants in the metabolic networks (Figures 4.7 and 4.8) and the higher accumulation of metabolites seen in the heatmap (Figure 4.4). On the other hand, the early stages of plant development are characterized by a rather poor secondary metabolism. Plants up to one-month old showed the lowest levels of metabolite accumulation and gene expression overall (except for GAO), while intermediate levels are observed in two to four-month-old plants, and the highest levels in metabolite accumulation are usually found in adult or nearly adult individuals. Altogether, our results suggest that in young yacón plants the biosynthesis and accumulation of CGAs and flavonoids is mostly regulated by external environmental factors (e.g., solar radiation), and as the plant grows older, the environment has an attenuated effect on the biosynthesis of flavonoids but not on CGAs. Therefore, although CGAs and flavonoids might be both involved in UV protection in yacón leaves, the biosynthesis of each chemical class seems to respond differently relative to the plant developmental stage. On the other hand, the presence of STLs appears to be relatively constant along the development of yacón, characterized by high levels of GAO expression in the early stages of leaf development and subsequent accumulation in capitate glandular trichomes. Therefore, our results demonstrate that the secondary metabolism of yacón is regulated by a complex interplay between both plant development and abiotic environmental conditions.

In conclusion, this study reports the developmental and environmental regulation of the secondary metabolism in a medicinally and economically important Andean crop of which scarce information regarding genomic and molecular traits is available. Our comprehensive study of the secondary metabolism of yacón leaves not only allowed the identification of several compounds previously unreported in this species (Appendix 8) but also improves the understanding of the regulatory factors and possible adaptive roles of different chemical classes of secondary metabolites under changing environmental conditions in Andean taxa. Furthermore, mass

spectrometry-based molecular networking allowed the visualization of the chemical diversity of yacón and the detection of several metabolites already described in this species, in addition to some putative novel analogues. Lastly, the sequences of the CHS, HQT and GAO genes reported herein provide the first information about the putative genes involved in the biosynthesis of flavonoids, CGAs and STLs in the genus *Smallanthus*. This finding might effectively promote the discovery of new genes involved in the secondary metabolism of yacón and its closely related taxa in the future. However, the functional characterization of such genes is still necessary to unambiguously determine their biosynthetic capabilities in model organisms.

CONCLUSIONS

The research detailed in this dissertation integrates metabolomic, biogeographic, taxonomical, evolutionary, chemical, molecular and ecological information for the study of the subtribe Espeletiinae and its sister genus *Smallanthus*. Metabolomic analyses by UHPLC-UV-HRMS of 211 samples and correlation with biogeographic data provided metabolomic evidence for allopatric segregation in Espeletiinae by identifying quantitative metabolic differences at higher taxonomic levels and lower geographic scales compared to the previous reports. Species displayed characteristic metabolic fingerprints related to their country of origin in a global scale, to their páramo massifs in a regional scale and to their páramo complexes in a local scale, revealing inter- and intraspecific metabolic variations. Additionally, this study allowed establishing chemotaxonomic relationships in Espeletiinae based on whole-metabolome comparisons, providing metabolomic evidence to support a putative segregation of the genus *Espeletia* into two different genera based on their characteristic metabolic fingerprints: one comprising Colombian taxa and the other Venezuelan representatives. We also provided the first ancestral states reconstructions of chemical traits in Andean taxa, demonstrating complex scenarios of chemical evolution with some traits representing direct descent from ancestral taxa, other traits resulting from evolutionary innovation, and several traits being gained and lost multiple times implying convergent evolution. Amplification of GAO in species-producing and species-lacking STLs suggested that the capacity of biosynthesizing STLs is putatively present and evolutionary conserved in the whole subtribe Espeletiinae. Lastly, metabolomic analyses of *S. sonchifolius* in combination with gene expression data provided relevant information towards the understanding of the regulatory factors shaping the metabolic fingerprints of Andean taxa, demonstrating that the secondary metabolism of yacón is regulated by a complex interplay between both plant development and abiotic environmental conditions and providing the first information about the genes involved in the biosynthesis of flavonoids, chlorogenic acids and STLs in the genus *Smallanthus*, which might effectively promote the discovery of new genes involved in the secondary metabolism of other related taxa in the future.

REFERENCES

AKULA, R.; RAVISHANKAR, G. A.; RAMAKRISHNA, A.; RAVISHANKAR, G. A. Influence of abiotic stress signals on secondary metabolites in plants. **Plant Signaling & Behavior**, v. 6, n. 11, p. 1720–1731, 2011.

AMBRÓSIO, S. R.; OKI, Y.; HELENO, V. C. G.; CHAVES, J. S.; NASCIMENTO, P. G. B. D.; LICHSTON, J. E.; CONSTANTINO, M. G.; VARANDA, E. M.; DA COSTA, F. B. Constituents of glandular trichomes of *Tithonia diversifolia*: Relationships to herbivory and antifeedant activity. **Phytochemistry**, v. 69, n. 10, p. 2052–2060, 2008.

AMREHN, E.; ASCHENBRENNER, A. K.; HELLER, A.; SPRING, O. Localization of sesquiterpene lactone biosynthesis in cells of capitate glandular trichomes of *Helianthus annuus* (Asteraceae). **Protoplasma**, v. 253, n. 2, p. 447-455, 2015.

ASCHENBRENNER, A. K.; AMREHN, E.; BECHTEL, L.; SPRING, O. Trichome differentiation on leaf primordia of *Helianthus annuus* (Asteraceae): morphology, gene expression and metabolite profile. **Planta**, v. 241, n. 4, p. 837–846, 2015.

BAO, W.; QU, Y.; SHAN, X.; WANG, Y. Screening and validation of housekeeping genes of the root and cotyledon of *Cunninghamia lanceolata* under abiotic stresses by using quantitative real-time PCR. **International Journal of Molecular Sciences**, v. 17, n. 8, p. 1198, 2016.

BARTHLOTT, W.; MUTKE, J.; RAFIQPOOR, M. D.; KIER, G.; KREFT, H. Global centres of vascular plant diversity. **Nova Acta Leopoldina**, v. 92, n. 342, p. 61–83, 2005.

BARTLEY, G. E.; AVENA-BUSTILLOS, R. J.; DU, W-X.; HIDALGO, M.; CAIN, B.; BREKSA, A. P. Transcriptional regulation of chlorogenic acid biosynthesis in carrot root slices exposed to UV-B light. **Plant Gene**, v. 7, p. 1-10, 2016.

BERRY, P. E.; CALVO, R. N. Wind pollination, self-incompatibility and altitudinal shifts in pollination systems in the high andean genus *Espeletia* (Asteraceae). **American Journal of Botany**, v. 76, p. 1602–1614, 1989.

BOMBO, A. B.; APPEZZATO-DA-GLÓRIA, B.; ASCHENBRENNER, A. K.; SPRING, O. Capitate glandular trichomes in *Aldama discolor* (Heliantheae - Asteraceae): morphology, metabolite profile and sesquiterpenes biosynthesis. **Plant Biology**, v. 18, n. 3, p. 455–462, 2015.

BRUNO, M.; ROSSELLI, S.; PIBIRI, I.; PIOZZI, F.; BOND, M. L.; SIMMONDS, M. S. J. Semisynthetic derivatives of *ent*-kauranes and their antifeedant activity. **Phytochemistry**, v. 58, n. 3, p. 463–474, 2001.

CALABRIA, L. M.; EMERENCIANO, V. P.; SCOTTI, T.; MABRY, T. J. Secondary chemistry of Compositae. In: FUNK, V. A.; SUSANNA, A.; STUESSY, T. F.; BAYER, R. J. (Eds.). **Systematics, Evolution, and Biogeography of Compositae**. Vienna: International Association for Plant Taxonomy, 2009. p. 73–88.

CARNAT, A.; HEITZ, A.; FRAISSE, D.; CARNAT, A. P.; LAMAISON, J. L. Major dicaffeoylquinic acids from *Artemisia vulgaris* L. **Fitoterapia**, v. 71, n. 5, p. 587–589, 2000.

CHOI, Y. H.; KIM, H. K.; LINTHORST, H. J. M.; HOLLANDER, J. G.; LEFEBER, A. W. M.; ERKELENS, C.; NUZILLARD, J. M.; VERPOORTE, R. NMR Metabolomics to revisit the tobacco mosaic virus infection in *Nicotiana tabacum* leaves. **Journal of Natural Products**, v. 69, n. 5, p. 742–748, 2006.

CLÉ, C.; HILL, L. M.; NIGGEWEG, R.; MARTIN, C. R.; GUISEZ, Y.; PRINSEN, E.; JANSEN, M. A. K. Modulation of chlorogenic acid biosynthesis in *Solanum lycopersicum*; consequences for phenolic accumulation and UV-tolerance. **Phytochemistry**, v. 69, n. 11, p. 2149–2156, 2008.

COLL-ARÁOZ, M. B.; MERCADO, M. I. M. I.; GRAU, A.; CATALÁN, C. A. N. *Ent*-kaurane derivatives from the root cortex of yacon and other three *Smallanthus* species (Heliantheae, Asteraceae). **Biochemical Systematics and Ecology**, v. 38, n. 5, p. 1042–1048, 2010.

CONTRERAS-ORTIZ, N.; ATCHISON, G. W.; HUGHES, C. E.; MADRIŇÁN, S. Convergent evolution of high elevation plant growth forms and geographically structured variation in Andean *Lupinus* (Fabaceae). **Botanical Journal of the Linnean Society**, v. 187, n. 1, p. 118–136, 2018.

COOPER-DRIVER, G. A.; LEQUESNE, P. W. Diterpenes as insect antifeedants and growth inhibitors - role in *Solidago* species. In: WALLER, G. R. (Ed.). **Allelochemicals: role in agriculture and forestry**. Washington: American Chemical Society Symposium 330, American Chemical Society, 1987. p. 535–550.

CRONQUIST, A. The Compositae Revisited. **Brittonia**, v. 29, n. 2, p. 137, 1977.

CUATRECASAS, J. A new subtribe in the Heliantheae (Compositae): Espeletiinae.

Phytologia, v. 35, p. 43–61, 1976.

CUATRECASAS, J. Speciation and radiation of the Espeletiinae in the Andes. In: VUILLEUMIER, F.; MONASTERIO, M. (Eds.). **High Altitude Tropical Biogeography**. New York, USA: Oxford University Press, 1986. p. 267–303.

CUATRECASAS, J. **A systematic study of the subtribe Espeletiinae: Heliantheae, Asteraceae**. New York, USA: The New York Botanical Garden Press, 2013.

DANINO, O.; GOTTLIEB, H. E.; GROSSMAN, S.; BERGMAN, M. Antioxidant activity of 1,3-dicaffeoylquinic acid isolated from *Inula viscosa*. **Food Research International**, v. 42, n. 9, p. 1273–1280, 2009.

DE VOS, R. C.; MOCO, S.; LOMMEN, A.; KEURENTJES, J. J. B.; BINO, R. J.; HALL, R. D. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. **Nature Protocols**, v. 2, n. 4, p. 778–791, 2007.

DELGADO, G. T. C.; TAMASHIRO, W. M. DA S. C.; MAROSTICA, M. R. J.; PASTORE, G. M.; JUNIOR, M. R. M.; PASTORE, G. M. Yacon (*Smallanthus sonchifolius*): A functional food. **Plant Foods for Human Nutrition**, v. 68, n. 3, p. 222–228, 2013.

DIAZGRANADOS, M. A nomenclator for the frailejones (Espeletiinae Cuatrec., Asteraceae). **PhytoKeys**, v. 16, p. 1–52, 2012.

DIAZGRANADOS, M.; BARBER, J. C. Geography shapes the phylogeny of frailejones (Espeletiinae Cuatrec., Asteraceae): a remarkable example of recent rapid radiation in sky islands. **PeerJ**, v. 5, p. e2968, 2017.

EHRLICH, P. R.; RAVEN, P. H. Butterflies and plants: A study in coevolution. **Evolution**, v. 18, n. 4, p. 586, 1964.

ERNST, M.; SILVA, D. B.; SILVA, R. R.; VÊNCIO, R. Z. N.; LOPES, N. P. Mass spectrometry in plant metabolomics strategies: from analytical platforms to data acquisition and processing. **Natural Product Reports**, v. 31, n. 6, p. 784-806, 2014.

FAGUA, J. C.; GONZALEZ, V. H. Growth rates, reproductive phenology, and pollination ecology of *Espeletia grandiflora* (Asteraceae), a giant Andean caulescent rosette. **Plant Biology**, v. 9, n. 1, p. 127–135, 2007.

FERNANDEZ, C. E. VIEHMANNOVA, I.; BECHYNĚ, M.; LACHMAN, J.; MILELLA, L.; MARTELLI, G. The cultivation and phenological growth stages of yacon

[*Smallanthus sonchifolius* (Poepp. et Endl.) H. Robinson. **Agricultura Tropica et Subtropica**, v. 40, n. 3, p. 71–77, 2007.

FIEHN, O. Metabolomics – the link between genotypes and phenotypes. **Plant Molecular Biology**, v. 48, n. 1-2, p. 155–171, 2002.

GALLON, M. E.; MONGE, M.; CASOTI, R.; DA COSTA, F. B.; SEMIR, J.; GOBBO-NETO, L. Metabolomic analysis applied to chemosystematics and evolution of megadiverse Brazilian Vernonieae (Asteraceae). **Phytochemistry**, v. 150, p. 93–105, jun. 2018.

GARCÍA, N., CALDERÓN, E., GALEANO, G. Frailejones. In: CALDERÓN, E., GALEANO, G., GARCÍA, N. (Ed.). **Libro Rojo de Plantas de Colombia**. Bogotá D.C., Colombia: IAVH, 2005. p. 225–385.

GENTA, S. B.; CABRERA, W. M.; MERCADO, M. I.; GRAU, A.; CATALÁN, C. A.; SÁNCHEZ, S. S. Hypoglycemic activity of leaf organic extracts from *Smallanthus sonchifolius*: Constituents of the most active fractions. **Chemico-Biological Interactions**, v. 185, n. 2, p. 143–152, 2010.

GENTRY, A. H. Neotropical floristic diversity: Phytogeographical connections between Central and South America, pleistocene climatic fluctuations, or an accident of the Andean orogeny? **Annals of the Missouri Botanical Garden**, v. 69, n. 3, p. 557-593, 1982.

GIVNISH, T. J. Adaptive radiations and molecular systematics: issues and approaches. In: GIVNISH, T. J.; SYTSMA, K. J. (Eds.). **Molecular Evolution and Adaptive Radiation**. Cambrige: Cambridge University Press, 1997. p. 1–54.

GÓMEZ-GUTIÉRREZ, M. C.; PENNINGTON, R. T.; NEAVES, L. E.; MILNE, R. I.; MADRIÑÁN, S.; RICHARDSON, J. E. Genetic diversity in the Andes: variation within and between the South American species of *Oreobolus* R. Br. (Cyperaceae). **Alpine Botany**, v. 127, n. 2, p. 155–170, 2017.

GÖPFERT, J. C.; HEIL, N.; CONRAD, J.; SPRING, O. Cytological development and sesquiterpene lactone secretion in capitate glandular trichomes of sunflower. **Plant Biology**, v. 7, n. 2, p. 148–155, 2005.

GÖPFERT, J. C.; MACNEVIN, G.; RO, D. K.; SPRING, O. Identification, functional characterization and developmental regulation of sesquiterpene synthases from sunflower capitate glandular trichomes. **BMC Plant Biology**, v. 9, n. 1, p. 86, 2009.

GRACE, S. C.; LOGAN, B. A. Energy dissipation and radical scavenging by the plant phenylpropanoid pathway. **Philos. Trans. R. Soc. Lond. B. Biol. Sci.**, v. 355, p. 1499–1510, 2000.

GRAU, A.; REA, J. Yacon. *Smallanthus sonchifolius* (Poepp. & Endl.) H. Robinson. In: HERMANN, M.; HELLER, J. (Eds.). **Andean roots and tubers: Ahipa, arracacha, maca and yacon. Promoting the conservation and use of underutilized and neglected crops.** 21. ed. Rome, Italy: Institute of Plant Genetics and Crop Plant Research, Gatersleben/ International Plant Genetic Resources Institute., 1997. p. 255.

HALL, T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucleic Acids Symposium Series**, v. 41, p. 95–98, 1999.

HARBORNE, J. **Introduction to Ecological Biochemistry**. Third ed. London: Academic Press, 1989.

HENDRY, G. A. F. Fructan and the ecology and evolution of the Compositae. In: CALIGARI, P. D. S.; HIND, D. J. N. (Eds.). **Compositae: Biology & Utilization. Proceedings of the International Compositae Conference, Kew, 1994, vol. 2**. London: Royal Botanic Gardens, Kew, 1996. p. 121–128.

HOOGHIEMSTRA, H.; CLEEF, A. M. Development of the high Andean Páramo flora and vegetation. In: VUILLEUMIER, F.; MONASTERIO, M. (Eds.). **High Altitude Tropical Biogeography**. Oxford: Oxford University Press, 1986. p. 153–201.

HOOGHIEMSTRA, H.; RAN, E. T. H. Late and middle pleistocene climatic change and forest development in Colombia: pollen record Funza II (2-158 m core interval). **Palaeogeography, Palaeoclimatology, Palaeoecology**, v. 109, n. 2-4, p. 211–246, 1994.

HOOGHIEMSTRA, H.; VAN DER HAMMEN, T. Quaternary Ice-Age dynamics in the Colombian Andes: developing an understanding of our legacy. **Philosophical Transactions of the Royal Society of London**, v. 359, n. 1442, p. 173–180, 2004.

HOOGHIEMSTRA, H.; WIJNINGA, V. M.; CLEEF, A. M. The paleobotanical record of Colombia: Implications for biogeography and biodiversity. **Annals of the Missouri Botanical Garden**, v. 93, n. 2, p. 297–325, 2006.

HUGHES, C.; EASTWOOD, R. Island radiation on a continental scale: Exceptional

rates of plant diversification after uplift of the Andes. **Proceedings of the National Academy of Sciences**, v. 103, n. 27, p. 10334–10339, 2006.

JAKOBSEN, H.; OLSEN, C. Influence of climatic factors on emission of flower volatiles in situ. **Planta**, v. 192, n. 3, p. 365-371, 1994.

JANSEN, M. A.; GABA, V.; GREENBERG, B. M. Higher plants and UV-B radiation: balancing damage, repair and acclimation. **Trends in Plant Science**, v. 3, n. 4, p. 131–135, 1998.

JEFFREY, C. Compositae: Introduction with key to tribes. In: KADEREIT, J. W.; JEFFREY, C. (Eds.). **Families and Genera of Vascular Plants, Flowering Plants, Eudicots, Asterales, vol. VIII.** Berlin: Springer-Verlag, 2007. p. 61–87.

JULKUNEN-TIITTO, R.; NENADIS, N.; NEUGART, S.; ROBSON, M.; AGATI, G.; VEPSALAINEN, J.; ZIPOLI, G.; NYBAKKEN, L.; WINKLER, B.; JANSEN, M. A. K. Assessing the response of plant flavonoids to UV radiation: an overview of appropriate techniques. **Phytochemistry Reviews**, v. 14, n. 2, p. 273-297, 2015.

KIM, J. Y.; CHO, J. Y.; MA, Y. K.; PARK, K. Y.; LEE, S. H.; HAM, K. S.; LEE, H. J.; PARK, K. H.; MOON, J. H. Dicaffeoylquinic acid derivatives and flavonoid glucosides from glasswort (*Salicornia herbacea* L.) and their antioxidative activity. **Food Chemistry**, v. 125, n. 1, p. 55–62, 2011.

KIM, W.; PEEVER, T. L.; PARK, J.-J.; PARK, C.-M.; GANG, D. R.; XIAN, M.; DAVIDSON, J. A.; INFANTINO, A.; KAISER, W. J.; CHEN, W. Use of metabolomics for the chemotaxonomy of legume-associated *Ascochyta* and allied genera. **Scientific Reports**, v. 6, n. 1, p. 20192, 2016.

KOLÁŘ, F.; DUŠKOVÁ, E.; SKLENÁŘ, P. Niche shifts and range expansions along cordilleras drove diversification in a high-elevation endemic plant genus in the tropical Andes. **Molecular Ecology**, v. 25, n. 18, p. 4593–4610, 2016.

KRASENSKY, J.; JONAK, C. Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. **Journal of Experimental Botany**, v. 63, n. 4, p. 1593–1608, 2012.

KRASTANOV, A. Metabolomics – the state of art. **Biotechnology & Biotechnological Equipment**, v. 24, n. 1, p. 1537–1543, 2010.

KRENN, L.; MIRON, A.; PEMP, E.; PETR, U.; KOPP, B. Flavonoids from *Achillea nobilis* L. **Zeitschrift für Naturforschung C**, v. 58, n. 1–2, p. 11-16, 2003.

LACHMAN, J.; FERNÁNDEZ, E. C.; ORSÁK, M. Yacon [*Smallanthus sonchifolia* (Poepp. et Endl.) H. Robinson] chemical composition and use - a review. **Plant, Soil and Environment**, v. 49, n. 6, p. 283–290, 2003.

LALLEMAND, L. A.; ZUBIETA, C.; LEE, S. G.; WANG, Y.; ACAJJAOUI, S.; TIMMINS, J.; MCSWEENEY, S.; JEZ, J. M.; MCCARTHY, J. G.; MCCARTHY, A. A. A structural basis for the biosynthesis of the major chlorogenic acids found in coffee. **Plant Physiology**, v. 160, p. 249–260, 2012.

LIVINGSTON, D. P.; HINCHA, D. K.; HEYER, A. G. Fructan and its relationship to abiotic stress tolerance in plants. **Cellular and Molecular Life Sciences**, v. 66, n. 13, p. 2007–2023, 2009.

LU, Y. H.; YUAN, M.; GAO, X. W.; KANG, T. H.; ZHAN, S.; WAN, H.; LI, J. H. Identification and validation of reference genes for gene expression analysis using quantitative PCR in *Spodoptera litura* (Lepidoptera: Noctuidae). **PloS one**, v. 8, n. 7, p. e68059, 2013.

LUEBERT, F.; WEIGEND, M. Phylogenetic insights into Andean plant diversification. **Frontiers in Ecology and Evolution**, v. 2, article 27, 2014.

LUTEYN, J. L. **Páramos: a checklist of plant diversity, geographical distribution and botanical literature**. New York, USA: New York Botanical Garden Press, 1999.

MACÍAS, F. A.; FERNÁNDEZ, A.; VARELA, R. M.; MOLINILLO, J. M. G.; TORRES, A.; ALVES, P. L. C. A. Sesquiterpene lactones as allelochemicals. **Journal of Natural Products**, v. 69, n. 5, p. 795–800, 2006.

MACÍAS, F. A.; TORRES, A.; MOLINLLO, J. G.; VARELA, R. M.; CASTELLANO, D. Potential allelopathic sesquiterpene lactones from sunflower leaves. **Phytochemistry**, v. 43, n. 6, p. 1205–1215, 1996.

MADDISON, W. P.; MADDISON, D. R. **Mesquite: a modular system for evolutionary analysis**, v. 62, p. 1103-118, 2008.

MADRIÑÁN, S.; CORTÉS, A. J.; RICHARDSON, J. E. Páramo is the world's fastest evolving and coolest biodiversity hotspot. **Frontiers in genetics**, v. 4, article 192, 2013.

MARTUCCI, M. E. P.; DE VOS, R. C. H.; CAROLLO, C. A.; GOBBO-NETO, L. Metabolomics as a potential chemotaxonomical tool: application in the genus *Vernonia* schreb. **PloS one**, v. 9, n. 4, p. e93149, 2014.

MARTUCCI, M. E. P.; LOEUILLE, B.; PIRANI, J. R.; GOBBO-NETO, L.

Comprehensive untargeted metabolomics of Lychnnophorinae subtribe (Asteraceae: Vernonieae) in a phylogenetic context. **PloS one**, v. 13, n. 1, p. e0190104, 2018.

MOGLIA, A.; LANTERI, S.; COMINO, C.; ACQUADRO, A.; DE VOS, R.; BEEKWILDER, J. Stress-Induced biosynthesis of dicaffeoylquinic acids in globe artichoke. **Journal of Agricultural and Food Chemistry**, v. 56, n. 18, p. 8641–8649, 2008.

MONASTERIO, M.; SARMIENTO, L. Adaptive radiation of *Espeletia* in the cold andean tropics. **Trends in Ecology & Evolution**, v. 6, n. 12, p. 387–391, 1991.

MYERS, N.; MITTERMEIER, R. A.; MITTERMEIER, C. G.; DA FONSECA, G. A. B.; KENT, J. Biodiversity hotspots for conservation priorities. **Nature**, v. 403, n. 6772, p. 853–858, 2000.

NEVADO, B.; ATCHISON, G. W.; HUGHES, C. E.; FILATOV, D. A. Widespread adaptive evolution during repeated evolutionary radiations in New World lupins. **Nature Communications**, v. 7, p. 12384, 2016.

NEVADO, B.; CONTRERAS-ORTIZ, N.; HUGHES, C.; FILATOV, D. A. Pleistocene glacial cycles drive isolation, gene flow and speciation in the high-elevation Andes. **New Phytologist**, v. 219, n. 2, p. 779–793, 2018.

NGUYEN, D. T. **Biochemical and evolutionary studies of sesquiterpene lactone metabolism in the sunflower (Asteraceae) family**. [PhD thesis.] University of Calgary, 2013.

NGUYEN, D. T.; GÖPFERT, J. C.; IKEZAWA, N.; MACNEVIN, G.; KATHIRESAN, M.; CONRAD, J.; SPRING, O.; RO, D.-K. Biochemical conservation and evolution of germacrene A oxidase in Asteraceae. **The Journal of biological chemistry**, v. 285, n. 22, p. 16588–98, 2010.

NGUYEN, D. T.; FARALDOS, J. A.; VARDAKOU, M.; SALMON, M.; O'MAILLE, P. E.; RO, D.-K. Discovery of germacrene A synthases in *Barnadesia spinosa*: The first committed step in sesquiterpene lactone biosynthesis in the basal member of the Asteraceae. **Biochemical and Biophysical Research Communications**, v. 479, n. 4, p. 622–627, 2016.

NIGGEWEG, R.; MICHAEL, A. J.; MARTIN, C. Engineering plants with increased levels of the antioxidant chlorogenic acid. **Nature Biotechnology**, v. 22, n. 6, p. 746–754, 2004.

OLIVER, S. Systematic functional analysis of the yeast genome. **Trends in Biotechnology**, v. 16, n. 9, p. 373–378, 1998.

PADILLA-GONZÁLEZ, G. F. **Impressão digital metabólica do gênero Espeletia e sua correlação com dados filogenéticos e biogeográficos**. [Master thesis] Universidade de Sao Paulo, 2014.

PADILLA-GONZÁLEZ, G. F.; DIAZGRANADOS, M.; CCANA-CCAPATINTA, G.; CASOTI, R.; DA COSTA, F. B. Caffeic acid derivatives and further compounds from *Espeletia barclayana* Cuatrec. (Asteraceae, Espeletiinae). **Biochemical Systematics and Ecology**, v. 70, p. 291–293, 2017.

PADILLA-GONZÁLEZ, G. F.; DIAZGRANADOS, M.; DA COSTA, F. B. Biogeography shaped the metabolome of the genus *Espeletia*: a phytochemical perspective on an Andean adaptive radiation. **Scientific Reports**, v. 7, article 8835, 2017.

PADILLA-GONZÁLEZ, G. F.; DIAZGRANADOS, M.; OLIVEIRA, T. B.; CHAGAS-PAULA, D. A.; DA COSTA, F. B. Chemistry of the subtribe Espeletiinae (Asteraceae) and its correlation with phylogenetic data: an in silico chemosystematic approach. **Botanical Journal of the Linnean Society**, v. 186, n. 1, p. 18–46, 2018.

PADILLA-GONZALEZ, G. F.; DOS SANTOS, F. A.; DA COSTA, F. B. Sesquiterpene lactones: more than protective plant compounds with high toxicity. **Critical Reviews in Plant Sciences**, v. 35, n. 1, p. 18–37, 2016.

PANERO, J. L.; CROZIER, B. S. Macroevolutionary dynamics in the early diversification of Asteraceae. **Molecular Phylogenetics and Evolution**, v. 99, p. 116–132, 2016.

PAYYAVULA, R. S.; SHAKYA, R.; SENGODA, V. G.; MUNYANEZA, J. E.; SWAMY, P.; NAVARRE, D. A. Synthesis and regulation of chlorogenic acid in potato: Rerouting phenylpropanoid flux in HQT-silenced lines. **Plant Biotechnology Journal**, v. 13, p. 551–564, 2015.

PEIFFER, M.; TOOKER, J. F.; LUTHE, D. S.; FELTON, G. W. Plants on early alert: glandular trichomes as sensors for insect herbivores. **New Phytologist**, v. 184, p. 644–656, 2009.

PÉREZ-ESCOBAR, O. A.; CHOMICKI, G.; CONDAMINE, F. L.; KARREMANS, A. P.; BOGARÍN, D.; MATZKE, N. J.; SILVESTRO, D.; ANTONELLI, A. Recent origin and rapid speciation of Neotropical orchids in the world's richest plant biodiversity hotspot. **New Phytologist**, v. 215, n. 2, p. 891–905, 2017.

PETRUĽOVÁ, V.; DUČAIOVÁ, Z.; REPČÁK, M. Short-term UV-B dose stimulates production of protective metabolites in *Matricaria chamomilla* leaves. **Photochemistry and Photobiology**, v. 90, n. 5, p. 1061–1068, 2014.

PICMAN, A. K. Biological activities of sesquiterpene lactones. **Biochemical Systematics and Ecology**, v. 14, n. 3, p. 255–281, 1986.

PIOZZI, F.; PASSANNA, S.; PATERNOS, M. P.; SPRIO, V. Kauranoid diterpenes in *Espeletia grandiflora*. **Phytochemistry**, v. 10, n. 5, p. 1164–1166, 1971.

PIOZZI, F.; PASSANNANTI, S.; MARINO, M. L.; SPRIO, V. Structure of grandiflorenic acid. **Canadian Journal of Chemistry**, v. 50, p. 109–112, 1972.

PIOZZI, F.; SPRIO, V.; PASSANNANTI, S.; MONDELLI, R. Structure of grandiflorolic acid. **Gazzetta Chimica Italiana**, v. 98, p. 907–910, 1968.

PLUSKAL, T.; CASTILLO, S.; VILLAR-BRIONES, A.; ORESIC, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. **BMC bioinformatics**, v. 11, article 395, 2010.

POUCHON, C.; FERNÁNDEZ, A.; NASSAR, J. M.; BOYER, F.; AUBERT, S.; LAVERGNE, S.; MAVÁREZ, J. Phylogenomic analysis of the explosive adaptive radiation of the *Espeletia* complex (Asteraceae) in the tropical Andes. **Systematic Biology**, v. 0, n. 0, p. 1-20, 2018.

RANGEL-CH, O. **La región de vida paramuna**. Bogotá, Colombia: Instituto de Ciencias Naturales - Instituto Alexander von Humboldt, 2000.

RAUSCHER, J. T. Molecular phylogenetics of the *Espeletia* complex (Asteraceae): Evidence from nrDNA ITS sequences on the closest relatives of an Andean adaptive radiation. **American Journal of Botany**, v. 89, n. 7, p. 1074–1084, 2002.

RISTAINO, J. B.; GROVES, C. T.; PARRA, G. R. PCR amplification of the Irish potato famine pathogen from historic specimens. **Nature**, v. 411, n. 6838, p. 695–697, 2001.

RODRIGUEZ, E.; TOWERS, G. H. N.; MITCHELL, J. C. Biological activities of sesquiterpene lactones. **Phytochemistry**, v. 15, n. 11, p. 1573–1580, 1976.

ROSSELLÓ-MORA, R.; LUCIO, M.; PEÑA, A.; BRITO-ECHEVERRÍA, J.; LÓPEZ-LÓPEZ, A.; VALENS-VADELL, M.; FROMMBERGER, M.; ANTÓN, J.; SCHMITT-KOPPLIN, P. Metabolic evidence for biogeographic isolation of the extremophilic bacterium *Salinibacter ruber*. **The ISME Journal**, v. 2, n. 3, p.

242–253, 2008.

ROZEMA, J.; VAN DE STAAIJ, J.; BJÖRN, L. O.; CALDWELL, M. UV-B as an environmental factor in plant life: stress and regulation. **Trends in Ecology & Evolution**, v. 12, n. 1, p. 22–28, 1997.

SAMPAIO, B. L.; EDRADA-EBEL, R.; DA COSTA, F. B. Effect of the environment on the secondary metabolic profile of *Tithonia diversifolia*: a model for environmental metabolomics of plants. **Scientific Reports**, v. 6, article 29265, 2016.

SILVA, D. B.; TURATTI, I. C. C.; GOUVEIA, D. R.; ERNST, M.; TEIXEIRA, S. P.; LOPES, N. P. Mass spectrometry of flavonoid vicenin-2, based sunlight barriers in *Lychnophora* species. **Scientific Reports**, v. 4, article 4309, 2014.

SMITH, A. P. **Growth and Population Dynamics of *Espeletia* (Compositae) in the Venezuelan Andes**. 48. ed. Washington: Srnithsonian Contributions to Botany, 1981.

SONNANTE, G.; D'AMORE, R.; BLANCO, E.; PIERRI, C.L.; De PALMA, M.; LUO, J.; TUCCI, M.; MARTIN, C. Novel hydroxycinnamoyl-coenzyme A quinate transferase genes from artichoke are involved in the synthesis of chlorogenic acid. **Plant Physiology**, v. 153, p. 1224–1238, 2010.

SUMNER, L. W.; AMBERG, A.; BARRETT, D.; BEALE, M. H.; BEGER, R.; DAYKIN, C. A.; FAN, T. W.-M.; FIEHN, O.; GOODACRE, R.; GRIFFIN, J. L.; HANKEMEIER, T.; HARDY, N.; HARNLY, J.; HIGASHI, R.; KOPKA, J.; LANE, A. N.; LINDON, J. C.; MARRIOTT, P.; NICHOLLS, A. W.; REILY, M. D.; THADEN, J. J.; VIANT, M. R. Proposed minimum reporting standards for chemical analysis. **Metabolomics**, v. 3, n. 3, p. 211–221, 2007.

SUMNER, L. W.; MENDES, P.; DIXON, R. A. Plant metabolomics: large-scale phytochemistry in the functional genomics era. **Phytochemistry**, v. 62, n. 6, p. 817–836, 2003.

SUZUKI, R.; SHIMODAIRA, H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. **Bioinformatics (Oxford, England)**, v. 22, n. 12, p. 1540–1542, 2006.

TANG, X. L.; WANG, H. Y.; SHAO, C. Y.; SHAO, H. B. Reference gene selection for qPCR normalization of *Kosteletzkya virginica* under salt stress. **BioMedical Research International**, v. 2015, article 823806, 2015.

TRIBA, M. N.; LE MOYEC, L.; AMATHIEU, R.; GOOSSENS, C.; BOUCHEMAL, N.;

NAHON, P.; RUTLEDGE, D. N.; SAVARIN, P. PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters. **Molecular BioSystems**, v. 11, n. 1, p. 13–19, 2015.

TURNER, B. L. Fossil history and geography. In: HEYWOOD, V. H.; HARBORNE, J. B.; TURNER, B. L. (Eds.). **The Biology and Chemistry of the Compositae vol. 1**. London: Academic Press, 1977. p. 21–40.

VÁSQUEZ, D. L. A.; BALSLEV, H.; HANSEN, M. M.; SKLENÁŘ, P.; ROMOLEROUX, K. Low genetic variation and high differentiation across sky island populations of *Lupinus alopecuroides* (Fabaceae) in the northern Andes. **Alpine Botany**, v. 126, n. 2, p. 135–142, 2016.

VILLAS-BÔAS, S. G.; RASMUSSEN, S.; LANE, G. A. Metabolomics or metabolite profiles? **Trends in Biotechnology**, v. 23, n. 8, p. 385–386, 2005.

WANG, J.; GAO, H.; ZHAO, J.; WANG, Q.; ZHOU, L.; HAN, J.; YU, Z.; YANG, F. Preparative separation of phenolic compounds from *Halimodendron halodendron* by high-speed counter-current chromatography. **Molecules**, v. 15, n. 9, p. 5998–6007, 2010.

WANG, M.; CARVER, J. J.; PHELAN, V. V.; SANCHEZ, L. M.; GARG, N.; PENG, Y.; NGUYEN, D. D.; WATROUS, J.; KAPONO, C. A.; LUZZATTO-KNAAN, T.; PORTO, C.; BOUSLIMANI, A.; MELNIK, A. V.; MEEHAN, M. J.; LIU, W-T.; CRÜSEMANN, M.; BOUDREAU, P. D.; ESQUENAZI, E.; SANDOVAL-CALDERÓN, M.; KERSTEN, R. D.; PACE, L. A.; QUINN, R. A.; DUNCAN, K. R.; HSU, C-C.; FLOROS, D. J.; GAVILAN, R. G.; KLEIGREWE, K.; NORTHEN, T.; DUTTON, R. J.; PARROT, D.; CARLSON, E. E.; AIGLE, B.; MICHELSEN, C. F.; JELSBAK, L.; SOHLENKAMP, C.; PEVZNER, P.; EDLUND, A.; MCLEAN, J.; PIEL, J.; MURPHY, B. T.; GERWICK, L.; LIAW, C-C.; YANG, Y-L.; HUMPF, H-U.; MAANSSON, M.; KEYZERS, R. A.; SIMS, A. C.; JOHNSON, A. R.; SIDEBOTTOM, A. M.; SEDIO, B. E.; KLITGAARD, A.; LARSON, C. B.; BOYA P, C. A.; TORRES-MENDOZA, D.; GONZALEZ, D. J.; SILVA, D. B.; MARQUES, L. M.; DEMARQUE, D. P.; POCIUTE, E.; O'NEILL, E. C.; BRIAND, E.; HELFRICH, E. J. N.; GRANATOSKY, E. A.; GLUKHOV, E.; RYFFEL, F.; HOUSON, H.; MOHIMANI, H.; KHARBUSH, J. J.; ZENG, Y.; VORHOLT, J. A.; KURITA, K. L.; CHARUSANTI, P.; MCPHAIL, K. L.; NIELSEN, K. F.; VUONG, L.; ELFEKI, M.; TRAXLER, M. F.; ENGENE, N.;

KOYAMA, N.; VINING, O. B.; BARIC, R.; SILVA, R. R.; MASCUCH, S. J.; TOMASI, S.; JENKINS, S.; MACHERLA, V.; HOFFMAN, T.; AGARWAL, V.; WILLIAMS, P. G.; DAI, J.; NEUPANE, R.; GURR, J.; RODRÍGUEZ, A. M. C.; LAMSA, A.; ZHANG, C.; DORRESTEIN, K.; DUGGAN, B. M.; ALMALITI, J.; ALLARD, P-M.; PHAPALE, P.; NOTHIAS, L. F.; ALEXANDROV, T.; LITAUDON, M.; WOLFENDER, J-L.; KYLE, J. E.; METZ, T. O.; PERYEA, T.; NGUYEN, D-T.; VANLEER, D.; SHINN, P.; JADHAV, A.; MÜLLER, R.; WATERS, K. M.; SHI, W.; LIU, X.; ZHANG, L.; KNIGHT, R.; JENSEN, P. R.; PALSSON, B. Ø.; POGLIANO, K.; LININGTON, R. G.; GUTIÉRREZ, M.; LOPES, N. P.; GERWICK, W. H.; MOORE, B. S.; DORRESTEIN, P. C.; BANDEIRA, N. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. **Nature Biotechnology**, v. 34, n. 8, p. 828–837, 2016.

WINK, M. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. **Phytochemistry**, v. 64, n. 1, p. 3–19, 2003.

WITTEN, I. H.; EIBE, F.; HALL, M. A. **Data Mining: Practical Machine Learning Tools and Techniques**. Third Edit ed. San Francisco, USA: Morgan Kaufmann, 2005.

APPENDICES

Appendix 1. Samples of Espeletiinae investigated in this project.

ID	species	coll num	paramo massif	paramo complex	date	season	elev (m)	lon	lat
1	*E. annemariana* Cuatrec.	MD3756	Boyacá	Tota - Bijagual - Mamapacha	8/19/2009	Rainy season	3220	5.53119	-72.72407
1	*E. annemariana* Cuatrec.	MD3757	Boyacá	Tota - Bijagual - Mamapacha	8/19/2009	Rainy season	3217	5.53007	-72.72447
2	*E. arbelaezii* Cuatrec.	MD3794	Boyacá	Guantiva - La Rusia	9/5/2009	Rainy to Dry Season	3193	6.10750	-72.84257
2	*E. arbelaezii* Cuatrec.	MD3799	Boyacá	Guantiva - La Rusia	9/5/2009	Rainy to Dry Season	3267	6.17685	-72.77435
2	*E. arbelaezii* Cuatrec.	GFPG021	Cundinamarca	Rabanal y Rio Bogota	1/14/2015	Dry season	3290	5.23472	-73.52417
3	*E. argentea* Humb. & Bonpl.	GFPG020	Cundinamarca	Rabanal y Rio Bogota	1/14/2015	Dry season	3290	5.21000	-73.54833
3	*E. argentea* f. phaneractis (S. F. Blake) Cuatrec.	MD3551	Cundinamarca	Guerrero	12/10/2007	Dry season	3268	5.00883	-74.18704
3	*E. argentea* f. phaneractis (S. F. Blake) Cuatrec.	MD3570	Cundinamarca	Guerrero	12/16/2007	Dry season	3638	5.22265	-74.01019
3	*E. argentea* Humb. & Bonpl.	MD3571	Cundinamarca	Guerrero	12/16/2007	Dry season	3670	5.22458	-74.02458
3	*E. argentea* Humb. & Bonpl.	GFPG016	Cundinamarca	Guerrero	1/11/2015	Dry season	3525	5.21111	-74.03861
3	*E. argentea* Humb. & Bonpl.	GFPG031	Cundinamarca	Cruz Verde - Sumapaz	1/22/2015	Dry season	3328	4.57972	-73.99556
3	*E. argentea* Humb. & Bonpl.	GFPG024	Cundinamarca	Guerrero	1/20/2015	Dry season	3465	5.00944	-74.20361
3	*E. argentea* Humb. & Bonpl.	GFPG036	Cundinamarca	Cruz Verde - Sumapaz	1/29/2015	Dry season	3459	4.18667	-74.20250
4	*E. ariana* S. Díaz & Rodr.-Cabeza	MD3820	Boyacá	Pisba	9/9/2009	Rainy season	3575	5.91343	-72.62026
4	*E. ariana* S. Díaz & Rodr.-Cabeza	MD3821	Boyacá	Pisba	9/9/2009	Rainy season	3631	5.92375	-72.62589
5	*E. aristeguietana* Cuatrec.	MD4081	Venezuela	Niquitao	7/23/2010	n.i.	2302	9.33698	-70.33977
5	*E. aristeguietana* Cuatrec.	MD4082	Venezuela	Niquitao	7/23/2010	n.i.	2291	9.33727	-70.33958
6	*E. azucarina* Cuatrec.	MD3801	Boyacá	Guantiva - La Rusia	9/6/2009	Rainy to Dry Season	4050	6.10772	-72.91202
6	*E. azucarina* Cuatrec.	MD3803	Boyacá	Guantiva - La Rusia	9/6/2009	Rainy to Dry Season	4063	6.10131	-72.91899
7	*E. barclayana* Cuatrec.	MD3537	Cundinamarca	Guerrero	12/10/2007	Dry season	3511	5.01160	-74.20332
7	*E. barclayana* Cuatrec.	MD3546	Cundinamarca	Guerrero	12/10/2007	Dry season	3471	5.01007	-74.20296

7	*E. barclayana* Cuatrec.	MD3566	Cundinamarca	Guerrero	12/16/2007	Dry season	3722	5.21132	-73.99896
7	*E. barclayana* Cuatrec.	GFPG023	Cundinamarca	Guerrero	1/20/2015	Dry season	3440	5.01111	-74.20333
9	*E. boyacensis* Cuatrec.	MD3633	Boyacá	Guantiva - La Rusia	12/28/2007	Dry season	3410	5.90129	-73.07196
9	*E. boyacensis* Cuatrec.	MD3644	Boyacá	Guantiva - La Rusia	12/28/2007	Dry season	3846	5.94692	-73.07827
9	*E. boyacensis* Cuatrec.	MD3792	Boyacá	Guantiva - La Rusia	9/4/2009	Rainy to Dry Season	3543	5.85225	-73.14240
9	*E. boyacensis* Cuatrec.	MD3582	Boyacá	Iguaque - Merchan	12/20/2007	Dry season	3618	5.68577	-73.43876
9	*E. boyacensis* Cuatrec.	GFPG003	Boyacá	Iguaque - Merchan	12/28/2013	Dry season	3384	5.68806	-73.45722
9	*E. boyacensis* Cuatrec.	GFPG006	Boyacá	Iguaque - Merchan	12/28/2013	Dry season	3469	5.68694	-73.45250
10	*E. brachyaxiantha* S. Díaz	MD3662	Boyacá	Guantiva - La Rusia	12/30/2007	Dry season	3966	5.93996	-73.00319
10	*E. brachyaxiantha* S. Díaz	MD3663	Boyacá	Guantiva - La Rusia	12/30/2007	Dry season	3966	5.93991	-73.00305
11	*E. brassicoidea* Cuatrec.	MD3843	Santanderes	Jurisdicciones - Santurban	10/2/2009	Rainy to Dry Season	2639	7.34613	-72.61639
11	*E. brassicoidea* Cuatrec.	MD3856	Santanderes	Jurisdicciones - Santurban	10/2/2009	Rainy to Dry Season	2796	7.33866	-72.60732
12	*E. cayetana* Cuatrec.	MD3579	Boyacá	Iguaque - Merchan	12/20/2007	Dry season	3582	5.68399	-73.44213
12	*E. cayetana* Cuatrec.	GFPG004	Boyacá	Iguaque - Merchan	12/28/2013	Dry season	3416	5.68639	-73.45472
13	*E. cayetana* Cuatrec.	GFPG026	Cundinamarca	Guerrero	1/20/2015	Dry season	3443	5.00917	-74.20750
14	*E.* cf. *cabrerensis*	SLC610	Cundinamarca	Cruz Verde - Sumapaz	9/21/2015	Rainy season	3653	4.32639	-74.13806
14	*E.* cf. *cabrerensis*	SLC667	Cundinamarca	Cruz Verde - Sumapaz	10/15/2015	Rainy season	3734	5.82500	-74.41944
15	*E.* cf. *tapirophila*	SLC664	Cundinamarca	Cruz Verde - Sumapaz	10/15/2015	Rainy season	3734	5.82500	-74.41944
16	*E. cleefii* Cuatrec.	MD3996	Boyacá	Guantiva - La Rusia	1/10/2008	Dry season	4307	6.37759	-72.27270
16	*E. cleefii* Cuatrec.	MD3693	Boyacá	Cocuy	1/7/2008	Dry season	4209	6.40006	-72.27355
16	*E. cleefii* Cuatrec.	MD3696a	Boyacá	Cocuy	1/10/2008	Dry season	4307	6.37759	-72.27270
17	*E. congestiflora* Cuatrec.	MD3640	Boyacá	Guantiva - La Rusia	12/28/2007	Dry season	3627	5.92310	-73.08931
18	*E. conglomerata* A. C. Sm.	MD3867	Santanderes	Almorzadero	10/4/2009	Rainy to Dry Season	3864	6.96302	-72.67341
19	*E. crassa* sp. nov. Diazgranados	MD3645	Boyacá	Guantiva - La Rusia	12/28/2007	Dry season	3981	5.95774	-73.07857
19	*E. crassa* sp. nov. Diazgranados	MD3646	Boyacá	Guantiva - La Rusia	12/28/2007	Dry season	3953	5.95774	-73.07857
20	*E. cuniculorum* Cuatrec.	MD3932	Venezuela	Sierra la Culata	10/25/2009	n.i.	3898	8.79200	-71.00542
21	*E. curialensis* Cuatrec.	MD3812	Boyacá	Pisba	9/8/2009	Rainy season	3373	5.91744	-72.52695
22	*E. delicatissima* sp. nov. Diazgranados	MD3576	Cundinamarca	Cruz Verde - Sumapaz	12/18/2007	Rainy to Dry Season	3597	4.32757	-74.13463

22	*E. delicatissima* sp. nov. Diazgranados	MD3577	Cundinamarca	Cruz Verde - Sumapaz	12/18/2007	Rainy to Dry Season	3591	4.32858	-74.1357
22	*E. delicatissima* sp. nov. Diazgranados	MD3576	Cundinamarca	Cruz Verde - Sumapaz	12/18/2007	Rainy to Dry Season	3597	4.32757	-74.13463
22	*E. delicatissima* sp. nov. Diazgranados	MD3577	Cundinamarca	Cruz Verde - Sumapaz	12/18/2007	Rainy to Dry Season	3591	4.32858	-74.1357
23	*E. discoidea* Cuatrec.	MD3804	Boyacá	Guantiva - La Rusia	9/6/2009	Rainy to Dry Season	3869	6.09403	-72.91146
23	*E. discoidea* Cuatrec.	MD3674	Boyacá	Guantiva - La Rusia	12/30/2007	Dry season	3528	6.06629	-72.90034
24	*E. dugandii* Cuatrec.	MD3872	Santanderes	Almorzadero	10/4/2009	Rainy to Dry Season	3144	6.84148	-72.65555
24	*E. dugandii* Cuatrec.	MD3873	Santanderes	Almorzadero	10/4/2009	Rainy to Dry Season	3155	6.83977	-72.65653
25	*E. episcopalis* S. Díaz & Rodr.-Cabeza	MD3817	Boyacá	Pisba	9/8/2009	Rainy season	3496	5.93538	-72.56063
26	*E. estanislana* Cuatrec.	MD3869	Santanderes	Almorzadero	10/4/2009	Rainy to Dry Season	3785	6.94694	-72.67689
26	*E. estanislana* Cuatrec.	MD3904	Santanderes	Jurisdicciones - Santurban	10/12/2009	Rainy to Dry Season	3712	7.09185	-72.80930
27	*E. formosa* S. Díaz & Rodr.-Cabeza	MD3811	Boyacá	Pisba	9/8/2009	Rainy season	3333	5.91750	-72.52400
27	*E. formosa* S. Díaz & Rodr.-Cabeza	MD3815	Boyacá	Pisba	9/8/2009	Rainy season	3389	5.91978	-72.53227
28	*E. frontinoensis* Cuatrec.	MD3834	Central	Frontino - Urrao	9/25/2009	Rainy season	3752	6.46215	-76.09913
28	*E. frontinoensis* Cuatrec.	MD3837	Central	Frontino - Urrao	9/25/2009	Rainy season	3677	6.44942	-76.08608
29	*E. grandiflora* Humb. & Bonpl.	GFPG039	Cundinamarca	Cruz Verde - Sumapaz	12/11/2014	Dry season	3423	4.42028	-74.12389
29	*E. grandiflora* Humb. & Bonpl.	MD3554	Cundinamarca	Cruz Verde - Sumapaz	12/15/2007	Dry season	3363	4.55447	-73.99613
29	*E. grandiflora* Humb. & Bonpl.	MD3557	Cundinamarca	Cruz Verde - Sumapaz	12/15/2007	Dry season	3415	4.55050	-73.99027
29	*E. grandiflora* Humb. & Bonpl.	MD3559	Cundinamarca	Cruz Verde - Sumapaz	12/15/2007	Dry season	3408	4.55111	-73.99070
29	*E. grandiflora* Humb. & Bonpl.	GFPG033	Cundinamarca	Cruz Verde - Sumapaz	1/27/2015	Rainy to Dry Season	3637	4.43389	-74.19056
29	*E. grandiflora* Humb. & Bonpl.	GFPG030	Cundinamarca	Cruz Verde - Sumapaz	1/22/2015	Rainy to Dry Season	3328	4.58000	-73.99583
29	*E. grandiflora* Humb. & Bonpl.	GFPG025	Cundinamarca	Guerrero	1/20/2015	Dry season	3460	5.00694	-74.20306
29	*E. grandiflora* Humb. & Bonpl.	MD3615	Cundinamarca	Chingaza	12/22/2007	Rainy season	3643	4.68051	-73.78597
29	*E. grandiflora* Humb. & Bonpl.	GFPG037	Cundinamarca	Cruz Verde - Sumapaz	1/29/2015	Rainy to Dry Season	3556	4.18750	-74.20444
29	*E. grandiflora* Humb. & Bonpl.	GFPG038	Cundinamarca	Cruz Verde - Sumapaz	1/29/2015	Rainy to Dry Season	3778	4.11000	-74.25139
30	*E. idroboi* Cuatrec.	MD4090	Central	Guanacas-Puracé-Coconucos	8/10/2011	Rainy season	3342	2.36765	-76.37868

30	*E. idroboi* Cuatrec.	MD4092	Central	Guanacas-Puracé-Coconucos	8/10/2011	Rainy season	3447	2.35886	-76.35050
31	*E. incana* Cuatrec.	MD3641	Boyacá	Guantiva - La Rusia	12/28/2007	Dry season	3704	5.93140	-73.08175
31	*E. incana* Cuatrec.	MD3656	Boyacá	Tota - Bijagual - Mamapacha	12/29/2007	Dry season	3896	5.70295	-72.79561
32	*E. jajoensis* Aristeg.	MD4077	Venezuela	Sierra Nevada	7/22/2010	n.i.	3129	9.06999	-70.62128
32	*E. jajoensis* Aristeg.	MD4079	Venezuela	Sierra Nevada	7/22/2010	n.i.	3096	9.06995	-70.62028
33	*E. jaramilloi* S. Díaz	MD3805	Boyacá	Pisba	9/7/2009	Rainy season	3615	5.88796	-72.67240
33	*E. jaramilloi* S. Díaz	MD3818	Boyacá	Pisba	9/8/2009	Rainy season	3576	5.94502	-72.58105
34	*E. killipii* Cuatrec. var. chisacana Cuatrec.	MD3572	Cundinamarca	Cruz Verde - Sumapaz	12/18/2007	Rainy to Dry Season	3508	4.32291	-74.13638
34	*E. killipii* Cuatrec. var. chisacana Cuatrec.	MD3575	Cundinamarca	Cruz Verde - Sumapaz	12/18/2007	Rainy to Dry Season	3629	4.32406	-74.13012
34	*E. killipii* Cuatrec.	GFPG034	Cundinamarca	Cruz Verde - Sumapaz	1/29/2015	Rainy to Dry Season	3549	4.29091	-74.20683
35	*E. litocrassa* sp. nov. Diazgranados	MD3624	Cundinamarca	Chingaza	12/22/2007	Rainy season	3455	4.73932	-73.84095
35	*E. litocrassa* sp. nov. Diazgranados	MD3628	Cundinamarca	Chingaza	12/22/2007	Rainy season	3445	4.73839	-73.84215
35	*E. litocrassa* sp. nov. Diazgranados	MD3618	Cundinamarca	Chingaza	12/22/2007	Rainy season	3481	4.74193	-73.83978
35	*E. litocrassa* sp. nov. Diazgranados	MD3619	Cundinamarca	Chingaza	12/22/2007	Rainy season	3481	4.74193	-73.83978
35	*E. litocrassa* sp. nov. Diazgranados	MD3620	Cundinamarca	Chingaza	12/22/2007	Rainy season	3486	4.74191	-73.84019
35	*E. litocrassa* sp. nov. Diazgranados	MD3621	Cundinamarca	Chingaza	12/22/2007	Rainy season	3478	4.74126	-73.8401
35	*E. litocrassa* sp. nov. Diazgranados	MD3622	Cundinamarca	Chingaza	12/22/2007	Rainy season	3467	4.74042	-73.84043
35	*E. litocrassa* sp. nov. Diazgranados	MD3623	Cundinamarca	Chingaza	12/22/2007	Rainy season	3464	4.73984	-73.8404
35	*E. litocrassa* sp. nov. Diazgranados	MD3624	Cundinamarca	Chingaza	12/22/2007	Rainy season	3455	4.73932	-73.84095
35	*E. litocrassa* sp. nov. Diazgranados	MD3625	Cundinamarca	Chingaza	12/22/2007	Rainy season	3451	4.73892	-73.84118
35	*E. litocrassa* sp. nov. Diazgranados	MD3627	Cundinamarca	Chingaza	12/22/2007	Rainy season	3451	4.73874	-73.84166
35	*E. litocrassa* sp. nov. Diazgranados	MD3628	Cundinamarca	Chingaza	12/22/2007	Rainy season	3445	4.73839	-73.84215
36	*E. lopezii* Cuatrec. var. major Cuatrec.	MD3649	Boyacá	Tota - Bijagual - Mamapacha	12/29/2007	Dry season	3531	5.70844	-72.82110

36	*E. lopezii* Cuatrec.	MD3699	Boyacá	Cocuy	1/11/2008	Dry season	4012	6.35683	-72.33491
36	*E. lopezii* Cuatrec.	GFPG022	Cundinamarca	Rabanal y Rio Bogota	1/14/2015	Dry season	3290	5.21583	-73.52722
37	*E. monguana* sp. nov. Diazgranados	MD3765	Boyacá	Tota - Bijagual - Mamapacha	8/22/2009	Rainy season	3942	5.79575	-72.72621
37	*E. monguana* sp. nov. Diazgranados	MD3767	Boyacá	Tota - Bijagual - Mamapacha	8/22/2009	Rainy season	3907	5.79551	-72.73973
38	*E. multicongestiflora* sp. nov. Diazgranados	MD3751	Boyacá	Tota - Bijagual - Mamapacha	8/17/2009	Rainy season	3403	5.52165	-72.85474
39	*E. murilloi* var. nov. Cuatrec. var. *alba* Diagranados	MD3773	Boyacá	Guantiva - La Rusia	8/25/2009	Dry to Rainy season	3724	5.95433	-73.01142
39	*E. murilloi* Cuatrec.	MD3740	Boyacá	Tota - Bijagual - Mamapacha	8/15/2009	Rainy season	3278	5.56292	-73.12767
39	*E. murilloi* var. nov. Cuatrec. var. *iguaquei* Diagranados	MD3583	Boyacá	Iguaque - Merchan	12/20/2007	Dry season	3603	5.68695	-73.43684
39	*E. murilloi* Cuatrec.	GFPG009	Boyacá	Iguaque - Merchan	12/28/2013	Dry season	3591	5.68889	-73.43972
39	*E. murilloi* var. nov. Cuatrec. var. *alba* Diagranados	MD3592	Cundinamarca	Rabanal y Rio Bogota	12/21/2007	Dry season	3190	5.20817	-73.54821
39	*E. murilloi* var. nov. Cuatrec. var. *alba* Diazgranados	MD3592	Cundinamarca	Rabanal y Rio Bogota	12/21/2007	Dry season	3190	5.20817	-73.54821
40	*E. nana* Cuatrec.	MD4001	Venezuela	Niquitao	12/28/2009	n.i.	3047	9.08698	-70.37332
40	*E. nana* Cuatrec.	MD4007	Venezuela	Niquitao	12/29/2009	n.i.	3718	9.11763	-70.48717
41	*E. nemekenei* Cuatrec.	MD3632	Boyacá	Guantiva - La Rusia	12/28/2007	Dry season	3403	5.90098	-73.07161
41	*E. nemekenei* Cuatrec.	MD3639	Boyacá	Guantiva - La Rusia	12/28/2007	Dry season	3595	5.92242	-73.08448
42	*E. occidentalis* A. C. Sm.	MD3827	Central	Belmira	9/23/2009	Rainy season	3142	6.62629	-75.64692
42	*E. occidentalis* A. C. Sm.	MD3828	Central	Belmira	9/23/2009	Rainy season	3218	6.63301	-75.64763
43	*E. oirensis* sp. nov. Diazgranados	MD3924	Santanderes	Tama	10/15/2009	Rainy to Dry Season	3153	7.39453	-72.38863
44	*E. oswaldiana* S. Díaz	MD3758	Boyacá	Tota - Bijagual - Mamapacha	8/19/2009	Rainy season	2861	5.50365	-72.74306
45	*E. paipana* S. Díaz & Pedraza	MD3788	Boyacá	Guantiva - La Rusia	9/4/2009	Rainy to Dry Season	3538	5.85311	-73.14345
45	*E. paipana* S. Díaz & Pedraza	MD3789	Boyacá	Guantiva - La Rusia	9/4/2009	Rainy to Dry Season	3530	5.85311	-73.14345
46	*E. pescana* (Diaz S.) S. Diaz	MD3659	Boyacá	Tota - Bijagual - Mamapacha	12/29/2007	Dry season	3795	5.70448	-72.80810
46	*E. pescana* (Diaz S.) S. Diaz	MD3760	Boyacá	Tota - Bijagual - Mamapacha	8/20/2009	Rainy season	3810	5.69464	-72.78730
47	*E. pisbana* S. Díaz & Rodr.-Cabeza	MD3814	Boyacá	Pisba	9/8/2009	Rainy season	3382	5.91992	-72.53231

48	*E. praefrontina* Cuatrec.	MD3832	Central	Frontino - Urrao	9/25/2009	Rainy season	3773	6.46111	-76.10016
48	*E. praefrontina* Cuatrec.	MD3835	Central	Frontino - Urrao	9/25/2009	Rainy season	3604	6.45530	-76.09002
49	*E. roberti* Cuatrec.	MD3888	Santanderes	Jurisdicciones - Santurban	10/10/2009	Rainy to Dry Season	3493	7.84869	-73.22305
49	*E. roberti* Cuatrec.	MD3889	Santanderes	Jurisdicciones - Santurban	10/10/2009	Rainy to Dry Season	3473	7.84823	-73.22252
50	*E. rositae* Cuatrec.	MD3800	Boyacá	Guantiva - La Rusia	9/6/2009	Rainy to Dry Season	3837	6.10400	-72.89922
51	*E. saboyensis* sp. nov. Diazgranados	MD3782	Boyacá	Iguaque - Merchan	9/2/2009	Dry season	3364	5.71461	-73.81660
51	*E. saboyensis* sp. nov. Diazgranados	MD3783	Boyacá	Iguaque - Merchan	9/2/2009	Dry season	3339	5.71154	-73.81597
52	*E. schultesiana* Cuatrec.	MDCama	Central	La Cocha - Patascoy	1/9/2009	Rainy season	3500	1.16252	-77.09879
53	*E. schultzii* Wedd.	PS10185	Venezuela	Sierra Nevada	10/10/2007	n.i.	4120	8.83222	-70.832500
53	*E. schultzii* Wedd.	MD3942	Venezuela	Sierra la Culata	10/27/2009	n.i.	4273	8.88423	-70.86298
53	*E. schultzii* Wedd.	MD3945	Venezuela	Sierra Nevada	10/28/2009	n.i.	3409	8.59244	-71.01806
54	*E. semiglobulata* Cuatrec.	MD4071	Venezuela	Sierra la Culata	7/21/2010	n.i.	4072	8.90565	-70.86231
55	*E. standleyana* A. C. Sm.	MD3868	Santanderes	Almorzadero	10/4/2009	Rainy to Dry Season	3837	6.96418	-72.67314
55	*E. standleyana* A. C. Sm.	MD3840	Santanderes	Jurisdicciones - Santurban	10/1/2010	Rainy to Dry Season	3419	7.10837	-72.97531
56	*E. steyermarkii* Cuatrec.	MD3861	Santanderes	Almorzadero	10/3/2009	Rainy to Dry Season	2800	7.06884	-72.65411
56	*E. steyermarkii* Cuatrec.	MD3862	Santanderes	Almorzadero	10/3/2009	Rainy to Dry Season	2801	7.06884	-72.65411
57	*E. summapacis* Cuatrec.	MD3735	Cundinamarca	Cruz Verde - Sumapaz	8/14/2009	Rainy season	3891	4.28734	-74.21742
57	*E. summapacis* Cuatrec.	MD3737	Cundinamarca	Cruz Verde - Sumapaz	8/14/2009	Rainy season	3838	4.28010	-74.21471
58	*E. susensis* sp. nov. Diazgranados	MD3748	Boyacá	Tota - Bijagual - Mamapacha	8/16/2009	Rainy season	3762	5.43450	-72.95312
59	*E. tenorae* cf. Aristeg.	MD4011	Venezuela	Niquitao	12/29/2009	n.i.	3664	9.11404	-70.48299
60	*E. tetrarbelaezii* sp. nov. Diazgranados	MD3706	Boyacá	Guantiva - La Rusia	1/12/2008	Dry season	3260	6.13705	-72.77567
61	*E. toquensis* sp. nov. Diazgranados	MD3746	Boyacá	Tota - Bijagual - Mamapacha	8/15/2009	Rainy season	3487	5.55333	-73.10761
61	*E. toquensis* sp. nov. Diazgranados	MD3746	Boyacá	Tota - Bijagual - Mamapacha	8/15/2009	Rainy season	3487	5.55333	-73.10761
62	*E. totensis* sp. nov. Diazgranados	MD3752	Boyacá	Tota - Bijagual - Mamapacha	8/17/2009	Rainy season	3406	5.52159	-72.85478
63	*E. tunjana* Cuatrec.	MD3742	Boyacá	Tota - Bijagual - Mamapacha	8/15/2009	Rainy season	3444	5.55172	-73.11167
64	*E. ulotricha* Cuatrec.	MD3970	Venezuela	Parque Nacional Dinira	12/23/2009	n.i.	3483	9.55653	-70.11703

65	*E. uribei* Cuatrec.	MD3617	Cundinamarca	Chingaza	12/22/2007	Rainy season	3479	4.74181	-73.83966
65	*E. uribei* Cuatrec.	MD3626	Cundinamarca	Chingaza	12/22/2007	Rainy season	3450	4.73879	-73.84115
66	*E. weddellii* Sch. Bip. ex Wedd.	MD3937	Venezuela	Sierra la Culata	10/27/2009	n.i.	4239	8.86994	-70.84398
67	*Es. corymbosa* (Humb. & Bonpl.) Cuatrec.	GFPG015	Cundinamarca	Guerrero	1/11/2015	Dry season	3304	5.19167	-74.02556
67	*Es. corymbosa* (Humb. & Bonpl.) Cuatrec.	GFPG028	Cundinamarca	Cruz Verde - Sumapaz	1/22/2015	Dry season	3344	4.58028	-73.99583
67	*Es. corymbosa* (Humb. & Bonpl.) Cuatrec.	MD3552	Cundinamarca	Cruz Verde - Sumapaz	12/15/2007	Dry season	3171	4.61039	-74.02419
67	*Es. corymbosa* (Humb. & Bonpl.) Cuatrec.	MD3565	Cundinamarca	Cruz Verde - Sumapaz	12/15/2007	Dry season	3303	4.56342	-73.98688
68	*Es. garciae* (Cuatrec.) Cuatrec.	GFPG008	Boyacá	Iguaque - Merchan	12/28/2013	Dry season	3596	5.68389	-73.44472
68	*Es. garciae* (Cuatrec.) Cuatrec.	MD3587	Boyacá	Iguaque - Merchan	12/20/2007	Dry season	3600	5.68633	-73.43790
68	*Es. garciae* (Cuatrec.) Cuatrec.	MD3580	Boyacá	Iguaque - Merchan	12/20/2007	Dry season	3609	5.68371	-73.44030
68	*Es. garciae* (Cuatrec.) Cuatrec.	MD3586	Boyacá	Iguaque - Merchan	12/20/2007	Dry season	3600	5.6864	-73.43774
69	*Es. garciae* (Cuatrec.) Cuatrec.	GFPG019	Cundinamarca	Rabanal y Rio Bogota	1/14/2015	Dry season	3290	5.21111	73.55028
69	*Es. garciae* (Cuatrec.) Cuatrec.	GFPG017	Cundinamarca	Guerrero	1/11/2015	Dry season	3611	5.22639	-74.03111
70	*Es. pleiochasia* (Cuatrec.) Cuatrec.	MD3631	Boyacá	Guantiva - La Rusia	12/28/2007	Dry season	3126	5.87988	-73.06394
70	*Es. pleiochasia* (Cuatrec.) Cuatrec.	MD3630	Boyacá	Guantiva - La Rusia	12/28/2007	Dry season	3084	5.87398	-73.06346
70	*Es. pleiochasia* (Cuatrec.) Cuatrec.	GFPG007	Boyacá	Iguaque - Merchan	12/28/2013	Dry season	3531	5.68611	-73.45028
71	*Es. rabanalensis* S. Díaz & Rodríguez-Cabeza	MD3591	Cundinamarca	Rabanal y Rio Bogota	12/21/2007	Dry season	3191	5.20814	-73.54867
72	*L. neriifolius* (Bonpl. ex Humb.) Ernst	JBB	Santanderes	Jurisdicciones - Santurban	10/7/2011	Rainy to Dry Season	2500	7.4275556	-72.582239
72	*L. neriifolius* (Sch. Bip. ex Wedd.) Ernst.	PS10116	Venezuela	Vargas	10/7/2007	n.i.	2230	10.548889	-66.867500
73	*L. sp1*	PS10432	Venezuela	Chorro el indio	10/19/2007	n.i.	3310	8.1619445	-71.898889
74	*L. sp2*	PS10608	Venezuela	Sierra Nevada	10/24/2007	n.i.	3490	8.5602778	-71.091945
75	*L. sp3*	PS10302	Venezuela	Sierra la Culata	10/14/2007	n.i.	2830	9.0377778	-70.856667
76	*L. sp4*	PS10348	Venezuela	Chorro el indio	10/17/2007	n.i.	3070	8.0213889	-71.968889

77	*L. sp5*	PS10510	Venezuela	Sierra Nevada	10/19/2007	n.i.	3140	8.3263889	-71.306667
78	*L. sp6*	PS10525	Venezuela	Parque Nacional Dinira	10/23/2007	n.i.	2770	9.5952778	-70.108889
79	*L. sp7*	PS10527	Venezuela	Parque Nacional Dinira	10/22/2007	n.i.	3230	9.5863889	-70.119444
80	*L. sp8*	PS10576	Venezuela	Parque Nacional Dinira	10/21/2007	n.i.	2400	9.6833333	-70.166667
81	*P. glandulosus* (Cuatrec.) Cuatrec.	MD3581	Boyacá	Iguaque - Merchan	12/20/2007	Dry season	3617	5.68459	-73.43949
81	*P. glandulosus* (Cuatrec.) Cuatrec.	MD3581b	Boyacá	Iguaque - Merchan	12/20/2007	Dry season	3617	5.68459	-73.43949
82	*R. sp1*	PS10372	Venezuela	Chorro el indio	10/17/2007	n.i.	3460	8.0275000	-71.959166
83	*R. sp2*	PS10607	Venezuela	Sierra Nevada	10/24/2007	n.i.	3460	8.5647223	-71.090000
84	*R. sp3*	PS10333	Venezuela	Chorro el indio	10/16/2007	n.i.	3130	8.2530556	-71.873333
85	*R. sp4*	PS10518	Venezuela	Sierra Nevada	10/19/2007	n.i.	3140	8.3263889	-71.306667
86	*R. sp5*	PS10532	Venezuela	Parque Nacional Dinira	10/23/2007	n.i.	3180	9.5880556	-70.118889
87	*R. sp6*	PS10545	Venezuela	Parque Nacional Dinira	10/23/2007	n.i.	2650	9.5941667	-70.104167
91	*E. sp1*	PS10435	Venezuela	Chorro el indio	10/19/2007	n.i.	3310	8.1619445	-71.898889
92	*E. sp2*	PS10439	Venezuela	Chorro el indio	10/18/2007	n.i.	3360	7.9247223	-72.078889
93	*E. sp3*	PS10462	Venezuela	Chorro el indio	10/18/2007	n.i.	3100	7.9494445	-72.081112
94	*E. sp4*	PS10604	Venezuela	Sierra Nevada	10/24/2007	n.i.	3950	8.5402778	-71.075278
95	*E. sp5*	PS10605	Venezuela	Sierra Nevada	10/24/2007	n.i.	3950	8.5402778	-71.075278
96	*E. sp6*	PS10606	Venezuela	Sierra Nevada	10/24/2007	n.i.	3950	8.5402778	-71.075278
97	*E. sp7*	PS10636	Venezuela	Sierra Nevada	10/25/2007	n.i.	3630	8.6841667	-70.878889
98	*E. sp8*	PS10186	Venezuela	Sierra Nevada	10/10/2007	n.i.	4120	8.8322223	-70.832500
99	*E. sp9*	PS10188	Venezuela	Sierra Nevada	10/10/2007	n.i.	4240	8.8583334	-70.825834
100	*E. sp10*	PS10280	Venezuela	Sierra Nevada	10/14/2007	n.i.	3730	9.0294445	-70.584445
101	*E. sp11*	PS10309	Venezuela	Sierra Nevada	10/14/2007	n.i.	4280	8.8758334	-70.845556
102	*E. sp12*	PS10330	Venezuela	Chorro el indio	10/16/2007	n.i.	3370	8.2486111	-71.881945
103	*E. sp13*	PS10346	Venezuela	Chorro el indio	10/16/2007	n.i.	3200	8.2522222	-71.878333
104	*E. sp14*	PS10373	Venezuela	Chorro el indio	10/17/2007	n.i.	3460	8.0275000	-71.959166
105	*E. sp15*	PS10495	Venezuela	Sierra Nevada	10/19/2007	n.i.	3140	8.3263889	-71.306667
106	*E. sp16*	PS10500	Venezuela	Sierra Nevada	10/19/2007	n.i.	3140	8.3263889	-71.306667

107	*E. sp17*	PS10501	Venezuela	Sierra Nevada	10/19/2007	n.i.	3140	8.3263889	-71.306667
108	*E. sp18*	PS10514	Venezuela	Sierra Nevada	10/19/2007	n.i.	3140	8.3263889	-71.306667
109	*E. sp19*	PS10519	Venezuela	Sierra Nevada	10/19/2007	n.i.	3140	8.3263889	-71.306667
110	*E. sp20*	PS10520	Venezuela	Parque Nacional Dinira	10/23/2007	n.i.	3210	9.5905556	-70.121666
111	*E. sp21*	PS10523	Venezuela	Parque Nacional Dinira	10/22/2007	n.i.	2850	9.5938889	-70.110833
112	*E. sp22*	PS10524	Venezuela	Parque Nacional Dinira	10/22/2007	n.i.	3160	9.5883333	-70.118056
113	*E. sp23*	PS10543	Venezuela	Parque Nacional Dinira	10/22/2007	n.i.	2700	9.5955556	-70.107223
114	*E. sp24*	PS10546	Venezuela	Parque Nacional Dinira	10/22/2007	n.i.	2730	9.5950000	-70.107778
115	*E. sp25*	PS10554	Venezuela	Sierra Nevada	10/24/2007	n.i.	3400	8.9841667	-70.547222
116	*E. sp26*	PS10555	Venezuela	Sierra Nevada	10/24/2007	n.i.	3550	9.0025000	-70.543333
117	*E. sp27*	PS10557	Venezuela	Sierra Nevada	10/24/2007	n.i.	3350	8.9769444	-70.551111
118	*E. sp28*	PS10561	Venezuela	Sierra Nevada	10/24/2007	n.i.	3350	8.9769444	-70.551111

n.i.: no information available

Appendix 2. Samples of *Espeletia* investigated in the biogeographic correlation with the páramo complexes.

Cod	Cod. coleta	Espécie	Complexo de páramo
1	GFPG016A	*E. argentea*	Páramo de Guerrero
2	GFPG016B	*E. argentea*	Páramo de Guerrero
3	GFPG024	*E. argentea*	Páramo de Guerrero
4	GFPG020A	*E. argentea*	Páramo de Rabanal
5	GFPG020B	*E. argentea*	Páramo de Rabanal
6	GFPG020C	*E. argentea*	Páramo de Rabanal
7	GFPG031A	*E. argentea*	Páramo de Sumapaz
8	GFPG031B	*E. argentea*	Páramo de Sumapaz
9	GFPG036	*E. argentea*	Páramo de Sumapaz
10	GFPG006	*E. boyacensis*	Páramo de Iguaque
11	MD3582	*E. boyacensis*	Páramo de Iguaque
13	MD3582	*E. boyacensis*	Páramo de Iguaque
14	MD3633	*E. boyacensis*	Páramo de Guantiva
15	MD3644	*E. boyacensis*	Páramo de Guantiva
16	MD3792	*E. boyacensis*	Páramo de Guantiva
17	MD3554	*E. grandiflora*	Páramo de Cruz Verde
18	MD3557	*E. grandiflora*	Páramo de Cruz Verde
19	MD3559	*E. grandiflora*	Páramo de Cruz Verde
20	GFPG037	*E. grandiflora*	Páramo de Sumapaz
21	GFPG038	*E. grandiflora*	Páramo de Sumapaz
22	MD3616	*E. grandiflora*	Páramo de Chingaza
23	MD3616	*E. grandiflora*	Páramo de Chingaza
24	MD3616	*E. grandiflora*	Páramo de Chingaza
25	GFPG025A	*E. grandiflora*	Páramo de Guerrero
26	GFPG025B	*E. grandiflora*	Páramo de Guerrero

Appendix 3. Parameters used in the software MZmine.

1. Raw data import

```
Raw data methods > Raw data import
```
2. Raw data filtering:

```
Raw data filtering > Scan by scan filtering
Filter: Savitzky-Golay filter
Number of datapoints: 5
```
3. Mass detection:

```
Peak Picking > Mass detection
MS level 1
Set filters: (+ or - depending on polarity)
Mass detector: Exact mass
Noise level: 1.0E6
```
4. FTMS shoulder peaks filter:

```
Peak Picking > FTMS shoulder peaks filter
Mass resolution: 70000
Peak model function: Lorentzian extended
```
5. Chromatogram builder:

```
Peak Picking > Chromatogram builder
Min time span (min): 0.2
Min height: 5.0E6
m/z tolerance: 0.002 m/z or 5.0 ppm
```
6. Chromatograms deconvolution:

```
Peak list processing > Chromatogram deconvolution
Wavelets (XCMS)
S/N threshold: 10
Wavelet scales: 0.25 - 5.0
Peak duration range: 0 - 5.0
```
7. Removing of isotopes:

```
Isotopes > Isotopic peaks grouper
m/z tolerance: 0.002 m/z or 5.0 ppm
Retention time tolerance: 0.7 min
Maximum charge: 2
Representative isotope: Most intense
```
8. Alignment:

```
Alignment > Ransac Aligner
m/z tolerance: 0.002 m/z or 5.0 ppm
Retention time tolerance: 0.7 min
RT time tolerance after correction: 0.3
Ransac interations: 0
Minimum number of points: 50%
```

Threshold value: 10
9. Gap filling:

Gap filling > Peak finder
Intensity tolerance: 30%
m/z tolerance: 0.002 m/z or 5.0 ppm
Retention time tolerance: 0.3 min
Retention time correction
10. Filtering:

Filtering > Duplicate peaks filter
m/z tolerance: 0.002 m/z or 5.0 ppm
Retention time tolerance: 0.3 min
11. Identification:

Identification > Fragment search
Retention time tolerance: 0.7 min
m/z tolerance: 0.002 m/z or 5.0 ppm
Max fragment peak height: 50.0 %
Min MS2 peak height: 1.5E6

Identification > Adduct search
Retention time tolerance: 0.7 min
m/z tolerance: 0.002 m/z or 5.0 ppm
Max relative adduct peak height: 50.0%

Identification > Complex search
Retention time tolerance: 0.7 min
m/z tolerance: 0.002 m/z or 5.0 ppm
Max complex peak height: 50.0%

12. Identification using a custom database

Identification > Custom database search

Appendix 4. Decision trees of the positive (A) and combined (B) datasets showing metabolomic correlation with biogeographic data.

Appendix 5. Detailed information about all dereplicated metabolites in species of Espeletiinae.

	Rt (min)	Substance	Negative ionization pseudomolecular ion (*m/z*)	Negative MS/MS	Positive ionization pseudomolecular ion (*m/z*)	Positive MS/MS	Confidence level*
1	0.92	gluconic acid or isomers	[M - H]⁻ 195.05013	-	-	-	3
2	0.98	quinic acid	[M - H]⁻ 191.05534	-	-	-	1
3	1.00	xylonic acid or isomers	[M - H]⁻ 165.03951	-	-	-	3
4	1.02	altraric acid	[M - H]⁻ 209.02950	-	-	-	2
5	1.02	glycosylated metabolite (C$_9$H$_{16}$O$_9$)	[M - H]- 267.07233	-	-	-	3
6	1.06	di-acetyl altraric acid	[M - H]⁻ 235.04552	-	-	-	2
7	1.11	malic acid or hydroxypropanedioic acid	[M - H]⁻ 133.01317	-	-	-	3
8	3.14	protocatechuic acid	[M - H]⁻ 153.01825	-	-	-	2
9	3.22	3-caffeoylquinic acid	[M - H]⁻ 353.08759	191.05539bp	[M + H]⁺ 355.10217, [(M + H) - QA]⁺ 163.03888	-	1
10	4.52	5-caffeoylquinic acid	[M - H]⁻ 353.08780	191.05539bp	[M + H]⁺ 355.10226, [(M + H) - QA]⁺ 163.03893	-	1
11	5.08	caffeic acid	[M - H]⁻ 179.03426	135.04410bp	-	-	1
12	5.87	1,3-dicaffeoylquinic acid	[M - H]⁻ 515.11902	353.08771, 335.07791, 191.05518bp, 179.03401, 135.04395	[M + H]⁺ 517.13336	-	1
13	6.10	5-feruloylquinic acid	[M - H]⁻ 367.10355	191.05544	[M + H]⁺ 369.11786	-	1
14	6.25	quercetin-3-di-*O*-glycoside	[M - H]⁻ 625.14142	463.08905, 300.02789bp, 271.02539	[M + H]⁺ 627.15546 [(M + H) - glycoside]⁺ 465.10278, [(M + H) - 2xglycosides]⁺ 303.04962	-	2
15	6.45	coumarinic acid	[M - H]⁻ 163.03905	119.04901bp	-	-	1
16	6.70	methyl 5-*O*-caffeoylquinate	[M - H]⁻ 367.10394	179.03430, 161.02359, 135.04413bp	[M + H]⁺ 369.11813	-	1

17	6.94	quercetin-3-*O*-arabinoglucoside	[M - H]⁻ 595.13110	300.02780bp, 271.02551, 151.00275	[M + H]⁺ 597.14539	-	1
18	7.04	scopoletin	[M - H]⁻ 191.03427	176.01077bp, 148.01563	[M + H]⁺ 193.04964	-	1
19	7.16	rutin	[M - H]⁻ 609.14575	301.03403, 300.02744bp, 271.09766, 245.04865	[M + H]⁺ 611.15997 [(M + H) - methylpentoside]⁺ 465.10248, [(M + H) - methylpentoside-hexoside]⁺ 303.04953	-	1
20	7.49	quercetin-3-*O*-galactoside	[M - H]⁻ 463.08838	300.02789bp, 271.02515, 255.03008, 151.00275	[M + H]⁺ 465.10272, [(M + H) - glycose]⁺ 303.04959	-	1
21	7.54	2-formyl tetrahydropyran-3-yl acetate	[M - H]⁻ 177.06525, [(M - H) - CO₂]⁻ 127.07523	-	-	-	2
22	7.64	Isorhamnetin-3-*O*-rutinoside	[M - H]⁻ 623.16217	608.13953, 315.05124bp, 300.02792, 299.02005, 171.06548	[M + H]⁺ 625.17633	-	1
23	7.96	quercetin-3-*O*-pentoside	[M - H]⁻ 433.07727	300.02750bp, 271.02496, 255.02992, 151.00266	[M + H]⁺ 435.09158bp, [(M + H) - arabinose]⁺ 303.04935	-	2
24	7.99	3-*O*-methylquercetin 7-*O*-glucopyranoside	[M - H]⁻ 477.10373	462.08078, 315.05118, 299.01978bp, 271.02512	[M + H]⁺ 479.11777	-	1

25	8.01	methyl 3-O-caffeoyl-4-O-feruloylquinate	[M - H]⁻ 543.15106	-	[M + H]⁺ 545.16528, [(M + H) - H₂O]⁺ 527.15454, [(M + H) - C₇H₈O₂]⁺ 421.11270, [(M + H) - H₂O -FER]⁺ 351.10733, [(M + H) - FER - 2(H₂O)]⁺ 333.08618 [(M + H) - C₁₇H₁₉O₉]⁺ 177.05457, [(M + H) - C₁₈H₂₁O₉]⁺ 163.03885	-	2
26	8.10	1,5-dicaffeoylquinic acid	[M - H]⁻ 515.11963	353.08820, 335.07724, 191.05542bp, 179.03429, 135.04411	[M + H]⁺ 517.13428	-	1
27	8.16	tricaffeoylaltraric acid isomer	[M - H]⁻ 695.12665	533.09436, 371.06226, 209.02983bp	-	-	3
28	8.37	3,5-dicaffeoylquinic acid	[M - H]⁻ 515.11908	353.08762, 191.05511bp, 179.03401, 135.04395	[M + H]⁺ 517.13342	-	1
29	8.47	quercetin-7-O-glycoside	[M - H]⁻ 463.08905	301.03546bp, 151.00284	[M + H]⁺ 465.10287, [(M + H) - glycose]⁺ 303.04922	-	2
30	8.66	3,4-dicaffeoylquinic acid	[M - H]⁻ 515.11945	353.08801, 191.05534, 179.03409, 173.04471bp, 135.04404	[M + H]⁺ 517.13422	-	1
31	9.12	tricaffeoylaltraric acid isomer	[M - H]⁻ 695.12476	533.09357, 371.06177, 209.02945bp	-	-	3
32	9.32	3,4-dihydroxyphenyl caffeate	[M - H]⁻ 287.05624	125.02322bp	[M + H]⁺ 289.07043	-	2
33	9.42	tricaffeoylaltraric acid isomer	[M - H]⁻ 695.12531	533.09375, 371.06201, 209.02960bp	-	-	3

34	9.72	2,3,5 or 2,4,5-tricaffeoylaltraric acid	[M - H]⁻ 695.12439	533.09344, 371.06146, 209.02939bp	[M + H]⁺ 697.13916	-	1
35	9.92	quercetin-3-O-methylpentoside-hexoside	[M - H]⁻ 609.12524	463.08878bp, 301.03455, 300.02771, 271.02521, 151.00272	[M + H]⁺ 611.13934, [(M + H) - methylpentoside-hexoside]⁺ 303.04944	-	2
36	10.00	trihydroxy-pentamethoxyflavone	[M - H]⁻ 419.09885	-	-	-	2
37	10.14	isorhamnetin 3-O-(acetylglycoside)	[M - H]⁻ 519.11475	504.09186, 315.05133, 299.01993bp, 271.02509	[M + H]⁺ 521.12915	-	2
38	10.59	luteolin	[M - H]⁻ 285.04059	175.03934, 151.00276, 133.02846bp	[M + H]⁺ 287.05505	-	1
39	10.60	tiliroside	[M - H]⁻ 593.13159	447.09290, 285.04031bp	[M + H]⁺ 595.14520	-	1
40	10.65	quercetin	[M - H]⁻ 301.03506	271.02487, 255.02972, 243.02966, 151.00256bp	[M + H]⁺ 303.04984	-	1
41	10.94	3,4,5-tri-O-caffeoylquinic acid	[M - H]⁻ 677.15228	515.12024, 353.08820bp, 335.07770, 191.01918, 179.03426, 135.04413	[M + H]⁺ 679.16541	-	1
42	11.17	3-methoxy quercetin	[M - H]⁻ 315.05087	300.02701, 271.02451bp, 255.02948, 243.02948	[M + H]⁺ 317.06461	-	1
43	11.29	biflavonoid ($C_{32}H_{22}O_{14}$)	[M - H]⁻ 629.09418	614.07074, 599.04755, 329.06686, 271.02490bp	[M + H]⁺ 631.10828	-	3

44	11.97	apigenin	[M - H]⁻ 269.04578	151.00269, 117.03344bp	[M + H]⁺ 271.06027	-	1
45	12.27	kaempferol	[M - H]⁻ 285.04050	-	[M + H]⁺ 287.05481	-	1
46	12.44	hesrepetin	[M - H]⁻ 301.07196	-	[M + H]⁺ 303.08618	-	1
47	12.62	5-hydroxyanthraquinone-1,3-dicarboxylic acid	[M - H]⁻ 311.02036	-	[M + H]⁺ 313.03442	-	2
48	12.94	pentahydroxy-tetramethoxyflavone	-	-	[M + H]⁺ 359.11276	344.08768, 329.06573, 298.08383bp, 283.06049, 255.06546	2
49	12.94	dimethoxyflavonoid ($C_{17}H_{14}O_7$)	[M - H]⁻ 329.06705	314.04358, 299.01987, 271.02490bp, 243.02971	[M + H]⁺ 331.08112	-	3
50	13.40	putative biflavonoid-protocatechuic acid ($C37H22O18$)	[M - H]⁻ 753.07391	299.01999bp	[M + H]⁺ 755.08801	-	2
51	13.64	isoetin	[M - H]⁻ 301.07178	299.02011, 165.01848, 135.04407bp	[M + H]⁺ 303.08633	-	2
52	14.48	dimethoxyquercetin	[M - H]⁻ 329.06683	314.04343, 299.01987bp, 271.02496, 243.02988	[M + H]⁺ 331.08075	-	2
53	14.71	putative diterpene ($C_{20}H_{32}O_4$)	[M - H]⁻ 335.22321	-	[M + H]⁺ 337.23712, [(M + H) - H₂O]⁺ 319.22672, [(M + H) - 2xH₂O]⁺ 301.21613	-	3
54	14.72	enhydrin	-	-	[M + ACN]⁺ 506.20184, [M + K]⁺ 503.13150, [M + Na]⁺ 487.15741, [M + NH₄]⁺ 482.20209, [M + H]⁺ 465.17563	405.15414, 349.12796, 288.28961, 257.08063, 229.08582	1
55	15.07	biflavonoid ($C31H20O14$)	[M - H]⁻ 615.07874, [(M - H) - QER - CH₃]⁻ 299.01990	299.01974bp, 271.02509, 243.03000	[M + H]⁺ 617.09271	-	3

56	15.29	dihydroxy-methoxyflavanone	[M - H]⁻ 285.07736	-	[M + H]⁺ 287.09177	167.03395bp, 147.04413, 119.04940	2
57	15.44	pinocebrin	[M - H]⁻ 255.06641	185.06021, 171.04408bp, 151.00281	[M + H]⁺ 257.08102	-	1
58	15.75	putative diterpene (C₂₀H₂₆O₃)	[M - H]⁻ 313.18106	-	[M + H]⁺ 315.19559	-	3
59	16.13	fluctuadin	-	-	[M + Na]⁺ 457.14679, [M + NH₄]⁺ 452.19144, [M + H]⁺ 435.16489	375.14365bp, 289.10696, 257.08075, 229.08589	2
60	16.14	putative diterpene (C₂₀H₃₀O₄)	[M - H]⁻ 333.20718	-	[M + H]⁺ 335.22122, [(M + H) - H₂O]⁺ 317.21103, [(M + H) - 2xH₂O]⁺ 299.20050	-	3
61	17.00	santin	[M - H]⁻ 343.08267	-	[M + H]⁺ 345.09720	330.07315, 315.04990bp, 287.05518, 259.06027	2
62	18.89	hydroxy-*ent*-kauren-18-oic acid isomer 4	[M - H]⁻ 317.21237	-	[M + H]⁺ 319.22617, [(M + H) - H₂O]⁺ 301.21579	-	2
63	20.12	polymatin B	-	-	[M + K]⁺ 471.14169, [M + Na]⁺ 455.16782, [M + NH₄]⁺ 450.21146, [M + H]⁺ 433.18588	373.16470, 291.12283, 273.11221bp, 241.08594	2
64	20.26	hydroxy-*ent*-kauren-18-oic acid isomer 3	[M - H]⁻ 317.21228	-	[M + H]⁺ 319.22647, [(M + H) - H₂O]⁺ 301.21600, [(M + H) - 2xH₂O]⁺ 283.20566	-	3
65	20.42	putative diterpene (C₁₈H₃₀O₃)	[M - H]⁻ 293.21274	-	[(M + H) - H₂O]⁺ 277.21658	-	3
66	20.83	hydroxy-*ent*-kauren-18-oic acid isomer 2	[M - H]⁻ 317.21268	-	[M + H]⁺ 319.22708, [(M + H) - H₂O]⁺ 301.21661	-	3
67	22.03	hydroxy-*ent*-kauren-18-oic acid isomer 1	[M - H]⁻ 317.21246	287.20209	[M + H]⁺ 319.22742	-	3
68	23.35	putative diterpene (C₂₀H₃₀O₂)	-	-	[M + H]⁺ 303.23252, [(M + H) - H₂O]⁺ 285.22144	-	3

#	RT	Compound	Negative		Positive		Level*
69	27.63	grandiflorenic acid	-	-	[M + H]$^+$ 301.21558, [(M + H) - H$_2$O]$^+$ 283.21558, [(M + H) - H$_2$O - CH$_2$O$_2$]$^+$ 255.21036	-	2
70	28.12	16-hydroxy kauran-19-al	-	-	[M + H]$^+$ 305.24707, [(M + H) - H$_2$O]$^+$ 287.23676, [(M + H) - 2xH$_2$O]$^+$ 269.22626	-	2
71	28.14	putative triterpene (C$_{29}$H$_{42}$O$_6$)	[M - H]$^-$ 485.28284	-	[M + H]$^+$ 487.29816	-	3
72	28.25	*ent*-kaurenoic acid	[M - H]$^-$ 301.21747	-	[M + H]$^+$ 303.23160	-	1
73	34.23	putative triterpene (C$_{29}$H$_{40}$O$_6$)	[M - H]$^-$ 485.28265	-	-	-	3
74	35.69	putative triterpene (C$_{29}$H$_{46}$O$_6$)	[M - H]$^-$ 491.34128	-	-	-	3

CAF: caffeoyl moiety; bp: base peak; ACN: acetonitrile adduct; QER: quercetin; FER: feruloyl moiety; QA: quinic acid.
* According to the Metabolomics Standard Initiative (SUMNER et al., 2007)

**Appendix 6.** NMR and mass spectra of the isolated metabolites
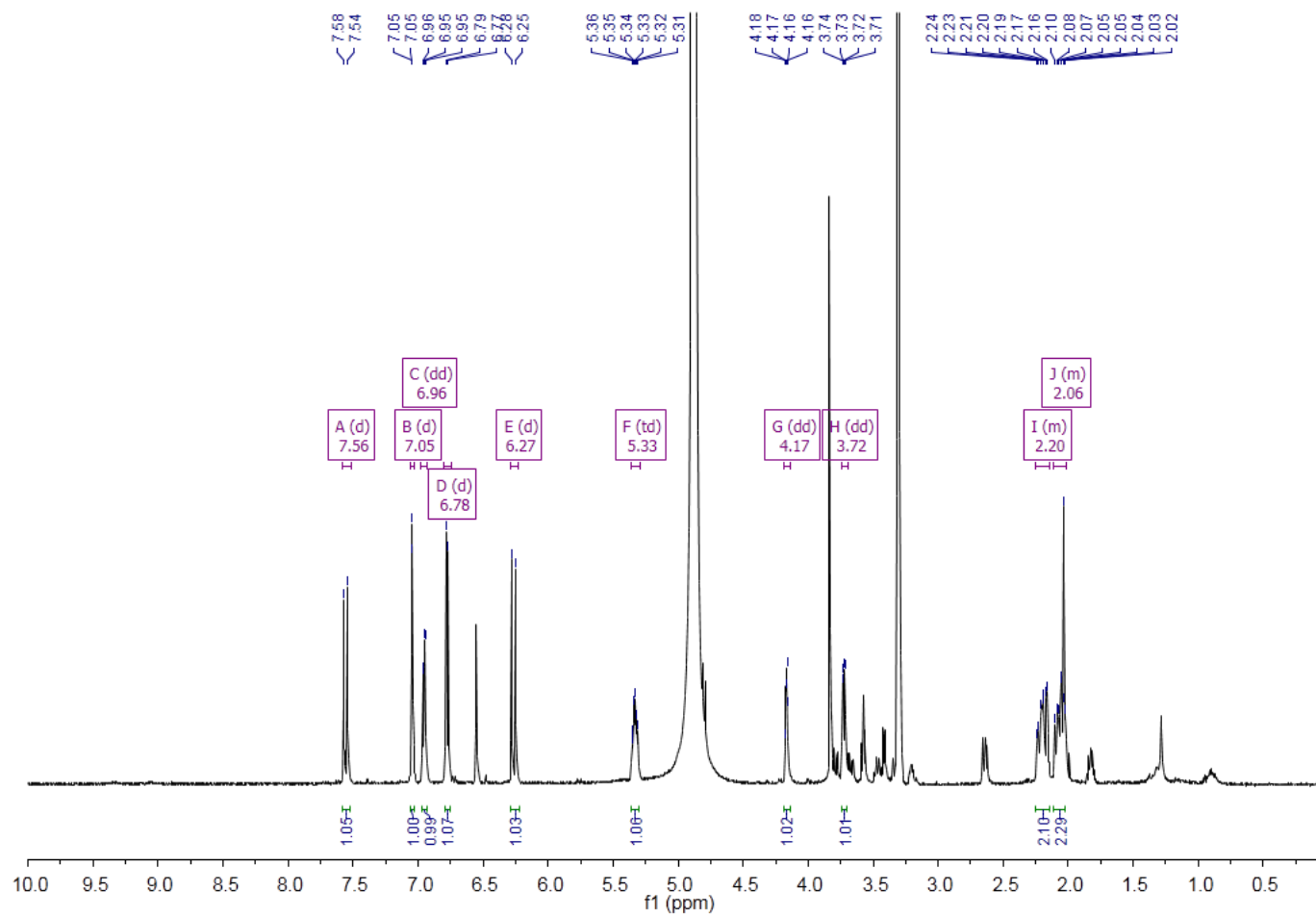
Figure A1. ^{1}H-NMR spectrum of 5-O-(E)-caffeoylquinic acid (500 MHz, CD$_3$OD).
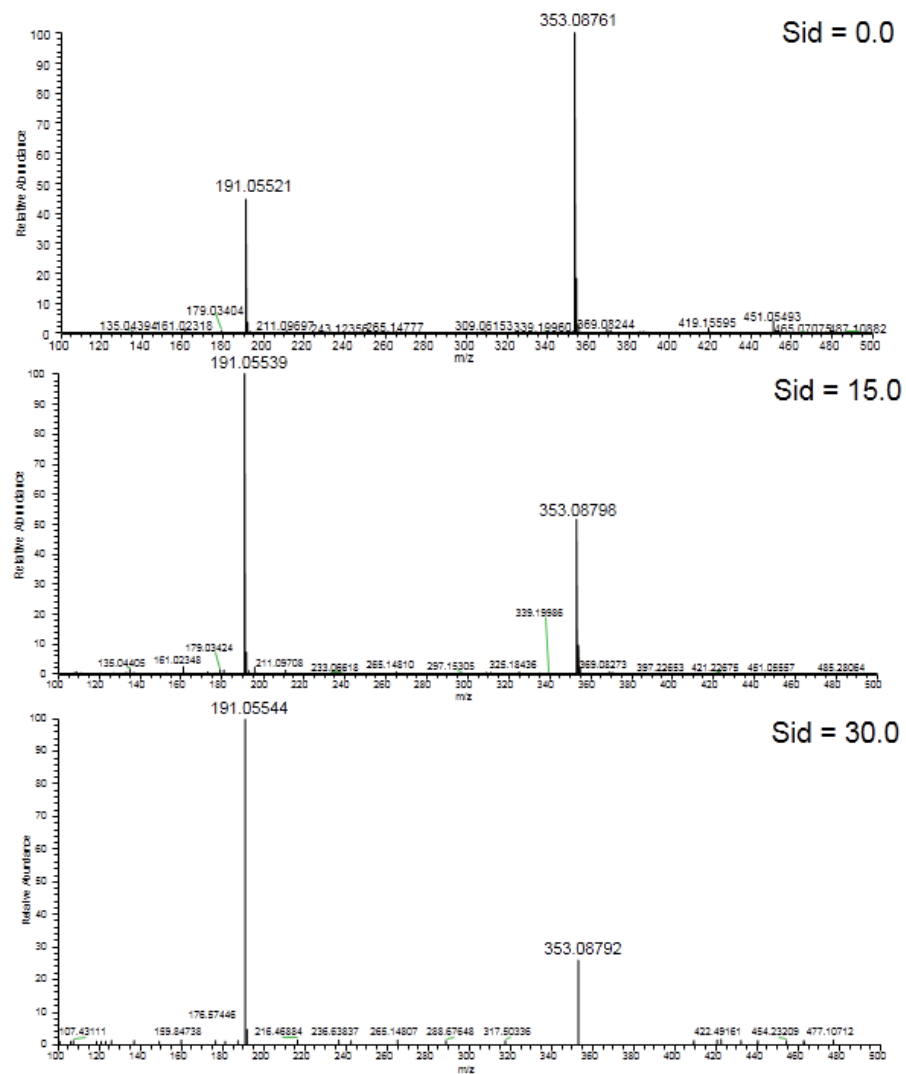
Figure A2. Negative mode HR-ESI-MS/MS spectra of 5-O-(E)-caffeoylquinic acid with different fragmentation energies.
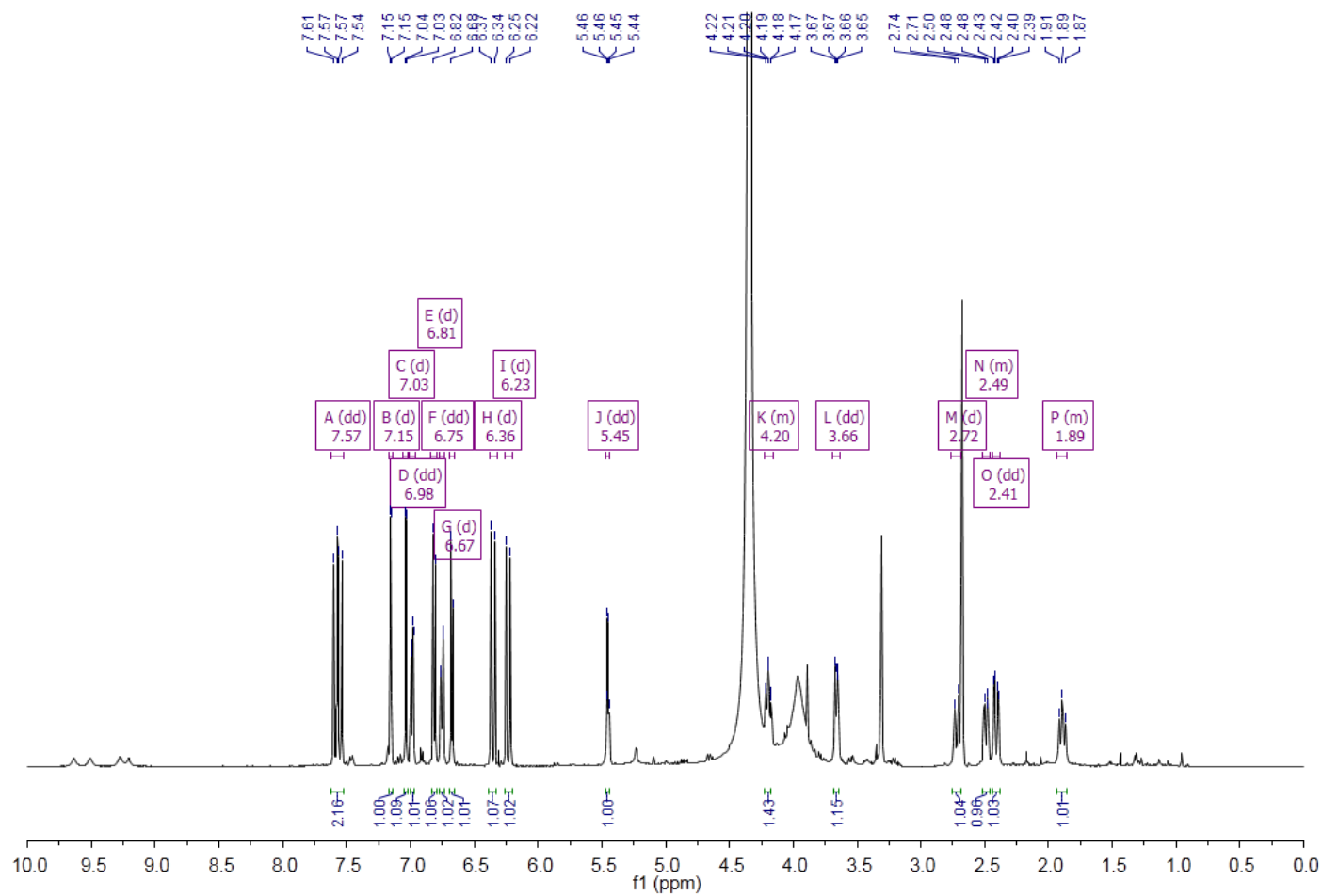
**Figure A3.** ^{1}H-NMR spectrum of 1,3-di-O-(E)-caffeoylquinic acid (500 MHz, CD$_3$OD).
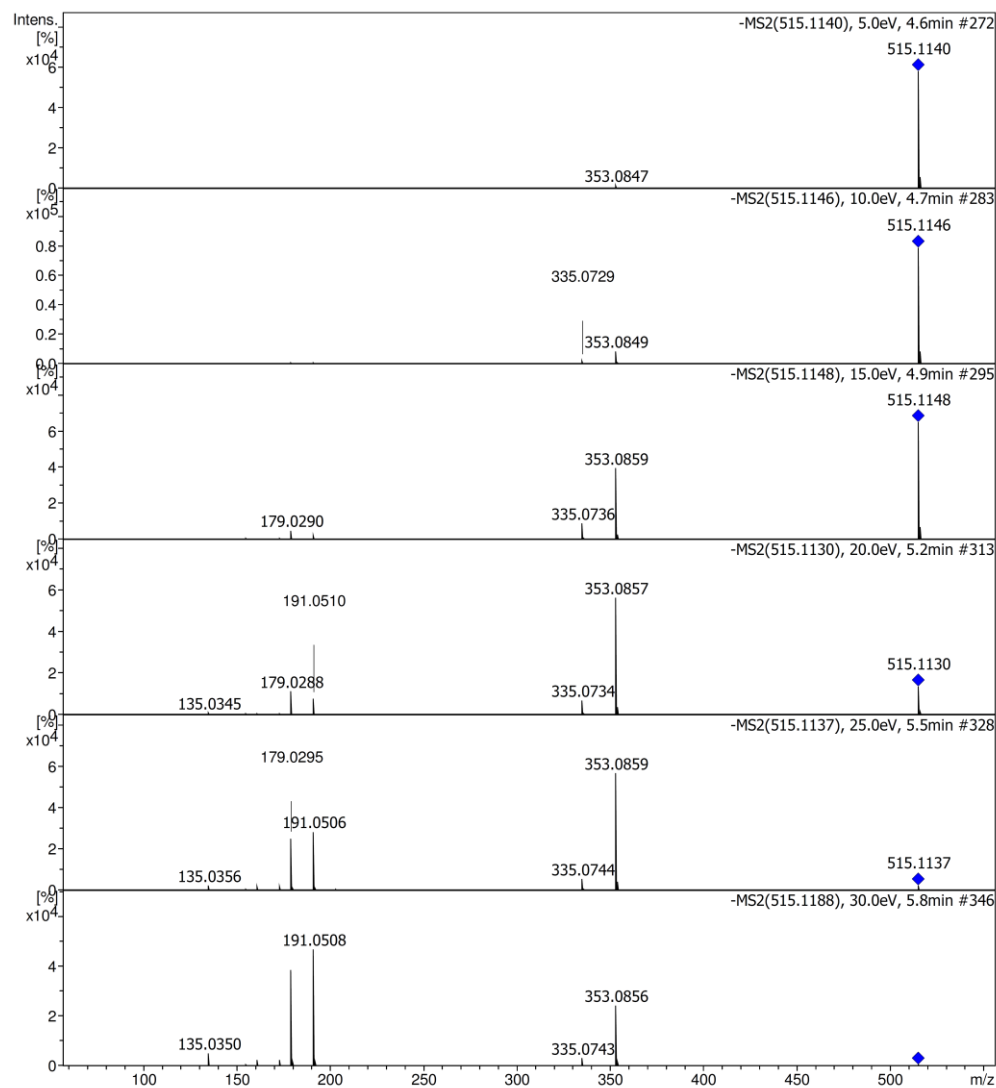
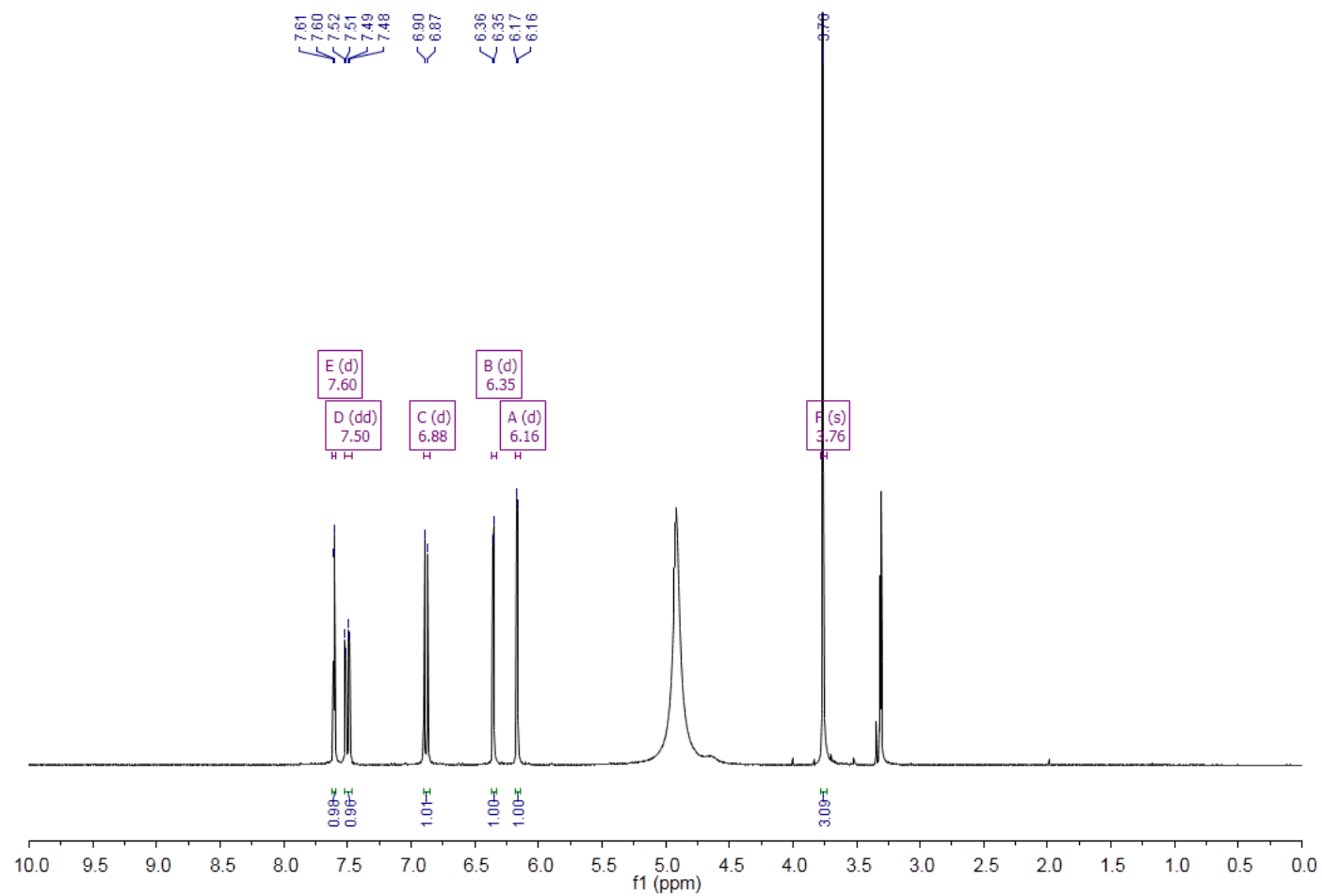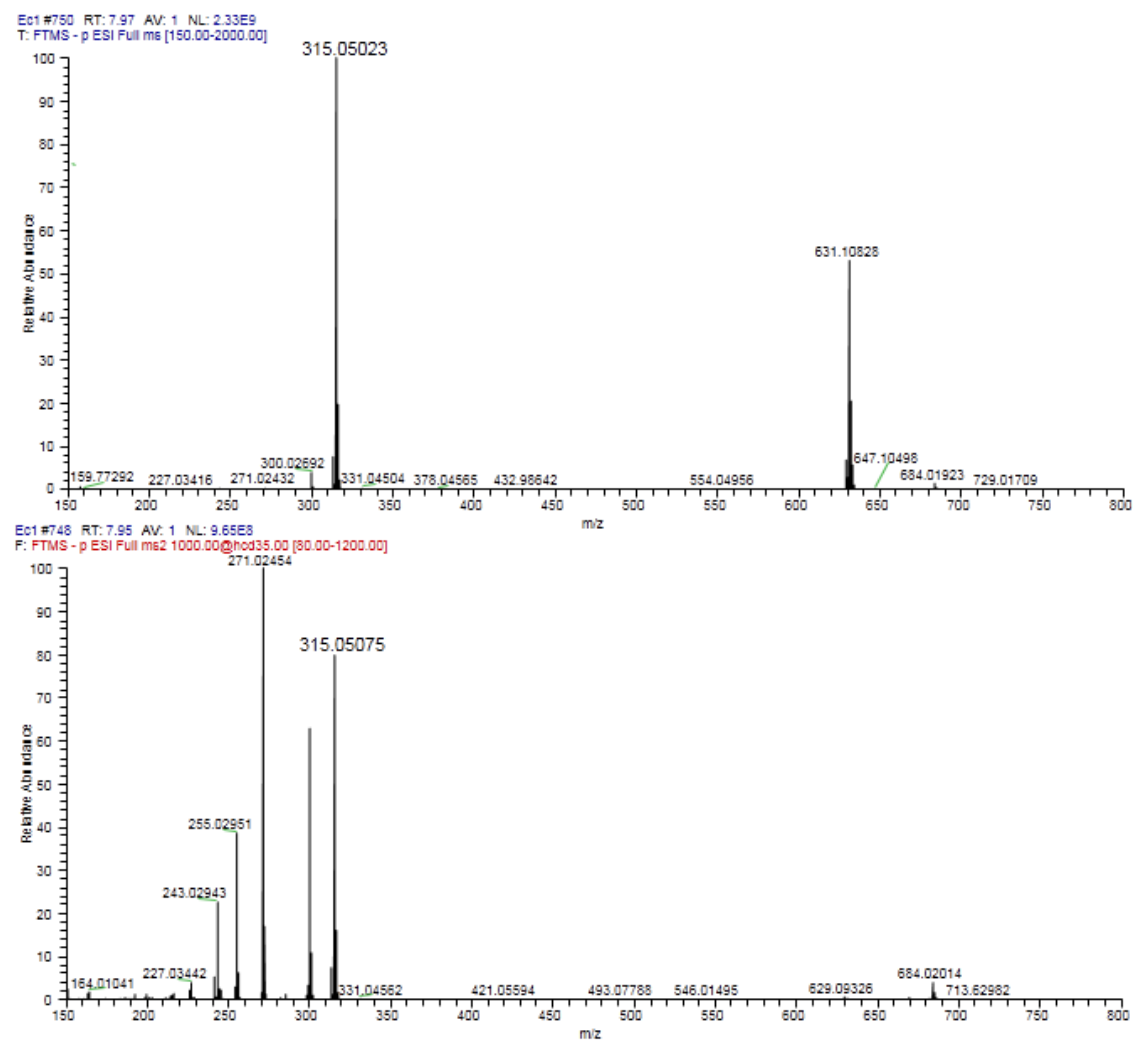**Figure A4.** Negative mode HR-ESI-MS/MS spectra of 1,3-di-*O*-(*E*)-caffeoylquinic acid with different fragmentation energies.
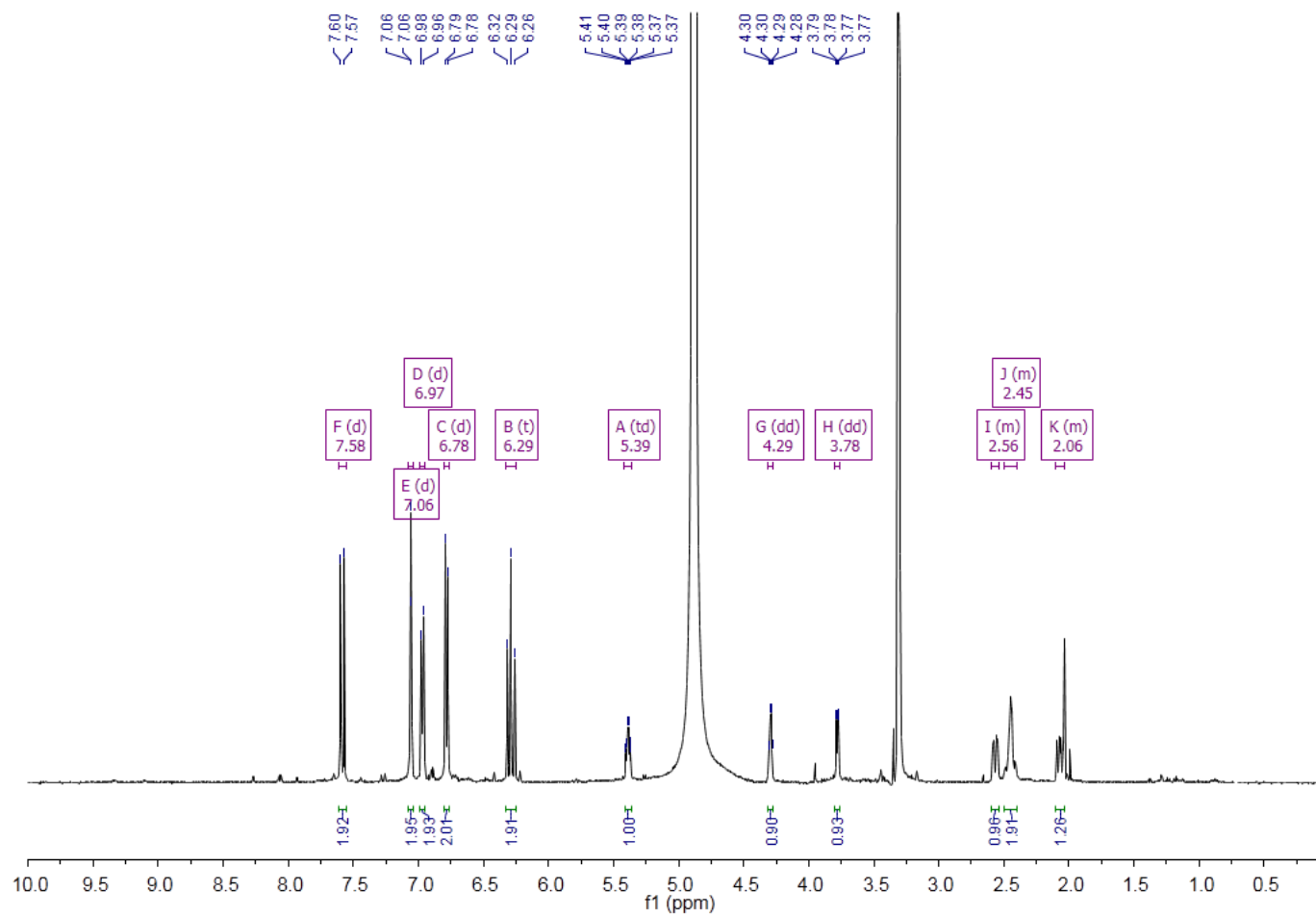
Figure A5. 1H-NMR spectrum of 1,5-di-*O*-(*E*)-caffeoylquinic acid (500 MHz, CD3OD).
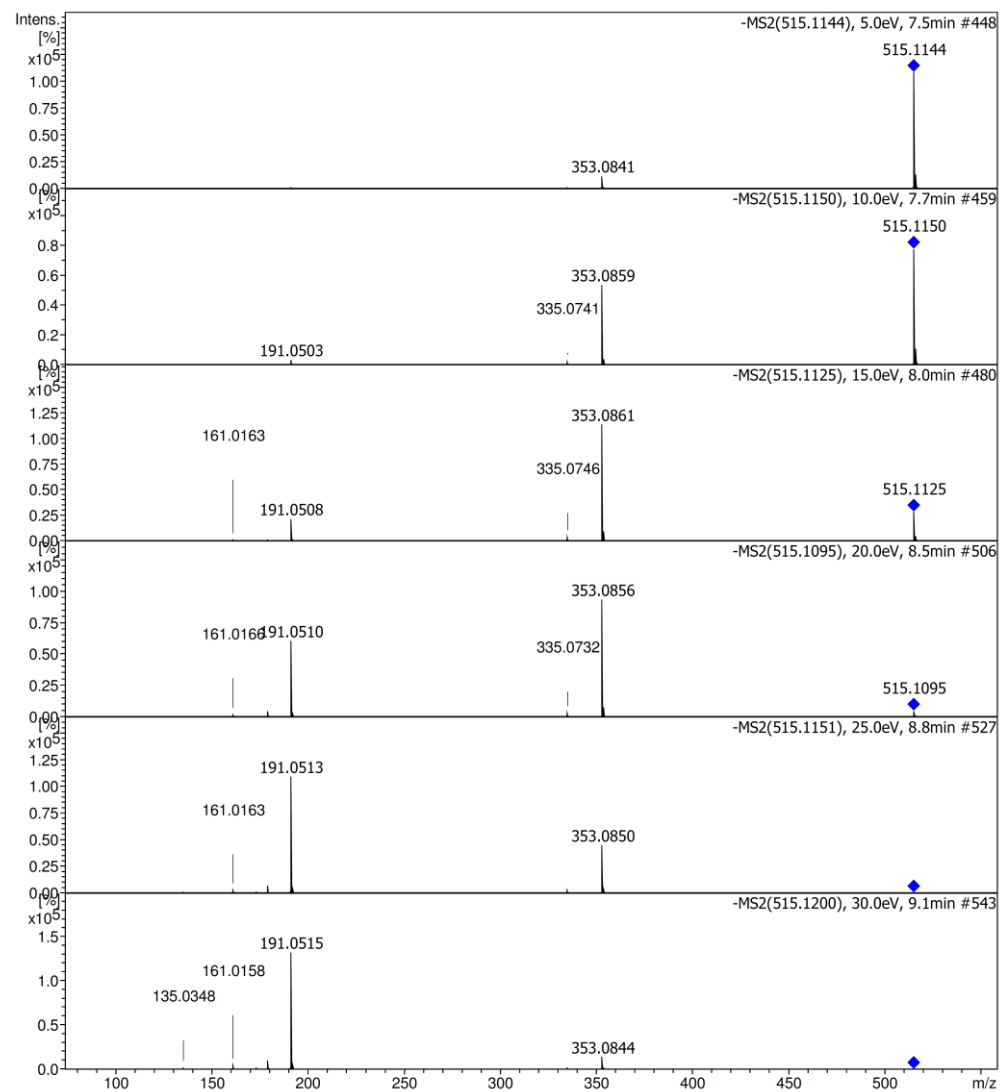
Figure A6. Negative mode HR-ESI-MS/MS spectra of 1,5-di-*O*-(*E*)-caffeoylquinic acid with different fragmentation energies.
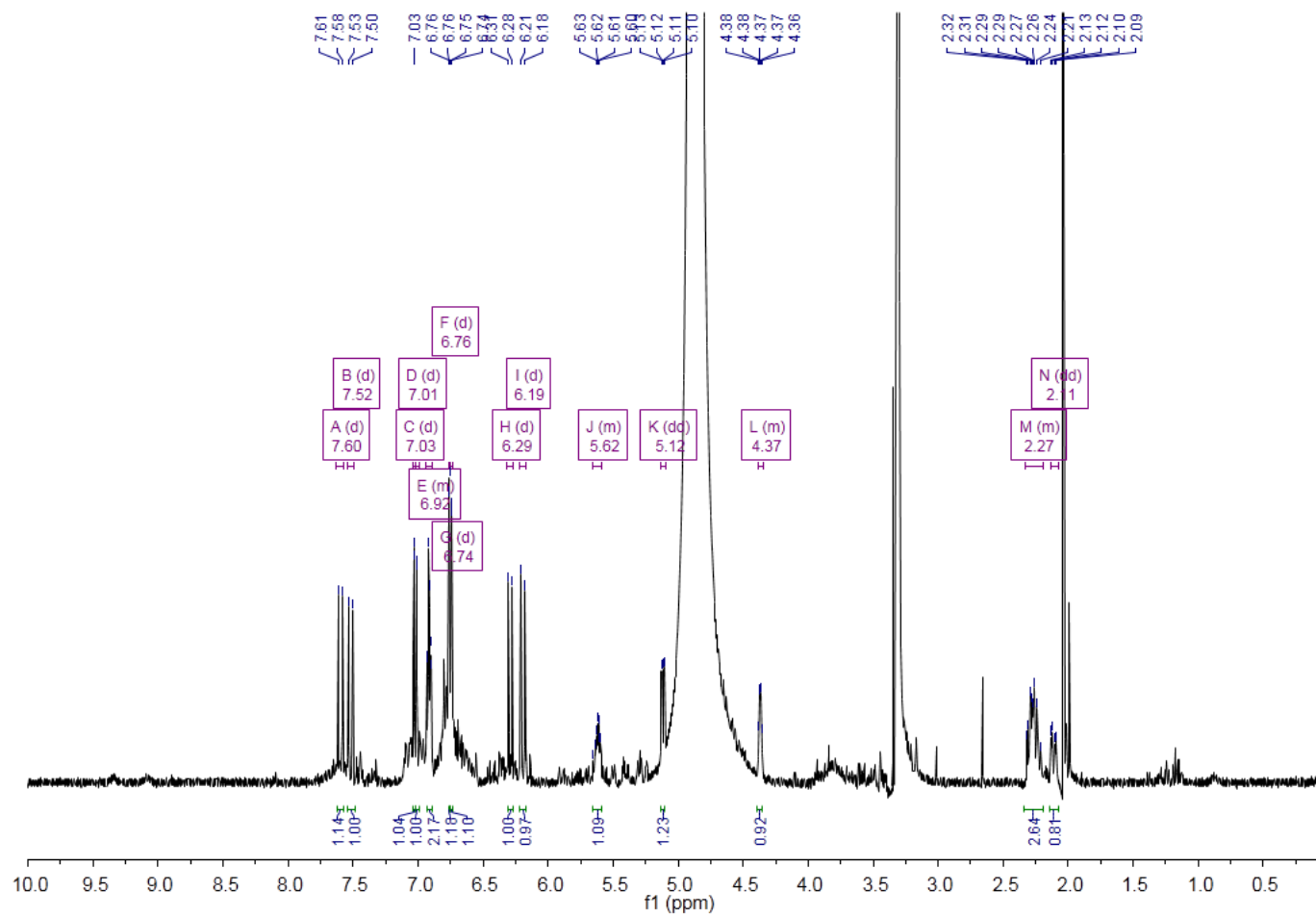
Figure A7. 1H-NMR spectrum of 3,4-di-*O*-(*E*)-caffeoylquinic acid (500 MHz, CD3OD).
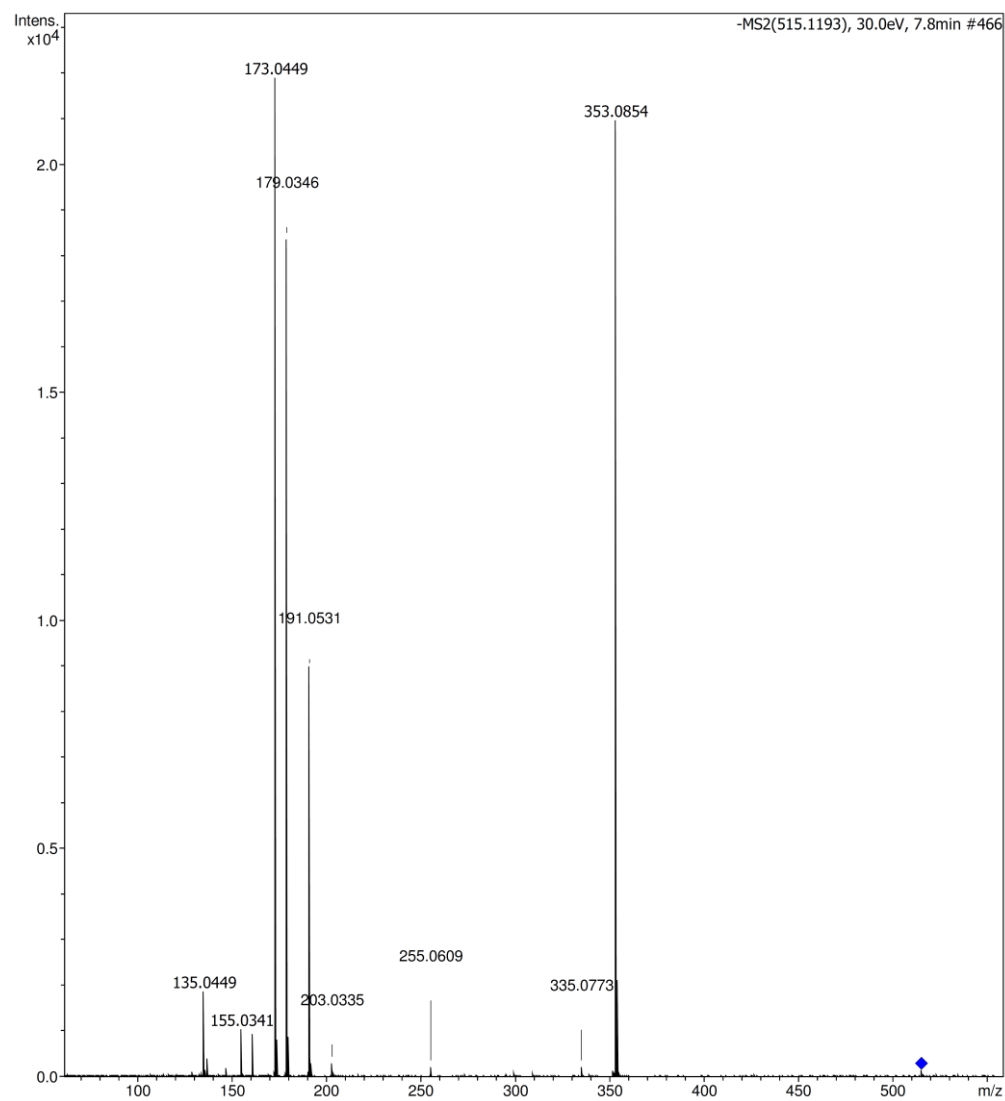
Figure A8. Negative mode HR-ESI-MS/MS spectrum in of 3,4-di-*O*-(*E*)-caffeoylquinic acid.
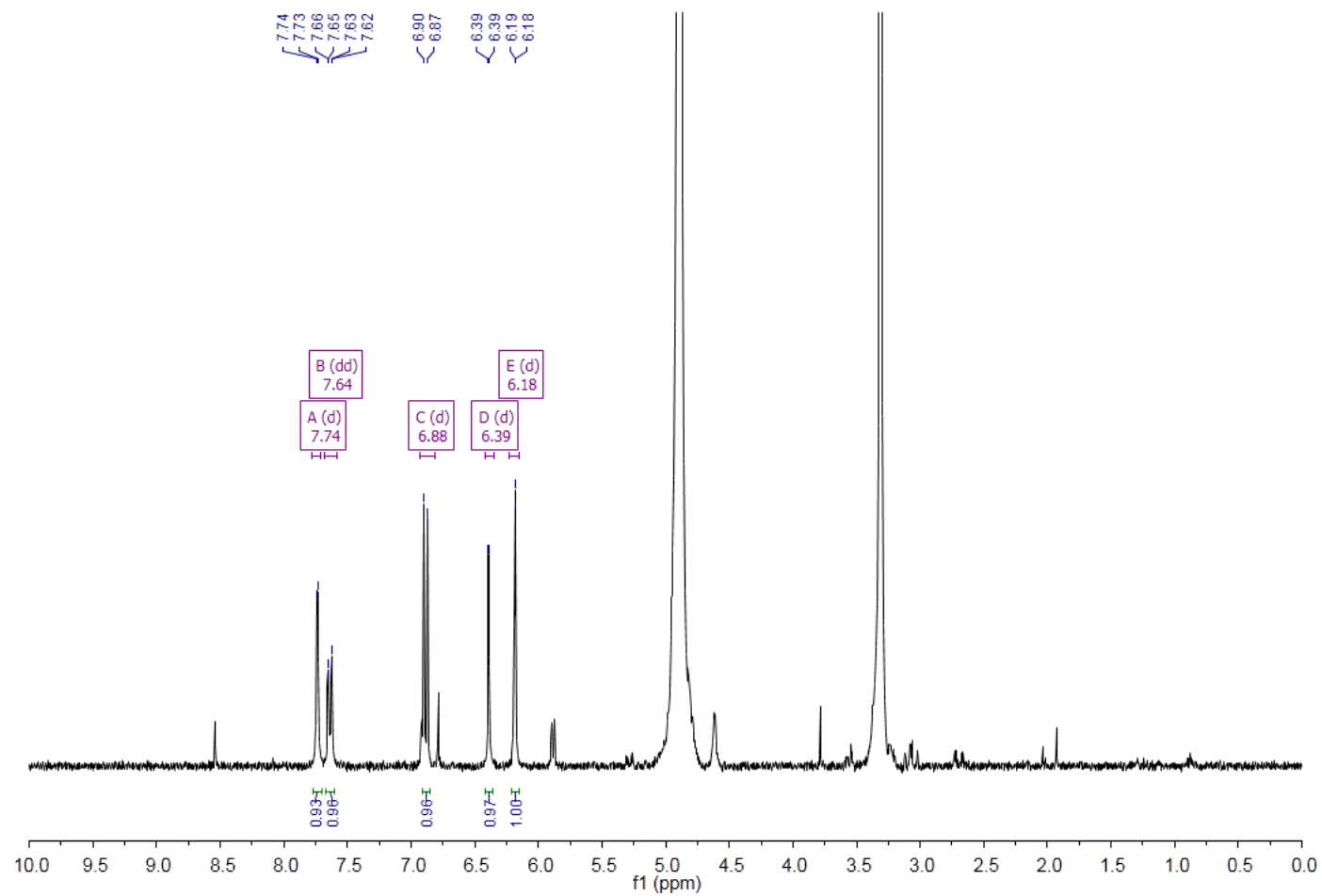
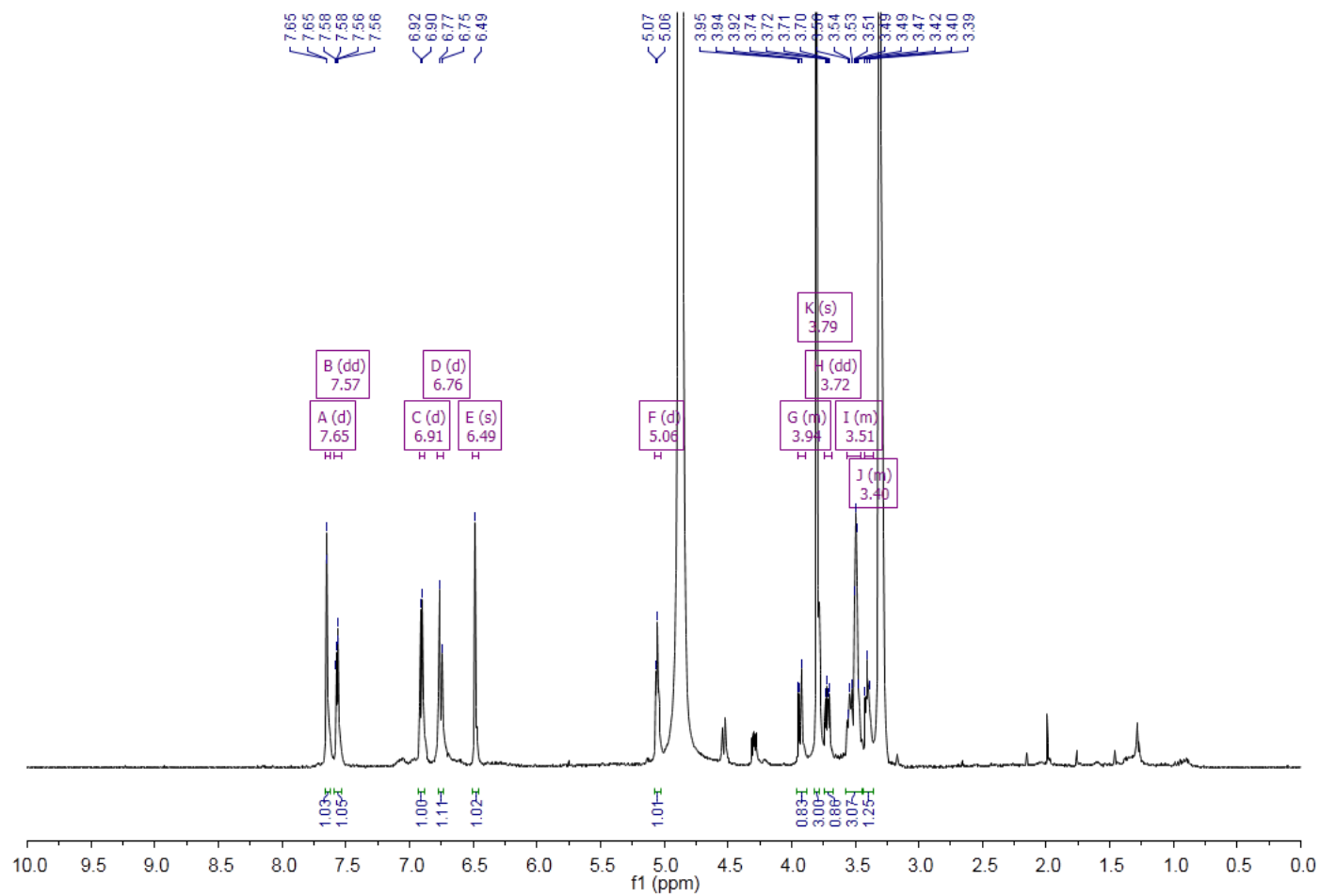**Figure A9.** 1H-NMR spectrum of quercetin (300 MHz, CD3OD).
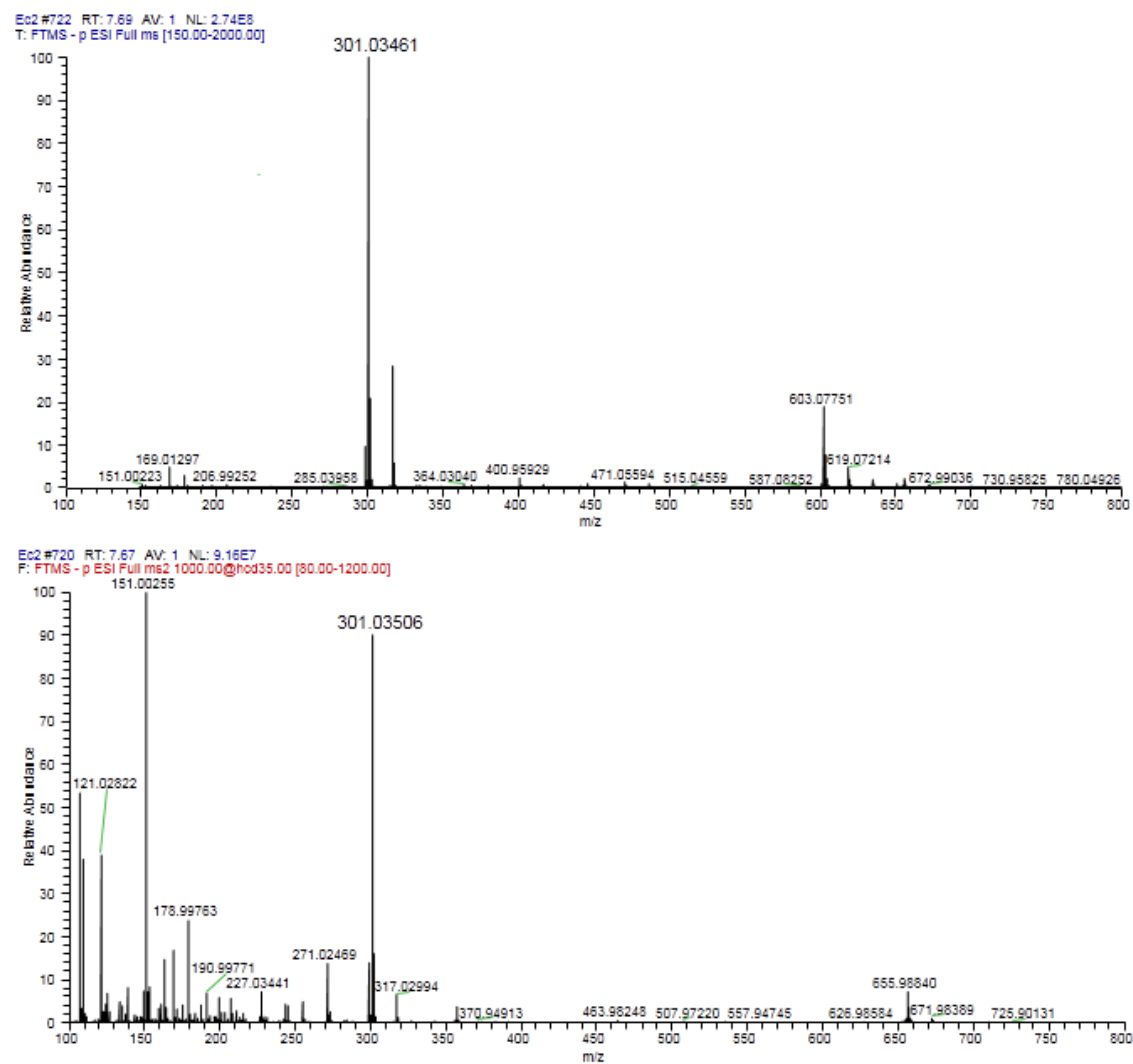
Figure A10. Negative mode HR-ESI-MS and MS/MS spectra of quercetin.

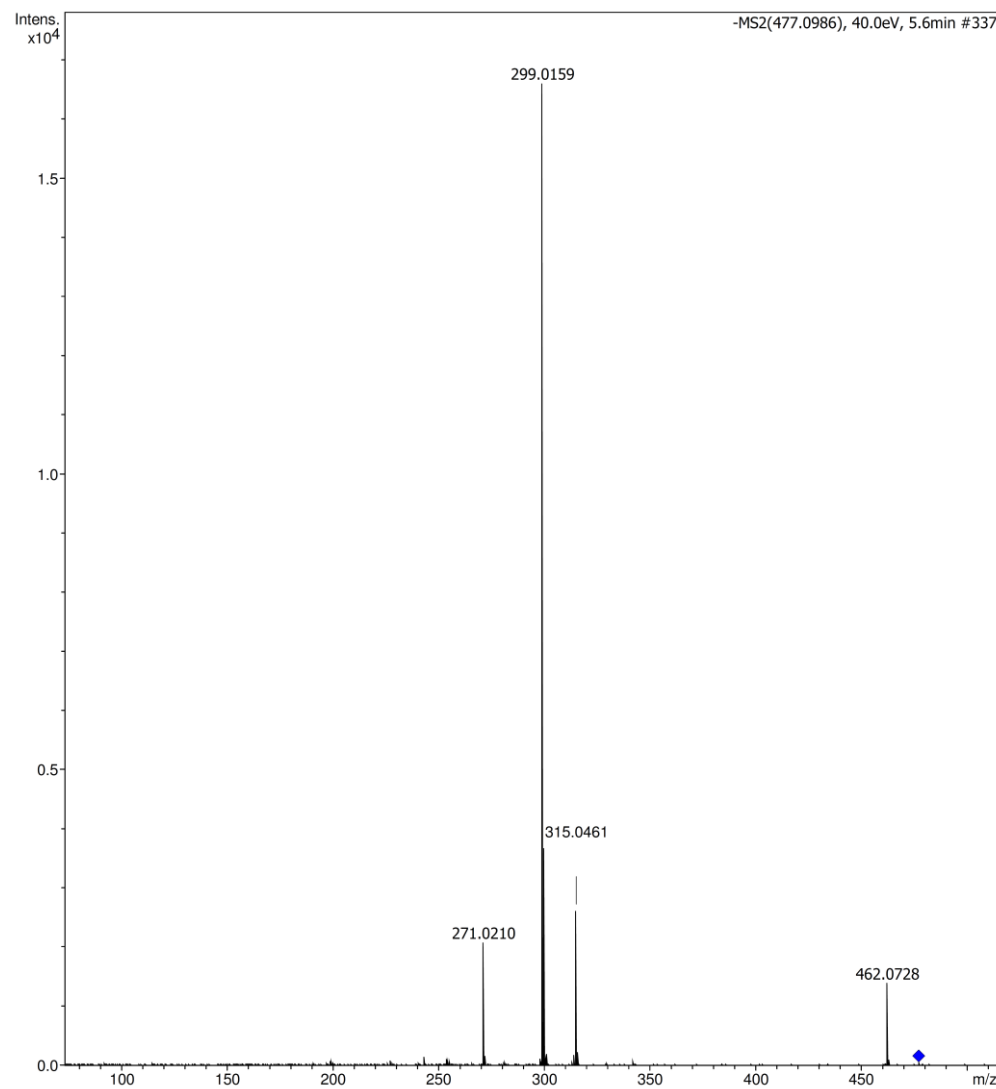**Figure A11.** 1H-NMR spectrum of 3-*O*-methylquercetin (300 MHz, CD3OD).

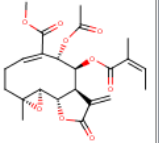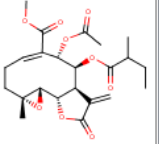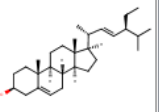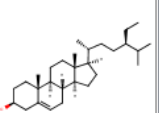**Figure A12.** Negative mode HR-ESI-MS and MS/MS spectra of 3-*O*-methylquercetin.
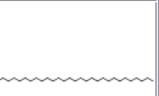
Figure A13. 1H-NMR spectrum of 3-*O*-methylquercetin 7-*O*-*β*-glucopyranoside (500 MHz, CD3OD).

Figure A14. Negative mode HR-ESI-MS/MS spectrum of 3-*O*-methylquercetin 7-*O*-*β*-glucopyranoside.

Appendix 7. Screenshot of the secondary metabolite database of the genus *Smallanthus*.

#	structure	Species	Authority	Plant ...	Country ...	State or ...	Locality of ...	GPS	Herbarium ...	Compound s...	Type of Ske...	Usual name	Sistematic nam...	IUPAC name	Molecular ...	CAS number	Monoisotop...
20		Smallanthus connatus	(Spreng.) H. Rob.	Aerial Parts	Argentina	Tucuman	Horco Molle - Yerba Buena	N.R.	Herbarium of the Instituto Miguel Lillo (LIL 607374)	Sesquiterpene lactones	Melampolide	longipilin acetate	Oxireno[9,10]cyclc 6-(acetyloxy)-1a,2,' methyl ester, (1aR	methyl (1S,2R,4R,7E,9S,10S,	C23H28O9	71135-27-8	448.17333
21		Smallanthus connatus	(Spreng.) H. Rob.	Aerial Parts	Argentina	Tucuman	Horco Molle - Yerba Buena	N.R.	Herbarium of the Instituto Miguel Lillo (LIL 607374)	Sesquiterpene lactones	Melampolide	N.R.	(4R*, 5S*, 6S*, 7R*, 9S*)-4?,5?-epoxy-8	methyl (1S,2S,4R,7E,9S,10S,	C23H30O9	000-00-13	450.18899
22		Smallanthus connatus	(Spreng.) H. Rob.	Aerial Parts	Argentina	Tucuman	Horco Molle - Yerba Buena	N.R.	Herbarium of the Instituto Miguel Lillo (LIL 607374)	Triterpenes	Phytosterol	stigmasterol	Stigmasta-5,22-die (3?,22E)-	(1S,2R,5S,10S,11S,14	C29H48O	83-48-7	412.37052
23		Smallanthus connatus	(Spreng.) H. Rob.	Aerial Parts	Argentina	Tucuman	Horco Molle - Yerba Buena	N.R.	Herbarium of the Instituto Miguel Lillo (LIL 607374)	Triterpenes	Phytosterol	?-sitosterol	Stigmast-5-en-3-ol; (3?)-	(1S,2R,5S,10S,11S,14	C29H50O	83-46-5	414.38617
24		Smallanthus connatus	(Spreng.) H. Rob.	Aerial Parts	Argentina	Tucuman	Horco Molle - Yerba Buena	N.R.	Herbarium of the Instituto Miguel Lillo (LIL 607374)	Alkane	Acyclic	nonacosane	nonacosane	nonacosane	C29H60	630-03-5	408.46950

Appendix 8. Detailed information about all dereplicated metabolites in yacón leaves.

	Rt (min)	Substance	Negative ionization pseudomolecular ion (*m/z*)	Negative MS/MS	Positive ionization pseudomolecular ion (*m/z*)	Positive MS/MS	Confidence level[a]
1	1.26	turanose	[M - H]⁻ 341.10880	341.10873, 179.05533, 89.02308bp	[M + Na]⁺ 365.10504	-	2
2	1.26	stachyose	[M - H]⁻ 665.21417	383.11972, 221.06633, 179.05528, 161.04459bp, 101.02312	-	-	2
3	1.27	raffinose	[M - H]⁻ 503.16156	221.06641, 179.05538bp, 101.02331, 89.02328	-	-	2
4	1.38	quinic acid*	[M - H]⁻ 191.05533	191.05519, 85.02828	-	-	1
5	1.38	altraric acid*	[M - H]⁻ 209.02991	209.02992, 191.01892, 133.01320, 85.02824bp	-	-	2
6	1.61	acetoxy glucuronic acid*	[M - H]⁻ 235.04559	131.03387, 115.00256, 113.02327, 103.00249, 85.02824, 75.00745bp	-	-	2
7	4.24	monocaffeoylaltraric acid isomer*	[M - H]⁻ 371.06192, [(M - H) - CAF]⁻ 209.02968	209.02983bp, 191.01920, 133.01303, 85.02824	-	-	3
8	4.57	monocaffeoylaltraric acid isomer*	[M - H]⁻ 371.06189, [(M - H) - CAF]⁻ 209.02963	209.02986bp, 191.01913, 133.01315, 85.02824	-	-	3
9	5.37	monocaffeoylaltraric acid isomer*	[M - H]⁻ 371.06186	209.02971bp, 191.01897, 133.01286, 85.02818	-	-	3
10	6.22	1-*O*-caffeoylquinic acid	[M - H]⁻ 353.08774	191.05542bp,	[M + H]⁺ 355.10239	-	2

11	6.68	5-O-caffeoylquinic acid	[M - H]⁻ 353.08810	179.03387 191.05544bp, 179.03423	[M + H]⁺ 355.10242	-	1
12	8.16	dicaffeoylaltraric acid isomer	[M - H]⁻ 533.09418	371.06192, 209.02957bp, 191.01901	-	-	3
13	8.19	rutin	[M - H]⁻ 609.14679	301.03534, 300.02771bp	[M + H]⁺ 611.16040, [(M + H) - rhamnose]⁺ 465.10248, [(M + H) - rhamnose - glucose]⁺ 303.04984bp	-	1
14	8.22	7-O-methoxy gossypetin 3-O-rutinoside*	[M - H]⁻ 639.15717	330.03818bp, 315.01489, 287.01962	[M + Na]+ 663.15265, [M + H]+ 641.17102, [(M + H) - rhamnose]+ 495.11288, [(M + H) - rhamnose - glucose]+ 333.06033	-	2
15	8.56	quercetin-3-O-galactoside	[M - H]⁻ 463.08841	301.03503, 300.02774bp, 271.02478	-	-	1
16	8.64	dicaffeoylaltraric acid isomer	[M - H]⁻ 533.09387	371.06226, 209.02979bp, 191.01903, 85.02818	-	-	3
17	8.67	pentahydroxy-methoxyflavone glycoside*	[M - H]⁻ 493.09854	-	[M + H]⁺ 495.11334	333.06027bp, 318.03699	2
18	9.09	quercetin-O-pentoside	[M - H]⁻ 433.07773	301.03516, 300.02771bp, 271.02484	-	-	2
19	9.12	dicaffeoylaltraric acid isomer	[M - H]⁻ 533.09418	371.06198, 209.02956bp, 191.01894	-	-	3
20	9.28	quercetin-O-pentoside	[M - H]⁻ 433.07797	301.03537, 300.02768bp, 271.02487	-	-	2
21	9.38	3,4-di-O-caffeoylquinic acid	[M - H]⁻ 515.11902	353.08786, 335.07745, 191.05548, 179.03421, 173.04469bp, 135.04407	-	-	1
22	9.41	dicaffeoylacetoxy glucuronic acid*	[M - H]⁻ 559.10895	397.07745, 235.04562,	-	-	2

No	RT	Compound	[M - H]⁻		[M + H]⁺		
				179.03410, 161.02345bp, 135.04408			
23	9.6	dicaffeoylaltraric acid isomer	[M - H]⁻ 533.09406	371.06189, 209.02953bp, 191.01889	-	-	3
24	9.86	3,5-di-O-caffeoylquinic acid	[M - H]⁻ 515.11969	353.08795, 191.05539bp, 179.03419, 135.04413	[M + H]⁺ 517.13416	-	1
25	10.26	4,5-di-O-caffeoylquinic acid	[M - H]⁻ 515.11964	353.08795, 191.05531, 179.03424, 173.04471bp	[M + H]⁺ 517.13416	-	1
26	10.93	tricaffeoylaltraric acid isomer	[M - H]⁻ 695.12561	533.09387, 371.06213, 209.02965bp, 191.01903	-	-	3
27	11.17	tricaffeoylaltraric acid isomer	[M - H]⁻ 695.12579	533.09503, 371.06235, 209.02988bp,191.01921	-	-	3
28	11.6	tricaffeoylaltraric acid isomer	[M - H]⁻ 695.12598	533.09351, 371.06186, 209.02977bp,191.01898	-	-	3
29	13.11	tricaffeoylaltraric acid isomer	[M - H]⁻ 695.12561	533.09442, 371.06198, 209.02977bp,191.01900	-	-	3
30	14.44	trihydroxy trimethoxyflavone*	[M - H]⁻ 359.07709	344.05362bp, 329.03030, 314.00677, 301.03537, 286.01212	-	-	2
31	15.65	enhydrin	-	-	[M + H + ACN]⁺ 506.20157, [M + Na]⁺ 487.15689, [M + NH₄]⁺ 482.20203, [M + H]⁺ 465.17508	405.15396, 349.12766, 289.10672bp, 229.08568	1
32	16	trihydroxy dimethoxyflavone	[M - H]⁻ 329.06665	314.04337bp, 299.01978, 271.02463	-	-	2

| No	RT	Name	[M-H]⁻ neg	neg fragments	pos adducts	pos fragments	
33	16.16	polymatin A	-	-	[M + Na]⁺ 413.15704, [M + H]⁺ 391.17490	373.16428, 291.12247, 273.11194bp, 241.08568, 213.09085	2
34	17.22	fluctuadin	-	-	[M + H + ACN]⁺ 480.22302, [M + Na]⁺ 457.14697, [M + NH₄]⁺ 452.19162, [M + H]⁺ 435.16513	375.14395bp, 289.10703, 257.08084, 229.08597	2
35	17.42	uvedalin	-	-	[M + H + ACN]⁺ 490.20709, [M + Na]⁺ 471.16260, [M + H]⁺ 449.18042	389.15930, 273.11215bp, 241.08583, 213.09096	1
36	17.56	8β-epoxyangeloyloxy-9α-ethoxy-14-oxo-acanthospermolide	-	-	[M + H + ACN]⁺ 450.24860, [M + Na]⁺ 427.17285, [M + NH₄]⁺ 422.21719, [M + H]⁺ 405.19092	359.14874, 289.14331, 243.10162bp, 215.10672	2
37	18.2	uvedalin aldehyde	-	-	[M + Na]⁺ 441.15182, [M + NH₄]⁺ 436.19650, [M + H]⁺ 419.16998	359.14874, 259.09647bp, 231.10149, 213.09094	2
38	18.8	(1Z,4E)-8β-angeloyloxy-germacra-1(10),4,11(13)-trien-6α,12-olide-14-oic acid	-	-	[M + H + ACN]⁺ 406.22223, [M + Na]⁺ 383.14636, [M + NH₄]⁺ 378.19101, [M + H]⁺ 361.16437	261.11215, 243.10167bp, 215.10677	2
39	18.94	longipilin acetate	-	-	[M + H + ACN]⁺ 494.23825, [M + K]⁺ 487.13635, [M + Na]⁺ 471.16260, [M + NH₄]⁺ 466.20728, [M + H]⁺ 449.18036	389.15952bp, 289.10687, 257.08066, 229.08586	1
40	19.15	polymatin B aldehyde	-	-	[M + H + ACN]⁺ 448.23279, [M + K]⁺ 441.13089, [M + Na]⁺ 425.15686, [M + NH₄]⁺ 420.20151, [M + H]⁺ 403.17502	371.32666, 343.15384, 243.10156bp, 215.10663	2
41	20.53	smaditerpenic acid C	[M - H]⁻ 349.23846	331.22824, 269.22809, 221.15411,	-	-	2

				113.02325, 69.03327			
42	22.17	polymatin B	-	-	[M + H + ACN]+ 478.24332, [M + K]+ 471.14151, [M + Na]+ 455.16763, [M + NH4]+ 450.21198, [M + H]+ 433.18585	373.16437, 291.12244, 273.11200bp, 241.08575, 213.09090	1
43	23.17	smaditerpenic acid E	[M - H]- 377.23328	59.01253bp	[M + Na]+ 401.22971	-	2
44	24.81	smaditerpenic acid F	[M - H]- 391.24905	59.01256bp	[M + Na]+ 415.24536	-	2

* New report for the species

CAF: caffeoyl moiety; bp: base peak; ACN: acetonitrile adduct.

a According to the Metabolomics Standard Initiative (SUMNER et al., 2007)

Appendix 9. Melt curves of all primers used in the RT-qPCR analyses.
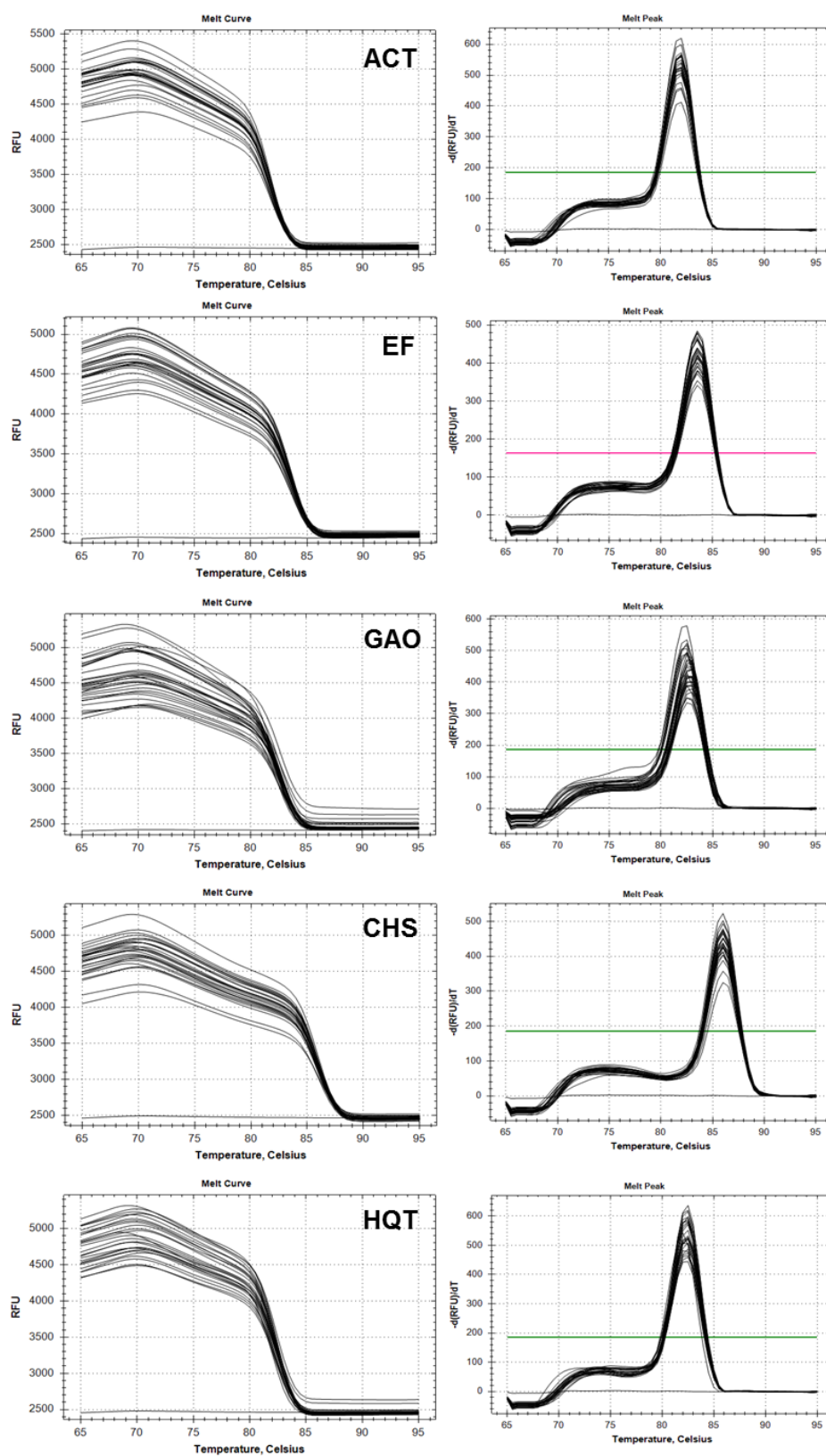
Appendix 10. q-PCR primers efficiency calculated using four serial dilutions of 1:5.